Top-Down Proteomic Analysis of Protein Pharmaceuticals, Mixtures of Protein Complexes, and Ribosomal Protein Extracts by Capillary Zone Electrophoresis – Mass Spectrometry

by Arseniy M. Belov

B.S. in Biochemistry, Western Washington University, Bellingham, WA

A dissertation submitted to

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

October 12th, 2017

Dissertation co-directed by

Dr. Jeffery Agar
Associate Professor of Chemistry and Chemical Biology

Dr. Alexander Ivanov
Associate Professor of Chemistry and Chemical Biology

and Dr. Barry L. Karger
James L. Waters Chair Professor *Emeritus* of Chemistry and Chemical Biology

ProQuest Number: 10690387

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.



ProQuest 10690387

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Acknowledgements

First and foremost, I would like to thank Professor Barry L. Karger for the guidance, support, education, and patience that he has given me for the duration of my time at the Barnett Institute of Chemical and Biological Analysis of Northeastern University. Beginning as a laboratory technician (November 2012 – August 2013) and then as a PhD student (September 2013 – July 2016), Professor Karger was my mentor and research advisor. Professor Karger's ability to stay up-to-date with techniques, instrumentation, and scientific literature has been a key attribute enabling his success over the many decades of his career. Moreover, his deep understanding of the fundamentals of separation science, his ability to integrate knowledge from a variety of fields and disciplines, and his awareness of the "big picture" are qualities that I hope to demonstrate for the rest of my career and life. Due to his *Emeritus* status as of July 2016, Professor Karger could not be my official advisor for the final year-and-a-half of my PhD and for my thesis defense. However, Professor Karger remained on my dissertation committee and co-directed my PhD defense.

I would like to sincerely thank Dr. Jeffrey Agar, who agreed to step in as my official graduate advisor for the period of July 2016 through my PhD defense (October 2017). Next, I would like to thank Dr. Alexander Ivanov for the mentorship, continued support, and training that he has provided me since I first came to Northeastern University in 2012. After obtaining tenure track status in August 2017, Dr. Ivanov was a member on my graduate committee and co-directed my PhD defense in October 2017 along with Dr. Agar and Professor Karger. During my five years at Northeastern University, Dr. Ivanov has consistently involved me in novel research projects, for which I am grateful. In turn, these projects have paved way for the development of my own research and career interests. Dr. Ivanov's adventurous attitude and deep understanding of analytical science, in particular mass spectrometry, are key attributes that have contributed to his success. I hope to retain these qualities in my future scientific endeavors as well. I would also like to thank Dr. Zhaohui "Sunny" Zhou for agreeing to participate as a member of my dissertation committee. I began learning from Dr. Zhou first in his protein pharmaceuticals course, which has also paved way for the development of my career interests.

I would like to thank Dr. David Bush and Dr. Shujia "Daniel" Dai for the training that they provided me and for the many enlightening discussions that we had. Dr. David Bush was instrumental in the research presented in Chapter 2. I extend my gratitude to the other members of Professor Karger's group: Dr. Simion Kreimer, Dr. Yuanwei "Abby" Gao, Somak Ray, Dr. Siyuan "Serah" Liu, Dr. Xianzhe "Jason" Wang, Dr. Siyang "Peter" Li, Dr. Vennela Mullangi, and Dr. Krishan Kumar. I would also like to sincerely thank all the friends and colleagues at Northeastern University and the Greater Boston Area who helped and encouraged me over a period of 5 years.

Lastly, I would like to extend a special gratitude to our collaborators, who have constantly offered support in the form of materials and supplies, as well as intellectual

contributions in the form of technical expertise and research guidance. These include Dr. Marcia Santos, Ted Amazeen, Chitra Ratnayake, Mark Flocco (SCIEX), Dr. Li Zang (Biogen), Dr. Rosa Viner (Thermo Fisher Scientific), and Dr. St. John Skilton and Dr. Marshall Bern (Protein Metrics).

Abstract of Dissertation

Mass spectrometry has emerged as a high accuracy, high sensitivity approach for the analysis of proteins. A conventional scheme for protein analysis by mass spectrometry involves reduction, alkylation, proteolytic digestion, followed by analysis either by electrospray ionization – mass spectrometry (ESI-MS) directly or with a prior separation by liquid chromatography (LC). This approach, termed bottom-up proteomics, offers the advantage of high sensitivity and allows for the analysis of complex mixtures of tens, hundred, and even thousands of proteins. However, it is well-known that proteins arising from the same gene may be chemically-different species as a result of sequence truncation, alternative splicing variants, and/or different post-translation modifications (PTMs). Such species are defined as proteoforms. A fundamental problem with the bottom-up approach is thus the inability to identify proteoforms, which may even possess multiple PTMs at once. In the course of proteolytic digestion, as a protein is converted into a mixture of peptides, associations between different PTMs are lost. In essence, what was once a single proteoform composed of numerous PTMs has become a set of smaller peptides scrambled within a mixture of other peptides from other proteoforms. Moreover, peptide fragments with bound PTMs may not contain basic amino acid residues and would not exhibit protonation to be detected by MS, resulting in losses and underrepresentation of such PTMs. An alternative approach to protein analysis by mass spectrometry is therefore top-down proteomics, in which proteins are analyzed in their intact forms. A further extension of top-down proteomics – native mass spectrometry (native MS) – involves the analysis of proteins at near-physiological conditions, where all-levels of protein structure, including any non-covalent associations, are maintained. The ability to perform top-down proteomics, and especially native MS, is a consequence of the numerous advancements in analytical instrumentation that have occurred over the past several decades.

However, the complexity of protein samples, especially for top-down proteomic applications, necessitates a high resolution separation component prior to characterization by ESI-MS, preferably in an online format. Conventional techniques include LC and ion mobility spectrometry. Reversed phase LC in particular, which separates analytes according to differences in hydrophobicity, is perhaps the most widely-used mode of LC for the separation of proteins online with ESI-MS. An alternative approach, however, involves capillary zone electrophoresis (CZE). CZE enables analyte separation according to charge and hydrodynamic volume differences in the liquid phase. Essentially, CZE may be viewed as a complementary technique that separates analytes according to a different mechanism than LC. Advancements in CZE instrumentation, capillary coatings, and interfaces with ESI-MS that have occurred over the past several decades have enabled the application of this technique for top-down proteomics.

In Chapter 2, a relatively small protein pharmaceutical possessing a single glycosylation site – recombinant human interferon-β1 (rhINF-β1) (Avonex, Biogen) – was analyzed under denaturing conditions by top-down CZE-ESI-MS. A positively-charged, cross-linked

polyethylenimine (cPEI) coating was administered and used on a commercial sheathless system (CESI, SCIEX) coupled to the Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). The analyses resulted in the identification of 138 individual proteoforms of rhINF-β1, of which 55 were quantitated. Charge species due to sialylation and deamidation were separated by CZE. Moreover, isobaric species due to positional isomers of a sialic acid on biantennary glycans were resolved by CZE. Fragmentation by high energy collision-induced dissociation (HCD) and electron transfer dissociation (ETD) contributed to the identification of many proteoforms. Reduced trapping pressure in the HCD cell of the Orbitrap Elite mass spectrometer enabled for accurate, highly sensitive detection of intact proteins. Quantitative intact analysis of rhINF-β1 revealed a potential correlation between methionine-truncation and deamidation, as well as between deamidation and glycosylation structures. This chapter is based on a 2016 publication in *Analytical Chemistry* [1].

In Chapter 3, analysis by sheathless CZE-ESI-MS was extended towards the characterization of a therapeutic monoclonal antibody (mAb). The cPEI coating, however, did not result in sufficient resolution of proteoforms on the mAb either in an intact form or after limited digestion (middle-down). Consequently, an alternative positively-charged coating (1-(4-iodobutyl) 4-aza-1-azoniabicyclo[2,2,2] octane iodide (M7C4I)), enabling a lower reverse electroosmotic flow (EOF) for higher resolution separations, was used. In middle-down analyses, charge-based proteoforms of the mAb were separated due to deamidation. Partial separation was achieved between species differing in glycosylation as a result of hydrodynamic volume differences. Tandem mass spectrometry with HCD fragmentation was used to confirm identities of many of the proteoforms. Analysis of the intact molecule revealed mAb populations differing in the extent of glycosylation, with two, one, or zero glycosylation sites per mAb. These populations were resolved baseline by CZE as a result of differences in hydrodynamic volume. Moreover, trace quantities of dissociated free chains of the mAb were detected and resolved by CZE. Quantitation of various populations of the intact mAb was achieved and corroborated between several experimental replicates. These results were further validated in analysis of the mAb performed under native conditions on a polyacrylamide-coated capillary coupled to the Q Exactive Plus MS *via* the CESI interface.

In Chapter 4, the sheathless CZE-ESI-MS platform was used for the analysis of numerous proteins under native conditions in one of the first demonstrations of CZE coupled online to native MS. Bare-fused silica (BFS) and polyacrylamide-coated capillaries, available commercially from SCIEX, were coupled to Orbitrap mass spectrometers in this application. First, a simple mixture of 3 proteins and associated homodimeric/homotetrameric protein complexes was analyzed under native conditions on a BFS capillary coupled to the Exactive EMR mass spectrometer (Thermo Fisher Scientific). The use of a near-physiological buffer allowed for the mitigation of non-specific adsorption of proteins to the capillary surface. In several instances, protein complexes were partially-resolved from their corresponding monomers by CZE. The use of all-ion-fragmentation allowed for the characterization of mixture

components, including prominent proteoforms for each protein/protein complex. A more complex mixture of over 11 proteins was then analyzed, this time on a polyacrylamide-coated capillary coupled to the Exactive EMR mass spectrometer. Utility of an uncharged coating suppressing EOF enabled further mitigation of non-specific adhesion of proteins to the capillary surface. All major mixture components were identified, and in several cases, separations of charge-based proteoforms from unmodified variants, as well as partial separations of protein complexes from their corresponding monomers were observed. Trastuzumab was then analyzed under native conditions on the same system, resulting in the identification of major glycosylation states of the mAb, including hemiglycosylated species. Moreover, Trastuzumab dimers were identified at approximate abundances of less than 3%. Native analysis revealed partial dissociation of individual chains of the mAb, suggesting that a trace quantity of light chains are bound non-covalently instead of *via* disulfide bonds; these forms were separated from intact mAb structures in the CZE domain. Finally, the protein extract from *E. coli* ribosomes was analyzed under native conditions by sheathless CZE-MS using the polyacrylamide-coated capillary coupled to the Q Exactive Plus Orbitrap MS (Thermo Fisher Scientific). As a result, over 90% of the protein groups from *E. coli* ribosomes were identified in a combinatorial analysis involving native and denaturing conditions. A single analysis under native conditions resulted in 42 identified protein groups, including some 144 proteoforms. Several potential non-covalent complexes between ribosomal proteins were identified by the native approach. Native CZE-MS was demonstrated as a viable technique for the analysis of single protein isolates, such as mAbs, simple mixtures of proteins and protein complexes, and for the protein isolate from an entire cellular organelle – the ribosome. This chapter is based on a 2017 publication in the *Journal of the American Society for Mass Spectrometry* [2].

CZE-MS offers a complementary approach for the analysis of a variety of protein analytes. In the first investigation, we started with a single protein of 22 kDa with a single glycosylation site. The second investigation involved a more complex protein – a therapeutic mAb candidate – of ~150 kDA and up to two glycosylation sites. Considerable optimization in sample preparation, as well as capillary coatings and analytical conditions was required for the characterization of this analyte. In the third investigation, sample complexity was extended further towards the analysis of protein mixtures and a cellular organelle. This required further method development at all stages of the analytical workflow, beginning with sample preparation, separation and detection parameters, instrumentation, and data analysis strategies. In summary, we have sought to use online CZE-MS in a top-down proteomic format to demonstrate the applicability of this technique for the analysis of a variety of samples of different complexities. Our efforts have resulted in characterization of numerous protein analytes, revealing information concerning these proteins that could not be obtained *via* more conventional bottom-up proteomic approaches. Lastly, we have demonstrated one of the first examples of the application of online CZE-MS under native conditions.

Table of Contents

Acknowledgements 2

Abstract of Dissertation 4

List of Abbreviations 11

Chapter 1: Introduction to Top-Down Proteomics and Capillary Zone Electrophoresis – Mass Spectrometry 12

1.0. Abstract 13

1.1. Heterogeneity of Proteins 14

1.2. Proteins as Pharmaceuticals 14

1.3. Analytical Techniques for Protein Characterization 16

1.4. Overview of Mass Spectrometry 18

1.5. Introduction to Analytical Separations 20

1.5.1. Liquid Chromatography 20

1.5.2. Capillary Zone Electrophoresis 21

1.5.3. Coupling Separations Online to ESI-MS 23

1.5.4. Coupling of Capillary Zone Electrophoresis to Mass Spectrometry *via* Sheath-Flow Junctions 24

1.5.5. Coupling of Capillary Zone Electrophoresis to Mass Spectrometry *via* Sheathless Junctions 25

1.5.6. Capillary Zone Electrophoresis in Microfluidic Chips 27

1.6. Proteomic Analysis by Mass Spectrometry 28

1.6.1. Bottom-up Proteomics 28

1.6.2. Top-Down Proteomics 30

1.6.3. MS Instrumentation for Top-Down Proteomics 32

1.6.4. Techniques for the Fragmentation of Intact Proteins 35

1.7. Separations of Intact Proteins for Top-Down Proteomics 39

1.7.1. Conventional Techniques and Liquid Chromatography 39

1.7.2. Ion Mobility Spectrometry 39

1.7.3. Intact Protein Separations by Capillary Zone Electrophoresis 40

1.7.4. Preconcentration Techniques for Intact Proteins in Capillary Zone Electrophoresis 41

1.8. Data Acquisition and Analysis Strategies for Top-Down Proteomics 43

1.9. Analytical Instrumentation 45

1.9.1. Q Exactive Plus MS 46

1.9.2. Exactive EMR MS 48
1.9.3. LTQ Orbitrap Elite MS 49
1.9.4. Sheathless CESI 50
1.10. Conclusions 52
Chapter 2: Ultra-High Resolution CZE-MS for Quantitative Characterization of Intact Biopharmaceutical Proteins: Proteoforms of Interferon-*β*1 53
2.0. Abstract 54
2.1. Introduction 54
2.2. Experimental 56
2.2.1. Chemicals and Materials 56
2.2.2. Buffer exchange and TCEP reduction of rhIFN-β1 56
2.2.3. Enzymatic removal of sialic acid and galactose 56
2.2.4. PEI coating and cross-linking procedure 57
2.2.5. CZE-MS 57
2.2.6. Data Processing 58
2.3. Results and Discussion 58
2.3.1. CZE-MS Characterization 59
2.3.2. CZE-MS of RhIFN-*β*1 67
2.3.3. CZE-MS Separations of Exoglycosidase-Treated RhIFN-*β*1 74
2.3.4. Relative Quantitation of RhIFN-*β*1 Proteoforms 80
2.3.5. Correlation of Protein Modifications 85
2.3.6. CZE-MS Analysis of Reduced rhIFN-*β*1 88
2.3.7. Top-Down CZE-MS^2 of the Des-1 RhIFN-*β*1 Modification 89
2.4. Conclusions 93
2.5. Acknowledgments 94
Chapter 3: Complementary Middle-Down and Intact Monoclonal Antibody Proteoform Characterization by Capillary Zone Electrophoresis – Mass Spectrometry 95
3.0 Abstract 96
3.1 Introduction 96
3.2. Experimental 98
3.2.1 Materials and Chemicals 98
3.2.2. Sample Preparation 99

3.2.3. CZE Methods 99
3.2.4. MS Instrumentation 100
3.2.5. Data Analysis 101
3.3 Results and Discussion 101
3.3.1. Fundamentals of CZE Operation Using Positively-Charged Capillary Coatings 101
3.3.2. Middle-Down Analysis: Products of Proteolytic Digestion and Reduction 102
3.3.3. Middle-Down Analysis: CZE-MS Separation Parameters 103
3.3.4. Middle-Down Analysis: Analysis of *IdeS*-Digested MAb 105
3.3.5. Middle-Down Analysis: Analysis of *IdeS*-Digested, Reduced MAb 111
3.3.6. Intact CZE-MS Characterization of the MAb 120
3.4. Conclusions 129
3.5. Acknowledgements 129
Chapter 4: Analysis of Proteins, Protein Complexes, and Organellar Proteomes Using Sheathless Capillary Zone Electrophoresis – Native Mass Spectrometry 130
4.0. Abstract 131
4.1. Introduction 131
4.2. Experimental 135
4.2.1. Chemicals and Materials 135
4.2.2. Preparation of the High-Purity Trastuzumab Standard 135
4.2.3. Preparation of Mixtures of Proteins and Protein Complexes 135
4.2.4. Preparation of the Ribosomal Protein Extract 136
4.2.5. Capillary Zone Electrophoresis Instrumentation and Methods 136
4.2.6. Mass Spectrometry 137
4.2.7. Data Analysis 138
4.2.8. Bottom-Up NanoLC-MS-Based Proteomic Analysis 138
4.3. Results and Discussion 139
4.3.1. Analytical Workflow 139
4.3.2. Native CZE-MS of Mixtures of Complex-Forming Proteins 140
4.3.3. Native CZE-MS of a Monoclonal Antibody 185
4.3.4. Native CZE-MS of the *E. coli* Ribosomal Extract 189
4.4. Conclusions 275
4.5. Acknowledgements 276

References 277

List of Abbreviations

CZE – capillary zone electrophoresis

BGE – background electrolyte

EOF – electroosmotic flow

tITP – transient isotachophoresis

μ_{APP} – solute apparent mobility

μ_{FS} – solute free solution mobility

μ_{EOF} – electroosmotic flow mobility

ζ – zeta potential

LC – liquid chromatography

RPLC – reverse phase liquid chromatography

R_S – resolution

MS – mass spectrometry

MS^1 – single stage mass spectrometry

MS^2 – tandem mass spectrometry

ESI – electrospray ionization

MALDI – matrix-assisted laser desorption ionization

CID – collision-induced dissociation

HCD – higher energy collision-induced dissociation

ECD – electron capture dissociation

ETD – electron transfer dissociation

UVPD – ultraviolet photodissociation

FTICR MS – Fourier transform ion cyclotron resonance mass spectrometer

TOF MS – time-of-flight mass spectrometer

Native MS – native mass spectrometry

Monoclonal antibody – mAb

PTM – post-translational modification

Immunoglobulin G – IgG

rhINF-β1 – recombinant human interferon-β1

cPEI – cross-linked polyethylenimine

M7C4I - 1-(4-iodobutyl) 4-aza-1-azoniabicyclo[2,2,2] octane iodide

Chapter 1: Introduction to Top-Down Proteomics and Capillary Zone Electrophoresis – Mass Spectrometry

1.0. Abstract

In biological systems, individual proteins exist not as single entities, but rather, as ensembles of chemically diverse species differing in primary structures and/or post-translational modifications (PTMs). Such species are defined as proteoforms. In the case of a prospective protein pharmaceutical, such as a monoclonal antibody, proteoforms with subtle structural differences are known to exhibit profound differences in biological activities. High resolution high mass accuracy mass spectrometry (MS) is a highly sensitive approach to the analysis of both intact isolated proteins and complex protein mixtures. While the conventional liquid chromatography (LC) – tandem MS approach is routinely applied to the analysis of proteolytic digests of complex protein mixtures, also referred to as bottom-up proteomics, it does not reveal individual proteoforms. Instead, the bottom-up proteomic approach results in the identification and quantitation of peptides with specific PTMs and is restricted to the identification of protein groups. Composition of individual proteoforms is unraveled by an alternative approach, to be referred to as top-down proteomics. In top-down proteomics, proteins are analyzed undigested and sequence composition and PTMs are identified through fragmentation in the mass spectrometer. This approach typically incorporates liquid phase separation and MS detection of an ensemble of intact proteins followed by the sequencing of mass-to-charge selected proteoforms to the constituent backbone fragments. Furthermore, analysis of protein complexes, which encompasses several protein subunits held together by non-covalent interactions under native conditions, provides information on protein-protein interactions as well as on the structure and composition of protein complexes.

Given the profound heterogeneity of proteins, high resolution separations are a requisite prior to MS detection for the successful characterization of intact proteins. While different modes of LC have historically been implemented most often in intact protein separations, capillary zone electrophoresis (CZE) represents a complementary approach. In CZE, analytes are separated in the liquid phase according to differences in charge and hydrodynamic volume. This chapter is an introduction to different analytical approaches applied to protein analysis, with an emphasis on CZE-MS. Fundamental aspects of MS are reviewed, including ionization techniques for transferring analytes from the liquid to the gas phase, followed by analytical separations and MS detection in the context of protein analysis. Advancements in analytical instrumentation, separation techniques, and data acquisition and analysis strategies applicable to intact protein characterization are described. Finally, analytical instrumentation and experimental techniques used in the following chapters are reported. The chapter is concluded with a brief overview of the projects described in detail in the ensuing chapters.

1.1. Heterogeneity of Proteins

The central dogma of molecular biology states that for every living organism, genetic information is stored as DNA. Through the process of transcription, this genetic information is converted into messenger RNA (mRNA), which is then converted into protein *via* translation. It is the proteins that are responsible for a great diversity of functions enabling life. However, proteins are far more complex than the heterogeneity stemming from combinations of the 20 common amino acids. An unexpected result of the human genome project was the discovery of approximately ~20,300 genes, as opposed to the anticipated ~100,000 [3]. This led to the realization that the complexity and variability of biological machinery exceeds that provided by the genetic code. Such complexity and variation may be explained in large part by the heterogeneity of protein structures [4].

Biological heterogeneity may be explained by events at the levels of DNA, RNA, and protein. For example, allelic variation may result in the production of different mRNA from the same gene. Alternative splicing events in eukaryotic mRNA may result in the excision of non-coding parts of the mRNA, thereby resulting in the production of different proteins from the same mRNA. In proteins, different post-translational modifications (PTMs) result in additional structural variation, potentially leading to the formation of species with different biological activities. The potential for biological recombination resulting in added protein heterogeneity is enhanced at each level of recombination, from DNA to mRNA to protein. A result is the production of chemically different forms of the same protein, originally stemming from the same gene, with variation in structure and often, biological activity. Such species have been defined as "proteoforms" [4].

However, in organisms, the biological machinery enabling life is often more complex than individual proteins, or even proteoforms. Biological processes are dependent on the interactions of different macromolecules, which may be proteins, RNA, lipids, or DNA [5]. For example, at the protein level, cadres of different proteins interact to form complexes held together by non-covalent interactions and are paramount for the normal function of biological systems [6]. Such complexes may be combinations of individual proteoforms of the same protein, or they may be combinations of proteins stemming from different genes. Alternatively, interactions between different proteins and RNA, such as ribosomal RNA (rRNA), result in the proper assembly of entire organelles, such as ribosomes [7]. In each case, each macromolecular complex represents an additional level of complexity and structural heterogeneity.

1.2. Proteins as Pharmaceuticals

Increasingly, protein pharmaceuticals are being implemented for treatment of disease and medical conditions. Recombinant DNA technology has enabled the production of human

proteins in host organisms such as bacteria, resulting in structural modifications for enhanced biological activity and large scale production [8]. Recombinant DNA technology, first developed in the 1970's, initiated a revolution in the pharmaceutical industry and facilitated the development and implementation of proteins as pharmaceutical products, termed "biologics" [9]. Since the introduction of the first biologic in the US in the early 1980's, over 90 commercial biologics have been FDA approved [10].

A key category of biologics are monoclonal antibodies (mAbs). Rapid advancements in the understanding of disease at the molecular level are resulting in a continued rise of therapeutic mAb development. Currently, over forty-seven FDA-approved mAbs are commercially-available on the US market, with a projection of up to 70 mAbs by 2020. Worldwide, mAb sales are expected to reach $125 billion by 2020 [10, 11]. A major advantage of mAbs is their ability to bind specifically and with high affinity to a broad range of molecules. A consequence of a lack of superfluous interactions is a significantly diminished chance of unexpected safety issues in human clinical trials, especially as compared with other pharmaceuticals. As a result, for many novel disease targets, mAb products are often the first candidates advancing to clinical trials [12].

Immunoglobulin G (IgG) antibodies have been the predominant type of mAbs employed as therapeutic proteins [13]. IgGs are composed of aggregated sets of two identical light and heavy chains linked by combinations of disulfide bonds and/or other non-covalent interactions, resulting in a ~150 kDa protein. The structure of most IgGs resembles a traditional "Y" form, with two F(ab) fragments linked to the F_C fragment *via* the hinge region. A schematic representation of a mAb is shown in **Figure 1-1**. Each F(ab) fragment is composed of a light chain and the variable (VH) and constant (CH1) regions of the heavy chain. The complementarity-determining regions (CDRs), which are the portions of the mAb implicated in antigen binding, are located at the N-terminal ends of each F(ab) fragment. The F_C fragment, composed of constant regions 2 and 3 (CH2 and CH3, respectively) of the heavy chain, is involved in the dimerization of the light and heavy chain doublets. A common mAb PTM is N-glycosylation at Asn_{297} of the CH2 region, resulting in considerable proteoform heterogeneity from the combination of glycans at the two glycosylation sites [14]. Other PTMs, further increasing proteoform heterogeneity, include Met_1-loss, formation of pyroglutamic acid, C-terminal lysine clipping, carboxymethylation of lysine, oxidation, deamidation, disulfide bond formation, glycation, single site amino acid replacement, and sequence truncation, amongst others [13, 15-18] .

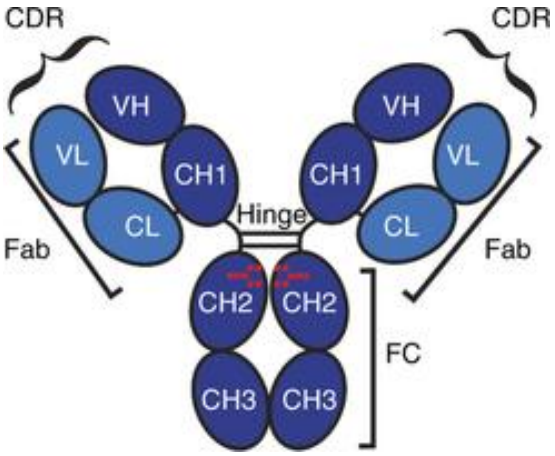


Figure 1-1. Schematic representation of a monoclonal antibody with labeled regions. Each light chain consists of a variable (VL) and common region (CL). Each heavy chain consists of a variable region (VL) and three common regions (CH1, CH2, and CH3). Each light chain-heavy chain dimer is held together by disulfide bonds in the hinge region. The variable regions of light and heavy chains constitute the complementarity-determining regions (CDR). Individual domains of the mAb are labeled in the schematic (F(ab) and F_C). Lastly, glycosylation is represented by red sugars bound to the CH2 region. A mAb may possess glycosylation at both CH2 regions, at one of two CH2 regions, or be fully aglycosylated. Figure reprinted from [19] with permission.

In projects outlined in this dissertation, two mAbs were investigated. In Chapter 3, an IgG1-type mAb (Biogen, Cambridge, MA) is described. In Chapter 4, analysis of intact humanized monoclonal anti-Her2 IgG1 kappa antibody (Trastuzumab, Herceptin; Genentech, South San Francisco, CA) is presented.

1.3. Analytical Techniques for Protein Characterization

Given the structure-function relationship of proteins, analytical characterization is important to the understanding of function and biological activity. For protein pharmaceuticals, structural and quantitative characterization is also necessary for effective drug design, clinical trial approval and subsequent product commercialization. Protein characterization involves the identification and quantitation of individual proteoforms, assessment of structure and function, and the identification of aggregation and possible associations with other molecules. These tasks require powerful analytical techniques [20].

Numerous analytical strategies exist for characterization of protein structure. For example, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are viable strategies for the determination of protein structure and molecular composition [21]. Edman degradation, which is a classical technique that involves sequential characterization of protein primary structure, is of limited application and is time-consuming, requires highly-pure single-protein isolates, and is limited in the identification and characterization of PTMs [22]. Other methods, such as dynamic light scattering and circular dichroism, also enable a degree of assessment of protein structure [23, 24]. Microscopy-based techniques enable visualization of proteins and structural assessment. However, these techniques offer insufficient sensitivity for the detection of PTMs and differentiation of many proteoforms. Studies have shown that even highly pure isolates of a single protein pharmaceutical – recombinant human interferon-β1 (rhINF-β1; Avonex, Biogen) – exist as hundreds of individual proteoforms distributed across a broad dynamic range [1, 25, 26]. Such proteoforms may even be combinations of several different PTMs, differentially affecting biological activity [1]. With both X-ray crystallography and NMR spectroscopy, the vast majority of these proteoforms are indistinguishable from each other without a prior high resolution separation and preconcentration, as many proteoforms are present at abundances below the limit of detection thresholds of these techniques. Moreover, the need for high purity isolates present at sufficient concentrations precludes high-throughput analyses of biological systems and proteomes with these techniques [27, 28]. With dynamic light scattering and circular dichroism, while conformational assessments may be achieved, characterization of individual residues is unattainable. Similarly, with various microscopy methods, while visualization enables determination of the folding and basic topography, such as protein quaternary structure [29], identification of individual residues and PTMs is unfeasible.

An additional consideration involves the need to analyze proteins under native, or near-physiological conditions. Under such conditions, interactions between different proteins are preserved. For example, a plethora of proteins in biology exist as aggregates held together by non-covalent forces. In such cases, characterization of subunit composition is imperative to the understanding of biological function [6]. With protein pharmaceuticals, such as mAbs, aggregation resulting from protein destabilization and unfolding negatively imparts biological activity and may induce immunogenicity following administration to patients [30]. Such structures, representing aggregates of mAb pharmaceuticals, are held together by non-covalent interactions and require native conditions for detection. With X-ray crystallography, which requires isolation of individual proteins or protein complexes, protein analysis under native conditions may not always be possible. A similar limitation exists with NMR spectroscopy [28].

An alternative means of protein analysis involves characterization by mass spectrometry (MS). While molecular resolution is inferior with MS as compared to X-ray crystallography and NMR spectroscopy, MS has notable advantages. For example, MS enables high-throughput analyses, detection of many analytes simultaneously, characterization under native conditions, and higher sensitivity [28].

1.4. Overview of Mass Spectrometry

MS enables the detection of analytes in the gas phase and measures their mass-to-charge ratios (*m/z*). In MS analysis, analytes are first transferred from the liquid to gas phases *via* ionization, such as electrospray ionization (ESI) [31] or matrix-assisted laser desorption ionization (MALDI) [32]. These ionization techniques, developed in the late 1980's, represented breakthroughs in biological molecule characterization. Previously, the unavailability of soft ionization techniques that could enable the transfer of biological macromolecules to the gas phase without excessive degradation and fragmentation limited the application of MS mostly to elemental analysis and the characterization of small and thermostable molecules. The introduction of ESI and MALDI facilitated the application of MS to biological macromolecule characterization, especially proteomics [33].

Following ionization, analytes are accumulated and injected into the mass analyzer for detection. The resultant signal intensities, measured in *m/z* ratios (*Thomsons* (*Th*)) and recorded on a mass spectrum, are proportional to the number of ions injected into the mass analyzer. Each ionized analyte observed in the mass spectrum is referred to as the molecular ion. In addition to a molecular ion, *m/z* clusters corresponding to naturally-occurring isotopes of the molecular ion are also observed. The relative intensities of isotopic peaks are proportional to the abundancies of elemental isotopes observed in nature. In MS, resolving power is the quotient of the analyte mass-to-charge ratio, *m/z*, and the analyte peak width at half maximum of the peak intensity, $\Delta m/z$. With sufficient mass accuracy and resolving power, the mass and charge of ionized analytes may be discerned from adjacent isotopic peaks [34].

Though composition of a molecular ion may be determined from elemental isotope ratios, e.g. for isobaric species composed of identical *m/z* ratios but different molecular compositions, intact measurements alone are insufficient. Fragmentation of injected analytes (precursor ions) results in the formation of informative fragment ions (product ions) that could then be used for the molecular characterization of the isobaric precursor ions [34]. In an example experiment, following ionization, a survey scan (MS^1 scan) of precursor ions is first performed. Following *m/z*-based selection and fragmentation of the specific precursor ions, a second scan (MS^2 scan) is then conducted to determine the *m/z* ratios of product ions. Such an experiment with alternating precursor and product ion scans is referred to as a tandem MS (MS/MS) experiment. While conventional MS^1 or MS/MS experiments result in the determination of molecular compositions and analyte sequences, they are incapable of elucidating the spatial arrangements of molecules within such species.

In the context of proteomics analysis, ESI is the most commonly-used ionization approach. In ESI, an acidic volatile liquid is sprayed from a narrow-bore needle, or an ESI tip. As a high positive (or negative) voltage is applied to the ESI tip, liquid droplets emanating from

the tip are deformed, producing a Taylor cone [35]. Sputtering of the Taylor cone results in the formation of a mist of positively (or negatively)-charged droplets, which are directed towards the counter electrode, which is the inlet of a mass spectrometer. Application of a heated gas, such as nitrogen, results in droplet desolvation. Following sufficient desolvation, the increased charge within droplets induces a Coulombic explosion, resulting in fission and formation of smaller droplets. Successive cycles of desolvation and droplet fission give rise to the transfer of analytes into the gas phase. These gas phase analytes are then directed into the inlet of the mass spectrometer by the appropriate electric fields [36].

In ESI experiments with protein mixtures, accumulation of excess positive charge results in the protonation of analytes at multiple sites. Under denaturing conditions (i.e. when using acidic buffers), analyte *m/z* ratios are reduced due to the increased accessibility to multiple protonation sites characterized by higher basicity. As a result, multiply-charged species are formed during ESI and the total analyte signal is distributed across multiple channels corresponding to the individual charge states. For proteins, which may range in molecular weights from 4,000 Da, in the case of some ribosomal proteins [37], to approximately 150,000 Da, in the case of mAbs [38], ESI enables detection of intact species on most commercial instruments. Depending on the instrument, signal distribution spans across multiple charge states in the *m/z* range between 100 - 4,000 *Th*.

While the development and commercialization of ESI paved way for the analysis of biomolecules by mass spectrometry, operation of ESI at nL/min flow rates (nanoESI) greatly expanded applicability of the ESI-MS to samples of even greater complexity [39]. Initially, several experiments performed in the early 1990's demonstrated that increased sensitivity could be attained by operation of ESI at flow rates of several hundreds of nL/min [40, 41]. In 1994, it was demonstrated that a pulled glass capillary ESI emitter with an inner diameter of a few micrometers at the tip could sustain ESI at flow rates of approximately 20-30 nL/min [42, 43]. In nanoESI, the size of the initial droplet arising from the Taylor cone is considerably smaller in comparison to that of the standard (or micro) ESI. The latter requires a longer series of Coulombic explosions for the formation of gas-phase ions. Consequently, desolvation is less effective than that in the nanoESI regime, resulting in decreased sensitivity. Over the past several decades, the evolution of ESI emitters has resulted in the development of robust needles that could routinely sustain operation of nanoESI. These emitters could be used in direct infusion experiments and could also be coupled to front-end separation techniques operated at ultra-low flow rates of tens of nL/min. As a result, sensitivity was improved dramatically and extended the applicability of ESI-MS towards proteomic analyses of increasing complexity [39].

An alternative to ESI for the analysis of biomolecules is MALDI. In MALDI, analytes are dissolved in a UV-absorbing matrix and placed on a stage for analysis by MS. As the solvent dries out, the matrix crystallizes and the analytes are thereby incorporated into the matrix crystals. A UV laser light is then used to vaporize small regions of the matrix, thereby resulting

in transfer of analytes into the gas phase. Ionization occurs as a result of protonation from the acidic environment of the matrix and from the dilute acid included in the sample [32].

In general, MALDI-MS is a less sensitive approach than ESI-MS due to limited ionization efficiency. In addition, in MALDI-MS, throughput is diminished as samples must be collected on a stage and crystallized with the matrix. Relative to the ESI, MALDI produces mostly singly-charged species [44]. In the context of intact protein analysis, ionization without reduction of the *m/z* ratio due to multiple charge states, as in the ESI, precludes the utility of MALDI with numerous commercial mass spectrometers. With the notable exception of time-of-flight instruments, most commercial mass spectrometers have an upper detection limit of 4,000 – 6,000 *m/z*. Detection of many singly-charged proteins, as generated by MALDI, results in ions exceeding the above-specified upper detection *m/z* range.

In conventional ESI-MS and MALDI-MS experiments, analytes are ionized from an acidic, volatile buffer and their *m/z* ratios are recorded on a mass spectrum. Additional characterization may be achieved by fragmenting analytes and recording the *m/z* ratios of fragment ion species, thereby elucidating molecular composition and sequence. However, for complex mixtures, a significant proportion of analytes may be undersampled by a mass spectrometer [45]. An example of a complex mixture is a cellular lysate characterized by hundreds, if not thousands, of different proteins within a broad dynamic range. In an analysis of such a mixture by ESI-MS, analytes may be represented within a broad continuum of abundances and may also possess different ionization efficiencies. The most abundant analytes with the greatest ionization efficiencies will accumulate the most charge in ESI and be significantly oversampled by the mass spectrometer. In effect, these analytes may overshadow less abundant species present in the sample on an ESI mass spectrum [45]. A viable solution involves separation of a complex mixture prior to characterization by ESI-MS. Such a mixture may be simplified *via* the separation of components according to differences in hydrophobicity, or charge and hydrodynamic volume. Simplification may result in more manageable sub-populations of the mixture that could then be analyzed separately by the mass spectrometer, thereby significantly increasing the depth of coverage and enabling identifications of low abundance species.

1.5. Introduction to Analytical Separations

1.5.1. Liquid Chromatography

In liquid chromatography (LC), analytes are separated in the liquid phase according to differences in physical properties such as hydrophobicity, charge, hydrodynamic volume, or the presence of affinity tags. In an LC experiment, analytes are first dissolved in a mobile phase. Next, they are injected onto a column containing a stationary phase. Separation occurs as a result

of differences in affinities of analytes for the stationary phase. As a result of mass transfer effects, analytes migrate through the column in bands approximately resembling Gaussian distributions. A variety of detectors, such as ultraviolet (UV) or mass spectrometers, may be coupled to LC. Following elution from the chromatographic column, analyte bands are recorded as chromatographic peaks by the detector of choice in a plot of intensity vs. time, referred to as a chromatogram [46].

Numerous modes of LC enable different interactions between analytes and the stationary phase. The most common mode is reversed-phase LC (RPLC), where analytes are separated according to differences in hydrophobicity [47]. Alternatively, hydrophilic interaction LC (HILIC) separates analytes according to differences in hydrophilicity. In this case, the stationary phase is a hydrophilic polymer [47, 48]. Other techniques, such as size exclusion chromatography (SEC) enable separation based on hydrodynamic volume [47]. In SEC, analytes are injected into a column packed with porous particles with varying pore diameters. Molecules with larger hydrodynamic radii bypass the porous particles and elute first. Interactions based on properties other than hydrodynamic volume are mitigated in SEC. Ion exchange (IEX) chromatography [47] is a mode of LC that enables separation based on differences in charge. In IEX, the stationary phase may either be positively or negatively charged. Analytes of opposite charge are attracted to the stationary phase *via* ionic interactions and separated according to the strengths of their affinities. In hydrophobic interaction chromatography (HIC) [47], which is similar to RPLC, analytes are also separated according to differences in hydrophobicity. However, analytes are applied to the stationary phase in a high salt buffer and are eluted by means of a hydrophobic eluent. In HIC, the more hydrophobic analytes elute last. Lastly, in affinity chromatography [47], the stationary phase is composed of an affinity resin with functional groups specific for certain analytes. As analytes are injected into a mobile phase, those possessing the corresponding functional groups will be retained, while others will flow through the column undisturbed. An eluent may then be introduced to remove the bound solutes from the stationary phase [46].

1.5.2. Capillary Zone Electrophoresis

An alternative liquid phase separation technique is capillary zone electrophoresis (CZE) [49, 50]. In CZE, analytes are separated according to differences in charge and hydrodynamic volume in an electrically-conductive buffer called the background electrolyte (BGE). A conventional CZE experiment involves separation in an open tubular capillary, micro-, or nano-fluidic channel with surfaces composed of charged silanol groups. As a result of charged functional groups on the capillary surface, oppositely-charged solvent molecules are attracted to form an electric bilayer. As a potential is applied from one end of the capillary to the other, an electric field is generated. Introduction of an electric field results in the displacement of the electric bilayer, resulting in a bulk flow of liquid, referred to as the electroosmotic flow (EOF),

towards the pole of opposite charge [49, 50]. In CZE, analyte mobilities, referred to as apparant mobilities (μ_{APP}), are a sum of analyte free solution mobilities (μ_{FS}) and the EOF mobility (μ_{EOF}). This relationship is summarized in **Equation 1**. Analyte μ_{FS}'s are characterized by the displacement of charged analytes towards poles of opposite charges, while μ_{EOF} is the displacement of the electric bilayer towards the pole of opposite charge. **Equations 2** and **3** summarize the relationships for μ_{APP} and μ_{EOF}, respectively.

$$\boldsymbol{\mu_{APP} = (\mu_{FS} + \mu_{EOF})}$$

Equation 1: Apparent mobility is defined as the sum of analyte free solution mobility (μ_{FS}) and EOF mobility (μ_{EOF}).

$$\boldsymbol{\mu_{FS} = \frac{q}{6\pi\eta R}}$$

Equation 2: An expression for the free solution mobility (μ_{FS}), which depends on the solute's net charge (q), Stokes radius (R), and viscosity (η).

$$\boldsymbol{\mu_{EOF} = \frac{\varepsilon\zeta}{\eta}}$$

Equation 3: An expression for the EOF mobility (μ_{EOF}), which relates to the relative permittivity (q) and viscosity (η) of the buffer solution, as well as the zeta potential of the electric bilayer (ζ).

The separation efficiency of analytical techniques, such as LC and CZE, is best modeled by the van Deemter relationship [47, 51], shown in **Equation 4**. The height equivalent of a theoretical plate (HETP) relates the peak width and the distance in which that peak has traveled. The plate count (N), also referred to as column efficiency, is a measure of the column resolving power and is related to HETP *via* **Equation 5**. A lower value of HETP therefore results in higher separation efficiency. The terms A, B, and C are constants representing different contributing factors to band broadening. In LC separations, the A-term of the van Deemter equation reflects the contribution to band broadening from Eddy diffusion, which represents different path lengths of analytes through the mobile phase. The B-term represents the contribution from longitudinal diffusion, or Brownian motion of analytes as they move through the column. Lastly, the C-term, representing the contribution of stationary phase mass transfer, describes differences in path lengths of analytes through the stationary phase [47]. The terms described above represent contributing factors to band broadening in LC separations, ultimately affecting separation efficiency. With CZE, given the absence of stationary phases and packed columns, the A and C terms of the van Deemter equation are eliminated, thereby resulting in unprecedentedly high plate counts compared to the conventional LC experiments [52].

$$\boldsymbol{HETP = \frac{L}{N} = a(d_p) + \frac{b}{u} + c(d_p)^2 u}$$

Equation 4: The height equivalent of a theoretical plate (HETP) is dependent on capillary/column length (L) divided by the efficiency (N). HETP is composed of three principle terms. The first term ($a(d_p)^2$) represents band broadening due to non-uniform flow through the stationary phase; the first term is proportional to the particle diameter (d_p). The second term (b/u) represents analyte longitudinal diffusion and is inversely proportional to flow rate. The third term ($c(d_p)^2u$) represents the contribution to HETP from band broadening within the mobile phase and pores of the stationary phase; the third term is proportional to particle diameter and the square of the flow rate.

$$\boldsymbol{N = 5.54\left(\frac{t_r}{w_{0.5}}\right)}$$

Equation 5: From the half peak height method, the efficiency (N) of an analyte depends on the migration time (or retention time in LC) (t_r) divided by the peak width at half height ($w_{0.5}$).

1.5.3. Coupling Separations Online to ESI-MS

In the context of proteomic analysis, inclusion of analytical separations prior to MS detection presents several notable advantages. For example, with complex mixtures composed of hundreds or even thousands of components distributed across a broad dynamic range, separation based on physical properties results in a simplification of the mixture, thereby enabling more comprehensive characterization by MS and identification of low-abundance analytes. Without prior analytical separations, direct infusion of complex mixtures into a mass spectrometer results in a detection bias towards the most abundant analytes and those with the highest ionization efficiencies. Additionally, separations enable increased time for MS/MS to be performed, as the mixture is broken up into more manageable sub-populations that could be analyzed in a sequential fashion in their order of elution by the mass spectrometer, thus increasing the depth of analysis and the number of MS/MS events [46, 53]. Even for a sample composed of a single highly pure analyte, such as an intact glycoprotein, individual proteoforms may be indistinguishable in the MS domain based on *m/z* and require separation based on other physical properties prior to MS detection for successful identification [1]. Coupling of separation techniques and mass spectrometry also offers the advantages of superior throughput, mitigation of analytical variance between replicate experiments, and fewer sample losses, as less sample handling steps are involved [46, 53].

However, a fundamental constraint in regards to coupling separations to mass spectrometry is the incompatibility of liquid phase separation techniques, such as CZE and LC, to mass spectrometry, which is operated under vacuum. Direct hyphenation of LC to mass

spectrometry must therefore be achieved *via* an interface that transfers analytes from the liquid into the gas phase with minimal losses, preserves the physical properties of the analyte, and removes a significant proportion of the separation medium, or mobile phase. Moreover, the interface junction must also allow for proper operation of both components in tandem. While a variety of ionization techniques are available enabling direct hyphenation, the most commonly-used involve atmospheric pressure ionization. Of these, ESI is most-often used for biomolecules such as proteins. Several modes of LC, such as RPLC and HILIC, are well-suited for direct coupling to ESI-MS. In these cases, the mobile phases are composed of largely volatile components, facilitating ionization and allowing for direct coupling to MS [54]. IEX and SEC may also be coupled directly to ESI-MS under conditions where ESI-compatible mobile phases with reduced ionic strengths are used [55, 56]. Other modes of LC, such as HIC and affinity LC, for example, must be performed in an offline fashion, under which separations are performed separately and individual fractions are collected, which are then analyzed independently by MS. With HIC and affinity LC, mobile phases typically contain non-volatile components that must be removed prior to ESI-MS [46].

1.5.4. Coupling of Capillary Zone Electrophoresis to Mass Spectrometry *via* Sheath-Flow Junctions

In regards to CZE, direct hyphenation remained a challenge that was resolved in the late 1980's *via* the introduction of a sheath-flow interface [57]. In CZE, a closed circuit is required for the generation of an electric field across the separation capillary. Also, in the conventional CZE interface design performed in a micro-capillary format, the capillary termini are immersed in anodic and cathodic reservoirs, containing electrically-conductive solutions, thereby precluding direct connection with the ESI needle and module. Successful operation of ESI-MS requires a closed circuit for the application of a high voltage, necessary for the electrospray to occur, in addition to the delivery of a stable flow of liquid into the ESI needle. The fundamental challenge involved with coupling of CZE and MS was therefore the hyphenation of the individual closed circuits required for both components and delivery of a constant flow rate of liquid to the ESI needle [58]. These issues were resolved by the introduction of a sheath-flow interface. A schematic of an early CZE-MS interface is shown in **Figure 1-2**. In such a setup, an electrical connection between CZE and ESI modules is established at their interface, as the BGE flown in the separation capillary is mixed with a sheath liquid flown coaxially from an independent, metal capillary immersed in a reservoir filled with an ESI-compatible, electrically-conductive solution [57]. However, CZE is fundamentally a nanoflow separation technique characterized by flow rates ranging from tens to hundreds of nL/min. In early CZE-ESI-MS interfaces, the coaxial sheath liquid was delivered at flow rates ranging from 5-10 μL/min. Therefore, the CZE eluent was diluted considerably at the junction between the CZE capillary and the ESI-MS system [59].

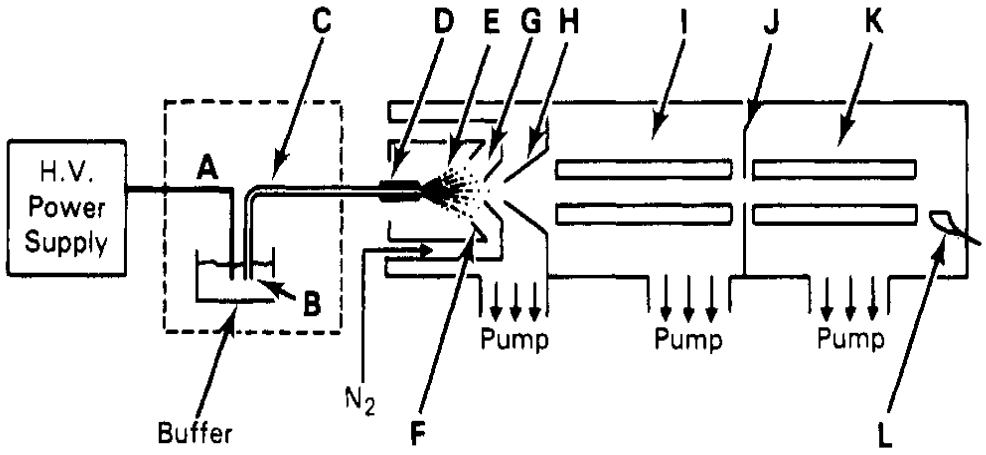


Figure 1-2. Schematic of the original sheath-flow interface for online CZE-ESI-MS. An electrically-insulating sampling box (A) is immersed in an anodic solution containing the BGE for CZE. The conductive capillary for the delivery of the sheath liquid (B) and the separation capillary (C) containing the BGE are immersed in the anodic reservoir. Both, separation and conductive capillaries are joined at the CZE/ESI junction, containing the cathode and electrospray needle (D). As an electrospray is generated (E), gas-phase ions then enter the mass spectrometer *via* the focusing ring (F), nozzle (G), and skimmer (H). Ions are then transmitted through an RF-only quadrupole (I), after which they pass through an ion entrance aperture (J), and are subsequently analyzed by the quadrupole mass spectrometer (K). A channeltron electron multiplier (L) is used to record an ESI-mass spectrum. This interface is reprinted from [57] with permission and represents the original sheath-liquid CZE-ESI-MS interface design.

1.5.5. Coupling of Capillary Zone Electrophoresis to Mass Spectrometry *via* Sheathless Junctions

Considerable improvements in CZE-ESI-MS interfaces have been achieved over the last decade [49, 50]. In particular, an interface was reported in which a hollow stainless steel needle was used as both the CZE outlet vial and ESI emitter. The separation capillary was inserted inside the hollow needle. A separate capillary, connected to a reservoir containing ESI-compatible solution, was connected to the hollow needle *via* a tee junction. In effect, sheath-liquid was delivered to the junction at flow rates of 0.1 μL/min or higher, resulting in considerable improvement over the early sheath-flow junctions [60]. Yet another sheath-flow interface has been reported capable of delivering coaxial sheath-liquid flow rates in the nL/min range, thereby resulting in significantly less dilution at the CZE-ESI-MS junction [61, 62]. Instead of using pressure to generate coaxial sheath-liquid flow, the sheath-liquid was supplied to the junction electrokinetically, which mitigates dilution effects at the interface between CZE and

ESI-MS. Reports with this setup have shown performance comparable to LC-based platforms [63].

Since the original design of the sheath-liquid junction in the 1987 [57], a number of improvements have been made, ultimately resulting in superior characteristics such as robustness, performance, and sensitivity [58]. The latter were achieved by decreasing the flow rates of the coaxial sheath-liquid [64]. A prospective alternative to sheath-flow interfaces are sheathless junctions. A number of designs have been reported. One such design incorporated a commercially-available capillary with an integrated, metal-coated ESI emitter [64]. The terminal end of the setup is passed through a sleeve containing an electrically-conductive, ESI-compatible liquid, which comes in contact with the outer surface of the ESI emitter, thereby enabling the closure of the independent electrical circuits for CZE and ESI-MS. The interface demonstrated attomole-level sensitivity, experimental reproducibility, robustness, and could be readily assembled [64]. Another novel, sheathless design involved a chemically-etched, porous capillary tip, which also acts as an ESI emitter [65]. A schematic of this interface is shown in **Figure 1-3**. In this setup, the terminal end of the separation capillary, containing the porous region, is inserted through a reservoir containing an ESI-compatible, electrically-conductive solution. This solution is pumped into the reservoir *via* a separate capillary. However, since only the porous end of the separation capillary is housed in the reservoir, the passage of ions from the reservoir through the pores in the separation capillary is achieved without dilution of the BGE. A number of groups have utilized this interface in proteomics experiments [1, 66-70]. The interface is sold commercially *via* SCIEX (Toronto, Canada) and marketed for CZE-MS applications as a part of SCIEX's CESI platform.

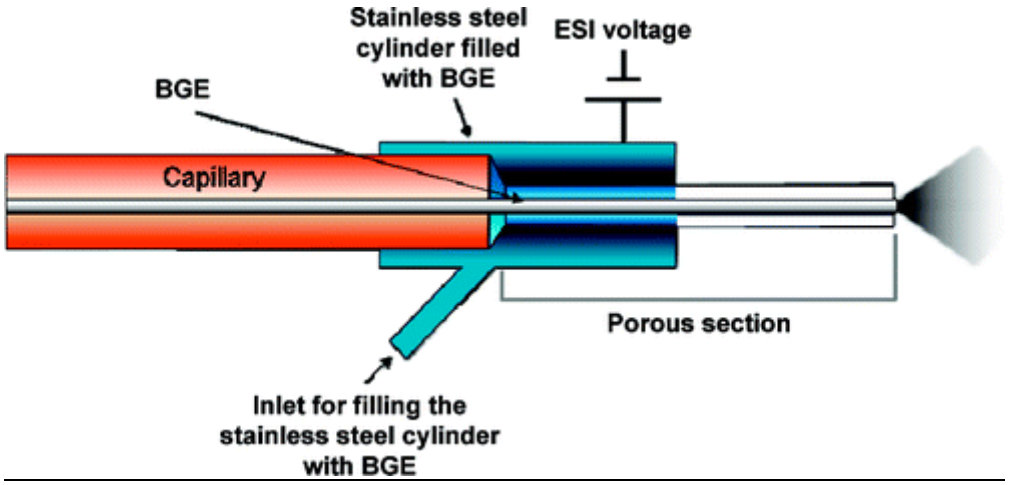


Figure 1-3. Representation of an etched porous nanosprayer for online CZE-ESI-MS. Part of the commercial CESI platform from SCIEX, the sheathless sprayer is composed of a

separation capillary with an HF-etched tip for nanoESI and a capillary used to deliver the conductive liquid. These components are joined at the CZE/ESI junction, containing a reservoir for the conductive solution. Analytical separations are performed in the separation capillary, culminating with the etched nanosprayer. A portion of the nanosprayer is housed in the reservoir. Pores formed *via* HF-etching in the terminal end of the separation capillary allow for the passage of ions into the separation capillary form the conductive solution-filled reservoir without dilution of analytes. This also allows for the closure of circuits for CZE and ESI. This schematic is reprinted from [67] with permission.

1.5.6. Capillary Zone Electrophoresis in Microfluidic Chips

An alternative to the micro-capillary format for CZE experiments are CZE separations performed on microfluidic chips. One example involves commercial microchips sold through 908 Devices (Boston, MA). A number of publications involving this interface have been reported [71-73]. A distinct advantage of such systems includes administration of capillary coating in the gas phase through chemical vapor deposition [71]. Utility of CZE for the analysis of charged biological macromolecules, such as intact proteins, depends in part on the use of appropriate coatings preventing non-specific adhesion to the separation channel walls. While capillary coatings for CZE are to be discussed later, it should be mentioned that the utility of chemical vapor deposition for coating administration results in considerable improvements in coating homogeneity, thereby resulting in diminished longitudinal diffusion of analytes represented by a smaller B-term in **Equation 4** [71]. Microchip design, shown in **Figure 1-4**, was based on a cross geometry and contained four functional elements: (1) an injection reservoir, (2) a BGE reservoir, (3) a waste reservoir, and (4) a serpentine separation channel. Connection between the CZE component and the ESI-MS junction was enabled by administration of a sheath-liquid, delivered electroosmotically to the ESI outlet *via* a separate channel within the microchip, thereby linking the sheath liquid reservoir with the ESI outlet [71]. Fundamentally, the microchip design represents a sheath-liquid interface between CZE and ESI-MS with limited dilution of analytes at the junction enabled by electroosmotic delivery of sheath liquid to the ESI outlet. Given the low flow rate of the sheath liquid, sensitivity of this interface is expected to be comparable to the most advanced sheathless systems. However, studies involving head-to-head comparisons are necessary to validate this point.

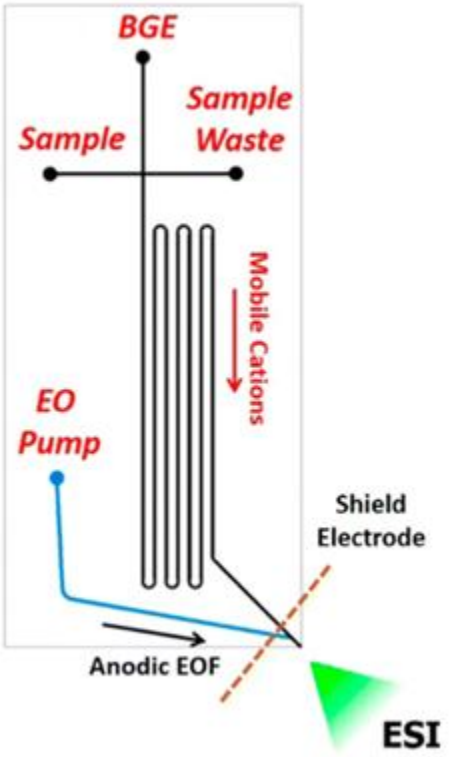


Figure 1-4. Schematic representation of the microchip used in online CZE-ESI-MS. The tee design of the microchip contains inlets for the introduction of BGE and sample, as well as a sample waste reservoir. The 23 cm-long separation channel (black) is coated with APS-PEG. Application of an appropriate electric field initiates CZE through the separation channel. A separate PEG-coated channel (blue), containing the sheath-liquid solution, is delivered in parallel to the CZE/ESI junction at the lower right edge of the microchip, thereby allowing for CZE operation and ESI. This schematic is reprinted from [72] with permission.

1.6. Proteomic Analysis by Mass Spectrometry

1.6.1. Bottom-up Proteomics

In the context of protein characterization by ESI-MS, the most straightforward experiment involves direct infusion of an intact protein or protein mixture into the MS, enabling the identification of molecular weights of the most abundant species. Numerous proteoforms and corresponding PTMs may also be revealed *via* delta mass shifts of their molecular weights relative to the unmodified variants [74]. Utility of MS/MS increases the depth of profiling, allowing for the determination of protein sequence and confirmation of proteoform identities determined previously from accurate mass. However, in such an experiment, especially if a

complex mixture composed of hundreds or even thousands of proteins is involved, identifications are restricted to only the most abundant protein groups and proteoforms. Limited sensitivity hinders identifications of lower-abundance species due to ionization suppression at the ESI source [74].

Consequently, a more advanced approach involves the analysis of proteolytic fragments of a single protein or protein mixture, to be referred to as bottom-up protein characterization. A typical experiment involves reduction and alklylation, followed by proteolytic digestion. The resultant proteolytic fragments are then separated by LC and characterized by ESI-MS/MS. In this approach, ESI mass spectra of individual proteins are represented by the signals of charged peptides, resulting in more comprehensive characterization by MS. In the bottom-up approach, individual PTMs may be identified more readily than in intact protein analysis, as the depth of coverage is increased for a combination of smaller peptides in comparison to an intact protein [74]. The approach also allows for superior determination of PTM localization, as individual amino acid residues on peptides are more accessible for fragmentation than on an intact protein, where they may be localized within the interior of a moiety [74]. Bottom-up proteomics may be performed readily in a variety of ways. For example, numerous modes of LC, allowing for the separation of peptides based on different physical properties, are compatible with ESI-MS in an online format, thereby allowing for direct analysis of a proteolytic peptide mixture in a single experiment by online LC-ESI-MS/MS. Alternatively, fractions of separated polypeptide bands may be collected. These fractions are then either infused directly into the mass spectrometer, or further separated using a different mode of LC, resulting in a two-dimensional separation [46]. An example of such an experiment involves analysis of a reduced, alkylated, and digested isolate of a mixture of proteins first by IEX LC, followed by online RPLC-ESI-MS/MS analysis [75]. Identification of individual protein groups is performed by comparing the identified peptides to sequences of the theoretical proteins contained within a genome-specific database. Identifications of individual PTMs are carried out by matching fragmentation mass spectra of peptides to the spectral libraries of the predicted theoretical peptides with known PTMs. A variety of commercial search algorithms and statistical approaches are available, enabling confident identifications of proteins from tandem mass spectra of proteolytic digests. Recent advancements in analytical instrumentation, separation science, and bioinformatics have resulted in considerable progress in bottom-up analysis of protein mixtures. The most advanced RPLC-ESI-MS/MS analyses have reported confident identifications of over 6,000 proteins from a 1 µg injection of a HeLa cell lysate [76-79] and over 9,000 proteins identifications from analysis of the *Drosophila melanogaster* proteome [80]. In comparison, state-of-the-art CZE-ESI-MS/MS approach has resulted in the identification of over 10,000 peptides, corresponding to ~2,100 protein groups [81].

While bottom-up proteomics is the most conventional method for analyzing complex mixtures of proteins by mass spectrometry, a number of inherent disadvantages exist. For example, bottom-up MS involves numerous sample preparation steps, thereby resulting in the

introduction of considerable artificial heterogeneity and analytical variance. An example includes the introduction of oxidation to a protein sample from exposure to air in the laboratory environment from numerous tube-to-tube transfers. Alternatively, denaturation of proteins with urea may result in carbamilation of the unfolded protein. While the bottom-up approach allows for sensitive detection of low abundance species and identification of a limited number of PTMs, fundamentally, individual proteoforms may not be revealed. For example, a single proteoform may be a combination of numerous PTMs. As the protein is reduced, alkylated, digested, and thereby simplified to a combination of peptides, the resultant bottom-up spectra do not allow for the identification and characterization of PTM combinations endogenous to the original proteoform [82]. From a biological standpoint, studies have shown that abundances of individual PTMs may be correlated to and may affect biological activity. For example, for rhINF-β1, levels of deamidation were correlated to specific glycosylation states [1]. Finally, non-covalent associations between proteins are lost in the course of bottom-up analyses. A considerable number of biological functions depend on the proper interaction of different proteins. For example, ribosomes, which facilitate the conversion of an mRNA to a protein *via* translation, are composed of over 50 proteins, depending on the organism, assembled together non-covalently with ribosomal RNA. In the bottom-up approach, associations between individual ribosomal proteins are lost as the sample is reduced and digested [37]. Even for an isolate of a highly-pure single protein, bottom-up proteomics results in a complete loss of information concerning any potential aggregates or bound ligands, labile PTMs that dissociate following exposure to denaturing conditions, such as O-phosphorylation, and correlation between different PTMs present on the same proteoform. In other words, bottom-up proteomics, in large part, is the study of peptides, which correlate only in part to real proteoforms driving biological activity.

1.6.2. Top-Down Proteomics

An alternative to the bottom-up approach is based on the liquid phase separation of a mixture of intact proteins followed by electrospray ionization and characterization by tandem mass spectrometry. This method is referred to as top-down proteomics. With this approach, intact proteins possessing numerous PTMs are directly identified and quantitated. Following successful fragmentation in a mass spectrometer, PTMs are localized down to specific amino acid residues. Top-down proteomics performed under near-physiological conditions enables characterization of native protein complexes held together by non-covalent interactions. Furthermore, labile PTMs and non-covalently-bound ligands are also detected.

However, top-down proteomics also presents unique analytical challenges not encountered in the bottom-up approach. One challenge pertains to protein solubility. Up to one third of the proteins encompassed in the human genome are hydrophobic membrane proteins. Consequently, these species are often underrepresented in top-down studies due to their low solubilities [83, 84]. Conversely, in the bottom-up proteomics, as hydrophobic proteins are

reduced and digested, a fraction of the resultant peptides are sufficiently soluble and possess adequate ionization efficiencies to enable detection of the hydrophobic protein [85]. Nevertheless, while the characterization of hydrophobic proteins *via* top-down proteomics remains an analytical challenge, several recent studies have reported on the use of MS-compatible detergents, micelles, or lipids for the stabilization of these structures in solution and subsequent ionization in the gas phase by ESI. These techniques were applied to protein multimers, in addition to protein-ligand complexes [86-88].

Another challenge of top-down proteomics pertains to MS detection of proteins exceeding approximately 70,000 Da. For these analytes, MS instrumentation with extended detection ranges is required. For example, for proteins such as mAbs, the resultant *m/z* distributions on an ESI mass spectrum may exceed detection ranges of some commercial mass spectrometers. Additionally, for high resolution Fourier Transform mass spectrometers (FTMS), such as the Fourier Transform Ion Cyclotron Resonance MS (FTICR-MS) [89] and the Orbitrap™ MS [90], resolving power decreases with an increase in *m/z*, resulting in diminished performance with higher molecular weight analytes. While advances in MS instrumentation for top-down proteomics are to be discussed in the ensuing section, it should be noted that such improvements have been achieved on a broad spectrum of contemporary mass spectrometers, including FTICR MS, Orbitrap MS, and time-of-flight (TOF) MS instruments over the last two decades [74].

In top-down applications, the use of ESI presents both advantages and challenges. On one hand, for small to medium-sized proteins (e.g. 4-55 kDa), ESI enables generation of multiply-charged species, resulting in the formation of multiple charge states and a distribution of *m/z's* on an ESI mass spectrum range amenable for analysis by most commercial mass spectrometers. Formation of multiply-charged ions also allows for more successful fragmentation of intact proteins, which is to be discussed later. On the other hand, generation of multiple charge states by ESI also results in "dilution" of the total analyte signal distributed across those charge states, thereby decreasing the signal-to-noise (S/N) ratio. This is particularly important for low-abundance species in a complex mixture, as insufficient S/N may preclude their detection by MS. A potential solution involves protein separation and ionization under near physiological, or native conditions (native MS).

With the growing understanding that protein-protein interactions are directly correlated to biological activity, the domain of native MS has been expanding over the past several decades [91, 92]. Analysis of intact proteins under native conditions allows for the detection of any non-covalent interactions and mitigates the incorporation of artificial modifications administered during sample preparation. In native ESI-MS at physiological pH of ~7, proteins assume more folded (native) conformations, making accessibility to the basic amino acid residues more restricted for protons from an ESI buffer. As a result, less protonation occurs as compared to that of the standard ESI performed under denaturing (more acidic) conditions. Consequently, native ESI-MS spectra are characterized by lower charge states and a diminished number of total charge

states, thereby yielding higher S/N [93, 94]. However, a central topic to native MS is whether or not proteins maintain their native conformations following ESI. It was hypothesized that proteins detected by native ESI-MS may not exhibit native structures and are, in fact, partially unfolded during the conversion from the condensed to gas phases [91]. A key experiment involved the comparison of protein collisional cross-sections (Ω), established *via* ion mobility spectrometry (IMS), with the corresponding Stokes radii (R_S), identified *via* dynamic light scattering (DLS), which is a technique that allows for the measurement of analyte size distributions in the condensed phase. For a variety of proteins and non-covalent protein complexes in the range of 15-500 kDa, Ω and R_S were found to be closely-related, confirming that protein conformation is preserved during ESI [95]. Moreover, a growing body of evidence, obtained with a variety of techniques, confirms conformational similarity between protein structures in the condensed and gas phases [96-104].

Top-down proteomics, performed under native conditions requires state-of-the art analytical instrumentation. Furthermore, a limited number of separation techniques that could be coupled to ESI-MS online allow for the operation under native conditions. Solutions to these analytical challenges are to be discussed in the ensuing sections on MS instrumentation for top-down proteomics (**sections 1.6.3-1.6.4**) and separations of intact proteins for top-down analysis (**section 1.7**).

1.6.3. MS Instrumentation for Top-Down Proteomics

Historically, the considerable analytical challenges described above have limited the application of top-down proteomics to analyses of complex mixtures. However, advancements in instrumentation, separation science, and bioinformatics have resulted in increased applications of the top-down approach to protein analysis over the past two decades. In intact protein analyses, high resolution and high mass accuracy are necessary for separation and accurate assignment of spectral peaks arising from multiple intact proteoforms. For example, extremely high resolving power may be required to identify and distinguish deamidated species from unmodified proteoforms, differing by 1 Da. Similarly, such capabilities are necessary for the identification of disulfide bridges, which represent a difference of 2 Da from the unmodified variant. High resolution and mass accuracy are also essential for the identification and accurate assignment of fragmentation spectra containing hundreds of fragment ions from a variety of proteins in a single mixture. Therefore, improvements in performance of high resolution MS instrumentation have greatly facilitated top-down proteomics.

Considerable interest in top-down analysis has been generated by FTICR MS instrumentation, which has come a long way from the first experiment reporting detection of an intact protein in 1989 [105]. In FTICR MS, ions are excited to their cyclotron orbits within a strong magnetic field. Excitation results in formation of spatially-coherent ion packets, which

oscillate at the reduced cyclotron frequencies, inversely proportion to their *m/z*. Ion clouds of spatially-resolved analytes induce image charge signals on the detection electrodes, which are then recorded by an acquisition system. The resultant transient signals are converted from a time domain to a frequency domain using a fast Fourier transform. Frequency spectra are converted to mass spectra using a calibration equation, which accounts for the magnetic field strength as well as for the space and image charge effects [89]. With an increase in the magnetic field strength, the mass resolving power and dynamic range of FTICR MS improve linearly and quadratically, respectfully. In the context of top-down proteomics, higher magnetic field strength resulted in broader use of FTICR MS instrumentation for intact protein analysis [106, 107], culminating in the introduction of a state-of-the-art 21 Tesla magnet [108]. Other enhancements to FTICR MS instrumentation include an electrodynamic ion funnel interface [109], a quadrupole mass filter [110], and external accumulation of ions prior to FTICR detection [111-113]. These modifications resulted in improved sensitivity, resolution, and dynamic range and were implemented on a 9.4 Tesla instrument [82, 114], which has since been used for a variety of top-down proteomic applications [115-117]. A further modification involved the development of a linear ion trap quadrupole (LTQ) FTICR MS, initially implemented on a 7 Tesla magnet [118] and subsequently utilized in a number of top-down proteomics experiments [119-121]. Further analytical improvements were enabled by a higher strength 12 Tesla [122-124] and 14.5 Tesla magnets [125].

An alternative to the use of high magnetic field strength for high resolution MS is an electrostatic trap, also based on image charge detection. The electrostatic ion trap, named an Orbitrap™ MS, was first introduced in 2000 [126] and then commercialized by Thermo Fisher Scientific in 2005 [127]. As compared to the high field FTICR, the Orbitrap does not require liquid helium, has a drastically smaller footprint, exhibits higher ion transfer efficiency from an external radio frequency (RF) ion trap to the electrostatic trap, as well as narrower temporal spread of ions in packets injected into the electrostatic trap. The latter two features result in higher overall sensitivity and compatibility with advanced signal processing techniques, such an enhanced fast Fourier transform (eFT) [128]. The detection component of an Orbitrap MS is composed of a central spindle electrode and coaxial outer barrel-like electrodes. As ions are injected into the region between the central and detection electrodes off the equator, they cycle about the central electrode on elliptical trajectories while also oscillating along the central electrode axis. The image charge induced on the detection electrodes represents a time-domain signal, which is converted to a frequency domain using an eFT. In the electrostatic ion trap, the frequency of ion axial motion is correlated to their *m/z*'s by an inverse square root dependence. Similar to FTICR, the frequency domain is converted to the *m/z* domain using a calibration equation [82]. Continual improvements in Orbitrap MS instrumentation have resulted in the commercialization of numerous configurations of the Orbitrap mass analyzer, such as the LTQ-Orbitrap MS [129], the quadrupole-Orbitrap MS (QExactive) [130], and the quadrupole -LTQ-Orbitrap MS (Orbitrap Fusion) by Thermo Fisher Scientific [131]. A particular advantage of Orbitrap MS over FTICR MS in the context of top-down analysis involves the relationship

between *m/z* and resolving power. For FTICR MS, resolving power decreases proportionally to *m/z*, while for Orbitrap MS, resolving power decreases proportionally to $(m/z)^{1/2}$. Consequently, for higher *m/z* precursors and an identical ion acquisition time (e.g. 1 second), the Orbitrap MS produced higher resolving power than FTICR MS [132].

A number of recent instrument modifications have also been made to Orbitrap mass analyzers to facilitate improved analysis of higher mass species under native conditions, such as monoclonal antibodies [133] and non-covalent protein assemblies approaching 1 MDa [134, 135]. In particular, instrumentation modifications to a commercial Q Exactive system have enabled multi-stage fragmentation of non-covalent protein complexes to individual subunits, and then down to the peptide level, thereby enabling the characterization of protein complex stoichiometry and primary structure of individual subunits [135]. A demonstration of the capabilities of this instrument was reported recently on a tetrameric 150 kDa protein, where native MS analysis enabled first, the ejection of individual subunits from the tetramer, followed by subunit fragmentation and characterization of specific phosphorylation sites across different proteoforms subjected to different biological growth conditions [136]. In this example, MS analysis enabled the characterization of all levels of protein structure, beginning with quaternary associations of the tetramer, down to primary structure of individual subunits.

An alternative to the image charge detection based -FTICR MS and Orbitrap MS instrumentation are time-of-flight mass spectrometers (TOF MS). In TOF MS, following ionization at atmospheric pressure, ions are transferred in RF-guides through several stages of differential pumping, focused into the TOF extraction region, and then injected into a field-free region equipped with an ion mirror [137]. Analyte *m/z* ratios are elucidated from their flight times between the extraction region and TOF detector [33]. The last several decades have seen significant improvements in performance of commercial TOF MS instrumentation, especially in terms of mass accuracy and resolving power [33, 138]. Popular commercial configurations of TOF MS instruments include tandem TOF MS (TOF-TOF MS), quadrupole TOF MS (Q-TOF MS), and dual quadrupole TOF MS (Q-Q-TOF MS). For the latter, inclusion of quadrupoles prior to the TOF drift tube allows for precursor ion selection followed by fragmentation, thereby enabling MS/MS analysis [33]. For the past several decades, the domain of top-down proteomics has been dominated by the TOF MS due to the virtually unlimited upper *m/z* range that these instruments can detect [134, 138, 139]. Nevertheless, these instruments possess certain limitations in the context of intact protein analysis. For example, one limitation on conventional instruments is the inability to select precursor ions in the quadrupole exceeding 3,000 – 4,000 *m/z* [138]. The other limitation is the reduced detection efficiency at higher *m/z*, which requires significant post-acceleration (> 20 kV) between the field-free region and the detector, particularly important in the native-MS studies [140, 141]. Nevertheless, a recent study reported on a Q-TOF MS with a quadrupole capable of selecting species up to 22,000 *m/z*, thereby enabling selection and fragmentation of large intact proteins, such as non-covalent complex assemblies exceeding 60 kDa [142].

While TOF instruments have theoretically unlimited *m/z* range and fast rates of spectrum acquisitions, they are inferior to Orbitrap instruments, such as the Q Exactive™ HF Orbitrap™ MS, in terms of sensitivity, dynamic range, as well as mass accuracy and resolving power [33]. As a comparison between Orbitrap MS and Q-TOF MS in the context of large intact protein analysis, a recent study reported on the detection of a 0.8 MDa 14-mer – the *E. coli* chaperonin GroEL – under native conditions. ATP and ADP molecules, representing less than 0.1% of the molecular weight of the GroEL, were incubated in the presence of the latter. The resultant mixture – one of ATP and GroEL and one of ADP and GroEL – was subsequently analyzed *via* direct infusion mode on an Orbitrap MS and on a Q-TOF MS. A modified Orbitrap system was capable of distinguishing between ATP- and ADP-bound states of the GroEL, while the Q-TOF MS only registered peak broadening upon addition of substrates to GroEL [134]. As TOF MS instruments are only capable of detecting average *m/z* values, they were incapable of distinguishing between substrate-unbound and substrate-bound states of the 0.8 MDa GroEL [134].

1.6.4. Techniques for the Fragmentation of Intact Proteins

While developments in FTICR MS, Orbitrap MS, and TOF MS instrumentation described above have enabled improved detection of large intact proteins, confident identification of such species strongly depends on the generation of informative MS^2 spectra. For complex samples encompassing many proteins and for proteomic samples of unknown composition, detection of intact ESI-MS spectra is often insufficient for protein identification. Consequently, a variety of fragmentation techniques for intact proteins have been developed. These include collision-induced dissociation (CID) and higher energy collisional dissociation (HCD), electron transfer dissociation (ETD) and electron capture dissociation (ECD), and ultraviolet photodissociation (UVPD).

In CID, analyte molecules present in the gas phase are accelerated by an applied electrical potential to a desired kinetic energy and collide with molecules of a neutral gas, such as argon, helium, or nitrogen. As a result of the collision between analyte and gas, a portion of analyte kinetic energy is converted into internal energy, which induces fragmentation of the amide bonds in the molecular ion and results in the formation of *b*- and *y*-type product ions [143], shown in **Figure 1-5**. However, in CID, activation of analyte molecules often proceeds *via* multiple collisions with gas molecules, thereby resulting in a rapid vibrational redistribution of internal energy. As the internal energy reaches the fragmentation threshold, the weakest bonds are preferentially cleaved regardless of whether they are present in the polypeptide backbone or side chains. Consequently, activation of proteins and peptides with CID results in an uneven fragmentation of the polypeptide chain, abundant internal fragment formation, and often results in a loss of PTMs from product ions following fragmentation [144]. Moreover, an additional limitation of CID – performed typically on ion trap mass analyzers – is that these instruments

cannot maintain parent ions and the lowest *m/z* product ions (bottom third of MS of MS^2 spectrum) stable at the same time, thereby resulting in a loss of the latter species during MS/MS analyses [145].

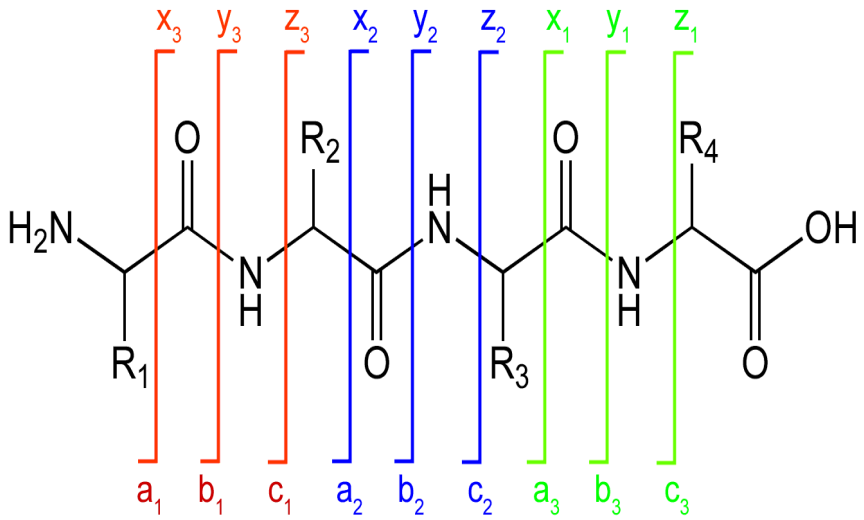


Figure 1-5. Cleavage locations on a generic tripeptide for CID/HCD, ECD/ETD, and UVPD. Fragmentation *via* CID/HCD results in the production of *b*- and *y*-type ions; ECD/ETD results in the production of *c*- and *z*-type ions, while UVPD results in the production of *a*-, *b*-, *c*-, *x*-, *y*-, and *z*-type ions. Schematic reprinted from https://en.wikipedia.org/wiki/Peptide_sequence_tag with permission.

Similar to CID, HCD also results in the fragmentation of the molecular ion into *b*- and *y*-type product ions. HCD is a fragmentation technique exclusive to Orbitrap MS and occurs in a multipole collision cell adjacent to the C-trap of the instrument [146]. In comparison to CID, HCD involves higher activation energies and shorter activation times [144]. In Orbitrap MS, ion detection occurs in the Orbitrap and not in the HCD multipole. Following fragmentation, the resultant ion packet is transferred to the C-trap, and then to the Orbitrap for detection. Therefore, HCD fragmentation does not suffer from the lower *m/z* cutoff as in CID performed on ion trap instruments. Consequently, even though various PTMs may still be lost from the polypeptide backbone following fragmentation with HCD, certain diagnostic signals characteristic of these PTMs may be detected in tandem MS spectra [144].

An alternative to CID /HCD is ECD fragmentation. Originally developed by the McLafferty group at Cornell University [147, 148], ECD is a soft fragmentation technique in which multiply-charged gas phase ions acquire an odd electron [149]. While the exact mechanism of fragmentation is still under debate, it is understood that exothermic electron

capture results in cleavage of the peptide bond, thereby producing *c*- and *z*-type ions depicted in **Figure 1-5** [149]. In ECD, precursor ions are bombarded with low energy electrons, since electron capture is highest at a minimum kinetic energy difference between the analyte ion and the electron. As the rate of electron capture is proportional to the square root of the precursor ion charge, multiply charged species are preferred in ECD [150]. Due to being a nonergodic process [151], ECD fragmentation of peptide backbone occurs before redistribution of internal energy and subsequent cleavage of the weakest bonds within polypeptides, as in CID. As a result, fragmentation with ECD results in a more uniform cleavage across peptide bonds, while preserving amide linkages and PTMs in product ions [149]. The ECD fragmentation technique is largely confined to FTICR-MS, as fragmentation occurs directly in the magnetic field of the ICR cell, which enables radial confinement of the electrons as they interact with precursor ions [149].

Similar to ECD, ETD involves the fragmentation of multiply-charged captions by transferring electrons to them, resulting in the generation of *c*- and *z*-type ions [152]. However, in ETD, electrons are transferred to the analyte ions from radical anions, instead of the direct transfer from irradiation observed in ECD. ETD is effectively the application of ECD beyond the FTICR MS and is used widely in TOF MS, ion trap MS, and Orbitrap MS instruments. Similar to ECD, ETD results in non-random polypeptide backbone fragmentation and is suitable for the analysis of labile PTMs [144]. However, in comparison to CID and especially HCD, ETD does not always produce complete fragmentation, especially of higher molecular weight species such as intact proteins. Instead, charge-reduced species are frequently observed. A possible explanation involves the incomplete overlap of analyte ion and reagent ion clouds, resulting in a mitigated interaction and formation of charge-reduced product ions. One solution involves sequential fragmentation with ETD and then HCD (EThcD), or ETD and the CID (ETciD) [153, 154]. These analysis modes are available exclusively on LTQ-Orbitrap MS [154, 155] instruments and the Orbitrap Fusion [153] and involve fragmentation with ETD in the LTQ, subsequent fragmentation of product ions in the HCD cell, followed by analysis in the Orbitrap. Product ions produced *via* ETD fragmentation are collided with neutral gas molecules in the HCD cell, resulting in the break-up of charge-reduced species into ETD-characteristic *c*- and *z*-type ions. Studies with EThcD have shown superior fragmentation of larger intact proteins than ETD or HCD alone [153].

An alternative fragmentation technique to the “slow heating” methods of CID and HCD and electron transfer-based methods of ECD and ETD is UVPD. Recent studies have reported on superior fragmentation and sequence coverage with UVPD than with CID/HCD or ECD/ETD in the context of intact proteins [37, 156]. UVPD results in the fragmentation of the amide backbone of proteins/polypeptides, thereby generating *a*-, *b*-, *c*-, *x*-, *y*-, and *z*-type ions, as shown in **Figure 1-5**. While the exact mechanism of UVPD is still being debated, it is understood that the amide backbone serves as an effective chromophore for the absorption of UV photons at a wavelength of 193 nm, resulting in high energy activation and subsequent fragmentation [156]. UVPD has been implemented in a wide array of instruments, including ion trap MS, TOF MS,

FTICR MS, and Orbitrap MS instruments [156]. A major advantage of UVPD is that it exhibits no charge-state dependence, as with ECD/ETD, where fragmentation of larger species is limited and results in the formation of charge-reduced products. Unlike in CID/HCD, UVPD enables the preservation of PTMs and allows for the identification and localization of specific PTM sites [37].

With CID/HCD, ECD/ETD, and UVPD, fragmentation of analytes occurs during latter stages of analysis in the mass spectrometer; with UVPD or ECD on FTICR MS instruments, for example, fragmentation occurs in the ICR cell following initial scans of precursor ions. In the case of HCD on Orbitrap MS instruments, fragmentation occurs in the HCD multipole after numerous events in the mass spectrometer, such as mass selection in the quadrupole mass filter, ion accumulation in the C-trap, initial precursor scans in the Orbitrap analyzer, etc. An alternative means of fragmentation is surface-induced dissociation (SID), which is a front-end fragmentation technique that relies on collisions of analytes with surfaces [157]. Originally developed in the laboratory of Graham Cooks of Purdue University in the late 1970's/early 1980's [158-160], SID resembles CID/HCD in that it involves the conversion of kinetic energy into internal energy *via* inelastic collisions of analyte ions. However, in SID, a surface replaces the neutral gas as the collision target. An advantage of SID over analogous CID/HCD techniques is that surfaces represent larger collision targets and therefore result in greater energy transfer after collision with analyte ions. Collisions with neutral gasses as in CID/HCD result in limited energy transfer due to the limited surface areas and low mass of neutral gas molecules. With surfaces, the collisional cross-section is significantly greater than for comparatively smaller gas ions, thereby allowing for superior energy transfer [157]. In the context of intact protein analyses, SID has been demonstrated to allow for superior fragmentation of larger species than ECD/ETD [161]. Especially in the context of native protein analysis, applications with SID have demonstrated that individual subunits of a multiprotein complex can be dislodged without subsequent unfolding of the dissociated monomer, as seen with CID/HCD [161]. For analysis of monomeric proteins, utility of SID has shown that proteins may be more effectively desolvated of non-covalently-bound adducts, resulting in sharper MS peaks with enhanced signal-to-noise ratios [134, 161] .

Advances in fragmentation techniques described above have greatly extended the capabilities of top-down proteomics towards the characterization of large proteins structures – and even non-covalent protein assemblies – in the tandem mass spectrometry mode. Limitations in fragmentation capabilities of intact proteins were for a long time considered a major bottleneck preventing the more widespread application of top-down analyses [74] , as detection of intact proteins alone is insufficient for sequence characterization and for the identification and localization of individual modifications. While further advances are necessary, current capabilities enable a variety of applications in top-down MS on many contemporary mass spectrometers.

1.7. Separations of Intact Proteins for Top-Down Proteomics

1.7.1. Conventional Techniques and Liquid Chromatography

The above-described advancements of MS instrumentation have facilitated the increased application of top-down analysis of proteins over the past two decades [74, 82]. However, analysis of complex mixtures of many proteins requires a high resolution separation component prior to characterization by MS. Moreover, even for analyses of single proteins, high resolution separations are required to resolve individual proteoforms so that they could then be distinguished by MS. Recent advancements in LC and CZE technologies have enabled improved separations of protein mixtures and single proteins for subsequent characterization by tandem MS.

For intact proteins, a variety of LC techniques may be coupled directly to ESI-MS in an online manner. For example, RPLC-ESI-MS and HILIC-ESI-MS are widely used methods for intact protein analysis [46, 48]. It should be noted that considerable improvement in separation performance with RPLC in particular has been achieved by the introduction of submicron particles in packed capillary columns [162]. Utility of IEX LC with mild salt gradients has also been shown to enable separations of intact proteins, based on differences in charge, in an online fashion with ESI-MS [163]. Other techniques, such as SEC, affinity LC, and HIC require offline operation, where a separation is performed and fractions are collected, which may then be analyzed by ESI-MS [46]. Gel-based techniques, such as sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), performed in both denaturing and native formats, enable separations based on molecular weight [46]. Two-dimensional SDS-PAGE (2D SDS-PAGE) separates proteins according to molecular weight and isoelectric point (pI) [164]. Native polyacrylamide gel electrophoresis (Native PAGE) separates proteins according to electrophoretic mobility under near-physiological conditions [165]. However, typically for SDS-PAGE, 2D SDS-PAGE, and Native PAGE, following separation, individual bands of interest are manually excised from the gel, and the gel polymer is removed prior to analysis by ESI-MS [166]. Conversely, gel elution liquid fraction entrapment electrophoresis (GELFrEE) enables separation of intact proteins in a gel-based format but allows for liquid phase recovery of separated species, thereby bypassing the requirement for further sample pretreatment prior to MS analysis [167]. A slightly modified format of the GELFrEE system (native GELFrEE) has also been shown to be a viable technique for intact protein separations under native conditions [168].

1.7.2. Ion Mobility Spectrometry

An alternative to LC and gel-based techniques for the separations of intact proteins is ion mobility spectrometry (IMS). In IMS, analytes are separated in the gas phase based on differences in size and shape [169]. Hyphenation of IMS to MS was reported first in 1962 [170].

However, more widespread application of IMS-MS for the analysis of proteins has been gaining traction only recently, largely due to the unavailability of necessary instrumentation in decades prior [171]. IMS is effectively a pulsed gas phase separation technique ideally suited for TOF MS. Conversely, hyphenation of IMS to high-resolving power MS instrumentation, such as Orbitrap MS, is an instrument design challenge due to the disparity between the comparatively fast IMS separation and slow Orbitrap MS acquisition. Consequently, successful coupling of IMS to Orbitrap MS been achieved only recently [172, 173]. Studies involving IMS-MS have shown the general applicability of this technique for the analysis of proteins under native, or near physiological conditions [169]. Fundamentally, IMS-MS represents an orthogonal separation method in comparison to LC techniques for the analysis of biomolecules. An example of such orthogonality involves separation by IMS of glycopeptides differing only in terms of the sites of glycosylation; separation of these species was not achieved by any mode of LC [174]. Lastly, IMS is well-suited for the separation of analytes with conformational differences. For example, variants of a single protein differing in conformation while possessing identical *m/z* ratios are unsuitable for separation by any mode of LC that could be coupled online to MS. Conversely, separation and characterization of such species may be achieved routinely in an online format by IMS-MS [175, 176].

1.7.3. Intact Protein Separations by Capillary Zone Electrophoresis

An alternative to the liquid phase separation techniques and gas phase IMS described above for the analysis of intact proteins is CZE. CZE – a liquid phase separation mode of capillary electrophoresis where individual analytes migrate through the capillary in zones – is a highly efficient separation technique capable of generating hundreds of thousands of plates in a single experiment [177]. CZE, like IMS, is an orthogonal separation technique in comparison to more conventional LC. With RPLC, HILIC, HIC, and IEX, separation of analytes occurs as a result of differences in interactions with stationary phases. In particular, with RPLC, separation is dictated by differences in hydrophobicity. With CZE, analytes are resolved by differences in charge and hydrodynamic volume. A different mechanism of separation is thereby present for CZE offering complementary and/or orthogonal analysis to LC.

The most straightforward proteomic CZE experiment involves the separation of peptides on a bare-fused silica (BFS) capillary, with a choice of either UV or ESI-MS detection [178]. A BFS capillary operated at acidic conditions contains a surface composed of negatively-charged silanol groups. For small analytes such as peptides, interactions with the negatively-charged capillary surface are mitigated, as peptides possess relatively few surface charges. However, in the context of intact protein analysis, one of the main limitations of CZE pertains to adsorption of proteins to the capillary surface *via* electrostatic bonds, hydrogen bonds, and hydrophobic interactions [178]. Such adsorption results in surface contamination and causes a non-uniform ζ potential across the capillary, thereby resulting in sample losses, band broadening, and peak

asymmetry [179]. Consequently, development of capillary coatings for CZE is paramount to the prevention of non-specific protein adhesion. A variety of charged and uncharged coatings have been developed [180]. With these, application of the coating may result in either static (covalently bonded) or dynamic (ionically-bonded) interactions with the surface of the BFS capillary [178]. For online experiments with ESI-MS, static coatings are preferable, as dynamic coatings may result in dissociation of the coating reagent and subsequent contamination of the mass spectrometer. A number of reviews of capillary coatings for CZE have been published [50, 180, 181].

1.7.4. Preconcentration Techniques for Intact Proteins in Capillary Zone Electrophoresis

Another limitation of CZE is the relatively limited sample loadability [182]. CZE is fundamentally a nanoscale technique, which serves as an advantage in that relatively little sample is consumed. However, injection of nanoscale quantities of peptides or intact proteins may result in limited sensitivity during MS analysis and the signal loss of low abundance analytes. One solution involves the utility of transient isotachophoresis (tITP) [182]. In tITP, analytes may be preconcentrated in the capillary as a result of differences in conductivity between the BGE and sample solution. A schematic summarizing tITP is shown in **Figure 1-6**. Briefly, in tIPT, a capillary is filled with sample, leading electrolyte (LE), and trailing electrolyte (TE). LE and TE typically have a common counterion and different co-ions. The ionic mobility of LE counterions is greater than the ionic mobility of co-ions from the TE. Following the application of a voltage across the capillary, the sample plug is preceded by a high electrical field in the TE and a low electrical field in the LE, resulting in the focusing of solutes at the TE/LE interface. In normal-mode CZE, separation performance is dictated by sample volume injections not exceeding, as a rule-of-thumb, 1% of the total volume of the capillary. Injection of greater sample amounts results in compromised resolution [183]. However, in tITP, provided that a conductivity difference exists between the BGE and sample buffer, sample injections close to 100% of the total capillary volume may be achieved, thereby resulting in significant increases in sample loadability. For example, a 100-fold preconcentration using tITP in comparison to normal-mode CZE has been reported [184].

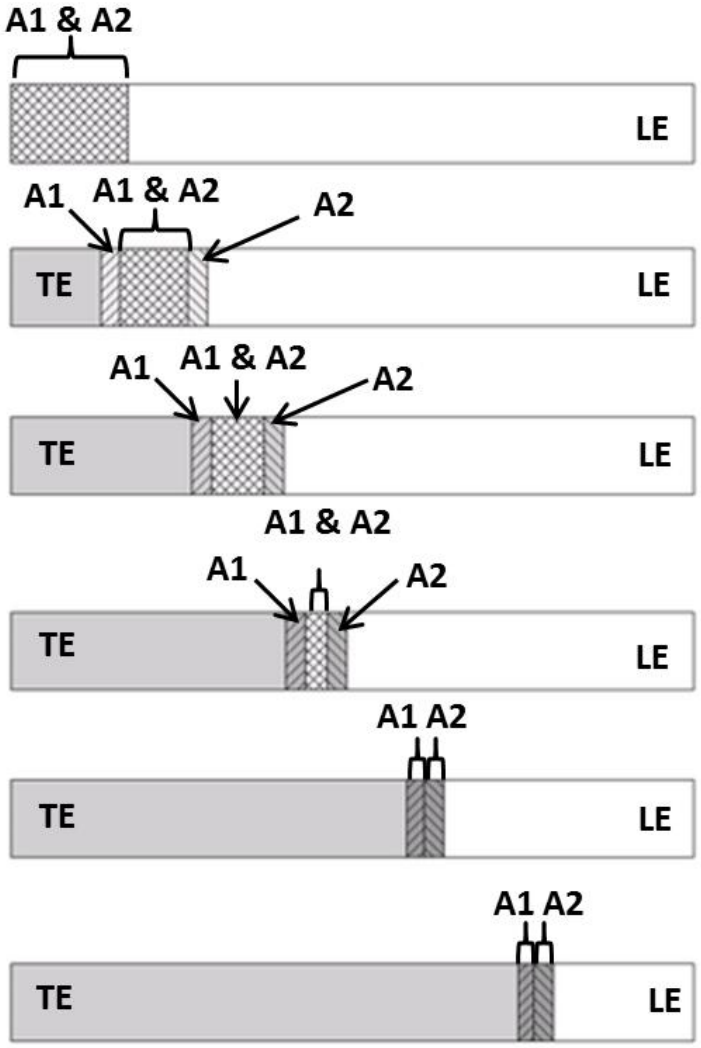


Figure 1-6. Schematic representation of tITP separation of two analytes, A1 and A2. In tITP, a broad zone corresponding to a mixture of analytes A1 and A2 is injected into a capillary containing a leading electrolyte (LE) solution. Following the injection of a trailing electrolyte (TE), separation of A1 and A2 occurs. Progressively darker shading in A1 and A2 zones reflect the concentration of these solutes during tITP.

Alternative techniques enabling increased loadability in CZE include field-amplified sample stacking (FASS), pH-mediated stacking, and chromatographic preconcentration [183]. In FASS, analyte preconcentration occurs as a result of differences in conductivities between sample matrix and the BGE. The sample matrix is typically composed of a low-conductivity solution, whereas the BGE is a high-conductivity solution. As a result, analyte velocities are higher as they migrate through the sample matrix. As analytes reach the sample matrix/BGE interface, their velocities decrease, resulting in a stacking of the analytes into narrow bands [183]. In CZE applications utilizing FASS, studies have reported on a doubling of sample

stacking efficiencies [185] and 1000-fold increases in sensitivity [186], resulting in considerable improvements in loadability over conventional CZE.

In pH-mediated sample stacking, preconcentration occurs as a result of conductivity differences between the sample matrix and BGE, similar to FASS. However, in FASS, the sample matrix is typically composed of water or low-conductivity buffers, whereas in applications involving pH-mediated sample stacking, the sample matrix composition is variable. In pH-mediated sample stacking, a low conductivity sample matrix is created during the course of the preconcentration process. First, the sample is electrokinetically injected into the capillary; in electrokinetic injections, the sample is injected into the capillary by EOF, thus favoring injection of smaller, charged analytes; comparatively, in hydrodynamic injections, sample is injected by pressure, resulting in no bias in the chemical composition of analytes injected into the capillary. Following electrokinetic injection of the sample, anions in the sample matrix are displaced weak acid anions from the BGE. An electrokinetic injection of a plug of strong acid solution is then performed, resulting in the migration of the cations from the acid plug through the sample matrix, thereby titrating the weak acid anions of the BGE in the sample zone. An uncharged, low conductivity sample zone is thus created. Sample cations are then able to migrate quickly through the sample zone until they reach the sample matrix/BGE boundary, where they stack into narrow bands. Using the pH-mediated sample stacking method, 300-fold improvements in limits of detection of biological samples have been reported [187].

An alternative means of increasing loadability in CZE is to use chromatographic preconcentration, where large volume, low concentration samples are deposited onto a solid stationary phase and eluted by smaller volume of the mobile phase [183]. Typically, such a procedure involves offline operation of a solid phase extraction (SPE) column for analyte preconcentration, followed by CZE in the second dimension. However, a variety of interfaces have been developed allowing for online operation of SPE and CZE [183]. Sensitivity improvements of 400- to 500-fold have been observed using online SPE-CZE interfaces [188, 189].

1.8. Data Acquisition and Analysis Strategies for Top-Down Proteomics

Improvements and breakthroughs in MS instrumentation and separation science have not yet been matched by similar progress in bioinformatics for the top-down/intact characterization of proteins. As such, bioinformatics remains a major bottleneck of top-down proteomics [190]. Unlike for bottom-up proteomics, for which a variety of tools exist enabling the routine analysis of individual proteins and complex samples, bioinformatics tools for top-down proteomics remain relatively limited, as analytical instrumentation improvements enabling more routine characterization of intact proteins have only been introduced relatively recently.

Identification of intact proteins in top-down proteomic/intact MS experiments may be achieved from intact protein molecular weight, determined from deconvolution of intact ESI

mass spectra, or from the analysis of fragment ion spectra, obtained from MS/MS experiments. For the analysis of ESI mass spectra of intact proteins, a variety of deconvolution algorithms and software packages are available from major vendors of MS instrumentation [85]. Deconvolution involves the arrangement of ESI-MS peaks into isotopic envelopes, thereby allowing for the identification of charge and molecular weight. Proteins may then be elucidated manually from comparisons of deconvoluted molecular weights to molecular weights of theoretical species.

A variety of algorithms are available facilitating the deconvolution of ESI mass spectra of intact proteins. The first automated computer algorithm reported for the deconvolution of complex ESI-MS spectra of peptides and proteins – THRASH – was developed by the McLafferty group at Cornell University in 2000 [191]. In THRASH, a subtractive peak-finding routine is used for the location and identification of potential isotopic clusters in an ESI mass spectrum. The THRASH algorithm also allows for the calculation of levels of background noise and the determination of charge states using the Fourier transform/Patterson method [191]. A variety of software tools for deconvolution of ESI mass spectra based on the THRASH algorithm are available [192, 193]. An extension of the THRASH algorithm for deconvolution of ESI mass spectra is the Xtract algorithm, adapted for the characterization of ESI mass spectra acquired on Orbitrap instruments using high resolving power settings [44, 194]. However, one drawback to both THRASH and Xtract is the requirement for isotopically-resolved peaks in the ESI mass spectrum [194]. This represents a constraint in the characterization of larger proteins, such as monoclonal antibodies, where the utility of the required resolving power for successful deconvolution by THRASH and Xtract results in decreased S/N. To this end, the ReSpect algorithm was developed, allowing for deconvolution of ESI mass spectra without the constraint of isotopically-resolved peaks. The ReSpect algorithm is essentially a maximum entropy approach to deconvolution of ESI mass spectra and involves first, a spectral deconvolution step, which is then used as a basis to compute the most likely zero-charge spectrum, in addition to associated mass errors [195].

However, fundamental limitations of intact protein characterization from deconvoluted ESI mass spectra involve the inability to distinguish between isobaric proteoforms, proteins in complex mixtures, proteins of unknown composition, the inability to determine the localization of PTMs, or even identify proteoforms composed of numerous combinations of PTMs [44]. For the successful characterization of proteins by intact mass alone, *a priori* knowledge of mixture components is a requisite. Thus, characterization of intact proteins through the deonvolution of intact ESI mass spectra is often limited to relatively simple mixtures and is unsuitable for complex proteomic samples of unknown composition, cellular lysates, or even protein biopharmaceuticals composed of hundreds of proteoforms [44]. For such samples, protein characterization may be achieved from the analysis of fragment ion spectra acquired in top-down proteomic applications.

In top-down analysis involving the fragmentation of intact proteins, a number of data acquisition strategies exist by which precursor ions may be selected for fragmentation. One of

such means involves the specification of *m/z* ranges for fragmentation prior to the MS/MS experiment, also referred to as targeted analysis. Several strategies exist for targeted analysis, such as selected reaction monitoring (SRM), where an ion is selected for fragmentation in the first stage of the MS analysis, and a resultant product ion is then selected for fragmentation in the subsequent stage of the MS experiment [196]. Another strategy is multiple reaction monitoring (MRM), which involves the application of SRM to multiple product ion species from one of more precursor ions [196]. Lastly, in parallel reaction monitoring (PRM), where a precursor ion is selected for fragmentation in the first stage of the experiment and all subsequent product ions are monitored in the second phase [196]. An alternative to targeted analysis is data-dependent acquisition (DDA), which involves on-the-fly selection and fragmentation of individual precursor ions identified in the MS^1 stage *via* a set of user-predefined rules [196]. Lastly, data-independent acquisition (DIA) involves selection and co-fragmentation of all ions within a specified *m/z* range independent of intensity [196].

Numerous software platforms are available for the analysis of top-down proteomic data. These tools are summarized on the website of the Consortium for Top-Down Proteomics (http://www.topdownproteomics.org/resources/software). However, only those tools used in the ensuing chapters are described in this section. Amongst these, ProSitePTM, developed by the Kelleher group at Northwestern University and commercialized by Thermo Fisher Scientific, enables the analysis of top-down MS/MS data and accounts for mass differences arising from polymorphisms, alternate splicing, and PTMs [197]. Another tool for top-down protein analysis is Byonic, developed by Protein Metrics (San Carlos, CA). Byonic enables searches of MS/MS spectra against an inputted database and a user-specified number of variable modifications. Moreover, Byonic enables user-control of the total number of various modification types. For example, the number of occurrences of PTMs such as oxidation and deamidation may be limited to one or two, while allowing for a greater number of potential biological modifications such as phosphorylation and methylation, which tend to cluster together [198].

1.9. Analytical Instrumentation

This section presents the analytical instrumentation used in the ensuing chapters. For MS, these instruments include the Q Exactive Plus MS, the Exactive (extended mass range) EMR MS, and the Orbitrap Elite MS (all Thermo Fisher Scientific). The Q Exactive Plus MS incorporates a quadrupole and the Orbitrap analyzer, while the Orbitrap Elite encompasses a linear trap quadrupole (LTQ), quadrupole, and an Orbitrap analyzer. A schematic of the Q Exactive Plus MS is shown in **Figure 1-7**; as schematic of the Orbitrap Elite is shown in **Figure 1-8**. The Exactive EMR is similar to the Q Exactive Plus; however, instead of a quadrupole enabling mass selection, the Exactive EMR contains a multipole enabling ion transmission only. On the other hand, Exactive EMR instrument mass range is up to *m/z* 20,000, while Q Exactive

Plus upper *m/z* range is limited to *m/z* 6,000.A schematic of the Exactive EMR MS is shown in **Figure 1-9**.

For capillary electrophoresis, the CESI 8000 (SCIEX) instrument was used in all performed experiments. All capillaries used for CZE-MS were part of the sheathless CESI platform from SCIEX and included commercial capillaries with uncharged surfaces, BFS capillaries, and BFS capillaries coated with M7C4I. A general schematic of the CESI technology is shown in **Figure 1-10**.

1.9.1. Q Exactive Plus MS

A schematic of the Q Exactive Plus MS with major components labeled is shown in **Figure 1-7**. In Orbitrap mass spectrometers, ionized analytes are first transmitted through the upstream stacked ring ion guide (S-lens). In the S-lens, ion focusing by radio frequency (RF) electric fields in the low-millibar range allows for a significant decrease of ion losses, resulting in an enhanced ion transmission of up to an order of magnitude [199, 200]. Following transmission through the S-lens, ions are collisionally-cooled in the inject and bent flatapoles and then transmitted through the analytical quadrupole. The quadrupole hyperbolic rods utilizes an RF- and DC-fields to selectively transmit ions within a user-specified *m/z* range, while all other *m/z* ranges are destabilized and filtered out. The Q Exactive Plus MS is equipped with an improved quadrupole, allowing for superior ion transmission in comparison to Q Exactive MS and Orbitrap Elite MS variants. The Q Exactive Plus MS quadrupole allows for the isolation of ions down to an isolation width of 0.4 *Th* at *m/z* 400 [130]. However, one limitation is the inability to select precursor ions exceeding ~2,500 *m/z*. The quadrupole is followed by an exit lens combined with a split lens, which acts as a gate for the incoming ion packet. An octopole then brings ions into an RF-only curved axis quadrupole, referred to as the C-trap. The C-trap is interfaced to a gas-filled HCD cell, which allows for ion fragmentation and collisional cooling [130]. The C-trap allows for accumulation of an ion packet in accordance with the automatic gain control (AGC) function. After the AGC is reached, the ion packet is accelerated to a high kinetic energy and injected tangentially into the Orbitrap analyzer. Ions are then trapped in an orbit around the inner electrode by an electrostatic field and oscillate along its z-axis according to frequencies inversely proportional to the square root of their *m/z*'s. The resultant image current is measured by the outer electrode and converted to a frequency domain from a time domain by enhanced fast Fourier transform (eFT). The identified *m/z* values are then recorded as a mass spectrum. The mass resolving power is *m/z*-dependent and, given a non-decaying transient, is proportional to the product of acquisition time and the frequency of axial oscillation, diminishing as $(m/z)^{1/2}$ [130].

The Q Exactive Plus MS possesses improved resolution as a result of the eFT algorithm, enabling a resolving power of 140,000 at *m/z* 400 from a 1024 ms transient duration [130]. The

Q Exactive Plus MS also allows for intact protein mode operation, which is characterized by the reduced pressure trapping of intact protein molecules in the C-trap of the instrument. For intact proteins, optimum analysis on Orbitrap instruments would involve trapping and collisional cooling at higher pressure and detection in the Orbitrap at lower pressure (enhanced vacuum, less transient decay). However, gas pressures in the C-trap/HCD cell and Orbitrap may not be manipulated independently. Consequently, for intact proteins not exceeding the range of 50 kDa, optimum operation involves low pressure trapping in the C-trap, thereby allowing for detection at an enhanced vacuum in the Orbitrap. For proteins exceeding 50 kDa, optimum operation is attained by trapping ions in the C-trap at standard gas pressure (3.5×10^{-5} mbar).

In the projects outlined in this dissertation, the Q Exactive Plus MS was operated in a variety of modes, including intact protein scans at high and low resolution, direct infusion mode, and MS/MS mode involving top "N" DDA acquisition. For the latter, following initial precursor ion scans, MS^2 scans were performed where a user-specified number of precursor ions were fragmented in the HCD cell; the resultant fragments were then sent back into the C-trap and then into the Orbitrap for analysis. Selection of precursors was performed on-the-fly by a set of user-specified rules relating to signal intensity and *m/z* isolation width.

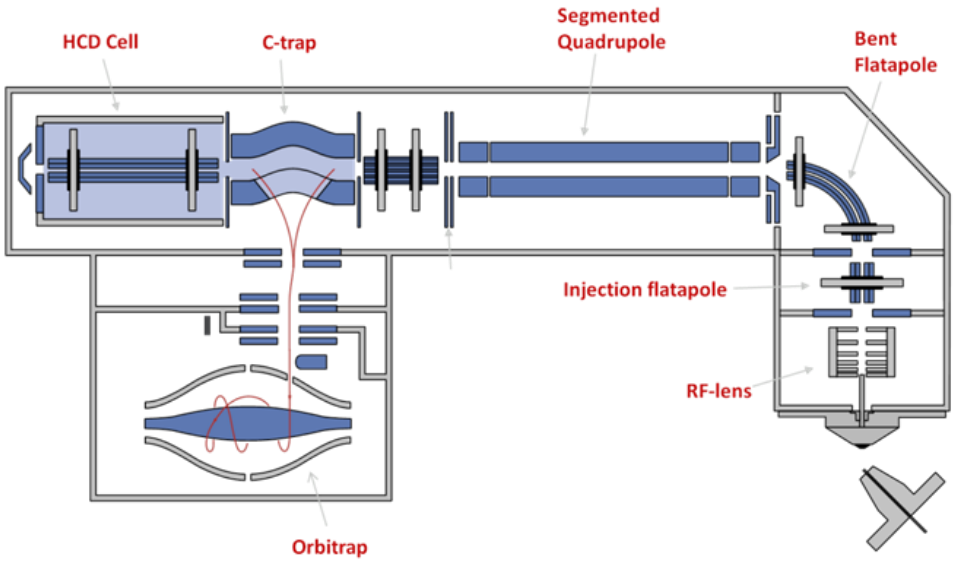


Figure 1-7. Schematic showing different components of the QExactive Plus MS. Schematic reprinted from [130] with permission.

1.9.2. Exactive EMR MS

A schematic of components of the Exactive EMR MS is shown in **Figure 1-8**. While similar in design to the Q Exactive Plus MS, the Exactive EMR MS possesses several key modifications. First, instead of a quadrupole capable of transmitting ions up to *m/z* 6,000, as in the QExactive Plus MS, the Exactive EMR MS possesses a multipole and allows for the transmission and ion trapping of species up to *m/z* 20,000. This is a key modification allowing for characterization of large proteins and multimeric protein complexes under native conditions, where the intact protein species possess considerably less charge than under standard denaturing conditions. Another key feature of the Exactive EMR MS is the ability to control gas pressure in the HCD cell. The ability to fine-tune collisional gas pressure in the HCD cell allows for the optimization of trapping and collisional cooling of large intact proteins, thereby improving S/N dramatically for large analytes. Several recent studies have involved analysis of intact mAbs, macromolecular assemblies of non-covalent protein complexes approaching molecular weights of 1.0 MDa, and even intact ribosomal subunits under native conditions using the Exactive EMR MS [7, 134, 201]. Lastly, the Exactive Plus EMR allows for all-ion fragmentation (AIF) in the HCD cell. In AIF, all detected precursors are fragmented *via* HCD, resulting in the generation of *b*- and *y*-type fragment ions. Precursor fragmentation occurs without on-the-fly selection and is unrelated to precursor ion intensity.

In projects outlined in this dissertation, the Exactive EMR MS was used for the analysis of mixtures of protein standards containing numerous non-covalent assemblies of protein complexes.

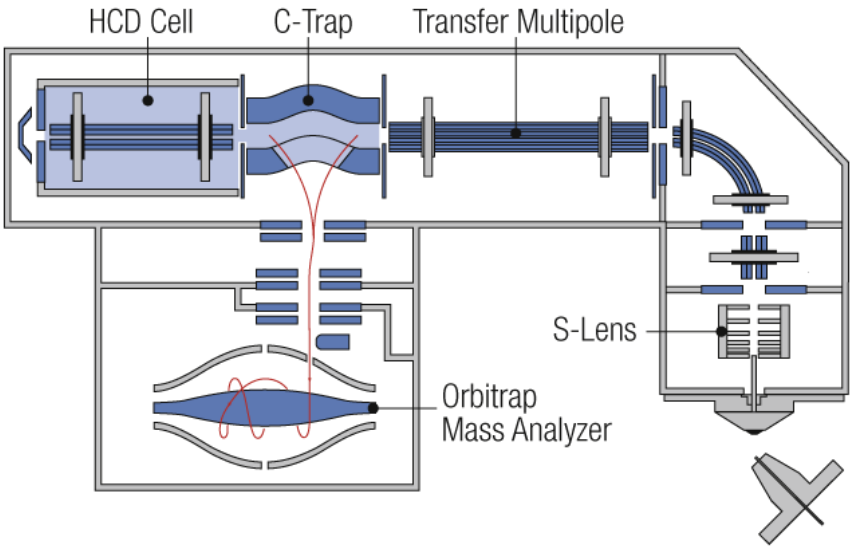


Figure 1-8. Schematic showing components of the Exactive EMR MS. Figure reprinted from [134] with permission.

1.9.3. LTQ Orbitrap Elite MS

As shown in the schematic in **Figure 1-9**, the LTQ Orbitrap Elite MS is a hybrid instrument composed of two mass analyzers – the LTQ and the Orbitrap. Various instrument components such as the S-lens, quadrupole, C-trap, etc., were presented in the above sections and will not be described here. In the LTQ, ion packets are trapped by a combination of a quadrupole RF and DC voltages. These voltages are applied to the ion trap to filter out all but the desired ion populations. Ion packets are collected until a user-controlled capacity is reached. Populations of ions are then be ejected sequentially in the radial directions by the application of an RF voltage ramp. Ejected ions strike an electron multiplier, which records signals proportional to the number of charged species in the ejected ion population.

The Orbitrap Elite MS enables detection of intact protein in low pressure mode, or "protein mode". In this mode of operation, instead of trapping and collisionally cooling intact proteins in the C-trap, proteins are trapped and collisionally cooled in the HCD cell at lower pressures than normal. This allows for a superior vacuum in the Orbitrap, which allows for longer transient times of intact protein analytes. However, this mode of analysis is only useful up to approximately 50 kDa species; for species exceeding this molecular weight, optimum signal is established under higher trapping pressure conditions. An additional important feature for intact analysis is enhanced resolving power (480,000 at *m/z* 200), allowing for resolution of isobaric species of small to medium-sized proteins.

The Orbitrap Elite MS allows for the fragmentation of ions in the LTQ *via* CID or ETD. These ions may be analyzed in the LTQ and/or sent to the C-trap, and then into the Orbitrap for detection. Alternatively, the instrument allows for HCD fragmentation in the HCD cell, in which case the resultant product ions are sent first into the C-trap, and then into the Orbitrap. Finally, the instrument allows for operation of EThcD, in which ETD fragmentation occurs in the LTQ and the resultant fragment ions are then fragmented *via* HCD in the HCD cell, after which they are sent to the C-trap and the into the Orbitrap for analysis [202].

In the projects outlined in this dissertation, the Orbitrap Elite MS was operated in a variety of modes for the characterization of single proteins and complex mixtures. Both, standard and intact protein modes were used, along with ETD and HCD fragmentation.

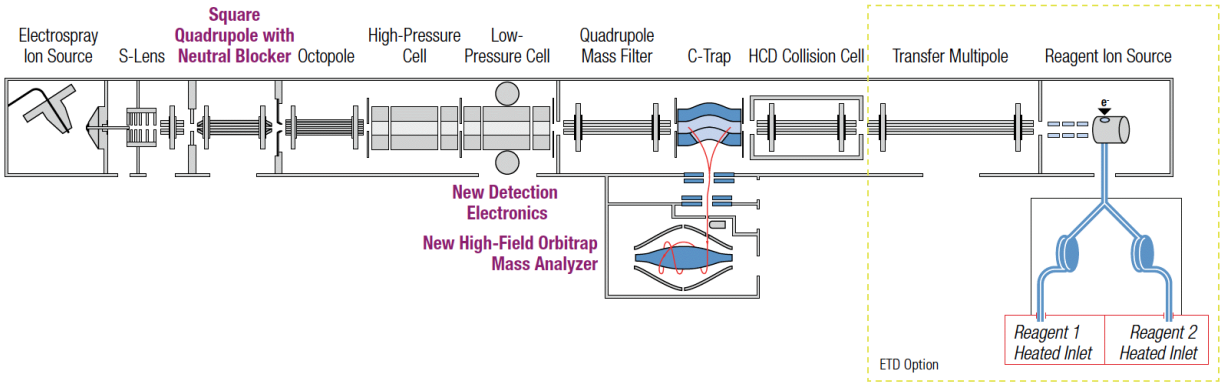


Figure 1-9. Schematic showing components of the Orbitrap Elite MS, including the ETD cell. Figure reprinted from [202] with permission.

1.9.4. Sheathless CESI

For all CZE-MS experiments outlined in projects in this dissertation, the commercial CESI platform (SCIEX) was used. **Figure 1-10** shows a schematic of the interface capillary cartridge (OptiMS) and CESI 8000 instrument. The sheathless interface, as described in previous sections of this dissertation, is composed of 2 capillaries: A separation capillary and a conductive capillary. The separation capillary contains the BGE and is used for CZE operation. The separation capillary contains a HF acid-etched, porous sprayer tip with a constant inner diameter (30 μm) and a reduced outer diameter at the emitter end. Moreover, the etched end contains micropores, while allow for the diffusion of small ions without dilution in the separation capillary. The conductive capillary is used to deliver ions to the conductive liquid reservoir, which houses a portion of the terminal end of the separation capillary. Flow of ions from the conductive liquid reservoir into the separation capillary allows for a closure of the circuits for necessary for both, capillary electrophoresis and ESI. Given that no dilution at the CZE/ESI junction occurs with this interface, depending on the specific capillary coating and mode of operation, typical flow rates for this device range from 10's to 100's of nL/min [67].

For operation of sheathes CESI-MS, a commercial CESI 8000 instrument was used. The CESI 8000 module allows for operation of CZE as a front-end separation technique to ESI-MS. The module allows for housing of OptiMS cartridges and possesses a power supply for CZE and ESI-MS. The module also allows for delivery of supplemental pressures to both, separation and conductive capillaries in OptiMS cartridges. The instrument also includes an integrated module for ESI-MS, containing adapter housing for the OptiMS sprayer. All online CZE-MS experiments performed in projects outlined in this dissertation were performed on the CESI 8000 module.

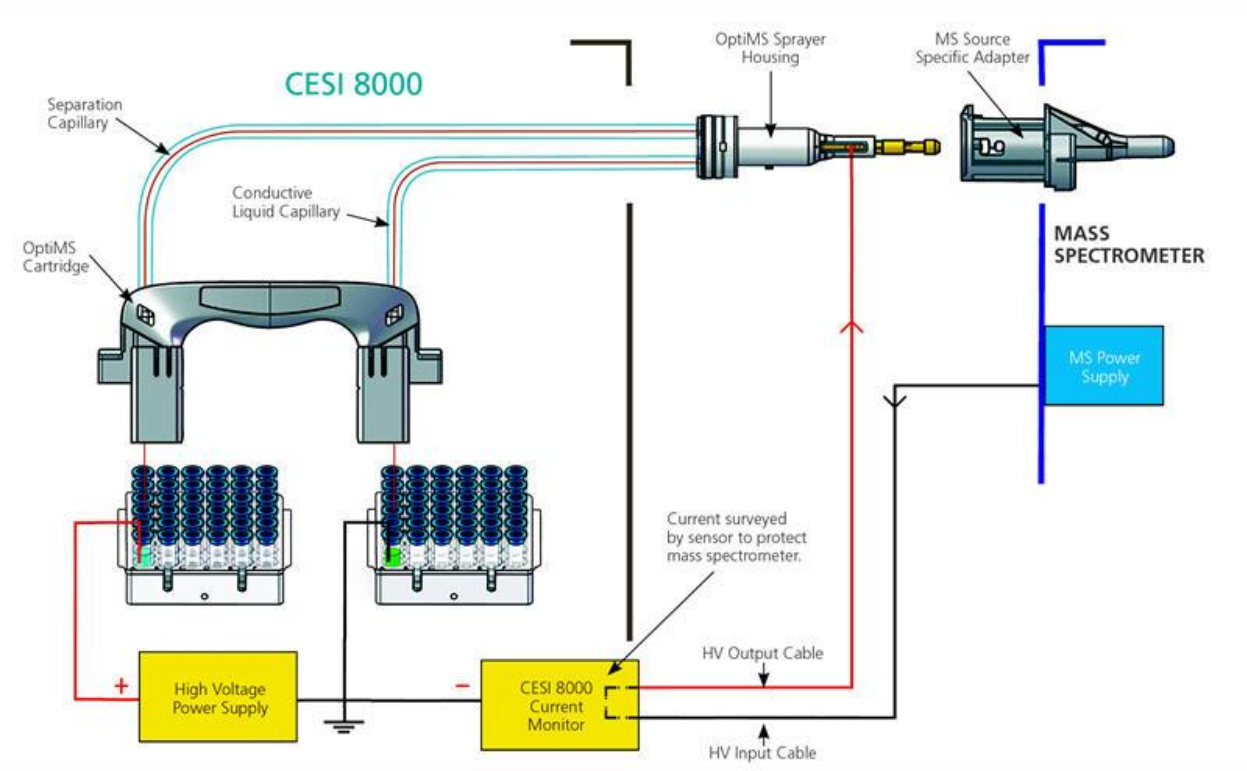


Figure 1-10. Schematic representation of the CESI technology. An OptiMS cartridge is composed of a separation capillary with an etched sprayer tip for nanoESI, as well as a conductive capillary used to deliver conductive solution. The CESI 8000 instrument is used to generate voltage and pressure for CZE operation. The OptiMS cartridge is inserted into the MS source adapter, connected to the mass spectrometer. This schematic was reprinted form (https://sciex.com/ce-features-and-benefits/cesi-technology).

Design flexibility of the OptiMS cartridges allow for the application of a variety of capillary coatings. In projects outlined in this dissertation, BFS capillaries were modified with cross-linked polyethylenimine (cPEI) and used for the analysis of an intact protein pharmaceutical with a single glycosylation site (**Chapter 2**). BFS capillaries were also modified with reagents resulting in positively-charged coatings at acidic conditions (1-(4-iodobutyl) 4-aza-1-azoniabicyclo[2,2,2] octane iodide (M7C4I)). M7C4I-coated capillaries were then used for analyses of intact mAbs and middle-down products of mAbs (**Chapter 3**). Unmodified BFS capillaries were used for the separation of intact mAbs under native conditions, as well as for the separation of mixtures of native proteins and protein complexes (**Chapter 4**). Capillaries with uncharged surfaces, which are a recently-introduced commercial product from SCIEX, were used for native separations of intact monoclonal antibodies. These capillaries were also used for the analysis of mixtures of native proteins and protein complexes, as well as for the analysis of

the protein extract from the *E. coli* ribosome under both, native and denaturing conditions (**Chapter 4**).

1.10. Conclusions

This chapter has introduced a variety of approaches used for characterization of biotherapeutics, such as mAbs, as well as proteins and protein complexes in biological mixtures. The complexity of protein analytes necessitates characterization *via* high-throughput, high-sensitivity capabilities offered by mass spectrometry. Moreover, biological heterogeneity necessitates front-end separation techniques that could effectively resolve intact proteins, including individual proteoforms, prior to characterization by MS. Advances over the past several decades in development of nano-flow liquid-phase separation techniques, MS instrumentation, and bioinformatics strategies for large intact proteins have enabled top-down proteomics of samples of increasing complexity. Analysis of proteins under native conditions represents an extension of intact/top-down analysis and enables characterizing of proteins and complexes of proteins in their native states. As an alternative to the more conventional LC-based methods for the analysis of intact proteins, CZE offers a complementary mode of analysis based on a different mechanism of separation. Furthermore, CZE offers the advantage of online compatibility with nanoESI-MS, enables high sensitivity, high resolution separations, and allows for liquid-phase separations under native conditions.

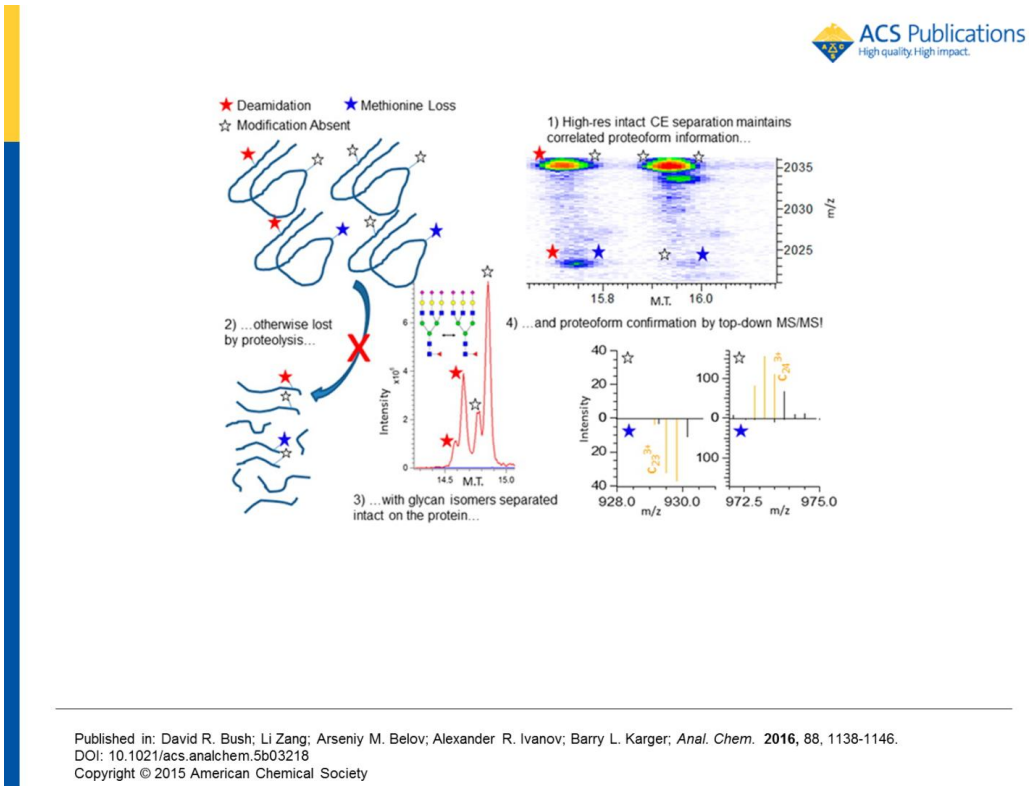
The ensuing chapters in this dissertation will describe several projects performed on the sheathless CESI – MS platform using a variety of different capillary coatings for CZE coupled to various Orbitrap instruments. **Chapter 2** will present the analysis of proteoforms of intact rhINF-β1 (Avonex; Biogen, Cambridge, MA) on a sheathless cPEI – Orbitrap Elite system under denaturing conditions. **Chapter 3** presents the characterization of a biotherapeutic mAb, currently in phase III clinical trials, in both, an intact and middle-down format using the sheathless CESI-MS system with positively-charged capillaries (M7C4I) coupled to the QExactive Plus MS. Lastly, **Chapter 4** will present the development of an analytical platform for the characterization of intact proteins and multimeric protein complexes under native conditions by online CZE-MS. Both sheathless BFS and capillaries with uncharged coatings coupled to the Exactive EMR and/or Q Exactive Plus were used in this application.

Chapter 2: Ultra-High Resolution CZE-MS for Quantitative Characterization of Intact Biopharmaceutical Proteins: Proteoforms of Interferon-β1

David R Bush[1], Li Zang[2], Arseniy M Belov[1], Alexander R Ivanov[1], and Barry L Karger[1*]

[1]Barnett Institute, Northeastern University, 360 Huntington Ave, Boston MA 02115

[2]Analytical Development Department, Biogen, Cambridge MA 02142



Published in: David R. Bush; Li Zang; Arseniy M. Belov; Alexander R. Ivanov; Barry L. Karger; *Anal. Chem.* **2016,** 88, 1138-1146.
DOI: 10.1021/acs.analchem.5b03218
Copyright © 2015 American Chemical Society

***My specific contributions to the work presented in this chapter included assistance in experimental design, enzymatic reactions of rhINF-β1 and characterization by CZE-MS, and data analysis.**

2.0. Abstract

New and improved methods are required for the enhanced characterization of complex biopharmaceuticals, especially those with charge and glycan heterogeneity. The use of ultra-high resolution separation and mass spectrometry (MS) analysis of intact proteoforms can contribute significantly to the characterization of such proteins, many of which are glycoproteins. Here, we report on the use of capillary zone electrophoresis (CZE) coupled *via* a commercial CESI sheathless interface to an Orbitrap Elite MS for the intact analysis of recombinant human interferon-β1 (Avonex, rhIFN-β1), a biopharmaceutical with complex glycosylation at a single N-linked site. Using a cross-linked polyethylenimine-coated capillary column, column efficiencies between 350,000 and 450,000 plates were produced, allowing for separation based on charge and subtle hydrodynamic volume differences. A total of 138 proteoforms were found, and 55 were quantitated. Charge species due to deamidation and sialylation, among others, were resolved by CZE. Given the very high column efficiency, isobaric positional isomers of sialic acid on biantennary glycan antennae were resolved and confirmed by the digestion of Avonex with sialidase. Further, triantennary positional isomers (antenna on α(1-3) or α(1-6) arms) were separated as observed by sialidase and galactosidase digestion. Proteoforms of the N-terminal cleavage of methionine were detected by precursor molecular weight and top-down ETD analysis of the reduced form of the protein. A total of more than 50 proteoforms have been found. Furthermore, quantitative analysis revealed potential correlations between the methionine loss with the relative amount of the deamidated form, as well as the extent of deamidation with glycan structure. We demonstrate that ultra-high resolution CZE separation of intact glycoprotein species coupled to MS has significant potential for the in-depth characterization and quantitative analysis of biopharmaceutical proteoforms.

2.1. Introduction

Protein biopharmaceuticals belong to an important and continually growing sector of the pharmaceutical industry. In 2012, almost half of the new drugs approved by the FDA and 5 out of the 10 top-selling drugs were biopharmaceuticals [203]. In contrast to synthetic small molecule drugs, biopharmaceuticals are produced from living organisms, leading to complex mixtures of proteoforms [204]. Sources of heterogeneity can include, among others, a variety of glycoforms, disulfide bond scrambling, oxidation, deamidation, sequence mutation and N- and C-terminal alteration [205]. Detailed characterization and quantitation of proteoforms is important, as specific modifications can affect the efficacy and safety of the drug product. For example, oxidation has been linked to protein aggregation, which has potential for immunogenicity [205-207]; sialic acid content of glycans can improve [205, 208] or lower efficacy [205, 209]; IgG fucosylation can limit Fc receptor binding [205, 210]; and deamidation can reduce activity [205, 211-213].

Common methods currently employed for analyzing biopharmaceutical products include peptide mapping for determining sequence, disulfide linkage and post-translational modifications (PTMs) at the peptide level, and glycan profiling. While the information provided by these bottom-up approaches is valuable, there are limitations: (1) Multiple sample preparation steps can introduce modifications in the protein product [214-220], (2) specific proteoforms, i.e. the exact modifications of the intact protein molecules, are not identified or quantitated. Intact MS (MS^1 of the undigested glycoprotein) and top-down MS [221-227] (fragmentation of the intact protein in the mass spectrometer) can be conducted with minimal sample preparation and can reveal individual proteoform structures. This latter information is important since the specific proteoforms determine protein function. Additionally, the correlated presence of multiple modifications on individual proteoforms can potentially lead to additional functional information.

The analysis of intact proteoforms is a growing area of interest as mass spectrometers have become ever higher resolving and sensitive. The need for high resolution separation prior to MS is essential where multiple proteoforms of biopharmaceuticals, some of which are isobaric or differ by only 1 or a few Daltons and/or are in low abundance. Efforts are being made to develop LC columns suitable for high resolution intact protein separation [228]. At the same time, capillary zone electrophoresis (CZE) represents an alternative and complementary separation approach based on charge and hydrodynamic volume, with the potential for ultra-high efficiency [229-231].

While the coupling of CE to MS has developed more slowly than that of LC to MS, recent strides have been made using sheath flow [232, 233] and sheathless interfacing [234]. A commercially available sheathless CESI source combines high resolution separation of CE with high resolution mass spectrometry. Because there is no sample dilution, as occurs in sheath-flow systems, the sheathless interface is more sensitive [235-237]. We [238] and others [229, 239] have demonstrated the potential of high resolution CZE coupled to high resolution MS to analyze intact complex glycoproteins. Recently, groups have shown that CZE-MS of intact proteins (MS^1) can be used to analyze complex protein mixtures [233, 240-242], but much development remains in this area as well as top-down analysis of glycoproteins.

In this work, we apply intact, top-down CZE-MS to Avonex, a commercial recombinant human interferon-β1a (rhIFN-β1) biopharmaceutical expressed from CHO cells used to treat multiple sclerosis [243]. RhIFN-β1 has known modifications, such as complex glycosylation, deamidation and succinimide, as well as oxidation [208, 229, 244, 245]. Here, CZE-MS is demonstrated to successfully characterize and quantitate individual molecular proteoforms with high-resolution accurate mass analysis. We use a capillary positive coating, generating between 350,000 and 450,000 total plates, with high migration reproducibility, based on apparent mobility of solutes. Although intact protein analysis of 80 proteoforms of rhIFN-β1 by CZE-MS has been reported [229], in the present study, we have significantly improved the separation and analysis of the proteoforms, leading to the determination of 138 individual proteoforms, some of

which were identified *via* the selective digestion by several exoglycosidases. Besides elucidating the main glycoforms of Avonex, we have observed intact separation of glycan positional isomers of the biantennary mono-sialic acid glycan as well as triantennary glycans with the third antenna attached to either the α(1-3) or α(1-6) arm. The spectra of 63 forms were deconvoluted, and 55 have been quantitated. We show that the extent of N-terminal methionine loss (des-1) and deamidation appear to be correlated, as well as the extent of deamidation with the glycan structure. The results of this work demonstrate the significant potential of high resolution CZE-MS for top-down analysis of glycosylated biotherapeutics.

2.2. Experimental

2.2.1. Chemicals and Materials

TCEP-HCl was purchased from Pierce Biotechnology (Rockford, IL), and Bio-gel P6 size exclusion columns were obtained from Bio-Rad (Hercules, CA). Trimethoxysilylpropyl modified polyethylenimine (PEI) in isopropanol (50%) was purchased from Gelest (Morrisville, PA). Avonex was obtained from Biogen Idec. Glyko sialidase A, β(1-4,1-6) galactosidase, and reaction buffer (250 mM sodium phosphate, pH 6.0) was obtained from Prozyme (Hayward, CA). 18.2 MΩ water was produced in house on a Milli-Q Integral water purification system (EMD-Millipore, Billerica MA). Acetic acid, ammonium hydroxide and 1,4-butanediol diglycidyl ether (BUDGE) were purchased from Sigma-Aldrich (St. Louis, MO).

2.2.2. Buffer exchange and TCEP reduction of rhIFN-β1

Biogel P6 size exclusion spin columns were equilibrated by rinsing with 500 µL of Milli-Q water three times. 75 µL of Avonex (0.27 mg/mL rhIFN-β1, 20 mM sodium acetate, 150 mM arginine HCl, pH 4.8) was loaded on the column and centrifuged at 1000 g for 4 minutes. Eluent was concentrated by a Speed-Vac to approximately 1.2 mg/mL for top-down experiments and 2 mg/mL for low-injection volume experiments. For top-down CZE-MS, TCEP was added to the buffer exchanged sample to a final concentration of 5 mM and allowed to react at room temperature for 30 minutes prior to analysis.

2.2.3. Enzymatic removal of sialic acid and galactose

For each experiment, Avonex (2 x 75 µL) was buffer exchanged, as described above, and concentrated to 12 µL by Speed-Vac. Then, 4 µL of reaction buffer and 2 µL of either sialidase

A, β(1-4, 1-6) galactosidase or both enzymes were added and incubated for 3 hours at 37 °C. After incubation, the reaction was refrigerated at 4 °C prior to CZE-MS.

2.2.4. PEI coating and cross-linking procedure

An etched porous OptiMS bare fused silica capillary (90 cm, 150 μm OD, 30 μm ID) from SCIEX (Brea, CA) was coated with cPEI as follows [246]. While the porous tip was immersed in Milli-Q water, the same water was flowed through both the separation and conductive capillaries for 10 min at 100 psi. Using again 100 psi, 0.1 M NaOH was pumped through the column for 20 min, followed by 0.1 M HCl for 10 min, and Milli-Q water for 10 min. The porous tip was then immersed in MeOH, and MeOH was washed at 100 psi through the separation capillary for 20 min and the conductive capillary for 5 min. Immediately following the MeOH rinse, a 10 % v/v PEI solution, made by diluting the commercial PEI solution with MeOH, was flowed through the separation capillary for 150 minutes at 20 psi. The coating solution remained in the column for 14-16 hrs, with the inlet in the coating solution and the etched tip outlet in methanol, after which the column was rinsed with methanol at 50 psi for 5 minutes. The coated capillary was cross-linked by passing a 10% v/v solution of BUDGE in dioxane for 5 hours at 20 psi. The inlet of the separation capillary was then carefully attached to an N_2 tank *via* a microtight union (Idex Health and Science, Oak Harbor, WA). N_2 was pressurized to 100 psi, and the cartridge was placed in an oven equilibrated at 80° C for 1 hour. Following the cross-linking, the separation capillary was rinsed at 50 psi with MeOH for 45 min, Milli-Q water for 45 min, 1 M NaCl for 10 min, and BGE (background electrolyte) for 90 min. With the exception of the cross-linking in the oven, all steps were carried out at room temperature.

2.2.5. CZE-MS

CZE was conducted on a CESI 8000 instrument (SCIEX), coupled to an LTQ Orbitrap Elite mass spectrometer (ThermoFisher Scientific, San Jose, CA) through an adapter for the OptiMS sprayer tip that is attached to the commercial nESI source supplied with the Orbitrap Elite MS. Except where stated, the BGE was 3% acetic acid pH 2.5, and the separation voltage was 20 kV in reverse polarity. Between runs, cPEI capillaries were rinsed with 1M NaCl for 2 minutes at 50 psi, followed by rinsing the capillary and conductive line rinsed for 7 and 2 minutes, respectively, with BGE at 50 psi. Injection conditions were either 1 psi for 10 s or 0.5 psi for 5 s, the specific condition indicated in the figure captions. After injection, a small plug of BGE was injected into the separation capillary at 0.5 psi for 5 seconds. The nESI conditions consisted of the sprayer tip positioned approximately 5 mm from the transfer capillary of the Orbitrap Elite MS with an electrospray potential of 1.6 kV.

For intact MS^1–only scans, the resolution was 120,000 at 400 *m/z* with 1 microscan and an AGC target of 2e6. For the repeatability experiments, the MS resolution was 30,000 at 400 *m/z*. For top-down MS^2 scans, the resolution was 60,000 at 400 *m/z*. For top-down ETD fragmentation, 5 microscans were acquired with an AGC target of 1e6, a reaction time of 25 ms and a target reagent current of 1e6. For all experiments, the Orbitrap Elite MS was operated in the "protein mode" with a reduced pressure difference between the HCD cell and the C-trap of $10e^{-10}$ mTorr.

2.2.6. Data Processing

Top-down data analysis was performed using both the high resolution MS^1–only scans and MS^2 scans as follows: (1) Monoisotopic masses and charge states of a peak of interest were determined from high resolution intact MS^1 by averaging the spectra over a peak and exporting a deconvoluted *.raw* file using the XTRACT algorithm in Xcalibur (v. 3.0.63, ThermoFisher Scientific). (2) Extracted ion electropherograms (XIEs) of targeted m/z in the low resolution MS^1 segment of the top-down scan were used to locate the area over which to average targeted MS^2. The resulting averaged spectra were then exported into a *.raw* file that was then converted to an *.mgf* file using the MSconvert utility of ProteoWizard (v. 3.0.4472) [247]. (3) The monoisotopic *m/z* and charge state, determined from the intact MS^1 scan, were applied to the header of the individual *.mgf* file. (4) Each *.mgf* file was analyzed by Byonic (v. 2.0-25, Protein Metrics, San Carlos, CA) against the sequence of recombinant human interferon β1. ETD and HCD spectra were searched with a 5 Da precursor mass tolerance, 15 ppm fragment mass tolerance, 1 deamidation (N) and 1 glycan (N) from the comprehensive human N-glycan database ('Human 360.txt') supplied with the Byonic.

2.3. Results and Discussion

CZE separation of numerous rhIFN-β1 proteoforms was achieved under acidic conditions using a cPEI-coated capillary coupled to a high resolution Orbitrap Elite MS. In the following, we first examine the CZE platform and analytical procedures that lead to ultra-high efficiency and precise reproducibility. Next, we detail the various proteoforms determined and the use of exoglycosidases and MS^2 to add and validate structures. We then turn to quantitation of the proteoforms and suggest that some modifications appear to be correlated.

2.3.1. CZE-MS Characterization

We separated the proteoforms of Avonex using acidic buffer conditions (3% acetic acid, pH 2.5) in order to protonate the various proteoforms for favorable CZE separation and electrospray ionization. We covalently coated the bare fused silica capillary with cPEI in contrast to dynamic polymeric coatings such as polybrene or hydroxypropylmethylcellulose, which require frequent recoating and can potentially contaminate the MS source [248-251]. Although the coating is positively charged and rhIFN-β1 has up to four sialic acids, the net charge of the various proteoforms is positive at pH 2.5, reducing the possibility of protein interaction with the coating. In addition, given the positively-charged surface, the EOF migrated from the negatively charged electrode at the inlet to the porous ESI tip in contact with the positively charged electrode. The net positively-charged proteins at pH 2.5 will electrophorese toward the negative electrode at the inlet but, at the same time, be pulled toward the outlet by the electroosmotic flow. The capillary is thus effectively lengthened with the slowest electrophoretically-migrating species (i.e. least positive and largest) eluting first from the capillary and fastest eluting last.

The run-to-run repeatability of rhIFN-β1 separation on a cPEI capillary is shown in the **Figure 2-1**, for four repeated injections. The RSD of the migration time is 0.5% and, in general, the RSD remained below 2% for run-to-run replicates as shown in **Table 2-1**. However, the EOF varies over the number of injections or column-to-column as seen with the migration time of the neutral marker. To compensate for this, we have determined the apparent mobililty, μ_{APP}, of the various proteoforms (see **Equations 1-3** for the determination of μ_{APP}, μ_{FS}, and μ_{EOF}). **Table 2-1** lists the proteoforms for which the various mobilities have been determined for run-to-run, day-to-day, and column-to-column. As can be seen, the reproducibility of the apparent mobilities of individual proteoforms is excellent, less than 0.6% RSD for different columns over multiple days. **Figure 2-2** is an ion density map showing the various proteoforms to be discussed below.

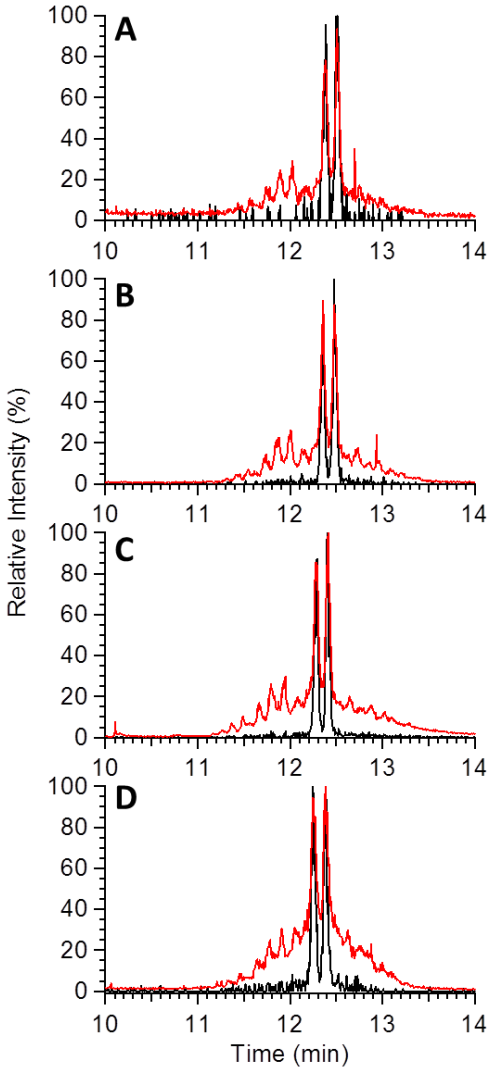


Figure 2-1: Repeatability of Avonex separation on a cPEI-coated capillary. The red traces are total ion electropherograms and the corresponding black traces the extracted ion electropherograms for the N54120 glycoform with two peaks each for the earlier migrating deamidated (*) and later migrating unmodified proteoforms (**).

Capillary and Injection Date	Acetaminophen			
	t_o (min)	t_o RSD	μ_{EOF} (m^2/Vs)	μ_{EOF} RSD
Capillary 1, Day 2	9.08	0.53	-8.28E-08	-0.58
	9.10		-8.26E-08	
	9.02		-8.34E-08	
Capillary	9.13		-8.23E-08	

1, Day 3	9.14		-8.23E-08				
	9.15		-8.22E-08				
Capillary 2, Day 1	9.55	0.98	-7.84E-08	-1.07			
	9.69		-7.72E-08				
	9.75		-7.67E-08				
Capillary 2, Day 2	9.81		-7.62E-08				
	9.76		-7.66E-08				
	9.79		-7.63E-08				
Capillary and Injection Date	**N54120 (N→D)**						
	t_e (min)	t_e RSD	μ_{FS} (m^2/Vs)	μ_{APP} (m^2/Vs)	Average μ_{APP}	% Error from Average	μ_{APP} RSD
Capillary 1, Day 2	15.62	1.09	3.64E-08	3.64E-08	3.64E-08	0.63	0.50
	15.71		3.65E-08	3.65E-08			
	15.37		3.62E-08	3.62E-08			
Capillary 1, Day 3	15.65		3.60E-08	3.60E-08			
	15.82		3.65E-08	3.65E-08			
	15.84		3.64E-08	3.64E-08			
Capillary 2, Day 1	17.16	2.31	3.64E-08	3.64E-08	3.66E-08	0.13	0.63
	17.77		3.66E-08	3.66E-08			
	17.99		3.66E-08	3.66E-08			
Capillary 2, Day 2	18.39		3.70E-08	3.70E-08			
	17.96		3.65E-08	3.65E-08			
	18.10		3.66E-08	3.66E-08			
Capillary and Injection Date	**N54120**						
	t_e (min)	t_e RSD	μ_{FS} (m^2/Vs)	μ_{APP} (m^2/Vs)	Average μ_{APP}	% Error from Average	μ_{APP} RSD
Capillary 1, Day 2	15.82	1.09	3.71E-08	3.71E-08	3.70E-08	0.89	0.49
	15.93		3.72E-08	3.72E-08			
	15.59		3.69E-08	3.69E-08			
Capillary 1, Day 3	15.86		3.67E-08	3.67E-08			
	16.04		3.71E-08	3.71E-08			
	16.06		3.71E-08	3.71E-08			
Capillary 2, Day 1	17.54	2.32	3.73E-08	3.73E-08	3.75E-08	0.53	0.58
	18.17		3.76E-08	3.76E-08			
	18.39		3.76E-08	3.76E-08			
Capillary 2, Day 2	18.80		3.79E-08	3.79E-08			
	18.35		3.74E-08	3.74E-08			
	18.51		3.75E-08	3.75E-08			
Capillary	**N54110 (N→D) 1**						

and Injection Date	t_e (min)	t_e RSD	μ_{FS} (m^2/Vs)	μ_{APP} (m^2/Vs)	Average μ_{APP}	% Error from average	μ_{APP} RSD
Capillary 1, Day 2	16.21	1.11	3.82E-08	3.82E-08	3.81E-08	0.36	0.51
	16.32		3.83E-08	3.83E-08			
	15.96		3.81E-08	3.81E-08			
Capillary 1, Day 3	16.23		3.78E-08	3.78E-08			
	16.43		3.82E-08	3.82E-08			
	16.45		3.82E-08	3.82E-08			
Capillary 2, Day 1	17.94	1.98	3.83E-08	3.83E-08	3.83E-08	0.19	0.58
	18.62		3.86E-08	3.86E-08			
	18.83		3.85E-08	3.85E-08			
Capillary 2, Day 2	18.81		3.80E-08	3.80E-08			
	18.77		3.83E-08	3.83E-08			
	18.97		3.85E-08	3.85E-08			
Capillary and Injection Date	**N54110 1**						
	t_e (min)	t_e RSD	μ_{FS} (m^2/Vs)	μ_{APP} (m^2/Vs)	Average μ_{APP}	% Error from average	μ_{APP} RSD
Capillary 1, Day 2	16.42	1.14	3.88E-08	3.88E-08	3.87E-08	0.58	0.52
	16.54		3.89E-08	3.89E-08			
	16.16		3.87E-08	3.87E-08			
Capillary 1, Day 3	16.44		3.84E-08	3.84E-08			
	16.65		3.88E-08	3.88E-08			
	16.67		3.88E-08	3.88E-08			
Capillary 2, Day 1	18.25	2.34	3.90E-08	3.90E-08	3.92E-08	0.51	0.50
	18.93		3.92E-08	3.92E-08			
	19.12		3.91E-08	3.91E-08			
Capillary 2, Day 2	19.58		3.95E-08	3.95E-08			
	19.12		3.90E-08	3.90E-08			
	19.28		3.91E-08	3.91E-08			
Capillary and Injection Date	**N54110 (N→D) 2**						
	t_e (min)	t_e RSD	μ_{FS} (m^2/Vs)	μ_{APP} (m^2/Vs)	Average μ_{APP}	% Error from average	μ_{APP} RSD
Capillary 1, Day 2	16.59	1.17	3.93E-08	3.93E-08	3.92E-08	0.45	0.54
	16.7		3.94E-08	3.94E-08			
	16.32		3.91E-08	3.91E-08			
Capillary 1, Day 3	16.6		3.88E-08	3.88E-08			
	16.83		3.93E-08	3.93E-08			
	16.85		3.93E-08	3.93E-08			

Capillary 2, Day 1	18.38	2.41	3.93E-08	3.93E-08	3.95E-08	0.33	0.50
	19.10		3.96E-08	3.96E-08			
	19.33		3.95E-08	3.95E-08			
Capillary 2, Day 2	19.75		3.98E-08	3.98E-08			
	19.29		3.94E-08	3.94E-08			
	19.46		3.95E-08	3.95E-08			
Capillary and Injection Date	**N54110 2**						
	t_e (min)	t_e RSD	μ_{FS} (m^2/Vs)	μ_{APP} (m^2/Vs)	Average μ_{APP}	% Error from average	μ_{APP} RSD
Capillary 1, Day 2	16.83	1.16	3.99E-08	3.99E-08	3.98E-08	0.53	0.53
	16.94		4.00E-08	4.00E-08			
	16.54		3.97E-08	3.97E-08			
Capillary 1, Day 3	16.83		3.94E-08	3.94E-08			
	17.06		3.99E-08	3.99E-08			
	17.07		3.99E-08	3.99E-08			
Capillary 2, Day 1	18.70	2.45	4.00E-08	4.00E-08	4.02E-08	0.42	0.49
	19.44		4.03E-08	4.03E-08			
	19.69		4.02E-08	4.02E-08			
Capillary 2, Day 2	20.12		4.05E-08	4.05E-08			
	19.67		4.01E-08	4.01E-08			
	19.78		4.01E-08	4.01E-08			

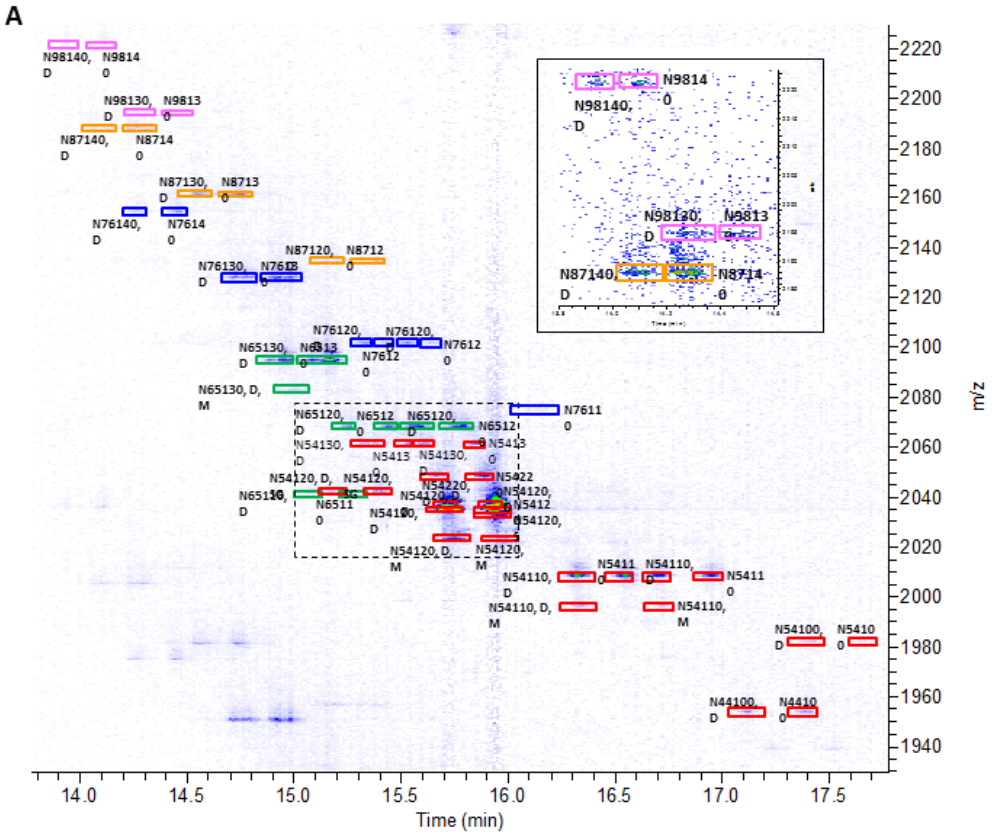
Table 2-1B

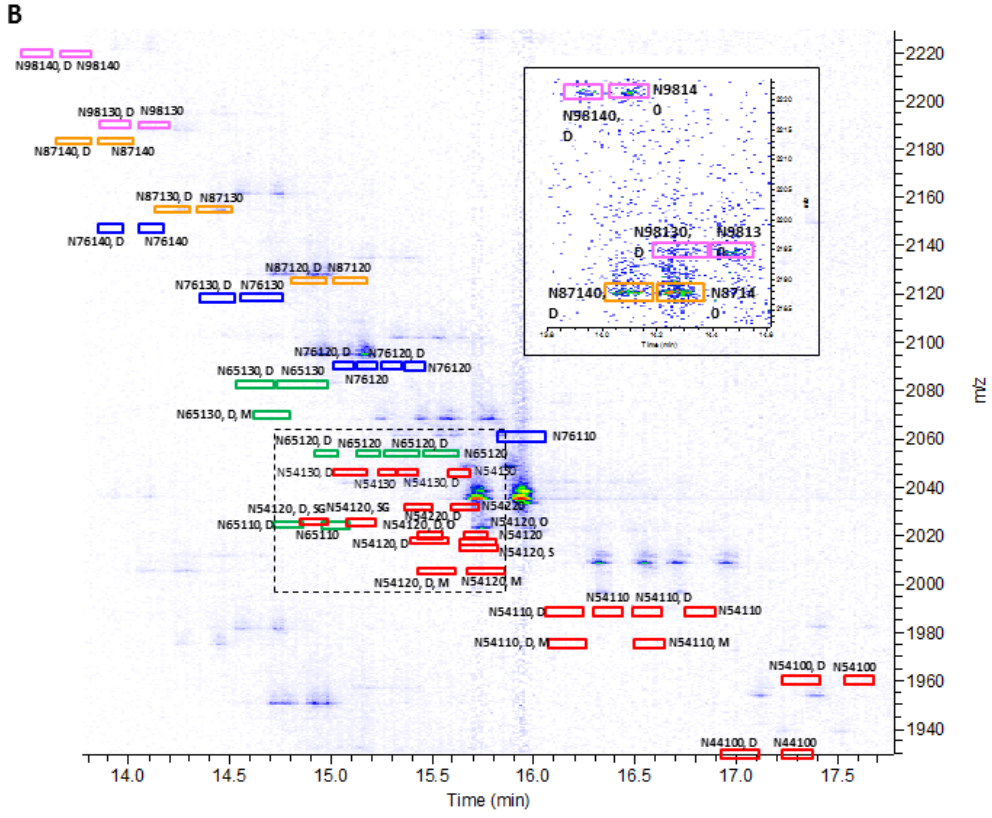
Capillary and Injection Date	**Acetaminophen**			
	t_o (min)	t_o RSD	μ_{EOF} (m^2/Vs)	μ_{EOF} RSD
Capillary 2, Day 1	9.08	0.53	-8.28E-08	-0.58
	9.10		-8.26E-08	
	9.02		-8.34E-08	
Capillary 2, Day 2	9.13		-8.23E-08	
	9.14		-8.23E-08	
	9.15		-8.22E-08	

Capillary and Injection Date	**N64130 (N→D) a**						
	t_e (min)	t_e RSD	μ_{FS} (m^2/Vs)	μ_{APP} (m^2/Vs)	Average μ_{APP}	% Error from Average	μ_{APP} RSD
Capillary 2, Day 1	NA	1.32	NA	NA	3.37E-08		0.35
	14.88		3.38E-08	3.38E-08		0.26	
	14.58		3.35E-08	3.35E-08		0.51	

Capillary 2, Day 2	NA		NA	NA			
	14.99		3.38E-08	3.38E-08		0.26	
	15.00		3.37E-08	3.37E-08		0.03	
Capillary and Injection Date	**N64130 (N→D) b**						
	t_e (min)	t_e RSD	μ_{FS} (m^2/Vs)	μ_{APP} (m^2/Vs)	Average μ_{APP}	% Error from Average	μ_{APP} RSD
Capillary 2, Day 1	14.85	1.04	NA	3.39E-08	3.39E-08	0.03	0.48
	14.95		3.40E-08	3.40E-08		0.46	
	14.65		3.38E-08	3.38E-08		0.28	
Capillary 2, Day 2	14.91		NA	3.36E-08		0.82	
	15.06		3.40E-08	3.40E-08		0.40	
	15.07		3.40E-08	3.40E-08		0.21	
Capillary and Injection Date	**N64130 a**						
	t_e (min)	t_e RSD	μ_{FS} (m^2/Vs)	μ_{APP} (m^2/Vs)	Average μ_{APP}	% Error from average	μ_{APP} RSD
Capillary 2, Day 1	14.98	1.03	3.43E-08	3.43E-08	3.43E-08	0.12	0.47
	15.07		3.45E-08	3.45E-08		0.43	
	14.77		3.42E-08	3.42E-08		0.24	
Capillary 2, Day 2	15.03		3.40E-08	3.40E-08		0.82	
	15.18		3.44E-08	3.44E-08		0.35	
	15.19		3.44E-08	3.44E-08		0.16	
Capillary and Injection Date	**N64130 b**						
	t_e (min)	t_e RSD	μ_{FS} (m^2/Vs)	μ_{APP} (m^2/Vs)	Average μ_{APP}	% Error from average	μ_{APP} RSD
Capillary 2, Day 1	15.05	1.06	3.46E-08	3.46E-08	3.46E-08	0.02	0.46
	15.16		3.48E-08	3.48E-08		0.47	
	14.85		3.45E-08	3.45E-08		0.26	
Capillary 2, Day 2	15.12		3.43E-08	3.43E-08		0.77	
	15.27		3.47E-08	3.47E-08		0.38	
	15.28		3.47E-08	3.47E-08		0.19	

Tables 1A-B: Repeatability and Reproducibility of μ_{APP} of rhIFN-β1 proteoforms. (A) Biantennary forms. **(B)** Triantennary forms.


A
N98140, D
N9814 0
N98130, D
N9813 0
N87140, D
N8714 0
N87130, D
N8713 0
N76140, D
N7614 0
N87120, D
N8712 0
N76130, D
N7613 0
N76120, D
N7612 0
N65130, D
N65130, D, M
N65120, D
N6512 0
N54130, D
N5413 0
N54120, D, M
N54120, D
N5412 0
N65110, D
N6511 0
N7611 0
N54110, D
N54110, D, M
N54110, M
N5411 0
N54100, D
N5410 0
N44100, D
N4410 0
N9814
N98140, D
N98130, D
N9813 0
N87140, D
N8714 0
m/z
Time (min)
14.0
14.5
15.0
15.5
16.0
16.5
17.0
17.5
1940
1960
1980
2000
2020
2040
2060
2080
2100
2120
2140
2160
2180
2200
2220


B
N98140, D
N98130, D
N87140, D
N87130, D
N76140, D
N87120, D
N76130, D
N76120, D
N65130, D
N65130, D, M
N65120, D
N76110
N54130, D
N54120, D, SG
N65110, D
N54120, D, M
N54110, D
N54110, D, M
N54110, M
N54100, D
N44100, D
N9814
N98140, D
N98130, D
N9813 0
N87140, D
N8714 0
m/z
Time (min)
14.0
14.5
15.0
15.5
16.0
16.5
17.0
17.5
1940
1960
1980
2000
2020
2040
2060
2080
2100
2120
2140
2160
2180
2200
2220

Figures 2-2 A-C: Ion density map of CZE-MS separation of rhIFN-β1. CZE-MS was performed for a 1.2 mg/mL rhIFN-β1 in 50 mM ammonium acetate pH 4.5, injected hydrodynamically with 0.5 psi of pressure for 10s (7.6 nL). **(A)** The map is focused on the charge 11+ species and shown as $\log_2$(intensity). The insert is a magnified view of region between 13.8 and 14.6 minutes and 2181-2224 *m/z* showing the low abundance N981x0 and N87140 peaks with intensity plotted linearly. **(C)** Magnified view of region between 14.8 and 16.75 minutes and 2021-2073 *m/z* showing the low abundance N54xx0 forms shown as $\log_2$(intensity). Glycans are highlighted with colors corresponding to the number of LacNAc: 2 red, 3 green, 4 blue, 5 orange, 6 pink. Glycan nomenclature: N(Hex)(HexNAc)(Fuc)(NeuAc)(NeuGc); Modifications are denoted as follows: D (Deamidation), S (Succinimide), O (Oxidation), M (N-methionine loss), SG (sulfated glycan). Peaks deconvoluted within 5 ppm of the theoretical mass are shown in **bold** and are listed in **Table 2-2**.

High column efficiency is essential for the separation of closely-related proteoforms. **Table 2-2** provides plate counts for a variety of proteoforms, showing that plate counts are between 350,000 and 450,000. An example of the separation efficiency and peak symmetry can be seen in **Figures 2-1** and **2-3**. Such large plate counts lead to the separation of subtly-different isobaric proteoforms, as shown below.

Glycan	t_e (min)	$N_{1/2}$
Acetaminophen	9.15	636200
N44100 Deam	17.25	384200
N44100	17.52	422000
N54110 Deam 1	16.45	390000
N54110 Deam2	16.84	451300
N54110 1	16.67	473800
N54110 2	17.07	394100
N54120 Deam	15.84	413200
N54120	16.06	424800
N65130 1 *	15.18	367200
N65130 2 *	15.27	427600
N76140 Deam	14.37	352100
N76140	14.56	469800

Table 2-2: Plate numbers and resolution of rhIFN-β1 proteoforms. Analytical separation metrics were calculated for proteoforms with unique glycoform structures with no potential for

glycan isomers to influence the analysis. The exception is the first and second non-deamidated peaks for the N65130 glycoform, which were not separated at baseline.

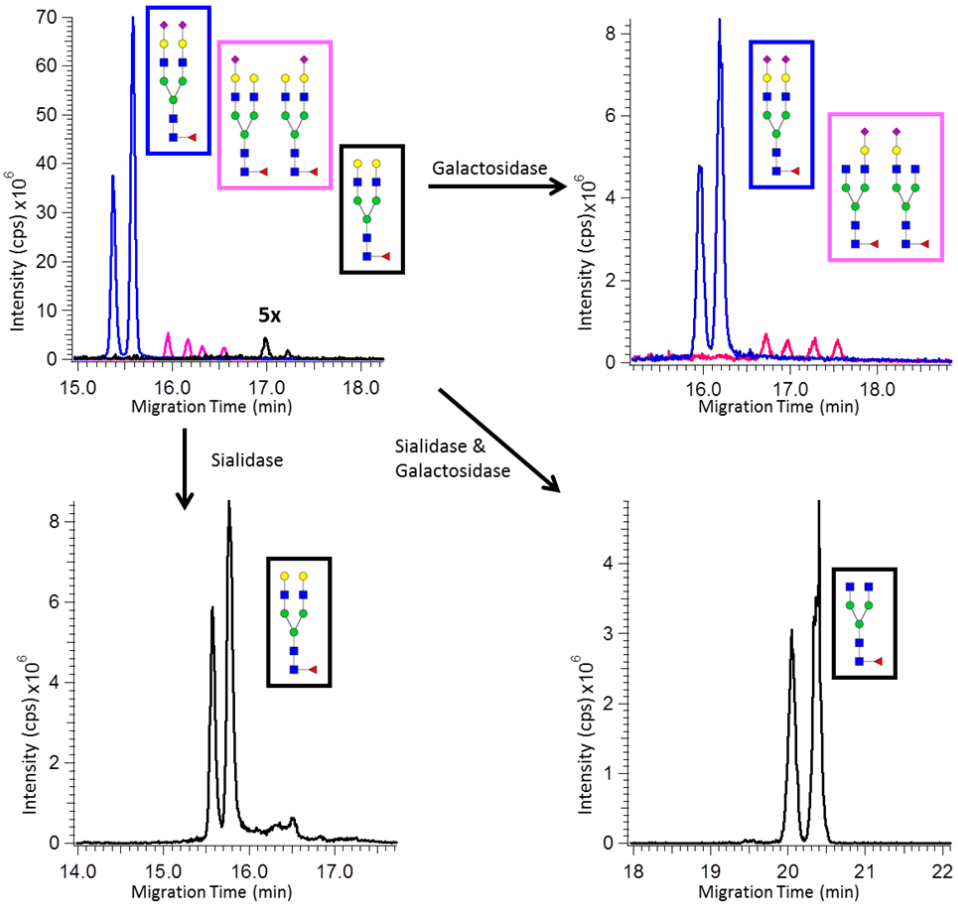
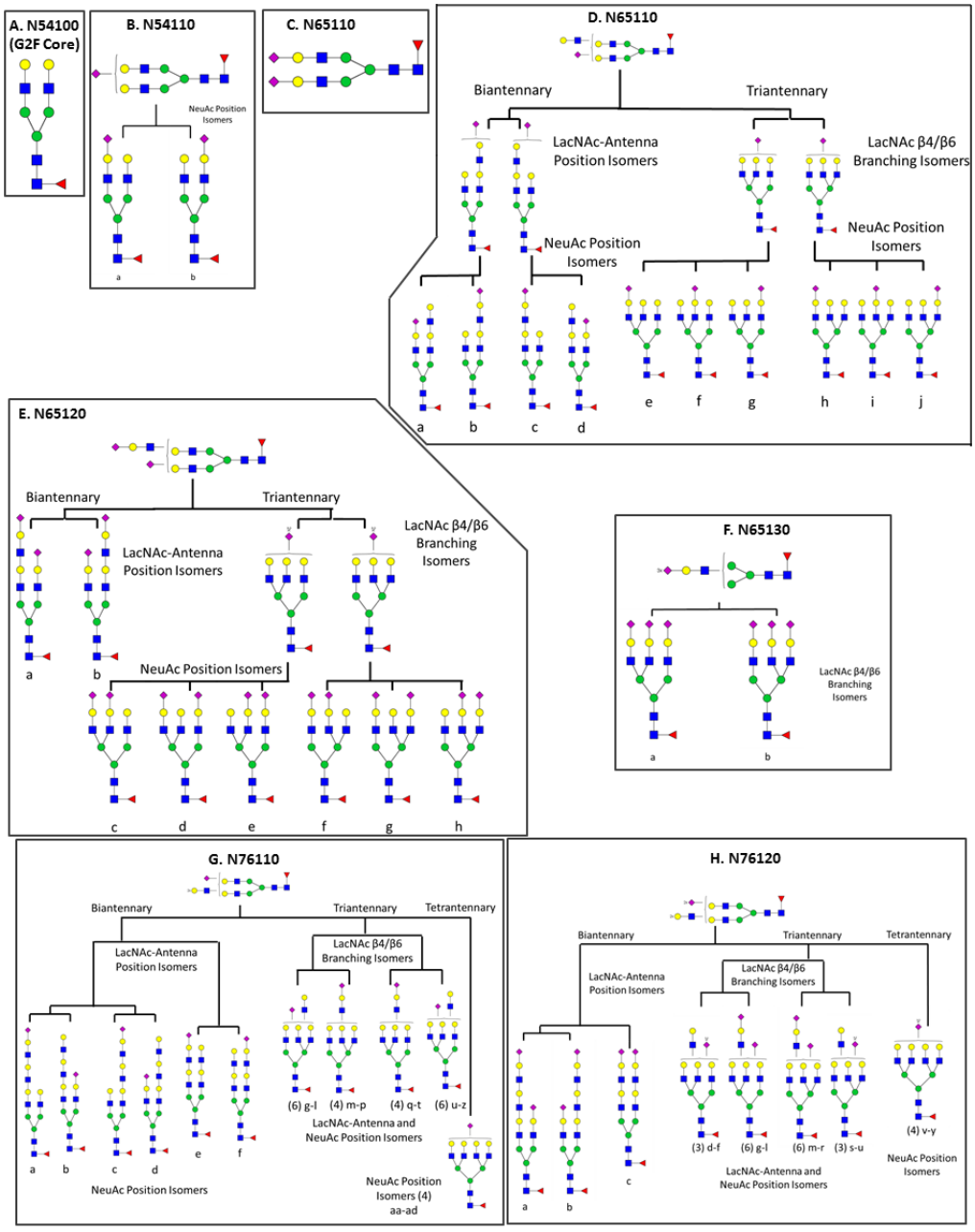


Figure 2-3: Extracted ion electropherograms of the N54xxx (2 LacNAc) glycoforms of rhIFN-β1 with and without exoglycosidase treatment. Injection conditions for the different treatments are stated in their respective figure captions of the ion density maps (**Figures 2-2**, **Figures 2-6 – 2-8**). Each channel corresponds to the top six charge states (z = +11 to +16) of the indicated forms. Possible glycan structures corresponding to the extracted signals peaks are outlined with the same color as the peaks.

2.3.2. CZE-MS of RhIFN-β1

N-glycans produces in CHO cells consist of the biantennary G2F core polysaccharide (**Figure 2-4**) [252], to which are attached, among others, repeating N-acetyllactosamine (LacNAc) extensions and up to four LacNAc antennae, many of which are terminated with sialic

acid (NeuAc) with very small amounts of (NeuGc) [253-255]. The mass spectrometric analysis of native glycans cannot distinguish between hexoses (Hex) as well as N-acetylhexosamines (HexNAc). We therefore have selected a previously developed notation to describe glycan compositions [256], where the first character denotes an N- or O-glycan, and each number represents the number of Hex, HexNAc, Fuc, NeuAc, and NeuGc units, respectively. Annotated structures are shown in **Figure 2-4**.



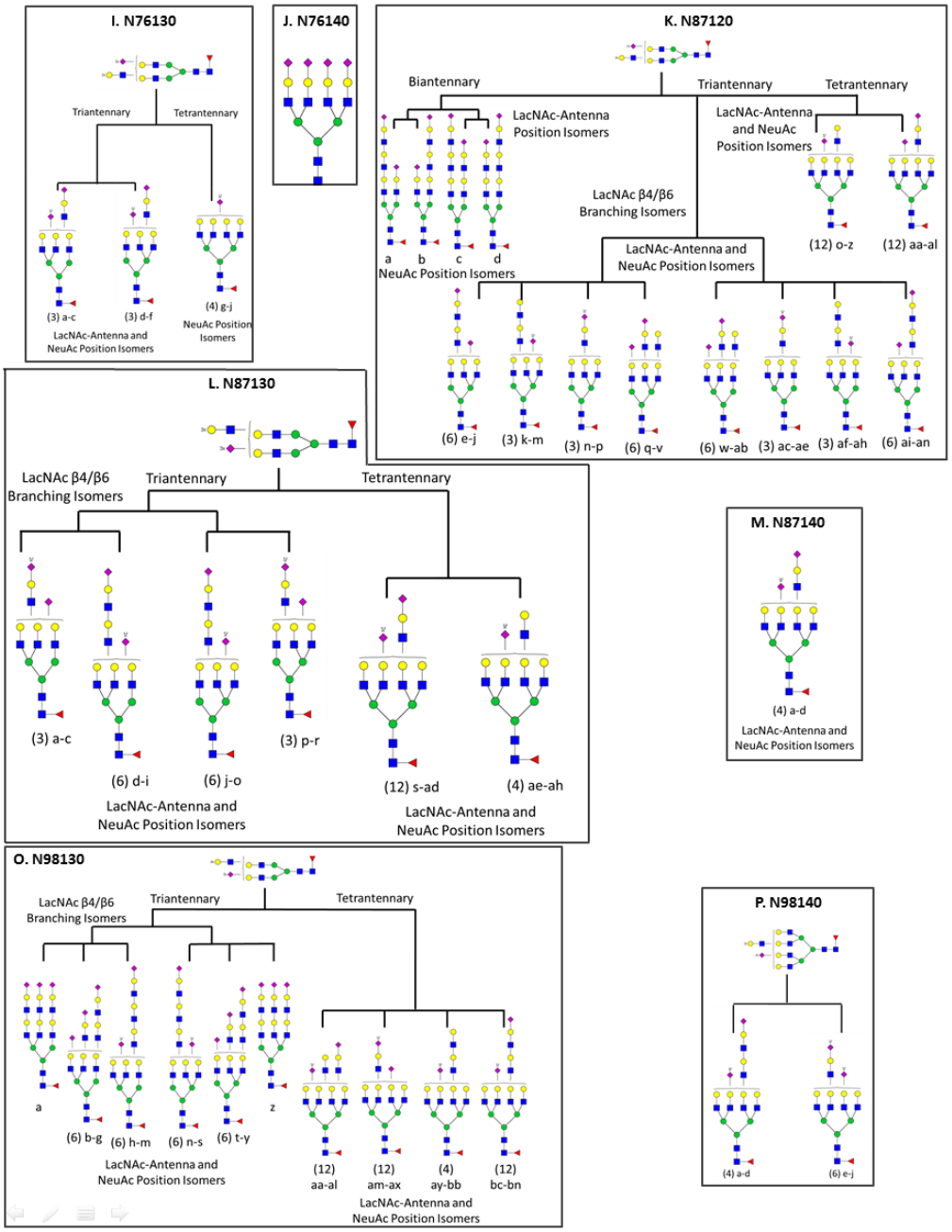


Figure 2-4: **Isomers of rhIFN-*β*1 glycan structures.** For every unique glycan composition detected by MS^1 analysis, structures are organized by number of antenna, position of the third antenna for triantennary structures, attachment position of LacNAc extensions, and attachment position of NeuAc attachment. For structures with a large number of possibilities due to the latter two features, not all permutations of those features are explicitly shown. The labeling conventions have been described previously [257], where each N- or O- denotes an N- or O-glycan, and each number represents the number of Hex, HexNAc, Fuc, NeuAc, and NeuGc, respectively.

The ion density map of intact rhIFN-β1 on the cPEI coated capillary is shown in **Figure-2-2A** and an expanded section in **Figure 2-2B**, with major proteoforms highlighted according to their sialic acid content. In total, 138 proteoforms were found. Sixty three proteoform identifications based on deconvoluted MS^1 accurate mass data (5 ppm) are detailed in **Table 2-3**, and the remainder were based on their relative mass and apparent mobility. Extracted ion electropherograms (EIE) of the major glycoforms with 2-5 LacNAc units are shown in **Figures 2-3** and **2-5**, **Figures 2-9 – 2-10**.

Proteoform		Monoisotopic MH+		
Glycan	PTM	Theo.	Obs.	Error (ppm)
N34000	N→D, mixed	21309.76	21309.67	4.3
N44000		21472.81	21472.74	3.2
	N→D	21473.79	21473.70	4.3
N54100		21781.92	21,781.91	0.6
	N→D	21782.91	21,782.91	0.3
N54110	N→D (1)	22074.01	22074.10	4.2
	Des-1 (1)	21942.97	21942.96	0.2
	(1)	22073.02	22072.99	1.4
	N→D (2)	22074.01	22073.92	3.7
	Des-1 (2)	21942.97	21942.96	0.2
	(2)	22073.02	22072.99	1.5
N54120		22364.12	22364.14	1.0
	Des-1	22233.08	22233.11	1.2
	Succ	22347.09	22347.18	4.1
	Ox	22380.11	22380.14	1.0
	SG	22444.08	22444.10	1.2
	N→D	22365.10	22365.13	1.4
	N→D, Des-1	22234.06	22234.08	0.6
	N→D, Succ	22348.08	22348.10	1.2
	N→D, Ox	22381.10	22381.18	3.5
	N→D, SG	22445.06	22445.14	3.5
N54220		22510.18	22510.09	3.8
	N→D	22511.16	22511.10	2.8
N54130	N→D mixed	22656.19	22656.23	1.7
N65110		22438.16	NA‡	NA
N65120	N→D (1)	22730.23	22730.19	1.7
	(1)	22729.25	22729.23	0.7
	N→D (2)	22730.23	22730.24	0.5
	(2)	22729.25	22729.21	1.8

N65130		23020.34	23020.41	3.0
	N→D, Des-1	22890.29	22890.31	0.9
	N→D	23021.33	23021.37	1.8
N76110	mixed	22656.22	22656.10	5.5
N76120	mixed	23095.37	23095.35	0.5
N76130		23385.48	23385.51	1.2
	N→D, Des-1	23255.42	23255.36	2.5
	N→D	23386.46	23386.53	3.0
N76140		23676.57	23676.55	0.9
	N→D	23677.56	23677.58	0.9
N87120	mixed, N→D	23460.50	23460.46	1.8
N87130		23750.61	23750.63	0.8
	N→D	23751.59	23751.66	3.0
N87140		24041.70	24041.74	1.5
	N→D	24042.69	24042.76	3.2
N98130	mixed, N→D	24116.72	24116.74	0.5
N98140		24406.84	24406.79	2.1
	N→D	24407.82	24407.83	0.3

Table 2-3: Summary of identified glycoforms and protein modifications of intact rhIFN-*β*1. Modifications: N→D, deamidation; Des-1, N-methionine loss; Ox, Oxidation; Succ, succinimide; SG, sulfated glycan. Glycoforms with more than one pair of deamidated/non-deamidated species are numbered in order of migration. 'Mixed' indicates that the proteoforms for a given glycoform could not be accurately isolated for quantitation and the theoretical MH^+ used in calculating the error was that of the indicated PTM. ǂ N65110 overlaps slightly in migration and *m/z* space with the N54120 SG peaks, shown in the ion density map in **Figure 2-2B**.

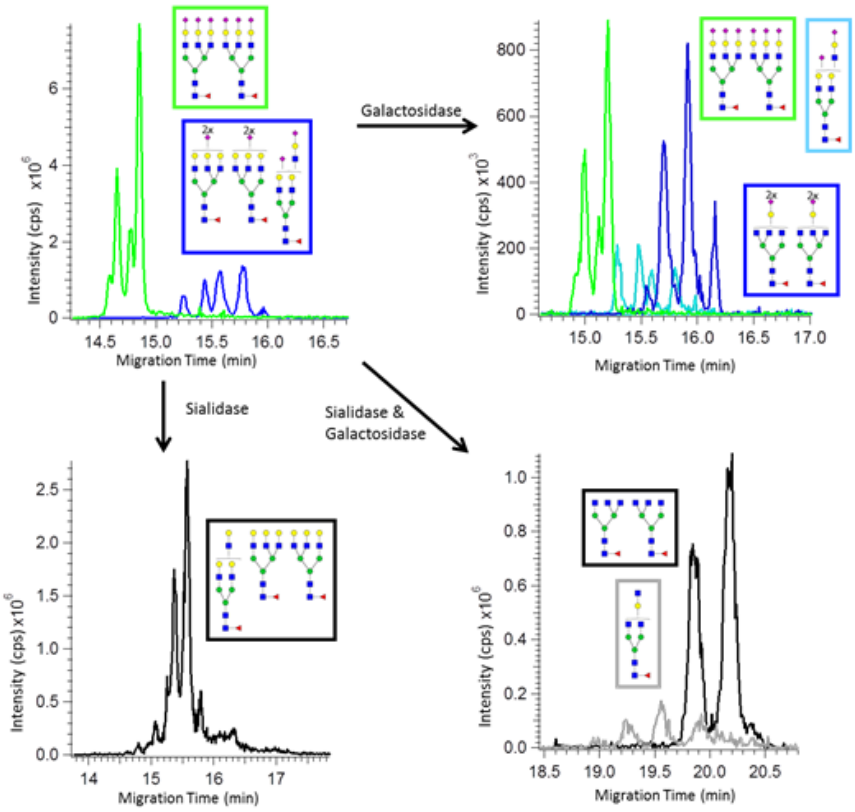


Figure 2-5: Extracted ion electropherograms of the N65xxx (3 LacNAc) glycoforms of rhIFN-β1 with and without exoglycosidase treatment. Injection conditions for the different treatments are stated in their respective figure captions of the ion density maps (**Figure 2-1**, **Figures 2-6 – 2-8**). Each channel corresponds to the top six charge states (z = +11 to +16) of the indicated forms. Possible glycan structures corresponding to the extracted signals peaks are outlined with the same color as the peaks.

As can be seen in **Figure 2-2** and **Table 2-3**, for a given glycan form, as expected, the migration time decreased with the number of sialic acids present. At the same time, for a given number of sialic acids, there are significant changes in migration due to the structure of the glycan on the proteoform resulting in hydrodynamic volume differences. The crystal structure of rhIFN-β1 [258] reveals that the glycosylation site is on the surface of the protein. Given the overall flexibility of glycans, variations in carbohydrate structure can potentially have a significant impact on hydrodynamic volume and thus migration time. For example, with additional HexNAcHex units, migration time is seen to decrease due to an increase in hydrodynamic volume (e.g., N54120 > N65120 > N76420, **Figures 2-2 – 2-3**, and **Figure 2-5**). Glycoforms are generally found in pairs of 2 peaks, which are attributed to non-deamidated and singly deamidated species, based on accurate mass (the single deamidation occurs at N25, based

on peptide mapping). Note, that without separation, it would be very difficult to identify the deamidated peak, which differs by only 1 Da, from the non-modified form using top-down proteomic analysis.

Three less-common glycans were found in **Figure 2-2** and listed in **Table 2-3**. First, a glycan with a second Fuc, N54220, has been observed in CHO cells as a sialyl Lewis-X (SLe^x) structure [253]. The associated deamidated and non-deamidated peaks migrate slightly slower than the N54210 form, as would be expected given the increase in hydrodynamic volume imparted by an additional neutral Fuc. Secondly, a +80 Da modification from the N54120 form was observed. This addition could be a sulfation or phosphorylation modification, but evidence from exoglycosidase digests described later points to a sulfated glycan. Either of the modifications would result in the pronounced increase in μ_{APP} due to the addition of a negative charge. Thirdly, while not commonly observed, the N54130 glycoforms (biantennary with three sialic acids) in **Figure 2-2** have been found to exist [259].

Some proteoforms are separated into more than two pairs of peaks (see **Figures 2-2B, Figures 2-3, 2-5**, **Figure 2-9**). The most prominent example is the N54110 species where two isobaric pairs of deamidated/non-deamidated peaks are found (see **Figures 2-2B** and **2-3**). Based again on accurate mass (**Table 2-3**), the four peaks are in the order: deamidated, non-deamidated, deamidated, and non-deamidated. Since the glycan is biantennary with a single NeuAc, the two pairs of deamidated/non-deamidated species are most likely due to the separation of positional isomers with the sialic acid residing on either the α(1-3) or α(1-6) linked mannose antenna (N54110a and N54110b, Scheme S-1B). This result is validated below with the exoglycosidase digest. Further, compared to the N54110 forms, the N54120 with 2 sialic acids cannot form positional isomers and thus only two peaks are found. The N54130 glycoforms, shown in **Figure 2-2B**, also have the two possible forms which, along with deamidation, provide two pairs of peaks. Another peak splitting occurs with N65130 with 3 sialic acids terminating a trianntenary structure (**Figure 2-3**), of which, as shown in **Figure 2-4F**, the only possible difference between N65130a and N65130b is attachment of the third antenna to either the α(1-3) or α(1-6) arm.

Positional isomer separation by CZE, while observed for glycans [260], has, to our knowledge, not been shown for intact proteins of 25,000 Da or more. Taking the free solution mobility ratios for the above positional isomers, we find that the hydrodynamic differences for the mono-sialic acid-biantennary proteoforms differ by 2.5% and for the triantennary with 3 sialic acids by only 0.9% (see **Table 2-1**). The ability to separate such small hydrodynamic volume differences is due to the ultra-high efficiency of the separations (> 350,000 plates).

Besides glycan proteoforms, 4 PTMs were observed in **Table 2-3**. In addition to deamidation, an associated succinimide was also found based on accurate mass. Oxidation was also observed, the presence of which is known from peptide mapping analysis (data not shown). The fourth detected PTM contained a -131 Da mass difference, which is assigned as an N-

terminal methionine loss (indicated as M in **Figure 2-2**). To confirm the des-1 structure, we reduced rhINF-β1 and performed top-down CZE-MS^2. ETD and HCD fragmentation gave *c*- and *b*-ions, respectively, that validated the des-1 assignment (**Figure 2-15**). It is important to note that a previous CZE-MS study of rhINF-β1 [229] did not report the N-methionine loss. As a consequence, several of the glycan structures reported in the paper were misassigned as glycans not commonly found in CHO cells (e.g., N64110 and N64100). When multiple PTMs and numerous glycan structures exist on a complex glycoprotein, structures corresponding to similar molecular masses can exist, potentially leading to misassignment. High resolution separation/mass spectrometry and the ability to conduct MS^2 analysis (top-down) is thus important.

2.3.3. CZE-MS Separations of Exoglycosidase-Treated RhIFN-β1

To explore more deeply the separation of the glycan isomers of the intact protein, we incubated rhINF-β1 with sialidase, galactosidase, or both enzymes to selectively remove sialic acid and/or galactose. We believe this to be the first-reported application of multiple exoglycosidase treatments to elucidate glycan structure intact on a protein of 25 kDa by CZE-MS. The ion density map for the treated samples are shown in **Figures 2-6 – 2-8** and extracted ion electropherograms of the high abundance glycoforms in **Figures 2-3** and **2-5**, **and Figures 2-9 – 2-10**.

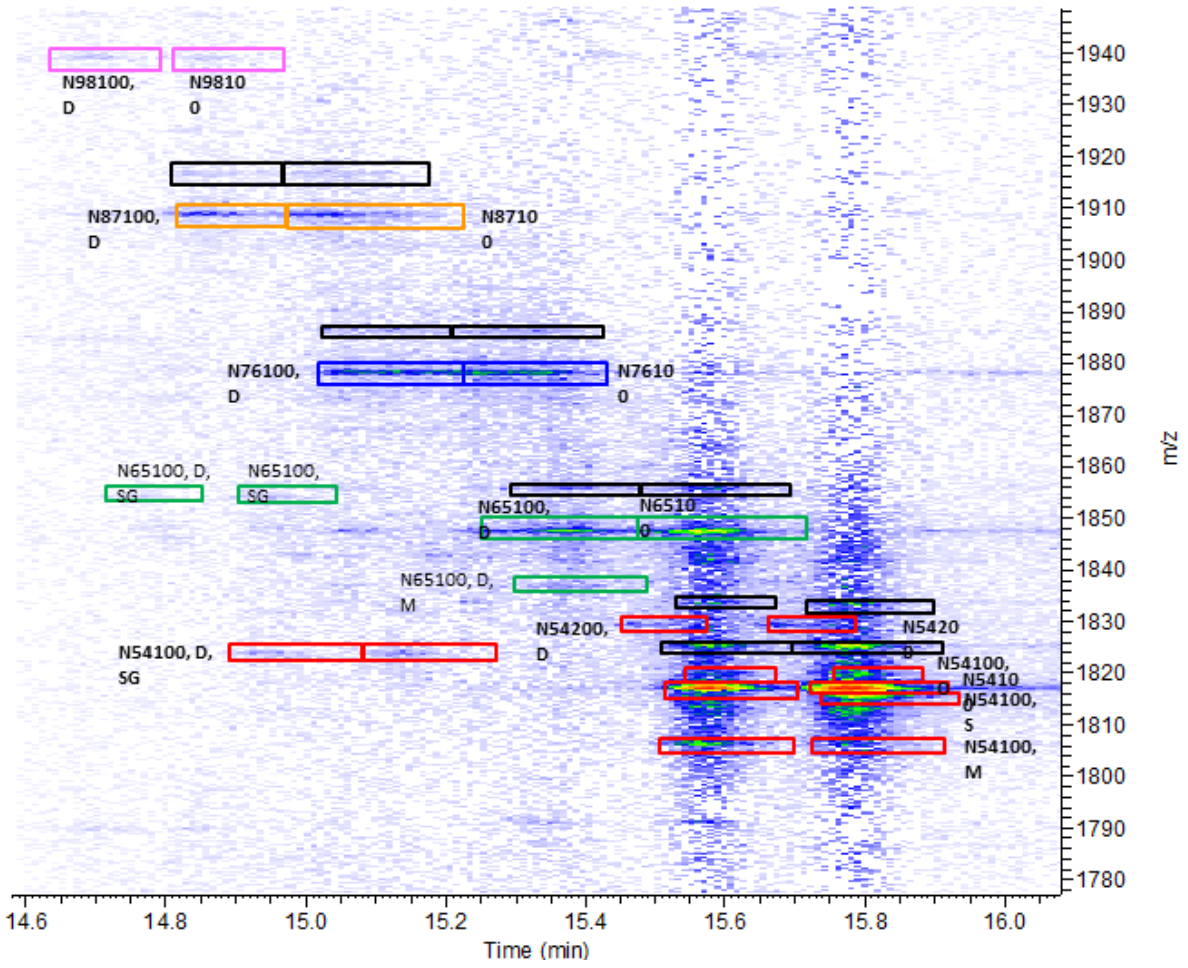


Figure 2-6: Ion density map of CZE-MS separation of sialidase-treated rhIFN-β1. CZE-MS was performed on a 1.2 mg/mL rhIFN-β1 in 50 mM ammonium acetate, pH 4.5 injected with 0.5 psi of pressure for 10 s (7.5 nL). The map is focused on the charge 12+ species and shown as $\log_2$(intensity). Glycans are highlighted with colors corresponding to the number of LacNAc: 2 red, 3 green, 4 blue, 5 orange, 6 pink. Peaks highlighted in black are +98 Da, likely due to complexation with phosphate in exoglycosidase digestion buffer. Glycan nomenclature: N(Hex)(HexNAc)(Fuc)(NeuAc)(NeuGc); Modifications are denoted as follows: D Deamidation, S Succinimide, O Oxidation, M N-methionine loss, SG sulfated glycan.

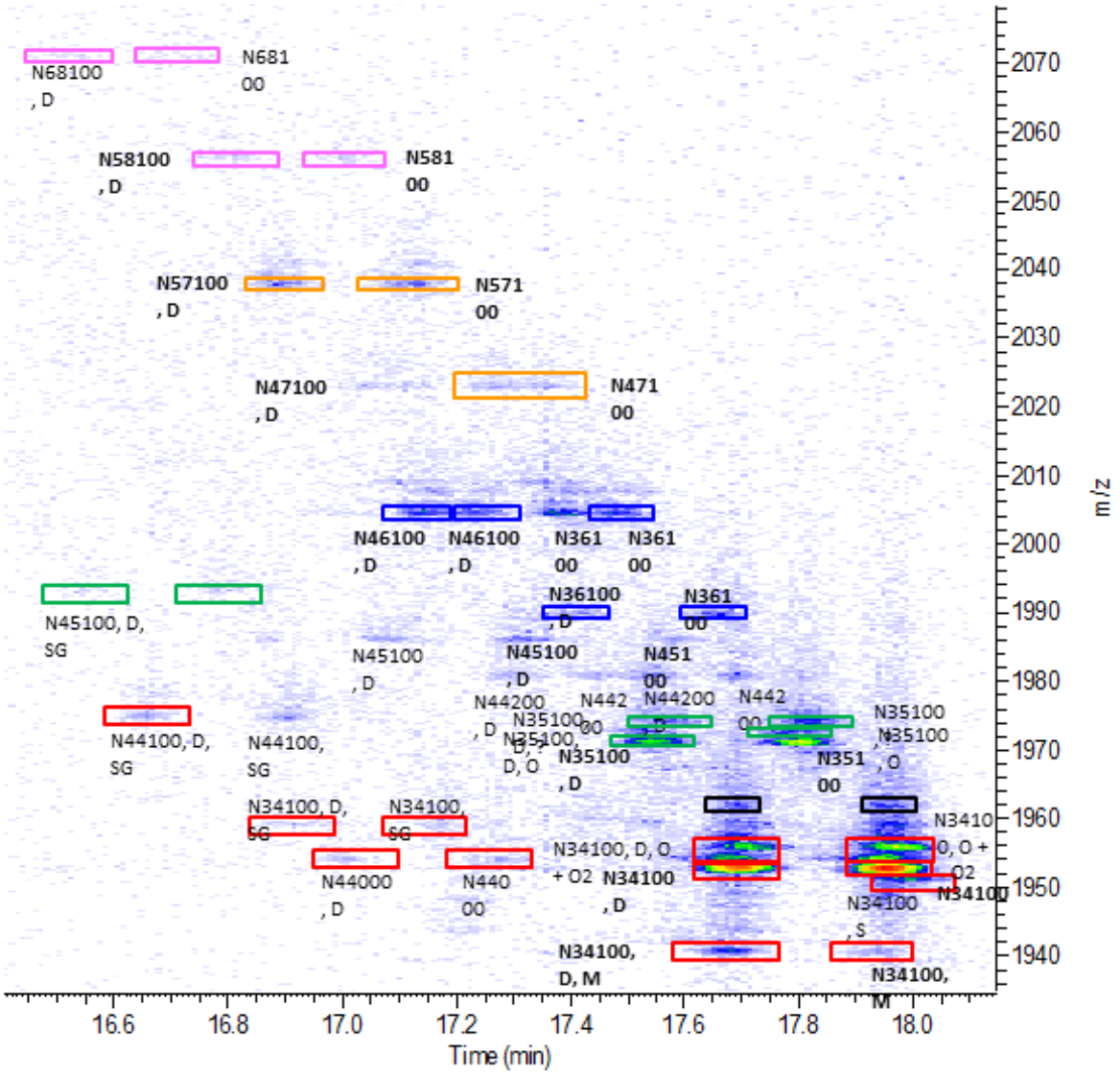


Figure 2-7: Ion Density map of CZE-MS separation of sialidase and galactosidase-treated rhIFN-β1. CZE-MS was performed on a 1.2 mg/mL rhIFN-β1 in 50 mM ammonium acetate pH 4.5 was injected with 0.5 psi of pressure for 10s (7.6 nL). The map is focused on the charge 11+ species and shown as $\log_2$(intensity). Glycans are highlighted with colors corresponding to the number of LacNAc: 2 red, 3 green, 4 blue, 5 orange, 6 pink. Peaks highlighted in black are +98 Da, likely due to complexation with phosphate in exoglycosidate digestion buffer. Glycan nomenclature: N(Hex)(HexNAc)(Fuc)(NeuAc)(NeuGc); Modifications are denoted as follows: D Deamidation, S Succinimide, O Oxidation, M N-methionine loss, SG sulfated glycan.

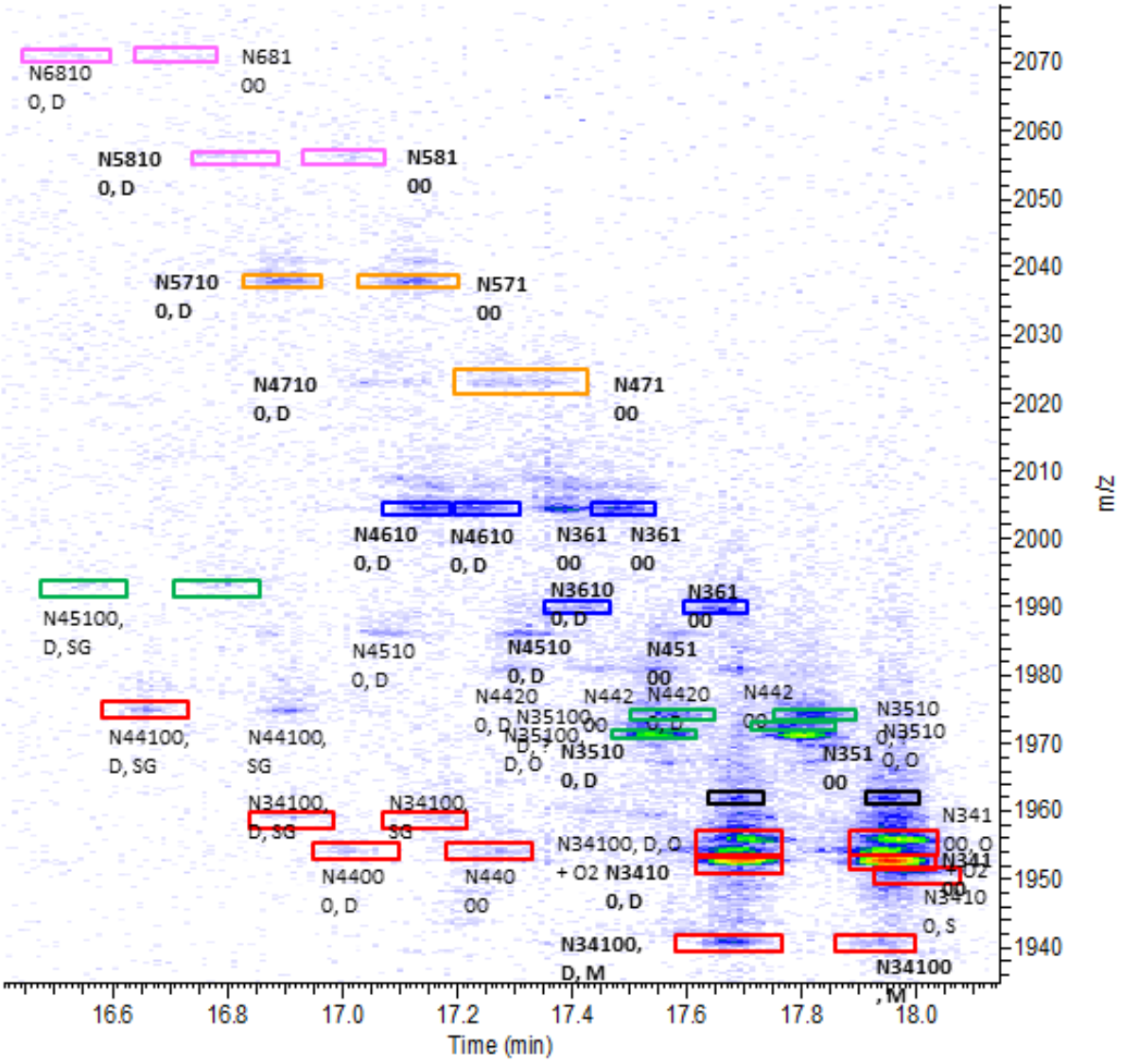


Figure 2-8: Ion Density map of CE-MS separation of galactosidase-treated rhIFN-*β*1. CZE-MS was performed on a 0.8 mg/mL rhIFN-β1 in 50 mM ammonium acetate pH 4.5 was injected with 0.5 psi of pressure for 10s (7.6 nL). The map is focused on the charge 11+ species and shown as $\log_2$(intensity). Glycans are highlighted with colors corresponding to the number of LacNAc: 2 red, 3 green, 4 blue, 5 orange. Glycan nomenclature: N(Hex)(HexNAc)(Fuc)(NeuAc)(NeuGc); Modifications are denoted as follows: D Deamidation, S Succinimide, O Oxidation, M N-methionine loss, SG sulfated glycan.

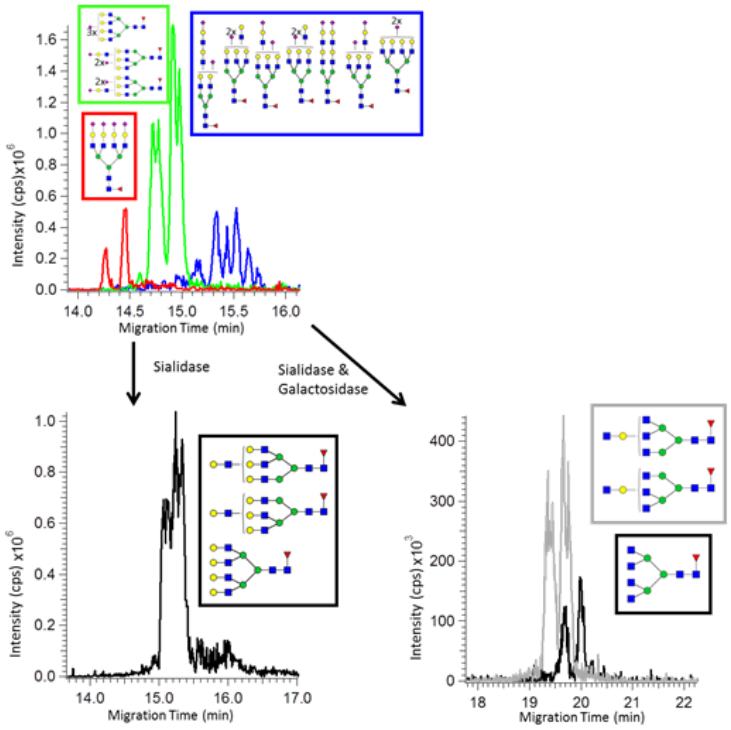


Figure 2-9: Extracted Ion Electropherograms of the N76xxx (4 LacNAc) glycoforms of rhIFN-*β*1 with and without exoglycosidase treatment. Injection conditions for the different treatments are stated in their respective figure captions of the ion density maps (**Figure 2-2**, **Figure 2-6 – 2-8**). Each channel corresponds to the top six charge states (z = 11-16+) of the indicated forms. Possible glycan structures corresponding to the extracted signals peaks are outlined with the same color as the peaks.

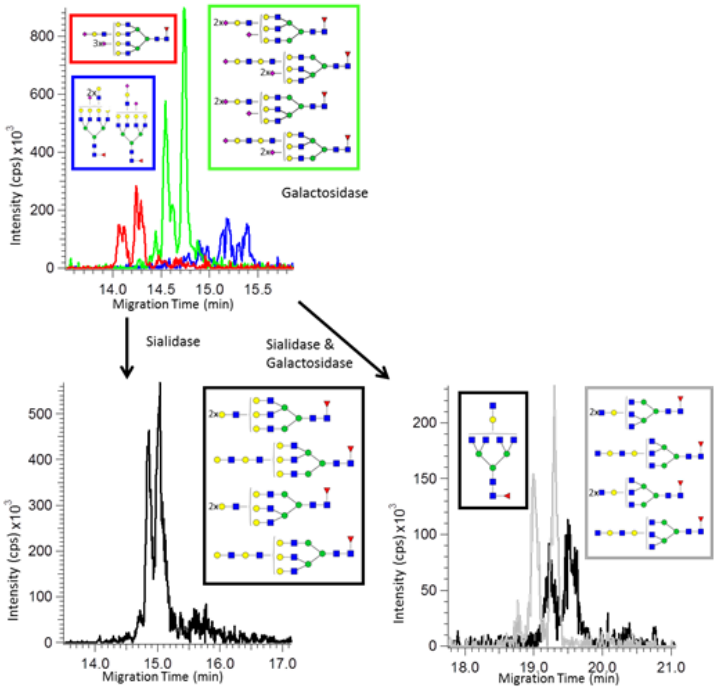


Figure 2-10: Extracted Ion Electropherograms of the N87xxx (5 LacNAc) glycoforms of rhIFN-β1 with and without exoglycosidase treatment. Injection conditions for the different treatments are stated in their respective figure captions of the ion density maps (**Figure 2-2**, **Figure 2-6 – 2-8**). Each channel corresponds to the top six charge states (z = +11-16) of the indicated forms. Possible glycan structures corresponding to the extracted signals peaks are outlined with the same color as the peaks.

Consider first the biantennary glycan structures shown in **Figure 2-3**. Sialidase and sialidase + galactosidase treatments collapse all forms (N54120, N54110, N54100) into two peaks, i.e. non- deamidated and deamidated. Importantly, the fact that N54110 (one sialic acid) converted from 4 to 2 peaks is strong evidence that positional isomers are separated in the non-treated intact protein sample. For the galacosidase treatment, the N54120 form remains unchanged as the galactoses are protected by NeuAc; however, as expected, N54110 and N54100 are digested to the N44110 and N34100 since the unprotected galactoses are removed.

The possibilities for structures with 3-LacNAc are more numerous, as the potential attachment of a third antenna to either the α(1-3) or α(1-6) mannose or a poly-LacNAc at the end of a biantennary structure becomes possible (see **Figure 2-5**). With respect to the triantennary 3-sialic acid (N65130) forms, galactosidase digestion, as expected, does not alter the EIE pattern. However, sialidase + galactosidase digestion leads to two major isobaric peaks (N35100) which strongly supports the positional isomer separation of the α(1-3) and α(1-6) structures.

Examining next the 2-sialic acid N65120 structures in the untreated sample, more than 4 isobaric peaks are observed. As shown in **Figure 2-5**, treatment with galactosidase yields primarily the triantennary N55120 structure with an unprotected galactose cleaved, and additionally, the biantennary N65120 structure with both galactoses being protected by sialic acid. Further, the galactosidase digest results in 2 pairs of deamidated/non-deamidated peaks from the attachment of the third LacNAc to either the α(1-3) or α(1-6) arm. Thus, the N65120 structure is a complex mixture of positional isomers of tri- and biantennary glycans. The exoglycosidase treatment clearly aids in the elucidation of these complex structures.

2.3.4. Relative Quantitation of RhIFN-β1 Proteoforms

We next conducted quantitative analysis of all proteoforms whose MS spectra were deconvoluted. Relative quantitation was achieved by summing the EIE integrated signal of the six charge states with the highest ion intensities and dividing by the sum of the total signal for all proteoforms. The signal extraction window was centered (+65%/-35%) around the theoretical average masses for the given charge state within mass tolerance windows of 1.4, 1.3, 1.25, 1.15, 1.0 Da (z = +11 to +16, respectively). Identified empirical formulas of rhIFN-β1 that could be accurately quantitated are presented in **Table 2-4**. MS spectral deconvolution was possible for all of these forms. Some low level features, although discernable, could not be quantitated due to overlapping signal with other glycoforms.

Untreated			
Proteoform		Average	*RSD
Glycan	PTM		
N34000	-	0.17	>20
N34000	N→D	0.19	>20
N44000	-	0.58	>20
N44000	N→D	0.54	>20
N54100	-	0.1	>20
N54110	N→D	0.21	>20
N54110	N→D (1)	2.2	7.9
N54110	N→D (1)	1.9	15
N54110	N→D (2)	1.2	>20
N54110	N→D (2)	1.2	>20
N54110	Des-1 N→D (1)	0.46	9.9
N54110	Des-1, (1)	0.093	>20
N54110	Des-1 N→D (2)	0.28	>20
N54110	Des-1, (2)	0.076	>20

N54120	-	31	3
N54120	Des-1	1.4	16
N54120	Succ.	5.1	2.9
N54120	N→D	16	5.7
N54120	Des-1, N→D	3.2	7.7
N54120	Succ, N→D	1.5	6.5
N54120 SG≠	-	NA	NA
N54220	N→D	NA	NA
N54220	-	2.5	10
N54220	N→D	1.2	9.2
N65110 ≠	mixed	NA	NA
N65120	1	0.48	18
N65120	2	0.84	7.7
N65120	3	1.5	6.8
N65130	N→D (a)	0.6	11
N65130	N→D (b)	2.1	7.1
N65130	(a)	1.5	13
N65130	(b)	4.5	9.5
N65130	Des-1 N→D (a+b)	0.38	4.1
N76110	mixed	1.4	17
N76120	mixed	2.8	8
N76130	-	3.1	9.9
N76130	N→D	1.7	6.5
N76140	-	0.6	>20
N76140	N→D	0.23	>20
N87120	1	0.16	>20
N87120	2	0.34	>20
N87120	3	0.4	13
N87120	4	0.3	9.8
N87130	-	1	6.5
N87130	N→D	0.67	>20
N87140	-	0.4	>20
N87140	N→D	0.2	>20
N98130	-	0.16	11
N98130	N→D	0.11	>20
N98140	-	0.13	>20
N98140	N→D	0.09	>20

Table 2-4: **Percent relative quantitation of CZE-MS separation of intact rhINF-*β*1 proteoforms**. Modifications: N→D, deamidation; Des-1, N-methionine loss; Succ, succinimide; SG, sulfated glycan. Glycoforms with more than one pair of deamidated/non-deamidated species

are numbered in order of migration. 'Mixed' indicates that the deamidated/non-deamidated proteoforms for a given glycoform could not be accurately isolated for quantitation and the theoretical MH^+ used in calculating the error was that of the indicated PTM. * N = 6. Quantitation was not performed on N54120 SG and N65110 as they overlap slightly in migration and *m/z* space as shown in the ion density map of **Figure 2-2B**.

Table 2-5 provides quantitative information on various exoglycosidase digest products. Note again that the relative abundance represents detailed information on individual proteoforms (glycan plus PTMs). As in **Table 2-4**, the error in %RSD is well below 20% except for vanishingly small species. The relative abundances of the digest products arising from an untreated species are consistent with the relative abundances shown in **Table 2-4** (i.e., the sum of the relative amounts of all digested forms are consistent with the untreated forms from which they arise). As seen in **Table 2-5**, the selective glycan release allows us to determine in detail the specific levels of a variety of proteoforms, not available from the untreated intact data.

Table 2-5A: Sialidase Treatment					
Proteoform			Monoisotopic MH+		
Glycan	PTM	MT* (min)	Theo.	Obs.	Error (ppm)
N54110	-	15.78	21781.92	21781.94	0.7
N54110	Des-1	15.8	21650.88	21650.96	3.7
N54110	Succ	15.8	21764.9	21764.89	0.4
N54110	Ox	15.78	21797.92	21797.92	0.2
N54110	80+ Daltons	15.13	21861.88	21861.88	0.2
N54110	N→D	15.57	21782.91	21782.94	1.6
N54110	N→D, Des-1	15.58	21651.87	21651.94	3.4
N54110	N→D, Succ	15.57	21765.88	21765.9	0.9
N54110	N→D, Ox	15.58	21798.9	21798.94	1.5
N54110	N→D, 80+ Dalons	14.94	21862.87	21862.88	0.9
N54200	-	15.67	21927.98	21927.92	2.7
N54200	N→D	15.49	21928.97	21928.95	0.9
N65100	-	15.56	22147.06	22147.13	3.2
N65100	N→D	15.36	22148.04	22148	1.8
N76100	-	15.29	22512.19	22512.19	0.2
N76100	N→D	15.09	22513.17	22513.19	0.9
N87100	-	15.04	22877.32	22877.24	3.5
N87100	N→D	14.83	22878.3	22878.32	0.8
N98100	-	14.85	23242.45	23242.45	0.1

N98100	N→D	14.69	23243.44	23243.42	0.8
Table 2-5B: Galactosidase Treatment					
Proteoform			**Monoisotopic MH+**		
Glycan	**PTM**	**MT* (min)**	**Theo.**	**Obs.**	**Error (ppm)**
N34100	-	21.57	21457.82	21457.83	0.4
N34100	N→D	21.13	21458.8	21458.81	0.4
N54120	-	19.36	22364.12	22364.13	0.7
N54120	N→D	19.06	22365.1	22365.13	0.9
N54120	Des-1	19.48	22233.07	22237.12	2.1
N54120	Des-1 N→D	19.12	22234.06	22234.04	0.6
N44110	N→D (1)	20.09	21911.95	21911.91	2
N44110	1	20.44	21910.97	21910.99	1.1
N44110	N→D (2)	20.83	21911.95	21912.08	5.7
N44110	2	21.19	21910.97	21910.99	1.2
N44110	Des-1 N→D (1)	20.11	21780.91	21780.94	1.3
N44110	Des-1, (1)	20.43	21779.93	21779.93	0.1
N44110	Des-1 N→D (2)	20.86	21780.91	21780.94	1.2
N44110	Des-1, (2)	21.25	21779.93	NA	NA
N65120	N→D (1)	18.37	22730.23	22730.26	1.1
N65120	1	18.64	22729.25	22729.26	0.4
N65120	N→D (2)	18.48	22730.23	22730.24	0.3
N65120	2	19.12	22729.25	NA	NA
N65120	3	19.35	22729.25	NA	NA
N55120	N→D	19.01	22568.18	22568.19	0.7
N55120	1	19.27	22567.19	22567.19	0.3
N55120	2	19.63	22567.19	22567.21	0.6
N65130	-	18.2	23020.34	23020.39	2.1
N65130	N→D	17.92	23021.33	23021.36	1.2
N66120	-	18.98	23932.33	22935.35	1.2
N66120	N→D	18.62	22933.31	22933.33	0.7
N76130	-	17.65	23385.47	23385.55	3.3
N76130	N→D	17.91	23386.46	23386.52	2.6
N66130	-	18.18	23233.42	23233.48	2.3
N66130	N→D	17.89	23224.41	23224.42	0.5
N76140	-	16.98	23676.7	23676.61	1.7
N76140	N→D	17.25	23677.55	NA	NA
N87130	-	17.61	23750.61	23750.54	2.9
N87130	N→D	17.36	23751.59	23751.62	1.3
N77130	-	17.82	23588.55	23588.58	0.9
N77130	N→D	17.6	23589.54	NA	NA

N87140	-	17.03	24041.7	24041.74	1.5
N87140	N→D	16.76	24042.69	NA	NA

Table 2-5C: Sialidase and Galactosidase Treatment					
Proteoform			**Monoisotopic MH+**		
Glycan	**PTM**	**MT* (min)**	**Theo.**	**Obs.**	**Error (ppm)**
N34100	-	17.79	21457.82	21457.84	1.2
N34100	N→D	17.69	21458.8	21458.86	2.6
N34100	Des-1	17.94	21326.78	21326.83	2.3
N34100	Des-1 N→D	17.68	21327.76	21327.78	0.8
N35100	-	17.8	21660.9	21660.91	0.7
N35100	N→D	17.53	21661.88	21661.89	0.3
N45100	-	17.6	21822.95	21822.97	0.9
N45100	N→D	17.32	21823.93	21823.92	0.5
N45100	N→D	17.06	21823.93	21823.92	0.5
N36100	-	17.65	21863.98	21864.03	2.4
N36100	N→D	17.4	21864.96	21864.97	0.4
N46100	-	17.38	22026.03	22026.07	1.8
N46100	-	17.5	22026.03	22026.07	1.8
N46100	N→D	17.15	22027.01	22027.03	0.9
N46100	N→D	17.23	22027.01	22027.03	0.9
N47100	-	17.04	22229.11	22229.12	0.4
N47100	-	17.34	22229.11	22229.12	0.4
N47100	N→D	17.04	22230.09	22230.11	0.6
N47100	N→D	17.12	22230.09	22230.11	0.6
N57100	-	17.13	22391.16	22391.18	1
N57100	N→D	16.89	22392.15	22392.18	1.3
N58100	-	17.02	22594.24	22594.26	1
N58100	N→D	16.81	22595.23	22595.24	0.8

Tables 2-5 A-C: Quantitation of exoglycosidase digests of rhIFN-*β*1. Modifications: N→D, deamidation; Des-1, N-methionine loss; Succ, succinimide; SG, sulfated glycan. Glycoforms with more than one pair of deamidated/non-deamidated species are numbered in order of migration. 'Mixed' indicates that the deamidated/non-deamidated proteoforms for a given glycoform could not be accurately isolated for quantitation and the theoretical MH^+ used in calculating the error was that of the indicated PTM. * N = 4. ** N =5.

As seen in **Table 2-4**, all features with total relative abundance higher than 1.2% of total peak intensity, a total of 55 forms, were quantitated with less than 20% RSD (n = 6). The major species is N54120 and the 2-sialic acid-biantennary structure consisting of the unmodified form plus deamidation and other PTMs, representing 58% of the total of all forms. The next most abundant glycoforms are N65130 (9%) and followed by N54110 (7.5%). These results are in general agreement with published data [25] as well as peptide mapping (data not shown). However, **Table 2-4** provides significantly more structural detail than has previously been documented. Importantly, as many proteoforms consist of multiple modifications (glycans and PTMs), this quantitative information would not be available by bottom-up methods.

2.3.5. Correlation of Protein Modifications

From the quantitative data for the intact forms shown in **Table 2-4**, we can examine whether the appearance of a specific modification potentially correlates with one or more modifications. Consider the N54120 glycoform, the most abundant form present. Examining the deamidation and des-1 modifications (see **Figure 2-2B**), we present in **Table 2-4** and **Table 2-6** the relative percentages of the combination of 4 proteoforms (deamidated, non-deamidated, des-1 non-deamidated, des-1 deamidated) for the untreated and exoglycosidase-treated samples. It can be seen that the des-1 deamidated proteoform is more abundant than the des-1 non-deamidated form while the deamidated form is less abundant than the non-deamidated (unmodified) form.

Table 2-6A Untreated		(N→D)	
		+	-
Des-1	-	31	60
	+	6.2	2.7

Table 2-6B Sialidase		(N→D)	
		+	-
Des-1	-	28	57
	+	8.6	6.8

Table 2-6C Sialidase & Galactosidase		(N→D)	
		+	-
Des-1	-	37	55
	+	6.0	2.3

Table 2-6D Galactosidase		(N→D)	
		+	-
Des-1	-	40	61

	+	8.2	3.9

Table 2-6: Percent deamidation and des-1 methionine loss for most abundant glycoforms of rhIFN-*β*1. Untreated and Galactosidase: N54120, Sialidase: N54100, Galactosidase and Sialidase: N34100

If we assume that des-1 and deamidation are formed from the unmodified proteoform, then the form with two modifications (des-1 and deam.) cannot be more abundant than either of the single modified forms until the time when one of the single modified forms is higher in abundance than the unmodified form. This holds true even when the modifications occur sequentially. However, the form with the two modifications is higher in abundance than the des-1 form, even while the deamidated form is lower in abundance than the unmodified form (**Table 2-6**). The reason for the potential correlation could be a result of either preferential enrichment during batch processing or to a synergistic effect between the two modifications. More studies are necessary to examine this potential correlation further.

As a potential second correlation, it appears that the size of the glycan influences the extent of deamidation, as the fraction of the deamidated species relative to the non-deamidated species appears to decrease with glycan size (see **Table 2-6**). Specifically, for the untreated and galactosidase-treated samples (**Tables 2-7**) all glycans of size N54120 and larger had a deamidation fraction of 40% or lower. The N54110 pairs and smaller glycans, on the other hand, are 50% or greater. In fact N54100 is roughly 70% deamidated, whereas N54110 is 53% and N54120 is only 34%. Note also, for a given glycan structure, the increase in the extent of deamidation when Des-1 is present, up to 80% or more when des-1 is present. It should be pointed out that more work is necessary to verify such correlations and to understand underlying causes, but this type of information from intact quantitative protein analysis presents intriguing possibilities.

Table 2-7A Untreated			
Proteoform		Average	RSD *
Glycan	Modifications		
N44000		50.3	11.5
N54100		69.7	9.2
N34000		53.9	4.9
N54110	(1)	53.9	8.7
	(2)	51.6	5.2
	Des-1 (1)	83.3	5.0
	Des-1 (2)	79.3	3.8
N54120		34.3	4.0
	Des-1	69.9	4.6
	Succ.	22.9	5.3
N54220		32.1	5.3
N65130ǂ		35.4	6.6
N76130		35.8	6.7
N76140		28.3	5.5
N87130		41.4	3.3
N87140		33.5	6.5
N98130		40.4	10.3
N98140		40.0	8.7

Table 2-7B Galactosidase			
Proteoform		Average	RSD **
Glycan	Modifications		
N44000		51	7.9
N44110	(1)	58	1.7
	(2)	49	5.4
N54120		40	11
	Des-1	68	11
N65130		35	6.6
N76130		39	6.5

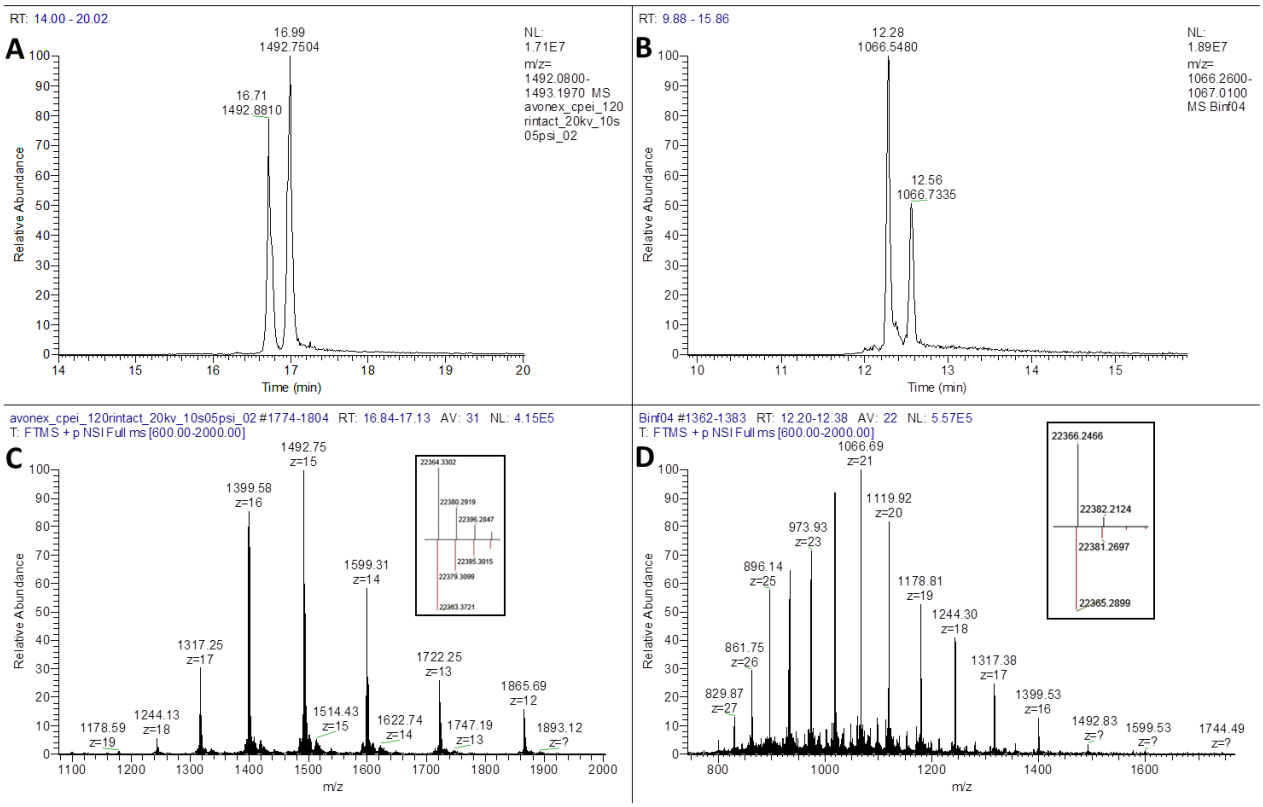
Table 2-7C Galactosidase + Sialidase			
Proteoform		Average	RSD ***
Glycan	Modifications		
N34100		41	0.52
	Des-1	72	3.9
N35100		37	1.0
N36100		41	2.8
N46100		41	5.6
N47100		39	4.4
N57100		47	5.8

Table 2-7: **Percent deamidation for glycoforms of rhINF-*β*1.** Modifications: Des-1, des-1 methionine loss; Succ, succinimide; SG, sulfated glycan. Glycoforms with more than one pair of deamidated / non-deamidated species are numbered in order of migration. * N = 6. ** N =5. *** N =4. ǂ Includes both α(1-3) and α(1-6) peaks of N65130. Averages are stated as the percent of the deamidated form of the total of the deamidated and non-deamidated forms.

2.3.6. CZE-MS Analysis of Reduced rhIFN-β1

In all the above studies, analysis was conducted with the single disulfide of Avonex in the oxidized form. In order to facilitate MS^2 characterization of intact species, we examined CZE-MS of rhIFN-β1 in the presence of the reducing agent, TCEP, in the sample buffer. If the disulfide bridge is left intact, fragmentation within and around the disulfide bonded region would be limited; reduction should facilitate fragmentation. Fresh TCEP solution was allowed to react with the sample for 30 min at room temperature, followed immediately by injection and CZE separation, without removal of the reducing agent or alkylation of the free cysteines.

Extracted ion electropherograms of the CZE separation and precursor charge state profiles for reduced and non-reduced major form, N54120, are shown in **Figures 2-11 A-D**. The precursor mass of the reduced forms (deamidated and non-deamidated) increased by 2 Da, relative to the oxidized forms, as expected. Opening of the disulfide bond leads to a more flexible denatured state and increased hydrodynamic volume. As a result, the reduced forms migrated faster in the reverse polarity mode (apparent mobility lower). Also, the reduced forms of the deamidated and non-deamidated rhIFN-β1 migrated in inverse order to their corresponding non-reduced forms.



Figures 2-11 A-D: **Comparison of non-reduced and TCEP reduced rhIFN-β1**. The extracted ion electropherogram for the **(A)** non-reduced and **(B)** reduced N54120. MS^1 of the non-reduced

N54120 peaks are shown in **(C)** and **(D)**; inserts show the deconvoluted mass for the averaged spectrum under the first and second migrating peaks (P1 and P2).

Reduction is seen to affect the charge states observed for the precursor ions. In the non-reduced species, the highest charge state is 19+, while in the reduced form, it is 28+, which corresponds to the fully protonated state, i.e., the total number of basic sites on the protein (11 K, 11 R, 5 H, N-terminus; **Figures 21C and D**). In addition, the charge state for the most intense ion increased from 15+ in the non-reduced to 21+ in the reduced sample, with the latter resulting in a larger number of charge states, as expected. Reduction led to a more open (denatured) rhIFN-β1 structure, allowing complete protonation and thus higher charge states of the protein.

2.3.7. Top-Down CZE-MS^2 of the Des-1 RhIFN-β1 Modification

Top-down CZE-MS^2 by both ETD and HCD fragmentation was performed on the reduced rhIFN-β1 N54120 form by targeting the isotope envelopes centered on the most abundant ion species of the tentative untruncated and des-1 proteoforms (1066.6 and 1060.6 *m/z*, respectively for z = +21). The MS^2 spectra were averaged across the corresponding CZE peaks and analyzed by Byonic, applying methionine loss, deamidation and glycosylation (*via* a human glycan database from Byonic) as variable modifications. Complete high resolution ETD and HCD MS^2 spectra acquired in the Orbitrap Elite (60,000 @ 400 *m/z*) are shown in the **Figure 2-12-24**. In **Figure 2-13**, it can be seen that HCD produces high levels of glycan fragmentation products at *m/z* less than 900 *Th*. In **Figure 2-14** we amplify the region between 900 and 1400 *Th* to reveal many *b*-and *y*-ions. **Figure 2-15** summarizes the primary sequence fragmentation pattern from three replicates of the untruncated (A) and des-1 (B) species where both ETD and HCD have been combined. Total sequence coverage was high at 76% and 80% for the unmodified and des-1 species, respectively. Only *c*-, *z*-, *b*-, and *y*-ions have been considered in this analysis in **Figure 2-15** (no neutral ion loss, etc.). It should be noted that the *c*-ions in the ETD spectra of the two species have a fragment number that differs by one amino acid residue, e.g., c_{13} of the untruncated form and c_{12} of the des-1 form correspond to the cleavage between the same residues Ser-Asn. This result is validates the intact MS^1 results that the modification is an N-terminal methionine loss. Since the deamidated form only differed from the non-modified form by 1 Da, we were not able to assign this site of PTM modification. However, it is well defined by the peptide map to be N23. **Figure 2-15** shows that detailed structural information can be obtained from top-down MS^2 ETD and HCD analysis from backbone fragmentation of rhIFN-β1.

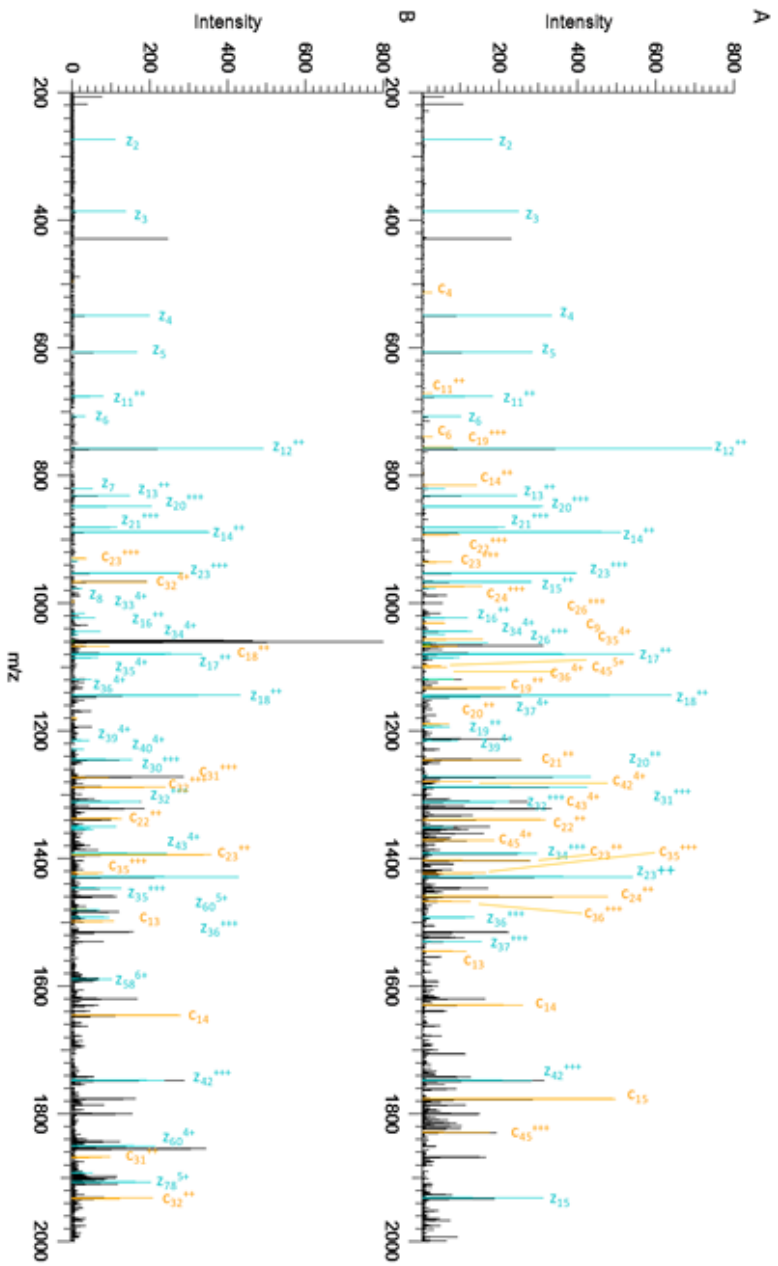


Figure 2-12: **Comparison of top-down ETD spectra of rhINF-β1.** Comparison of the N54120 (A) and des-1 N54120 (B) forms of rhIFN-β1.

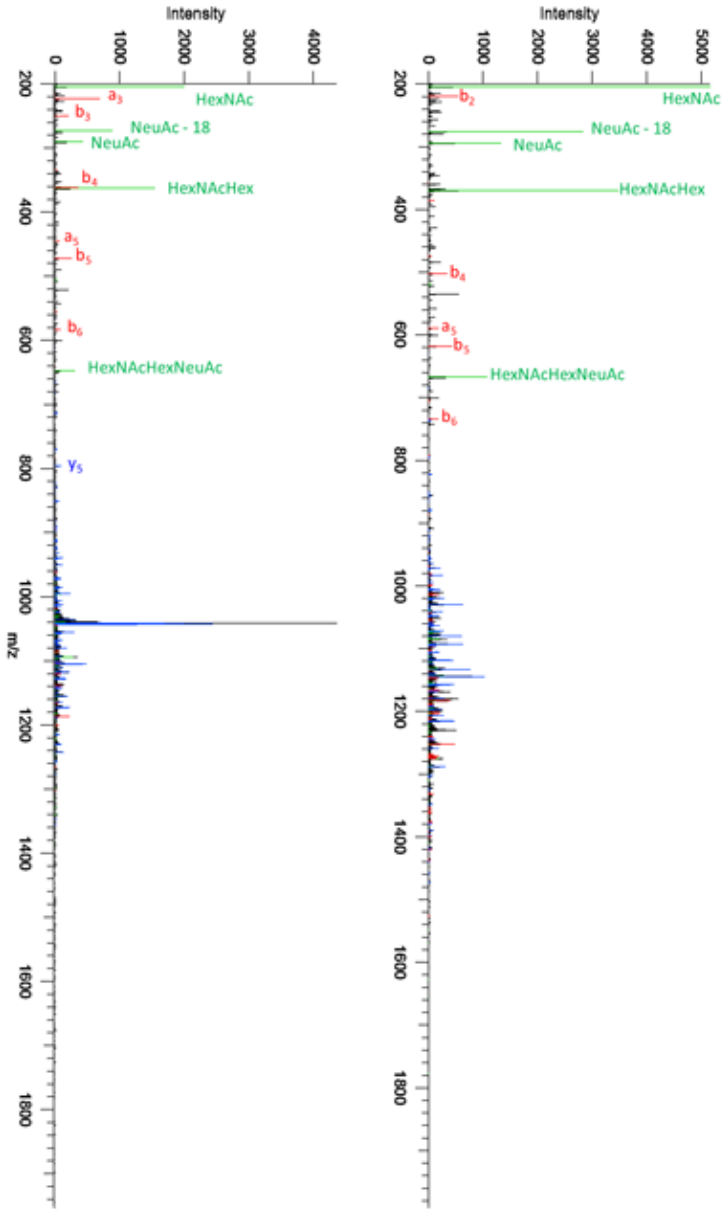


Figure 2-13: Comparison of the top-down HCD spectra of rhIFN-*β*1. Comparison of the N54120 (A) and des-1 N54120 (B) forms of rhIFN-*β*1. Only peaks between 200 and 900 m/z are annotated, an expanded view of the region between 900 and 1400 *m/z*.

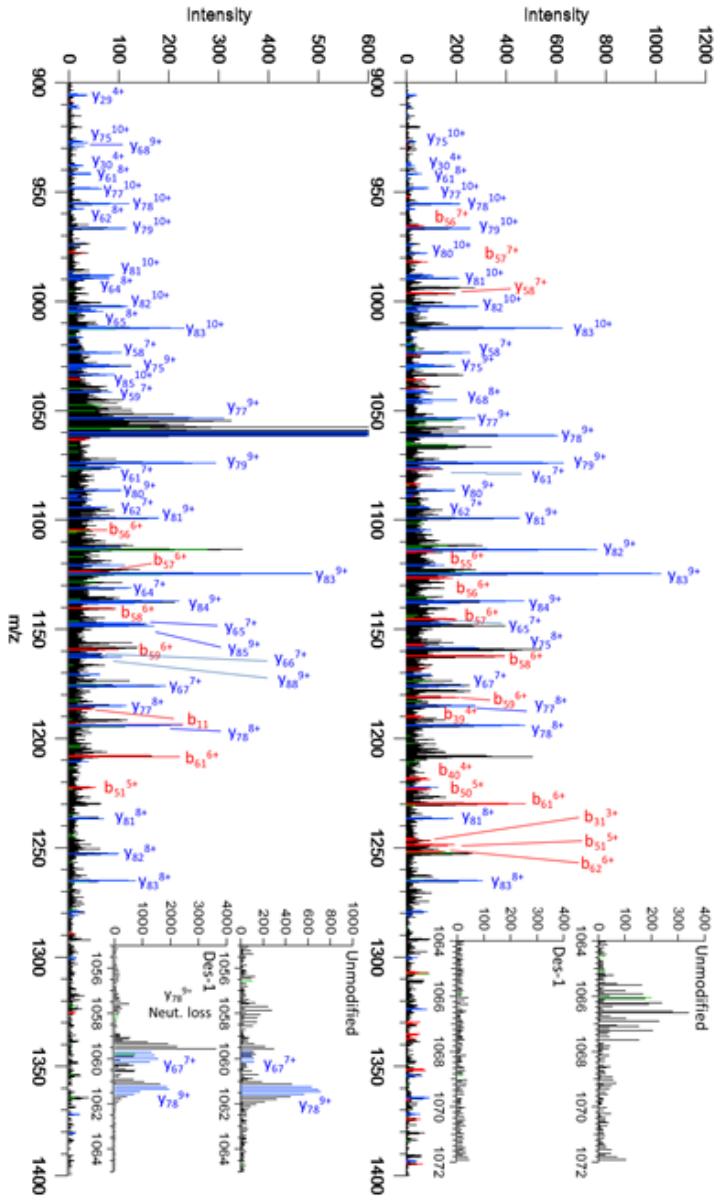


Figure 2-14: Expanded region of the top-down HCD spectra in Fig. 2-13. Top and bottom inserts show the regions surrounding the isolation window for the unmodified and des-1 forms.

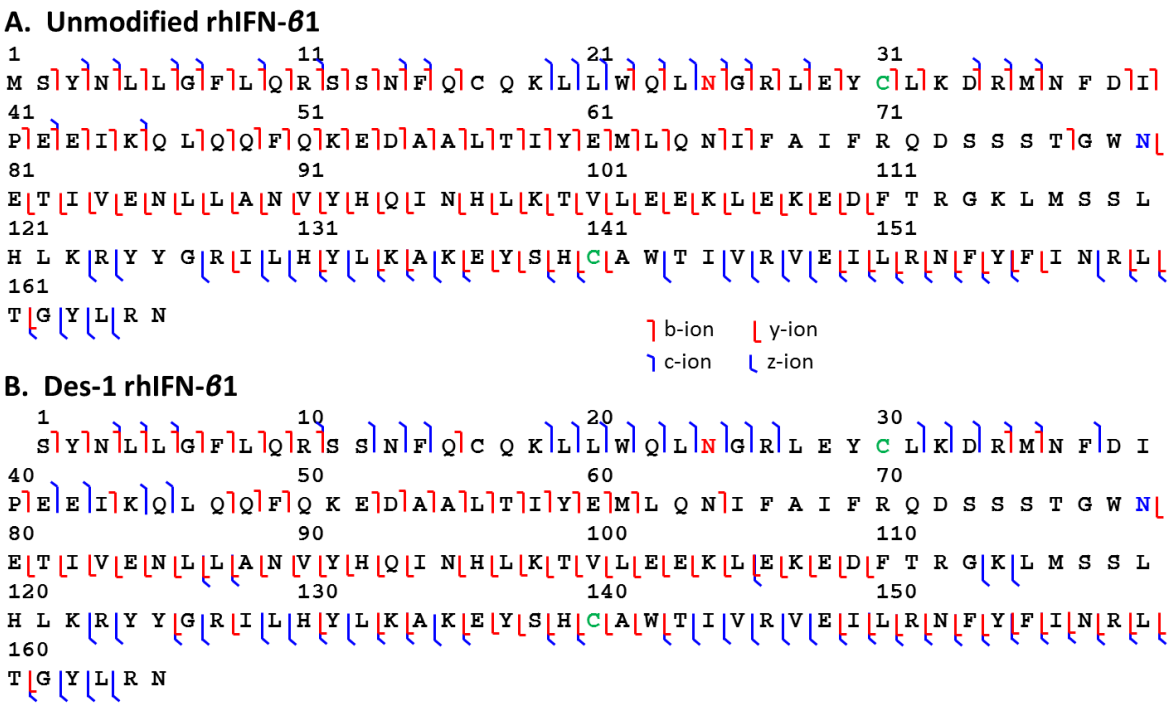


Figure 2-15: Summary of identified ETD and HCD fragmentation peaks for the unmodified and des-1 reduced N54120 rhIFN-*β*1. The following known modifications are indicated with color: disulfide bonded cysteines (**C**), deamidation site (**N**), glycan site (**N**). Ions observed in the spectra of the unmodified forms are in blue, des-1 in red. The coverage represents the sum of identified fragments from 3 experiments for both proteoform.

2.4. Conclusions

This paper demonstrated the power of high resolution CZE separation coupled to high resolution MS for intact protein analysis. High efficiency separation is needed to resolve basic forms, e.g., positional isomers, and those with small mass differences, e.g., deamidation. Given the potential for large proteoform combinations due to multiple glycan structures and PTMs, care must be exercised in identification-based intact mass. We used two forms of fragmentation (top-down MS/MS and exoglycosidase digestion) to aid in the elucidation of structures. Of course, proteins with multiple glycan sites would be more complicated. Nevertheless, the combination of high efficiency protein separation coupled to high resolution MS^1 and MS^2, along with a variety of enzymatic treatments should prove useful in the advancing the characterization of intact biopharmaceutical proteins. Ultimately, a combination of intact/top-down and bottom-up (and even middle-down) analyses will provide the most information on structure and heterogeneity of the protein. The study of intact and top-down MS analyses leads to the potential of finding correlations between individual glycoforms. A recent study showed that the number of C-terminal lysine and pyroglutamic acid moieties on mAbs influenced the size and processing rate

of glycans [261]. This result, along with potential correlations shown in this chapter, suggests that a detailed study of biopharmaceuticals by intact and top-down MS may lead to new ways of characterizing these complex therapeutics.

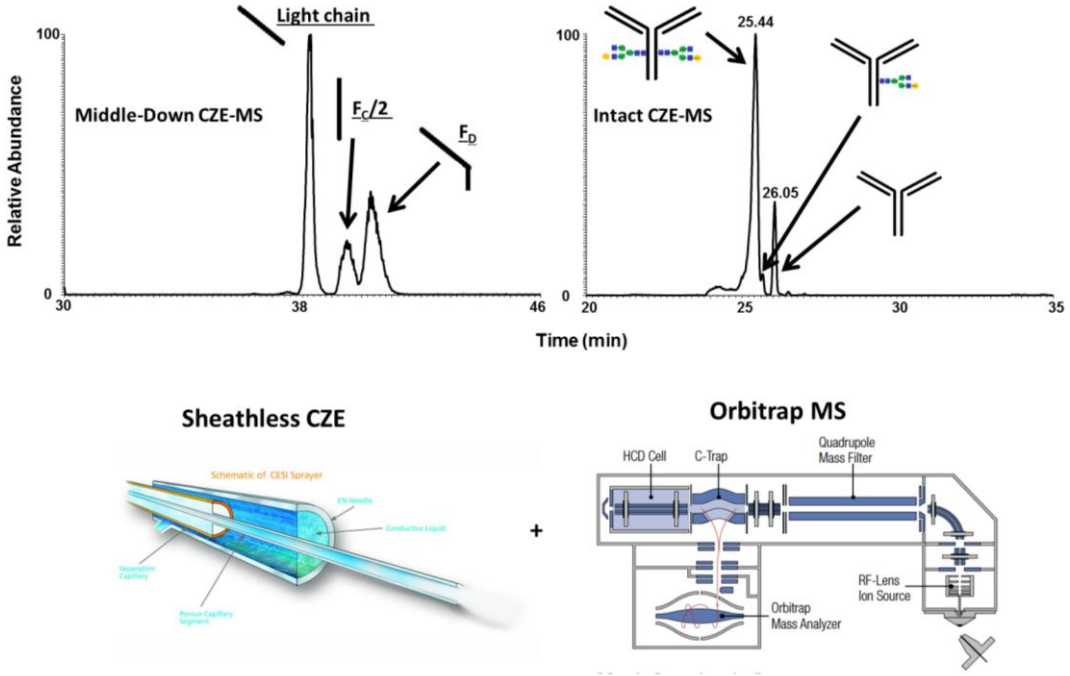
2.5. Acknowledgments

We would like to acknowledge Biogen Biopharmaceutical Development Department for their support through a Sponsored Research Agreement and Thermo Fisher Scientific through a technology alliance. We also thank Dr. Marcia Santos, Dr. Jeff Chapman, Dr. Andras Guttman, and Dr. Chitra Ratnayake at SCIEX, Dr. Neloni Ranmali Wijeratne at USC, Dr. Rosa Viner at Thermo Fisher Scientific, and Dr. Marshall Bern at Protein Metrics.

Chapter 3: Complementary Middle-Down and Intact Monoclonal Antibody Proteoform Characterization by Capillary Zone Electrophoresis – Mass Spectrometry

Arseniy M. Belov[1], Li Zang[2], Roberto Sebastiano[3], Marcia R. Santos[4], David R. Bush[5], Alexander R. Ivanov[1], and Barry L. Karger[1*]

[1] Barnett Institute of Chemical and Biological Analysis, 360 Huntington Ave., Boston, Massachusetts 02115, United States
[2] Analytical Development Department, Biogen, Cambridge, Massachusetts 02142, United States
[3] Department of Chemistry, Material and Chemical Engineering "Giulio Natta", Polytechnic of Milan, Via Mancinelli, 7 e 20131 Milan, Italy
[4]SCIEX, 250 South Kraemer Blvd., Brea, CA 92821, United States
[5] Genedata Inc, 1 Cranberry Hill, Lexington, Massachusetts 02421, United States



***My specific contributions to the work presented in this chapter included active involvement in experimental design, execution of all reported experiments, all data analysis, and manuscript preparation.**

3.0 Abstract

High resolution capillary zone electrophoresis – mass spectrometry (CZE-MS) has been of increasing interest for the analysis of biopharmaceuticals. In this work, a combination of the middle-down and intact CZE-MS analyses has been implemented for the characterization of a biotherapeutic monoclonal antibody (mAb) with a variety of post-translational modifications (PTMs) and glycosylation structures. Middle-down and intact CZE separations were performed in an acidified methanol-water background electrolyte on a capillary with a positively-charged coating (M7C4I) coupled to an Orbitrap mass spectrometer using a commercial sheathless interface (CESI). Middle-down analysis of the *IdeS*-digested mAb provided characterization of PTMs of digestion fragments. High resolution CZE enabled separation of charge variants corresponding to 2X-deamidated, 1X-deamidated, and non-deamidated forms at baseline resolution. In the course of the middle-down CZE-MS analysis, separation of glycoforms of the $F_C/2$ fragment was accomplished because of hydrodynamic volume differences. Several identified PTMs were confirmed by CZE-MS^2. Incorporation of TCEP reducing agent in the CZE injection buffer resulted in successful analysis of reduced forms without the need for alkylation. CZE-MS studies on the intact mAb enabled baseline separation of the 2X-glycosylated, 1X-glycosylated, and aglycosylated populations as a result of hydrodynamic volume differences. The presence of trace quantities of dissociated light and heavy chains were also detected in the intact protein analysis. Analysis of the mAb under native conditions verified identifications achieved *via* intact mass analysis and allowed for quantitative confirmation of proteoforms. The CZE-MS analysis of different mAb proteoforms was found to be complementary to the more conventional LC-MS approach.

3.1 Introduction

Monoclonal antibodies (mAbs) are a major class of protein pharmaceuticals, with currently over 50 FDA-approved, commercially-available drugs [10]. Moreover, a large number of additional therapeutic mAbs are currently under development, including antibody-drug conjugates, bispecific antibodies, and antibody carriers [262]. Monoclonal antibodies, as well as other protein pharmaceuticals, consist of mixtures of proteoforms, which may differ in glycosylation and other post-translational modifications (PTMs), sequence truncation, or other modifications that can occur at various stages of production, purification, or storage [4, 263]. Given that these proteoforms may result in differences in biological activity, stability, or toxicity, comprehensive characterization of the protein therapeutic is necessary for drug approval and safety.

Conventional mAb characterization has typically been conducted by "bottom-up" liquid chromatography (LC) – mass spectrometry (MS), based on characterization of product peptides

from proteolytic enzyme digestion. However, the bottom-up approach may lead to sample preparation modifications and, importantly, precludes the ability to identify individual proteoforms, often consisting of combinations of multiple modifications [44, 264]. Thus, methods for intact and middle-down protein analysis are being developed [19, 265].

Analysis of intact mAbs is typically performed in order to determine protein molecular mass, extent of oligomerization, and additional structural information, such as the presence of free chains [266], glycosylation [1, 25, 267], as well as various PTM combinations [1, 268]. On conventional Orbitrap™ instruments, MS resolution is inversely proportional to the square root of *m/z*, which makes baseline resolution of higher molecular weight species challenging. This is applicable to intact mAbs, which possess multiple glycosylation sites and exhibit signals in the relatively high *m/z* range (~*m/z* 4000-7000) [19]. In addition to the intact analysis, characterization *via* digestion of the ~150 kDa mAb into large fragments (middle-down) has also been explored [265, 269]. Middle-down analysis of mAbs, characterized by lower molecular weight precursors, has the potential to resolve subtle differences in the *m/z* domain, as well as to detect and identify low-abundance PTMs. Moreover, precursor ion selection with an analytical quadrupole on a commercial Q Exactive™ Orbitrap mass spectrometer is confined to an upper *m/z* limit of 2,500 *Th* [270], thereby precluding the ability to select the most abundant charge states of an intact mAb for MS^2 analysis. Conversely, limited mAb digestion (e.g., using *IdeS* enzyme) and ionization under denaturing conditions yield both the precursor and fragment ion species, which are amenable to analysis on a commercial Orbitrap™ MS [267]. In this study, both the middle-down and intact analyses have been explored.

Critical to the success of MS analysis of a heterogeneous mixture of a mAb is separation prior to MS detection. Typically, various 1D or 2D capillary liquid chromatography (LC) approaches are employed either online or offline to MS. These approaches include reversed-phase liquid chromatography (RPLC) [267], size exclusion chromatography (SEC) [271], ion exchange chromatography (IEX) [163], hydrophobic interaction chromatography [272], or hydrophilic interaction liquid chromatography [273]. An alternative high resolution technique is capillary zone electrophoresis (CZE), which separates analytes based on differences in charge and hydrodynamic volume. CZE has been demonstrated to be a high efficiency separation method for the analysis of proteins [25, 26, 274, 275]. In contrast to LC (except low resolution SEC), CZE does not involve interaction of the analyte with the stationary phase, thus representing a complementary separation approach.

A critical component of coupling CZE to ESI-MS is the interface junction. The most straightforward CZE-ESI-MS junction involves a sheath-liquid, and there have been successful reports of protein analysis using this approach in both microcapillary and microchip formats [60, 61, 276]. An alternative approach is the use of a sheathless junction, in which the background electrolyte (BGE) is electrosprayed directly from the separation capillary without dilution from the sheath liquid [65, 67]. A commercial sheathless system (CESI) has been introduced and shown to be highly sensitive for the analysis of peptides and proteins, including glycoproteins [1, 25, 26]. A recent study from our laboratory using the CESI interface has demonstrated the

separation of recombinant human interferon beta-1 (rhINF-β1, ~23,000 Da), where 138 proteoforms were detected and 55 quantitated [1]. In another study, mixtures of native proteins and protein complexes, a monoclonal antibody, and the protein extract from the *E. coli* ribosomes were characterized using CESI interface in one of the first examples of native CZE coupled online to ESI-MS [2].

In the present study, middle-down and intact CZE-MS has been explored for the analysis of a therapeutic monoclonal antibody of the immunoglobulin type 1 (referred to as "mAb" throughout this manuscript). Experiments were performed with a positively-charged capillary surface (1-(4-iodobutyl) 4-aza-1-azoniabicyclo[2,2,2] octane iodide (M7C4I) [39]) using an acidic methanol/water background electrolyte (BGE). Middle-down analysis was performed *via* digestion with *IdeS* (FabRICATOR) protease with and without subsequent reduction. Major PTMs of the mAb, including glycosylation, oxidation, deamidation, pyroglutamation, carboxymethylation of lysine, and glycation have been observed. Confirmation of many of the proteoform identities was achieved by middle-down CZE-MS^2 using higher collision energy dissociation (HCD). Separations of charge variants arising from 2X-deamidated, 1X-deamidated, and non-deamidated species, as well as other proteoforms differing in hydrodynamic volume were attained. With respect to the intact mAb, we have also achieved separations of a variety of glycosylated structures, including 2X-glycosylated, 1X-glycosylated, and fully aglycosylated populations. In addition, trace quantities of non-covalently bound light and heavy chains were observed under the denaturing conditions of the BGE. Different forms of the intact mAb observed under denaturing conditions (M7C4I capillary) were also quantitated and with the observed abundances were verified in the analysis performed under native conditions (polyacrylamide-coated capillary). These results demonstrate the potential of CZE-MS as a complementary approach to LC-MS in the middle-down and intact characterization of mAbs.

3.2. Experimental

3.2.1 Materials and Chemicals

Methanol, isopropanol, acetic acid, formic acid, sodium hydroxide, sodium tetraborate, boric acid, sodium chloride, and TCEP-HCl were obtained from Sigma Aldrich (St. Louis, MO, USA). Purified mAb stocks (51.1 mg/mL) were used as samples. The (M7C4I) reagent was synthesized according to previous reports [277]. *IdeS* was purchased from Genovis (Lund, Sweden). Micro Bio-Spin 6 columns, used for sample clean-up, were purchased from BioRad (Hercules, CA). All sheathless bare fused silica (BFS) and polyacrylamide-coated capillaries were provided by SCIEX (Brea, CA).

3.2.2. Sample Preparation

For middle-down CZE-MS with *IdeS* digestion, the mAb in formulation was first diluted to 1 mg/mL in 50 mM ammonium acetate, pH 6.5. Next, *IdeS* was added according to the manufacturer's specifications and allowed to digest the protein for 30 minutes at 37^0 C, followed by clean-up with Micro Bio-Spin 6 columns. For subsequent reduction, TCEP-HCl was added to the cleaned-up sample to a final concentration of 5 mM, and reduction was performed for 60 minutes at room temperature. For intact CZE-MS, the mAb was diluted to 1 mg/mL in 50 mM ammonium acetate, pH 4.0, cleaned-up with Micro Bio-Spin 6 columns, immediately followed by sample injection. For native CZE-MS, the mAb was diluted to 1 mg/mL in 40 mM ammonium acetate, pH 4.0 from formulation. The mAb was then cleaned up with Micro Bio-Spin 6 columns, immediately followed by sample injection.

3.2.3. CZE Methods

All CZE-MS operations were performed on a CESI 8000 instrument (SCIEX, Brea, CA). For the application of the M7C4I coating onto a bare fused silica (BFS) capillary, previously-described procedures [278, 279] were followed with minor modifications. Briefly, a sheathless BFS capillary was conditioned with 0.1 M sodium hydroxide for 15 minutes at 100 psi. Next, water (15 minutes, 100 psi), followed by 25 mM sodium tetraborate, pH 9.8 (3 minutes, 100 psi) was flown through the capillary. A 4 mM solution of M7C4I in 25 mM sodium tetraborate, pH 9.8, was then flushed through the capillary at 75 psi for 3 minutes, followed by a water rinse (100 psi for 3 minutes). Lastly, the coated capillaries were then conditioned with BGE at 100 psi for 10 minutes, and BGE was flushed through the conductive capillary at 75 psi for 5 minutes. For middle-down experiments, the optimum BGE was determined to be 50% methanol, 1% formic acid. For intact experiments, the optimum BGE was 10% isopropanol, 0.2% formic acid.

Even though the M7C4I coating is a covalent coating, we found that the reagent is removed slowly over time, resulting in changes in the electroosmotic flow (EOF). To account for changes in EOF, we employed a washing procedure prior to each new separation that involved brief rinses with the M7C4I reagent in 25 mM sodium tetraborate, pH 9.8 (75 psi, 3 min). These were followed with water (100 psi, 3 min) and BGE rinses (100 psi, 10 min), for a total washing time of 16 minutes. After a few days of continual use, we completely recoated the capillary, as described above.

All CZE methods involving the M7C4I-coated capillary employed -20 kV with ramp times of 1 minute. Sample injections were performed at 1 psi for 10 seconds, corresponding to a 1.7 nL injection volume. Following each separation, a 5 minute ramp down time to -1 kV was programmed in order to protect the integrity of the porous sprayer tip.

For middle-down CZE-MS/CZE-MS^2 experiments with a BGE of 50% methanol, 1% formic acid, the optimum spray potential was determined to be 1.3 kV and was employed to

maintain stable nanoESI. For intact CZE-MS experiments involving the 10% isopropanol, 0.2% formic acid BGE, a 1.4 kV spray potential was used.

For native CZE-MS analysis of the mAb, a polyacrylamide-coated capillary (SCIEX, Brea, CA) was used. Prior to each injection, a series of rinses of separation and conductive lines were first performed. For the separation capillary these included: 0.1 M hydrochloric acid (100 psi, 5 min), DDI water (100 psi, 5 min), followed by the 40 mM ammonium acetate, pH 7.5 BGE (100 psi, 10 minutes); the conductive capillary was rinsed with a 10% acetic acid at 75 psi for 4 min. Sample injection was performed at 2.5 psi for 15 seconds, corresponding to approximately 6 nL of sample (0.8% of the capillary volume). Separations were performed for 30 minutes at 20 kV with a ramp time of one minute. Due to the significant suppression of EOF on polyacrylamide-coated capillaries, a supplemental pressure of 3 psi was applied to generate flow of solvent and facilitate the migration of analytes towards the MS detector. At the end of the run, a ramp-down from 20 kV to 1 kV over five minutes was performed back to the start conditions for the next injections.

3.2.4. MS Instrumentation

For middle-down analysis, a Q Exactive™ Plus Orbitrap™ MS (Thermo Fisher Scientific, Bremen, Germany) was used. Middle-down analysis on the Q Exactive Plus was performed with the intact protein mode enabled, allowing for low pressure trapping (1.80e-5 mbar in the C-trap/HCD cell) and collisional cooling of proteins in the C-trap. For intact mAb analysis, MS detection was performed under the standard pressure conditions.

For both middle-down and intact CZE-MS, the optimum in-source collision-induced dissociation energy (in-source CID) was established to be 50 V x charge, allowing for enhanced desolvation of the precursor ions. For middle-down schemes on the Q Exactive™ Plus, analysis was performed with an automatic gain control (AGC) of 10^6, a maximum injection time of 250 ms, a resolving power of either 17,500 or 140,000 at *m/z* 200, an inlet transfer capillary temperature of 110^0 C, and an averaging over 3 microscans. The nominal resolving power of 17,500 was used in CZE-MS experiments, while 140,000 was selected for precursor ion scans in CZE-MS^2 experiments. For CZE-MS experiments with intact forms, detection settings were identical to those described above, using a nominal MS resolving power of 17,500.

For CZE-MS^2 experiments, a "Top 10" method was used. Precursor scans were performed as described above. For MS^2 scans, instrument resolving power was set to 140,000 at *m/z* 200 with averaging over 5 microscans. AGC was $5x10^5$, with a maximum injection time of 250 ms. An isolation window of 6 *Th* was selected, and the optimum normalized collision energy was determined to be 23 V x charge.

For native CZE-MS analysis of the mAb, the sheathless polyacrylamide-coated capillary was coupled to the Q Exactive^{TM} Plus Orbitrap^{MS} MS. A 1.4 kV spray potential was determined to be the optimum ESI voltage. Protein collisional cooling and trapping were performed in the HCD cell of the instrument, as opposed to the C-trap in the intact (denaturing) and middle-down

experiments. Give the relatively high *m/z* range for an intact mAb under native conditions (5,000 – 7,000 *m/z*), detection required optimization of the in-source CID voltage and corresponding voltages in the HCD cell. For native analysis, an in-source CID energy of 40 eV x charge, in addition to an HCD energy of 75 eV x charge were used for optimum collisional cooling and desolvation. Trapping and collisional cooling were also performed at elevated gas pressures in the HCD cell, corresponding to approximately 1.02e-4 mbar for native CZE-MS experiments. This trapping pressure is considerably higher than standard trapping pressure used in intact (denaturing) (4.37e-5 mbar) and reduced trapping pressure (1.80e-5 mbar) used for middle-down analyses. For native CZE-MS, an AGC of 2.0e5 and a maximum injection time of 50 msec were used. Finally, the instrument was scanned from 1,500 – 10,000 *m/z* at a resolving power of 17,500 at 200 *m/z*.

3.2.5. Data Analysis

Deconvolution of CZE-MS experiments was achieved with BioPharma Finder 1.0 (Thermo Fisher Scientific, San Jose, CA). For experiments performed at 17,500 resolving power, the Sliding Window algorithm with ReSpect was used. For isotopically-resolved data with 140,000 resolving power, the Sliding Window algorithm using Xtract was employed.

For analysis of CZE-MS^2 experiments acquired on the Q Exactive™ Plus, Byonic (version 2.5.6) (Protein Metrics, San Carlos, CA) was utilized. Searches were performed using fully specific digestion with a maximum of 2 missed cleavages. A 10 ppm precursor mass tolerance and a 20 ppm fragment mass tolerance were used in all searches, along with a 1% false discovery rate (FDR). Common and expected PTMs were manually added to the search bar and used as dynamic modifications. Modifications included oxidation, pyroglutamation, deamidation, N-terminal methionine-loss, glycation, di-oxidation, and carboxymethylation of lysine. A database of 38 common biantennary glycans was also included in the searches.

For deconvolution of native CZE-MS, Intact Mass software version 1.6 (Protein Metrics) was used. A deconvolution range corresponding to 5,000 – 6,600 *Th* was selected with a 30 ppm tolerance. A range corresponding to 17 to 20 minutes of the electropherogram was selected for deconvolution. The above deconvolution range was selected in order to allow for accurate deconvolution and quantitation of 2X-glycosylated, 1X-glycosyalted, and aglycosyalted populations of the mAb.

3.3 Results and Discussion

3.3.1. Fundamentals of CZE Operation Using Positively-Charged Capillary Coatings

Critical aspects to large fragment or intact protein analysis by CZE-MS are the capillary coating and BGE. Use of uncoated bare-fused silica capillaries under acidic BGE conditions can result in significant band broadening and poor recovery due to adsorption of positively-charged

proteins to the negatively-charged silanol groups on the capillary surface [274]. Numerous neutral and positively-charged capillary coatings have been developed to overcome the adsorption issue [268, 274]. On a capillary with a positively-charged coating and an acidic BGE, the drag force exerted on positively-charged analytes by the electroosmotic bulk flow (EOF) and the electrostatic field force, causing analyte migration towards the cathode, act in opposite directions. **Equation 6** defines the resolution (R_S) for separation of solutes "a" and "b" in the presence of EOF. R_S is determined by the number of theoretical plates (N), the free solution mobilities of solutes "a" and "b" (μ_{FSa} and μ_{FSb}, respectively), their average free solution mobility ($\overline{\mu}_{FS}$), and the EOF mobility (μ_{EOF}) [47]. The latter two parameters are opposite in sign. **Equation 6** shows that reducing the sum of $\overline{\mu}_{FS}$ and μ_{EOF} results in an increase in R_S.

$$R_S = \frac{\sqrt{N}(\mu_{FSb} - \mu_{FSa})}{4(\overline{\mu}_{FS} + \mu_{EOF})}$$

Equation 6: Resolution between two solutes, "a" and "b".

For individual analytes, the apparent solute mobility, μ_{APP}, as well as μ_{FS} and μ_{EOF} are governed by **Equations 1-3,** introduced in Chapter 1. It can be seen from **Equations 1-3** that an increase in the solute's Stokes' radius, which is proportional to the hydrodynamic volume, results in a decrease of the free solution mobility, μ_{FS}, and an increase in the apparent solute mobility, μ_{APP}. Therefore, for a given net charge, solutes with larger hydrodynamic volumes would exhibit shorter migration times than more compact counterparts. Similarly, higher net charge solutes are characterized by higher free flow mobilities, giving rise to longer migration times.

3.3.2. Middle-Down Analysis: Products of Proteolytic Digestion and Reduction

In this work, middle-down analysis of the mAb involving digestion with *IdeS* was explored. **Figure 3-1** in the Supporting Information Section shows the expected products from this digestion**.** Briefly, digestion with *IdeS* resulted in cleavage between Gly residues in the hinge region of the antibody, yielding $F(ab')_2$ (98.4 kDa) and two $F_C/2$ (~24 kDa) fragments. Subsequent reduction resulted in the dissociation of $F(ab')_2$, yielding two light chains (Lc) (23.2 kDa) and two F_D fragments (26.0 kDa), in addition to the two $F_C/2$ fragments.

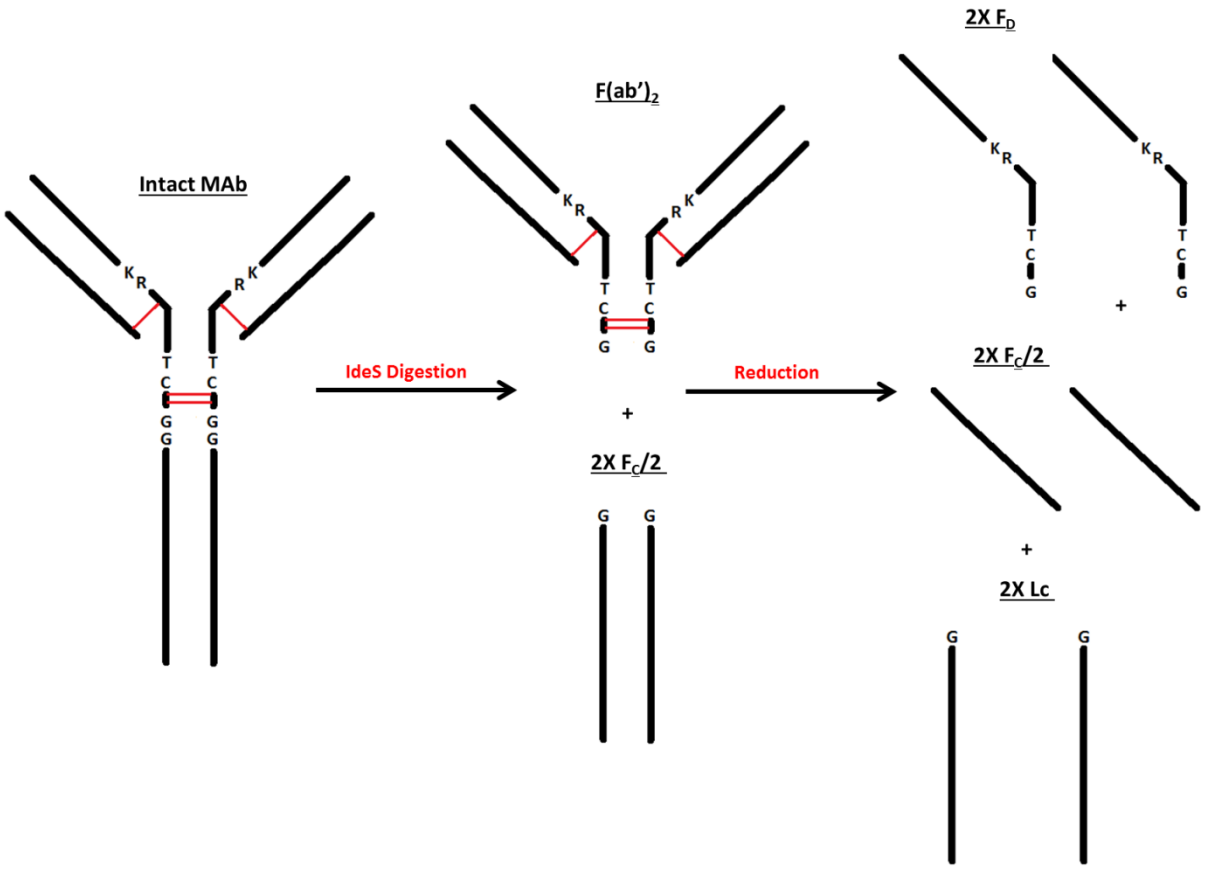


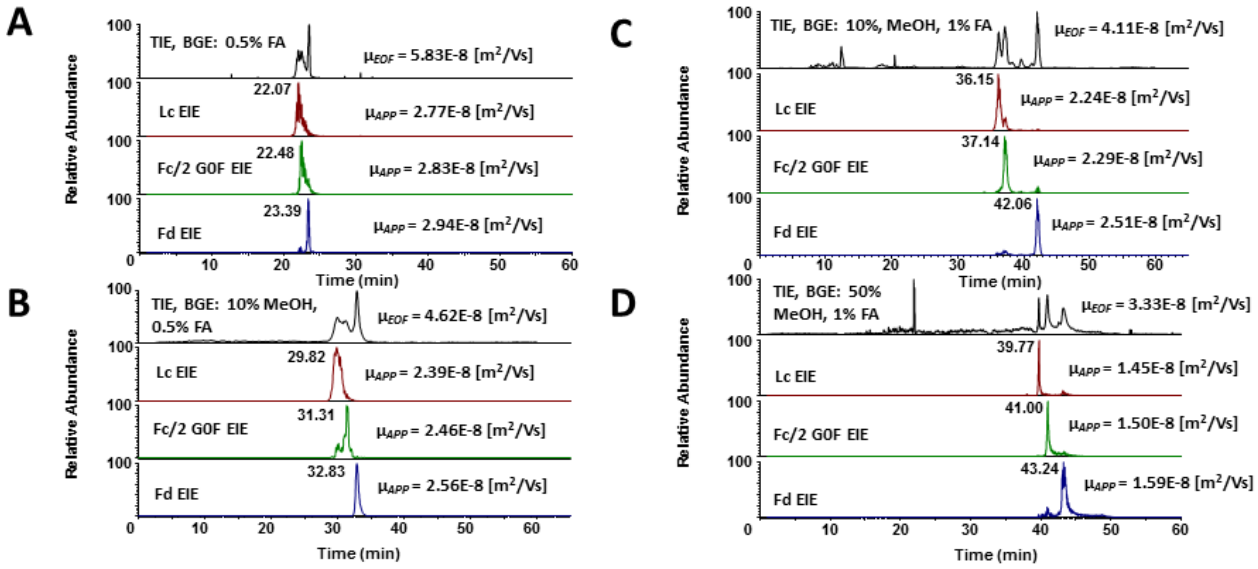
Figure 3-1. Middle-down fragments of a monoclonal antibody following digestion with *IdeS* protease. Middle-down schemes performed in this study for the mAb involved either digestion with *IdeS* or digestion with *IdeS* with subsequent reduction with TCEP-HCl.

3.3.3. Middle-Down Analysis: CZE-MS Separation Parameters

Despite a recent success in the characterization of rhINF-β1 using a positively charged cross-linked polyethyleneimine (cPEI) coating [1], intact and middle-down analyses of the mAb under the same conditions did not yield separation of various proteoforms, as a two- to three-fold difference between μ_{FS} of individual analytes of the intact and middle-down products of the mAb relative to μ_{EOF} was observed on the cPEI-coated capillary under the acidic conditions (results not shown). Hence, we aimed at increasing separation performance (R_s) by reducing the EOF (see **Equation 6**). A positively-charged capillary coating with a lower density of charged groups relative to cPEI was thus selected (M7C4I) [280, 281]. In comparison to cPEI, the M7C4I-coated capillary has been reported to generate a lower reverse EOF [278, 282].

We then explored a variety of BGEs under acidic conditions to optimize separation on the M7C4I-coated capillary. As an example, **Figure 3-2** shows separations of the *IdeS*-digested,

reduced mAb in different BGEs; corresponding μ_{APP} values are indicated for each digestion product. Using a BGE of 0.5% formic acid, poor separation of digestion fragments Lc, F_D, and G0F F_C/2 was obtained. Improvement occurred as 10% methanol, 1% formic acid was used. Interestingly, lowering the formic acid concentration to 0.5% in 10% methanol significantly degraded the separation performance as a result of the higher value of μ_{APP}. Further increases in methanol with 1% formic acid improved the separation, reaching the optimum at 50% methanol. A less polar organic solvent, isopropanol, was also tested. Up to 50% isopropanol with 1% formic acid was utilized. While resolution was slightly improved, it was found that the analysis times were almost doubled over that of the 50% methanol, 1% formic acid BGE (results not shown). As a result, the BGE consisting of 50% methanol, 1% formic acid was selected, as it provided adequate separation performance of *IdeS*-digested, reduced mAb within a reasonable analysis time. Furthermore, in agreement with others [279], reproducibility was found to be in the range of 2-3 percent CV. The selection of the BGE for the separation of intact proteoforms of the mAb will be discussed later in the section on intact analysis.

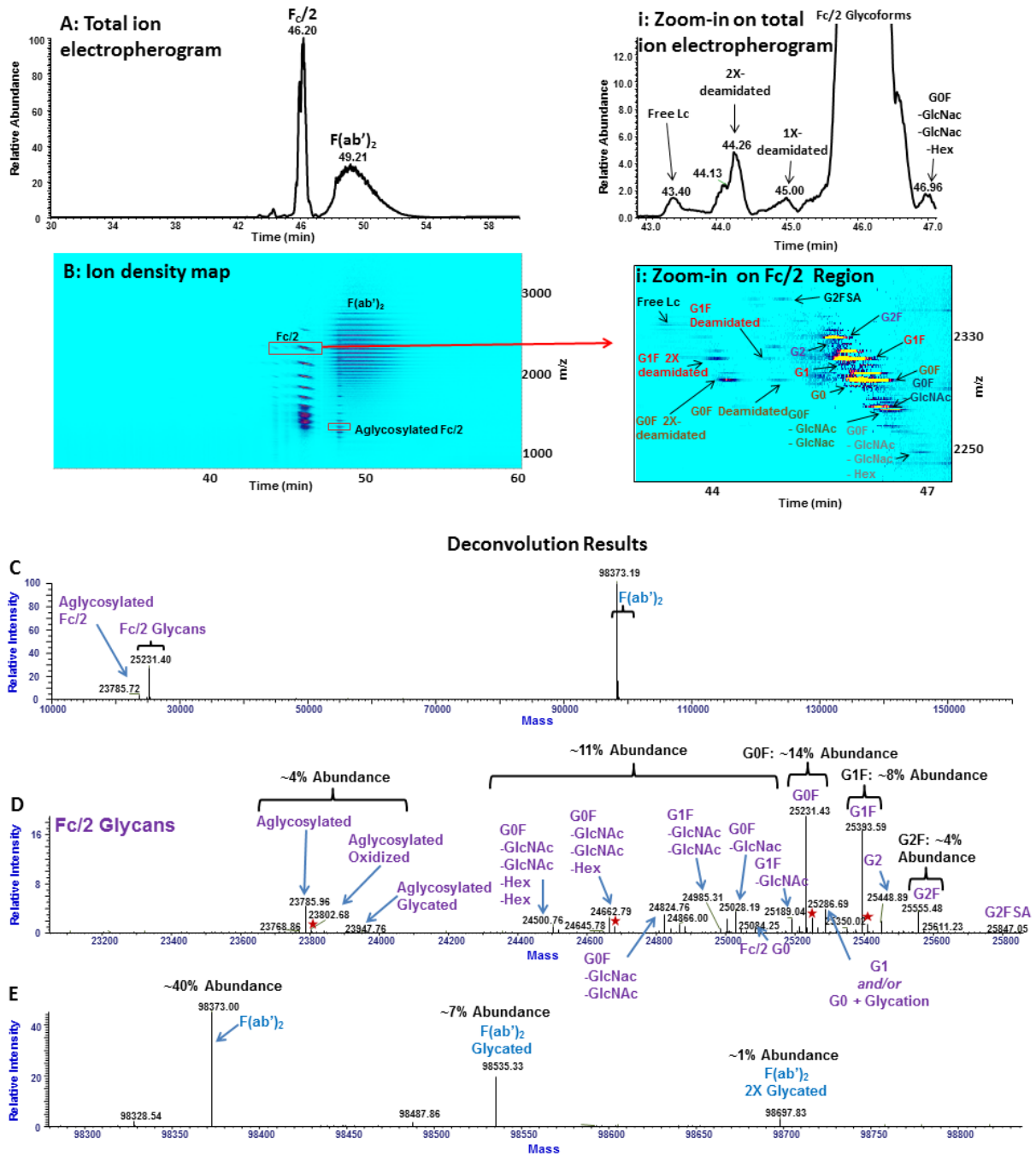


Figures 3-2 A-D. **Comparison of separation efficiencies for the *IdeS*-digested, reduced mAb at different background electrolytes (BGEs).** Each experiment is represented by a total ion electropherogram (TIE) trace, in addition to the extracted ion electropherograms (EIEs) of the most abundant charge states of Lc, F_C/2 G0F, and F_D species for the *IdeS*-digested, reduced mAb. EOF mobility (μ_{EOF}) is also annotated for each trace. (**A**) Separation performed in a 0.5% formic acid (FA) BGE. (**B**) Separation performed in the 10% methanol (MeOH), 0.5% FA BGE. (**C**) Separation in the 10% MeOH, 1% FA BGE. (**D**) Separation in the 50% MeOH, 1% FA BGE.

3.3.4. Middle-Down Analysis: Analysis of *IdeS*-Digested MAb

IdeS-digested mAb was analyzed on the sheathless M7C4I coated-capillary – Q Exactive™ Plus system with a 50% methanol, 1% formic acid BGE, as shown in **Figure 3-3**. Separation between the major digestion fragments, $F_C/2$ and $F(ab')_2$, was achieved at baseline resolution (**Figures 3-3 A-B**). Deconvolution of the CZE-MS electropherogram resulted in the identification of numerous forms for these two digested fragments with approximate relative abundances indicated in **Figures 3-3 C-E**. For the ~24 kDa $F_C/2$ fragment, 15 different glycoforms were identified, in addition to the aglycosylated form. Sequential additions of two deamidations were detected for G0F and G1F glycoforms, in addition to a single deamidation for the G2F glycoform. For several glycoforms of $F_C/2$ and aglycosylated $F_C/2$ species, low-level oxidations, as well as oxidized and glycated forms were identified (**Figure 3-3D**).

For the *IdeS*-digested mAb experiment, relative ratios of a variety of proteoforms are reported in **Figures 3-3 D-E**. As an example, the G0F glycoform was found to be the most abundant $F_C/2$ species at roughly 14%. The G1F was the second-most abundant glycoform, at approximately 8%, followed by the G2F glycoform (4%). Proteoforms of the aglycosylated $F_C/2$ were also present at approximately 4% (**Figure 3-3D**). Quantitative measurements were found to be within 1% RSD between 3 experimental replicates of the *IdeS*-digested, non-reduced G0F and G1F glycoforms of the $F_C/2$ fragment of the mAb (results not shown).



Figures 3-3 A-E. Characterization of the *IdeS*-digested mAb. (**A**) The total ion electropherogram. (**A(i)**) 2X- and 1X-deamidated $F_C/2$ forms are resolved from the 0X-deamidated species and are shown as an inset to the total ion electropherogram. (**B**) The ion density map matching the electropherogram in (**A**). Inset (**B(i)**) shows a zoomed-in view on the map in (**B**) representing the 11+ charge state of the $F_C/2$ species. (**C**) Deconvoluted mass spectrum corresponding to the $F_C/2$ and $F(ab')_2$ fragments. (**D**,**E**) Insets into the regions corresponding to $F_C/2$ and $F(ab')_2$ fragments, respectively, in (**C**). Various oxidized proteoforms

in **(D)** are labeled as red stars. Approximate abundances of $F_C/2$ and $F(ab')_2$ proteoforms are shown.

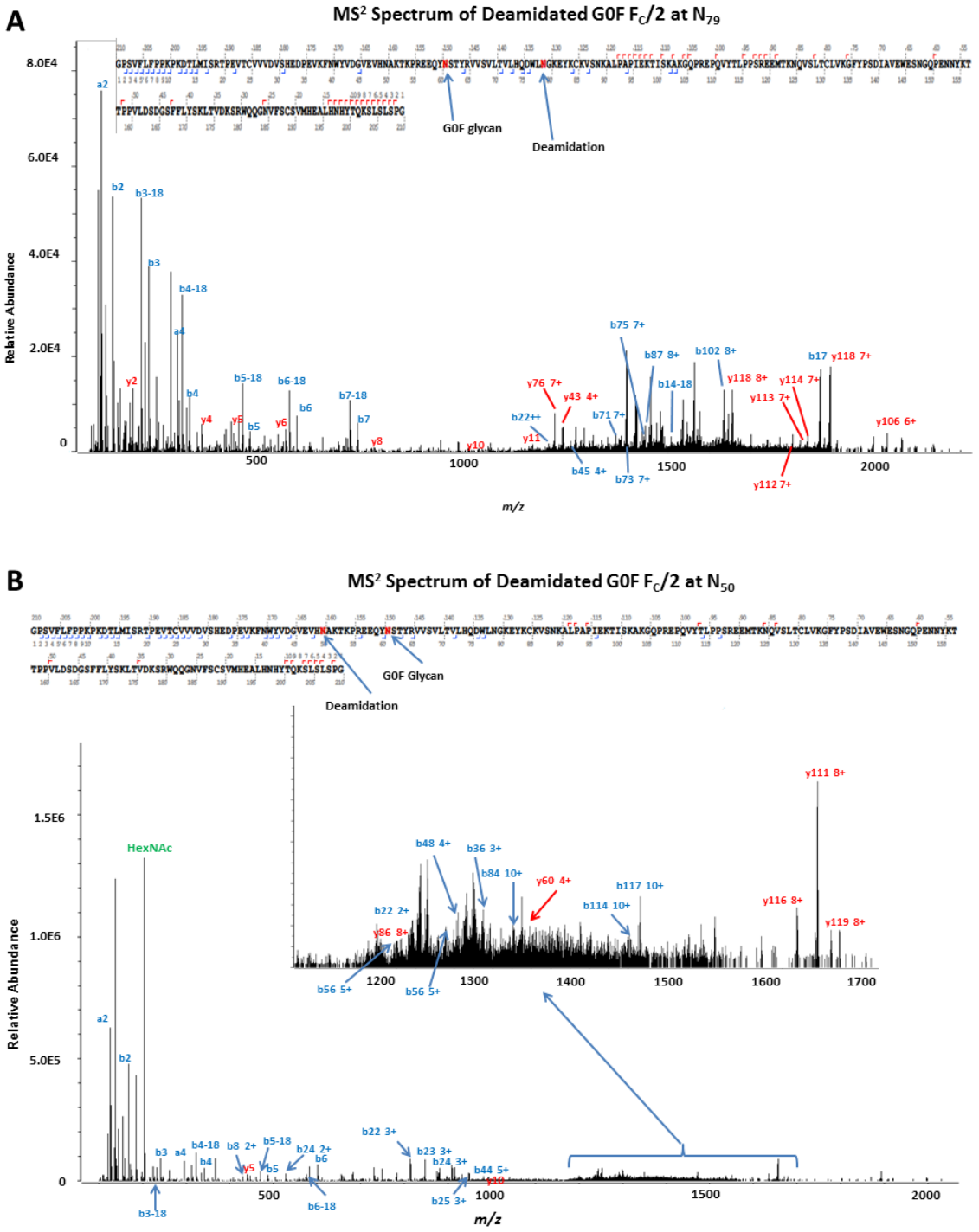
As shown in the ion density map in **Figure 3-3B(i)**, partial separations of individual glycoforms of the $F_C/2$ fragment were achieved. $F_C/2$ proteoforms possessing larger glycans (e.g. G2F) migrated faster than $F_C/2$ proteoforms with smaller glycans (e.g. G0F). Aglycosylated $F_C/2$ migrated last in comparison to all other $F_C/2$ proteoforms. $F_C/2$ proteoforms with larger glycans are characterized by larger Stokes' radii (R) in comparison to $F_C/2$ proteoforms with smaller glycans, and especially the aglycosylated $F_C/2$ form. As R is inversely proportional to μ_{FS} (see Eq**.** 3), migration times are shorter for net positively-charged analytes with larger hydrodynamic volumes. As the majority of identified $F_C/2$ proteoforms contained neutral, biantennary glycans, their migration order is expected to be governed by variations in the hydrodynamic volumes.

As shown in the inset in the total ion electropherogram in **Figure 3-3A(i)** and in the ion density map in **Figure 3-3B(i),** baseline CZE resolution was achieved between 2X-deamidated, 1X-deamidated, and non-deamidated glycoforms of $F_C/2$. Deamidation in proteins results in an increase in the molecular weight by 0.96 Da and a reduction of charge by one, as asparagine is converted to a mixture of aspartic acid and iso-aspartic acid [283]. As a 0.96 Da molecular weight increase does not yield a detectable increase in hydrodynamic volume, separations of the deamidated forms were likely to be effected by differences in the net charge. In CZE, μ_{FS} is proportional to the analyte charge (q) (see Eq. 3). Since the net positively-charged analytes possess free solution mobilities in the direction of the cathode maintained at -20 kV, species with the lower positive charge, migrate faster. The 2X-deamidated forms of glycosylated $F_C/2$ possess a "-1" charge difference in comparison to the 1X-deamidated variants, which, in turn, have a "-1" charge difference in comparison to the unmodified glycoforms of $F_C/2$. The experimentally observed migration order for the glycoforms of $F_C/2$ was found to be in accord with their net charges, revealing shorter migration times of the least positively-charged 2X-deamidated species, followed by the 1X-deamidated and then by the unmodified glycoforms.

Figure 3-3D shows that several oxidized $F_C/2$ glycoforms as well as an oxidized and a glycated form of the aglycosylated $F_C/2$ were identified by deconvolution. In the case of oxidation, no charge variation is expected. For glycation under acidic conditions, no charge variation is expected, as the Amadori product maintains the same charge as the unglycated variant. Consistently, CZE separation of the oxidized forms of glycosylated $F_C/2$ species from the unmodified variants could not be obtained (**Figure 3-3B(i)**). Similarly, separation of oxidized aglycosylated $F_C/2$ species was not achieved from aglycosylated $F_C/2$. In addition, the glycated aglycosylated $F_C/2$ proteoform was unresolved from the aglycosylated form. In all of these cases, CZE separations were impeded by the lack of charge variations under the acidic conditions. Differences in hydrodynamic volume, represented by a +16 Da change for oxidized forms and +162 Da change for glycated forms, were found to be insufficient to enable CZE separation.

A CZE-MS^2 "Top N" method at 23 V higher collision-induced dissociation energy (HCD) was then used to confirm and determine localization of PTMs attached to specific amino

acid residues. For example, for the G0F $F_C/2$ fragment, deamidations at residues N_{79} (**Figure 3-4A**) and N_{50} (**Figure 3-4B**) were detected. 1X-deamidation assigned previously to the G1F $F_C/2$ protein by accurate mass was also confirmed by CZE-MS^2. CZE-MS^2 enabled confirmation of most biantennary glycans ascribed to the $F_C/2$ fragment from previous CZE-MS experiments. CZE-MS^2 also resulted in the identification of a glycated form (at residue K_{90}) of the G0F $F_C/2$ fragment (**Figure 3-5**). Analysis of CZE-MS data resulted in the identification of a G0 $F_C/2$ form, in addition to a G1 $F_C/2$ proteoform. However, identification of a glycated form of G0 $F_C/2$ by CZE-MS^2 suggests that the peak corresponding to the G1 $F_C/2$ proteoform could in part be a combination of G1 $F_C/2$ and glycated G0 $F_C/2$, represented by the same delta mass shift. This is shown in **Figure 3-3D**.



Figures 3-4 A-B. Confirmation of Glycoform Deamidations by CZE-MS^2. (**A**) CZE-MS^2 HCD mass spectrum of the deamidated G0F $F_C/2$ fragment. Glycosylation and deamidation were confidently identified at residues N_{60} and N_{79}, respectively. (**B**) Confirmation of deamidation at residue N_{50} of the G0F $F_C/2$ proteoform by CZE-MS^2. Fragmentation was performed in the HCD cell at a collision energy of 23 V. The MS^2 spectrum shows the identified *a*-, *b*-, and *y*-fragment ions. Annotations were generated using Byonic software at a mass measurement accuracy of 20

ppm. The inset in (**B**) shows a portion of the mass spectrum representing fragment ions used to identify deamidation. The top panels in (**A**) and (**B**) show sequence coverage obtained by CZE-MS^2.

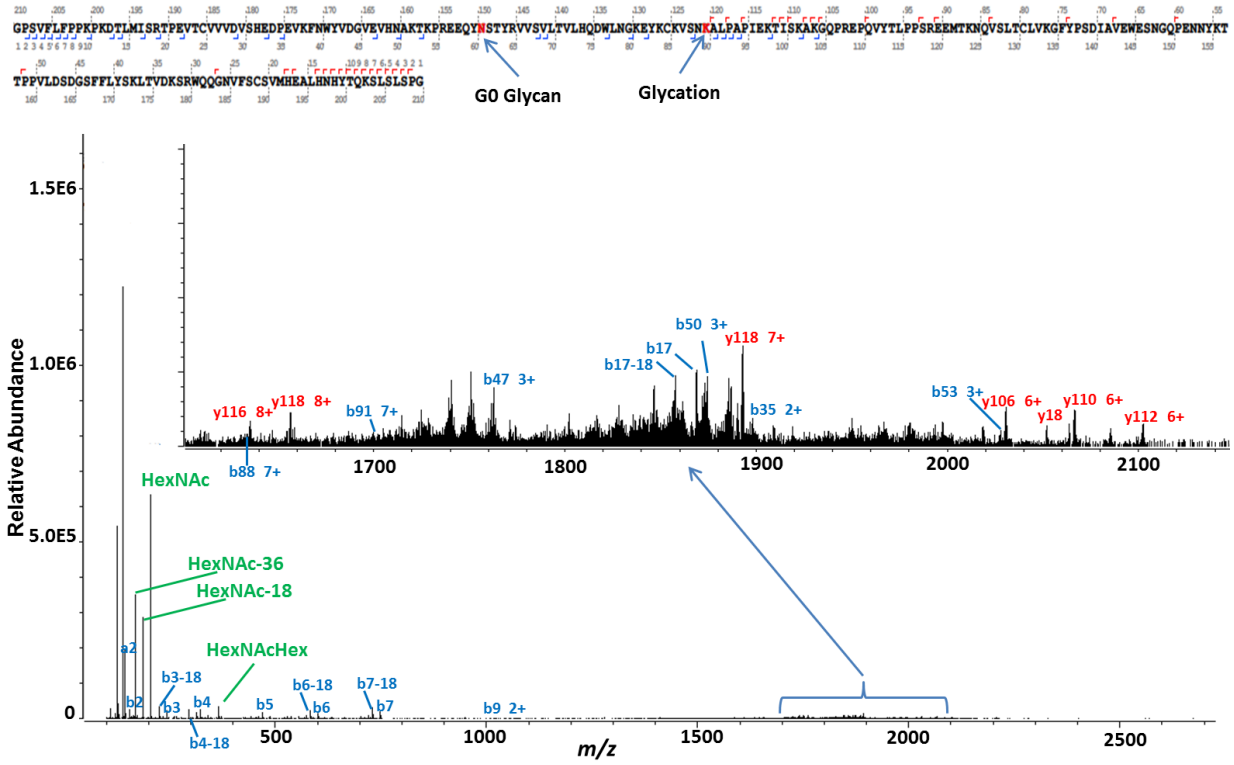


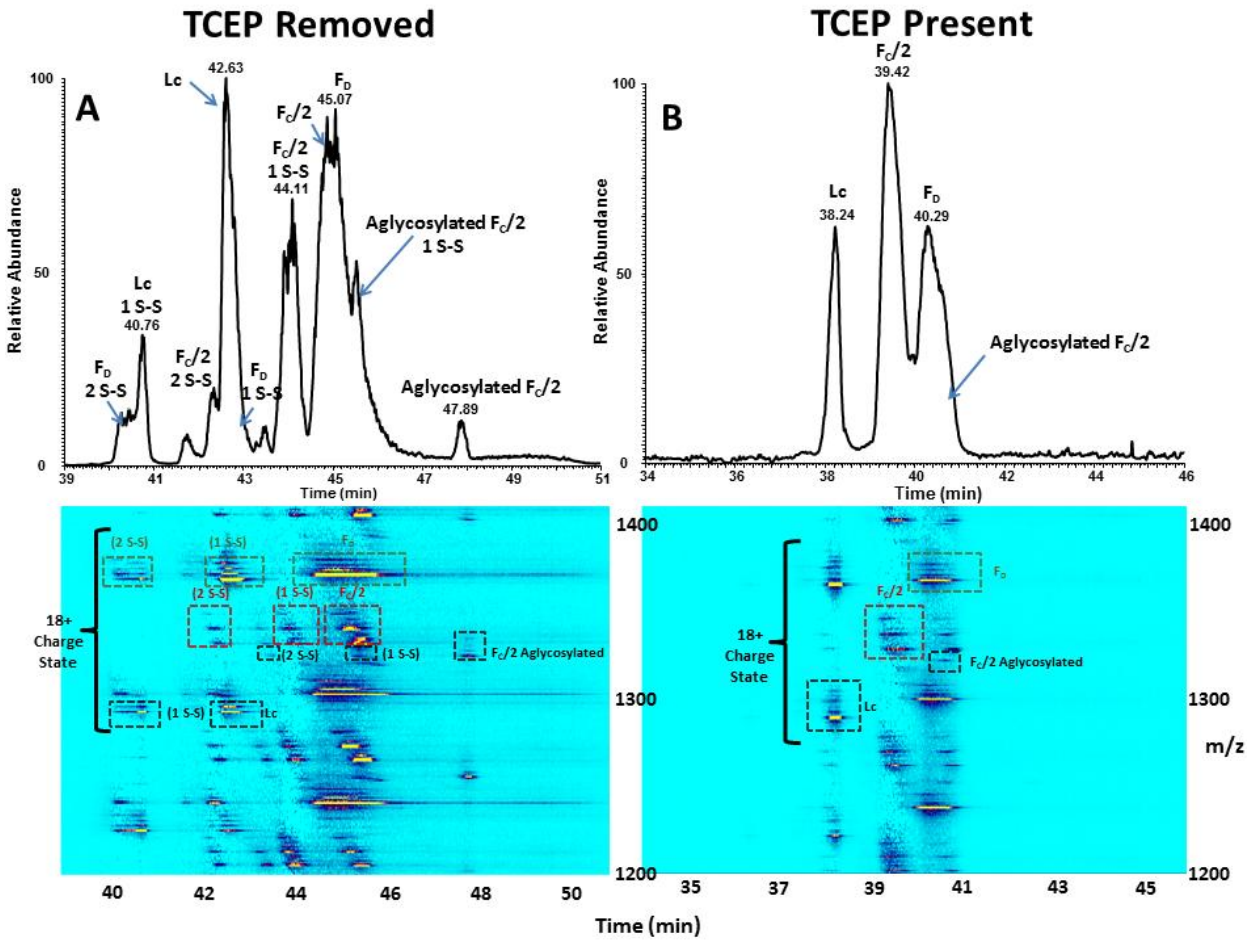
Figure 3-5. Confirmation of glycation at residue K_{90} of the G0F F_C/2 proteoform by CZE-MS^2. The top panel shows the sequence coverage on the G0F F_C/2 fragment obtained using HCD fragmentation at a collision energy of 23 V x charge. The bottom panel shows the MS^2 spectrum with a number of *a*-, *b*-, *y*-ions, and sugars annotated with Byonic software at a mass measurement accuracy of 20 ppm. The inset shows a portion of the mass spectrum representing fragment ions used to identify glycation at residue K_{90}.

Turning to the 98.4 kDa $F(ab')_2$, two glycations were found after deconvolution of the electropherogram. Unmodified $F(ab')_2$ was detected at 40%, while 1X-glycated and 2X-glycated $F(ab')_2$ were present at roughly 7% and 1%, respectively (**Figure 3-3E**). However, the relatively large size of the $F(ab')_2$ precluded identification of other PTMs by MS, as small differences in *m/z* were unresolved by the mass spectrometer. Moreover, such forms exceeded the 2,500 *Th* limit of precursor selection by the analytical quadrupole of the MS. Characterization of proteoforms of $F(ab')_2$ required a further decrease of the 98.4 kDa $F(ab')_2$ protein into smaller fragments by means of reduction.

3.3.5. Middle-Down Analysis: Analysis of *IdeS*-Digested, Reduced MAb

Reduction of *IdeS*-digestion products, $F_C/2$ and $F(ab')_2$, decreased the molecular weights of species, allowing for improved characterization of the mAb. The formed products included Lc (23.2 kDa) and F_D (26.0 kDa), along with the reduced $F_C/2$ fragment (~24.0 kDa). However, reduction of *IdeS*-digestion products with TCEP-HCl resulted in the introduction of significant artificial heterogeneity if the reducing agent was removed from the sample buffer prior to injection into the capillary. **Figure 3-6** shows the total ion electropherograms and corresponding ion density maps of the mAb *IdeS*-digestion products with and without the reducing agent, TCEP-HCl. Exclusion of TCEP from the sample buffer prior to injection resulted in the formation of multiple isoforms of scrambled disulfides for each digestion fragment (**Figure 3-6A**). Conversely, inclusion of TCEP eliminated scrambled disulfide isoforms of *IdeS*-digestion products and resulted in a simplification of the electropherogram (**Figure 3-6B**). The use of TCEP also simplified sample preparation, as an alkylation step was not required. For the middle-down schemes described below, limited proteolytic digestions were followed by reduction with TCEP. The reduced samples were then buffer-exchanged into water containing up to 5 mM TCEP.

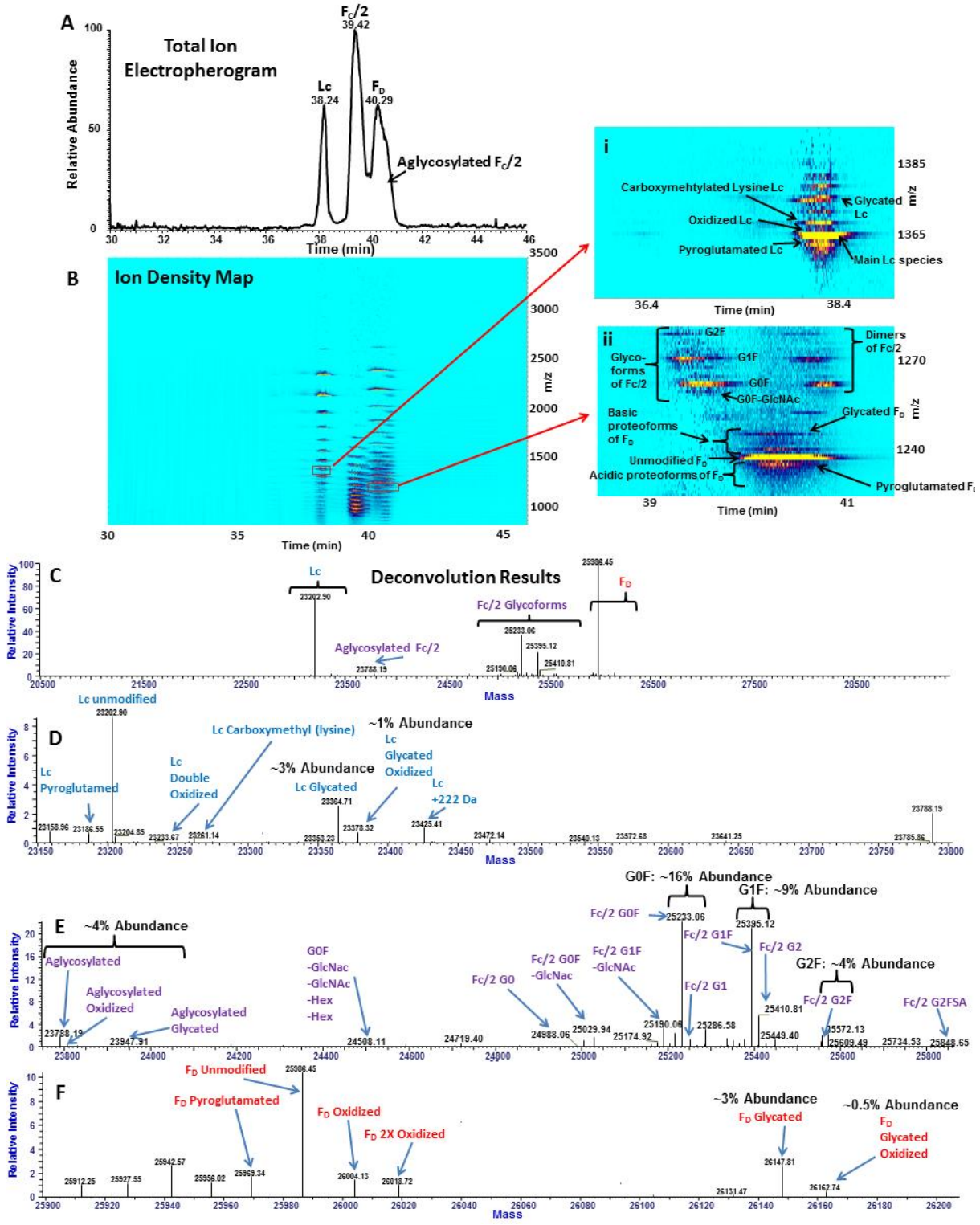
While the analysis of the *IdeS*-digested mAb facilitated the characterization of proteoforms of the $F_C/2$ fragment, reduction following digestion enabled a more comprehensive analysis of the $F(ab')_2$ fragment. The latter enabled characterization of proteoforms of the formed Lc and F_D fragments. **Figures 3-7 A-B** show the separation of the *IdeS*-digested, reduced mAb by CZE-MS in a 50% methanol/1% formic acid BGE. Shown in **Figures 3-7 C-F**, the deconvolution result for the entire electropherogram resulted in the identification of numerous proteoforms of Lc, F_D, and $F_C/2$ digestion fragments.



Figures 3-6 A-B. Inclusion of TCEP in sample buffer during CZE mitigates the formation of scrambled disulfide isoforms for Lc, $F_C/2$, and F_D fragments. (**A**) The total ion electropherogram obtained after removal of the reducing agent, TCEP-HCl, prior to sample injection is shown in the top panel and the corresponding ion density map is shown in the bottom panel. Scrambled disulfide species were detected for all the studied mAb fragments. (**B**) The total ion electropherogram obtained with the reducing agent (5 mM) in the sample buffer is shown in the top panel and the corresponding ion density map is shown in the bottom panel. No scrambled disulfide forms were observed, resulting in a simplified electropherogram without the need for alkylation. For both ion density maps, only species from the 18+ charge state are annotated.

While analysis of the *IdeS*-digested mAb facilitated the characterization of proteoforms of the $F_C/2$ fragment, reduction following digestion enabled a more comprehensive analysis of the $F(ab')_2$ fragment. The latter enabled characterization of proteoforms of the formed Lc and F_D fragments. **Figures 3-7A-B** show the separation of the *IdeS*-digested, reduced mAb by CZE-MS in a 50% methanol/1% formic acid BGE. Shown in **Figures 3-7C-F**, the deconvolution result for

the entire electropherogram resulted in the identification of numerous proteoforms of Lc, F_D, and $F_C/2$ digestion fragments.



Figures 3-7 A-F. CZE-MS analysis of the *IdeS*-digested, reduced mAb. (A) Total ion electropherogram of digestion fragments. (**B**) The corresponding ion density map with zoomed-in regions for Lc (**B(i)**) and F_D (**B(ii)**), shown as separate insets. (**C**) Deconvoluted spectrum, with zoomed-in views of regions corresponding to Lc (**D**), $F_C/2$ (**E**), and F_D (**F**).

For the Lc fragment shown in **Figure 3-7D**, proteoforms possessing pyroglutamation, oxidation, carboxymethylation of lysine, and glycation were observed. Depicted in **Figure 3-7B(i)**, the ion density map of the Lc 17+ charge state shows the resolution of the individual species in both, CZE and MS domains. In the case of glycation, as described previously, no charge variation between glycated and unmodified forms of Lc was observed. Similar to the aglycosylated $F_C/2$ species, glycation of the Lc fragment did not result in a sufficient hydrodynamic volume change to enable separation from unmodified Lc. As a result, glycated Lc co-migrated with unmodified Lc. Carboxymethylation of lysine is an advanced glycation end-product characterized by a +58 Da delta mass shift [284] with no accompanying change in charge under the acidic conditions, similar to glycation. Shown in **Figure 3-7B(i)**, the carboxymethylated lysine form of Lc was unresolved from the unmodified variant. Similarly to glycation, the variation in hydrodynamic volume due to carboxymethylation of lysine was found to be insufficient to enable CZE separation. For the same reason, double-oxidized forms of the Lc, as well as its glycated and oxidized proteoforms, were unresolved from the unmodified variant.

Pyroglutamation is a relatively common N-terminal PTM found on light and heavy chains in mAbs. Formation of pyroglutamic acid occurs both spontaneously and/or may be catalyzed by glutaminyl cyclase – an enzyme found in many organisms. Formation of pyroglutamic acid occurs as a result of a rearrangement of N-terminal glutamine or glutamate residues [285]. In the case of pyroglutamation from the N-terminal glutamine, the resultant species lose a single charge and, in turn, become more acidic. However, in the case of pyroglutamic acid formation from glutamate, the same charge is maintained under acidic conditions [285]. In the case of mAb1, a fraction of the N-terminal glutamate on the Lc fragment was found to be pyroglutamated, as seen in **Figure 3-7D**. As no charge variation occurred during the conversion from N-terminal glutamate to pyroglutamic acid, the pyroglutamated Lc could not be separated from the unmodified Lc (**Figure 3-7B(i)**).

Similar to the *IdeS*-digested, non-reduced middle-down analysis of mAb1 (see **Figure 3-3D**), a number of glycosylated $F_C/2$ proteoforms, aglycosylated $F_C/2$, oxidized aglycosylated $F_C/2$, and glycated aglycosylated $F_C/2$ species were observed for the *IdeS*-digested, reduced mAb1. **Figure 3-7E** shows all the identified proteoforms of the reduced $F_C/2$ fragment. $F_C/2$ glycoforms were identified in approximate abundances of 4% for aglycosylated $F_C/2$ species, 16% for G0F $F_C/2$, 9% for G1F $F_C/2$, and 4% for G2F $F_C/2$. Partial separation of numerous glycoforms of the $F_C/2$ fragment is demonstrated in **Figure 3-7B(ii)**. Similar to the *IdeS*-digested mAb1 experiment, $F_C/2$ proteoforms characterized by larger glycans (e.g. G2F) migrated faster due to the hydrodynamic volume effect. Separations of 2X-deamidated from 1X-deamidated,

from non-deamidated $F_C/2$ glycoforms were achieved as a result of differences in a single charge imparted by each deamidation. The least positively-charged, 2X-deamidated analytes, migrated first, followed by the 1X-deamidated and then by the unmodified $F_C/2$ glycoforms.

Analysis of the *IdeS*-digested, reduced mAb resulted in the identification of numerous proteoforms of the F_D fragment. Specifically, oxidized, double-oxidized, glycated, and combined glycated and oxidized forms were detected by deconvolution. **Figure 3-7F** shows the deconvolution result for the F_D region. For the pyroglutamated F_D form, a single charge is imparted as a result of the conversion of glutamine to pyroglutamic acid, along with a -17 Da delta mass shift [285]. Nevertheless, as shown in **Figure 3-7B(iv)**, the pyroglutamated F_D co-migrated with the unmodified F_D. Conceivably, the formation of cyclical pyroglutamic acid from N-terminal glutamine resulted in a decrease of the hydrodynamic volume for the F_D fragment that offset the change imparted by the loss of a single charge, resulting in co-migration of the two forms.

From deconvolution results of the *IdeS*-digested mAb1 analysis, approximately 7% of the $F(ab')_2$ was found to be glycated, along with 1% relative abundance for the 2X-glycosylated species (see **Figures 3-3 D-E**). These results were corroborated by the *IdeS*-digested, reduced mAb1 experiment, which revealed approximately half of the glycation attributed to the F_D (**Figure 3-7D**) and half to the Lc fragments (**Figure 3-7F**).

CZE-MS^2 analysis of the *IdeS*-digested, reduced mAb was performed in order to validate identifications of PTMs of major digestion fragments and to establish localization of the modified residue sites. A key limitation of the *IdeS*-digested mAb1 experiment was the inability to analyze the large $F(ab')_2$ protein by CZE-MS^2. Reduction of *IdeS*-digestion products of mAb1 facilitated the analysis of F_D and Lc fragments, as charge-state distributions of precursor ions were below the 2,500 *Th* threshold for selection for CZE-MS^2 analysis. From CZE-MS^2 analysis, with high confidence, two different oxidations were identified on the Lc fragment at R_{24} (**Figure 3-8**) and R_{108} (**Figure 3-9**) residues. The assignments of the two oxidations for the Lc fragment were corroborated by the CZE-MS deconvolution results illustrated in **Figure 3-7D**. In addition, N-terminal pyroglutamation for the F_D fragment was also confirmed by the CZE-MS^2 analysis (**Figure 3-10**).

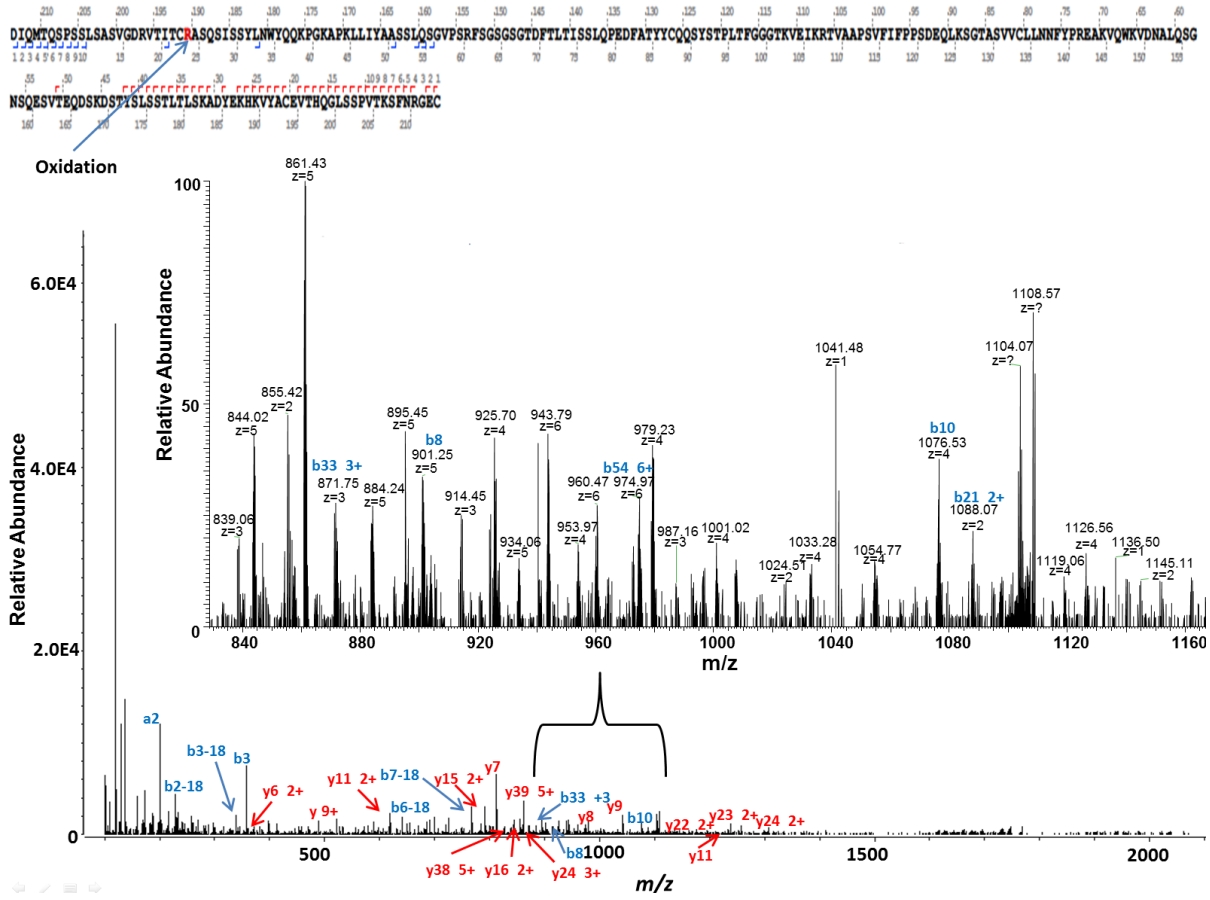


Figure 3-8. Identification of the oxidation site at residue R_{24} on the Lc fragment using CZE-MS^2 at a collision energy of 23 V x charge. The top panel shows the sequence coverage of the Lc fragment. The ESI-MS^2 spectrum shows the identified fragment ions, including those used to localize the PTM site. Identification was performed at a mass measurement accuracy of 20 ppm. The inset represents a portion of the mass spectrum, which contains the *b*-fragment ion series used to identify oxidation at residue R_{24}.

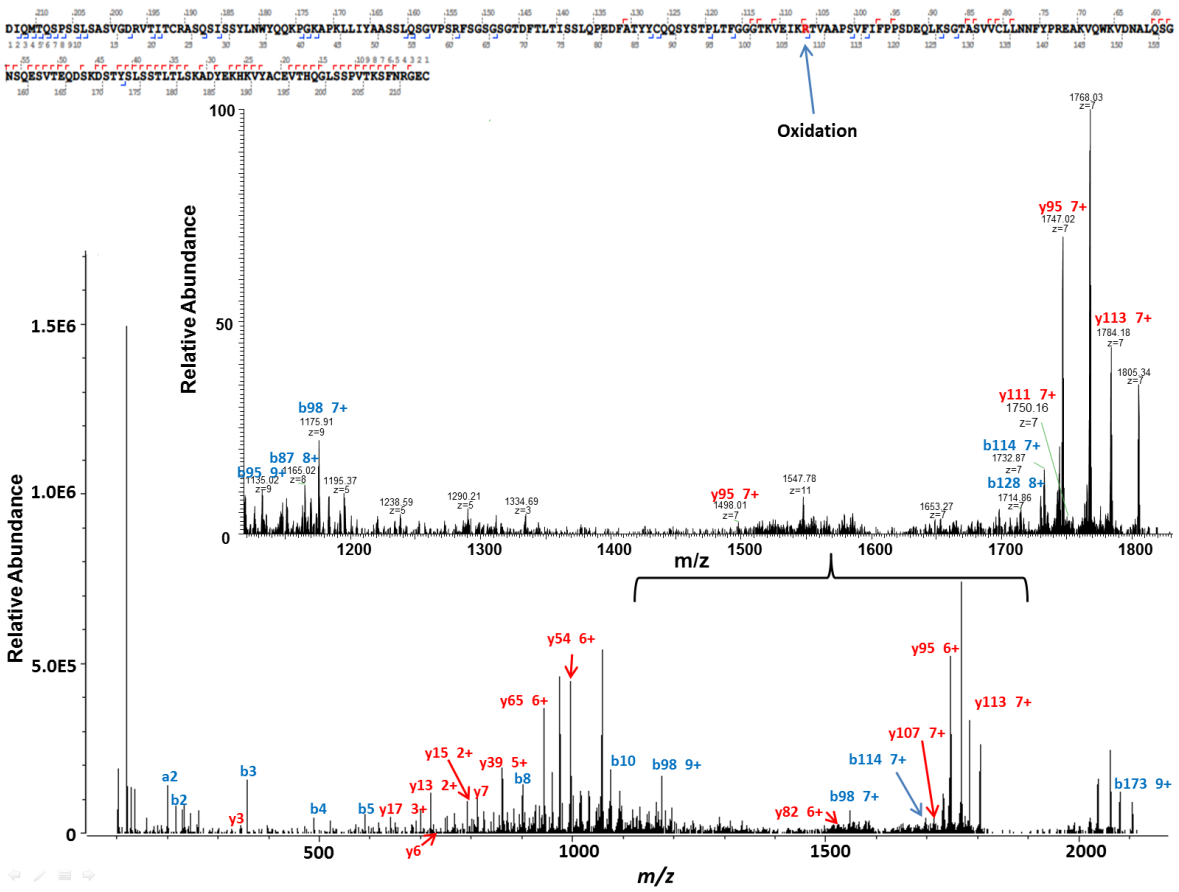


Figure 3-9. Identification of the oxidation site at residue R_{108} on the Lc fragment using CZE-MS^2 at collision energy of 23 V x charge. The top panel shows the sequence coverage of the Lc fragment. The ESI-MS^2 spectrum shows the identified fragment ions, including those used to localize the PTM site. Identification was performed at a mass measurement accuracy of 20 ppm. The inset shows a portion of the mass spectrum, which contains the *b*-fragment ion series used to identify oxidation at residue R_{108}.

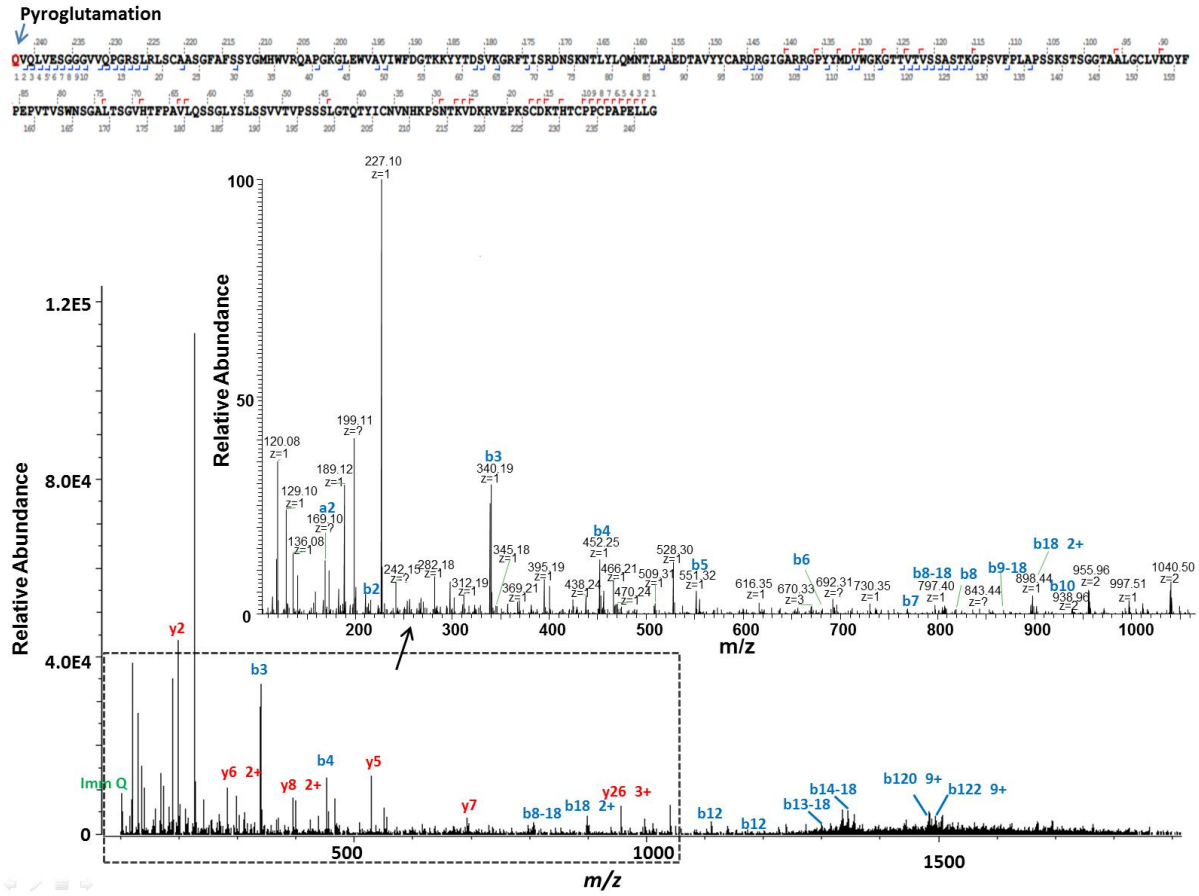
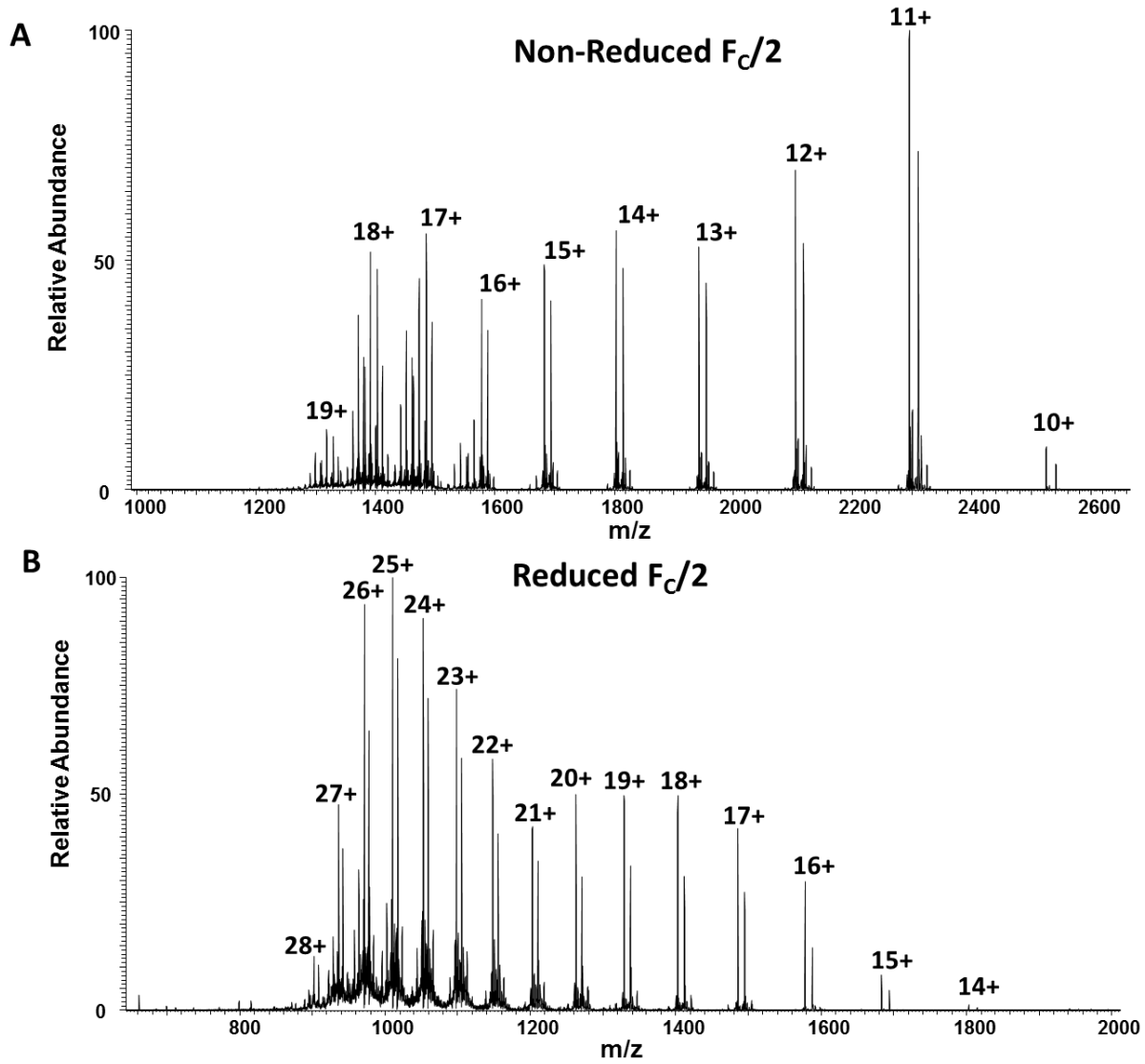


Figure 3-10. Confirmation of pyroglutamation at residue Q_1 on the F_D fragment by CZE-MS^2 at a collision energy of 23 eV x charge. The top panel shows the sequence coverage of the F_D fragment. The ESI-MS^2 spectrum shows the identified fragment ions, including those used to localize the PTM site. Identification was performed at a mass measurement accuracy of 20 ppm. The inset shows a portion of the mass spectrum, which contains the *b*-fragment ion series used to identify oxidation at Q_1.

A comparison of CZE separations of *IdeS*-digested mAb and *IdeS*-digested, reduced mAb revealed a difference in migration times of both the glycosylated and aglycosylated $F_C/2$ fragments. For example, in the case of the glycosylated $F_C/2$ proteoforms from the *IdeS*-digested sample, migration time was approximately 46.2 min (**Figures 3-3 A-B**). However, for the *IdeS*-digested, reduced sample, the migration time of the identical forms was approximately 39.5 min (**Figures 3-7 A-B**). The preceding may be explained by different degrees of unfolding as a result of disulfide bond cleavage. Two phenomena are observed simultaneously. On one hand, reduction results in protein unfolding and a greater exposure of internal residues to the outside environment. This is evident in part by different charge state distributions on ESI mass spectra detected for identical forms between *IdeS*-digestions of the mAb with and without reduction. **Figure 3-11** shows a comparison of ESI mass spectra of *IdeS*-non-reduced glycosylated $F_C/2$ and

IdeS-reduced glycosylated $F_C/2$ fragments. In the case of the non-reduced glycosylated $F_C/2$ species, the detected charge state distribution was constrained between $[M+11H]^{11}+$ and $[M+19H]^{19}+$ (highest abundance: $[M+11H]^{11+}$) species (**Figure 3-11A**). For the reduced glycosylated $F_C/2$ species, the distribution shifted to $[M+14H]^{14}+$ and $[M+28H]^{28+}$ charge states (highest abundance: $[M+25H]^{25+}$) (**Figure 3-11B**). While these charge state distributions may not be correlated directly to analyte charge in solution, a subtle increase in charge is nevertheless expected as a result of unfolding. On the other hand, protein unfolding from reduction is expected to result in a significant increase in the hydrodynamic volume. In CZE with a positively-charged capillary coating, species with larger hydrodynamic volumes experience faster migration times (see Eqs. 2-3). This trend is confirmed by our experiments with the reduced glycosylated $F_C/2$ proteoforms. Due to partial unfolding, the latter are characterized by increased hydrodynamic volumes and were found to migrate faster than the non-reduced variants. A similar trend was also observed for the aglycosylated $F_C/2$ species. These effects highlight the sensitivity of CZE to differences in analyte hydrodynamic volume.



Figures 3-11 A-B. Comparison of $F_C/2$ charge states in *IdeS*-digested, non-reduced and *IdeS*-digested, reduced mAb experiments. (A) Observed $F_C/2$ charge states in *IdeS*-digested, non-reduced mAb experiment. **(B)** Observed $F_C/2$ charge states in *IdeS*-digested, reduced mAb experiment.

3.3.6. Intact CZE-MS Characterization of the MAb

Following middle-down analysis, the mAb was analyzed by intact sheathless CZE-MS. With the middle-down approaches described above, digestion and reduction resulted in formation of $F_C/2$ fragments. Characterization of $F_C/2$ fragments revealed individual glycosylation states. However, an intact mAb may be composed of asymmetric glycans across its two $F_C/2$ domains. This information is not accessible by most middle-down analysis approaches involving reduction. Consequently, intact mAb analysis is an important complement to middle-down experiments allowing for the characterization of glycosylated structures of the mAb.

Minimal sample processing required for intact CZE-MS analysis is also a potential advantage for the analysis of biopharmaceutical protein composition in formulation. Furthermore, minimal sample processing allows for structural characterization of the mAb molecule as a whole and not its parts, as in the case of middle-down.

Initially, the BGE for middle-down experiments (50% methanol, 1% formic acid) was employed in the analysis of the intact mAb. For transient isotachophoresis to be performed, the intact mAb was diluted in a leading electrolyte consisting of 50 mM ammonium acetate at pH 4.0. However, no separation of different proteoforms of the intact mAb was observed (**Figure 3-12**). A reduction of the organic content, first down to 10% methanol, in addition to reduction of the formic acid concentration to 0.2%, resulted in improved separation of intact mAb proteforms, as well as in a decrease in the analysis time. Further improvement in separation was obtained by replacing methanol by isopropanol in the BGE mixture (results not shown). Increased viscosity of isopropanol in comparison to that of methanol resulted in slight decreases in μ_{APP}'s of major proteoforms of the intact mAb at the expense of a small increase in the analysis time. The 10% isopropanol, 0.2% formic acid BGE was thus selected for separation of intact proteins.

Figure 3-13 shows the results of CZE-MS analysis of the intact mAb under the optimized BGE conditions. Our experiments revealed the complete glycosylation profile of the mAb, in addition to elucidation of a trace quantity of Lc and heavy chain (Hc) bound non-covalently to the remainder of the molecule. **Figure 3-13A** shows the extracted ion electropherograms of major glycosylated proteoforms of the mAb. The eletropherograms are accompanied by an ion density map, shown in **Figure 3-13B**. Insets in **Figure 3-13B** highlight the individual regions of interest. **Figure 3-13B(i)** shows the 38+ charge state of the intact molecule, revealing all three detected populations of the mAb in both CZE and MS domains. **Figures 3-13B(ii)-(iii)** show species stemming from dissociation of light chains of the mAb. Species identified in **Figure 3-13B(ii)** correspond to a dissociated Lc with an uncharacterized +131 Da delta mass shift relative to unmodified Lc. Species shown in **Figure 3-13B(iii)** were presumed to be dimers of Lc. These structures, corresponding to 47,660 Da and 48,025 Da, are smaller than the theoretical molecular weight of free aglycosyalted heavy chain of the mAb (49,781 Da). Moreover, in the context of an intact mAb, a structure devoid of two light chains is more likely than a structure with a single dissociated heavy chain. However, additional confirmation is needed to validate this hypothesis. The deconvoluted electropherogram in **Figures 3-13 C-D** reveals the presence of three types of populations of mAb glycans. Namely, (1) 2X-glycosylated mAb, with glycans present on both $F_C/2$ domains; (2) 1X-glycosylated mAb, with glycans present on one of the two $F_C/2$ domains; and (3) aglycosylated mAb with no bound glycans. **Figure 3-14** shows reproducibility of quantitation results for 2X-glycosylated, 1X-glycosylated, and aglycosylated populations of the mAb across three experimental replicates. These results, combined with obtained ratios shown in **Figure 3-13C**, resulted in approximate ratios of the 2X-glycosylated mAb to the 1X-glycosylated, to the aglycosylated of 81.5%±2% to 6.5%±1.5% to 12%±0.5%, respectively. In addition to the unmodified form detected at 8%±1% relative abundance, up to two glycations

were detected for the aglycosylated mAb, with relative abundances of approximately 2%±1% for the 1X-glycated variant and 1% for the 2X-glycated proteoform (**Figures 3-13 C-D**).

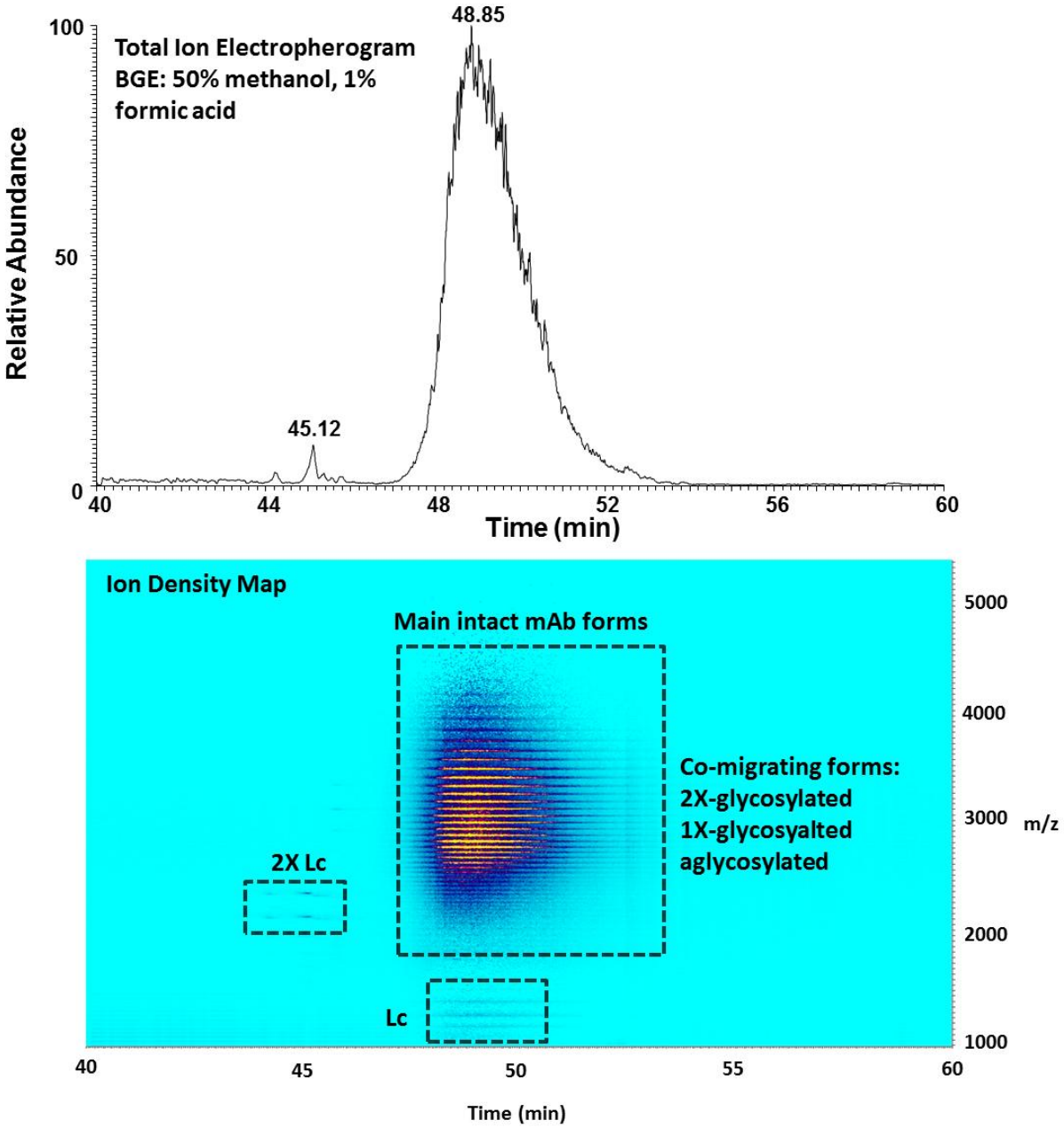
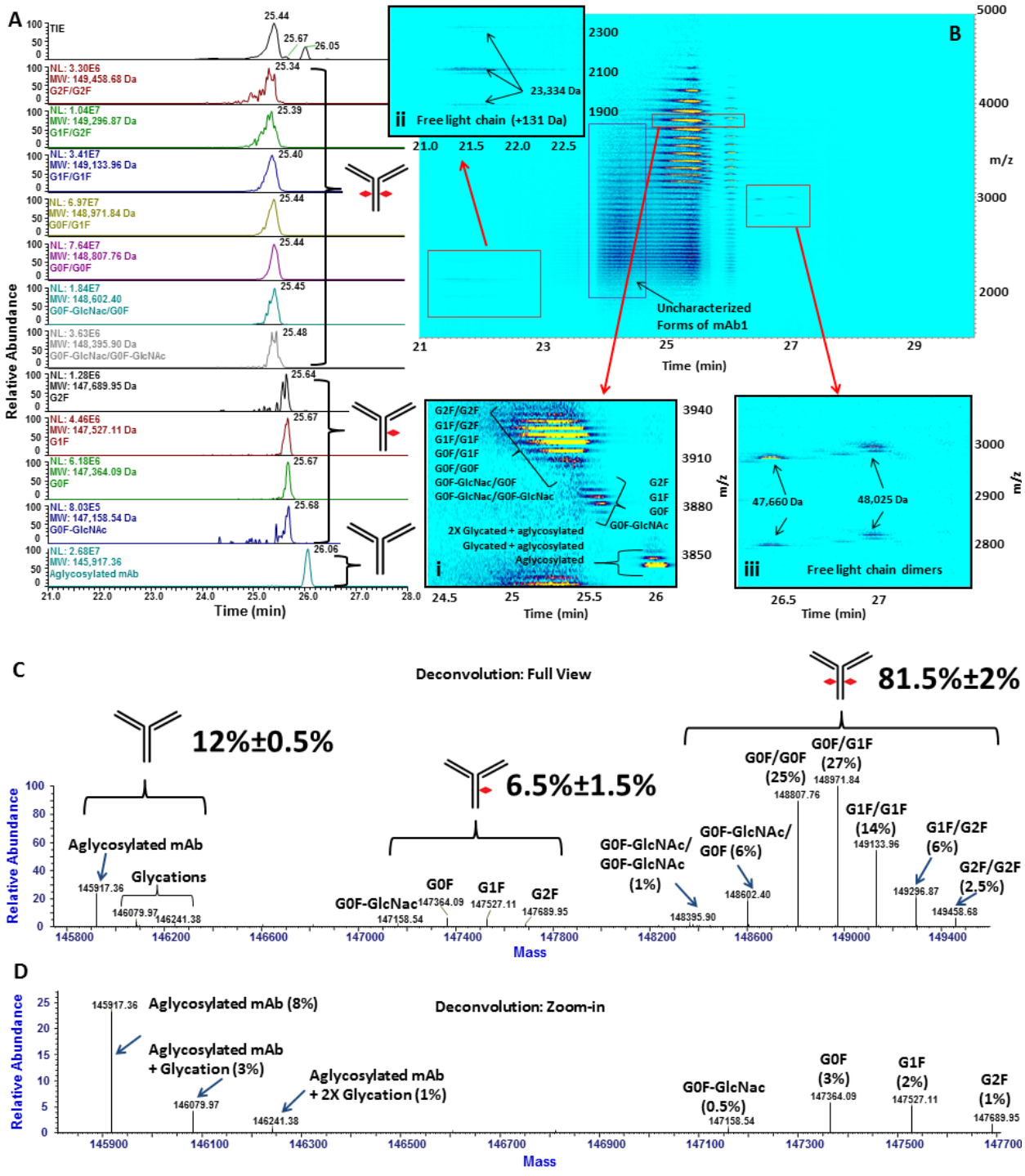
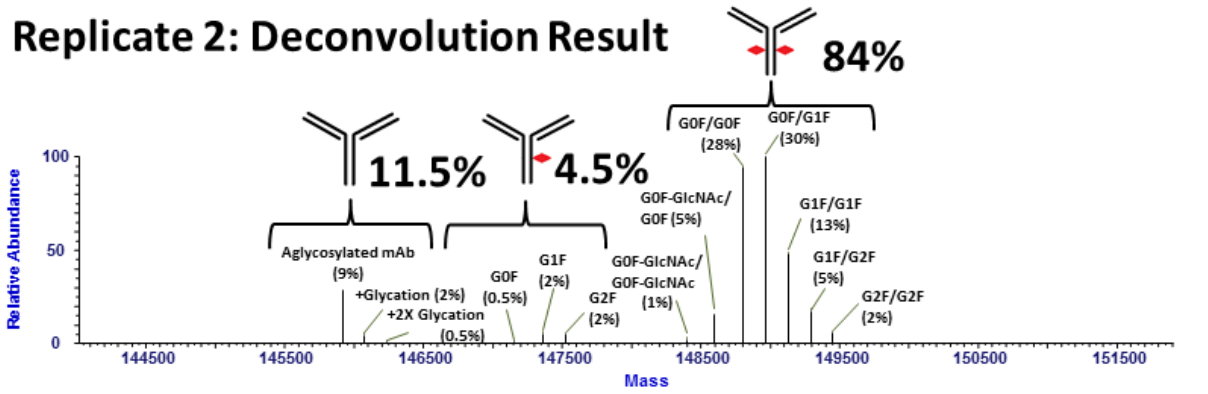


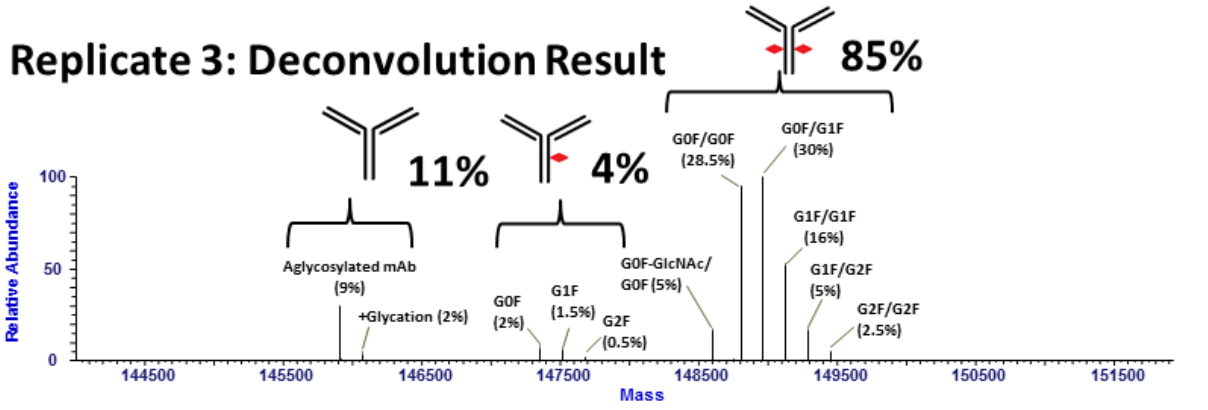
Figure 3-12. CZE-MS Analysis of the mAb in 50% methanol, 1% formic acid BGE. The top panel shows the total ion electropherogram, while the bottom panel is the ion density map.



Figures 3-13 A-D. Intact CZE-MS analysis of the mAb. (**A**) The total ion electropherogram (TIE), along with extracted ion electropherograms (EIEs) of major glycoforms of the mAb. (**B**) The ion density map for the CZE-MS experiment shown in (**A**). The inset (**B(i)**) shows the ion density map of the 38+ charge state of the mAb. CZE-MS signal intensities of glycoforms, such as 2X-glycosylated, 1X-glycosylated, and aglycosylated are illustrated. Insets **B(ii)**) and **B(iii)**)

show the detected heavy and light chains, respectively. (**C**) Deconvolution results showing major glycosylation states of the intact mAb for the electropherogram region between 24.7 minutes and 26.0 minutes. Quantitative ratios of each population of the mAb (2X-glycosylated, 1X-glycosylated, aglycosylated) are specified with relative standard deviations accounting for three additional experimental replicates (see **Figure S9**). (**D**) A zoomed-in view of the deconvolution spectrum to ~25% relative abundance shows the detected aglycosylated and 1X-glycosylated proteoforms of IgG1. In both (**C**) and (**D**), relative abundances of the identified proteoforms are annotated as percentages.





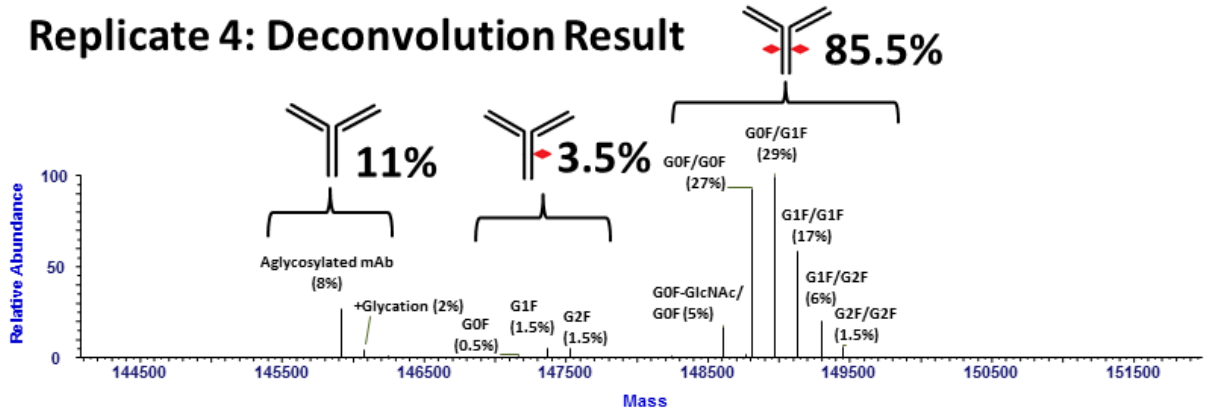
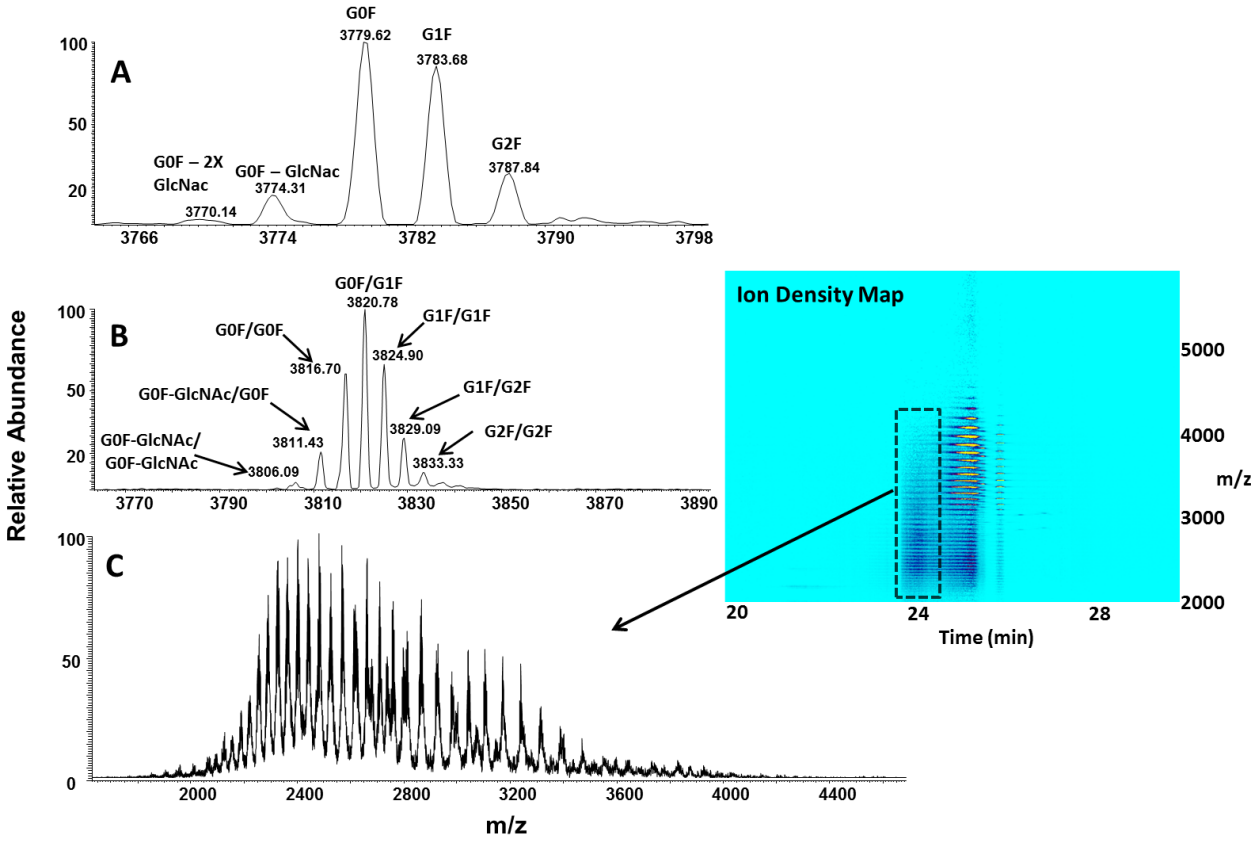


Figure 3-14. **Reproducibility of quantitation of 2X-glycosylated, 1X-glycosylated, and aglycosylated populations of the mAb.** An additional 3 experimental replicates are shown, revealing relative percentages of mAb glycosylation populations. Replicate 1 is considered as the experiment shown in **Figure 3-13**.

Baseline CZE separation was achieved for the mAbs with different extents of glycosylation. As the majority of detected glycan moieties are neutral structures, it is assumed that individual populations were resolved due to differences in the hydrodynamic volume. Relative migration proceeded in the order of the 2X-glycosylated, to the 1X-glycosylated, to the aglycosylated mAb, according to a decrease in the hydrodynamic volume. However, within individual populations of the intact mAb, species with glycoform variations could not be electrophoretically resolved under the utilized conditions, as demonstrated in **Figure 3-13B(i)**. A lack of resolution of individual glycoforms of the mAb may be attributed to an insufficient change in the hydrodynamic volume relative to the intact mAb. For example, the difference between hydrodynamic volumes of species such as G0F/G0F mAB and G0F/G1F mAb were found to be insufficient to be discerned by CZE. Conversely, in the middle-down experiments, individual glycoforms of the smaller $F_C/2$ fragment were partially resolved, presumably due to larger relative variations of the hydrodynamic volumes caused by different glycoforms.

Trace quantities of the dissociated Lc and Lc dimers were also detected, with migration times of 21.5 min and 26.5 min, respectively (**Figures 3-13 B(ii)-(iii)**). These species were detected at less than 1% relative abundances. Detection of the dissociated chains during intact analysis is likely due to the non-covalent associations between chains within the mAb. A trace quantity of the mAb is thereby composed of light chain domains bound *via* non-covalent interactions instead of disulfide bonds. The presence of such species has been reported previously, and such fragments are known to occur under native conditions [2, 286-288].

One of the difficulties in intact analysis is the inability to distinguish between small mass differences in higher *m/z* ranges. For the mAb investigated in this study, proteoforms corresponding to oxidation, deamidation, carboxymethylation of lysine, and pyroglutamation of different fragments of the *IdeS*-digested mAb – observed in middle-down CZE-MS – were undetected on the intact mAb by CZE-MS analysis. **Figure 3-15A** shows the ESI mass spectrum of the 39+ charge state for the 1X-glycosylated mAb. Only individual glycosylation states of the 1X-glycosylated mAb are observed, with limited MS resolution available for detection of smaller *m/z* shifts corresponding to the above modifications. The same trend is observed for the 2X-glycosylated mAb, shown in **Figure 3-15B**. Lastly, a distinct early-migrating CZE peak was detected between approximately 23.7 and 24.7 minutes. The integrated ESI mass spectrum for this entire region is shown in **Figure 3-15C**. The resultant mass spectrum contains a distribution of numerous species, presumably corresponding to acidic proteoforms and mAb structures without dissociated chains. However, a lack of resolution in the MS and CZE domains precluded accurate deconvolution and identification of these proteoforms.



Figures 3-15 A-C. Proteoform identifications limited to glycosylation of the intact mAb. (**A**) A zoom-in of the 39+ charge states of the CZE-MS spectra for the 1X-glycosylated mAB. (**B**) A zoom-in of the 39+ charge state of the 2X-glycosylated mAb. (**C**) The averaged mass spectrum is shown for the region between 23.7 min and 24.7 min. The inset shows the ion density map with the preceding region identified.

To verify abundances of various proteoforms of the mAb determined in intact analysis described above, we also characterized the intact mAb under native conditions using a sheathless polyacrylamide-coated capillary coupled to the Q Exactive Plus™ MS. The result of this experiment is shown in **Figure 3-16**. While the separation of individual glycosylation states (2X-glycosylated, 1X-glycosylated, and aglycosylated) was not achieved under native conditions (**Figures 3-16 A-B**), this approach resulted in the detection of mAb dimers at less than 1% relative abundance. These species co-migrated with the main forms of the mAb (**Figures 3-16 A-B**). Nevertheless, detection of aggregates is an important advantage of the analytical technique for characterization of mAbs. A previous study out of our lab has also reported on dimers detected in native CZE-MS analysis for a different mAb [2]. In addition to aggregates, species corresponding to dissociated light chain and light chain dimers, as well as intact mAb with a loss of two light chains were also observed. These forms were baseline-resolved from the main species of the mAb by CZE (**Figures 3-16 A-B**) and were present at less than 1% relative abundance. A plausible conclusion is that a trace quantity of light chain of the mAb is bound

non-covalently to the remainder of the molecule. Exchange from formulation buffer into the 40 mM ammonium acetate, pH 7.5 buffer likely resulted in dissociation of these chains.

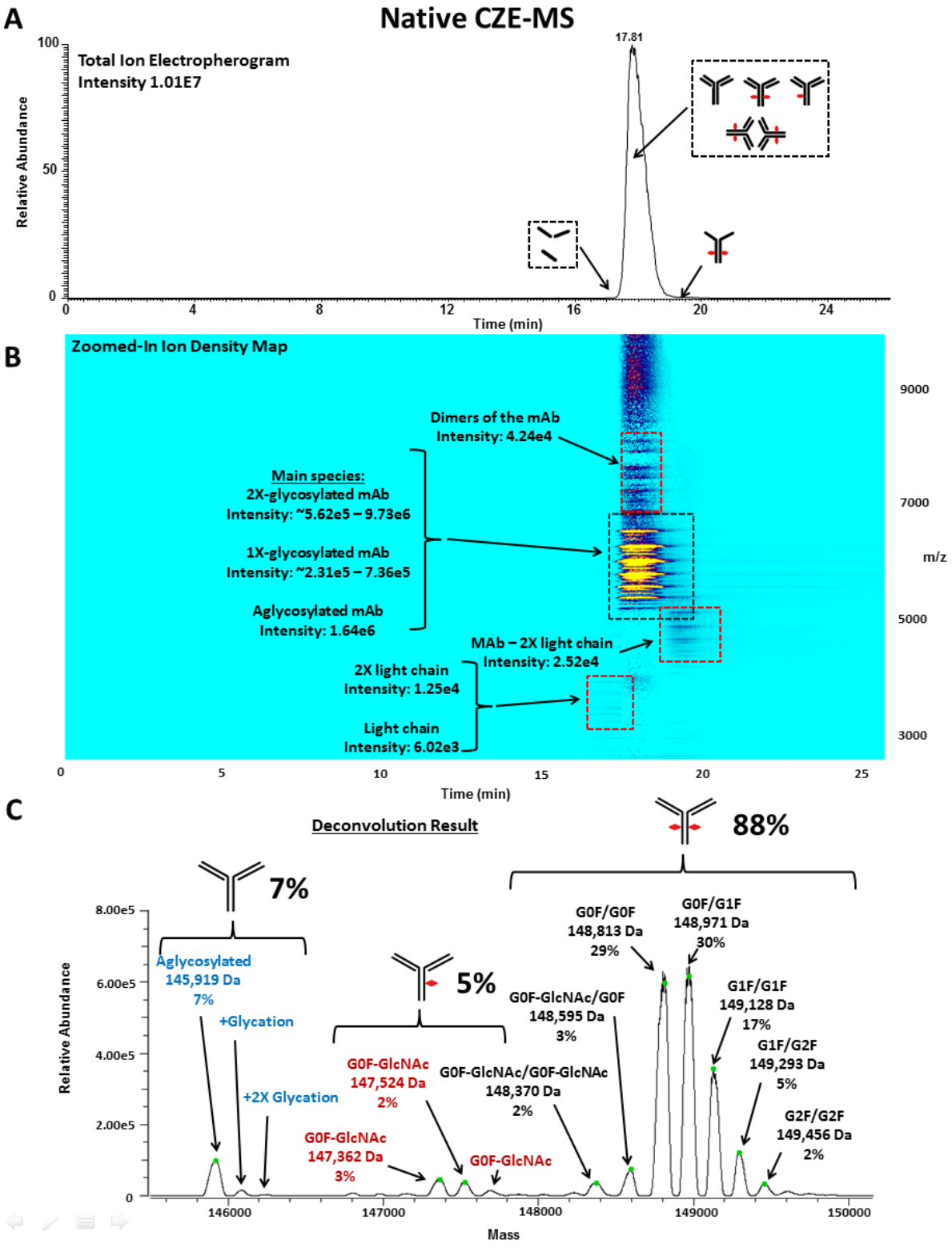


Figure 3-16 A-C. Characterization of major proteoforms of the mAb under native conditions on the sheathless polyacrylamide-coated capillary – QExactive Plus MS system. (**A**) The total ion electropherogram, with major forms of the mAb shown according to their migration times. (**B**) The ion density map of the experiment showing major species, as well as corresponding intensity values. (**C**) Deconvolution result of the main peak between 17 and 20 minutes. Approximate abundances of various proteoforms are indicated, as well as approximate ratios of aglycosylated, 1X-glycosylated, and 2X-glycosylated populations of the mAb. For those species that could not be identified by deconvolution, abundance values were not indicated.

Abundances of major forms of the mAb identified in denaturing conditions (see **Figure 3-13** and **Figure 3-14**) were verified by the abundance values resulted from deconvolution of the main electrophoretic peak observed in the native experiment. **Figure 3-16C** shows the deconvolution result for the main peak observed in **Figure 3-16A**. Relative abundances of individual proteoforms of the mAb were comparable between native and denaturing experiments and across two different deconvolution algorithms. For example, the aglycosylated mAb was present at approximately 7% relative abundance under native conditions and 8% under denaturing conditions. Overall, the ratio of 2X-glycosyalted to 1X-glycosylated to aglycosyalted mAb under native conditions was approximately 88% to 5% to 7%, respectively. These results are comparable to ratios of major glycosylation states of the mAb observed under denaturing conditions (see **Figure 3-13C** and **Figure 3-14**).

The CZE-MS technique in the analysis of mAbs represents a complementary approach to more conventional LC-based methods. For example, a number of mAb studies utilizing middle-down RPLC-MS have reported on the separation of mAb proteoforms differing in extents of oxidation [265, 289]. Such species exhibit differences in hydrophobicity and are thus amenable to separation by RPLC. However, other proteoforms characterized by differences in charge instead of hydrophobicity, such as deamidation, could not be resolved by RPLC [265, 289]. As an alternative to RPLC-MS for middle-down mAb characterization, online hydrophilic interaction liquid chromatography (HILIC)-MS was used to resolve $F_C/2$ glycosylation structures [273]. However, separation was limited to species exhibiting differences in hydrophilicities, and as such, only species differing in glycosylation were separated [273]. In our CZE-MS analysis, separations of deamidated proteoforms, partial separations of glycoforms of $F_C/2$, and separation of intact glycosylation states of the mAb were achieved. Thus, our approach is complementary to both RP and reported HILIC LC-MS for mAb characterization.

As an alternative to mAb analysis by online RP and HILIC LC-MS, cation exchange chromatography (CEX)-MS under native conditions (pH 7.0) resulted in separations of deamidated forms, species differing in C-terminal lysine truncation, and glycoforms differing in sialylation [290]. While a different mechanism of separation is present in acidic CZE, both approaches result in resolution of charge variants. However, CZE has an advantage of separating species with differences in hydrodynamic volumes, which makes it a complementary to CEX-MS. For example, isoforms corresponding to neutral biantennary glycans of $F_C/2$ were partially

resolved in this study. Such structures do not exhibit charge variation and would therefore not be separated by CEX approaches.

3.4. Conclusions

In this study, a mAb was analyzed by a middle-down sheathless CZE approach involving digestion of the protein by *IdeS* protease, with and without subsequent reduction. Individual proteoforms were identified by a combination of CZE with single-stage MS analysis and CZE-MS^2 with HCD fragmentation, enabling specific PTM site confirmation in numerous cases. Charge variants of the deamidated proteoforms were baseline separated by CZE. Species differing in hydrodynamic volume, represented by different glycoforms of the $F_C/2$ domain, were also resolved. These results were complemented by the intact mAb characterization, which revealed different populations of the mAb corresponding to 2X-glycosylated, 1X-glycosylated, and aglycosylated structures; these populations were resolved by CZE on the basis of hydrodynamic volume differences. Intact analysis revealed the presence of asymmetric glycans for the fully glycosylated protein, provided insights into the structural assembly of individual chains within the mAb, and allowed for quantitation of different intact proteoforms; these results were verified by CZE-MS performed under native conditions. Our CZE-MS-based approach to mAb characterization is complementary to the more conventional RPLC/HILIC-MS-based approaches, with an emphasis on the separation of species with charge and hydrodynamic volume differences instead of hydrophobicity/hydrophilicity. In addition, CZE-MS approach was found to be complementary to CEX-MS, which also allows for the separation of analytes with differences in charge.

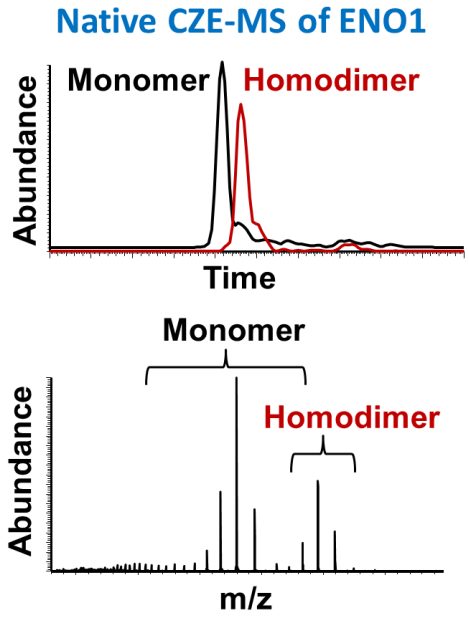
3.5. Acknowledgements

We acknowledge the Biogen Technical Development Department for their support through a Sponsored Research Agreement. This work was also supported by the NIH 1R01GM120272 award (ARI). We acknowledge Thermo Fisher Scientific for their support through a technology alliance. The authors also thank SCIEX for providing CESI capillaries used in this study, the CESI 8000 instrument for CZE, and for insightful discussions.

Chapter 4: Analysis of Proteins, Protein Complexes, and Organellar Proteomes Using Sheathless Capillary Zone Electrophoresis – Native Mass Spectrometry

Arseniy M. Belov[1], Rosa Viner[2], Marcia R. Santos[3], David M. Horn[2], Marshall Bern[4], Barry L. Karger[1] and Alexander R. Ivanov[1*]

1. Northeastern University, Barnett Institute of Chemical and Biological Analysis, Boston, MA 02115, USA
2. Thermo Fisher Scientific, San Jose, CA 95134, USA
3. SCIEX, Brea, CA 92821, USA
4. Protein Metrics, San Carlos, CA 94070, USA



***My specific contributions to the work presented in this chapter included active involvement in experimental design, execution of Trastuzumab and ribosomal protein experiments, data analysis, and manuscript preparation.**

4.0. Abstract

Native mass spectrometry (MS) is a rapidly advancing field in the analysis of proteins, protein complexes, and macromolecular species of various types. The majority of native MS experiments reported to-date has been conducted using direct infusion of purified analytes into a mass spectrometer. In this study, capillary zone electrophoresis (CZE) was coupled online to Orbitrap mass spectrometers using a commercial sheathless interface to enable high-performance separation, identification and structural characterization of limited amounts of purified proteins and protein complexes, the latter with preserved non-covalent associations under native conditions. The performance of both bare-fused silica and polyacrylamide-coated capillaries was assessed using mixtures of protein standards known to form non-covalent protein-protein and protein-ligand complexes. High-efficiency separation of native complexes is demonstrated using both capillary types, while the neutral-coated capillary showed better reproducibility and higher efficiency for more complex samples. The platform was then evaluated for the determination of monoclonal antibody aggregation and for analysis of proteomes of limited complexity using a ribosomal isolate from *E. coli*. Native CZE-MS, using accurate single stage and tandem-MS measurements, enabled identification of proteoforms and non-covalent complexes at femtomole levels. This study demonstrates that native CZE-MS can serve as an orthogonal and complementary technique to conventional native MS methodologies with the advantages of low sample consumption, minimal sample processing and losses, and high throughput and sensitivity. This study presents a novel platform for analysis of ribosomes and other macromolecular complexes and organelles, with the potential for discovery of novel structural features defining cellular phenotypes (e.g. specialized ribosomes).

4.1. Introduction

Native mass spectrometry (MS) enables the analysis of proteins, non-covalent assemblies of proteins, small ligand molecules and ions associated with proteins, and even macromolecular particles (e.g. viruses, ribonucleoprotein particles, exosomes) in their unaltered, biologically functional states under near-physiological conditions [93, 134, 291-294]. In native MS, non-covalent interactions are preserved, and tertiary and quaternary structures maintained [74, 295]. Recently, native MS has emerged as a powerful tool to provide information on topology, stoichiometry, function, dynamics, and composition of protein assemblies [74, 91, 296]. In more conventional top-down and intact MS characterization of proteins, non-covalent associations are lost as the protein is unfolded under denaturing conditions. Unfolding leads to an increased exposure of the protein to the aqueous buffer, thereby resulting in additional protonation. The resultant MS signal is then divided into a larger number of charge states in comparison to native MS conditions, resulting in complex spectra with poorly resolved ion species (especially those corresponding to molecular species with subtle structural differences), potential difficulties in data interpretation, and lower sensitivity (**Figure 4-1**). Importantly, the lower charge states under

native conditions lead to the increased *m/z* spacing (resolution) of ion clusters corresponding to different charge states, potentially resulting in acquisition of structurally more informative spectra in analysis of proteoforms and similar M_r high molecular mass species. Additionally, experimental reproducibility is improved in native MS experiments in comparison to non-native approaches (i.e. middle-down, bottom-up, and top-down), as less sample handling and processing is required.

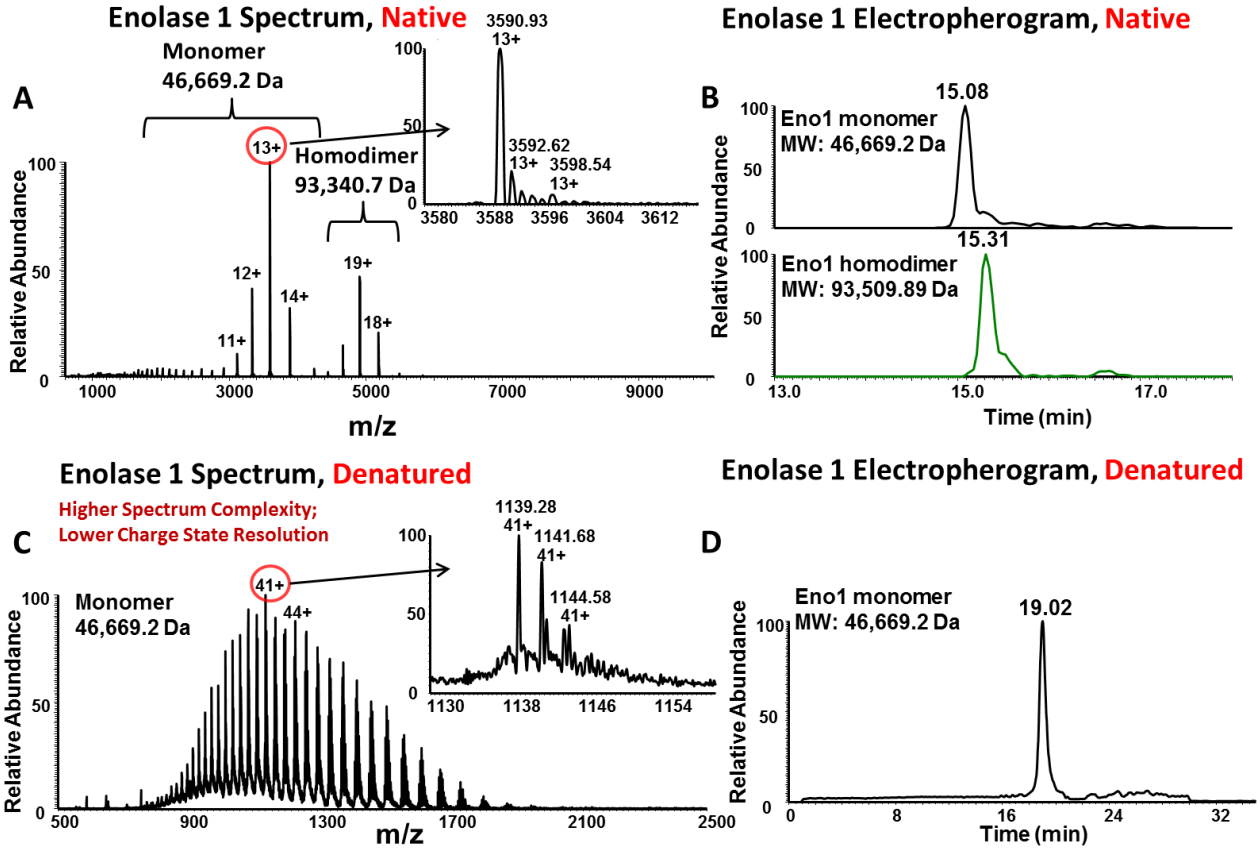


Figure 4-1. Native CZE-MS Allows for the Separation and Characterization of Protein Complexes. Online native CZE-MS resulted in the separation and detection of enolase 1 homodimer and monomer. CZE-MS[1] analysis under denaturing conditions resulted in the dissociation of the homodimer and detection of a single electrophoretic peak corresponding to the monomeric species.

Subtle differences in protein structure and conformation may result in substantial differences in biological function of a protein or therapeutic properties of a biopharmaceutical. In biological systems, a protein encoded by a single gene may exist as a population of different proteoforms, e.g., charge isoforms, molecules with a variety of post-translational modifications (PTMs), sequence variants, products resulting from truncation of protein, or alternative mRNA splicing [4, 263]. Additionally, chemical modifications of proteins may occur at various stages of sample procurement, preparation, storage and analysis. As a result, comprehensive

characterization of a single isolated protein – let alone a complex mixture – is highly challenging, requiring separation prior to high-resolution high mass accuracy MS analysis. We and others have recently demonstrated the power of CZE-MS in high-resolution separations of pharmaceutical glycoproteins [297-299]. In our study, a sheathless cross-linked polyethyleneimine-coated capillary coupled to the Orbitrap Elite MS was used to separate proteoforms of recombinant human interferon-β1 (Avonex), resulting in the detection of 138 proteoforms and the separation of deamidated, sialylated, truncated species, and positional glycoform isomers [299].

Nanoflow (<1 μL/min) and, especially, ultra-low flow (≤25 nL/min) separation techniques coupled to nanoESI are highly desirable for native and top-down analyses of protein structures where an MS signal is divided into multiple channels corresponding to ion species of different charge states [300, 301]. In native MS, where physiological or near-physiological conditions are required, the task of separating non-denatured proteins and protein complexes at high resolution is further complicated. Conventional separation techniques compatible with native MS include size exclusion chromatography (SEC) [302], ion exchange chromatography (IEX) [163], hydrophobic interaction chromatography (HIC) [303], affinity chromatography [304], capillary isoelectric focusing (cIEF) [305, 306], native gel-eluted liquid fraction entrapment electrophoresis (GELFrEE) [167, 307], native polyacrylamide gel electrophoresis [308], capillary zone electrophoresis (CZE) [67, 309] and flow field-flow fractionation (F4) [310, 311]. However, for many of these methods, online interfacing to a mass spectrometer is challenging or yet impossible. SEC, which can allow separation under native conditions, provides insufficient resolution of species with subtle structural differences. cIEF may lead to unfolding of proteins and dissociation of protein complexes at their pIs and require the use of denaturing sheath flow buffers to maintain stable electrospray, which may limit native MS applications to relatively stable complexes. The native GELFrEE approach, although viable for the separation of native multimeric complexes and proteins in mixtures, has not yet been directly interfaced with MS or demonstrated the capability of high-resolution separations of charge-based proteoforms and species with very subtle structural differences [167, 307].

In contrast to the above techniques, capillary zone electrophoresis (CZE) can achieve high-resolution separation under native state conditions. CZE separates molecular species according to differences in their electrophoretic mobilities, which are dependent on analyte charge and hydrodynamic volume. Furthermore, negligible sample and buffer quantities are required, and sample losses and carryover are minimal due to the open tubular capillary geometry [312]. Currently, there are several commercial solutions for interfacing CZE online to MS that can potentially be used for native MS [60, 61, 65, 67, 313]. The ability to separate non-covalent complexes from corresponding unbound complex-forming constituents can address a long-standing debate in the field of native ESI-MS on whether protein complexes and aggregates are formed in the gas phase during electrospray ionization, desolvation and ion manipulation in the mass spectrometer or the native MS data closely represent the structures of non-covalent complexes existing in aqueous solutions of biological systems [91, 314]. Such separation

supports the capability of native MS to infer the structures of non-covalent macromolecular complexes existing in-solution from the results of gas-phase native ESI-MS measurements [315, 316].

Historically, native MS has been largely limited to the domain of TOF mass analyzers due to the high upper *m/z* thresholds of detection of these instruments (up to 100,000 Th) [134]. However, native MS analyses with TOF analyzers suffer from the limited resolving power of these instruments. Recent studies have demonstrated that high resolution Orbitrap mass analyzers can be used for the characterization of native proteins and large protein complexes [134, 135, 317]. For example, a modified Exactive Plus Orbitrap instrument was applied to distinguish between ATP- and ADP-bound states of the 0.8 MDa *E. coli* chaperonin GroEL, a fourteen subunit complex [134]. It was further demonstrated that a modified Exactive Plus Orbitrap instrument could be used for native MS experiments of substrates with molecular weights approaching 10.0 MDa while still possessing sufficient resolution and sensitivity to detect small changes to the large molecular mass of the multisubunit protein complex or macromolecular assembly [134, 318]. Moreover, recent studies have shown that an Orbitrap-based instrument can efficiently perform ion isolation of the complex and fragmentation into individual monomeric subunits (MS^2) followed by subsequent isolation of subunit ions and fragmentation into backbone fragments (MS^3) [135, 317].

Considering that *a vast array of biological functions are conducted by protein assemblies rather than individual proteins*, investigation of functional protein complexes, their composition, architecture, stoichiometry, and structural features is of great importance to understanding biological systems. The complexity of proteins and protein assemblies, as well as the heterogeneity of proteoforms in biomedical samples, necessitates both a high-efficiency separation component and a sensitive mass spectrometer of a high resolving power and an appropriate *m/z* range. In the present work, we survey the potential of an online sheathless CZE-Orbitrap MS platform to investigate model mixtures of intact proteins and protein complexes, heterogeneous biopharmaceutical proteins, and limited complexity proteomic samples. The approach is initially evaluated using model proteins with non-covalently associated multisubunit and small ligand entities, a biopharmaceutical monoclonal antibody, and then explores the *E. coli* ribosomal proteome. The platform facilitated the detection of numerous proteoforms and non-covalent complexes, with structures being confirmed by CZE-MS^2 experiments. Particular molecular species could not be detected under denaturing conditions or in native direct infusion experiments. CZE separation, followed by higher energy collision-induced dissociation (HCD) MS^2 resulted in the identification of 42 ribosomal protein groups, 144 proteoforms and 679 proteoform-spectrum matches (PfSMs) in a single analysis of the *E. coli* ribosomal proteome under native conditions. Characterization of the same sample by native CZE-MS^1 and denaturing CZE-MS^2 resulted in identification of 9 additional ribosomal protein groups, corresponding to a total of 51 identifications, which indicated the complementarity of techniques. Importantly, native CZE-MS^2 resulted in a superior depth of proteomic profiling for the *E. coli* ribosomal proteome in comparison to denaturing CZE-MS^2. While extraction of proteins from intact

ribosomes involved a brief exposure to denaturing conditions prior to reconstitution into a buffer of a physiological pH, the resulting reconstituted protein extract was used as a model sample to assess the ability of native CZE-MS to survey complex biological samples that may contain non-covalent molecular assemblies that either survived the sample cleanup procedure or got re-associated and re-folded under native conditions. Several non-covalent associations of *E. coli* ribosomal proteins were elucidated using native CZE-MS analyses that could not otherwise be observed using conventional denaturing top-down or bottom-up approaches. Our work demonstrates the potential of CZE interfaced online with advanced mass spectrometry to enable native, high-resolution separations of proteins and protein complexes in their physiologically-relevant states and facilitate structurally-informative analyses *via* single stage and tandem MS under native non-denaturing conditions.

4.2. Experimental

4.2.1. Chemicals and Materials

All chemicals and materials used in this study were purchased from Sigma-Aldrich (St. Louis, MO) unless stated otherwise.

4.2.2. Preparation of the High-Purity Trastuzumab Standard

Trastuzumab (Herceptin, Genentech, South San Francisco, CA), kindly provided by Alain Beck (Centre d'Immunologie Pierre Fabre, St Julien-en-Genevois Cedex, France), was diluted in 25 mM ammonium acetate to a final concentration of 0.8 mg/mL and cleaned-up via Micro Bio-Spin size-exclusion columns with Bio-Gel P6 (SEC, Bio-Rad, Hercules, CA).

4.2.3. Preparation of Mixtures of Proteins and Protein Complexes

All protein standards were purchased from Sigma-Aldrich in lyophilized form. The following proteins were used: Enolase (P/N E6126; *S. cerevisiae*), carbonic anhydrase (P/N C3934; *Bos taurus*), alcohol dehydrogenase (P/N A8656; *S. cerevisiae*), protein disulfide isomerase (P/N P3818; *Bos taurus*), alpha crystallin (P/N C4163; *Bos taurus*), myoglobin (P/N M0630; *E. caballus*), cytochrome C (P/N C3131; *Bos taurus*), ribonuclease A (P/N R6513; *Bos taurus*), cytochrome C reductase type I (P/N C3381; *S. scrofa*), pyruvate kinase type III (P/N P9136; *Oryctolagus caniculus*), and fibrogen (P/N F8630; *Bos taurus*). All proteins were reconstituted in 25 mM ammonium acetate, pH 7.5 and mixed to the final concentrations specified below. Desalting was performed with Micro Bio-Spin 6 columns (Bio-Rad, Hercules, CA). A mixture of three proteins was prepared from enolase (9.18 μM), carbonic anhydrase (4.00 μM**)**, and alcohol dehydrogenase (23.5 μM). A mixture of eleven proteins and protein

complexes consisted of enolase (9.18 µM), carbonic anhydrase (4 µM), alcohol dehydrogenase (23.5 µM), pyruvate kinase type III (15.0 µM), alpha crystallin (21.6 µM), myoglobin (7.17 µM), cytochrome C (10.4 µM), ribonuclease A (17.8 µM), cytochrome C reductase type I (10.2 µM), protein disulfide isomerase (15.0 µM) and fibrogen (17.4 µM).

4.2.4. Preparation of the Ribosomal Protein Extract

The *E. coli* ribosomal protein extract was prepared as described [319, 320]. In brief, to remove RNA, 33 µL of cold, glacial acetic acid and 0.25 volumes (8.25 µL) of 100 mM magnesium chloride was added to 1 mg of the intact *E. coli* ribosomal isolate (13.3 µM; P0763S, New England Biolabs, Ipswich, MA). Next, the sample was spun-down at 10,000 RPM for 5 minutes; the resultant supernatant was collected and cleaned-up two times into 25 mM ammonium acetate, pH 7.5 using Micro Bio-Spin 6 SEC columns, according to manufacturer's protocol (Bio-Rad). The pH of the mixture was closely monitored throughout the prep and brought close to neutral pH.

4.2.5. Capillary Zone Electrophoresis Instrumentation and Methods

All capillary electrophoresis equipment, including bare fused silica (BFS) and neutral polyacrylamide-coated capillaries with hydrofluoric acid-etched emitters for sheathless nanoESI (aka CESI), were provided by SCIEX Separations (Brea, CA). A commercial CESI 8000 instrument was interfaced with either Exactive Plus extended mass range (EMR) or Q Exactive Plus mass spectrometers (Thermo Fisher Scientific) in all CZE-MS experiments. Both the BFS and polyacrylamide-coated capillaries were 30 µm inner diameter and 90 cm long, and contained an integrated sheathless etched porous nanospray tip.

For the analysis of the mixture of 3 standard proteins and complexes on the sheathless BFS – Exactive Plus EMR mass spectrometer, the capillary was first rinsed with 0.1 M sodium hydroxide (100 psi, 5 min), followed by 0.1 M hydrochloric acid (5 min) and nanopure water (5 min). Next, the separation capillary was conditioned with the BGE (25 mM ammonium acetate, pH 7.5) at 100 psi for 10 min. The conductive line capillary contained 3% acetic acid solution at 75 psi flow. Sample injection was performed at 2.5 psi for 15 seconds, corresponding to approximately 6 nL of sample (0.8% of the capillary volume). Separations were conducted for 60 minutes with a 20 kV potential applied across the capillary with a ramp time of one minute. At the end of the run, a ramp-down from 20 kV to 1 kV over five minutes was performed back to the start conditions for the next injections.

For the analysis of samples on the sheathless polyacrylamide-coated capillary, the column was first rinsed at 100 psi with 0.1 M hydrochloric acid (5 min), DDI water (5 min), followed by the BGE (40 mM ammonium acetate, pH 7.5) for 10 minutes; and the conductive line capillary was rinsed with a 3% acetic acid at 75 psi for 4 min. Injection was performed at 2.5

psi for 15 seconds, corresponding to approximately 6 nL of sample (0.8% of the capillary volume). Separations were performed for 60 minutes at 20 kV with a ramp time of one minute. Due to the significant suppression of EOF on polyacrylamide-coated capillaries, supplemental pressure was applied to generate flow of solvent and facilitate the migration of analytes towards the MS detector. For the analysis of the mixture of 11 proteins, 3 psi was used. For the ribosomal protein extract, separations were performed also at 3 psi supplemental pressure in CZE-MS^1 experiments or with 1 psi supplemental pressure applied for the first 15 minutes of the separation, followed by a supplemental pressure of 3 psi for the remainder of the run (in CZE-MS^2 experiments). A ramp-down from 20 kV to 1 kV over a five minute period was run after the conclusion of the separation.

4.2.6. Mass Spectrometry

All CZE-MS experiments performed on the Exactive Plus EMR (Thermo Fisher Scientific) used the following settings: a 1.5 kV spray potential, resolution of 35,000 at 200 Th, automatic gain control (AGC) of 3e6 and maximum injection time of 250 ms. In-source collision-induced dissociation (100 V x charge) was applied to provide sufficient desolvation without dissociation of non-covalently-bound subunits or fragmentation of covalent bonds. The instrument was operated in extended mass range mode (EMR). For AIF, fragmentation was performed with 150 V x charge in the HCD cell.

For the native CZE-MS^1 analysis of the *E. coli* (strain B/ BL21-DE3; UniProt) ribosomal protein extract, the Q Exactive Plus (Thermo Fisher Scientific) was operated with 1.3 – 1.5 kV spray potential, 100 eV for in-source collision-induced dissociation, 1e6 for target AGC, 250 ms maximum injection time, and a 17,500 resolution at 200 Th. For CZE-MS^2 experiments, the target AGC was set to 3e6, and the maximum injection time was 250 ms at a resolution of 140,000 at 200 Th. For native CZE-MS^2, the normalized collision energy (NCE) was set at 33%. For both native CZE-MS^1 and CZE-MS^2, the instrument was operated in the standard mode with a HV pressure of approximately 3.6E-05 mbar.

For denaturing CZE-MS^1, the MS detection settings were similar as the native experiments, with the exception of a lower in-source collision-induced dissociation energy (15 eV) and optimum NCE of 25% (for CZE-MS^2).

For native MS infusion experiments of Trastuzumab, the polyacrylamide-coated capillary and a CESI 8000 instrument were used as an infusion platform into the Q Exative Plus MS. Infusion was achieved by filling the capillary with Trastuzumab (2 mg/mL, 20 mM ammonium acetate, pH 8.0) *via* an injection of 100 psi for 1 minute. Next, the separation capillary was inserted into the buffer vial at the anodic end containing 20 mM ammonium acetate, pH 8.0. A l supplemental pressure of 2.5 psi and if needed a 20 kV potential were applied to facilitate flow of liquid towards the MS inlet (cathodic end). These conditions were identical to parameters in CZE-MS experiments for this sample. A 1.5 kV voltage was applied to facilitate stable nESI.

The inlet capillary temperature of the MS was 350 degrees Celsius and the rest of the MS parameters were set as defined above.

4.2.7. Data Analysis

Analysis of native CZE-MS^1 runs from all system/instruments was performed using both XCalibur and Protein Deconvolution 4.0 (Thermo Fisher Scientific). Deconvolution was performed using the Sliding Window approach with either ReSpect for isotopically unresolved data, or Xtract for high resolution CZE-MS^1 runs where isotopic resolution was achieved. Analysis of all-ion fragmentation (AIF) MS data acquired on an Exactive Plus EMR was performed using Byonic software (Version 2.5.6, Protein Metrics, San Carlos, CA) or manually, using fragment masses predicted by the Protein Prospector software (UCSF). It should be noted that for the analysis of AIF data acquired on an Exactive Plus EMR, which lacks a quadrupole and therefore the precursor ion selection capability for MS^2, spectra in the regions of the CZE-MS^1 run corresponding to migration times of separated or partially-separated proteins or protein complexes were averaged and analyzed individually. However, since baseline separation was not achieved for all mixture constituents, AIF-generated fragments from co-migrating species were often detected within the same AIF spectrum. For all CZE-MS^2 experiments, Byonic (Version 2.5.6, Protein Metrics, San Carlos, CA) was employed for data interpretation. Byonic searches were performed with fully-specific digestion specificity and a maximum of two missed cleavages, as well as a maximum of one rare and one common modification for each proteoform-spectrum match (*PfSM*). Also, a 20 ppm fragment mass tolerance and a 10 ppm precursor mass tolerance were used in all Byonic searches. Common and expected modifications for the model protein mixtures and ribosomal extract were manually inputted into Byonic and used as dynamic modifications. Dynamic modifications included oxidation (methionine), deamidation (asparagine), phosphorylation (histidine, lysine, arginine, serine, threonine, and tyrosine), acetylation (N-terminal and lysine), N-terminal methionine loss, coordination of magnesium (2+) and zinc (2+) ions (C-terminal, aspartate, and glutamate), methylation (lysine and arginine), dimethylation (lysine), and N-terminal pyroglutamation (glutamate and glutamine). For analysis of ribosomal proteins and mixtures of protein standards, a database containing 208 protein sequences that combined the results of bottom-up proteomic profiling of the *E. coli* ribosomal isolate that included sequences of identified ribosomal proteins (UniProt, *E. coli* strain B / BL21-DE3), ribosome-associated proteins and sequences of common contaminants was used.

4.2.8. Bottom-Up NanoLC-MS-Based Proteomic Analysis

NanoLC-MS proteomic profiling was performed in a similar fashion to that described in [321]. A fused silica capillary (75 μm internal diameter, 20 cm long) (Polymicro Technologies, Phoenix, AZ) was packed with Magic C18AQ, 3 μm diameter, 200 Å pore size beads from Michrom Bioresources (Auburn, CA, USA). Liquid chromatography (LC) separation was

performed on an Ultimate 3500 system from Thermo Fisher Scientific, and mass spectrometry data was acquired on an LTQ Orbitrap XL ETD (Tune Plus version 2.5.5) from Thermo Fisher Scientific. The sample was electrosprayed using a distal-coated, 20 μm internal diameter, 360 μm outer diameter tip with a 10 μm opening from New Objective (Woburn, MA, USA) connected directly to the column head using a Teflon sleeve.

Reversed-phase liquid chromatography was carried out using mobile phase A: 0.1% formic acid in water and mobile phase B: 0.1% formic acid in acetonitrile. Approximately 100 fmol of each trypsin digested protein was injected directly onto the analytical column with 2% B at 250 nL/min. The sample was loaded and desalted over 10 minutes in 2% B and separated at 250 nL/min flow rate using the following linear gradient: 2-37% B over 60 minutes, 37-95% B over 10 minutes, 95% B hold for 9 minutes, 95% to 2% B in 1 minute, 2% B hold for 10 minutes. Data were acquired on an LTQ Orbitrap XL mass spectrometer using parameters for conventional Top 10 analysis enabled through Xcalibur (v. 2.0.7). Full precursor scans were acquired over a 400-1700 Th range in the Orbitrap at 60k resolution (at 400 Th), with the AGC target set to 1e6 and 500 ms maximum ion accumulation time. The 10 highest intensity eligible precursors (Top 10) were sequentially selected in 2 Th wide isolation windows centered on the monoisotopic peak and activated for maximum 30 ms by CID at 35% normalized energy. In fragment ion accumulation, the AGC target was set to 200,000, the maximum injection time was 100 ms, and the ion trap was scanned at 0.5 Da resolution using dynamic exclusion of 60 seconds. Ions with a +1 charge state and lower than 500 intensity units were considered chemical noise.

The acquired data were analyzed in Proteome Discoverer 1.4 (Thermo Fisher Scientific) with Sequest HT peptide spectral match scoring and Percolator validation and filtering ($q < 0.01$) [322]. The search was conducted against a concatinated bovine, *E. coli*, equine, human, porcine, rabbit and *S. cerevisiae* (ATCC) UniProt databases containing canonical proteins and known variants from March 2015 appended with 47 common contaminants. The maximum precursor mass error was set to 20 ppm and the maximum fragment mass error was set to 0.6 Da. Carbamidomethylation of cysteine was set as a static modification, and oxidation of methionine, N-terminal acetylation and deamidation of glutamine and asparagine were set as dynamic modifications. Up to 2 missed tryptic cleavages were allowed. Amino acid sequences of all proteins identified in bottom-up experiments were consolidated into a single database, which was used for the identification of species in the intact CZE-MS^1 and top-down CZE-MS^2 approaches.

4.3. Results and Discussion

4.3.1. Analytical Workflow

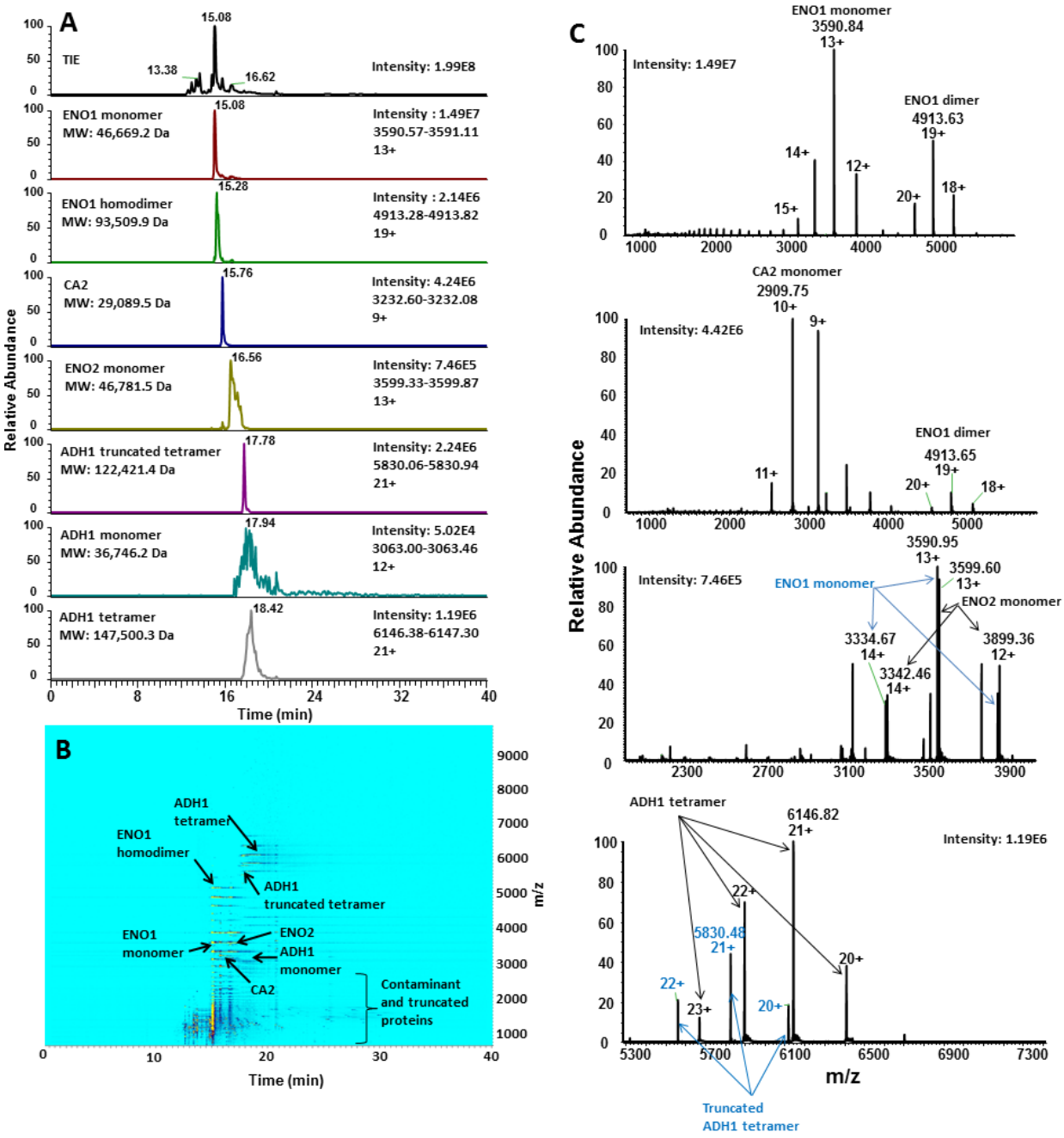
In this study, we demonstrate the power of CZE as an online high resolution separation technique compatible with native state MS analysis using: (1) mixtures of complex-forming

protein standards, (2) a monoclonal antibody (Trastuzumab), and (3) an extract of the *E. coli* ribosomal proteome. Both bare fused silica (BFS) and polyacrylamide-coated capillaries were used for CZE separations under native conditions interfaced online *via* a SCIEX sheathless porous interface (CESI) for high sensitivity [67]. The CESI interface enabled the online integration of high resolution CZE separations with MS conducted at ultra-low flow rates (measured at 10 – 20 nL/min) and without analyte dilution prior to ionization [65, 323]. The flow rates used in our CZE-MS experiments are substantially lower than those typically utilized in more conventional sheath-liquid CZE platforms [67]. Ultra-low flow rate separations coupled to nanoESI-MS result in improved ionization efficiency, enhanced ion transfer into the mass spectrometer, and decreased ion suppression, collectively leading to the enhanced MS detection sensitivity [300, 301, 324]. We found that the acidic buffer used in the conductive line of the interface (3% acetic acid) did not significantly alter the pH of the electrosprayed BGE (e.g. 20-40 mM ammonium acetate, pH 7.5- 8.5) , which was important for maintaining native conformations of analytes. This is attributed largely to the design of the sheathless porous CESI interface (SCIEX), which allows for the limited passage of only small ions (i.e. not liquid) from the conductive line into the separation capillary at the CZE/ESI-MS porous junction without analyte dilution [65, 67, 323]. While the limited transport of hydroxonium ions from the conductive line into the separation capillary at the porous junction allowed for closing the CZE and ESI electric circuits and ionizing analyte molecules as well as for the improved sensitivity, stability and performance of CZE-ESI MS, the pH of the resultant solution measured at the outlet of the capillary emitter was found within the native range (pH 6.5-7.5). Additionally, the porous junction is approximately 2 cm in length at the outlet of the separation capillary, which makes the exposure of the analyte molecules to this minor change of the pH very short (seconds). Nearly identical results were found when 10% acetic acid was used in the conductive line. Identification of proteins and protein complexes under native conditions involved high mass accuracy single stage intact MS (CZE-MS^1), followed by either all-ion fragmentation (AIF) on an Exactive Plus EMR or data-dependent high energy collision dissociation (DDA-HCD) CZE-MS^2 on a Q Exactive Plus. Peptide maps of all proteins were obtained *via* conventional bottom-up nanoLC-MS.

4.3.2. Native CZE-MS of Mixtures of Complex-Forming Proteins

A mixture of three protein standards known to form non-covalent multi-subunit homomeric complexes and complexes with coordinated metal-ion ligands (alcohol dehydrogenase (*S. cerevisiae*), enolase (*S. cerevisiae),* and carbonic anhydrase (*Bos taurus*)) was analyzed using a BFS capillary coupled to an Exactive Plus EMR MS (**Figure 4-2**). This experiment was performed as a proof-of-principle for online native CZE-MS using a relatively simple mixture of three proteins and associated complexes. According to the product specifications, alcohol dehydrogenase 1 (ADH1) was expected to be observed as a homotetramer with the molecular mass of 141 - 151 kDa; enolase (ENO1) as a homodimer of 93 kDa with

bound Mg^{2+} ions, and carbonic anhydrase 2 (CA2) as a 29 kDa complex with coordinated Zn^{2+} ions. Bottom-up proteomic analysis of the tryptic digested proteins revealed the presence of contaminants and isozymes, such as ENO1 and ENO2, as well as protein fragments, corresponding to truncated protein standards. Based on the results of bottom-up analysis and observed ion species in native CZE-MS experiments, theoretical molecular masses were calculated for monomers and complexes (see **Tables 4-1 and 4-2**). CZE-MS[1] was first conducted under native conditions using 25 mM ammonium acetate, pH 7.5 as the BGE, and the MS resolving power was set to 35,000 at 200 Th. Together, these CZE and MS conditions resulted in the separation and detection of the major components of the mixture between 12 and 32 minutes, including ENO1 homodimer and ADH1 tetramers (**Figure 1A-B**). ESI mass spectra of ENO1 monomer and homodimer, ENO2, CA2, and ADH1 tetramers are shown in **Figure 4-2C**. On the basis of accurate deconvoluted masses, the major mixture components, in addition to the most abundant proteoforms, were identified, as detailed next and in **Table 4-1**.



Figures 4-2 A-C. Native CZE-MS analysis of a mixture of three protein complex-forming standards. Monomers of ENO1 and ENO2, CA2, and ADH1 were detected and separated, as well as multimeric complexes for ENO1 (homodimer) and ADH1 (homotetramer and truncated homotetramer). For the separation performed on the sheathless BFS – Exactive Plus EMR system, extracted ion electropherograms (XIEs) for the most abundant ions of each protein/protein complex are shown (**A**). The separation and pattern and relative intensities can be observed on the ion density map (**B**). Examples of integrated MS spectra are shown for several major mixture components: ENO1 and ENO2, as well as truncated ADH1 and unmodified ADH1 tetramers (**C**).

Protein	Monoisotopic M	Average M	Average Mass (Deconvoluted)	# of Charge States	Migration Time	Mod (tentative)
Enolase 1 (S. cerevisiae)	46656.23	46684.38	46669.23	18	15.08	Met removed (-131.1961), I-->V (-14.0266)
Enolase 1 (S. cerevisiae)	46656.23	46684.38	46709.53	18	15.08	Met removed (-131.1961), I-->V (-14.0266), +Mg (+24.305 Da), +ammonium ion (+17.03Da)
Enolase 1 (S. cerevisiae)	46656.23	46684.38	46731.73	18	15.08	Met removed (-131.1961), I-->V (-14.0266), +Phospho (+79.98 Da), +ammonia loss (-17.03 Da)
Enolase 1 (S. cerevisiae)	93312.46	93368.75	93341.66	7	15.08	Homodimer, Met removed (-131.1961), I-->V (-14.0266)
Enolase 1 (S. cerevisiae)	93312.46	93368.75	93439.74	7	15.08	Homodimer, Met removed (-131.1961), I-->V (-14.0266), +4Mg (+24.305 Da each)

Protein	Monoisotopic M	Average M	Average Mass (Deconvoluted)	# of Charge States	Migration Time	Mod (tentative)
Enolase 1 (S. cerevisiae)	93312.46	93368.75	93395	5	15.08	Homodimer, Met removed (-131.1961), I-->V (-14.0266), +2Mg (+24.305 Da each)
Enolase 1 (S. cerevisiae)	93312.46	93368.75	93509.89	5	15.08	Homodimer, Met removed (-131.1961), I-->V (-14.0266), +3Mg (+24.305 Da each), +Phospho (+79.98 Da), +ammonium ion (+17.03Da)
Carbonic anhydrase 2 (bovine)	28964.67	28982.23	29086.81	9	15.78	Met removed (-131.1961),+Zn^{2+} (+65.382), acetylated (+42.0367)
Carbonic anhydrase 2 (bovine)	28964.67	28982.23	29024.15	5	15.78	Met removed (-131.1961), acetylated (+42.0367)

Protein	Monoisotopic M	Average M	Average Mass (Deconvoluted)	# of Charge States	Migration Time	Mod (tentative)
Carbonic anhydrase 2 (bovine)	28964.67	28982.23	29107.74	4	15.78	Met removed (-131.1961), acetylated (+42.0367), phosphorylated (+79.9799)
Alcohol dehydrogenase 1 (S. cerevisiae)	36825.67	36848.78	36745.28	19	17.46	Monomer, Met removed (-131.1961), acetylated (+42.0367),+2xZn 2+ (+65.382), I->V(-14.0266)
Alcohol dehydrogenase 1 (S. cerevisiae)	36825.67	36848.78	147500.3	9	20.74	Homoteramer, Met removed (-131.1961), acetylated (+42.0367),+2xZn^{2+} (+65.382) per subunit, I->V(-14.0266)
Enolase 2 (S. cerevisiae)	46753.97	46782.98	46781.87	7	16.56	Monomer, Met removed (-131.1961)

Table 4-1: Major components of the three protein mixture identified from CZE-MS[1] data acquired under native conditions. Identifications were achieved by manual matching of experimental and theoretical molecular masses.

CZE separation allowed structural analyses that could not be easily achieved using direct infusion. Partial separation was obtained between the ENO1 monomer (46,669.2 Da) and homodimer (93,509.9 Da), which were present at similar CZE-MS^1 signal intensities (**Figure 4-2 A, C and Figure 4-1**). The separation yielded sufficient peak resolution between ENO1 monomer and dimer to enable accurate quantitative analysis and MS deconvolution (**Figure 4-3)**. The latter is important for correct assignments of proteoforms and non-covalent binding (e.g. metal ions and multiple subunits) features that often cannot be analyzed by direct infusion. Both monomer and homodimer of ENO1 were also separated from ENO2 (46,781.5 Da), which has ~95% sequence homology with ENO1 and migrated as a broader peak (**Figure 4-3A and Table 4-1**). Considering the high sequence homology of these isozymes, the peak width may correlate with the conformational stability of a protein under the employed separation conditions or the complexity of the mixture of proteoforms. The corresponding ENO2 homodimeric complex was not observed.

For ADH1, monomer (36,746.2 Da) (8%) and homotetramer (147,500.3 Da) (92%) peaks were detected with slight shifts in migration times of peak maxima (**Figure 4-4A)**. Based on these observations, information about the stability of complexes in biological samples and shifts in dissociation/association rates can be potentially generated based on CZE-MS^1 quantitative characteristics of protein assemblies and corresponding monomeric subunits. In addition, another species of the ADH1 standard was also detected, likely corresponding to the homotetramer of truncated ADH1 (122,421.4 Da, experimental MW) (**Figure 4-2A, C and Table 4-1**). Similarly, the combination of CZE separation and high mass accuracy MS detection under native conditions followed by MS deconvolution led to identification of CA2 (P00921, theoretical MW: 28,982.6 Da). Numerous contaminants, including intact and lower molecular mass truncated protein species were also detected in the mixture by both CZE-MS^1 primarily within the 800 - 2,500 Th range (**Figure 4-2B**) and bottom-up nLC-MS analysis (see **Table 4-2**).

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
P68083	Myoglobin Equus caballus	94.8	20	4712.3	Met-loss [N-Term]; Met-loss+Acetyl [N-Term]
P61823	Ribonuclease pancreatic Bos taurus	72.7	11	3533.7	
P00924	Enolase 1 Saccharomyces cerevisiae	83.1	43	4150.9	
P11974	Pyruvate kinase PKM Oryctolagus cuniculus	81	54	4439.5	Met-loss+Acetyl [N-Term]
P00330	Alcohol dehydrogenase 1 Saccharomyces cerevisiae	65.2	30	4840	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
P00921	Carbonic anhydrase 2 Bos taurus	80	36	3246.8	Met-loss+Acetyl [N-Term]
G1SQD3	Pyruvate kinase Oryctolagus cuniculus	64.4	42	3540.3	
P00925	Enolase 2 Saccharomyces cerevisiae	84	38	2851	
F1SHL9	Pyruvate kinase (Fragment) Sus scrofa	58.1	42	3206.1	
F1N0H3	Carbonic anhydrase 2 (Fragment) Bos taurus	76.3	27	2676.7	
A5D984	Pyruvate kinase Bos taurus	56.7	34	2835.3	
P02170	Myoglobin Oryctolagus cuniculus	42.9	7	1559.4	Met-loss+Acetyl [N-Term]; Met-loss [N-Term]
P02470	Alpha-crystallin A chain Bos taurus	94.2	21	1896.5	Met-loss [N-Term]; Acetyl [N-Term]
P00331	Alcohol dehydrogenase 2 Saccharomyces cerevisiae	21	7	2108.9	
F1MAV0	Fibrinogen beta chain Bos taurus	72.3	43	1458.7	
F1MGU7	Fibrinogen gamma-B chain Bos taurus	81.3	37	1269.2	
F1MNQ4	Superoxide dismutase [Cu-Zn] Bos taurus	75.7	10	923.3	
F1RUN2	Serum albumin Sus scrofa	66.1	46	1032.1	
P00442	Superoxide dismutase [Cu-Zn] Bos taurus	97.4	14	1134.4	
A5PJE3	Fibrinogen alpha chain Bos taurus	55	39	761.7	
P02510	Alpha-crystallin B chain Bos taurus	92	16	921.2	Met-loss [N-Term]; Acetyl [N-Term]
P12709	Glucose-6-phosphate isomerase Saccharomyces cerevisiae	57.2	23	739.2	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
P00883	Fructose-bisphosphate aldolase A Oryctolagus cuniculus	85.4	33	648.5	Met-loss [N-Term]
P62894	Cytochrome Bos taurus	68.6	15	455.1	
A6H7J6	Protein disulfide-isomerase Bos taurus	60.6	34	452.5	
P00445	Superoxide dismutase [Cu-Zn] Saccharomyces cerevisiae	94.2	13	626.1	Met-loss [N-Term]
F1RVN1	Uncharacterized protein Sus scrofa	76.5	29	520	
A6QLL8	Fructose-bisphosphate aldolase Bos taurus	66.2	25	515.7	Met-loss [N-Term]
P02769	Serum albumin Bos taurus	61	33	425	
P04264	Keratin, type II cytoskeletal Homo sapiens	42.5	21	424.1	
F1SRS2	Uncharacterized protein Sus scrofa	55.3	7	456.5	
B8Y9S9	Embryo-specific fibronectin 1 transcript variant Bos taurus	37.2	48	404.1	
B8Y9T0	Cumulus cell-specific fibronectin 1 transcript variant Bos taurus	37.3	46	386.4	
Q5XLD3	Creatine kinase M-type Sus scrofa	69.6	24	460	
F7D4V9	Cytochrome c Equus caballus	51.4	11	298	
Q2UVX4	Complement C3 Bos taurus	47.7	50	342.2	
F1RX37	Fibrinogen beta chain Sus scrofa	23.8	11	297.3	
P02189	Myoglobin Sus scrofa	69.5	10	342.3	Met-loss [N-Term]; Met-loss+Acetyl [N-Term]

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
P07318	Beta-crystallin B1 Bos taurus	82.6	18	287.6	
F7BYR9	Calsequestrin Equus caballus	36.6	13	319.4	
B3CL06	Serotransferrin Sus scrofa	62.5	39	304.4	
P13696	Phosphatidylethanolamine-binding protein 1 Bos taurus	79.1	13	248.4	
F1RFY2	Beta-enolase Sus scrofa	62	19	250.4	
F1MDC1	Uncharacterized protein Bos taurus	33.7	21	208.6	
F1SM98	Uncharacterized protein Sus scrofa	70.3	14	247.4	
D0G7F6	Triosephosphate isomerase Sus scrofa	79	15	218.8	
P04178	Superoxide dismutase [Cu-Zn] Sus scrofa	79.1	9	241	
Q5E956	Triosephosphate isomerase Bos taurus	74.3	14	212.4	
F6Z5C4	Uncharacterized protein (Fragment) Equus caballus	68.2	13	252.7	Acetyl [N-Term]
P52556	Flavin reductase (NADPH) Bos taurus	78.6	12	238.1	
Q6QAQ1	Actin, cytoplasmic 1 Sus scrofa	53.9	14	200.8	Met-loss+Acetyl [N-Term]
F6TZS9	Triosephosphate isomerase Equus caballus	77.9	14	191.2	
P00669	Seminal ribonuclease Bos taurus	23.3	3	257.5	
F1RKG8	Uncharacterized protein Sus scrofa	79.1	14	202	
P00817	Inorganic pyrophosphatase Saccharomyces cerevisiae	60.3	13	187.7	
P00942	Triosephosphate isomerase Saccharomyces	73.4	12	179.3	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
	cerevisiae				
F6RUR3	Uncharacterized protein Bos taurus	53.9	18	172.7	
G1TKY2	Uncharacterized protein Oryctolagus cuniculus	35.2	7	136.6	
Q1JPG7	Pyruvate kinase Bos taurus	51.3	17	156.3	
P02522	Beta-crystallin B2 Bos taurus	79.5	13	136.8	Met-loss+Acetyl [N-Term]
P00171	Cytochrome b5 Bos taurus	55.2	7	124.2	
P37012	Phosphoglucomutase 2 Saccharomyces cerevisiae	51.3	18	137.1	Met-loss+Acetyl [N-Term]
P15497	Apolipoprotein A-I Bos taurus	56.6	20	120.9	
G3X6L9	Uncharacterized protein Bos taurus	4.3	7	143.7	
Q28085	Complement factor H Bos taurus	35.6	24	145.5	
Q9TTJ5	Regucalcin Bos taurus	66.6	15	142.2	
P35908	Keratin, type II cytoskeletal 2 epidermal Homo sapiens	26.1	11	119.3	
F1SKC4	Uncharacterized protein (Fragment) Sus scrofa	45.3	13	143.3	
F1RGX4	Uncharacterized protein Sus scrofa	71.8	6	142	
P04406	Glyceraldehyde-3-phosphate dehydrogenase Homo sapiens	38.8	8	100.3	
I3LRR4	Uncharacterized protein (Fragment) Sus scrofa	58.3	14	150.2	
P28801	Glutathione S-transferase P Bos taurus	87.1	12	138.4	Met-loss [N-Term]

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
P11708	Malate dehydrogenase, cytoplasmic Sus scrofa	32.3	8	113.6	
F6R1X9	Uncharacterized protein Equus caballus	22.8	7	107	
F1SLA0	ATP synthase subunit beta Sus scrofa	48.3	18	111.4	
P48644	Retinal dehydrogenase 1 Bos taurus	51.5	16	107.2	
Q3SZX4	Carbonic anhydrase 3 Bos taurus	50	11	81.2	
Q3ZCJ6	Epoxide hydrolase 1, microsomal (Xenobiotic) Bos taurus	55.9	20	107.8	
Q29551	Succinyl-CoA:3-ketoacid coenzyme A transferase 1, mitochondrial Sus scrofa	46.9	15	114.7	
P00503	Aspartate aminotransferase, cytoplasmic Sus scrofa	54.7	16	108.5	Met-loss [N-Term]
P19141	Beta-crystallin B3 Bos taurus	67.3	10	115.7	Met-loss+Acetyl [N-Term]
P02067	Hemoglobin subunit beta Sus scrofa	87.1	12	102.7	Met-loss [N-Term]
E1BEL7	Heat shock protein beta-1 Bos taurus	73.4	10	107.7	
A4FUZ1	Lactoylglutathione lyase Bos taurus	66.8	10	94.2	Met-loss+Acetyl [N-Term]
P00355	Glyceraldehyde-3-phosphate dehydrogenase Sus scrofa	52.9	10	116.6	
F6WPB4	Uncharacterized protein Equus caballus	73	9	105.6	
I3LSI3	Adenylate kinase 2, mitochondrial Sus scrofa	61.9	11	130.3	Met-loss [N-Term]
F1RII7	Hemoglobin subunit beta Sus scrofa	87.1	12	102.7	Met-loss [N-Term]

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
P35527	Keratin, type I cytoskeletal 9 Homo sapiens	35.6	10	121.8	
F1MB08	Alpha-enolase Bos taurus	35	10	91.7	
P26444	Beta-crystallin A2 Bos taurus	65	7	91	
P08059	Glucose-6-phosphate isomerase Sus scrofa	44.1	15	97.3	Met-loss+Acetyl [N-Term]
G3X8B1	Peptidyl-prolyl cis-trans isomerase Bos taurus	34.1	7	67.6	Met-loss+Acetyl [N-Term]
P06733	Alpha-enolase Homo sapiens	38.2	10	95.2	
F1N5Q6	Beta-crystallin A3 Bos taurus	78.1	13	101.3	
A5D7J6	CALR protein Bos taurus	20.9	6	102.7	
Q05B89	Putative GTP cyclohydrolase 1 type 2 Bos taurus	41.6	11	100.8	
F1MCK2	Uncharacterized protein Bos taurus	17.2	16	48.4	
F1SA62	Uncharacterized protein (Fragment) Sus scrofa	47.9	7	89.2	
Q0VCX2	78 kDa glucose-regulated protein Bos taurus	21.8	10	84.4	
G3N0V0	Uncharacterized protein (Fragment) Bos taurus	35.6	7	65.4	
F1MCF8	Uncharacterized protein Bos taurus	22.2	4	94.9	
F1N3Q7	Apolipoprotein A-IV Bos taurus	50.8	17	74.7	
P15253	Calreticulin Oryctolagus cuniculus	20.8	6	93.7	
P00359	Glyceraldehyde-3-phosphate dehydrogenase 3	57.8	10	95	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
	Saccharomyces cerevisiae				
Q2KJ32	Selenium-binding protein 1 Bos taurus	44.3	12	98.6	
E1BH06	Uncharacterized protein Bos taurus	17	15	88.5	
Q5E946	Protein deglycase DJ-1 Bos taurus	68.8	9	78.9	
F1SRC5	Aconitate hydratase, mitochondrial (Fragment) Sus scrofa	41.3	17	90.5	
P33097	Aspartate aminotransferase, cytoplasmic Bos taurus	49.4	15	83.9	Met-loss [N-Term]
O77834	Peroxiredoxin-6 Bos taurus	54.5	9	88.4	Met-loss [N-Term]
F2Z5B6	Tropomyosin alpha-1 chain Sus scrofa	27.1	11	84.3	
F1N0R5	von Willebrand factor Bos taurus	10.8	18	96.1	
F1MC45	Complement factor H (Fragment) Bos taurus	31.5	15	88.5	
P02070	Hemoglobin subunit beta Bos taurus	76.6	11	65.6	
P02561	Tropomyosin alpha-4 chain Equus caballus	29.8	10	76.7	
Q01361	Cytochrome P450 2D14 Bos taurus	42.8	14	64.2	
P06753-2	Isoform 2 of Tropomyosin alpha-3 chain Homo sapiens	29.8	9	75.5	
F6X6J0	Histone H2A (Fragment) Equus caballus	25.9	3	54.7	
F7A281	Uncharacterized protein Equus caballus	32.7	11	68.5	
F1RLH7	Uncharacterized protein Sus scrofa	57	6	59.6	
F1SAF0	Dihydrolipoyl dehydrogenase Sus scrofa	32.4	9	90.2	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
P46976	Glycogenin-1 Homo sapiens	19.7	8	83.6	Met-loss+Acetyl [N-Term]
B9A041	Malate dehydrogenase, cytoplasmic Homo sapiens	29	4	66.9	Met-loss [N-Term]
Q05905	Protein HRI1 Saccharomyces cerevisiae	50	10	63.1	
E1BLS8	Uncharacterized protein Bos taurus	17.4	11	77.8	
F7AHF3	Uncharacterized protein Equus caballus	14.9	4	63.4	
F6QXW2	Uncharacterized protein Equus caballus	58.9	8	68.5	
E1B726	Plasminogen Bos taurus	32	15	70.1	
F1S031	Uncharacterized protein Sus scrofa	61.8	8	56	
P00443	Superoxide dismutase [Cu-Zn] Equus caballus	40.3	4	78.2	
Q2HJD7	3-hydroxyisobutyrate dehydrogenase, mitochondrial Bos taurus	42	9	69	
F1RMN7	Hemopexin Sus scrofa	33.3	9	55.6	
Q05752	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 7 Bos taurus	42.5	5	56.8	
P82460	Thioredoxin Sus scrofa	80	6	53.4	
P68363	Tubulin alpha-1B chain Homo sapiens	37.3	9	68.2	
P18648	Apolipoprotein A-I Sus scrofa	26.4	8	59.8	
I3L6G6	ATP synthase-coupling factor 6, mitochondrial Sus scrofa	44.4	6	73.4	
P00366	Glutamate dehydrogenase 1, mitochondrial	23.7	10	52.4	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
	Bos taurus				
P79274	Long-chain specific acyl-CoA dehydrogenase, mitochondrial Sus scrofa	33.5	12	59.6	
Q95274	Thymosin beta-4 Sus scrofa	77.3	4	64.2	Met-loss+Acetyl [N-Term]
F6Q6W4	Uncharacterized protein (Fragment) Equus caballus	57.8	7	50.7	
F7ANC9	Uncharacterized protein (Fragment) Equus caballus	37	9	65	
F6Z587	Uncharacterized protein Equus caballus	29	12	62.1	
F1MY85	Complement C5a anaphylatoxin Bos taurus	14.1	13	60.1	
O02772	Fatty acid-binding protein, heart Sus scrofa	66.2	8	55.3	Met-loss+Acetyl [N-Term]
F1SDX9	Peroxiredoxin 2 Sus scrofa	67.2	7	75.6	Met-loss+Acetyl [N-Term]
K7GN40	Uncharacterized protein Sus scrofa	53.5	13	64	
P80276	Aldose reductase Sus scrofa	56	11	49	Met-loss+Acetyl [N-Term]
P55052	Fatty acid-binding protein, epidermal Bos taurus	40	6	58.4	
I3L5D5	Uncharacterized protein Sus scrofa	17.4	10	35.7	
F1SMZ7	Uncharacterized protein Sus scrofa	30.3	11	54.5	
P12763	Alpha-2-HS-glycoprotein Bos taurus	11.7	3	35.7	
G1SFG6	Glucose-6-phosphate isomerase Oryctolagus	29.9	8	50.4	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
	cuniculus				
P00336	L-lactate dehydrogenase B chain Sus scrofa	34.4	10	48.5	
F1RYU5	NAD(P)(+)--arginine ADP-ribosyltransferase Sus scrofa	32.3	7	68.1	
G1TKH3	Superoxide dismutase [Cu-Zn] Oryctolagus cuniculus	29.4	3	52.9	
P10599	Thioredoxin Homo sapiens	46.7	4	37.9	
F1MPK6	Uncharacterized protein (Fragment) Bos taurus	5.9	8	54.5	
F7AUY9	Uncharacterized protein (Fragment) Equus caballus	22.1	3	42	
F6ZEV8	Uncharacterized protein Equus caballus	39.1	2	35.3	
P12026	Acyl-CoA-binding protein Sus scrofa	59.8	4	37.6	
Q3MHL4	Adenosylhomocysteinase Bos taurus	22.9	8	47	
Q6Q781	Calpastatin Sus scrofa	22.8	11	55.3	
P26779-2	Isoform 2 of Prosaposin Bos taurus	12.9	7	40.6	
Q9BGI3	Peroxiredoxin-2 Bos taurus	45.7	6	53.4	
G1SWS9	Uncharacterized protein Oryctolagus cuniculus	15.2	8	41	
P02671	Fibrinogen alpha chain Homo sapiens	2.5	2	28.1	
P20821	Glycine cleavage system H protein, mitochondrial Bos taurus	53.2	5	44.6	
P01966	Hemoglobin subunit alpha Bos taurus	30.3	3	36.7	
Q58DC0	Serine/threonine-protein phosphatase CPPED1 Bos taurus	39.9	7	48.5	Met-loss+Acetyl [N-Term]

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
Q03247	Apolipoprotein E Bos taurus	32	8	40.2	
P11842	Beta-crystallin A4 Bos taurus	44.3	6	49.3	
Q9N0J6	Fructose-1,6-bisphosphatase isozyme 2 Oryctolagus cuniculus	43.7	8	36.2	
F1MLX9	Graves disease carrier protein Bos taurus	27.8	10	26.2	
P62958	Histidine triad nucleotide-binding protein 1 Bos taurus	60.3	4	42.8	
E1BBP7	Histone H4 Bos taurus	28.8	3	31	
F1RTN3	Moesin Sus scrofa	24.6	9	42.2	
P52175	Nucleoside diphosphate kinase A 2 Bos taurus	45.4	5	37.3	
Q2EN76	Nucleoside diphosphate kinase B Sus scrofa	39.5	5	38.5	
F6VEA3	Proteasome subunit alpha type Equus caballus	39.4	6	39.8	
G1SDF2	Protein-L-isoaspartate O-methyltransferase Oryctolagus cuniculus	35.4	7	37	
Q2KJD0	Tubulin beta-5 chain Bos taurus	20	5	44.6	
F1MR63	Uncharacterized protein Bos taurus	24.8	7	39.2	
F1SJX1	Uncharacterized protein Sus scrofa	34.5	7	46.4	
I3LUZ8	Uncharacterized protein Sus scrofa	57.1	7	44.2	
F1ST01	Uncharacterized protein Sus scrofa	16.5	5	35.8	
I3LLT5	Uncharacterized protein Sus scrofa	12.5	5	30.2	
A6QQT4	Aldehyde dehydrogenase Bos taurus	19.8	7	28.7	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
Q7SIH1	Alpha-2-macroglobulin Bos taurus	10.9	9	30.3	
F1MLB8	ATP synthase subunit alpha Bos taurus	23.1	8	36.1	
A1E295	Cathepsin B Sus scrofa	21.5	5	40.6	
P00126	Cytochrome b-c1 complex subunit 6, mitochondrial Bos taurus	48.4	4	36.8	
Q0VD27	Enolase-phosphatase E1 Bos taurus	55.2	7	42	Met-loss [N-Term]
P02081	Hemoglobin fetal subunit beta Bos taurus	33.1	5	22	
P00724	Invertase 2 Saccharomyces cerevisiae	14.8	5	35.7	
Q2VTP6	Peptidyl-prolyl cis-trans isomerase Sus scrofa	74.1	6	40.7	Met-loss [N-Term]
Q3ZBG0	Proteasome subunit alpha type-7 Bos taurus	41.9	7	39.4	
P02639	Protein S100-A1 Bos taurus	46.8	3	40.1	Met-loss+Acetyl [N-Term]
F1N3A1	Thrombospondin-1 Bos taurus	4.6	4	25.5	
G1T8Z0	Uncharacterized protein (Fragment) Oryctolagus cuniculus	24.7	5	30.8	
F6R9F1	Uncharacterized protein Bos taurus	17.9	8	20.9	
F1SFI5	Uncharacterized protein Sus scrofa	26	8	37.2	
F1SKY2	Uncharacterized protein Sus scrofa	43.8	8	35.5	
P06504	Beta-crystallin S Bos taurus	45.5	5	34.8	
Q58DK5	Delta-aminolevulinic acid dehydratase Bos taurus	47.4	8	34.1	
P10790	Fatty acid-binding protein, heart Bos taurus	31.6	5	24.2	Met-loss+Acetyl [N-Term]
P00435	Glutathione peroxidase 1	46.3	6	26.8	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
	Bos taurus				
P00360	Glyceraldehyde-3-phosphate dehydrogenase 1 Saccharomyces cerevisiae	45.8	7	30.3	
P33198	Isocitrate dehydrogenase [NADP], mitochondrial (Fragment) Sus scrofa	23.3	7	33	
Q8WZA0-2	Isoform 2 of Protein LZIC Homo sapiens	24.2	6	36	
Q6QRN6	NAD(P)H-hydrate epimerase Bos taurus	29.5	5	35.5	
G3MY67	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (Fragment) Bos taurus	20.5	8	32.8	
M0QW17	Uncharacterized protein (Fragment) Bos taurus	15.8	7	23	
F1RKM1	Uncharacterized protein (Fragment) Sus scrofa	22.1	5	39.5	
A0A0A0MY58	Uncharacterized protein (Fragment) Sus scrofa	65.1	4	27	
H9GW43	Uncharacterized protein Bos taurus	55.2	8	41.1	
F6ZT60	Uncharacterized protein Equus caballus	11.6	7	34.4	
F1SJ34	Uncharacterized protein Sus scrofa	37.5	4	34.2	
F1RSP5	Uncharacterized protein Sus scrofa	28.2	8	34.1	
F1S8Y5	Uncharacterized protein Sus scrofa	32.6	7	27	
F1SAB6	Acyl carrier protein Sus scrofa	21.2	4	24.1	
A8E652	CYP2C19 protein Bos taurus	15.5	5	32	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
Q3SZV7	Hemopexin Bos taurus	17.6	5	21.5	
Q3T0P6	Phosphoglycerate kinase 1 Bos taurus	24.9	6	29.8	
P40185	Protein MMF1, mitochondrial Saccharomyces cerevisiae	44.8	4	29	
I3LN42	Uncharacterized protein (Fragment) Sus scrofa	32.1	8	41.8	
E1BHY6	Uncharacterized protein Bos taurus	7.5	2	40.8	
F1MP67	Uncharacterized protein Bos taurus	20	7	29.1	
F6S2F8	Uncharacterized protein Equus caballus	13.6	2	24.4	Met-loss+Acetyl [N-Term]
F1RX35	Uncharacterized protein Sus scrofa	14.2	5	21.7	
P02662	Alpha-S1-casein Bos taurus	19.2	3	22.4	
E1BC10	Amine oxidase Bos taurus	12.5	6	21.5	
P37111	Aminoacylase-1 Sus scrofa	27	5	28.6	
P00889	Citrate synthase, mitochondrial Sus scrofa	24.1	6	28.3	
Q9N0F1	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial Sus scrofa	8.4	3	26	
P23196	DNA-(apurinic or apyrimidinic site) lyase Bos taurus	5.7	2	29	
P25786-2	Isoform Long of Proteasome subunit alpha type-1 Homo sapiens	29.7	6	26.2	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
P18203	Peptidyl-prolyl cis-trans isomerase FKBP1A Bos taurus	72.2	5	28.1	Met-loss [N-Term]
Q7SIB7	Phosphoglycerate kinase 1 Sus scrofa	25.7	6	26	
B5KJG2	Phosphoglycerate mutase 2 Sus scrofa	36	6	31	
F1SB60	Superoxide dismutase Sus scrofa	40.1	6	32.4	
F6YN05	Uncharacterized protein (Fragment) Equus caballus	48.1	5	23.9	
E1BGJ0	Uncharacterized protein Bos taurus	1.7	5	23.3	
F6WPT4	Uncharacterized protein Equus caballus	27.8	4	27.1	
F1RQU2	Uncharacterized protein Sus scrofa GN=HSPCB PE=3 SV=1	12.2	7	25.5	
P39873	Brain ribonuclease Bos taurus	13.2	3	17.9	
F1MEQ3	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial Bos taurus	8.3	3	21.3	
Q2EN74	Epidermal fatty acid-binding protein Sus scrofa	54.1	6	27	Met-loss+Acetyl [N-Term]
P23005	Gamma-crystallin F Bos taurus	29.3	3	34.2	
Q76LV2	Heat shock protein HSP 90-alpha Bos taurus	11.7	7	25.7	
A1XQS3	Mitochondrial NDUFA4 Sus scrofa	47.6	5	17.7	
G5E589	Proteasome subunit beta type Bos taurus	35.3	6	27.6	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
G1T4X8	Proteasome subunit beta type Oryctolagus cuniculus	42.8	5	32.2	Acetyl [N-Term]
G3X6N3	Serotransferrin Bos taurus	12.4	6	21.5	
F1MVK1	Uncharacterized protein (Fragment) Bos taurus	4.7	5	17.4	
F7AKP1	Uncharacterized protein Equus caballus	28.8	2	28.5	
G1U9C0	Uncharacterized protein Oryctolagus cuniculus	9.3	3	13.1	Met-loss [N-Term]
G1T7R2	Uncharacterized protein Oryctolagus cuniculus	16.1	3	22.1	Acetyl [N-Term]
K7GN83	Uncharacterized protein Sus scrofa	20.1	5	17.3	
I3LNG8	Uncharacterized protein Sus scrofa	17.1	5	18.3	
I3LG67	Uncharacterized protein Sus scrofa	63.5	4	25.4	
P51122	Acidic leucine-rich nuclear phosphoprotein 32 family member A Bos taurus	18.5	4	22.7	
F1SFI7	Alpha-2-HS-glycoprotein (Fragment) Sus scrofa	18	3	37.3	
F1S6Q7	ATP synthase subunit delta, mitochondrial Sus scrofa	37.5	3	18.7	
B9ZSM8	CD90 protein Sus scrofa	37.3	4	27	
F1MW44	Coagulation factor XIII A chain Bos taurus	16.3	7	22.7	
I3LJA6	Fibrinogen alpha chain (Fragment) Sus scrofa	20.8	3	24.5	
F1MYN5	Fibulin-1 Bos taurus	17.1	5	26.4	
A0A096P6K3	Gamma-crystallin A Bos taurus	29.3	3	31	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
P02526	Gamma-crystallin B Bos taurus	46.3	4	26.1	
Q8SPS7	Haptoglobin Sus scrofa	27.7	6	21.5	
G3X6L4	Hepatoma-derived growth factor Bos taurus	22.6	3	35.2	
G1SUJ3	Histone H2A Oryctolagus cuniculus	23.3	3	17.3	
Q148I8	Keratin 31 Bos taurus	12.7	4	18.1	
F1STY1	NADH dehydrogenase [ubiquinone] 1 subunit C2 Sus scrofa	29.5	3	17.5	
F1MJS4	Phosphoglucomutase-1 Bos taurus	17.8	6	18.5	
G3V5Z7	Proteasome subunit alpha type Homo sapiens	24.2	5	16.9	
G1SHV9	Proteasome subunit beta type (Fragment) Oryctolagus cuniculus	23.9	3	18.1	
A0A0A6YYL6	Protein RPL17-C18orf32 Homo sapiens	6.6	2	21.2	
F6UQC4	Uncharacterized protein Equus caballus	2.2	4	18.3	
F1SLX2	Uncharacterized protein Sus scrofa	6.8	7	20.9	
I3L650	Uncharacterized protein Sus scrofa	2.3	2	11.7	
F1S8C6	Uncharacterized protein Sus scrofa	14.5	4	19.2	
F1SC51	Uncharacterized protein Sus scrofa	26.7	5	17.4	Met-loss+Acetyl [N-Term]
P80227	Acylamino-acid-releasing enzyme Bos taurus	12.6	5	26.7	
P39954	Adenosylhomocysteinase Saccharomyces cerevisiae	21.4	4	18.2	
P49419	Alpha-aminoadipic semialdehyde	14.7	4	20.5	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
	dehydrogenase Homo sapiens				
P13861	cAMP-dependent protein kinase type II-alpha regulatory subunit Homo sapiens	23.8	6	21.1	Met-loss+Acetyl [N-Term]
Q2TBQ1	Coagulation factor XIII, B polypeptide Bos taurus	7	2	18.4	
I3LER5	Cytochrome c oxidase subunit 4 isoform 1, mitochondrial Sus scrofa	26	4	12	
Q05639	Elongation factor 1-alpha 2 Homo sapiens	7.3	3	16.4	
P15090	Fatty acid-binding protein, adipocyte Homo sapiens	50.8	5	14.7	Met-loss+Acetyl [N-Term]
Q29545	Inhibitor of carbonic anhydrase Sus scrofa	6.3	2	29.9	
P05787-2	Isoform 2 of Keratin, type II cytoskeletal 8 Homo sapiens	11.5	5	18.8	
Q29550	Liver carboxylesterase Sus scrofa	12.9	5	14.4	
A6H7E3	PDZ and LIM domain 1 Bos taurus	16.8	4	16.6	
Q8NC51	Plasminogen activator inhibitor 1 RNA-binding protein Homo sapiens	7.4	2	21	Met-loss [N-Term]
F6TCI4	Proteasome subunit alpha type Equus caballus	32.9	4	25	
Q07505	Putative carboxymethylenebutenolidase Saccharomyces cerevisiae	14.7	3	20.4	
F1RGJ3	Stress-70 protein, mitochondrial Sus scrofa	10.9	5	18.8	
G5E513	Uncharacterized protein (Fragment) Bos taurus	8.8	3	15.8	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
F1MFF5	Uncharacterized protein (Fragment) Bos taurus	41.7	2	16.2	
F2Z558	Uncharacterized protein (Fragment) Sus scrofa	14.5	3	17.3	
F1MKB7	Uncharacterized protein Bos taurus	36.1	4	19.2	
F1MH40	Uncharacterized protein Bos taurus	24.6	4	17.4	
K7GQB8	Uncharacterized protein Sus scrofa	9.6	6	20.8	
F1SMZ6	Uncharacterized protein Sus scrofa	43.1	4	21.9	
F1S0J8	Uncharacterized protein Sus scrofa	7.2	3	16.3	
F1SU52	Uncharacterized protein Sus scrofa	29.4	6	21.1	
I3LJ05	Uncharacterized protein Sus scrofa	25.1	3	18.9	
F1SCD0	Uncharacterized protein Sus scrofa	12.1	4	20.4	
O97764	Zeta-crystallin Bos taurus	20.9	3	30.5	
Q8HZJ8	11 beta-hydroxysteroid dehydrogenase type 1 Bos taurus	21.2	5	11.4	
Q5EA79	Aldose 1-epimerase Bos taurus	15.5	3	5	
A7Z066	Canx protein Bos taurus	6.4	3	14.6	
Q2KIT5	Cysteine-rich with EGF-like domain protein 2 Bos taurus	8.8	2	19.2	
Q5S3G4	Cytochrome c oxidase subunit 5B, mitochondrial Sus scrofa	39.5	4	15.5	
A1XQT2	Cytochrome c oxidase subunit 6C Sus scrofa	32.9	3	15.4	Met-loss+Acetyl [N-Term]
Q5E9D5	Destrin Bos taurus	24.2	3	17.9	
F1MH65	Filensin Bos taurus	7.4	3	9.6	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
I3LPP1	Fumarate hydratase, mitochondrial Sus scrofa	18.9	4	28	
Q0VCM4	Glycogen phosphorylase, liver form Bos taurus	6.9	4	12.2	
P54149	Mitochondrial peptide methionine sulfoxide reductase Bos taurus	26.2	4	16.3	
A6QNL5	PDIA6 protein (Fragment) Bos taurus	6.8	3	22.7	
F6Z688	Proteasome subunit alpha type Equus caballus	24.5	4	15	
Q0III8	RNASET2 protein (Fragment) Bos taurus	16.8	3	18.2	
F6V9Q3	Uncharacterized protein (Fragment) Equus caballus	9.6	2	13	
F1SGG2	Uncharacterized protein Sus scrofa	10.4	5	8.4	
I3LFV4	Uncharacterized protein Sus scrofa	27.5	3	20.1	Met-loss+Acetyl [N-Term]
F1SC20	Alpha-1B-glycoprotein Sus scrofa	7.4	3	12.5	
F1RL80	Branched-chain-amino-acid aminotransferase Sus scrofa	11.5	3	14.5	
P00727	Cytosol aminopeptidase Bos taurus	11.8	4	16.2	
Q6UAQ8	Electron transfer flavoprotein subunit beta Sus scrofa	23.1	4	14.9	
F1MDK9	Gamma-crystallin D Bos taurus	20.1	2	18.1	
B7NZQ6	GDP dissociation inhibitor 1 (Predicted) Oryctolagus cuniculus	8.5	2	20.8	
P00338-3	Isoform 3 of L-lactate dehydrogenase A chain Homo sapiens	7.5	2	12.4	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
P55290-4	Isoform 4 of Cadherin-13 Homo sapiens	4.6	2	9	
F1RGE3	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 2 Sus scrofa	23.2	2	11.5	
F1MHB8	Nicotinate-nucleotide pyrophosphorylase [carboxylating] Bos taurus	13	3	14.7	
F1MU19	Proteasome activator complex subunit 2 (Fragment) Bos taurus	15	3	15	
F1ST02	Proteasome subunit beta type Sus scrofa	14	3	14.1	
Q3MHN0	Proteasome subunit beta type-6 Bos taurus	28	3	18.3	
G5E5T5	Uncharacterized protein (Fragment) Bos taurus	17	3	16	
F6U5R1	Uncharacterized protein (Fragment) Equus caballus	12.9	3	17.8	
I3LSM7	Uncharacterized protein (Fragment) Sus scrofa	9.9	3	19.6	
F1SJP6	Uncharacterized protein (Fragment) Sus scrofa	20	2	8.4	
F1MM14	Uncharacterized protein Bos taurus	6.4	2	9.6	
E1BBY7	Uncharacterized protein Bos taurus	5.5	3	17.1	
F7CKI7	Uncharacterized protein Equus caballus	1.8	3	13.7	
F6SRG1	Uncharacterized protein Equus caballus	43.5	3	17.4	
G1TIR7	Uncharacterized protein Oryctolagus cuniculus	36.1	2	15	
G1SZ14	Uncharacterized protein Oryctolagus cuniculus	13.8	4	13.6	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
K7GQV5	Uncharacterized protein Sus scrofa	19.8	2	12.5	
F1SRB9	Uncharacterized protein Sus scrofa	11.1	3	11.5	
P49753	Acyl-coenzyme A thioesterase 2, mitochondrial Homo sapiens	8.9	3	9	
Q3ZC42	Alcohol dehydrogenase class-3 Bos taurus	12.3	2	17	
P34955	Alpha-1-antiproteinase Bos taurus	7.5	3	9.7	
F7BMK4	Cytochrome c oxidase subunit 6A, mitochondrial Equus caballus	29.9	2	7.3	
P22411	Dipeptidyl peptidase 4 Sus scrofa	6	3	12	
I3L5W3	Ficolin-2 Sus scrofa	11.5	2	11.1	
F5H039	Gephyrin Homo sapiens	5.4	3	10.2	
P26641-2	Isoform 2 of Elongation factor 1-gamma Homo sapiens	6.4	2	12.4	
O43852-3	Isoform 3 of Calumenin Homo sapiens	5.9	2	10.7	
F6TYW4	Malate dehydrogenase Equus caballus	11.2	3	11.2	
G1TZS1	Myosin light chain kinase, smooth muscle Oryctolagus cuniculus	1.2	2	7.5	
A5PJV7	NGFR protein Bos taurus	6.3	2	13.6	
F1MBI1	Proteasome subunit beta type Bos taurus	11.6	2	15.1	
E5RJR5	S-phase kinase-associated protein 1 Homo sapiens	22.7	3	11.2	
F7CIC0	Uncharacterized protein (Fragment) Equus caballus	4.9	2	9.7	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
F1S6G0	Uncharacterized protein (Fragment) Sus scrofa	3.8	2	15.9	
F1STC5	Uncharacterized protein (Fragment) Sus scrofa	45.9	2	16.1	
F1MN00	Uncharacterized protein Bos taurus	7.4	3	10.1	
E1BDQ1	Uncharacterized protein Bos taurus	17.6	2	13.5	
E1B8K6	Uncharacterized protein Bos taurus	7.8	3	12.6	
F7DZE7	Uncharacterized protein Equus caballus	9.2	2	9.8	
G1TTR4	Uncharacterized protein Oryctolagus cuniculus	17.3	2	13.9	
G1T5V2	Uncharacterized protein Oryctolagus cuniculus	16.7	3	11.4	
F1S3J0	Uncharacterized protein Sus scrofa	17.3	4	11.2	Met-loss [N-Term]
F1RZR6	Uncharacterized protein Sus scrofa	25.4	2	15.6	
F1SID7	Uncharacterized protein Sus scrofa	2.8	2	9.4	
F1N5M2	Vitamin D-binding protein Bos taurus	12	3	16.5	
F1N690	Acetyltransferase component of pyruvate dehydrogenase complex Bos taurus	4.8	2	8.1	
P07107	Acyl-CoA-binding protein Bos taurus	27.6	2	6.2	Met-loss+Acetyl [N-Term]
F1RQQ7	Alpha-1,4 glucan phosphorylase (Fragment) Sus scrofa	5.8	3	10	
I3LIV7	Alpha-1-acid glycoprotein Sus scrofa	11.7	2	8	
P56560	Amine oxidase [flavin-containing] B Bos taurus	8.8	2	10.4	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
Q95339	ATP synthase subunit f, mitochondrial Sus scrofa	23.9	2	4.2	Met-loss+Acetyl [N-Term]
A1A4J1	ATP-dependent 6-phosphofructokinase, liver type Bos taurus	3.6	2	8.7	
A3KMY8	Beta-glucuronidase Bos taurus	4.8	2	8.4	
P07688	Cathepsin B Bos taurus	10.7	2	9	
P23805	Conglutinin B os taurus	8.6	2	9.1	
A7MBI5	DPYSL3 protein Bos taurus	5.3	2	11.2	
G1TKN7	Enhancer of rudimentary homolog Oryctolagus cuniculus	27.9	2	10	Met-loss+Acetyl [N-Term]
B7NZI5	Eukaryotic translation initiation factor 6 Oryctolagus cuniculus	18.9	2	9.8	
J3KQE5	GTP-binding nuclear protein Ran (Fragment) Homo sapiens	10.7	2	6.7	
P00348	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial Sus scrofa	10.2	2	7.7	
P35221-2	Isoform 2 of Catenin alpha-1 Homo sapiens	3.1	2	4.8	
O15173-2	Isoform 2 of Membrane-associated progesterone receptor component 2 Homo sapiens	10.9	2	7.3	
Q0VD56	Myosin binding protein C, cardiac Bos taurus	2.9	2	6.6	
F1SRG2	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6 Sus scrofa	20.3	2	8.9	
P39721	Protein AIM2 Saccharomyces cerevisiae	10.2	2	4.8	
P38616	Protein YGP1 Saccharomyces cerevisiae	13.8	2	10.1	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
F6ZEH8	Serotransferrin Equus caballus	2.8	2	6.6	
Q92783	Signal transducing adapter molecule 1 Homo sapiens	6.1	2	10	
A0A087WUZ3	Spectrin beta chain, non-erythrocytic 1 Homo sapiens	1.6	2	10.2	
H9GW42	Uncharacterized protein (Fragment) Bos taurus	5.7	2	7	
F1S1Q7	Uncharacterized protein (Fragment) Sus scrofa	13.1	2	10.6	
F1N169	Uncharacterized protein Bos taurus	1.1	2	7.1	
F1MZE0	Uncharacterized protein Bos taurus	6.1	2	9.8	
E1BCE2	Uncharacterized protein Bos taurus	6	3	6.7	
F1N2K1	Uncharacterized protein Bos taurus	8.7	2	10.1	
G1TTY1	Uncharacterized protein Oryctolagus cuniculus	19.8	2	10.3	
G1SST9	Uncharacterized protein Oryctolagus cuniculus	4.9	2	6.9	
G1TDQ2	Uncharacterized protein Oryctolagus cuniculus	13.7	2	7.7	
I3LP02	Uncharacterized protein Sus scrofa	8.2	2	6.2	
F1STR1	Uncharacterized protein Sus scrofa	13.2	3	9.7	
F1S1B9	Uncharacterized protein Sus scrofa	14.9	3	7.9	
F1S4Y6	Uncharacterized protein Sus scrofa	49	3	7.4	Met-loss+Acetyl [N-Term]
F1RUC0	Uncharacterized protein Sus scrofa	6.6	2	4.8	
F1SLR1	Uncharacterized protein Sus scrofa	17.4	2	10.9	

Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
F1S4P7	Uncharacterized protein Sus scrofa	9.3	3	6.5	
F1SLF0	Uncharacterized protein Sus scrofa	12.7	2	12	
I3LQF8	Uncharacterized protein Sus scrofa	9.5	2	10.1	
Q3ZC41	Acetyl-Coenzyme A acyltransferase 1 Bos taurus	7.5	2	5.9	
D0G0B3	Acetyl-Coenzyme A acyltransferase 2 Sus scrofa	11.6	2	5.6	
G1SRM4	Acyl-protein thioesterase 1 (Fragment) Oryctolagus cuniculus	16	2	5.4	
F1SD45	cAMP-dependent protein kinase catalytic subunit alpha Sus scrofa	6	2	4	
F1SGS9	Catalase Sus scrofa	4.7	2	6	
F1N455	Dipeptidyl peptidase 1 Bos taurus	7.3	2	5.4	
P19120	Heat shock cognate 71 kDa protein Bos taurus	4.8	2	3.6	
F1MN74	Isocitrate dehydrogenase [NAD] subunit, mitochondrial Bos taurus	6.8	2	5.8	
F1MMK9	Protein AMBP Bos taurus	6.8	2	3.3	
Q9GJT2	S-formylglutathione hydrolase Sus scrofa	20.6	2	5.7	
G5E5C8	Transaldolase Bos taurus	8.3	2	4.7	
I3LNB4	Uncharacterized protein (Fragment) Sus scrofa	6.6	2	5.8	
I3LDA5	Uncharacterized protein (Fragment) Sus scrofa	10	2	8.7	

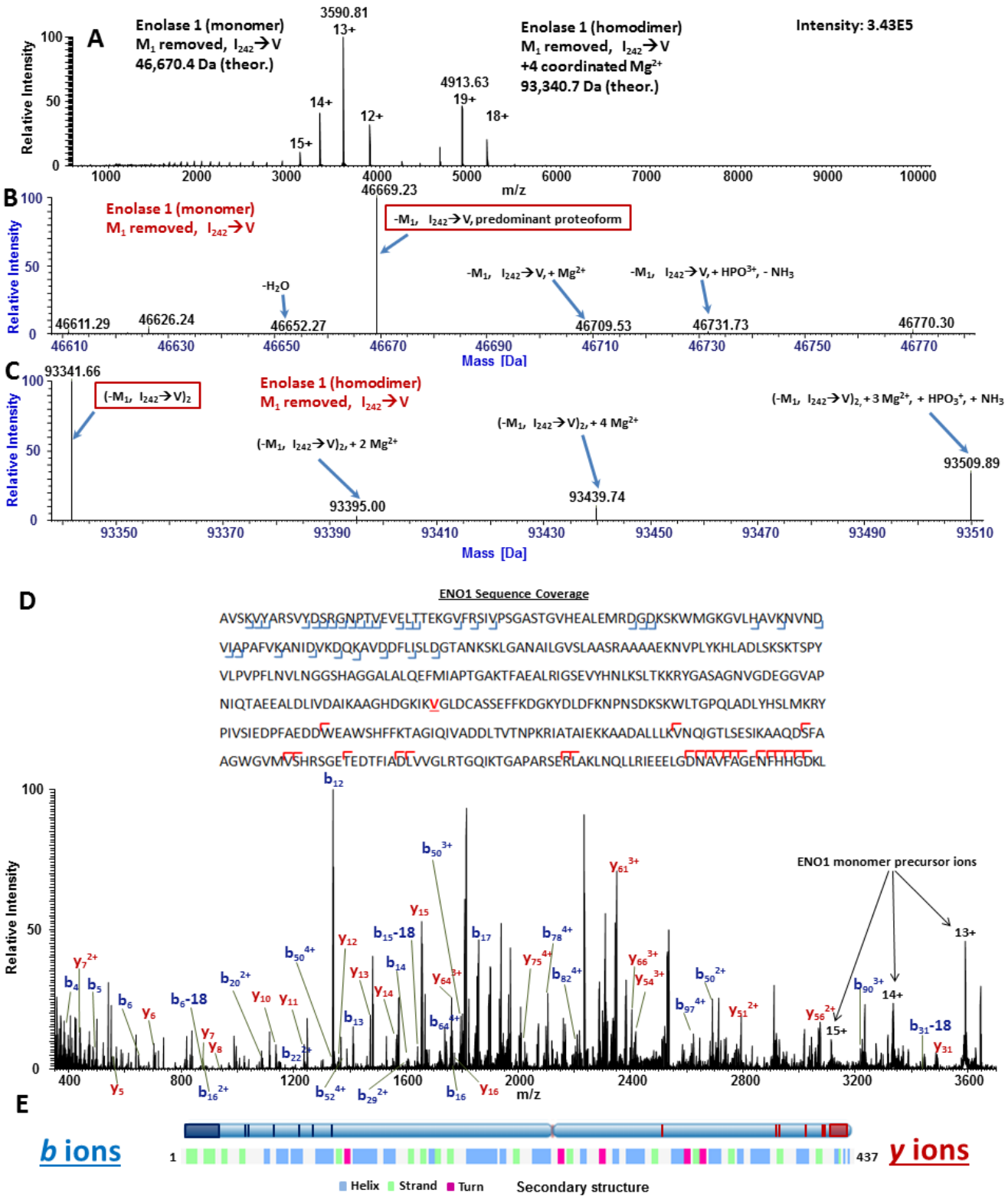
Accession	Description	Coverage %	# Peptides	Score Sequest	Mod(s)
F1RSG9	Uncharacterized protein (Fragment) Sus scrofa	11.1	2	4.6	
F1RMV5	Uncharacterized protein (Fragment)	30.2	2	4.6	
F1N076	Uncharacterized protein Bos taurus	3.4	2	6.5	
F6SJM8	Uncharacterized protein Equus caballus	4.6	2	4.7	
U3KMX4	Uncharacterized protein Oryctolagus cuniculus	9.7	2	1.8	
G1TY06	Uncharacterized protein Oryctolagus cuniculus	14.7	2	5.4	
G1U7Q6	Uncharacterized protein Oryctolagus cuniculus	15.8	2	5.1	
F1RHJ9	Uncharacterized protein Sus scrofa	6.3	2	5.4	
K7GP28	Uncharacterized protein Sus scrofa	7.8	2	6.2	
I3L5D0	Uncharacterized protein Sus scrofa	17.3	2	5.2	Met-loss+Acetyl [N-Term]
F1N0Y0	Uridine phosphorylase Bos taurus	13.9	2	5.2	

Table 4-2: Identification results for the mixture of 11 complex-forming protein standards using a conventional bottom-up nanoLC-MS analysis. From the number of identified species, heterogeneity of the mixture extended well beyond the original 11 protein standards.

Following the CZE-MS[1] experiments, all-ion fragmentation (AIF) scans of all detected precursors were conducted in the HCD cell of the Exactive Plus EMR. The AIF fragmentation resulted in the generation of HCD-characteristic *b*- and *y*-ions that validated the major components of the mixture identified by CZE-MS[1]. In spite of the sample complexity, separation, combined with accurate intact mass measurements and AIF, allowed the characterization of proteoforms and protein complexes of the mixture. The majority of observed molecular masses of proteins and protein complexes in this mixture (see below) were mapped to theoretically expected masses, based on protein sequences reported in UniProt within a 30 ppm

mass deviation (see **Table 4-1**). In a few cases, the mass deviation was slightly higher, which we attribute to incomplete desolvation or non-specific binding of adducts.

Using the ReSpect and Sliding Window algorithms of the Protein Deconvolution 4.0 software suite, identities of proteoforms of the ENO1 monomer and homodimer were determined (**Figure 4-3B, C**). Fragment ion *m/z* values were mapped using a tolerance of 20 ppm (FDR<1%). In **Figure 4-3B**, the major form of ENO1 is seen to be a previously reported proteoform with an N-terminal Met-loss and a likely Ile → Val substitution at position 242 of the canonical protein sequence (P00924) [325, 326]. However, confirmation of the exact location of this substitution was not possible *via* AIF-MS. It should be noted that the observed variants of ENO1 are consistent with a previously reported primary structure and known modifications [325, 326]. Minor ion species (below 5%) likely corresponded to other reported ENO1-associated structural features, such as phosphorylation, oxidation, and magnesium ion coordination [327] (**Table 4-1**). Additionally, products of incomplete ion desolvation and in-source decay were observed (**Figure 4-3B**). For the ENO1 homodimer, in addition to Met_1-truncation and the Ile_{242}→Val substitution, coordination of two magnesium ions per subunit was observed (**Figure 4-3C**), in agreement with the literature [328]. Confirmation of the identity of the Met-removed, Ile_{242} → Val variant as the most abundant ENO1 proteoform was achieved *via* mapping of the *m/z* values of fragment ions to the sequence with the above substitution and a fragment mass tolerance of 20 ppm (**Figure 4-3D, Figure 4-5**). However, localization of the exact substitution site was not achieved. Similar to previously-reported findings for other proteins studied under native conditions, as shown in **Figure 4-3E**, mapping of the most abundant *b*- and *y*- fragment ions onto the ENO1 secondary structure revealed by X-ray crystallography [329], expectedly demonstrated that the majority of identified fragments corresponded to N- and C-termini and turn loop regions, which are typically exposed to the surrounding aqueous environment and tend to contain polar and charged amino acid residues. **Figure 4-5** shows the CZE-AIF-MS spectrum for ENO1 homodimer. As expected, the AIF fragmentation patterns for ENO1 homodimer and monomer were very similar, likely because the majority of fragment ions originated from the monomer that was released from the homodimer in the process of HCD activation.



Figures 4-3 A-E. **Native CZE-MS characterization of enolase 1.** Averaged spectra across the entire migration time range corresponding to ENO1 electrophoretic peak resulted in the detection of both ENO1 monomer and homodimer (**A**). Deconvolution of the CZE-MS signal acquired

during of the migration of the electrophoretic peak corresponding to ENO1 monomer resulted in the identification of major proteoforms of ENO1 (**B**). Deconvolution of the CZE-MS signal acquired during the migration time range corresponding mostly to ENO1 homodimer resulted in the identification of proteoforms of the dimer (**C**). Averaged AIF spectra of the entire migration time range corresponding to the ENO1 electrophoretic peak resulted in the generation of *b*- and *y*- ions characteristic to ENO1 (**D**). Fragmentation loci were mapped as blue (*b*-ions) and red (*y*-ions) vertical marks and shaded boxes on the secondary structure of ENO1, revealing that for the native globular protein ENO1, HCD fragmentation occurred predominantly on the N- and C-termini and turn loop regions (not colored on the lower bar representing the secondary structure of ENO1 (UniProt) (**E**).

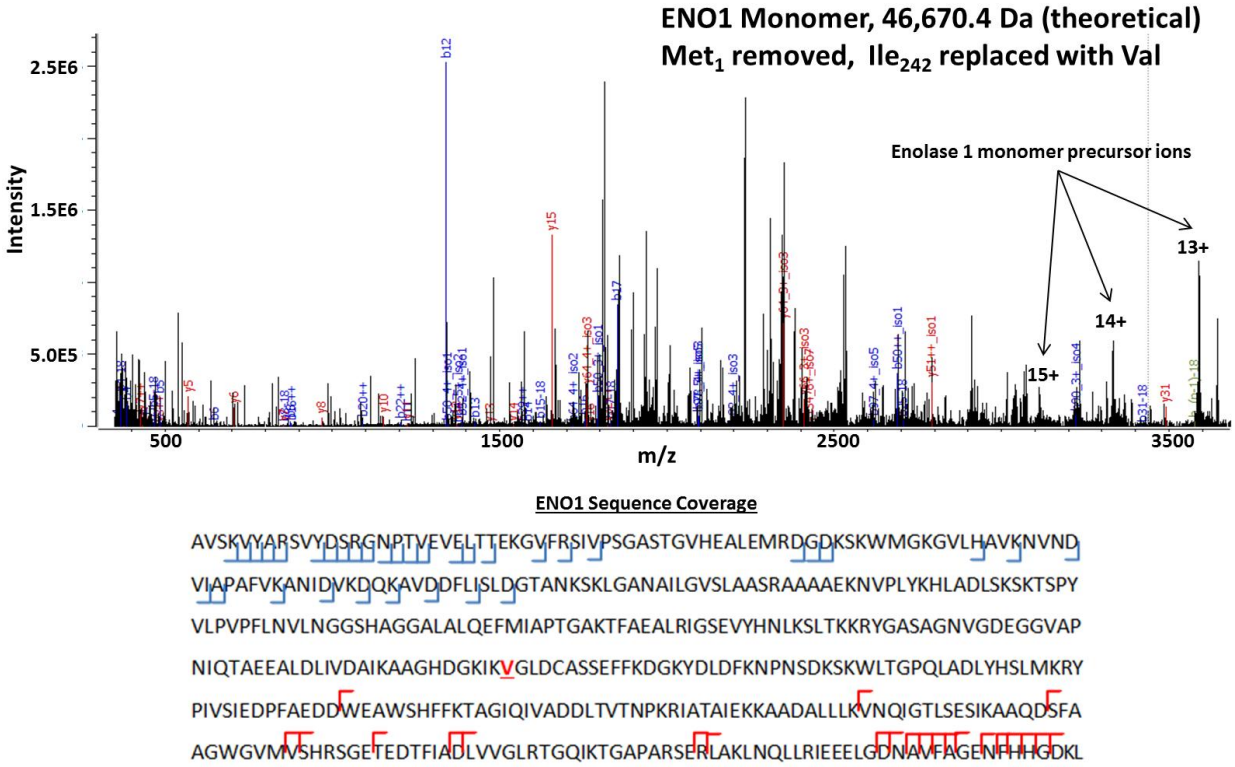


Figure 4-4: Fragmentation of ENO1 homodimers. AIF fragmentation of the ENO1 monomer peak resulted in the generation of sequence-characteristic *b*- and *y*-ions, shown in the top panel. The bottom panel shows fragment locations relative to the primary structure of ENO1.

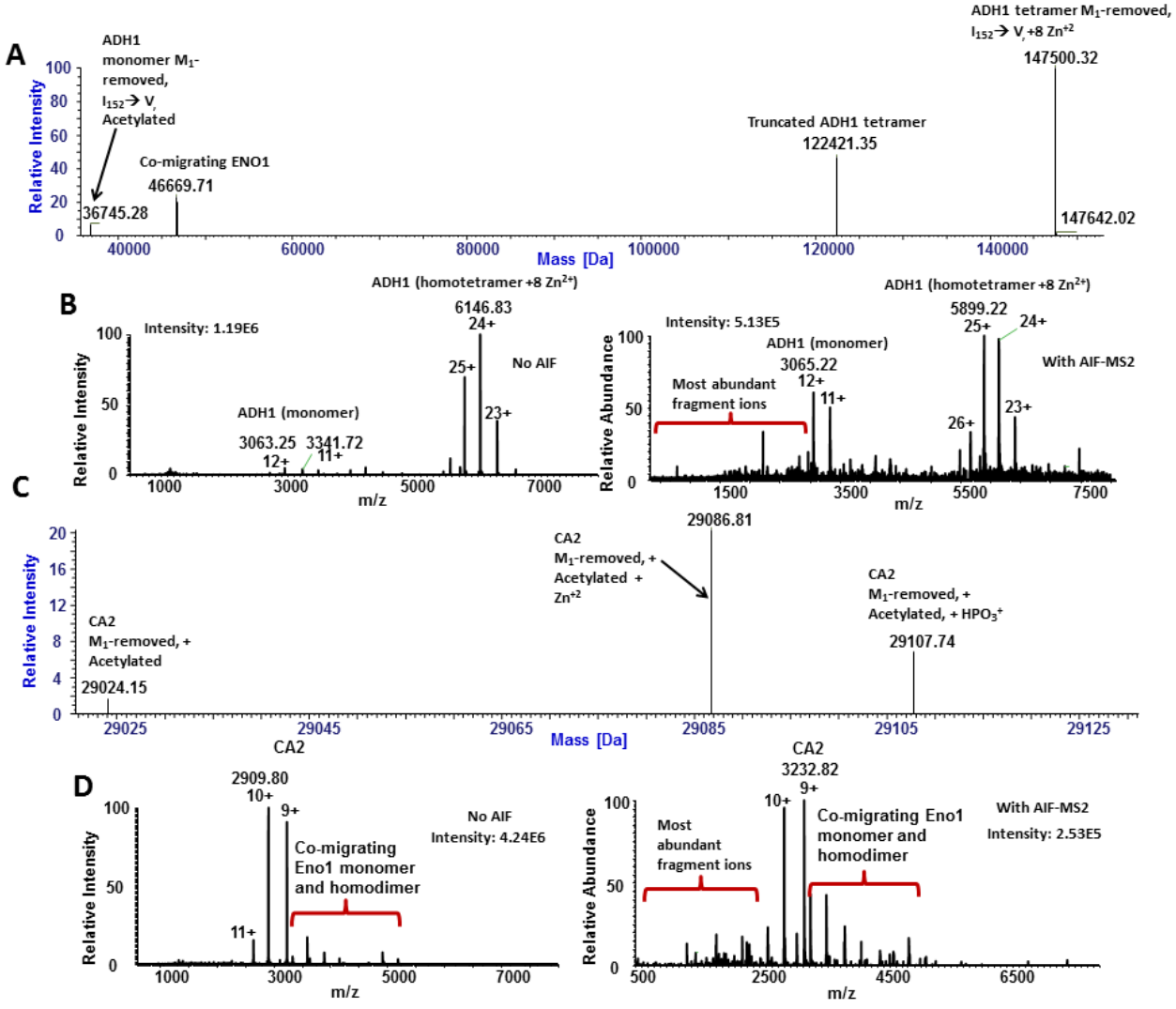
The combination of native CZE separation with MS^1 and AIF was also important for the characterization of ADH1 (**Figure 4-2**). Deconvolution of the ADH1 peak resulted in the detection of the ADH1 monomer with a MW of 36,745.3 Da, which corresponds to the N-terminal methionine truncated form, and the more prevalent homotetramer (12-fold higher) with a MW of 147,500.3 Da, which corresponds to the N-terminal methionine loss on each subunit

and a total of eight coordinated Zn^{2+} ions (two Zn^{2+} ions per subunit are required for appropriate folding and functional activity of ADH1) (**Figure 4-5A**) [330]. Additionally, a likely truncated form of the ADH1 tetramer (122,421.4 Da) was observed in the mixture, as well as the ADH1 standard. The experimental mass of the monomer closely matched a previously reported ADH1 variant (P00330) with Met_1-truncation and acetylation at the N-terminal Ser residue, along with a single substitution of Ile with Val at position 152 or 338 [330, 331]. This match resulted in a mass difference of 9 ppm between the experimentally determined and calculated average mass values (**Table 4-1**). Similar to enolase, AIF of ADH1 led to the validation of the major forms of ADH1 determined from intact mass measurements (**Figure 4-5B** and **Figure 4-6A**).

A previous study of ADH1 tetramers by direct infusion native MS on a Fourier transform ion cyclotron resonance (FTICR)-MS with electron capture dissociation (ECD) reported on the fragmentation of residues exclusively at the N-terminus of the protein [296]. The authors correlated this fragmentation pattern with the X-ray crystal structure of the ADH tetramer that showed that the N terminus of the protein is exposed and available for fragmentation, while the C terminus is hidden and involved in the interface of the subunits. Fragmentation coverage was also limited as a result of the formation of charge-reduced ADH1 tetramer species by ECD, resulting in no dissociation of ADH1 subunits from the tetrameric complex [296]. In our study, in addition to N-terminal fragments we also observed C-terminal fragment ions (**Figure 4-6A**). A different mechanism of fragmentation (HCD AIF accompanied with in-source CID) was used in our experiments, in addition to a different mass analyzer type (Orbitrap) and MS instrument, which could be responsible for these differences in fragmentation patterns. Also, in our study, the combination of in-source CID used for desolvation and AIF possibly promoted partial dissociation of ADH1 tetramers into monomers, which made both protein termini accessible for fragmentation. In a different study, a variety of fragmentation techniques, including collisionally activated dissociation (CAD), infrared multiphoton dissociation (IRMPD), in-source dissociation, and ECD were used for the top-down characterization of native ADH using an FTICR-MS instrument [332]. CAD and, especially IRMPD, resulted in the significant generation of sequence-informative N- and C-terminal fragment ions. These results, and particularly CAD results, are similar to those obtained in our experiments using HCD AIF accompanied with in-source CID. However, expectedly, sequence coverage was considerably more comprehensive using CAD and IRMPD in [332] than our approach, involving HCD AIF. In our study, similar to the previous reports [296, 332], the sequence origin of the detected fragment ions can be correlated to the outer surface regions of the complex and protein termini.

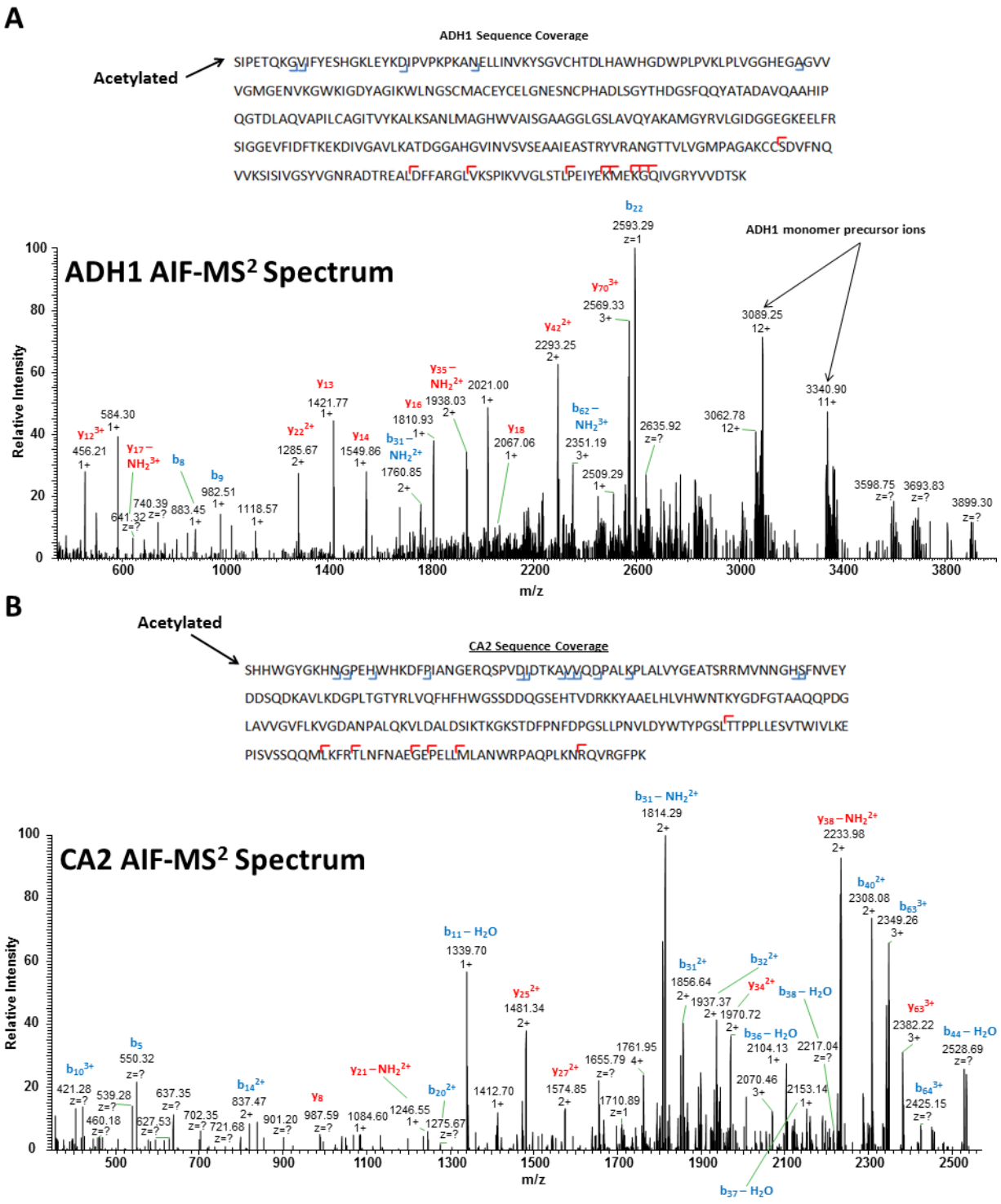
The characterization of proteoforms and complexes of CA2 (P00921, theoretical MW: 28,982.6 Da) was achieved; deconvolution revealed three major proteoforms of CA2 (**Figure 4-5C**). The most abundant species (experimental MW: 29,086.8 Da) closely matched the proteoform with a loss of the N-terminal methionine, N-terminal acetylation, and a coordinated Zn^{2+} ion (calculated MW: 29,089.7 Da). Both Met_1-truncation and coordination of Zn^{2+}-ions by three histidyl residues of CA2 were previously reported [333]. The second most abundant form was tentatively identified to contain N-terminal Met-loss, N-terminal acetylation, and

phosphorylation (experimental MW: 29,107.7 Da) (**Table 4-1**). The third detected form of CA2 was determined to possess N-terminal Met-loss and acetylation (experimental MW: 29,024.2 Da), with an abundance of less than 4% relative to the Zn^{2+}-coordinated species (**Figure 4-5C, Table 4-1**). From CZE-MS utilizing AIF, proteoforms of CA2 were fragmented into corresponding *b*- and *y*-ions (**Figure 4-5D**). Analysis of fragment ion spectra resulted in the confirmation of the identity of the most abundant proteoform (**Figure 4-6B**). Coordination of one Zn^{2+} ion by monomeric CA2 under native conditions and the primary structure of the most abundant form of CA2 – confirmed by the AIF MS – fragmentation pattern are also in good agreement with the crystal structure of this protein and previous reports [333].



Figures 4-5 A-D. Native CZE-MS analysis of alcohol dehydrogenase 1 and carbonic anhydrase 2. Deconvolution of the CZE-MS^1 data in the electrophoregram region corresponding to ADH1 (17.5-22 min) resulted in the detection of peaks corresponding to monomer, tetramer,

and truncated tetramer forms of ADH1 (**A**). MS fragmentation using HCD-AIF with a collision energy of 150 eV resulted in the fragmentation of ADH1 species, producing sequence-characteristic *b*- and *y*-ion fragments (**B**). Deconvolution of the electrophoretic peak corresponding to the CA2 (15.5-16.0 min) resulted in the identification of major proteoforms of CA2 (**C**). Fragmentation of the CA2 peak was achieved via HCD-AIF with collision energy of 150 eV, resulting in the generation of sequence-characteristic *b*- and *y*-ions (**D**).



Figures 4-6 A-B. CZE-MS with AIF confirmed identities of major mixture components of ADH1 and CA2. Annotated CZE-MS-AIF spectra are shown for ADH1 (**A**) and CA2 (**B**). For both proteins, AIF spectra were integrated across electrophoretic peaks corresponding to all associated proteoforms.

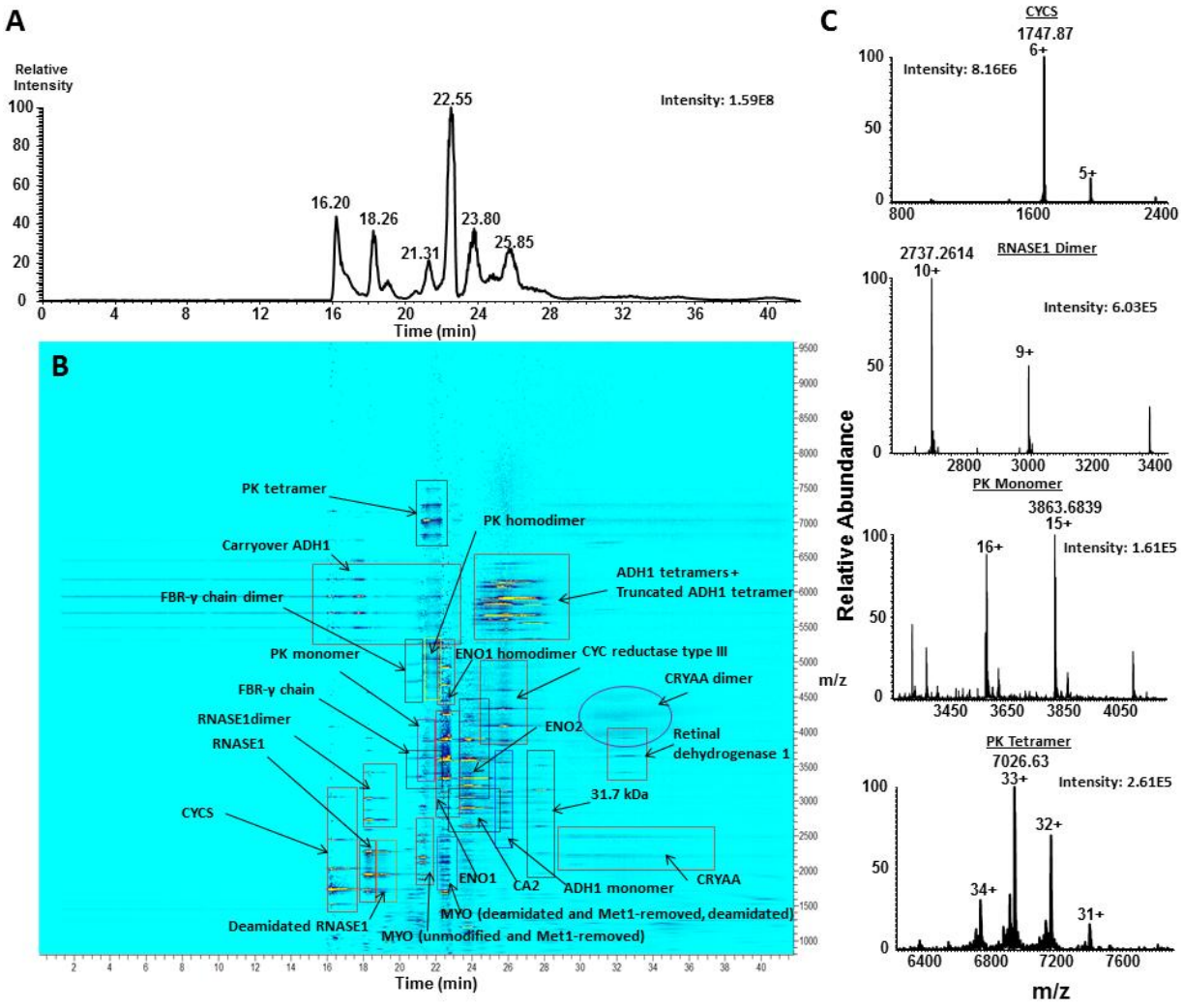
Following the characterization of a simple mixture of three proteins, a more complex mixture of eleven protein standards known to form non-covalent multisubunit protein complexes and complexes with coordinated metal ions was evaluated using the sheathless CZE-MS system. However, first, the components were analyzed separately using a conventional bottom-up nanoLC-MS^2 approach (FDR<1%). The sample heterogeneity extended well beyond the 11 commercial standards used to prepare the mixture, considering the number of contaminant proteins detected in the bottom-up approach (**Table 4-2**). Sequences of these proteins were assembled into a protein database to facilitate identification of proteins and protein complexes (including identification of predominant proteoforms and sequence variants) in subsequent native CZE-MS^1 experiments.

The BFS capillary demonstrated high separation performance in the analysis of the three protein mixture under native conditions because the choice of ammonium acetate, pH 7.0-8.5 as the BGE effectively suppressed adsorption of these specific proteins on the capillary surface. However, in CZE separations of intact proteins and protein complexes, the performance of BFS capillaries is often suboptimal due to the large number of surface charges – both on the protein and on the inner capillary surface – thereby leading to non-specific adsorption to the capillary wall and hence, potentially resulting in peak tailing, lower separation efficiency, and carryover issues [67, 334-336]. Thus, the separation performance of the BFS capillary was found inadequate for the mixture of 11 proteins and associated complexes as well as for higher complexity samples and high molecular mass proteins (e.g. mAbs). Presumably, ionic strength and salt concentration of the ammonium acetate BGE used in these experiments was insufficient for effectively preventing adsorption for a number of proteins of the mixture. Therefore, a recently-commercialized capillary with an uncharged polyacrylamide coating preventing protein adsorption [67] was used for analysis of a higher complexity samples.

The components of the mixture of 11 proteins and associated complexes migrated between 15 and 35 minutes and, as expected, many of them were detected in native CZE-MS but not in direct infusion analyses (**Figure 4-7, Figure 4-8**). Successful characterization of the mixture of 11 proteins and other higher complexity samples *via* direct infusion native MS was precluded by considerable overlapping of *m/z*'s of different proteins, resulting in difficult to interpret, less informative spectra. Moreover, in direct infusion native MS mode, difference in abundances and ionization efficiencies between individual analytes resulted in considerable signal suppression of lower abundance and lower ionization efficiency species.

Accurate mass measurements from the native CZE-MS^1 run were used for the identification of major forms of individual proteins. A higher molarity BGE (40 mM, relative to the 25 mM ammonium acetate we used previously), was utilized in these experiments to improve the separation and more-closely mimic physiological conditions, thereby improving the stability of non-covalent protein complexes in the mixture during the analysis. It should be noted that the

molarity of the ammonium acetate BGE used in this study was limited by the performance and the long-term stability of the porous sheathless emitter. The 40 mM ammonium acetate BGE did not result in any changes in spray stability and the lifetime of the emitter. Separation efficiency exceeding 150,000 theoretical plates per meter were readily observed for selected components of the mixture using the employed native conditions.



Figures 4-7 A-C. Characterization of the complex mixture on the sheathless cPA capillary-Exactive EMR system. The mixture of 11 complex-forming proteins was separated and characterized on the sheathless polyacrylamide-coated – Exactive EMR system and resulted in the detection of all 11 protein standards in their monomeric forms, in addition to multisubunit complexes for PK, ADH1, RNASE1, ENO1, CRYAA, and truncated ADH1 as well as contaminant proteins and truncated species. Panel **(A)** represents the total ion electropherogram for the separation, **(B)** represents the annotated ion density map, showing migration times (horizontal axis) of detected proteins vs. their corresponding *m/z* values (vertical axis), and (**C**)

represents native MS spectra for CYCS monomer, RNASE1 homodimer, PK monomer and PK homotetramer detected on the Exactive Plus EMR.

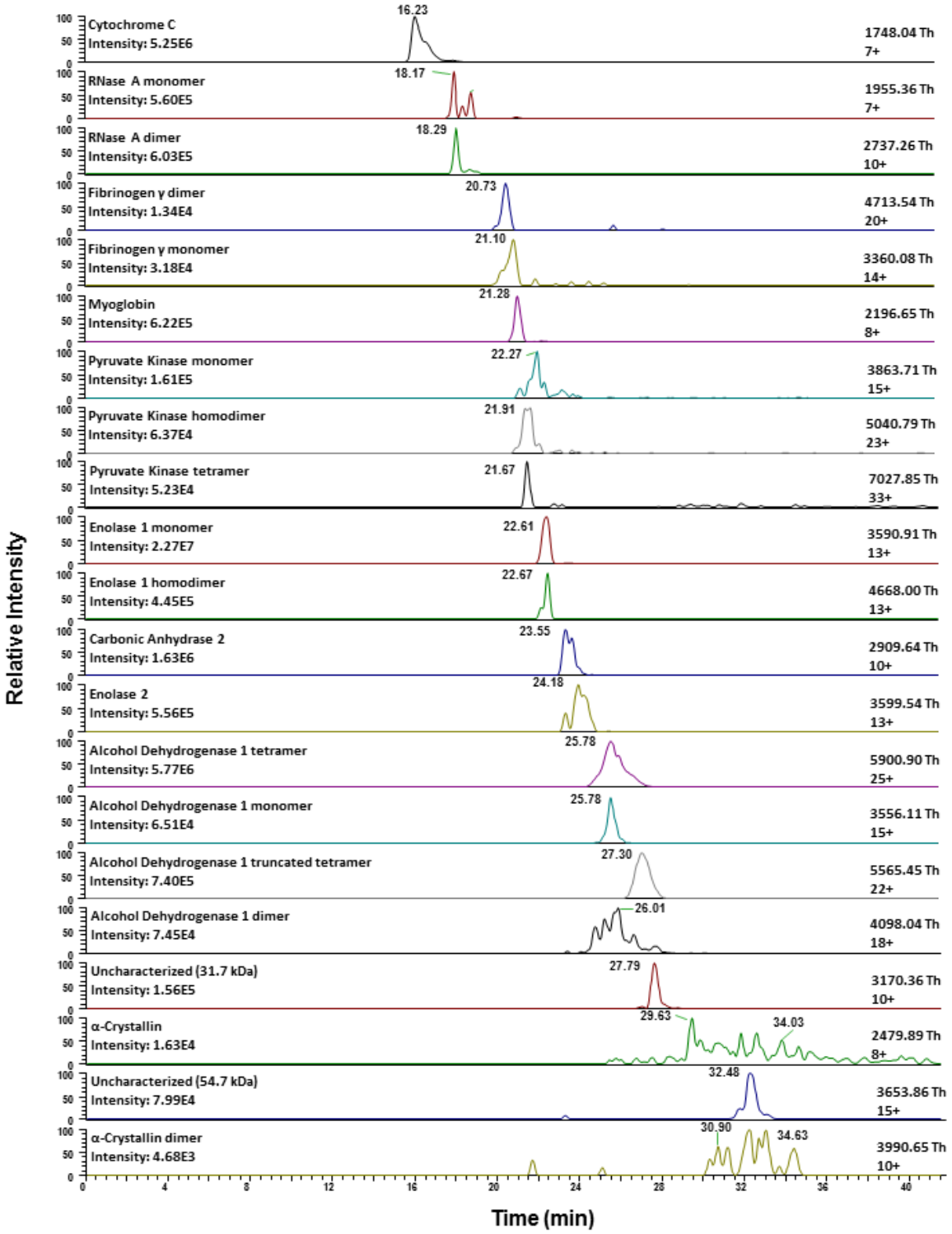


Figure 4-8: **Extracted ion electropherograms of components of the mixture of 11 proteins and associated complexes.** Identification of major mixture components was achieved *via*

deconvolution of the CZE-MS^1 experiment performed under native conditions. Extracted ion electropherograms of the most abundant charge states of major species are shown, along with corresponding signal intensities. Two uncharacterized contaminant proteins are also shown, corresponding to 31.7 kDa and 54.7 kDa, respectively.

Detection of large, multimeric protein complexes, such as ADH1 (~148 kDa) and pyruvate kinase (PK) (~231 kDa), was enabled by the Exactive Plus EMR. The enhanced gas flow into the HCD cell of the instrument allowed for collisional cooling and trapping of large proteins and complexes. The combination of advanced CZE- and MS-based technologies enabled the separation of individual proteins and detection of corresponding complexes (**Figure 4-7B** and **Figure 4-8**). Baseline CZE separation was achieved between isozymes of enolase – ENO1 and ENO2; however, ENO1 monomer and homodimer were not fully resolved. Similarly, ADH1 monomer, homodimer and homotetramer were not well-resolved electrophoretically, although the corresponding peaks were of different shapes and demonstrated noticeable offsets in migration times, which in combination with MS detection made the corresponding CZE-MS^1 migration patterns easy to distinguish. Also, the ADH1 monomer peak was considerably sharper relative to the experiments conducted on the BFS capillary. The peak intensity for the tetramer was approximately 90 and 80 fold higher than for the monomer and the dimer, respectively, which reflects the abundance ratios of these forms in solution. Examination of the extracted ion electropherograms corresponding to the most abundant charge states of these forms indicates small levels of in-source dissociation of the ADH1 complexes (the tetramer into the dimer and both the tetramer and the dimer into the monomer). Monomer (~58.0 kDa), homodimer (~116 kDa) and homotetramer (~232 kDa) PK species were detected as peaks of distinct shapes and with visible offsets in migration time of their peak maxima. In the case of the PK species, the monomeric species showed signal intensity approximately 2-3-fold higher relative to the dimer and tetramer peaks. Low levels of in-source dissociation of the multisubunit PK complex were also evident from the comparative examination of the extracted ion electropherograms for all three detected PK species. Other detected multisubunit complexes included RNASE1 dimer, alpha-crystallin (CRYAA) dimer, and fibrinogen-γ (FBR-γ) dimer. Baseline resolution was achieved between unmodified and deamidated myoglobin (MYO), as well as Met_1-removed and the deamidated form of Met_1-removed MYO. Similarly, deamidated RNASE1 was separated from the unmodified proteoform (**Figure 4-7, Figure 4-8**).The CZE separation allowed us to distinguish and reliably identify these near-isobaric species with a mass difference of close to 1 Da, and showed the potential for relative quantitation of such proteoforms and complexes, which would be considerably more challenging in direct infusion experiments.

4.3.3. Native CZE-MS of a Monoclonal Antibody

A recombinant, humanized monoclonal anti-Her2 IgG1 antibody, Trastuzumab, was examined under native conditions as a means to assess the heterogeneity of a biopharmaceutical protein and determine the potential presence of aggregates. To our knowledge, this experiment represents the first attempt of the characterization of an intact mAb using online native CZE-MS. A neutral polyacrylamide-coated capillary was coupled to the Exactive EMR Plus mass spectrometer and used for this experiment. **Figure 4-9A** shows the separation of the antibody sample components in a BGE of 20 mM ammonium acetate pH 8.0. The total ion electropherogram is shown in the top panel, followed by extracted ion electropherograms of major identified species. Predominant forms of the intact mAb, corresponding to 2X-glycosylated, 1X-glycosylated (hemi-glycosylated), and dimeric structures co-migrated at 24.4 min. Molecular species of approximately 101 kDa, presumably corresponding to the mAb with a loss of two light chains, were detected at 26.62 min and separated from the major forms of Trastuzumab. The corresponding dissociated light chains were detected as monomeric forms at 24.41 min and as dimeric forms at 22.83 min (**Figure 4-9A**). This supports previously reported observations that in some mAbs, individual chains are held together by non-covalent interactions instead of disulfide bonds [276, 286]. As a result, these chains may dissociate, especially if exposed to conditions other than the formulation buffer. Since these structures were separated by CZE, the dissociation of chains did not occur inside the mass spectrometer.

The spectrum integrated across the main electrophoretic peak (24.38 min) revealed the presence of the monomer and a trace (<1%) of a dimeric species (**Figure 4-9B, C**). As stated above, the mAb depleted of two light chains was also detected. As an illustration of this component, a zoomed-in view of the 19+ charge state of these forms is shown **Figure 4-9D**. A zoomed-in spectrum of the 24+ charge state of the main forms is shown in **Figure 4-9E**, revealing 1X- and 2X-glycosylated structures of the intact mAb. For the 2X-glycosylated forms, deconvolution of the MS signal from the entire electrophoretic migration window of the mAb sample constituents resulted in the identification of several biantennary glycoforms (**Figure 4-9F**). In addition, several dimers (assumed to be non-covalently associated) were also observed (**Figure 4-9G**). It is important to note that dimers of Trastuzumab could not be observed in denaturing CZE-MS (or LC-MS) experiments conducted by us and others [337].

The most prominent charge states, 24+ and 36+, were observed for the monomer and the dimeric (aggregated) mAb species, respectively (**Figures 4-9B, C**). Deconvolution of native CZE-MS^1 of Trastuzumab was performed using ReSpect with the Sliding Window algorithm. Depending on the target mass range (e.g. 100 – 300 kDa, 140 – 150 kDa, 280 – 300 kDa) and *m/z* range (e.g. 5,000 – 9,000 Th, 7,500 – 9,000 Th, 5,000 – 6,500 Th), the intensities of the resultant peaks corresponding to Trastuzumab dimers were ~140 - 250 fold lower than the intensity of the predominant peak corresponding to the monomer. Relative quantitation of the monomer and the aggregate can therefore be achieved using the native CZE-MS approach.

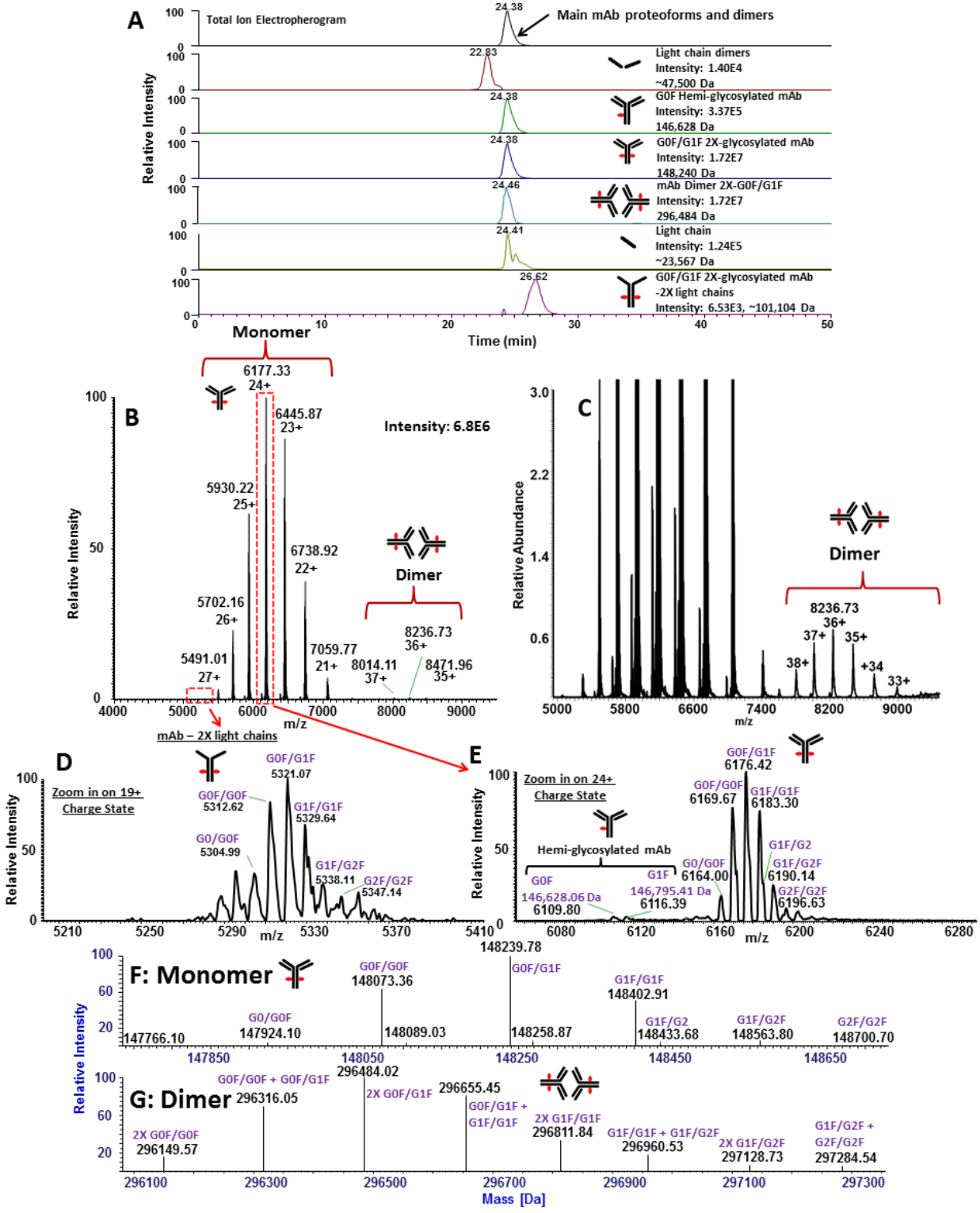
The average experimental molecular mass of the predominant form of the antibody is 148,073.36 Da (**Figure 4-9F**), which is a close match (15.81 Da greater) to the theoretical molecular mass of 148,057.55 Da calculated from the sequence reported in reference [338] with two G0F moieties and 16 disulfide bonds that corresponds to the molecular formula $C_{6560}H_{10132}N_{1728}O_{2090}S_{44}$. The difference between the theoretical and experimental molecular masses of the dimer was 34.48 Da, possibly due to strong coordination of one water molecule or replacement of a proton with an ammonium ion per monomer. It is worth noting that DrugBank (www.drugbank.ca/drugs/DB00072) has a different sequence for the heavy chain (HC) of the innovator version of Trastuzumab (Genentech), including DELTK for residues 360-364 of the HC instead of the EEMTK, and an extra P in the sequence with VEPPK instead of VEPK at residues 218-222 of the HC, as also reported by others [338, 339]. Additionally, the absence of the C-terminal K is yet another difference between the DrugBank sequence and the one reported in [338], which also agrees with our experimental findings.

From deconvolution of the MS^1 spectrum corresponding to the monomer, several combinations of biantennary glycans in the F_C region of the monomer were detected under native CZE-MS experiments (**Figure 4-9G**). These structures represent glycosylation at both $F_C/2$ domains corresponding to the expected G0/G0F, G0F/G0F, G0F/G1F, G1F/G1F, G1F/G2, G1F/G2F, and G2F/G2F forms. The presence of asymmetric glycosylation, characterized by different glycan compositions for both $F_C/2$ domains in monoclonal antibodies, has also been reported previously for mAbs, including Trastuzumab [337, 340]. Deconvolution of the MS^1 spectra of dimeric Trastuzumab, presented in **Figure 4-9G,** shows the composition of dimeric proteoforms. In addition to homodimers, characterized by symmetric Trastuzumab monomers (2X G0F/G0F, 2X G0F/G1F, 2X G1F/G1F, 2X G1F/G2F), heterodimers characterized by asymmetric monomers (G0F/G0F + G0F/G1F, G0F/G1F + G1F/G1F, G1F/G1F + G1F/G2F, G1F/G2F + G2F/G2F) were also detected (**Figure 4-9G**). Interestingly, dimers were determined to be combinations of 2X-glycosylated Trastuzumab forms, with no dimeric species containing partially deglycosylated subunits (i.e. 1X-glycosylated).

While identifications of the mAb components presented above could be also achieved by direct infusion native MS conducted using the same capillary and the CESI-MS platform at the identical flow rate and MS conditions, we expect that the CZE separation alleviates the ion suppression effects that are still present in the direct infusion mode experiments even at such low flow rates as 20-40 nL/min and improves the detection (and potentially relative quantitation) of minor sample components when combined with native MS.

It should be noted that a variety of other approaches exist for the identification of protein aggregation, including SEC coupled to multi-angle light scattering (MALS) [341] and NMR [342]. However, for SEC-MALS, detailed characterization of subunit composition may be not achieved. For example, individual glycosylation states of Trastuzumab dimers, as identified in this study, could not be characterized by SEC-MALS. For current NMR methodologies, while smaller protein dimers may be identified, mAbs such as Trastuzumab are too large for their

structure to be resolved with these approaches [343]. Thus, techniques such as native CZE-MS, enabling characterization of protein sequence and identification of higher order structure, can potentially provide important structural information. The ability to detect non-covalent assemblies of mAb molecules can be used to explore the pathways of aggregation (native and non-native) [344, 345] and assess the extent of protein aggregation, which is precluded by most other separation techniques that could not be performed online with native MS. The native CZE-MS approach can be potentially further extended to the relative quantitative or semi-quantitative analysis of aggregates in mAb-based immunotherapeutics, antibody-drug conjugates or other biopharmaceutical proteins, as well as different types of protein complexes in heterogeneous biological samples.



Figures 4-9 A-G. Native CZE-MS analysis of Trastuzumab. The total ion electropherogram is shown in the top panel, as well as identified components of the mAb represented by extracted ion electropherograms. An averaged spectrum across the entire electrophoretic peak resulted in the identification of charge states corresponding to monomeric and dimeric forms of Trastuzumab

(**B**). A zoomed-in spectrum revealed the *m/z* distribution of Trastuzumab dimers (**C**). A zoom-in of the 19+ charge state of species corresponding to the mAb with depleted two light chains is shown (**D**). A zoomed-in view of the 24+ charge state shows 2X-glycosylated and 1X-glycosylated (hemi-glycosylated) mAb structures (**E**). Deconvolution of Trastuzumab monomer peaks resulted in the identification of major glycoforms of the protein (**F**). Deconvolution also resulted in the characterization of dimeric forms, corresponding to combinations of glycoforms of Trastuzumab (**G**).

4.3.4. Native CZE-MS of the *E. coli* Ribosomal Extract

In this study, we also assessed the capabilities of native CZE-MS in the characterization of biological samples and proteomes of low complexity, using the protein isolate from the *E. coli* ribosomes. Ribosomes are non-membranous organelles that synthesize proteins from mRNA. These complex molecular machines are typically composed of over 50 distinct proteins with one to three rRNA molecules, depending on the organism. Ribosomes have traditionally been viewed as complex reproducibly-assembled ribozymes of a virtually fixed protein composition, whose function is to catalyze protein synthesis with a limited to non-existent ability to regulate this process [346]. Recent studies have revealed that ribosomes may be highly heterogeneous in their composition among various cell and tissue types [346-348]. These studies also indicated the possibility of variable protein-protein interactions within ribosomes that may alter their functional activities and specializations [346-350]. Detailed proteomic profiling of ribosomes and analysis of protein-protein interactions are necessary for elucidating possible mechanisms of specialized ribosomal activities. Native infusion MS analysis of the non-lysed ribosome and ribosomal subunits was recently demonstrated [318]. To our knowledge, however, no MS-based studies of ribosomal proteins performed with online separations under native conditions have been reported. In this work, we evaluated the capabilities of sheathless CZE with a commercial polyacrylamide-coated capillary coupled to a Q Exactive Plus mass spectrometer for characterization of individual intact proteoforms and non-covalent assemblies from a model sample of the *E. coli* ribosome protein extract under native conditions (**Figure 4-10 A-B**). Data dependent acquisition (Top N) was used to generate MS/MS spectra. The results were then compared to those of CZE-MS analysis performed under denaturing conditions (**Figures 4-10 C-D**).

The compiled sequence database of ribosomal proteins (RPs), in combination with native CZE-tandem MS, allowed the identification of 42 ribosomal protein groups, 603 proteoform-spectrum matches (PfSMs), and 137 proteoforms in a single experiment at <1% FDR (**Tables 4-3, 4-4**). An additional six RP identifications, corresponding to 48 in total, were made from deconvolution of native CZE-MS^1 data and manual matching of experimental and theoretical molecular masses (**Table 4-5**). The combination of native CZE-MS^1 and CZE-MS^2 resulted in identification of close to 90% of the expected *E. coli* RPs. The migration order of RPs was

closely-related to hydrodynamic volume, as smaller proteins (e.g. L36, L33) were detected first and larger proteins (e.g. S8, S9) last (**Figures 4-11 A-D**, **Figure 4-12**).

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
L10(3xL7)	54784.22	107710294.00	6	23.920-24.070	Native CZE-MS
	53933.78	11441232.25	5	30.190-30.760	Native CZE-MS
	53380.82	17469987.00	6	24.680-24.950	Native CZE-MS
	52655.68	13759078.69	5	23.860-24.000	Native CZE-MS
	51775.75	30250682.75	6	30.310-30.880	Native CZE-MS
	49708.77	37296832.75	6	24.070-24.910	Native CZE-MS
	49661.92	63525851.31	6	23.860-24.640	Native CZE-MS
	49518.41	19600258.84	7	23.860-24.000	Native CZE-MS
4x(L7L12)	49275.09	180192828.00	5	28.600-28.790	Native CZE-MS
	47592.30	12518507.38	6	32.210-33.740	Native CZE-MS
	47545.60	15683762.13	6	26.120-26.820	Native CZE-MS
L10(2xL7/L12)	43626.76	87729626.38	8	32.170-36.330	Native CZE-MS
	43451.64	13822108.38	6	32.520-33.470	Native CZE-MS
	43450.76	12671426.38	6	30.190-30.840	Native CZE-MS
	41062.04	24635193.00	5	24.530-24.800	Native CZE-MS
	39754.65	14661561.16	5	23.860-24.260	Native CZE-MS
	39082.33	18276580.50	4	27.800-28.030	Native CZE-MS
	38035.85	15739414.25	5	37.830-38.410	Native CZE-MS
	38035.11	37784358.75	4	24.680-24.990	Native CZE-MS
L6 dimer	37544.59	394313006.00	5	28.900-29.320	Native CZE-MS
	37425.29	19249161.81	4	33.630-35.760	Native CZE-MS
	37425.01	18490802.88	4	25.710-26.470	Native CZE-MS
	37353.40	17494721.50	4	25.710-26.360	Native CZE-MS
	36946.31	37315468.63	4	30.190-30.840	Native CZE-MS
	36912.43	14203769.19	4	30.500-31.140	Native CZE-MS
	35892.63	13314294.50	4	25.820-26.090	Native CZE-MS
	35825.38	18694936.00	5	24.680-25.210	Native CZE-MS
	35824.71	20932786.13	5	30.190-30.690	Native CZE-MS
	35291.31	586968264.00	5	28.220-28.710	Native CZE-MS
	34771.54	19180808.16	5	37.750-38.520	Native CZE-MS
	34737.52	68392405.81	5	34.730-38.680	Native CZE-MS
	34737.43	38425652.25	5	24.870-25.520	Native CZE-MS
	34338.70	33004451.75	4	25.670-26.820	Native CZE-MS
	33812.45	15986162.75	4	27.800-28.180	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	33812.22	26157549.50	4	24.720-24.990	Native CZE-MS
	33714.25	44680714.38	4	35.190-38.640	Native CZE-MS
	33680.96	12867691.50	4	26.240-26.890	Native CZE-MS
	33347.45	78728801.50	4	30.310-31.030	Native CZE-MS
	32723.70	12410932.16	5	36.480-38.640	Native CZE-MS
	32722.03	31027909.50	5	27.800-28.070	Native CZE-MS
	31761.80	42085354.56	4	32.790-36.410	Native CZE-MS
	31635.31	37323464.50	4	24.720-24.950	Native CZE-MS
	31601.64	14110928.75	4	25.630-26.050	Native CZE-MS
	31330.34	13157763.00	4	27.800-28.070	Native CZE-MS
	30121.94	15231007.75	4	25.630-25.780	Native CZE-MS
	30016.90	41543997.44	4	30.310-31.410	Native CZE-MS
	29751.23	30068649.88	4	35.340-36.710	Native CZE-MS
	29750.40	16955522.69	4	33.320-34.730	Native CZE-MS
	29747.04	17069984.06	4	32.210-34.730	Native CZE-MS
L2	29728.97	6484473790.00	5	32.640-38.680	Native CZE-MS
L2	29728.93	1726659232.00	7	24.450-25.100	Native CZE-MS
	29715.21	41375973.13	5	26.240-26.970	Native CZE-MS
	29446.04	17475574.50	4	25.670-25.940	Native CZE-MS
	28784.55	21650011.50	4	24.680-25.330	Native CZE-MS
	28754.08	20940324.19	4	36.940-38.410	Native CZE-MS
	28041.87	25704625.75	4	23.920-24.110	Native CZE-MS
	27824.70	75286226.22	4	32.290-37.600	Native CZE-MS
	27472.63	15069587.00	4	27.350-27.880	Native CZE-MS
	26925.61	92505813.38	5	32.290-38.560	Native CZE-MS
	26613.45	21097667.00	4	24.720-25.060	Native CZE-MS
	26055.36	20219169.50	4	34.500-37.130	Native CZE-MS
	26055.23	62014539.88	4	30.040-30.990	Native CZE-MS
	25912.15	13242559.00	4	24.830-25.520	Native CZE-MS
	25912.00	53392901.13	4	30.230-30.880	Native CZE-MS
	25872.03	133049904.00	4	24.950-25.250	Native CZE-MS
S3	25852.03	5812204638.00	15	24.800-26.890	Native CZE-MS
	25287.11	12766446.44	4	34.810-36.060	Native CZE-MS
	25261.40	16827501.25	4	25.780-26.050	Native CZE-MS
	24663.51	14626937.56	4	29.430-30.340	Native CZE-MS
	24617.51	99297752.00	6	28.330-28.600	Native CZE-MS
L1	24598.62	11174562972.00	13	27.990-29.130	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	24589.03	14138818.13	4	35.230-35.910	Native CZE-MS
	24581.71	89532974.00	8	28.370-28.750	Native CZE-MS
	24554.52	89122071.00	8	28.370-28.710	Native CZE-MS
	24391.29	12553084.25	4	30.500-30.690	Native CZE-MS
	24059.61	28844330.50	4	24.530-24.830	Native CZE-MS
	23993.30	23163547.75	5	30.340-30.760	Native CZE-MS
	23992.60	32833035.31	6	30.570-31.910	Native CZE-MS
	23988.46	12665814.69	4	36.900-38.250	Native CZE-MS
S4	23337.97	962992308.00	4	36.180-37.480	Native CZE-MS
S4	23337.96	896511188.00	6	33.930-35.340	Native CZE-MS
S4	23337.87	185127748.00	4	33.020-33.360	Native CZE-MS
S4	23337.63	78704493.44	6	38.450-39.150	Native CZE-MS
	23295.30	23337258.00	4	36.100-36.560	Native CZE-MS
L3	22257.68	1459085736.00	8	29.350-30.310	Native CZE-MS
	22253.23	14067587.25	4	26.550-27.230	Native CZE-MS
	22231.58	19578074.88	4	30.190-30.990	Native CZE-MS
	21725.68	67435845.75	4	24.720-25.670	Native CZE-MS
	21613.35	12142120.75	3	24.800-24.950	Native CZE-MS
	21611.53	12550481.63	3	37.020-38.100	Native CZE-MS
	21587.00	25499292.13	3	24.570-25.060	Native CZE-MS
	21385.16	34556461.50	3	28.260-28.600	Native CZE-MS
	21380.41	29589442.00	3	26.780-27.080	Native CZE-MS
	21331.42	30355572.25	5	25.630-26.120	Native CZE-MS
	21286.96	32278618.75	3	26.850-27.230	Native CZE-MS
	21073.50	32614705.69	3	26.050-26.780	Native CZE-MS
	21061.54	26160413.00	4	25.970-26.320	Native CZE-MS
	21035.69	49705336.22	6	33.660-38.250	Native CZE-MS
	20950.09	28513119.00	3	24.110-24.760	Native CZE-MS
	20669.57	15483834.00	3	25.100-25.330	Native CZE-MS
	20596.69	74475627.00	6	24.800-25.060	Native CZE-MS
	20570.50	18734902.25	3	30.000-30.990	Native CZE-MS
	20550.78	85700109.75	5	24.830-26.630	Native CZE-MS
	20550.71	11898012.38	5	39.480-59.870	Native CZE-MS
	20550.68	142383734.31	5	31.450-38.520	Native CZE-MS
	20550.58	11587996.25	4	27.690-28.290	Native CZE-MS
	20531.73	12031261.75	3	30.040-30.760	Native CZE-MS
	20530.91	24502237.75	5	29.430-30.990	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	20530.75	19891761.19	5	39.360-59.920	Native CZE-MS
	20530.71	161451979.06	5	31.560-38.680	Native CZE-MS
	20530.69	46271170.25	4	25.560-26.930	Native CZE-MS
	20530.40	64028960.38	4	27.500-28.450	Native CZE-MS
	20396.04	17052227.75	4	30.340-30.880	Native CZE-MS
	20387.89	26676975.13	3	33.630-37.750	Native CZE-MS
	20368.95	37856526.63	3	30.230-30.990	Native CZE-MS
	20211.48	14319452.84	3	32.170-34.050	Native CZE-MS
L5	20170.50	1330331482.44	9	30.230-33.630	Native CZE-MS
	19871.78	13131928.25	3	26.940-28.780	Native CZE-MS
	19870.35	18607553.75	3	28.570-28.760	Native CZE-MS
	19810.62	13038868.00	3	27.950-28.220	Native CZE-MS
	19751.69	101852359.13	3	24.680-26.930	Native CZE-MS
	19751.60	22624977.78	3	31.750-34.050	Native CZE-MS
	19751.26	12065340.75	3	29.920-30.910	Native CZE-MS
	19731.36	42126799.50	3	26.210-27.350	Native CZE-MS
	19731.23	22871236.94	3	29.810-30.840	Native CZE-MS
	19731.08	20332277.95	5	23.860-24.110	Native CZE-MS
	19545.35	29063897.00	3	24.260-24.530	Native CZE-MS
	19457.28	55756795.25	3	30.310-30.910	Native CZE-MS
	19371.94	12385012.13	5	25.630-26.780	Native CZE-MS
	19323.85	20335058.50	3	24.800-24.950	Native CZE-MS
	19306.86	150424736.63	4	24.150-26.050	Native CZE-MS
	19306.39	27413848.75	3	30.500-31.560	Native CZE-MS
	19104.69	64101981.00	3	25.060-25.210	Native CZE-MS
	18972.93	50466460.50	4	24.260-24.910	Native CZE-MS
	18952.28	44928863.50	4	24.640-24.910	Native CZE-MS
	18794.20	600123447.00	4	28.820-29.470	Native CZE-MS
L6	18772.48	7460249193.50	10	28.630-30.150	Native CZE-MS
	18757.10	647031582.00	3	28.750-29.580	Native CZE-MS
	18732.58	36219921.41	3	36.670-38.330	Native CZE-MS
	18728.92	703275297.00	3	28.790-29.620	Native CZE-MS
	18700.23	122598041.00	3	28.900-29.470	Native CZE-MS
	18234.79	12793423.78	3	30.570-31.750	Native CZE-MS
	18184.70	14234733.30	3	30.720-31.480	Native CZE-MS
	18098.90	11231188.63	3	32.600-34.350	Native CZE-MS
	18028.85	32551587.25	4	27.800-28.220	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	17990.18	187265974.00	3	25.630-26.470	Native CZE-MS
	17927.74	13215119.38	3	37.910-38.410	Native CZE-MS
	17849.95	24302120.00	4	24.530-24.910	Native CZE-MS
	17775.44	48621391.50	3	24.720-25.970	Native CZE-MS
	17747.96	374189216.00	3	24.260-24.530	Native CZE-MS
	17747.66	264638414.50	3	24.530-25.100	Native CZE-MS
	17740.11	13656095.00	5	24.800-25.250	Native CZE-MS
	17739.94	26318107.00	4	27.080-27.270	Native CZE-MS
	17739.54	159024619.25	5	30.230-32.440	Native CZE-MS
	17724.87	26043947.50	4	24.640-24.800	Native CZE-MS
	17722.74	104119036.69	3	24.530-27.230	Native CZE-MS
	17722.72	16779599.38	4	31.600-34.080	Native CZE-MS
	17722.53	93608858.13	3	30.190-30.990	Native CZE-MS
	17705.59	70765418.63	3	25.210-27.800	Native CZE-MS
L10	17560.70	38253062.00	3	24.530-24.950	Native CZE-MS
	17513.10	30031691.00	3	30.610-31.030	Native CZE-MS
	17493.92	492884774.00	4	30.190-32.370	Native CZE-MS
	17477.90	71790832.38	4	30.340-30.990	Native CZE-MS
	17477.03	184292203.50	4	30.230-30.910	Native CZE-MS
	17473.79	316198468.00	4	30.380-30.720	Native CZE-MS
	17473.33	1755891895.50	9	25.970-28.030	Native CZE-MS
	17472.86	519546534.25	6	29.730-32.560	Native CZE-MS
	17470.26	24875609.56	4	36.900-38.410	Native CZE-MS
	17470.21	15770413.50	3	26.240-26.700	Native CZE-MS
	17470.13	21061287.81	4	34.730-36.290	Native CZE-MS
	17458.43	520653692.00	3	26.210-26.930	Native CZE-MS
S7	17456.26	3286304348.00	10	30.080-30.840	Native CZE-MS
S7	17456.03	1627376711.91	10	30.150-32.290	Native CZE-MS
S5	17454.98	19186293.00	3	24.300-24.570	Native CZE-MS
	17451.65	30263224.00	4	27.800-28.220	Native CZE-MS
	17451.42	78131259.69	5	31.450-38.410	Native CZE-MS
	17440.67	695197718.44	10	30.150-31.870	Native CZE-MS
	17398.85	34217755.00	4	30.190-30.500	Native CZE-MS
	17273.46	27173306.50	4	30.460-30.840	Native CZE-MS
	17215.85	21286543.00	5	25.940-26.510	Native CZE-MS
	17162.89	12830319.25	3	36.750-38.100	Native CZE-MS
	17162.88	21760011.00	3	26.930-27.230	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	17110.32	19683003.50	4	24.260-24.490	Native CZE-MS
	17090.41	22884416.25	3	24.570-24.990	Native CZE-MS
	17049.15	46104896.25	4	26.850-27.540	Native CZE-MS
	16850.60	27139802.50	3	26.970-27.230	Native CZE-MS
	16633.79	24892056.00	3	24.110-24.680	Native CZE-MS
	16600.68	38220421.13	4	25.140-27.200	Native CZE-MS
	16568.90	54813254.88	4	24.530-25.670	Native CZE-MS
	16402.97	27766972.00	4	24.800-25.100	Native CZE-MS
	16399.13	95975052.75	5	28.180-28.940	Native CZE-MS
	16360.52	57431887.13	3	30.190-31.670	Native CZE-MS
	16313.31	51022848.81	3	26.850-27.610	Native CZE-MS
	16312.99	22593560.50	3	24.720-25.250	Native CZE-MS
	16199.35	11863398.13	3	26.090-26.780	Native CZE-MS
	16057.87	16584724.63	3	27.800-28.100	Native CZE-MS
	16057.76	30832874.88	3	32.370-35.870	Native CZE-MS
	16056.17	310568771.00	3	26.820-27.650	Native CZE-MS
	16040.17	626834564.88	7	25.860-27.880	Native CZE-MS
	16034.53	214051424.25	5	26.780-27.650	Native CZE-MS
	16033.42	93770793.00	4	26.820-27.800	Native CZE-MS
	16021.21	68289616.00	3	27.990-28.180	Native CZE-MS
L13	16018.70	226310225.00	4	28.330-30.340	Native CZE-MS
L13	16018.59	19376760.00	3	35.450-35.610	Native CZE-MS
L13	16018.39	209864148.49	3	38.450-59.920	Native CZE-MS
L13	16018.31	746458677.00	3	31.180-34.690	Native CZE-MS
L13	16018.29	7908289976.00	7	26.660-28.070	Native CZE-MS
	16017.57	159905357.50	5	28.070-28.860	Native CZE-MS
	16003.77	551027163.00	3	26.820-27.760	Native CZE-MS
	15975.97	206440580.00	4	26.820-27.650	Native CZE-MS
	15960.58	93039774.00	3	26.970-27.460	Native CZE-MS
	15959.73	65377738.00	3	26.820-27.500	Native CZE-MS
	15768.57	12417824.50	3	37.830-38.560	Native CZE-MS
	15649.77	46404548.50	6	27.420-28.290	Native CZE-MS
	15577.42	25708922.63	3	30.310-30.910	Native CZE-MS
	15504.71	25394339.25	3	26.550-27.200	Native CZE-MS
	15465.28	12971416.75	4	27.920-28.220	Native CZE-MS
	15409.31	22605009.63	4	25.250-26.660	Native CZE-MS
	15384.96	14262525.75	3	24.300-24.490	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	15327.64	36792505.50	3	32.480-32.790	Native CZE-MS
	15327.10	1823387593.50	3	32.600-36.860	Native CZE-MS
	15311.68	73002318.00	3	25.780-26.050	Native CZE-MS
	15311.35	11865426.25	3	37.400-37.750	Native CZE-MS
	15311.21	2043323097.00	3	32.480-36.900	Native CZE-MS
	15310.58	106399190.00	6	24.870-26.700	Native CZE-MS
L16	15283.64	18048214.50	3	24.420-24.570	Native CZE-MS
L16	15283.37	19024864.00	3	26.820-27.230	Native CZE-MS
	15141.95	33333881.00	3	24.530-24.910	Native CZE-MS
	15128.35	53950091.75	5	24.450-25.250	Native CZE-MS
	15114.84	55259696.00	4	24.380-24.570	Native CZE-MS
	15113.64	76153799.50	5	24.570-26.470	Native CZE-MS
	15113.60	35374709.50	4	32.440-37.170	Native CZE-MS
	15023.46	13269198.56	3	32.100-34.730	Native CZE-MS
	15019.90	24781531.00	3	26.820-27.080	Native CZE-MS
	14988.05	33253680.50	5	26.050-26.550	Native CZE-MS
L15	14968.90	71082809.50	5	25.710-26.090	Native CZE-MS
L15	14966.38	91884172.50	5	31.600-32.710	Native CZE-MS
L15	14966.19	3532690499.00	6	25.670-27.920	Native CZE-MS
L15	14966.15	657159104.00	6	25.820-25.970	Native CZE-MS
	14951.88	208058797.00	3	26.090-26.630	Native CZE-MS
	14904.35	26389280.47	4	32.290-35.490	Native CZE-MS
	14904.25	83605147.38	4	24.450-26.820	Native CZE-MS
	14889.24	18212527.50	3	24.420-24.910	Native CZE-MS
	14872.96	246498276.00	5	27.690-28.370	Native CZE-MS
L11	14869.90	1875745917.00	6	27.570-28.480	Native CZE-MS
	14864.23	85976746.00	3	36.100-37.640	Native CZE-MS
	14864.16	15456654.50	3	34.960-35.420	Native CZE-MS
S9	14855.89	78749334.00	3	27.800-28.030	Native CZE-MS
	14807.22	60621312.38	3	24.830-26.400	Native CZE-MS
	14776.83	75312206.50	3	30.310-30.880	Native CZE-MS
	14747.37	19671058.00	3	33.400-34.310	Native CZE-MS
	14746.54	48821706.00	3	33.360-35.260	Native CZE-MS
	14740.15	12670478.25	3	33.970-34.270	Native CZE-MS
	14739.86	15176988.75	4	32.640-32.790	Native CZE-MS
S9	14724.83	80594486.96	4	46.650-59.870	Native CZE-MS
S9	14724.72	4335435984.13	8	31.560-37.940	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
S9	14724.55	70605299.26	7	40.380-45.400	Native CZE-MS
S9	14724.36	14531521.34	4	39.440-40.630	Native CZE-MS
	14679.75	22231349.50	5	24.110-24.990	Native CZE-MS
	14676.65	14034059.69	3	25.710-26.630	Native CZE-MS
	14561.06	12872512.38	3	30.340-30.840	Native CZE-MS
	14540.64	33319336.13	3	30.190-30.690	Native CZE-MS
L17	14364.41	1903783470.50	5	33.930-37.670	Native CZE-MS
L17	14364.40	123221696.69	3	40.330-59.710	Native CZE-MS
	14227.40	40088660.00	4	27.420-27.650	Native CZE-MS
	14048.80	40439689.09	8	34.200-38.520	Native CZE-MS
	14048.70	14793876.44	4	25.630-26.700	Native CZE-MS
	14022.54	21183078.75	4	25.400-26.970	Native CZE-MS
	14022.41	31971894.63	3	29.170-30.310	Native CZE-MS
	14016.48	215358419.00	3	30.190-32.600	Native CZE-MS
S8	13995.05	2895695196.00	4	30.190-31.290	Native CZE-MS
	13904.39	37540903.50	3	25.630-25.940	Native CZE-MS
	13891.55	53944118.63	3	25.210-26.660	Native CZE-MS
	13879.23	35262448.50	3	24.830-25.180	Native CZE-MS
	13812.11	14658476.63	3	30.340-30.840	Native CZE-MS
	13747.96	11955307.25	3	24.420-24.950	Native CZE-MS
S11	13730.36	30748259.88	3	26.590-26.820	Native CZE-MS
S11	13729.87	29496898.50	3	32.440-33.060	Native CZE-MS
S11	13729.07	64110762.69	3	31.290-31.830	Native CZE-MS
S11	13729.04	62011404.29	3	40.120-59.920	Native CZE-MS
S11	13729.03	306068651.00	3	27.500-27.920	Native CZE-MS
S11	13728.85	184071011.44	3	31.070-36.030	Native CZE-MS
S11	13728.31	1297091197.00	6	25.400-26.630	Native CZE-MS
	13696.16	12381540.09	4	35.910-38.060	Native CZE-MS
	13657.04	17077274.25	4	35.610-38.060	Native CZE-MS
S12	13652.08	416349624.00	5	24.260-24.640	Native CZE-MS
S12	13649.94	546665336.00	5	24.070-24.260	Native CZE-MS
S12	13648.59	777818034.00	7	24.300-24.830	Native CZE-MS
	13643.93	90386737.25	3	24.150-24.680	Native CZE-MS
	13627.20	44963518.00	3	24.420-24.570	Native CZE-MS
	13604.63	18236489.38	5	25.370-26.120	Native CZE-MS
L14	13540.37	1901323243.00	5	24.070-24.490	Native CZE-MS
	13432.53	42629295.75	3	24.420-24.950	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
L20	13365.28	17664820.94	3	34.880-38.640	Native CZE-MS
	13098.23	45629362.00	4	24.450-24.950	Native CZE-MS
	13082.76	38180025.25	4	26.210-26.780	Native CZE-MS
	13031.32	52726791.25	3	36.860-37.170	Native CZE-MS
	13003.95	551813473.00	5	25.370-25.900	Native CZE-MS
L19	13001.43	2522005428.50	5	25.210-26.050	Native CZE-MS
	12968.01	46462453.52	3	53.420-59.920	Native CZE-MS
	12967.99	24880068.50	3	33.820-34.430	Native CZE-MS
	12967.82	1567946297.63	8	34.050-38.560	Native CZE-MS
	12967.67	69626075.03	3	39.780-45.450	Native CZE-MS
	12956.24	13361139.88	3	30.230-30.880	Native CZE-MS
	12956.14	39761050.88	3	24.530-25.400	Native CZE-MS
	12955.25	15954989.50	4	26.120-26.660	Native CZE-MS
	12954.75	67561846.63	4	23.920-24.910	Native CZE-MS
S13	12950.58	19666075.69	4	29.350-29.920	Native CZE-MS
	12943.13	17473191.69	4	32.440-35.300	Native CZE-MS
	12943.09	28758748.88	3	24.570-25.210	Native CZE-MS
	12926.46	73025071.00	4	24.950-25.780	Native CZE-MS
	12822.94	30765274.50	3	24.680-24.990	Native CZE-MS
	12799.65	12640709.88	4	25.780-26.700	Native CZE-MS
	12799.58	32881659.13	3	32.210-35.070	Native CZE-MS
	12799.57	11458372.81	3	35.870-37.170	Native CZE-MS
	12791.74	47327348.00	3	28.520-28.790	Native CZE-MS
	12787.21	15981949.50	4	26.930-27.270	Native CZE-MS
L18	12774.79	21011209.50	3	26.970-27.230	Native CZE-MS
L18	12774.52	19235196.13	3	35.760-36.330	Native CZE-MS
L18	12771.63	402760336.00	3	28.180-28.900	Native CZE-MS
L18	12770.99	21475999.75	3	29.770-30.150	Native CZE-MS
L18	12768.99	3095496417.25	5	27.920-32.750	Native CZE-MS
	12547.00	12218299.50	3	30.340-30.570	Native CZE-MS
	12535.93	40330534.69	3	29.470-31.030	Native CZE-MS
	12426.77	35179955.00	3	24.260-24.760	Native CZE-MS
	12396.80	24226216.94	3	32.520-34.500	Native CZE-MS
	12396.79	29639686.75	3	35.190-38.100	Native CZE-MS
	12396.56	13890846.38	3	30.230-30.720	Native CZE-MS
	12382.74	11253919.47	3	36.900-38.250	Native CZE-MS
	12228.49	126898774.00	3	27.920-28.070	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
L22	12226.13	20372507.00	3	31.640-32.330	Native CZE-MS
L22	12225.62	22554025.50	3	31.560-32.370	Native CZE-MS
L22	12225.27	345500300.00	4	27.760-28.290	Native CZE-MS
L22	12224.91	401910346.00	3	28.260-28.940	Native CZE-MS
	12187.05	54318741.44	4	30.190-30.990	Native CZE-MS
	12031.07	68993057.75	5	31.870-38.480	Native CZE-MS
	12031.04	13403197.75	4	25.860-26.820	Native CZE-MS
	12018.84	54852819.00	4	24.530-24.950	Native CZE-MS
L7	12006.02	16158787.13	6	34.810-38.370	Native CZE-MS
L7	12005.79	89822515.25	4	24.570-25.370	Native CZE-MS
	11898.18	16733407.00	3	24.530-24.990	Native CZE-MS
	11845.77	15548685.50	3	25.630-25.820	Native CZE-MS
	11766.32	13082522.00	3	33.170-34.620	Native CZE-MS
S10	11734.50	20415585.00	3	33.740-34.160	Native CZE-MS
	11706.52	19957432.50	5	32.370-34.880	Native CZE-MS
	11706.49	11340439.38	5	35.450-36.480	Native CZE-MS
S14	11449.05	14575406.82	3	50.040-59.920	Native CZE-MS
S14	11448.96	84039083.00	3	32.210-32.600	Native CZE-MS
S14	11448.84	1022251741.13	3	31.330-37.830	Native CZE-MS
S14	11446.23	19959316.75	4	25.710-26.550	Native CZE-MS
	11435.91	27949518.00	3	24.640-24.990	Native CZE-MS
	11303.79	165305752.75	4	30.150-31.560	Native CZE-MS
	11237.88	19863480.28	4	37.170-38.370	Native CZE-MS
	11225.90	15045492.72	5	23.860-24.760	Native CZE-MS
L23	11199.30	11807813.25	3	31.140-31.830	Native CZE-MS
L23	11198.83	26031533.63	3	31.290-31.980	Native CZE-MS
L23	11198.27	875728805.75	4	28.900-30.610	Native CZE-MS
L24	11186.87	510789122.00	3	26.820-27.390	Native CZE-MS
L24	11184.17	807205314.00	4	26.780-27.500	Native CZE-MS
	10995.64	117515269.50	3	24.110-25.100	Native CZE-MS
	10995.57	12066038.13	3	30.310-30.720	Native CZE-MS
	10906.43	39857170.00	3	26.090-27.270	Native CZE-MS
	10796.91	25463426.50	3	24.420-24.640	Native CZE-MS
	10796.83	19332287.19	3	34.390-38.020	Native CZE-MS
	10761.27	78593929.13	3	31.870-36.440	Native CZE-MS
	10758.71	62045558.38	3	27.650-28.290	Native CZE-MS
L25	10695.13	339346134.00	3	28.180-28.600	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
L25	10694.75	185489198.00	3	28.370-28.750	Native CZE-MS
L25	10692.72	357282672.00	4	28.100-28.750	Native CZE-MS
	10686.36	12329809.84	3	32.640-34.580	Native CZE-MS
	10577.25	20431056.75	3	28.070-28.900	Native CZE-MS
	10565.84	18295676.38	3	30.340-30.840	Native CZE-MS
	10545.04	24430321.75	3	28.900-29.540	Native CZE-MS
	10544.97	73593990.19	3	31.330-36.790	Native CZE-MS
	10544.79	12146001.38	3	30.080-30.570	Native CZE-MS
	10455.75	16681600.00	3	25.820-26.700	Native CZE-MS
	10455.68	42950074.50	3	24.380-24.680	Native CZE-MS
	10445.62	13051513.19	5	34.620-36.220	Native CZE-MS
S19	10298.22	109851928.75	4	26.780-27.350	Native CZE-MS
	10253.97	15625173.50	3	24.680-25.210	Native CZE-MS
	10005.44	13425809.75	3	27.950-28.100	Native CZE-MS
	9860.77	14115708.38	3	30.340-30.880	Native CZE-MS
	9785.33	18122927.06	3	36.790-38.250	Native CZE-MS
	9626.10	12881271.06	3	29.940-31.520	Native CZE-MS
S17	9571.42	21049370.38	3	29.860-32.940	Native CZE-MS
S20	9554.26	15765783.91	3	24.040-26.910	Native CZE-MS
	9445.60	52971728.13	3	24.570-25.100	Native CZE-MS
	9355.65	36003891.50	4	32.330-36.900	Native CZE-MS
S16	9189.41	100632956.25	3	29.090-30.150	Native CZE-MS
	9157.76	33648523.25	3	27.350-27.880	Native CZE-MS
	9085.33	11283203.75	3	30.340-30.990	Native CZE-MS
	9074.90	28306865.00	3	38.680-38.870	Native CZE-MS
	9013.66	19611837.00	3	25.370-25.900	Native CZE-MS
L27	8992.48	257779862.00	3	25.400-25.670	Native CZE-MS
L28	8876.75	209303875.00	3	24.300-25.370	Native CZE-MS
	8860.83	84628737.50	4	24.420-26.360	Native CZE-MS
	8860.00	70189090.00	3	24.640-24.830	Native CZE-MS
S18	8845.83	20943953.50	3	24.570-25.100	Native CZE-MS
	8797.45	117359137.75	3	24.450-26.360	Native CZE-MS
	8789.18	51584453.00	4	24.150-25.060	Native CZE-MS
	8780.33	21262262.50	3	24.530-24.800	Native CZE-MS
	8725.43	11769476.38	3	29.240-30.190	Native CZE-MS
	8725.27	11397622.17	3	36.940-38.640	Native CZE-MS
	8608.36	17306111.81	3	26.120-26.820	Native CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	8572.92	19050190.38	4	32.600-34.620	Native CZE-MS
	8572.90	13126993.38	3	36.900-38.410	Native CZE-MS
	8564.47	13691834.50	3	27.950-28.220	Native CZE-MS
S21	8348.45	14223738.38	3	37.940-38.560	Native CZE-MS
	8284.92	28080253.31	4	37.170-38.330	Native CZE-MS
	8284.68	23866736.13	4	24.570-26.090	Native CZE-MS
	8068.64	15757907.00	4	25.780-26.090	Native CZE-MS
	7969.21	14026430.27	3	37.020-38.410	Native CZE-MS
L31	7861.02	13317508.59	4	27.910-28.560	Native CZE-MS
	7636.28	14060749.75	4	24.680-24.830	Native CZE-MS
	7564.35	57628600.25	5	32.750-38.480	Native CZE-MS
	7508.20	25966699.38	3	24.150-24.830	Native CZE-MS
	7452.37	58539152.00	3	24.420-24.640	Native CZE-MS
	7444.55	15537822.25	3	24.830-25.330	Native CZE-MS
	7276.47	32302666.42	4	36.060-38.370	Native CZE-MS
	7276.26	27367688.69	4	29.920-30.880	Native CZE-MS
L35	7172.32	16002847.64	4	36.060-38.640	Native CZE-MS
	6604.10	15846258.13	3	35.870-38.060	Native CZE-MS
	6387.63	18147413.84	3	30.190-30.840	Native CZE-MS
L33	6250.72	20729247.00	3	25.970-27.230	Native CZE-MS
	6135.37	39522323.19	3	30.310-30.990	Native CZE-MS
	57880.08	6231022.56	8	13.920-14.590	Denaturing CZE-MS
	40915.76	22296264.88	17	13.920-14.110	Denaturing CZE-MS
	39087.71	13841884.13	7	15.060-16.760	Denaturing CZE-MS
	39087.32	20272817.31	4	17.500-20.040	Denaturing CZE-MS
	39015.01	17863386.88	4	13.400-13.750	Denaturing CZE-MS
	38940.84	6560249.88	12	14.280-14.480	Denaturing CZE-MS
	38720.36	12619115.25	12	12.150-12.350	Denaturing CZE-MS
	37542.06	16944235.41	18	14.180-14.790	Denaturing CZE-MS
	36784.11	5779094.17	5	15.680-16.820	Denaturing CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	35257.39	6053083.00	5	14.750-15.050	Denaturing CZE-MS
	35157.97	18606289.75	9	15.060-15.260	Denaturing CZE-MS
	33868.38	4853864.00	8	13.710-14.320	Denaturing CZE-MS
	32537.65	16787675.85	4	15.730-19.160	Denaturing CZE-MS
	31106.55	7599246.69	7	13.500-14.840	Denaturing CZE-MS
	29932.88	7510669.69	5	14.020-14.320	Denaturing CZE-MS
L2	29728.46	1029733081.00	20	12.310-12.610	Denaturing CZE-MS
L2	29726.88	231576506.00	21	12.090-12.350	Denaturing CZE-MS
	29705.77	20273838.00	11	12.350-12.550	Denaturing CZE-MS
	29684.98	62514880.50	12	12.350-12.550	Denaturing CZE-MS
	29683.20	18325134.38	9	12.150-12.350	Denaturing CZE-MS
	29652.45	54519504.63	8	12.350-12.550	Denaturing CZE-MS
	29549.70	3943309.94	5	15.730-16.870	Denaturing CZE-MS
	29512.57	5427396.30	6	15.530-16.820	Denaturing CZE-MS
	29511.65	5380595.22	6	17.760-19.940	Denaturing CZE-MS
	29448.20	55446845.50	5	12.670-13.020	Denaturing CZE-MS
	28778.64	8452837.50	5	14.590-14.790	Denaturing CZE-MS
	27995.36	15749205.19	5	17.450-20.100	Denaturing CZE-MS
S2	26627.71	60987078.01	21	13.660-14.320	Denaturing CZE-MS
	26609.57	10270243.53	14	13.920-14.220	Denaturing CZE-MS
	26474.81	24948613.06	15	14.070-14.370	Denaturing CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
S3	25851.33	915362325.00	18	13.290-13.590	Denaturing CZE-MS
	25832.77	39840688.50	6	13.290-13.540	Denaturing CZE-MS
	25748.99	15831991.00	6	13.290-13.490	Denaturing CZE-MS
	25365.42	6149184.38	6	14.960-15.160	Denaturing CZE-MS
	24855.29	4244849.91	10	15.680-15.980	Denaturing CZE-MS
L1	24597.61	2210374290.00	18	15.110-15.520	Denaturing CZE-MS
	24553.17	285257413.00	15	15.160-15.460	Denaturing CZE-MS
	24538.22	23935769.00	13	15.160-15.420	Denaturing CZE-MS
	23501.96	7134775.75	6	14.540-14.840	Denaturing CZE-MS
S4	23337.36	833552163.00	16	12.980-13.280	Denaturing CZE-MS
	23318.86	84604644.00	9	13.030-13.230	Denaturing CZE-MS
	22719.19	18298821.00	10	12.200-12.400	Denaturing CZE-MS
L3	22256.80	272557048.00	14	14.960-15.260	Denaturing CZE-MS
	22239.06	22980678.81	9	14.960-15.200	Denaturing CZE-MS
	22197.48	12423060.75	8	15.010-15.200	Denaturing CZE-MS
	22085.95	164118382.27	16	14.120-14.630	Denaturing CZE-MS
	22068.28	6657142.25	11	14.380-14.590	Denaturing CZE-MS
	22041.59	6005336.88	10	14.380-14.590	Denaturing CZE-MS
	22027.82	7750384.13	10	14.380-14.590	Denaturing CZE-MS
	21737.47	6636032.13	5	14.590-14.840	Denaturing CZE-MS
	21519.10	5222818.00	8	11.210-11.410	Denaturing CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	21460.61	5654841.56	4	15.010-15.570	Denaturing CZE-MS
	21132.37	5000938.86	6	14.070-14.530	Denaturing CZE-MS
	20439.84	40603478.00	6	11.840-12.040	Denaturing CZE-MS
	20305.63	15613376.50	4	14.800-15.000	Denaturing CZE-MS
	20292.37	3732757.81	3	14.020-14.220	Denaturing CZE-MS
L5	20169.60	396824433.75	13	14.540-14.890	Denaturing CZE-MS
	20152.66	21513935.50	12	14.640-14.840	Denaturing CZE-MS
	20125.31	13566141.75	11	14.640-14.840	Denaturing CZE-MS
	20123.01	5475522.75	10	11.260-11.460	Denaturing CZE-MS
	19717.02	4018259.88	4	15.010-15.200	Denaturing CZE-MS
	18803.07	13263873.63	6	14.750-15.050	Denaturing CZE-MS
	18789.64	9099813.25	5	14.750-15.000	Denaturing CZE-MS
L6	18771.84	1410244368.00	11	14.700-15.050	Denaturing CZE-MS
	18754.93	175132575.25	8	14.750-15.050	Denaturing CZE-MS
	18740.99	31799809.00	6	14.800-15.050	Denaturing CZE-MS
	18728.09	125150427.25	9	14.750-15.050	Denaturing CZE-MS
	18724.68	4642912.88	7	11.420-11.620	Denaturing CZE-MS
	18713.14	112571155.31	8	14.750-15.050	Denaturing CZE-MS
	18698.13	53649352.00	7	14.800-15.000	Denaturing CZE-MS
	18682.68	24176057.38	7	14.800-15.050	Denaturing CZE-MS
	18654.68	7388379.75	5	14.800-15.000	Denaturing CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	18641.35	3838328.02	3	17.500-18.640	Denaturing CZE-MS
	18641.00	5389729.38	3	9.547-10.220	Denaturing CZE-MS
	18640.91	3695599.41	3	15.790-16.760	Denaturing CZE-MS
	18640.46	4771613.28	6	11.160-11.780	Denaturing CZE-MS
	18640.19	17492759.44	5	13.970-14.940	Denaturing CZE-MS
	18604.48	28234812.91	10	11.160-11.410	Denaturing CZE-MS
	18392.97	10577556.83	5	13.550-14.690	Denaturing CZE-MS
	18305.29	30922981.00	4	13.140-13.330	Denaturing CZE-MS
	18207.99	3678493.45	3	15.680-16.920	Denaturing CZE-MS
	18137.50	3732839.10	4	16.000-16.760	Denaturing CZE-MS
	17985.03	4834006.31	4	11.990-12.190	Denaturing CZE-MS
	17983.81	44509220.00	10	11.470-11.720	Denaturing CZE-MS
	17949.87	4345148.06	4	15.010-15.200	Denaturing CZE-MS
	17865.92	13941909.88	8	11.680-11.880	Denaturing CZE-MS
	17615.81	40254440.88	9	15.420-15.730	Denaturing CZE-MS
	17597.46	16874976.50	6	15.480-15.680	Denaturing CZE-MS
	17579.86	453310460.75	13	15.420-15.780	Denaturing CZE-MS
L10	17563.33	43951670.25	9	15.480-15.730	Denaturing CZE-MS
	17547.49	9464661.38	11	15.680-15.940	Denaturing CZE-MS
	17506.21	9152586.75	4	15.320-15.620	Denaturing CZE-MS
	17475.43	28047939.09	6	15.320-15.880	Denaturing CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	17472.63	909718128.00	10	13.090-13.280	Denaturing CZE-MS
	17472.05	52190632.84	10	15.420-15.780	Denaturing CZE-MS
S7	17456.41	108719526.00	5	13.090-13.280	Denaturing CZE-MS
S5	17453.52	1457128164.00	11	15.270-15.570	Denaturing CZE-MS
	17428.22	89693540.00	5	13.090-13.280	Denaturing CZE-MS
	17411.25	85808558.31	9	15.320-15.570	Denaturing CZE-MS
	17396.10	79623291.25	8	15.320-15.570	Denaturing CZE-MS
	17381.05	70256712.00	6	15.320-15.570	Denaturing CZE-MS
	17359.20	24618403.00	8	11.260-11.460	Denaturing CZE-MS
	17327.92	4619150.75	5	11.630-11.820	Denaturing CZE-MS
	16638.41	8209584.50	4	15.420-15.680	Denaturing CZE-MS
	16540.33	9985825.38	3	14.960-15.880	Denaturing CZE-MS
	16185.11	5833419.23	3	17.610-19.990	Denaturing CZE-MS
	16163.07	4307790.03	4	17.450-18.800	Denaturing CZE-MS
	16116.59	23832252.25	5	12.830-13.280	Denaturing CZE-MS
	16034.14	70364362.00	6	13.140-13.390	Denaturing CZE-MS
L13	16017.83	1502346920.00	9	13.140-13.390	Denaturing CZE-MS
	15974.88	87582211.50	5	13.190-13.390	Denaturing CZE-MS
	15954.08	3979013.71	3	17.450-19.940	Denaturing CZE-MS
	15885.67	4836266.81	3	16.930-17.500	Denaturing CZE-MS
	15770.45	26581936.63	4	13.030-13.330	Denaturing CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	15768.25	44867815.88	10	16.670-16.980	Denaturing CZE-MS
	15763.26	9736908.13	5	15.110-15.460	Denaturing CZE-MS
	15430.70	63856557.00	12	15.010-15.260	Denaturing CZE-MS
	15361.70	9932872.69	6	12.250-12.870	Denaturing CZE-MS
	15327.33	14744525.00	4	12.870-13.130	Denaturing CZE-MS
	15326.51	284457788.00	8	12.670-12.920	Denaturing CZE-MS
	15310.61	348917278.50	8	12.670-13.020	Denaturing CZE-MS
	15301.43	199682534.75	11	14.900-15.260	Denaturing CZE-MS
L16	15285.75	5764395.25	6	15.010-15.200	Denaturing CZE-MS
	15266.67	30003337.50	5	12.670-12.870	Denaturing CZE-MS
	15172.52	87005589.13	11	14.440-15.160	Denaturing CZE-MS
	15063.77	15698053.50	4	13.190-13.650	Denaturing CZE-MS
L15	14965.71	939049612.56	8	13.500-14.430	Denaturing CZE-MS
	14949.89	37364958.50	3	13.550-13.750	Denaturing CZE-MS
	14921.92	39260699.00	5	13.550-13.750	Denaturing CZE-MS
	14901.56	4088367.84	5	15.840-16.760	Denaturing CZE-MS
	14886.58	10488826.50	7	15.160-15.940	Denaturing CZE-MS
L11	14886.06	44556422.27	8	15.420-16.870	Denaturing CZE-MS
L11	14884.74	5859356.69	6	14.540-15.160	Denaturing CZE-MS
L11	14869.76	1877042152.56	11	14.440-16.980	Denaturing CZE-MS
	14853.98	91368994.84	9	14.540-16.920	Denaturing CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	14852.90	12780275.25	5	15.060-15.730	Denaturing CZE-MS
	14840.34	6760285.13	8	14.490-14.790	Denaturing CZE-MS
	14839.90	17598456.91	8	15.790-16.870	Denaturing CZE-MS
	14826.26	86549421.31	9	14.490-16.870	Denaturing CZE-MS
	14811.59	29631019.89	8	15.680-16.920	Denaturing CZE-MS
	14811.44	8656265.50	6	14.490-14.840	Denaturing CZE-MS
	14796.62	17753151.94	6	15.940-16.760	Denaturing CZE-MS
S9	14771.98	10480842.50	4	15.480-15.780	Denaturing CZE-MS
S9	14724.40	490816407.00	9	12.670-12.970	Denaturing CZE-MS
	14380.13	19924363.75	5	12.770-12.970	Denaturing CZE-MS
L17	14364.04	419313300.25	8	12.720-13.130	Denaturing CZE-MS
	14347.97	32154714.00	5	12.720-12.970	Denaturing CZE-MS
	14030.89	67466761.00	7	15.480-15.730	Denaturing CZE-MS
	14010.96	20733607.69	8	15.530-15.830	Denaturing CZE-MS
S8	13994.72	828542513.84	10	15.480-15.880	Denaturing CZE-MS
	13979.00	64655096.50	7	15.480-15.730	Denaturing CZE-MS
	13964.53	14062836.00	6	15.530-15.730	Denaturing CZE-MS
	13950.89	10601234.56	7	15.580-15.780	Denaturing CZE-MS
	13936.33	11120109.44	4	15.480-15.680	Denaturing CZE-MS
	13892.07	7050904.63	6	15.530-15.730	Denaturing CZE-MS
	13745.57	24238415.25	6	11.210-11.720	Denaturing CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
S11	13726.12	144687967.13	7	12.200-12.810	Denaturing CZE-MS
S12	13648.87	457224544.50	8	11.680-12.300	Denaturing CZE-MS
S12	13647.32	205172038.00	9	11.160-11.720	Denaturing CZE-MS
	13631.52	18593635.63	7	11.210-11.720	Denaturing CZE-MS
	13603.45	11539098.38	8	11.210-12.300	Denaturing CZE-MS
	13589.01	9951615.50	4	11.520-11.720	Denaturing CZE-MS
	13574.41	8316903.75	4	11.520-11.720	Denaturing CZE-MS
L14	13540.31	180872377.75	8	14.280-14.590	Denaturing CZE-MS
	13538.38	75308956.00	5	12.930-13.130	Denaturing CZE-MS
	13482.76	7247972.75	5	14.330-14.530	Denaturing CZE-MS
	13458.07	4695193.07	5	14.180-15.200	Denaturing CZE-MS
	13365.30	345348223.50	8	12.040-12.400	Denaturing CZE-MS
L19	13001.30	421918500.25	6	13.030-13.280	Denaturing CZE-MS
S13	12950.60	282818130.00	5	12.350-12.610	Denaturing CZE-MS
	12867.25	43424449.38	5	12.870-13.590	Denaturing CZE-MS
	12846.80	8574054.56	4	11.730-12.610	Denaturing CZE-MS
	12805.34	62111800.00	6	13.400-13.590	Denaturing CZE-MS
	12794.73	6065451.30	3	17.550-18.640	Denaturing CZE-MS
L18	12768.96	1296104682.00	7	13.340-13.650	Denaturing CZE-MS
	12753.13	125121146.00	5	13.400-13.590	Denaturing CZE-MS
	12725.19	47196608.50	5	13.400-13.590	Denaturing CZE-MS

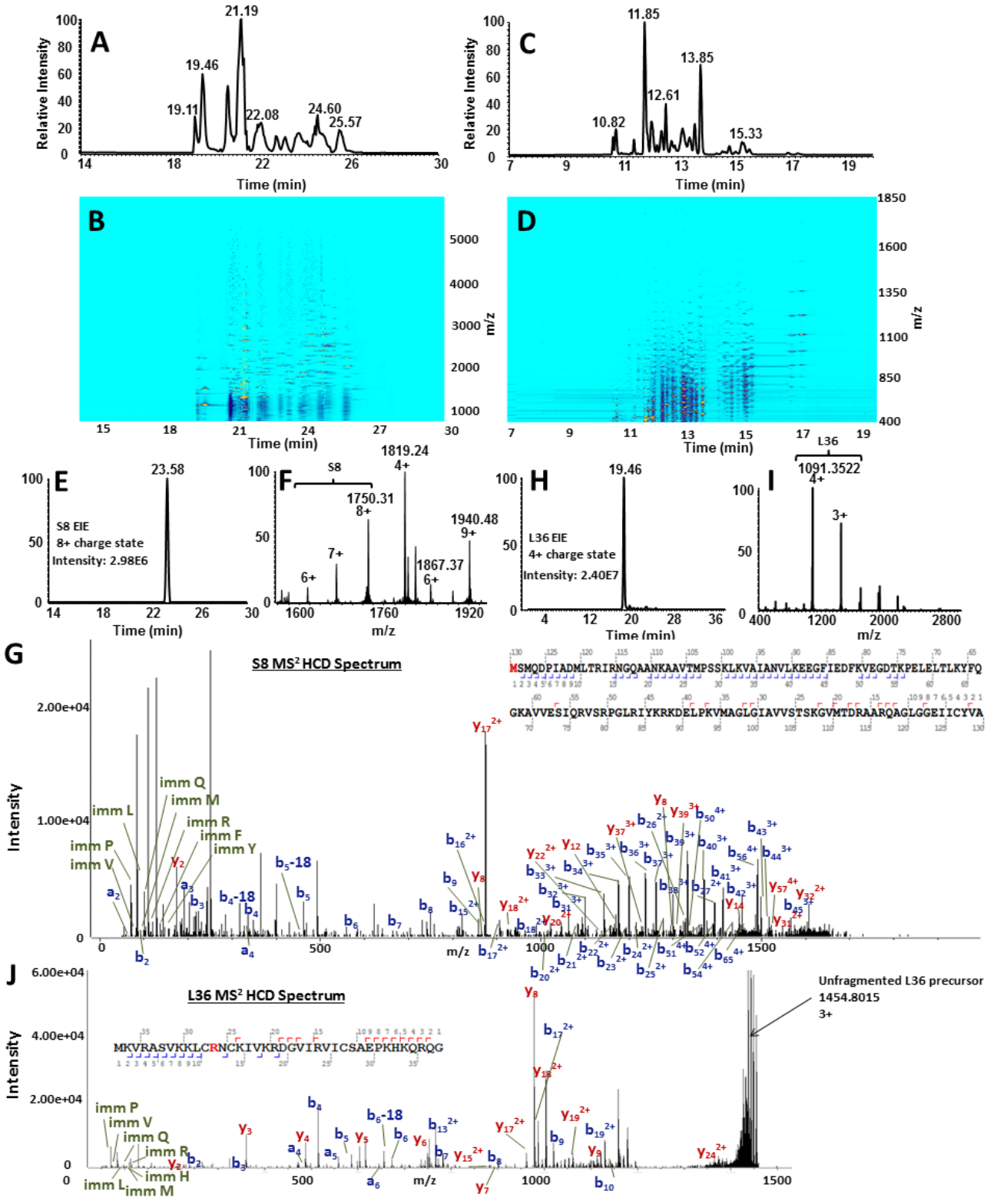
Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	12710.82	57665961.00	4	13.400-13.590	Denaturing CZE-MS
	12652.60	84901416.00	8	15.060-15.260	Denaturing CZE-MS
	12435.83	4422934.50	6	17.660-17.910	Denaturing CZE-MS
	12323.85	26504069.00	5	12.830-13.020	Denaturing CZE-MS
	12294.68	6238388.56	6	16.780-17.080	Denaturing CZE-MS
	12237.02	5201634.56	4	18.540-18.800	Denaturing CZE-MS
	12235.61	13387569.69	8	17.090-17.860	Denaturing CZE-MS
L22	12225.62	690475806.50	7	12.770-13.070	Denaturing CZE-MS
	12220.77	27723834.88	9	17.550-17.910	Denaturing CZE-MS
	12219.66	195138345.75	11	17.040-17.600	Denaturing CZE-MS
L7	12205.25	599049025.88	11	16.980-17.860	Denaturing CZE-MS
	12189.93	34792117.50	8	17.240-17.600	Denaturing CZE-MS
	12177.44	401887723.78	10	16.620-17.860	Denaturing CZE-MS
	12176.70	7676833.31	7	17.240-17.910	Denaturing CZE-MS
	12175.99	18880614.81	7	17.290-17.600	Denaturing CZE-MS
L7/L12	12163.18	415907160.16	10	16.670-17.650	Denaturing CZE-MS
L7/L12	12161.59	23075741.00	7	17.240-17.600	Denaturing CZE-MS
	12147.42	44006957.47	8	16.830-17.550	Denaturing CZE-MS
	12133.26	15670556.47	7	16.830-17.500	Denaturing CZE-MS
	12132.11	5605449.88	6	17.240-17.440	Denaturing CZE-MS
	12124.97	16088129.25	4	12.870-13.230	Denaturing CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	12119.20	23545598.06	8	16.830-17.180	Denaturing CZE-MS
	12105.24	12394159.44	7	16.830-17.180	Denaturing CZE-MS
	12090.14	5200681.00	4	16.880-17.130	Denaturing CZE-MS
	11979.70	6729690.72	3	17.550-18.690	Denaturing CZE-MS
	11750.54	20960308.00	6	14.490-14.790	Denaturing CZE-MS
S10	11734.84	557437970.75	7	14.020-14.790	Denaturing CZE-MS
	11719.43	29411352.94	9	14.490-15.260	Denaturing CZE-MS
	11676.85	14520549.38	6	14.490-14.740	Denaturing CZE-MS
S14	11448.64	195878158.00	5	12.310-12.500	Denaturing CZE-MS
	11290.03	5774069.50	3	13.500-14.010	Denaturing CZE-MS
L23	11198.43	697205888.00	6	13.240-13.440	Denaturing CZE-MS
L24	11183.82	190599944.00	5	13.140-13.330	Denaturing CZE-MS
L25	10692.55	507477452.00	5	13.550-13.800	Denaturing CZE-MS
	10298.55	169238863.00	4	11.990-12.240	Denaturing CZE-MS
	10235.21	32917611.50	4	12.980-13.180	Denaturing CZE-MS
S15	10136.88	439132412.00	6	12.980-13.180	Denaturing CZE-MS
S17	9552.60	120306507.00	3	11.890-12.090	Denaturing CZE-MS
S16	9189.86	777203660.00	5	13.760-13.960	Denaturing CZE-MS
S16	9189.54	55140495.14	4	13.710-14.270	Denaturing CZE-MS
S18	8992.47	46443304.75	4	11.630-12.140	Denaturing CZE-MS
	7844.28	8951025.25	3	6.740-6.938	Denaturing CZE-MS

Protein Name	M average (deconvoluted)	Sum Intensity	# of Charge States	RT Range	Experiment
	7754.44	17534255.50	4	13.030-13.230	Denaturing CZE-MS
	7434.98	56903906.38	4	14.490-15.460	Denaturing CZE-MS
	7427.07	13329592.19	4	15.940-16.760	Denaturing CZE-MS
	7420.13	5656678.39	4	15.940-16.350	Denaturing CZE-MS
	7370.74	133775840.00	3	13.760-13.960	Denaturing CZE-MS
	7331.29	16194283.75	5	16.830-17.080	Denaturing CZE-MS
L29	7272.64	3561006204.00	4	13.710-14.010	Denaturing CZE-MS
	7270.50	13046205.38	5	17.350-17.550	Denaturing CZE-MS
	7224.98	5140978.75	3	11.520-11.720	Denaturing CZE-MS
	6975.46	20130932.50	3	15.480-15.730	Denaturing CZE-MS
	6483.83	25412061.00	3	12.310-12.550	Denaturing CZE-MS
L33	6250.80	40278148.28	3	10.140-12.550	Denaturing CZE-MS
	6102.55	5411465.03	3	17.140-17.550	Denaturing CZE-MS
	6088.68	27893597.22	3	16.830-17.130	Denaturing CZE-MS
	6081.66	58625182.19	3	16.830-17.180	Denaturing CZE-MS

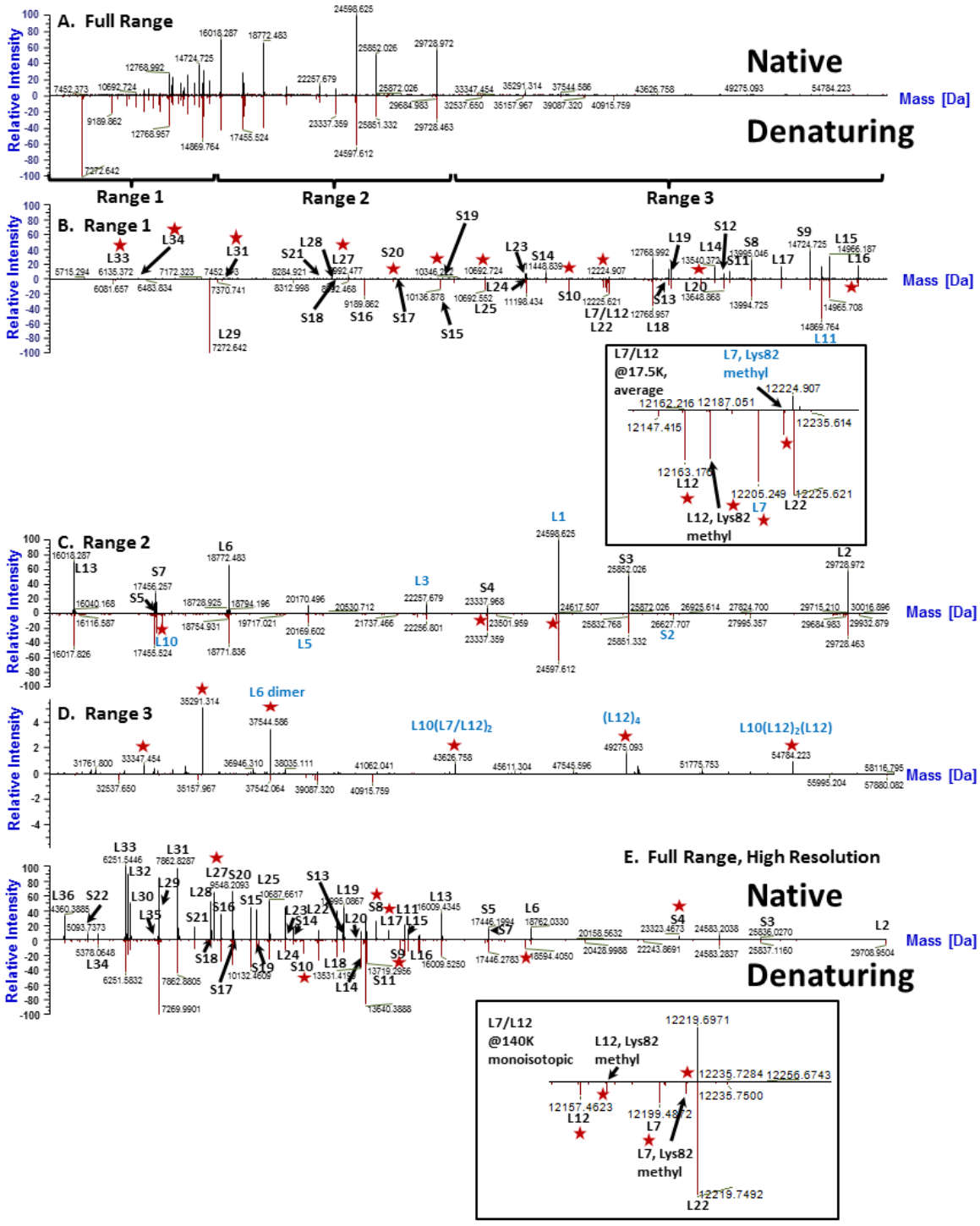
Table 4-5: Summary of all species identified from native and denaturing CZE-MS[1] analyses in the *E. coli* ribosomal organellar proteome. Deconvolution was performed with Protein Deconvolution 4.0 using the ReSpect scheme with the Sliding Window algorithm. Since non-isotopically-resolved data was used, deconvoluted species were reported as average molecular masses. Identifications of specific proteins were made on the basis of molecular mass comparisons to masses of ribosomal proteins identified in CZE-MS[2] (see Supplementary Table 4). Only identifications on non-modified ribosomal proteins and protein complexes were labeled in the left column of the table.

Next, we conducted conventional CZE-MS characterization under denaturing conditions (**Figure 4-10 C-D**), and a comparison of RP species identified through deconvolution of native and denaturing CZE-MS^1 data is shown in **Figure 4-11**. The number of molecular features detected at the low molecular weight portion of the deconvoluted spectrum (<15,000 Da) under both conditions CZE-MS^1 experiments was comparable (**Figure 4-11 A-B**). However, in the higher molecular weight region (>30,000 Da), the native condition resulted in the detection of a larger number of molecular ion species, potentially corresponding to non-covalent protein complexes (**Figure 4-11 A** and **C**). Protein-protein interactions, especially in the stalk complex of the ribosome, are not well-characterized, which made assignment of high molecular weight species difficult (**Figure 4-11D**) [351]. These results were then supplemented with a denaturing CZE-MS^2 experiment (**Table 4-4**) where a lower number of species were found. We explain this by the "dilution" of the analyte signal across more charge states and the reduced spacing between protein ion species of different charges as well as between species with subtle mass differences, resulting from higher net charges under denaturing conditions. Nevertheless, the denaturing CZE-MS^2 experiments resulted in identification of three additional RPs not found under native conditions, demonstrating the complementarity of the approaches. In total, the combination of native and denaturing CZE-MS approaches resulted in the identification of 51 out of 55 known proteins of the *E. coli* ribosome, close to 200 different proteoforms, and 679 PfSMs, which was unachievable in direct infusion experiments.



Figures 4-10 A-J. **Comparison of native and denaturing CZE-MS analysis of ribosomal proteins.** Combined results of native and denaturing CZE-MS analyses of ribosomal proteins on the polyacrylamide-coated capillary resulted in the total identification of 51 RPs. Panel (**A**)

shows the total ion electropherogram for the experiment performed under native conditions. The ion density map for the native experiment, showing the migration time (horizontal axis) vs. *m/z* (vertical axis), is shown in (**B**). The total ion electropherogram for the denaturing separation of RPs separation with a 3% acetic acid BGE (pH 2.5) is shown in (**C**) and the ion density map is shown in (**D**). An example of RP characterization using the developed workflow is demonstarated for S8 (14.0 kDa). Native CZE-MS^1 analysis resulted in the migration of S8 at 23.58 minutes, represented by the extracted ion electropherogram for the most prominent charge state (+8) (**E**). The ESI mass spectrum for the entire S8-corresponding electrophoretic peak is revealed in (**F**). Fragmentation of S8 *via* DDA-HCD resulted in unambiguous identification of S8 with a N-terminal Met-loss. Detected *b*- and *y*- ion distributions are shown in (**G**). Another example is shown for RP L36. The extracted ion electropherogram for L36 is shown in (**H**), followed by the corresponding ESI mass spectrum (**I**) and the CZE-MS^2 spectrum with identified fragment ion series (**J**).



Figures 4-11 A-E. **Comparison of ribosomal protein identificaiton between native and denaturing CZE-MS analyses.** CZE-MS[1] analyses of proteins extracted from the *E. coli* ribosomal isolate were conducted under native and denaturing conditions using ammonium acetate (pH 7.5) and 3% acetic acid (pH 2.5) BGEs, respectively. MS data acquisition was performed using either low or high resolving power (17,500 or 140,000 at 200 *Th*, respectively)

on a Q Exactive Plus instrument. (**A**). A comparison of deconvolution results between native (*all top panels*) and denaturing (*all bottom panels*) CZE-MS^1 analyses resulted in a similar distribution and quantity of identified proteins in mass Ranges 1 and 2 (4,000 – 15,000 Da and 15,000 – 31,000 Da) (**B-C**). However, for higher molecular weight species, native separations resulted in the identification of more peaks than the corresponding denaturing run (**D**). For CZE separations performed with high resolution MS data acquisition, numerous low MW proteins were identified that were not observed through deconvolution of low resolution spectra (**E**). Specifically, L36, S22, and L34 – all below 5,550 Da – were readily identified. Identifications of specific ribosomal proteins were achieved either *via* Byonic analysis of CZE-MS^2 data (*black labels*) or *via* manual interpretation of CZE-MS^1 deconvolution data (*blue labels*). Peaks elucidated by one set of conditions (native or denaturing) and not the other are annotated with red stars. For both low resolution and high resolution experiments, zoom-in views of the region corresponding to proteins L7/L12 and L22 (12,150 Da – 12,270 Da) are shown as separate inserts.

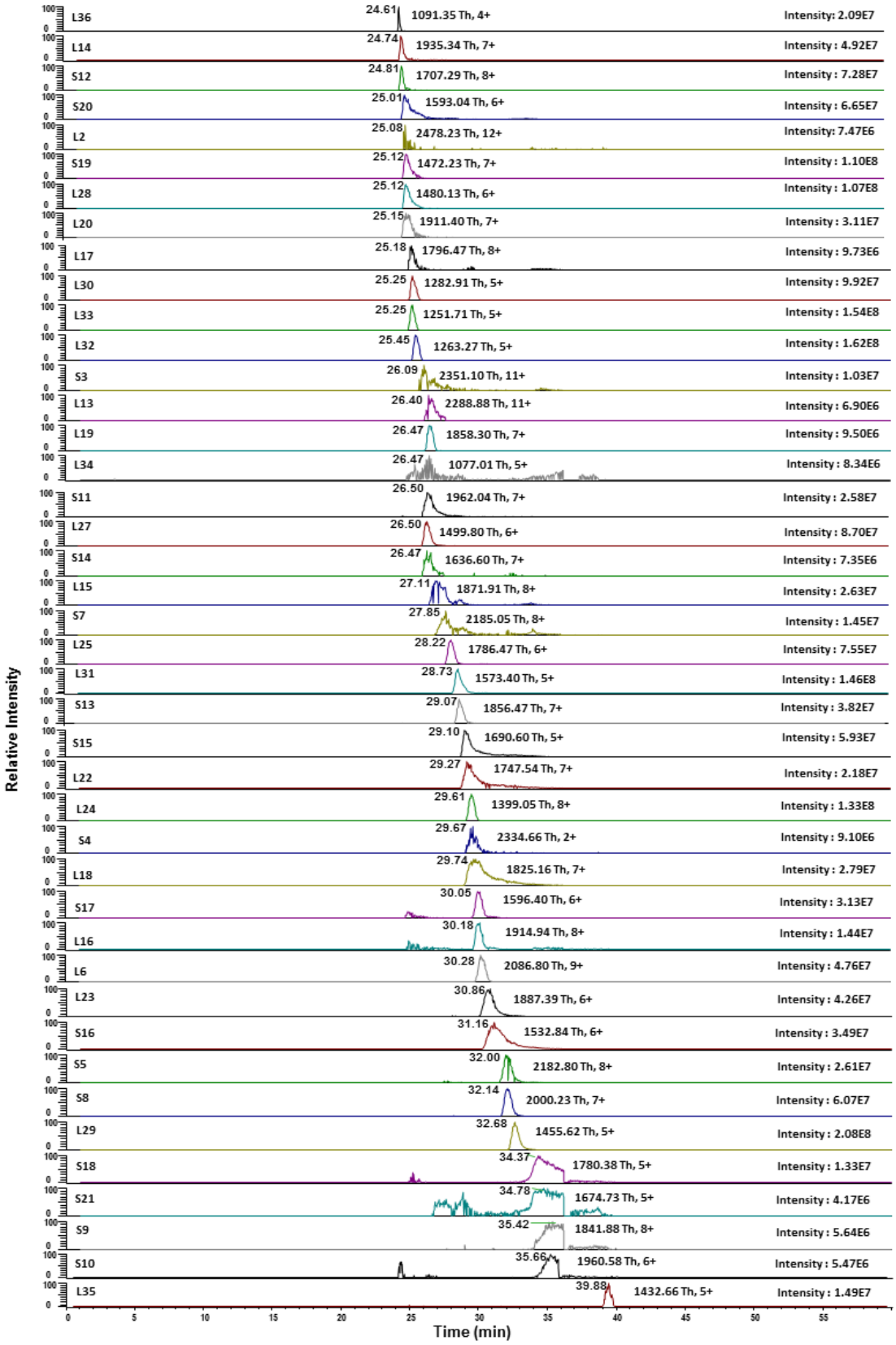

L36 24.61 1091.35 Th, 4+ Intensity: 2.09E7
L14 24.74 1935.34 Th, 7+ Intensity : 4.92E7
S12 24.81 1707.29 Th, 8+ Intensity : 7.28E7
S20 25.01 1593.04 Th, 6+ Intensity : 6.65E7
L2 25.08 2478.23 Th, 12+ Intensity: 7.47E6
S19 25.12 1472.23 Th, 7+ Intensity : 1.10E8
L28 25.12 1480.13 Th, 6+ Intensity : 1.07E8
L20 25.15 1911.40 Th, 7+ Intensity : 3.11E7
L17 25.18 1796.47 Th, 8+ Intensity : 9.73E6
L30 25.25 1282.91 Th, 5+ Intensity : 9.92E7
L33 25.25 1251.71 Th, 5+ Intensity : 1.54E8
L32 25.45 1263.27 Th, 5+ Intensity : 1.62E8
S3 26.09 2351.10 Th, 11+ Intensity : 1.03E7
L13 26.40 2288.88 Th, 11+ Intensity : 6.90E6
L19 26.47 1858.30 Th, 7+ Intensity : 9.50E6
L34 26.47 1077.01 Th, 5+ Intensity : 8.34E6
S11 26.50 1962.04 Th, 7+ Intensity : 2.58E7
L27 26.50 1499.80 Th, 6+ Intensity : 8.70E7
S14 26.47 1636.60 Th, 7+ Intensity : 7.35E6
L15 27.11 1871.91 Th, 8+ Intensity : 2.63E7
S7 27.85 2185.05 Th, 8+ Intensity : 1.45E7
L25 28.22 1786.47 Th, 6+ Intensity : 7.55E7
L31 28.73 1573.40 Th, 5+ Intensity : 1.46E8
S13 29.07 1856.47 Th, 7+ Intensity : 3.82E7
S15 29.10 1690.60 Th, 5+ Intensity : 5.93E7
L22 29.27 1747.54 Th, 7+ Intensity : 2.18E7
L24 29.61 1399.05 Th, 8+ Intensity : 1.33E8
S4 29.67 2334.66 Th, 2+ Intensity : 9.10E6
L18 29.74 1825.16 Th, 7+ Intensity : 2.79E7
S17 30.05 1596.40 Th, 6+ Intensity : 3.13E7
L16 30.18 1914.94 Th, 8+ Intensity : 1.44E7
L6 30.28 2086.80 Th, 9+ Intensity : 4.76E7
L23 30.86 1887.39 Th, 6+ Intensity : 4.26E7
S16 31.16 1532.84 Th, 6+ Intensity : 3.49E7
S5 32.00 2182.80 Th, 8+ Intensity : 2.61E7
S8 32.14 2000.23 Th, 7+ Intensity : 6.07E7
L29 32.68 1455.62 Th, 5+ Intensity : 2.08E8
S18 34.37 1780.38 Th, 5+ Intensity : 1.33E7
S21 34.78 1674.73 Th, 5+ Intensity : 4.17E6
S9 35.42 1841.88 Th, 8+ Intensity : 5.64E6
S10 35.66 1960.58 Th, 6+ Intensity : 5.47E6
L35 39.88 1432.66 Th, 5+ Intensity : 1.49E7
Relative Intensity
Time (min)
0 5 10 15 20 25 30 35 40 45 50 55

Figure 4-12. Extracted ion electropherograms of ribosomal proteins identified by native CZE-MS^2. Native CZE-MS^2 analysis of the protein extract from the *E. coli* ribosomal isolate with DDA-HCD-MS^2 fragmentation on the Q Exactive Plus MS resulted in identification of 42 distinct protein groups in a single run. Extracted ion electropherograms of the highest-abundance charge states and proteoforms of ribosomal proteins, designated "S" and "L" for their origins from small (30S) and large (50S) ribosomal subunits, respectively, are shown in order of migration. Signal intensities for extracted ions are also displayed.

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L2	R[+14]	29708.728	29725.994	-7.0	29726.246	29743.523	Native CZE-MS2
S3	K[+14]	25839.044	25851.152	-2.6	25854.149	25866.265	Native CZE-MS2
S4	K[+14]	23340.599	23338.611	-0.8	23354.147	23352.157	Native CZE-MS2
S4		23323.546	23324.596	-2.0	23337.083	23338.133	Native CZE-MS2
L6	D[+24]	18783.055	18786.113	-2.5	18793.760	18796.820	Native CZE-MS2
L6	E[+24]	18782.108	18786.113	0.4	18792.812	18796.820	Native CZE-MS2
L6	K[+14]	18780.034	18776.144	-6.6	18790.737	18786.845	Native CZE-MS2
L6	E[+24]	18775.983	18786.113	-5.2	18786.683	18796.820	Native CZE-MS2
L6		18762.094	18762.128	-1.8	18772.786	18772.820	Native CZE-MS2
L6		18762.070	18762.128	-3.1	18772.762	18772.820	Native CZE-MS2
L6		18762.070	18762.128	-3.1	18772.762	18772.820	Native CZE-MS2
L6		18762.053	18762.128	-4.0	18772.745	18772.820	Native CZE-MS2
L6		18762.008	18762.128	-6.4	18772.700	18772.820	Native CZE-MS2
L6		18761.061	18762.128	-3.4	18771.752	18772.820	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S7	R[+14]	17465.304	17477.430	-4.9	17475.187	17487.321	Native CZE-MS2
S7		17464.381	17463.414	-2.1	17474.263	17473.296	Native CZE-MS2
S7		17464.381	17463.414	-2.1	17474.263	17473.296	Native CZE-MS2
S7		17463.339	17463.414	-4.3	17473.221	17473.296	Native CZE-MS2
S7		17463.315	17463.414	-5.7	17473.197	17473.296	Native CZE-MS2
S5	.[+42]	17447.378	17446.299	4.3	17457.250	17456.171	Native CZE-MS2
S5	.[+42]	17446.482	17446.299	10.5	17456.354	17456.171	Native CZE-MS2
S5	.[+42]	17446.482	17446.299	10.5	17456.354	17456.171	Native CZE-MS2
S5	.[+42]	17446.394	17446.299	5.4	17456.265	17456.171	Native CZE-MS2
S5	.[+42]	17446.370	17446.299	4.0	17456.241	17456.171	Native CZE-MS2
S5	.[+42]	17446.317	17446.299	1.0	17456.189	17456.171	Native CZE-MS2
S5	.[+42]	17445.347	17446.299	2.9	17455.218	17456.171	Native CZE-MS2
S5	.[+42]	17445.347	17446.299	2.9	17455.218	17456.171	Native CZE-MS2
L13	K[+28]	16029.512	16037.581	-2.6	16038.500	16046.574	Native CZE-MS2
L13	R[+14]	16029.512	16023.565	-4.5	16038.500	16032.549	Native CZE-MS2
L13	K[+28]	16025.509	16037.581	-2.0	16034.494	16046.574	Native CZE-MS2
L13		16011.673	16009.549	7.3	16020.650	16018.525	Native CZE-MS2
L13		16011.673	16009.549	7.3	16020.650	16018.525	Native CZE-MS2
L13		16010.586	16009.549	2.1	16019.562	16018.525	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L13		16009.661	16009.549	7.0	16018.636	16018.525	Native CZE-MS2
L13		16009.625	16009.549	4.7	16018.600	16018.525	Native CZE-MS2
L13		16009.625	16009.549	4.7	16018.600	16018.525	Native CZE-MS2
L13		16009.571	16009.549	1.4	16018.547	16018.525	Native CZE-MS2
L13		16009.571	16009.549	1.4	16018.547	16018.525	Native CZE-MS2
L13		16009.571	16009.549	1.4	16018.547	16018.525	Native CZE-MS2
L13		16009.571	16009.549	1.4	16018.547	16018.525	Native CZE-MS2
L13		16009.571	16009.549	1.4	16018.547	16018.525	Native CZE-MS2
L13		16008.678	16009.549	8.2	16017.653	16018.525	Native CZE-MS2
L13		16007.527	16009.549	-1.0	16016.501	16018.525	Native CZE-MS2
L13		16007.527	16009.549	-1.0	16016.501	16018.525	Native CZE-MS2
L13		16007.527	16009.549	-1.0	16016.501	16018.525	Native CZE-MS2
L16	D[+64]	15337.292	15336.395	-6.9	15345.848	15344.950	Native CZE-MS2
L16	.[+42]	15320.398	15314.477	-6.5	15328.944	15323.019	Native CZE-MS2
L16	K[+42]	15319.518	15314.477	1.5	15328.063	15323.019	Native CZE-MS2
L16	.[+42]	15319.480	15314.477	-0.9	15328.026	15323.019	Native CZE-MS2
L16	.[+42]	15319.480	15314.477	-0.9	15328.026	15323.019	Native CZE-MS2
L16	.[+42]	15319.480	15314.477	-0.9	15328.026	15323.019	Native CZE-MS2
L16	.[+42]	15319.480	15314.477	-0.9	15328.026	15323.019	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L16	.[+42]	15319.390	15314.477	-6.8	15327.935	15323.019	Native CZE-MS2
L16	.[+42]	15319.325	15314.477	-11.0	15327.870	15323.019	Native CZE-MS2
L16	.[+42]	15319.315	15314.477	-11.7	15327.860	15323.019	Native CZE-MS2
L16	.[+42]	15318.386	15314.477	-6.8	15326.930	15323.019	Native CZE-MS2
L16	.[+42]	15318.386	15314.477	-6.8	15326.930	15323.019	Native CZE-MS2
L16	.[+42]	15318.386	15314.477	-6.8	15326.930	15323.019	Native CZE-MS2
L16	.[+42]	15318.386	15314.477	-6.8	15326.930	15323.019	Native CZE-MS2
L16	.[+42]	15318.386	15314.477	-6.8	15326.930	15323.019	Native CZE-MS2
L16	.[+42]	15318.386	15314.477	-6.8	15326.930	15323.019	Native CZE-MS2
L16	.[+42]	15317.502	15314.477	1.0	15326.046	15323.019	Native CZE-MS2
L16	.[+42]	15303.545	15314.477	6.8	15312.080	15323.019	Native CZE-MS2
L16	K[+14]	15303.417	15286.482	-8.0	15311.952	15295.007	Native CZE-MS2
L16	K[+14]	15303.417	15286.482	-8.0	15311.952	15295.007	Native CZE-MS2
L16	K[+14]	15303.417	15286.482	-8.0	15311.952	15295.007	Native CZE-MS2
L16	K[+14]	15303.417	15286.482	-8.0	15311.952	15295.007	Native CZE-MS2
L16	K[+14]	15303.417	15286.482	-8.0	15311.952	15295.007	Native CZE-MS2
L16	K[+14]	15303.417	15286.482	-8.0	15311.952	15295.007	Native CZE-MS2
L16	K[+14]	15303.340	15286.482	-13.0	15311.875	15295.007	Native CZE-MS2
L16	K[+14]	15302.513	15286.482	-1.5	15311.047	15295.007	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L16	K[+14]	15302.513	15286.482	-1.5	15311.047	15295.007	Native CZE-MS2
L16	K[+14]	15302.513	15286.482	-1.5	15311.047	15295.007	Native CZE-MS2
L16	K[+14]	15302.406	15286.482	-8.5	15310.941	15295.007	Native CZE-MS2
L16	K[+14]	15302.324	15286.482	-13.9	15310.858	15295.007	Native CZE-MS2
L16	K[+14]	15300.541	15286.482	0.8	15309.074	15295.007	Native CZE-MS2
L16	K[+14]	15300.541	15286.482	0.8	15309.074	15295.007	Native CZE-MS2
L16	K[+14]	15300.541	15286.482	0.8	15309.074	15295.007	Native CZE-MS2
L16	K[+14]	15300.541	15286.482	0.8	15309.074	15295.007	Native CZE-MS2
L16	K[+14]	15287.342	15286.482	-9.4	15295.867	15295.007	Native CZE-MS2
L16	K[+14]	15287.342	15286.482	-9.4	15295.867	15295.007	Native CZE-MS2
L16	K[+14]	15287.342	15286.482	-9.4	15295.867	15295.007	Native CZE-MS2
L16	K[+14]	15287.342	15286.482	-9.4	15295.867	15295.007	Native CZE-MS2
L16	R[+14]	15287.342	15286.482	-9.4	15295.867	15295.007	Native CZE-MS2
L16	K[+14]	15287.327	15286.482	-10.4	15295.852	15295.007	Native CZE-MS2
L15	K[+14]	14959.243	14972.330	-2.9	14967.563	14980.659	Native CZE-MS2
L15		14959.193	14958.315	-8.3	14967.514	14966.635	Native CZE-MS2
L15	K[+14]	14959.193	14972.330	-6.2	14967.514	14980.659	Native CZE-MS2
L15	K[+14]	14959.193	14972.330	-6.2	14967.514	14980.659	Native CZE-MS2
L15	K[+14]	14959.193	14972.330	-6.2	14967.514	14980.659	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L15		14958.241	14958.315	-5.0	14966.560	14966.635	Native CZE-MS2
L15		14958.230	14958.315	-5.6	14966.550	14966.635	Native CZE-MS2
S9		14718.962	14716.989	-2.3	14727.132	14725.158	Native CZE-MS2
L17	E[+64]	14419.544	14420.593	-3.1	14427.528	14428.577	Native CZE-MS2
L17	K[+42]	14400.653	14398.675	-2.0	14408.625	14406.646	Native CZE-MS2
L17	K[+28]	14399.512	14384.696	-16.3	14407.483	14392.658	Native CZE-MS2
L17	R[+14]	14380.598	14370.680	-8.1	14388.558	14378.634	Native CZE-MS2
L17	K[+28]	14379.657	14384.696	-1.6	14387.616	14392.658	Native CZE-MS2
L17	K[+14]	14379.604	14370.680	-7.4	14387.562	14378.634	Native CZE-MS2
L17	K[+28]	14378.655	14384.696	-1.4	14386.613	14392.658	Native CZE-MS2
L17	K[+28]	14378.655	14384.696	-1.4	14386.613	14392.658	Native CZE-MS2
L17	K[+28]	14377.655	14384.696	-1.2	14385.613	14392.658	Native CZE-MS2
L17	K[+14]	14377.614	14370.680	-6.2	14385.572	14378.634	Native CZE-MS2
L17	K[+28]	14377.614	14384.696	-4.0	14385.572	14392.658	Native CZE-MS2
L17	K[+14]	14377.587	14370.680	-8.1	14385.545	14378.634	Native CZE-MS2
L17	K[+28]	14376.652	14384.696	-1.2	14384.609	14392.658	Native CZE-MS2
L17	K[+28]	14376.652	14384.696	-1.2	14384.609	14392.658	Native CZE-MS2
L17	K[+14]	14374.633	14370.680	-4.2	14382.589	14378.634	Native CZE-MS2
L17		14357.636	14356.665	-2.2	14365.581	14364.609	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L17		14357.558	14356.665	-7.7	14365.503	14364.609	Native CZE-MS2
L17		14357.558	14356.665	-7.7	14365.503	14364.609	Native CZE-MS2
L17		14357.556	14356.665	-7.8	14365.501	14364.609	Native CZE-MS2
L17		14357.556	14356.665	-7.8	14365.501	14364.609	Native CZE-MS2
L17		14357.556	14356.665	-7.8	14365.501	14364.609	Native CZE-MS2
L17		14357.556	14356.665	-7.8	14365.501	14364.609	Native CZE-MS2
L17		14357.556	14356.665	-7.8	14365.501	14364.609	Native CZE-MS2
L17		14356.684	14356.665	1.4	14364.629	14364.609	Native CZE-MS2
L17		14356.684	14356.665	1.4	14364.629	14364.609	Native CZE-MS2
L17		14356.684	14356.665	1.4	14364.629	14364.609	Native CZE-MS2
L17		14356.672	14356.665	0.5	14364.617	14364.609	Native CZE-MS2
L17		14356.622	14356.665	-3.0	14364.566	14364.609	Native CZE-MS2
L17		14355.702	14356.665	2.8	14363.646	14364.609	Native CZE-MS2
L17		14355.697	14356.665	2.5	14363.641	14364.609	Native CZE-MS2
S8	K[+28]	14010.465	14015.519	-2.6	14018.194	14023.251	Native CZE-MS2
S8	K[+28]	14010.465	14015.519	-2.6	14018.194	14023.251	Native CZE-MS2
S8	K[+28]	14010.465	14015.519	-2.6	14018.194	14023.251	Native CZE-MS2
S8	K[+28]	14007.519	14015.519	1.9	14015.246	14023.251	Native CZE-MS2
S8	E[+24]	14007.519	14011.473	4.3	14015.246	14019.202	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S8	K[+14]	14005.518	14001.503	0.1	14013.244	14009.226	Native CZE-MS2
S8	R[+14]	14004.450	14001.503	-4.5	14012.175	14009.226	Native CZE-MS2
S8	E[+24]	14003.547	14011.473	7.2	14011.271	14019.202	Native CZE-MS2
S8		13988.532	13987.487	2.9	13996.247	13995.202	Native CZE-MS2
S8		13988.510	13987.487	1.4	13996.225	13995.202	Native CZE-MS2
S8		13988.501	13987.487	0.7	13996.216	13995.202	Native CZE-MS2
S8		13988.432	13987.487	-4.2	13996.147	13995.202	Native CZE-MS2
S8		13987.604	13987.487	8.4	13995.319	13995.202	Native CZE-MS2
S8		13987.604	13987.487	8.4	13995.319	13995.202	Native CZE-MS2
S8		13987.604	13987.487	8.4	13995.319	13995.202	Native CZE-MS2
S8		13987.604	13987.487	8.4	13995.319	13995.202	Native CZE-MS2
S8		13987.563	13987.487	5.4	13995.277	13995.202	Native CZE-MS2
S8		13987.563	13987.487	5.4	13995.277	13995.202	Native CZE-MS2
S8		13987.563	13987.487	5.4	13995.277	13995.202	Native CZE-MS2
S11	K[+14]	13722.285	13720.356	-5.7	13729.834	13727.904	Native CZE-MS2
S11	R[+14]	13721.277	13720.356	-6.0	13728.825	13727.904	Native CZE-MS2
S11	K[+14]	13721.276	13720.356	-6.1	13728.824	13727.904	Native CZE-MS2
S11	K[+14]	13721.257	13720.356	-7.4	13728.806	13727.904	Native CZE-MS2
S11	K[+14]	13721.257	13720.356	-7.4	13728.806	13727.904	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S11	K[+14]	13721.246	13720.356	-8.3	13728.794	13727.904	Native CZE-MS2
S11	K[+14]	13720.333	13720.356	-1.7	13727.880	13727.904	Native CZE-MS2
S12	K[+42]	13643.324	13640.479	-12.1	13650.824	13647.977	Native CZE-MS2
S12	K[+42]	13643.312	13640.479	-13.0	13650.812	13647.977	Native CZE-MS2
S12	K[+42]	13643.307	13640.479	-13.3	13650.807	13647.977	Native CZE-MS2
S12	K[+42]	13643.307	13640.479	-13.3	13650.807	13647.977	Native CZE-MS2
S12	K[+42]	13642.335	13640.479	-11.0	13649.834	13647.977	Native CZE-MS2
S12	K[+28]	13642.308	13626.499	-18.0	13649.807	13633.989	Native CZE-MS2
S12	K[+28]	13642.308	13626.499	-18.0	13649.807	13633.989	Native CZE-MS2
S12	K[+28]	13642.308	13626.499	-18.0	13649.807	13633.989	Native CZE-MS2
S12	K[+42]	13642.305	13640.479	-13.2	13649.804	13647.977	Native CZE-MS2
S12	K[+42]	13642.305	13640.479	-13.2	13649.804	13647.977	Native CZE-MS2
S12	K[+28]	13625.302	13626.499	-14.2	13632.791	13633.989	Native CZE-MS2
L14		13534.331	13533.458	-9.6	13541.762	13540.889	Native CZE-MS2
L14		13533.357	13533.458	-7.4	13540.788	13540.889	Native CZE-MS2
L14		13533.357	13533.458	-7.4	13540.788	13540.889	Native CZE-MS2
L20	K[+28]	13373.647	13386.670	1.5	13380.978	13394.010	Native CZE-MS2
L20		13358.731	13358.639	6.9	13366.053	13365.961	Native CZE-MS2
L20		13357.665	13358.639	2.2	13364.986	13365.961	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L19	K[+42]	13036.174	13037.168	0.7	13043.295	13044.290	Native CZE-MS2
L19	E[+24]	13019.083	13019.143	-4.6	13026.194	13026.253	Native CZE-MS2
L19		12996.137	12995.158	-1.9	13003.233	13002.253	Native CZE-MS2
L19		12996.137	12995.158	-1.9	13003.233	13002.253	Native CZE-MS2
S13	;[+24]	12995.186	12985.224	-5.5	13002.282	12992.313	Native CZE-MS2
L19		12995.119	12995.158	-3.0	13002.215	13002.253	Native CZE-MS2
L19		12995.096	12995.158	-4.8	13002.191	13002.253	Native CZE-MS2
L19		12995.090	12995.158	-5.2	13002.185	13002.253	Native CZE-MS2
L19		12995.090	12995.158	-5.2	13002.185	13002.253	Native CZE-MS2
L19		12995.088	12995.158	-5.4	13002.184	13002.253	Native CZE-MS2
L19		12995.086	12995.158	-5.6	13002.181	13002.253	Native CZE-MS2
L19		12995.086	12995.158	-5.6	13002.181	13002.253	Native CZE-MS2
L19		12995.086	12995.158	-5.6	13002.181	13002.253	Native CZE-MS2
L19		12995.086	12995.158	-5.6	13002.181	13002.253	Native CZE-MS2
S13	D[+24]	12983.205	12985.224	-1.0	12990.293	12992.313	Native CZE-MS2
S13	E[+24]	12983.159	12985.224	-4.5	12990.247	12992.313	Native CZE-MS2
S13	D[+24]	12982.243	12985.224	2.2	12989.330	12992.313	Native CZE-MS2
S13	D[+24]	12982.142	12985.224	-5.5	12989.230	12992.313	Native CZE-MS2
S13	K[+14]	12980.230	12975.255	-3.2	12987.316	12982.338	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S13	K[+28]	12979.099	12989.270	-10.6	12986.184	12996.362	Native CZE-MS2
S13	D[+24]	12978.353	12985.224	11.7	12985.438	12992.313	Native CZE-MS2
S13		12963.254	12961.239	0.6	12970.330	12968.313	Native CZE-MS2
S13		12963.178	12961.239	-5.2	12970.254	12968.313	Native CZE-MS2
S13		12963.151	12961.239	-7.3	12970.227	12968.313	Native CZE-MS2
S13		12962.290	12961.239	3.7	12969.365	12968.313	Native CZE-MS2
S13		12962.191	12961.239	-4.0	12969.266	12968.313	Native CZE-MS2
S13		12962.176	12961.239	-5.1	12969.251	12968.313	Native CZE-MS2
S13		12962.176	12961.239	-5.1	12969.251	12968.313	Native CZE-MS2
S13		12961.289	12961.239	3.8	12968.363	12968.313	Native CZE-MS2
S13		12961.271	12961.239	2.5	12968.345	12968.313	Native CZE-MS2
S13		12961.263	12961.239	1.9	12968.337	12968.313	Native CZE-MS2
S13		12961.216	12961.239	-1.8	12968.290	12968.313	Native CZE-MS2
S13		12961.200	12961.239	-3.0	12968.274	12968.313	Native CZE-MS2
S13		12960.177	12961.239	-4.5	12967.251	12968.313	Native CZE-MS2
L18		12763.039	12762.939	7.8	12769.990	12769.889	Native CZE-MS2
L18		12763.031	12762.939	7.2	12769.982	12769.889	Native CZE-MS2
L18		12763.018	12762.939	6.2	12769.969	12769.889	Native CZE-MS2
L18		12763.018	12762.939	6.2	12769.969	12769.889	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L22		12219.835	12219.777	4.8	12226.447	12226.389	Native CZE-MS2
L22		12219.794	12219.777	1.4	12226.406	12226.389	Native CZE-MS2
S14		11443.159	11443.231	-6.3	11449.287	11449.359	Native CZE-MS2
S14		11443.159	11443.231	-6.3	11449.287	11449.359	Native CZE-MS2
S14		11443.139	11443.231	-8.0	11449.267	11449.359	Native CZE-MS2
S14		11443.139	11443.231	-8.0	11449.267	11449.359	Native CZE-MS2
S14		11443.139	11443.231	-8.0	11449.267	11449.359	Native CZE-MS2
S14		11443.121	11443.231	-9.6	11449.249	11449.359	Native CZE-MS2
S14		11443.121	11443.231	-9.6	11449.249	11449.359	Native CZE-MS2
S14		11443.108	11443.231	-10.7	11449.236	11449.359	Native CZE-MS2
L23	K[+28]	11216.191	11221.261	-4.8	11222.177	11227.251	Native CZE-MS2
L23	E[+24]	11214.213	11217.215	0.7	11220.198	11223.202	Native CZE-MS2
L23	K[+14]	11210.237	11207.246	-1.7	11216.220	11213.226	Native CZE-MS2
L23	K[+14]	11210.235	11207.246	-1.9	11216.217	11213.226	Native CZE-MS2
L23	E[+24]	11210.200	11217.215	0.8	11216.183	11223.202	Native CZE-MS2
L23	K[+14]	11209.144	11207.246	-9.7	11215.125	11213.226	Native CZE-MS2
L23	;[+24]	11207.251	11217.215	6.2	11213.232	11223.202	Native CZE-MS2
L24	K[+14]	11202.340	11193.292	1.6	11208.317	11199.264	Native CZE-MS2
L24	K[+14]	11202.296	11193.292	-2.3	11208.274	11199.264	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L24	E[+24]	11201.310	11203.262	4.9	11207.287	11209.240	Native CZE-MS2
L24	K[+28]	11201.310	11207.308	2.0	11207.287	11213.288	Native CZE-MS2
L24	K[+14]	11200.320	11193.292	0.4	11206.296	11199.264	Native CZE-MS2
L24	K[+14]	11200.320	11193.292	0.4	11206.296	11199.264	Native CZE-MS2
L24	K[+28]	11200.320	11207.308	3.2	11206.296	11213.288	Native CZE-MS2
L23		11194.194	11193.230	-3.5	11200.166	11199.202	Native CZE-MS2
L23		11193.286	11193.230	5.0	11199.258	11199.202	Native CZE-MS2
L23		11193.281	11193.230	4.5	11199.252	11199.202	Native CZE-MS2
L23	K[+14]	11193.281	11207.246	7.3	11199.252	11213.226	Native CZE-MS2
L23		11193.244	11193.230	1.2	11199.216	11199.202	Native CZE-MS2
L23		11193.232	11193.230	0.1	11199.204	11199.202	Native CZE-MS2
L23		11193.227	11193.230	-0.3	11199.199	11199.202	Native CZE-MS2
L23		11193.224	11193.230	-0.6	11199.196	11199.202	Native CZE-MS2
L24		11181.351	11179.277	6.1	11187.315	11185.240	Native CZE-MS2
L24		11181.347	11179.277	5.7	11187.312	11185.240	Native CZE-MS2
L24		11180.367	11179.277	7.8	11186.331	11185.240	Native CZE-MS2
L24		11179.362	11179.277	7.6	11185.325	11185.240	Native CZE-MS2
L24		11179.362	11179.277	7.6	11185.325	11185.240	Native CZE-MS2
L24		11179.362	11179.277	7.6	11185.325	11185.240	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L24		11179.351	11179.277	6.7	11185.314	11185.240	Native CZE-MS2
L24		11179.351	11179.277	6.7	11185.314	11185.240	Native CZE-MS2
L24		11179.344	11179.277	6.0	11185.307	11185.240	Native CZE-MS2
L24		11179.317	11179.277	3.6	11185.280	11185.240	Native CZE-MS2
L24		11179.275	11179.277	-0.1	11185.238	11185.240	Native CZE-MS2
L24		11178.299	11179.277	2.3	11184.262	11185.240	Native CZE-MS2
L25	K[+14]	10707.778	10701.743	1.4	10713.447	10707.408	Native CZE-MS2
L25	K[+14]	10707.778	10701.743	1.4	10713.447	10707.408	Native CZE-MS2
L25	E[+24]	10700.712	10711.712	3.4	10706.377	10717.384	Native CZE-MS2
L25		10688.836	10687.727	9.9	10694.493	10693.384	Native CZE-MS2
L25		10688.803	10687.727	6.7	10694.460	10693.384	Native CZE-MS2
L25		10688.803	10687.727	6.7	10694.460	10693.384	Native CZE-MS2
L25		10688.803	10687.727	6.7	10694.460	10693.384	Native CZE-MS2
L25		10687.857	10687.727	12.1	10693.514	10693.384	Native CZE-MS2
L25		10687.842	10687.727	10.8	10693.499	10693.384	Native CZE-MS2
L25		10687.842	10687.727	10.8	10693.499	10693.384	Native CZE-MS2
L25		10687.842	10687.727	10.8	10693.499	10693.384	Native CZE-MS2
L25		10687.842	10687.727	10.8	10693.499	10693.384	Native CZE-MS2
L25		10687.842	10687.727	10.8	10693.499	10693.384	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L25		10687.834	10687.727	10.0	10693.491	10693.384	Native CZE-MS2
L25		10687.834	10687.727	10.0	10693.491	10693.384	Native CZE-MS2
L25		10687.834	10687.727	10.0	10693.491	10693.384	Native CZE-MS2
L25		10687.834	10687.727	10.0	10693.491	10693.384	Native CZE-MS2
L25		10687.812	10687.727	7.9	10693.469	10693.384	Native CZE-MS2
L25		10687.812	10687.727	7.9	10693.469	10693.384	Native CZE-MS2
L25		10687.769	10687.727	3.9	10693.426	10693.384	Native CZE-MS2
L25		10687.769	10687.727	3.9	10693.426	10693.384	Native CZE-MS2
L25		10686.796	10687.727	6.7	10692.452	10693.384	Native CZE-MS2
S19		10294.611	10293.682	-7.2	10300.022	10299.093	Native CZE-MS2
S19		10294.590	10293.682	-9.3	10300.001	10299.093	Native CZE-MS2
S19		10293.588	10293.682	-9.1	10298.999	10299.093	Native CZE-MS2
S19		10293.586	10293.682	-9.3	10298.997	10299.093	Native CZE-MS2
S19		10293.581	10293.682	-9.7	10298.992	10299.093	Native CZE-MS2
S19		10293.581	10293.682	-9.7	10298.992	10299.093	Native CZE-MS2
S19		10293.581	10293.682	-9.8	10298.992	10299.093	Native CZE-MS2
S19		10293.579	10293.682	-10.0	10298.990	10299.093	Native CZE-MS2
S19		10293.579	10293.682	-10.0	10298.990	10299.093	Native CZE-MS2
S19		10293.575	10293.682	-10.4	10298.985	10299.093	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S19		10292.658	10293.682	-2.0	10298.068	10299.093	Native CZE-MS2
S19		10292.658	10293.682	-2.0	10298.068	10299.093	Native CZE-MS2
S19		10292.658	10293.682	-2.0	10298.068	10299.093	Native CZE-MS2
S15		10132.574	10132.473	9.9	10137.884	10137.783	Native CZE-MS2
S15		10132.562	10132.473	8.8	10137.872	10137.783	Native CZE-MS2
S15		10132.546	10132.473	7.2	10137.856	10137.783	Native CZE-MS2
S15		10132.546	10132.473	7.2	10137.856	10137.783	Native CZE-MS2
S15		10132.539	10132.473	6.5	10137.849	10137.783	Native CZE-MS2
S15		10132.530	10132.473	5.6	10137.840	10137.783	Native CZE-MS2
S15		10132.530	10132.473	5.6	10137.840	10137.783	Native CZE-MS2
S15		10131.538	10132.473	6.7	10136.848	10137.783	Native CZE-MS2
S15		10131.429	10132.473	-4.0	10136.739	10137.783	Native CZE-MS2
S20	K[+14]	9571.171	9562.303	-17.0	9576.131	9567.258	Native CZE-MS2
S17		9568.327	9568.250	8.1	9573.286	9573.209	Native CZE-MS2
S17		9568.317	9568.250	7.0	9573.276	9573.209	Native CZE-MS2
S17		9568.315	9568.250	6.8	9573.274	9573.209	Native CZE-MS2
S17		9568.315	9568.250	6.8	9573.274	9573.209	Native CZE-MS2
S17		9567.304	9568.250	6.0	9572.262	9573.209	Native CZE-MS2
S20		9553.195	9548.287	-11.4	9558.144	9553.233	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S20	K[+14]	9549.212	9562.303	-5.0	9554.158	9567.258	Native CZE-MS2
S20		9548.236	9548.287	-5.4	9553.182	9553.233	Native CZE-MS2
S20		9548.210	9548.287	-8.1	9553.156	9553.233	Native CZE-MS2
S20		9548.210	9548.287	-8.1	9553.156	9553.233	Native CZE-MS2
S20		9548.206	9548.287	-8.6	9553.152	9553.233	Native CZE-MS2
S20		9548.204	9548.287	-8.8	9553.150	9553.233	Native CZE-MS2
S20		9548.202	9548.287	-8.9	9553.148	9553.233	Native CZE-MS2
S20		9548.199	9548.287	-9.2	9553.145	9553.233	Native CZE-MS2
S20		9548.194	9548.287	-9.8	9553.140	9553.233	Native CZE-MS2
S20		9547.262	9548.287	-2.3	9552.208	9553.233	Native CZE-MS2
S20		9546.200	9548.287	-8.4	9551.145	9553.233	Native CZE-MS2
S16	K[+42]	9224.953	9227.968	-0.6	9229.697	9232.714	Native CZE-MS2
S16	K[+42]	9224.953	9227.968	-0.6	9229.697	9232.714	Native CZE-MS2
S16	E[+24]	9209.948	9209.943	0.6	9214.683	9214.678	Native CZE-MS2
S16	D[+24]	9208.957	9209.943	2.0	9213.692	9214.678	Native CZE-MS2
S16	D[+24]	9207.988	9209.943	5.7	9212.722	9214.678	Native CZE-MS2
S16	E[+24]	9207.988	9209.943	5.7	9212.722	9214.678	Native CZE-MS2
S16	D[+24]	9207.967	9209.943	3.4	9212.701	9214.678	Native CZE-MS2
S16	D[+24]	9207.967	9209.943	3.4	9212.701	9214.678	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S16	D[+24]	9207.967	9209.943	3.4	9212.701	9214.678	Native CZE-MS2
S16	R[+14]	9207.967	9199.973	-3.6	9212.701	9204.702	Native CZE-MS2
S16	E[+24]	9207.967	9209.943	3.4	9212.701	9214.678	Native CZE-MS2
S16	D[+24]	9205.884	9209.943	-4.9	9210.617	9214.678	Native CZE-MS2
S16	R[+14]	9202.980	9199.973	-0.3	9207.711	9204.702	Native CZE-MS2
S16	R[+14]	9202.980	9199.973	-0.3	9207.711	9204.702	Native CZE-MS2
S16	R[+14]	9200.893	9199.973	-9.1	9205.622	9204.702	Native CZE-MS2
S16		9188.990	9185.957	2.4	9193.712	9190.678	Native CZE-MS2
S16		9188.990	9185.957	2.4	9193.712	9190.678	Native CZE-MS2
S16		9187.011	9185.957	5.5	9191.732	9190.678	Native CZE-MS2
S16		9187.011	9185.957	5.5	9191.732	9190.678	Native CZE-MS2
S16		9186.027	9185.957	7.6	9190.747	9190.678	Native CZE-MS2
S16		9186.025	9185.957	7.4	9190.745	9190.678	Native CZE-MS2
S16		9186.025	9185.957	7.4	9190.745	9190.678	Native CZE-MS2
S16		9186.021	9185.957	7.0	9190.742	9190.678	Native CZE-MS2
S16		9186.021	9185.957	7.0	9190.742	9190.678	Native CZE-MS2
S16		9186.019	9185.957	6.7	9190.739	9190.678	Native CZE-MS2
S16		9186.016	9185.957	6.4	9190.736	9190.678	Native CZE-MS2
S16		9186.016	9185.957	6.4	9190.736	9190.678	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S16		9186.016	9185.957	6.4	9190.736	9190.678	Native CZE-MS2
S16		9186.011	9185.957	5.9	9190.731	9190.678	Native CZE-MS2
S16		9186.011	9185.957	5.9	9190.731	9190.678	Native CZE-MS2
S16		9186.011	9185.957	5.9	9190.731	9190.678	Native CZE-MS2
S16		9186.010	9185.957	5.8	9190.730	9190.678	Native CZE-MS2
S16		9186.006	9185.957	5.2	9190.726	9190.678	Native CZE-MS2
S16		9186.006	9185.957	5.2	9190.726	9190.678	Native CZE-MS2
S16		9186.004	9185.957	5.0	9190.724	9190.678	Native CZE-MS2
S16		9186.003	9185.957	5.0	9190.723	9190.678	Native CZE-MS2
S16		9185.998	9185.957	4.4	9190.718	9190.678	Native CZE-MS2
S16		9185.990	9185.957	3.5	9190.710	9190.678	Native CZE-MS2
S16		9185.970	9185.957	1.4	9190.690	9190.678	Native CZE-MS2
S16		9185.939	9185.957	-2.0	9190.660	9190.678	Native CZE-MS2
L27		8988.894	8988.848	5.1	8993.491	8993.445	Native CZE-MS2
L27		8988.799	8988.848	-5.5	8993.396	8993.445	Native CZE-MS2
L27		8988.799	8988.848	-5.5	8993.396	8993.445	Native CZE-MS2
L27		8988.795	8988.848	-5.9	8993.393	8993.445	Native CZE-MS2
L27		8988.793	8988.848	-6.2	8993.390	8993.445	Native CZE-MS2
L27		8988.769	8988.848	-8.9	8993.366	8993.445	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L27		8987.777	8988.848	-7.6	8992.373	8993.445	Native CZE-MS2
S18	Y[+80]	8936.843	8930.780	4.8	8941.408	8935.341	Native CZE-MS2
S18	D[+64]	8914.854	8914.741	12.6	8919.405	8919.292	Native CZE-MS2
S18	E[+64]	8913.848	8914.741	12.3	8918.398	8919.292	Native CZE-MS2
S18	E[+64]	8913.848	8914.741	12.3	8918.398	8919.292	Native CZE-MS2
S18	E[+64]	8912.857	8914.741	13.8	8917.407	8919.292	Native CZE-MS2
S18	E[+64]	8912.857	8914.741	13.8	8917.407	8919.292	Native CZE-MS2
S18	.[+42]	8894.862	8892.824	3.5	8899.400	8897.361	Native CZE-MS2
S18	.[+42]	8893.863	8892.824	4.0	8898.401	8897.361	Native CZE-MS2
S18	.[+42]	8893.824	8892.824	-0.4	8898.362	8897.361	Native CZE-MS2
S18	.[+42]	8893.783	8892.824	-5.0	8898.321	8897.361	Native CZE-MS2
S18	.[+42]	8892.899	8892.824	8.5	8897.437	8897.361	Native CZE-MS2
S18	.[+42]	8892.899	8892.824	8.4	8897.436	8897.361	Native CZE-MS2
S18	.[+42]	8892.887	8892.824	7.1	8897.425	8897.361	Native CZE-MS2
S18	.[+42]	8892.882	8892.824	6.5	8897.419	8897.361	Native CZE-MS2
S18	.[+42]	8892.876	8892.824	5.8	8897.413	8897.361	Native CZE-MS2
S18	.[+42]	8892.876	8892.824	5.8	8897.413	8897.361	Native CZE-MS2
S18	.[+42]	8892.876	8892.824	5.8	8897.413	8897.361	Native CZE-MS2
S18	.[+42]	8892.875	8892.824	5.7	8897.412	8897.361	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S18	.[+42]	8892.867	8892.824	4.9	8897.405	8897.361	Native CZE-MS2
S18	.[+42]	8892.850	8892.824	2.9	8897.387	8897.361	Native CZE-MS2
S18	.[+42]	8892.845	8892.824	2.3	8897.382	8897.361	Native CZE-MS2
S18	.[+42]	8892.837	8892.824	1.4	8897.374	8897.361	Native CZE-MS2
L28		8870.833	8870.840	-0.9	8875.356	8875.364	Native CZE-MS2
L28		8870.755	8870.840	-9.6	8875.278	8875.364	Native CZE-MS2
L28		8870.753	8870.840	-9.8	8875.277	8875.364	Native CZE-MS2
L28		8870.753	8870.840	-9.9	8875.276	8875.364	Native CZE-MS2
L28		8870.750	8870.840	-10.2	8875.273	8875.364	Native CZE-MS2
L28		8870.750	8870.840	-10.2	8875.273	8875.364	Native CZE-MS2
L28		8870.750	8870.840	-10.2	8875.273	8875.364	Native CZE-MS2
L28		8870.742	8870.840	-11.1	8875.265	8875.364	Native CZE-MS2
S21		8364.752	8364.693	7.1	8368.960	8368.901	Native CZE-MS2
S21		8364.734	8364.693	4.9	8368.942	8368.901	Native CZE-MS2
S21		8364.724	8364.693	3.7	8368.932	8368.901	Native CZE-MS2
S21		8364.722	8364.693	3.4	8368.930	8368.901	Native CZE-MS2
S21		8364.646	8364.693	-5.6	8368.854	8368.901	Native CZE-MS2
S21		8364.637	8364.693	-6.7	8368.845	8368.901	Native CZE-MS2
S21		8364.620	8364.693	-8.7	8368.828	8368.901	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S21		8364.602	8364.693	-11.0	8368.809	8368.901	Native CZE-MS2
S21		8363.735	8364.693	5.4	8367.942	8368.901	Native CZE-MS2
L31	E[+24]	7884.919	7890.908	3.9	7888.828	7894.821	Native CZE-MS2
L31	E[+24]	7884.919	7890.908	3.9	7888.828	7894.821	Native CZE-MS2
L31	E[+24]	7884.919	7890.908	3.9	7888.828	7894.821	Native CZE-MS2
L31		7877.953	7866.923	-0.9	7881.857	7870.821	Native CZE-MS2
L31		7864.964	7866.923	6.0	7868.860	7870.821	Native CZE-MS2
L31		7864.948	7866.923	4.0	7868.844	7870.821	Native CZE-MS2
L31		7862.980	7866.923	9.0	7866.875	7870.821	Native CZE-MS2
L31		7862.980	7866.923	9.0	7866.875	7870.821	Native CZE-MS2
L31		7862.980	7866.923	9.0	7866.875	7870.821	Native CZE-MS2
L31		7862.977	7866.923	8.6	7866.872	7870.821	Native CZE-MS2
L31		7862.977	7866.923	8.6	7866.872	7870.821	Native CZE-MS2
L31		7862.977	7866.923	8.6	7866.872	7870.821	Native CZE-MS2
L31		7862.969	7866.923	7.6	7866.864	7870.821	Native CZE-MS2
L31		7862.945	7866.923	4.5	7866.840	7870.821	Native CZE-MS2
L29	E[+64]	7328.904	7333.918	0.3	7332.466	7337.483	Native CZE-MS2
L29	E[+64]	7328.904	7333.918	0.3	7332.466	7337.483	Native CZE-MS2
L29	K[+42]	7312.915	7312.001	-12.2	7316.467	7315.552	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L29	K[+42]	7308.961	7312.001	-4.0	7312.511	7315.552	Native CZE-MS2
L29	K[+42]	7308.961	7312.001	-4.0	7312.511	7315.552	Native CZE-MS2
L29	K[+42]	7307.934	7312.001	-7.3	7311.483	7315.552	Native CZE-MS2
L29	K[+42]	7307.902	7312.001	-11.7	7311.451	7315.552	Native CZE-MS2
L29	D[+24]	7294.000	7293.975	3.4	7297.540	7297.515	Native CZE-MS2
L29	E[+24]	7293.007	7293.975	4.9	7296.547	7297.515	Native CZE-MS2
L29	R[+14]	7293.007	7284.006	-3.9	7296.547	7287.540	Native CZE-MS2
L29	R[+14]	7293.007	7284.006	-3.9	7296.547	7287.540	Native CZE-MS2
L29	D[+24]	7292.985	7293.975	1.8	7296.525	7297.515	Native CZE-MS2
L29	D[+24]	7292.962	7293.975	-1.3	7296.502	7297.515	Native CZE-MS2
L29	E[+24]	7292.949	7293.975	-3.1	7296.489	7297.515	Native CZE-MS2
L29	E[+24]	7292.949	7293.975	-3.1	7296.489	7297.515	Native CZE-MS2
L29	R[+14]	7292.020	7284.006	-1.7	7295.559	7287.540	Native CZE-MS2
L29	D[+24]	7291.942	7293.975	-3.6	7295.481	7297.515	Native CZE-MS2
L29	E[+24]	7291.939	7293.975	-4.0	7295.478	7297.515	Native CZE-MS2
L29	E[+24]	7291.939	7293.975	-4.0	7295.478	7297.515	Native CZE-MS2
L29	R[+14]	7291.931	7284.006	-14.0	7295.470	7287.540	Native CZE-MS2
L29	E[+24]	7291.931	7293.975	-5.2	7295.470	7297.515	Native CZE-MS2
L29	E[+24]	7291.930	7293.975	-5.3	7295.469	7297.515	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L29	R[+14]	7291.930	7284.006	-14.1	7295.469	7287.540	Native CZE-MS2
L29	E[+24]	7291.918	7293.975	-6.9	7295.457	7297.515	Native CZE-MS2
L29	E[+24]	7291.918	7293.975	-6.9	7295.457	7297.515	Native CZE-MS2
L29	K[+14]	7291.918	7284.006	-15.7	7295.457	7287.540	Native CZE-MS2
L29	K[+14]	7291.918	7284.006	-15.7	7295.457	7287.540	Native CZE-MS2
L29	R[+14]	7291.011	7284.006	-2.5	7294.550	7287.540	Native CZE-MS2
L29	R[+14]	7291.009	7284.006	-2.8	7294.548	7287.540	Native CZE-MS2
L29	R[+14]	7289.990	7284.006	-4.9	7293.528	7287.540	Native CZE-MS2
L29	R[+14]	7286.023	7284.006	1.4	7289.558	7287.540	Native CZE-MS2
L29	R[+14]	7284.013	7284.006	0.9	7287.547	7287.540	Native CZE-MS2
L29		7270.971	7269.990	-3.2	7274.496	7273.515	Native CZE-MS2
L29		7270.971	7269.990	-3.2	7274.496	7273.515	Native CZE-MS2
L29		7270.971	7269.990	-3.2	7274.496	7273.515	Native CZE-MS2
L29		7270.971	7269.990	-3.2	7274.496	7273.515	Native CZE-MS2
L29		7270.971	7269.990	-3.2	7274.496	7273.515	Native CZE-MS2
L29		7270.971	7269.990	-3.2	7274.496	7273.515	Native CZE-MS2
L29		7270.971	7269.990	-3.2	7274.496	7273.515	Native CZE-MS2
L29		7270.967	7269.990	-3.7	7274.493	7273.515	Native CZE-MS2
L29		7270.963	7269.990	-4.2	7274.489	7273.515	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L29		7270.963	7269.990	-4.2	7274.489	7273.515	Native CZE-MS2
L29		7270.963	7269.990	-4.2	7274.489	7273.515	Native CZE-MS2
L29	;[+64]	7270.957	7262.881	6.8	7274.483	7266.402	Native CZE-MS2
L29		7270.957	7269.990	-5.0	7274.483	7273.515	Native CZE-MS2
L29		7270.957	7269.990	-5.0	7274.483	7273.515	Native CZE-MS2
L29		7270.078	7269.990	12.0	7273.603	7273.515	Native CZE-MS2
L29		7270.052	7269.990	8.5	7273.578	7273.515	Native CZE-MS2
L29		7270.052	7269.990	8.5	7273.578	7273.515	Native CZE-MS2
L29		7270.052	7269.990	8.5	7273.578	7273.515	Native CZE-MS2
L29		7270.050	7269.990	8.3	7273.576	7273.515	Native CZE-MS2
L29		7270.043	7269.990	7.2	7273.568	7273.515	Native CZE-MS2
L29		7270.042	7269.990	7.1	7273.568	7273.515	Native CZE-MS2
L29		7270.033	7269.990	5.9	7273.558	7273.515	Native CZE-MS2
L29		7270.024	7269.990	4.6	7273.549	7273.515	Native CZE-MS2
L29		7270.022	7269.990	4.3	7273.547	7273.515	Native CZE-MS2
L29		7269.943	7269.990	-6.6	7273.468	7273.515	Native CZE-MS2
L29		7269.940	7269.990	-7.0	7273.465	7273.515	Native CZE-MS2
L29		7269.124	7269.990	18.9	7272.649	7273.515	Native CZE-MS2
L29		7269.103	7269.990	16.0	7272.628	7273.515	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L29		7269.102	7269.990	15.8	7272.626	7273.515	Native CZE-MS2
L29		7269.086	7269.990	13.7	7272.611	7273.515	Native CZE-MS2
L29		7269.086	7269.990	13.7	7272.611	7273.515	Native CZE-MS2
L29		7269.074	7269.990	11.9	7272.598	7273.515	Native CZE-MS2
L29		7268.129	7269.990	20.0	7271.653	7273.515	Native CZE-MS2
L29	K[+42]	7236.041	7240.964	13.0	7239.545	7244.471	Native CZE-MS2
L35	T[+80]	7232.204	7234.207	0.5	7235.706	7237.710	Native CZE-MS2
L29	E[+24]	7226.055	7222.938	14.7	7229.553	7226.434	Native CZE-MS2
L35	D[+64]	7219.148	7218.169	-3.3	7222.642	7221.662	Native CZE-MS2
L35	D[+64]	7210.209	7218.169	9.4	7213.697	7221.662	Native CZE-MS2
L35	K[+42]	7198.241	7196.251	-2.3	7201.722	7199.730	Native CZE-MS2
L35	K[+42]	7197.187	7196.251	-9.4	7200.667	7199.730	Native CZE-MS2
L35	K[+42]	7193.195	7196.251	-6.5	7196.672	7199.730	Native CZE-MS2
L35	K[+42]	7192.244	7196.251	0.8	7195.720	7199.730	Native CZE-MS2
L35	K[+42]	7192.239	7196.251	0.2	7195.716	7199.730	Native CZE-MS2
L35	K[+42]	7192.228	7196.251	-1.4	7195.704	7199.730	Native CZE-MS2
L35	K[+42]	7192.228	7196.251	-1.4	7195.704	7199.730	Native CZE-MS2
L35	K[+42]	7192.196	7196.251	-5.9	7195.672	7199.730	Native CZE-MS2
L35	K[+42]	7192.194	7196.251	-6.0	7195.671	7199.730	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L35	K[+42]	7192.193	7196.251	-6.2	7195.670	7199.730	Native CZE-MS2
L35	K[+28]	7191.256	7182.272	-6.4	7194.732	7185.742	Native CZE-MS2
L35	K[+42]	7191.255	7196.251	2.9	7194.732	7199.730	Native CZE-MS2
L35	K[+42]	7191.200	7196.251	-4.8	7194.676	7199.730	Native CZE-MS2
L35	K[+14]	7177.327	7168.256	5.6	7180.794	7171.718	Native CZE-MS2
L35	D[+24]	7177.278	7178.226	7.8	7180.746	7181.694	Native CZE-MS2
L35	D[+24]	7177.277	7178.226	7.6	7180.744	7181.694	Native CZE-MS2
L35	;[+24]	7177.277	7178.226	7.6	7180.744	7181.694	Native CZE-MS2
L35	R[+14]	7177.252	7168.256	-4.8	7180.720	7171.718	Native CZE-MS2
L35	D[+24]	7177.252	7178.226	4.2	7180.720	7181.694	Native CZE-MS2
L35	D[+24]	7177.252	7178.226	4.2	7180.720	7181.694	Native CZE-MS2
L35	K[+14]	7177.252	7168.256	-4.8	7180.720	7171.718	Native CZE-MS2
L35	D[+24]	7177.249	7178.226	3.7	7180.717	7181.694	Native CZE-MS2
L35	D[+24]	7177.239	7178.226	2.3	7180.706	7181.694	Native CZE-MS2
L35	D[+24]	7177.236	7178.226	1.9	7180.703	7181.694	Native CZE-MS2
L35	D[+24]	7177.211	7178.226	-1.6	7180.678	7181.694	Native CZE-MS2
L35	K[+28]	7176.284	7182.272	4.4	7179.751	7185.742	Native CZE-MS2
L35	K[+28]	7176.284	7182.272	4.4	7179.751	7185.742	Native CZE-MS2
L35	D[+24]	7176.282	7178.226	8.8	7179.749	7181.694	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L35	K[+28]	7176.281	7182.272	4.1	7179.748	7185.742	Native CZE-MS2
L35	K[+28]	7176.281	7182.272	4.1	7179.748	7185.742	Native CZE-MS2
L35	K[+28]	7176.280	7182.272	4.0	7179.747	7185.742	Native CZE-MS2
L35	K[+14]	7176.274	7168.256	-1.3	7179.741	7171.718	Native CZE-MS2
L35	K[+14]	7176.274	7168.256	-1.3	7179.741	7171.718	Native CZE-MS2
L35	K[+14]	7176.274	7168.256	-1.3	7179.741	7171.718	Native CZE-MS2
L35	K[+28]	7176.266	7182.272	2.0	7179.733	7185.742	Native CZE-MS2
L35	K[+14]	7176.266	7168.256	-2.4	7179.733	7171.718	Native CZE-MS2
L35	D[+24]	7176.266	7178.226	6.6	7179.733	7181.694	Native CZE-MS2
L35	D[+24]	7176.259	7178.226	5.5	7179.725	7181.694	Native CZE-MS2
L35	D[+24]	7176.259	7178.226	5.5	7179.725	7181.694	Native CZE-MS2
L35	D[+24]	7176.259	7178.226	5.5	7179.725	7181.694	Native CZE-MS2
L35	D[+24]	7176.257	7178.226	5.3	7179.724	7181.694	Native CZE-MS2
L35	K[+28]	7176.252	7182.272	0.1	7179.719	7185.742	Native CZE-MS2
L35	K[+14]	7176.238	7168.256	-6.2	7179.705	7171.718	Native CZE-MS2
L35	K[+14]	7176.238	7168.256	-6.2	7179.705	7171.718	Native CZE-MS2
L35	D[+24]	7176.238	7178.226	2.7	7179.705	7181.694	Native CZE-MS2
L35	D[+24]	7176.238	7178.226	2.7	7179.705	7181.694	Native CZE-MS2
L35	K[+14]	7176.238	7168.256	-6.3	7179.705	7171.718	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L35	D[+24]	7176.223	7178.226	0.6	7179.690	7181.694	Native CZE-MS2
L35	K[+14]	7176.223	7168.256	-8.4	7179.690	7171.718	Native CZE-MS2
L35	K[+14]	7176.223	7168.256	-8.4	7179.690	7171.718	Native CZE-MS2
L35	K[+14]	7176.223	7168.256	-8.4	7179.689	7171.718	Native CZE-MS2
L35	R[+14]	7176.223	7168.256	-8.4	7179.689	7171.718	Native CZE-MS2
L35	K[+14]	7175.216	7168.256	-8.9	7178.682	7171.718	Native CZE-MS2
L35	K[+14]	7169.278	7168.256	2.5	7172.740	7171.718	Native CZE-MS2
L35	K[+14]	7169.266	7168.256	0.9	7172.729	7171.718	Native CZE-MS2
L35	R[+14]	7169.266	7168.256	0.9	7172.729	7171.718	Native CZE-MS2
L35	K[+14]	7169.266	7168.256	0.9	7172.729	7171.718	Native CZE-MS2
L35	K[+14]	7168.295	7168.256	5.4	7171.757	7171.718	Native CZE-MS2
L35	K[+14]	7168.292	7168.256	5.0	7171.754	7171.718	Native CZE-MS2
L35	K[+14]	7168.292	7168.256	5.0	7171.754	7171.718	Native CZE-MS2
L35	K[+14]	7168.292	7168.256	5.0	7171.754	7171.718	Native CZE-MS2
L35	K[+14]	7168.252	7168.256	-0.6	7171.714	7171.718	Native CZE-MS2
L35	K[+14]	7166.300	7168.256	7.0	7169.760	7171.718	Native CZE-MS2
L35		7156.280	7154.241	4.6	7159.735	7157.694	Native CZE-MS2
L35		7154.340	7154.241	13.8	7157.793	7157.694	Native CZE-MS2
L35		7154.340	7154.241	13.8	7157.793	7157.694	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L35		7154.340	7154.241	13.8	7157.793	7157.694	Native CZE-MS2
L35		7154.340	7154.241	13.8	7157.793	7157.694	Native CZE-MS2
L35		7154.326	7154.241	11.9	7157.779	7157.694	Native CZE-MS2
L35		7154.326	7154.241	11.9	7157.779	7157.694	Native CZE-MS2
L35		7154.326	7154.241	11.9	7157.779	7157.694	Native CZE-MS2
L35		7154.325	7154.241	11.8	7157.778	7157.694	Native CZE-MS2
L35		7154.325	7154.241	11.8	7157.778	7157.694	Native CZE-MS2
L35		7154.324	7154.241	11.6	7157.777	7157.694	Native CZE-MS2
L35		7154.324	7154.241	11.6	7157.777	7157.694	Native CZE-MS2
L35		7154.324	7154.241	11.6	7157.777	7157.694	Native CZE-MS2
L35		7154.321	7154.241	11.2	7157.774	7157.694	Native CZE-MS2
L35		7154.321	7154.241	11.2	7157.774	7157.694	Native CZE-MS2
L35		7154.310	7154.241	9.7	7157.763	7157.694	Native CZE-MS2
L35		7154.308	7154.241	9.4	7157.761	7157.694	Native CZE-MS2
L35		7154.308	7154.241	9.4	7157.761	7157.694	Native CZE-MS2
L35		7154.301	7154.241	8.5	7157.754	7157.694	Native CZE-MS2
L35		7154.301	7154.241	8.5	7157.754	7157.694	Native CZE-MS2
L35		7154.296	7154.241	7.7	7157.749	7157.694	Native CZE-MS2
L35		7154.296	7154.241	7.7	7157.749	7157.694	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L35		7154.296	7154.241	7.7	7157.749	7157.694	Native CZE-MS2
L35		7154.275	7154.241	4.9	7157.729	7157.694	Native CZE-MS2
L35		7154.274	7154.241	4.7	7157.727	7157.694	Native CZE-MS2
L35		7154.271	7154.241	4.2	7157.724	7157.694	Native CZE-MS2
L35		7154.271	7154.241	4.2	7157.724	7157.694	Native CZE-MS2
L35		7154.266	7154.241	3.5	7157.719	7157.694	Native CZE-MS2
L35		7154.266	7154.241	3.5	7157.719	7157.694	Native CZE-MS2
L35		7154.231	7154.241	-1.3	7157.684	7157.694	Native CZE-MS2
L35		7154.231	7154.241	-1.3	7157.684	7157.694	Native CZE-MS2
L35		7154.189	7154.241	-7.2	7157.642	7157.694	Native CZE-MS2
L35		7153.360	7154.241	17.2	7156.813	7157.694	Native CZE-MS2
L35		7153.360	7154.241	17.2	7156.813	7157.694	Native CZE-MS2
L35		7153.355	7154.241	16.5	7156.808	7157.694	Native CZE-MS2
L35		7153.355	7154.241	16.5	7156.808	7157.694	Native CZE-MS2
L35		7153.355	7154.241	16.5	7156.808	7157.694	Native CZE-MS2
L35		7153.332	7154.241	13.3	7156.785	7157.694	Native CZE-MS2
L30		6407.555	6407.612	-9.0	6410.542	6410.600	Native CZE-MS2
L30		6407.550	6407.612	-9.6	6410.538	6410.600	Native CZE-MS2
L30		6407.546	6407.612	-10.3	6410.534	6410.600	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L32		6312.351	6312.398	-7.5	6315.279	6315.326	Native CZE-MS2
L32		6312.343	6312.398	-8.8	6315.271	6315.326	Native CZE-MS2
L32		6312.339	6312.398	-9.3	6315.267	6315.326	Native CZE-MS2
L32		6312.333	6312.398	-10.4	6315.261	6315.326	Native CZE-MS2
L32		6312.333	6312.398	-10.4	6315.261	6315.326	Native CZE-MS2
L32		6312.333	6312.398	-10.4	6315.261	6315.326	Native CZE-MS2
L32		6312.332	6312.398	-10.5	6315.260	6315.326	Native CZE-MS2
L33	.[+42]	6269.313	6279.597	-39.9	6272.214	6282.505	Native CZE-MS2
L33	K[+14]	6251.560	6251.602	-6.8	6254.450	6254.492	Native CZE-MS2
L33	K[+14]	6251.555	6251.602	-7.5	6254.445	6254.492	Native CZE-MS2
L33	K[+14]	6251.554	6251.602	-7.6	6254.445	6254.492	Native CZE-MS2
L33	R[+14]	6251.542	6251.602	-9.7	6254.432	6254.492	Native CZE-MS2
L33	R[+14]	6251.539	6251.602	-10.2	6254.429	6254.492	Native CZE-MS2
L33	K[+14]	6251.537	6251.602	-10.4	6254.427	6254.492	Native CZE-MS2
L36		4362.384	4362.457	-16.6	4364.096	4364.169	Native CZE-MS2
L36		4360.388	4362.457	-14.2	4362.099	4364.169	Native CZE-MS2
L36	S[+80]	4318.368	4311.383	-9.0	4320.052	4313.063	Native CZE-MS2
L36	S[+80]	4318.368	4311.383	-9.0	4320.052	4313.063	Native CZE-MS2
L36	S[+80]	4318.368	4311.383	-9.0	4320.052	4313.063	Native CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L5	M[+16]	20173.656	20174.664	-0.2	20185.229	20186.236	Denaturing CZE-MS2
L5	M[+16], R[+16]	20171.662	20190.658	3.3	20183.233	20202.241	Denaturing CZE-MS2
L5	D[+22], M[+16]	20169.675	20196.633	6.5	20181.244	20208.220	Denaturing CZE-MS2
L6	M[+16], T[+80]	18834.164	18858.089	8.2	18844.901	18868.841	Denaturing CZE-MS2
L6	E[+62]	18826.935	18824.042	-6.2	18837.667	18834.772	Denaturing CZE-MS2
L6	E[+62]	18823.073	18824.042	1.8	18833.803	18834.772	Denaturing CZE-MS2
L6	E[+62]	18821.187	18824.042	8.3	18831.916	18834.772	Denaturing CZE-MS2
L6	E[+22]	18775.339	18784.098	14.5	18786.039	18794.803	Denaturing CZE-MS2
L6	E[+22], N[+1]	18774.154	18785.082	5.8	18784.853	18795.788	Denaturing CZE-MS2
L10	K[+100], M[+16]	17710.462	17686.368	0.8	17720.498	17696.389	Denaturing CZE-MS2
L10	K[+28]	17615.451	17598.388	0.3	17625.428	17608.355	Denaturing CZE-MS2
L10	K[+28]	17614.244	17598.388	-11.3	17624.220	17608.355	Denaturing CZE-MS2
L10	K[+28]	17583.451	17598.388	6.4	17593.408	17608.355	Denaturing CZE-MS2
L10	M[+16], R[+14]	17577.354	17600.367	3.6	17587.307	17610.335	Denaturing CZE-MS2
L10		17545.211	17570.357	-3.5	17555.144	17580.306	Denaturing CZE-MS2
L10	C[-1]	17542.164	17569.349	-5.4	17552.095	17579.297	Denaturing CZE-MS2
S5	M[+16], T[+80]	17503.574	17500.250	17.9	17513.481	17510.155	Denaturing CZE-MS2
S5	N[+1], S[+80]	17500.198	17485.239	-5.2	17510.103	17495.135	Denaturing CZE-MS2
S5	.[42], M[+16]	17489.371	17462.294	-0.8	17499.269	17472.175	Denaturing CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S5	.[42], M[+16]	17468.398	17462.294	4.8	17478.283	17472.175	Denaturing CZE-MS2
S5	M[-89][+16]	17466.344	17462.294	2.1	17476.228	17472.175	Denaturing CZE-MS2
S5	M[-89][+16]	17457.276	17462.294	-0.1	17467.155	17472.175	Denaturing CZE-MS2
S5	.[+42]	17454.331	17446.299	0.3	17464.207	17456.171	Denaturing CZE-MS2
L13	K[+14], M[+16]	16067.248	16039.560	-25.3	16076.260	16048.554	Denaturing CZE-MS2
L13	M[+16], R[+16]	16034.604	16041.539	5.5	16043.595	16050.535	Denaturing CZE-MS2
L13	N[+1]	16023.451	16010.533	-7.9	16032.435	16019.509	Denaturing CZE-MS2
L13	K[+28], N[+1]	16021.416	16038.565	-5.7	16030.399	16047.558	Denaturing CZE-MS2
L13	K[+28], N[+1]	16014.639	16038.565	9.7	16023.618	16047.558	Denaturing CZE-MS2
S6	K[+100], M[+16]	15305.531	15280.442	0.4	15314.067	15288.962	Denaturing CZE-MS2
S6	K[+100], M[+16]	15302.398	15280.442	-7.7	15310.933	15288.962	Denaturing CZE-MS2
S6		15173.371	15164.431	-5.9	15181.825	15172.879	Denaturing CZE-MS2
L15	E[+22]	14966.324	14980.284	5.8	14974.648	14988.618	Denaturing CZE-MS2
L15	K[+14]	14966.171	14972.330	-9.3	14974.496	14980.659	Denaturing CZE-MS2
L15	K[+28]	14964.298	14986.346	1.7	14972.622	14994.683	Denaturing CZE-MS2
L11	K[+100], M[+16]	14867.925	14851.863	0.6	14876.189	14860.116	Denaturing CZE-MS2
L11	K[+100]	14855.998	14835.868	4.3	14864.254	14844.111	Denaturing CZE-MS2
L11	K[+14], M[+16]	14029.343	14017.498	-14.0	14037.083	14025.231	Denaturing CZE-MS2
S8	K[+14]	13996.453	14001.503	-2.4	14004.173	14009.226	Denaturing CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S8	C[-1], D[+46]	13650.456	13643.448	-1.1	13657.961	13650.948	Denaturing CZE-MS2
S12	C[-1], D[+46]	13648.547	13643.448	6.0	13656.050	13650.948	Denaturing CZE-MS2
L14	K[+28]	13543.400	13561.489	-2.1	13550.838	13568.938	Denaturing CZE-MS2
L14	K[+28]	13541.453	13561.489	2.3	13548.889	13568.938	Denaturing CZE-MS2
L14	M[+16], S[+80]	13519.771	13498.378	23.9	13527.194	13505.788	Denaturing CZE-MS2
L20	K[+14], N[+1]	13368.664	13373.638	3.1	13375.992	13380.970	Denaturing CZE-MS2
L20	D[+22], N[+1]	13365.662	13381.592	9.2	13372.988	13388.929	Denaturing CZE-MS2
L20	D[+22], N[+1]	13364.616	13381.592	6.1	13371.942	13388.929	Denaturing CZE-MS2
S13	D[+46]	13013.256	13007.227	0.7	13020.363	13014.330	Denaturing CZE-MS2
L19	K[+14]	13002.171	13009.173	1.6	13009.271	13016.278	Denaturing CZE-MS2
S13	R[+16]	12975.624	12977.234	30.6	12982.707	12984.318	Denaturing CZE-MS2
S13	K[+28]	12969.195	12989.270	-0.7	12976.274	12996.362	Denaturing CZE-MS2
S13		12967.131	12961.239	-9.9	12974.209	12968.313	Denaturing CZE-MS2
L18	M[+16], R[+16]	12769.880	12794.928	2.8	12776.835	12801.899	Denaturing CZE-MS2
L7/L12	K[+14]	12304.451	12302.535	-7.4	12311.116	12309.199	Denaturing CZE-MS2
L7/L12	K[+100]	12270.440	12257.495	-8.1	12277.084	12264.131	Denaturing CZE-MS2
L7/L12		12267.467	12288.520	1.5	12274.109	12295.175	Denaturing CZE-MS2
L7/L12	E[+62], M[+16]	12250.435	12235.388	-0.3	12257.066	12242.009	Denaturing CZE-MS2
L22	K[+14], M[+16]	12243.662	12249.787	-8.6	12250.289	12256.418	Denaturing CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L7/L12	.[42], M[+16]	12237.437	12215.485	-9.9	12244.060	12222.094	Denaturing CZE-MS2
L22	E[+22]	12226.798	12241.746	8.4	12233.415	12248.372	Denaturing CZE-MS2
L7/L12	M[-89]	12220.512	12199.490	-4.0	12227.124	12206.089	Denaturing CZE-MS2
L7/L12	K[+42]	12215.217	12199.490	-26.7	12221.826	12206.089	Denaturing CZE-MS2
L7/L12	M[+16], R[+16]	12205.435	12189.469	-7.2	12212.038	12196.062	Denaturing CZE-MS2
L22	N[+1]	12205.236	12220.761	-38.8	12211.839	12227.373	Denaturing CZE-MS2
L7/L12	K[+14]	12195.492	12171.495	-6.9	12202.089	12178.077	Denaturing CZE-MS2
L7/L12	M[+16], R[+14]	12195.433	12187.490	-6.9	12202.030	12194.082	Denaturing CZE-MS2
L7/L12		12179.435	12157.479	-9.7	12186.021	12164.052	Denaturing CZE-MS2
L7/L12		12162.462	12157.479	-2.8	12169.038	12164.052	Denaturing CZE-MS2
S14	K[+100], N[+1]	11546.248	11544.231	0.9	11552.440	11550.422	Denaturing CZE-MS2
S14	K[+28]	11450.214	11471.262	1.9	11456.346	11477.408	Denaturing CZE-MS2
L23	K[+14]	11197.160	11207.246	-4.7	11203.134	11213.226	Denaturing CZE-MS2
L25	K[+42]	10737.223	10729.774	39.6	10742.910	10735.457	Denaturing CZE-MS2
L25		10696.992	10687.727	21.9	10702.654	10693.384	Denaturing CZE-MS2
L25	K[+14]	10693.673	10701.743	-4.1	10699.333	10707.408	Denaturing CZE-MS2
L25		10692.707	10687.727	-3.5	10698.367	10693.384	Denaturing CZE-MS2
L25	K[+28]	10691.848	10715.759	15.9	10697.507	10721.433	Denaturing CZE-MS2
L25		10689.735	10687.727	0.0	10695.392	10693.384	Denaturing CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
S19	D[+46]	10353.765	10339.669	4.7	10359.214	10345.109	Denaturing CZE-MS2
S19	K[+14], M[+16]	10348.648	10323.692	-12.4	10354.093	10329.122	Denaturing CZE-MS2
S19	K[+42], M[+16]	10347.594	10351.724	-11.3	10353.038	10357.171	Denaturing CZE-MS2
S19		10289.098	10293.682	42.1	10294.506	10299.093	Denaturing CZE-MS2
S19		10284.636	10293.682	-1.5	10290.042	10299.093	Denaturing CZE-MS2
S15		10138.417	10132.473	-7.5	10143.731	10137.783	Denaturing CZE-MS2
S15		10113.483	10132.473	7.2	10118.781	10137.783	Denaturing CZE-MS2
S20	D[+62]	9586.333	9610.201	22.1	9591.303	9615.185	Denaturing CZE-MS2
S20	M[+16], R[+16]	9570.185	9580.277	-6.1	9575.145	9585.243	Denaturing CZE-MS2
S20	K[+14], M[+16]	9555.298	9578.298	8.1	9560.249	9583.262	Denaturing CZE-MS2
S20		9555.263	9548.287	-5.0	9560.213	9553.233	Denaturing CZE-MS2
S20	M[+16], R[+16]	9523.285	9509.240	-0.2	9528.216	9514.162	Denaturing CZE-MS2
S16		9190.935	9185.957	-4.3	9195.658	9190.678	Denaturing CZE-MS2
L27	K[+42], N[+1]	9168.877	9162.920	-6.8	9173.587	9167.625	Denaturing CZE-MS2
S16	Q[+14]	9151.235	9128.936	24.6	9155.933	9133.621	Denaturing CZE-MS2
L27	;[+22]	8994.811	9010.818	5.2	8999.412	9015.428	Denaturing CZE-MS2
L28	D[+22], N[+1]	8874.748	8893.794	2.0	8879.274	8898.332	Denaturing CZE-MS2
L31	C[-1]	7884.835	7865.915	-18.4	7888.743	7869.812	Denaturing CZE-MS2
L31	C[-1]	7882.880	7865.915	-11.7	7886.787	7869.812	Denaturing CZE-MS2

Protein	Mod(s)	Observed (M+H), monoisotopic	Calc. mass (M+H), monoisotopic	Mass error (ppm)	M average: Calc. from Observed (M+H), using "averagine"	M average: Calc. from Calc. (M+H), using "averagine"	Experiment Type
L29	M[+16], S[+80]	7372.932	7365.951	-5.9	7376.521	7369.537	Denaturing CZE-MS2
L29		7274.936	7269.990	-9.7	7278.465	7273.515	Denaturing CZE-MS2
L29		7258.949	7269.990	-0.6	7262.467	7273.515	Denaturing CZE-MS2
L35	R[+16]	7158.200	7170.236	0.7	7161.656	7173.699	Denaturing CZE-MS2
L30	K[+28]	6427.567	6435.643	-7.7	6430.567	6438.648	Denaturing CZE-MS2
L32	K[+14], M[+16]	6348.200	6345.354	-25.9	6351.150	6348.303	Denaturing CZE-MS2
L32	D[+46]	6341.835	6358.386	-78.3	6344.781	6361.343	Denaturing CZE-MS2
L32	K[+14], N[+1]	6316.400	6327.398	6.1	6319.331	6330.335	Denaturing CZE-MS2
L32	K[+14]	6316.335	6326.414	-7.3	6319.265	6329.351	Denaturing CZE-MS2
L33	.[+14]	6254.537	6251.602	-12.0	6257.429	6254.492	Denaturing CZE-MS2
L33	.[+14]	6251.573	6251.602	-4.6	6254.463	6254.492	Denaturing CZE-MS2
L36	C[-1]	4362.389	4361.449	-14.5	4364.101	4363.161	Denaturing CZE-MS2

Table 4-4: Top-down analysis of the protein extract from the *E. coli* ribosome *via* CZE-MS^2 under native and denaturing conditions. All identified proteoforms are reported in this table. Searches performed with Byonic (FDR<0.01) provided monoisotopic molecular masses for all identified species. The determined monoisotopic molecular masses were manually converted to average molecular masses using the "averagine" model (Senko *et al*. JASMS 1995) [352].

The above identification of RPs from native and denaturing CZE-MS^2 experiments was achieved using Byonic software (Protein Metrics). In our searches, we used a fragment mass

tolerance of 20 ppm and an FDR<1%. As an example of RP identifications, in **Figure 4-10**, we show the extracted ion electropherogram (**Figure 4-10E**) and specific precursor ions for the ribosomal protein S8 (MW: 14.1 kDa) (**Figure 4-10F**). The corresponding CZE-MS^2 spectrum is then matched to theoretical peaks *in situ*, resulting in the identification of the protein with an N-terminal Met-truncation, which matched the corresponding experimental mass with an error of ~5 ppm (**Figure 4-10G**). Another example involves the ribosomal protein L36 (MW: 4.37 kDa). For L36, the extracted ion electropherogram is shown in **Figure 4-10H**, followed by the detected precursor ions, shown in **Figure 4-10I**. The CZE-MS^2 spectrum used to identify an oxidation with high confidence at Arg_{11} is then shown in **Figure 4-10J**. Other ribosomal and ribosome-associated proteins and their specific PTMs, as well as non-covalent complexes, were determined in an analogous manner and are shown in **Table 4-5**.

We also processed the CZE-MS^2 data using ProSightPC in addition to the analysis via Byonic. While the re-analysis of native experiments did not yield additional identifications, searches of the denaturing data resulted in the identification of an additional five ribosomal proteins (L24, L6, L14, L34, S10). However, the ProSightPC searches did not yield any additional RP identifications through the combination of native and denaturing approaches. Nevertheless, several ribosome-associated proteins and additional PfSMs corresponding to RPs, not previously found by Byonic, were identified by ProSightPC. The former included stationary-phase-induced ribosome-associated protein and cold shock protein A that were likely co-purified along with other ribosomal proteins during the isolation and protein extraction steps.

Identification of ribosomal extract components by both software tools was achieved using database searching of CZE-DDA-MS^2 against a database constructed on the basis of the results from the bottom-up LC-MS^2 analysis (i.e. bottom-up guided native MS analysis) (**Table 4-3**). Identification of individual proteins and protein complexes from a complex mixture would be considerably more challenging without a target database (e.g. limited samples) and may lead to a lower profiling depth and an increased FDR. Higher efficiency fragmentation techniques and more advanced bioinformatics approaches (both database- and spectral library-based in combination with MS deconvolution) are needed to address these limitations.

Accession	Description	Coverage %	# PSMs	MW [kDa]	Score Sequest	Mod(s)
C6EI71	30S ribosomal protein S1	69.7	390	61.121	1152.903	Met-loss [N-Term]; Met-loss+Acetyl [N-Term]; Acetyl [N-Term]
C6EGE5	30S ribosomal protein S10	35.0	10	11.728	23.45501	
C6EGG9	30S ribosomal protein S11	38.0	13	13.836	41.93071	

Accession	Description	Coverage %	# PSMs	MW [kDa]	Score Sequest	Mod(s)
C6EGC2	30S ribosomal protein S12	41.1	31	13.729	80.14011	
C6EGG8	30S ribosomal protein S13	67.8	31	13.091	87.22342	
C6EGF9	30S ribosomal protein S14	18.8	86	11.573	340.2495	
C6EH60	30S ribosomal protein S15	52.8	15	10.263	33.67196	Met-loss [N-Term]
C6EJY3	30S ribosomal protein S16	58.5	54	9.185	124.6592	
C6EGF5	30S ribosomal protein S17	22.6	40	9.698	164.0567	
C6ECE7	30S ribosomal protein S18	48.0	25	8.981	55.94032	
C6EGF0	30S ribosomal protein S19	68.5	58	10.424	130.9537	
C6EAB6	30S ribosomal protein S2	69.7	73	26.727	194.4119	
C6EB30	30S ribosomal protein S20	43.7	10	9.678	26.07211	
C6EHT0	30S ribosomal protein S21	33.8	12	8.495	25.02193	
C6EGF2	30S ribosomal protein S3	58.4	57	25.967	151.029	
C6EGH0	30S ribosomal protein S4	54.4	102	23.455	310.4655	
C6EGG3	30S ribosomal protein S5	62.9	73	17.534	188.782	
C6ECE9	30S ribosomal protein S6 Escherichia coli	71.0	80	15.163	301.7659	
C6EGC3	30S ribosomal protein S7 Escherichia coli	59.6	65	17.593	210.1965	
C6EGG0	30S ribosomal protein S8 Escherichia coli	50.8	49	14.118	128.1802	Met-loss [N-Term]
C6EH02	30S ribosomal protein S9 Escherichia coli	53.1	24	14.847	41.4954	Met-loss [N-Term]
C6EEN7	30S ribosomal subunit S22 Escherichia coli	42.2	9	5.093	34.73508	

Accession	Description	Coverage %	# PSMs	MW [kDa]	Score Sequest	Mod(s)
C6EE35	50S ribosomal protein L1 Escherichia coli	69.7	137	24.714	394.8446	
C6EE34	50S ribosomal protein L10 Escherichia coli	51.5	23	17.7	65.91596	
C6EE36	50S ribosomal protein L11 Escherichia coli	26.8	16	14.866	21.8178	
C6EH01	50S ribosomal protein L13 Escherichia coli	60.6	53	16.009	123.7657	
C6EGF6	50S ribosomal protein L14 Escherichia coli	43.9	40	13.532	76.96747	Met-loss [N-Term]; Acetyl [N-Term]
C6EGG5	50S ribosomal protein L15 Escherichia coli	49.3	31	14.957	73.28626	
C6EGF3	50S ribosomal protein L16 Escherichia coli	44.9	21	15.271	56.39864	
C6EGH2	50S ribosomal protein L17 Escherichia coli	33.9	15	14.356	36.88622	
C6EGG2	50S ribosomal protein L18 Escherichia coli	49.6	62	12.762	255.5871	
C6EJY6	50S ribosomal protein L19 Escherichia coli	60.0	81	13.125	198.445	
C6EGE9	50S ribosomal protein L2 Escherichia coli	54.6	52	29.842	125.3445	
C6ECP1	50S ribosomal protein L20 Escherichia coli	35.6	9	13.489	16.11481	
C6EH41	50S ribosomal protein L21 Escherichia coli	60.2	49	11.557	93.75118	Met-loss [N-Term]
C6EGF1	50S ribosomal protein L22	57.3	45	12.219	144.3749	
C6EGE8	50S ribosomal protein L23 Escherichia coli	15.0	5	11.192	6.212666	

Accession	Description	Coverage %	# PSMs	MW [kDa]	Score Sequest	Mod(s)
C6EGF7	50S ribosomal protein L24 Escherichia coli	65.4	151	11.309	418.9912	
C6EA12	50S ribosomal protein L25 Escherichia coli	68.1	64	10.687	127.0695	Met-loss [N-Term]
C6EH42	50S ribosomal protein L27 Escherichia coli	38.8	66	9.119	170.8322	
C6EDF4	50S ribosomal protein L28 Escherichia coli	47.4	16	9.001	32.1089	
C6EGF4	50S ribosomal protein L29 Escherichia coli	49.2	28	7.269	70.24713	
C6EGE6	50S ribosomal protein L3 Escherichia coli	60.8	69	22.23	239.0788	
C6EGG4	50S ribosomal protein L30 Escherichia coli	59.3	14	6.538	27.42896	
C6EE73	50S ribosomal protein L31 Escherichia coli	64.3	12	7.866	28.77341	
C6EHC5	50S ribosomal protein L32 Escherichia coli	63.2	53	6.442	145.7369	Met-loss [N-Term]
C6EDF5	50S ribosomal protein L33 Escherichia coli	56.4	12	6.368	25.10342	
C6EGA1	50S ribosomal protein L34 Escherichia coli	19.6	2	5.377	1.721421	
C6ECP0	50S ribosomal protein L35 Escherichia coli	20.0	2	7.284	1.808036	
C6EGG7	50S ribosomal protein L36 Escherichia coli	21.1	2	4.361	3.54205	
C6EGE7	50S ribosomal protein L4 Escherichia coli	52.7	22	22.073	53.1798	
C6EGF8	50S ribosomal protein L5	64.8	84	20.289	207.066	

Accession	Description	Coverage %	# PSMs	MW [kDa]	Score Sequest	Mod(s)
	Escherichia coli					
C6EGG1	50S ribosomal protein L6 Escherichia coli	64.4	77	18.892	275.0141	
C6EE33	50S ribosomal protein L7/L12 Escherichia coli	75.2	162	12.288	348.849	
C6ECE6	50S ribosomal protein L9 Escherichia coli	58.4	127	15.759	413.123	
C6EAB9	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	5.1	1	29.873	2.485701	
C6EF82	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	11.7	7	84.675	20.94484	
C6ELG9	6,7-dimethyl-8-ribityllumazine synthase	27.6	3	16.147	0	
C6ECY2	60 kDa chaperonin	15.7	7	57.293	18.67278	
C6EAU2	Acetyltransferase component of pyruvate dehydrogenase complex	19.5	11	66.055	36.39404	
C6EHC0	Acyl carrier protein	20.5	1	8.634	3.79841	
C6EB40	Chaperone protein DnaK	3.8	1	69.072	2.8065	
C6EAA7	Chaperone protein skp	9.9	1	17.677	3.146208	
C6EJZ4	Cold shock protein associated with 30S ribosomal	32.7	5	12.777	20.34704	

Accession	Description	Coverage %	# PSMs	MW [kDa]	Score Sequest	Mod(s)
	subunit					
C6EK69	Cold shock protein E	17.4	1	7.459	1.982509	
C6EC07	Cold-shock DNA-binding domain protein	46.4	3	7.398	10.0598	
C6EAU1	Dihydrolipoyl dehydrogenase	3.2	1	50.657	2.886779	
C6EE39	Elongation factor Tu	39.6	17	43.286	47.64586	
C6EFU6	Exoribonuclease 2	17.5	9	72.43	21.04273	
C6EGD0	Ferroxidase	18.4	3	18.483	6.057896	
C6EC50	Glyceraldehyde-3-phosphate dehydrogenase	19.0	7	35.51	11.78678	
C6EH24	Hpf protein	17.9	2	10.744	4.0231	
C6EE18	HU, DNA-binding transcriptional regulator, alpha subunit	16.7	1	9.529	2.556107	
C6EGV5	Isocitrate dehydrogenase [NADP]	11.3	2	45.741	1.663644	
C6EIL7	Peptide chain release factor 2	5.2	2	41.195	8.517668	
C6EII3	Phosphoglycerate kinase	4.4	1	41.093	3.600396	
C6EH47	Predicted RNA-binding protein	12.4	1	10.777	1.860347	
C6EAU3	Pyruvate dehydrogenase E1 component	15.6	11	99.636	34.98184	
C6EK56	Ribosomal silencing factor RsfS	75.2	11	11.575	36.35503	
C6EH58	Ribosome-binding factor A 1	35.3	5	15.145	14.38698	
C6EBZ8	RNA chaperone ProQ	32.8	7	25.847	17.04594	
C6EKF6	Serine	9.6	3	45.288	13.55227	

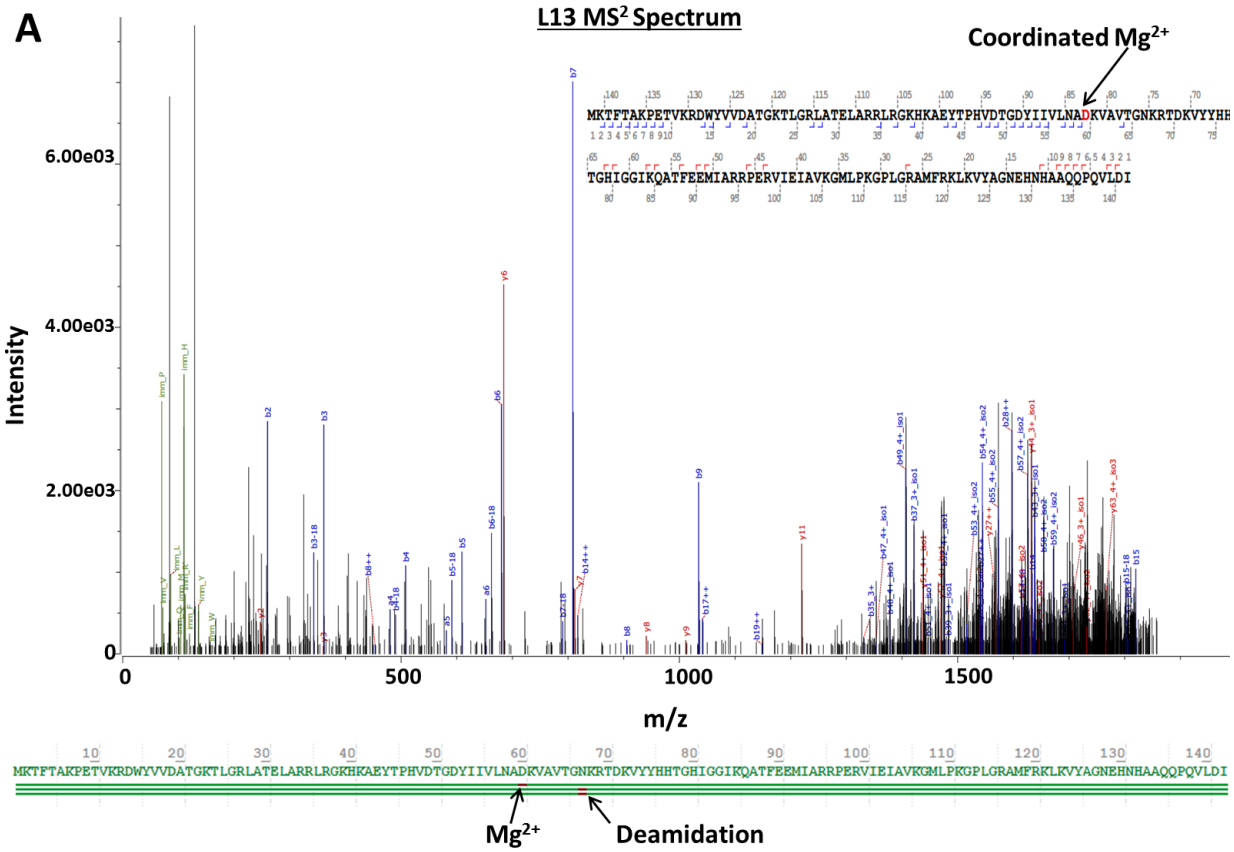
Accession	Description	Coverage %	# PSMs	MW [kDa]	Score Sequest	Mod(s)
	hydroxymethyl transferase					
C6EJX3	SsrA-binding protein	25.6	10	18.258	23.03089	
C6EJL0	Succinyl-CoA ligase [ADP-forming] subunit beta	2.8	1	41.367	2.468198	
C5W4P5	Translation initiation factor IF-3	36.1	8	20.551	22.56605	
C6EL29	Trigger factor	31.7	19	48.163	49.11785	
C6EEX9	Triosephosphate isomerase	5.5	1	26.955	2.757239	
C6EIF1	Uncharacterized protein	44.4	9	12.872	27.2605	
C6ECB5	UPF0307 protein YjgA	11.5	2	21.346	7.951681	

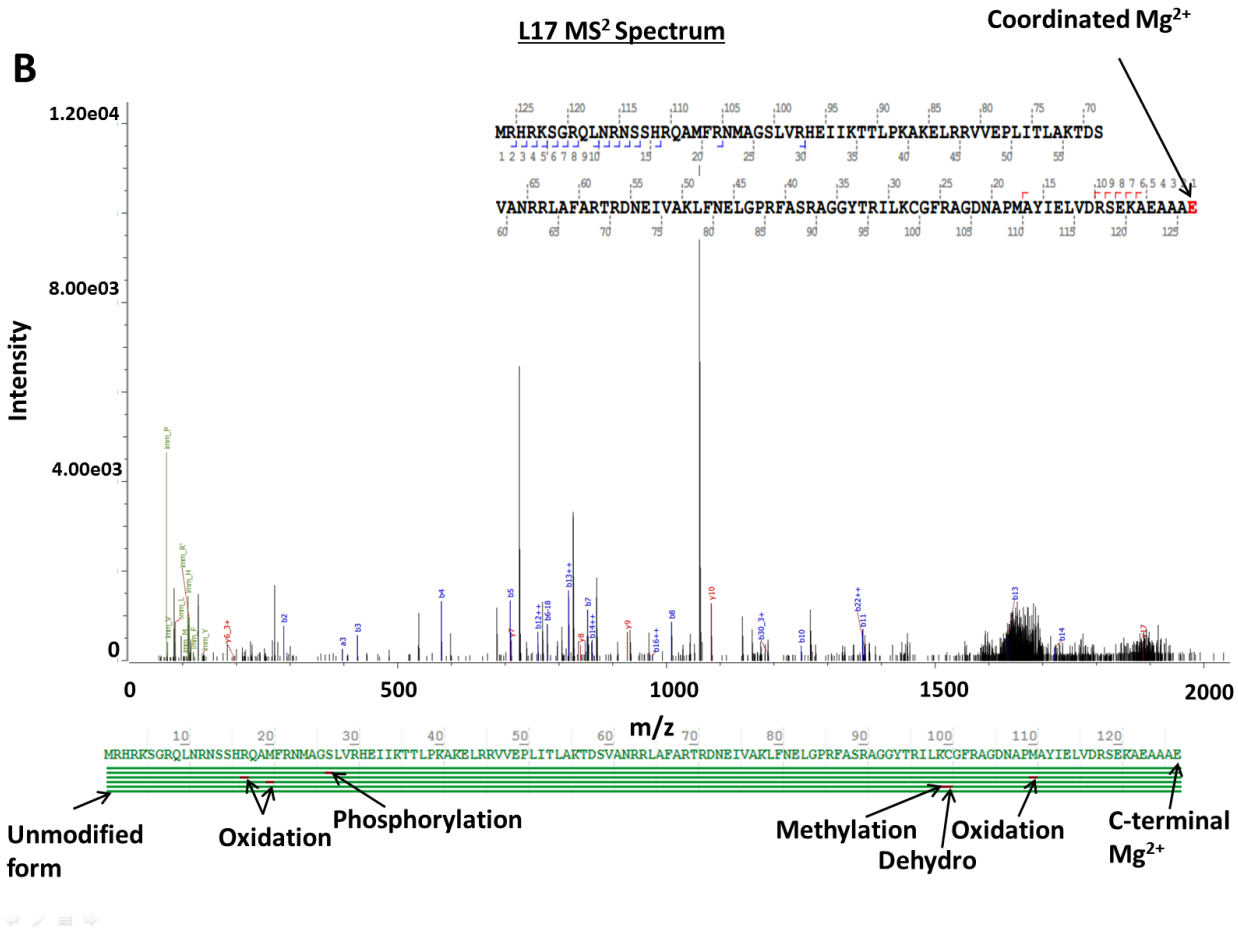
Table 4-5: Proteins identified in the analysis of the protein extract from the *E. coli* ribosome using a conventional bottom-up nanoLC-MS/MS approach. Identifications of ribosomal and ribosome-associated proteins were achieved from searches using Proteome Discoverer 1.4. All identified proteins are shown in this table (FDR<0.01, minimum two peptides).

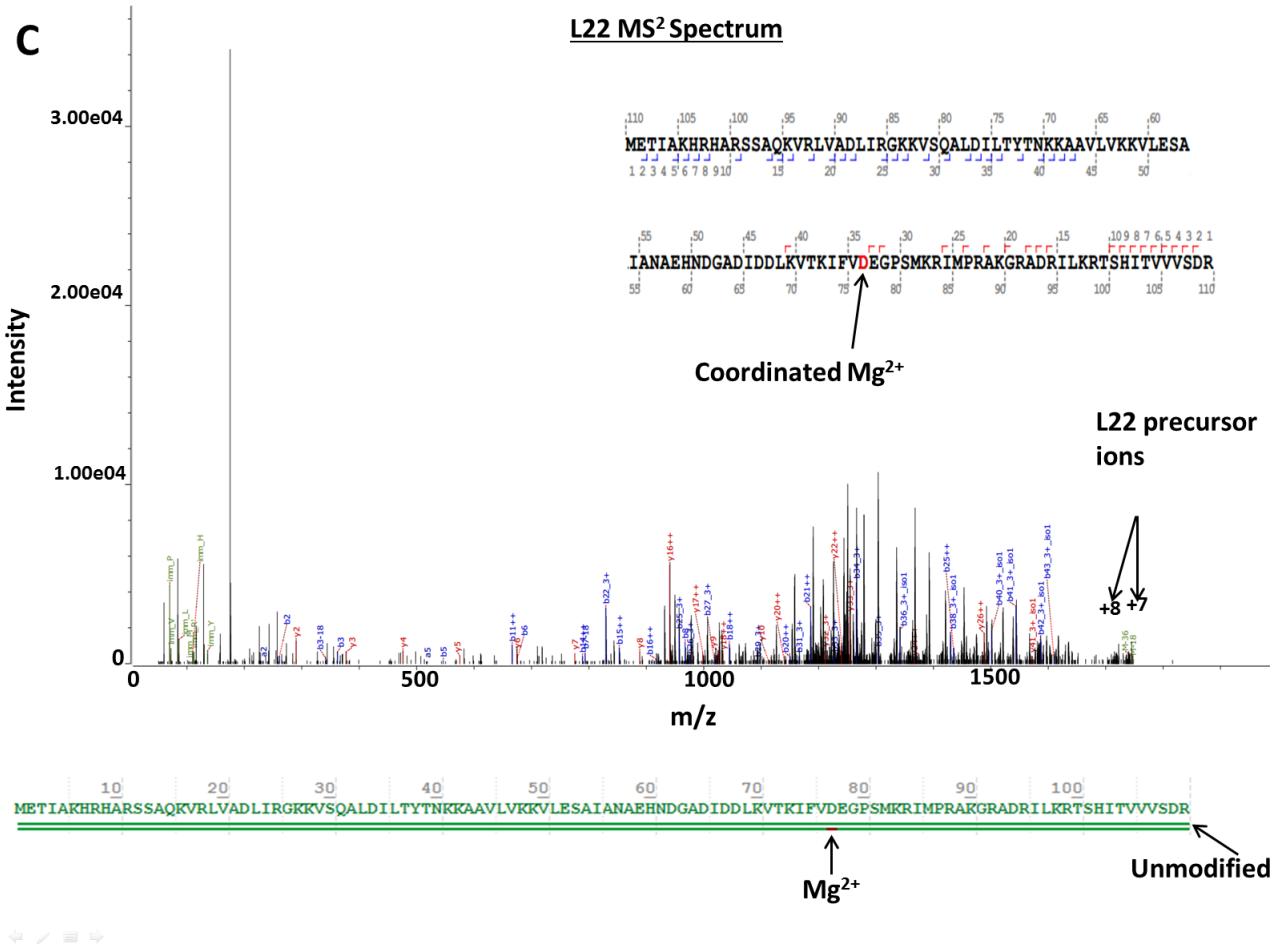
It should be mentioned that in this study, glacial acetic acid was used to precipitate ribosomal RNA for the extraction of RPs, similar to the approach described in [319]. A brief exposure to denaturing conditions was then followed by reconstitution of the protein extract into physiological pH buffer (ammonium acetate, pH 7.5). Even a brief exposure to denaturing conditions is expected to result in protein unfolding and dissociation of non-covalent interactions for at least some complexes. However, certain non-covalent interactions, including coordination of metal ions and multisubunit protein-protein interactions, can be stable at such conditions [353-355]. We anticipated that the reconstituted RP extracts would contain non-covalent complexes that were (a) stable enough to survive the denaturing conditions and (b) re-assembled after the buffer exchange to native conditions. While the latter reconstituted complexes may not exactly represent structural features characteristic to naïve ribosomes due to protein misfolding and scrambled interactions, we used the ribosomal protein extract as a model sample to assess the

capabilities of native CZE-MS in analysis of biological samples containing non-covalent protein complexes.

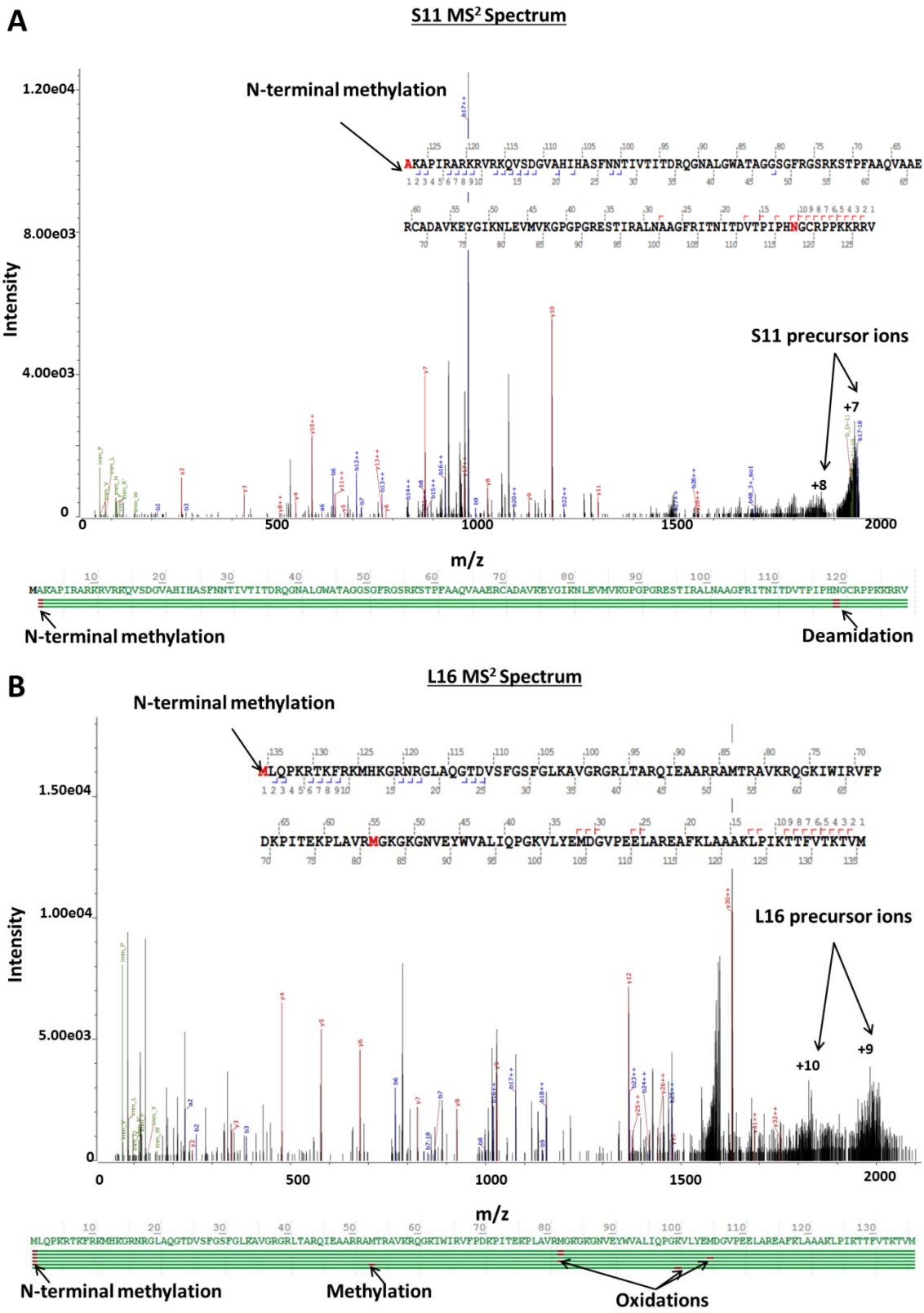
In contrast to analytical approaches requiring denaturing conditions, CZE-MS^2 characterization of the *E. coli* RP extract under native conditions resulted in the detection of multiple non-covalent protein-metal ion complexes. Metal ions (e.g. Mg^{2+}, Zn^{2+}) coordinated with specific ribosomal proteins are essential to the functional integrity of the ribosome [356]. CZE-MS^2 under native conditions allowed for the detection of ribosomal proteins L13, L17, and L22 with coordinated Mg^{2+} cofactors and L16, L29, L35, and S18 with coordinated Zn^{2+} cofactors. This was evident from both accurate intact mass measurements and fragmentation spectra, even though unambiguous localization of the coordination sites was often not possible from the acquired HCD spectra (**Figure 4-13 A-C**). Several PTMs were readily detected including oxidation, N-terminal methylation and acetylation, internal methylation and acetylation, N-terminal methionine truncation, phosphorylation, and deamidation. Methylation at the N-terminal α-amino group is an important PTM commonly found on large macromolecular assemblies, such as ribosomes [357]. N-terminal methylation of S11, L16 and L33 was unambiguously elucidated *via* native CZE-MS^2 analyses (**Figure 4-14 A-C**), which are in agreement with findings reported in a previous study [358]. N-terminal acetylation is also a common protein PTM that confers stability to proteins, in addition to being crucial for cellular localization, regulation, and function. N-terminal acetylation at S5 and S18 was reliably confirmed *via* native CZE-MS^2 (**Figure 4-15 A-B**), in agreement with previous reports [358-361]. In both cases, acetylation was observed on N-terminal Ala residues of the Met_1-truncated proteoforms, as expected. Numerous oxidation sites, N-terminally deamidated, and Met_1-truncated proteoforms of RPs were also observed. Methylation of Lys and Arg, as well as di-methylation of Lys, were also found on various internal residues of RPs (L25, S8, S3, S16, S12, L2, L16, L35, L24, L20, L17, and S4). However, localization of the aforementioned PTMs at specific residues was not possible in some cases due to incomplete sequence coverage from DDA-HCD MS^2. Similarly, a phosphorylated variant of L36 was also detected, with the corresponding PTM likely at Ser_6. All proteoforms identified from CZE-MS^2, as well as their match scores, determined and expected molecular masses, and sequence coverage values are summarized in **Table 4-4**. These results show that the native CZE-MS approach can be effectively used for detailed structural characterization of ribosomes and other macromolecular complexes to better understand their functional specialization and biology. However, alternative, gentler protein extraction techniques that do not require denaturing conditions should be used to more closely characterize biologically relevant complexes. In comparison, direct infusion conducted under native conditions does not readily allow for the identification of lower abundance proteoforms in such complex biological samples.

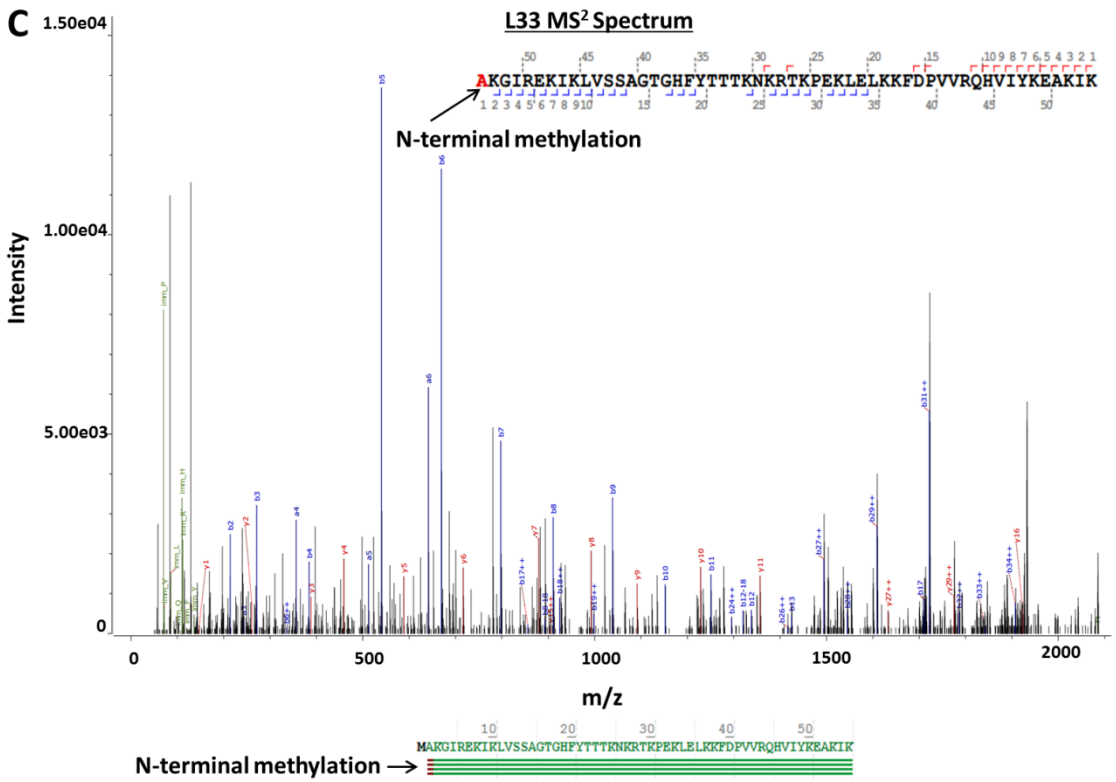

A
L13 MS² Spectrum
Coordinated Mg²⁺
MKTFTAKPETVKRDWYVVDATGKTLGRLATELARRLRGKHKAEYTPHVDTGDYIIVLNADKVAVTGNKRTDKVYYHH
TGHIGGIKQATFEEMIARRPERVIEIAVKGMLPKGPLGRAMFRKLKVYAGNEHNHAAQQPQVLDI
Intensity
6.00e03
4.00e03
2.00e03
0
0
500
1000
1500
m/z
MKTFTAKPETVKRDWYVVDATGKTLGRLATELARRLRGKHKAEYTPHVDTGDYIIVLNADKVAVTGNKRTDKVYYHHTGHIGGIKQATFEEMIARRPERVIEIAVKGMLPKGPLGRAMFRKLKVYAGNEHNHAAQQPQVLDI
Mg²⁺
Deamidation


B
L17 MS2 Spectrum
Coordinated Mg2+
MRHRKSGRQLNRNSSHRQAMFRNMAGSLVRHEIIKTTLPKAKELRRVVEPLITLAKTDS
VANRRLAFARTRDNEIVAKLFNELGPRFASRAGGYTRILKCGFRAGDNAPMAYIELVDRSEKAEAAAAE
Intensity
1.20e04
8.00e03
4.00e03
0
0
500
1000
1500
2000
m/z
MRHRKSGRQLNRNSSHRQAMFRNMAGSLVRHEIIKTTLPKAKELRRVVEPLITLAKTDSVANRRLAFARTRDNEIVAKLFNELGPRFASRAGGYTRILKCGFRAGDNAPMAYIELVDRSEKAEAAAAE
Unmodified form
Oxidation
Phosphorylation
Methylation
Dehydro
Oxidation
C-terminal Mg2+

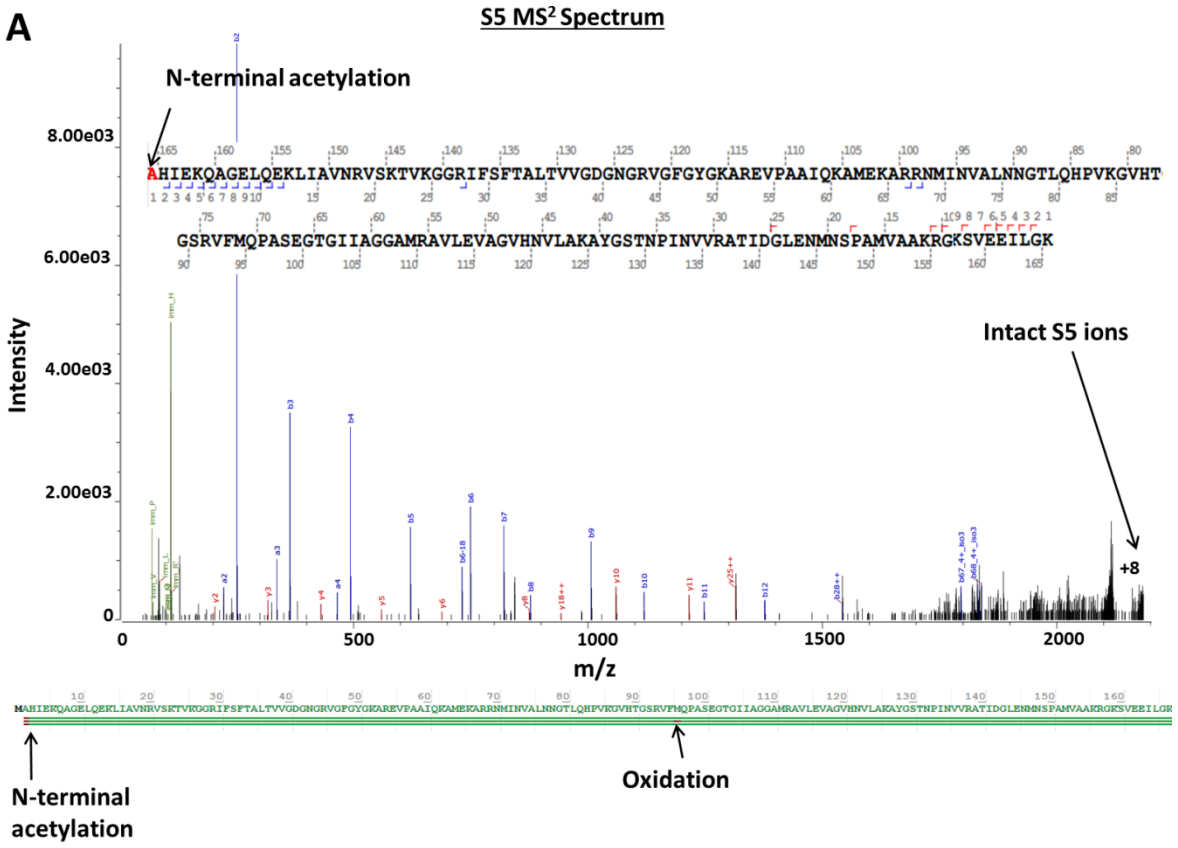


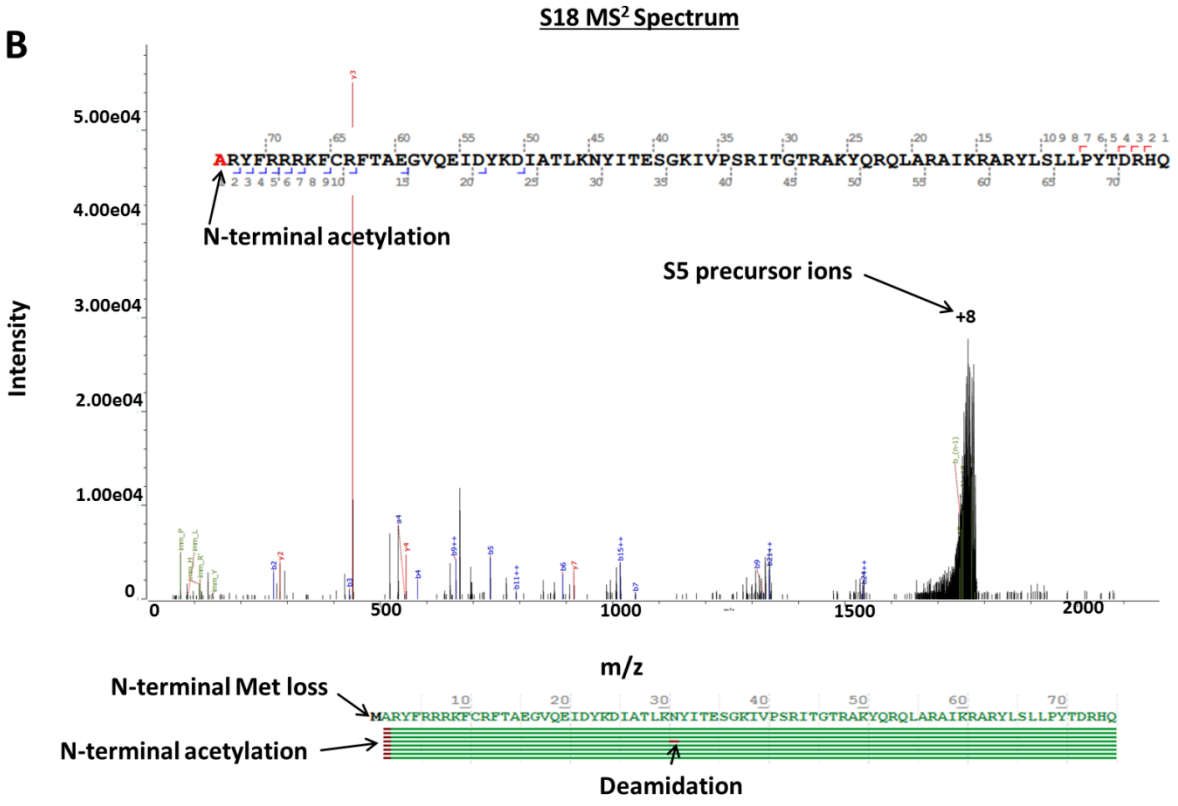
Figures 4-13 A-C. Ribosomal proteins with coordinated metal ions identified by native CZE-MS^2. Examples of ribosomal proteins with coordinated Mg^{2+} ions detected using CZE-MS^2 under native conditions: L13 (**A**), L17 (**B**), and L22 (**C**). Assignments of sequence-characteristic *b*- and *y*- ions are shown for an example spectrum. For each protein, in addition to annotated MS^2 spectra, primary structures with indicated fragmentation coverage are shown on top. Additionally, other detected *proteoform-spectrum matches* (PfSMs) with indicated PTMs are also shown below.


A
S11 MS2 Spectrum
N-terminal methylation
AKAPIRARKRVRKQVSDGVAHIHASFNNTIVTITDRQGNALGWATAGGSGFRGSRKSTPFAAQVAAE
RCADAVKEYGIKNLEVMVKGPGPGRESTIRALNAAGFRITNITDVTPIPHNGCRPPKKRRV
S11 precursor ions
+7
+8
Intensity
m/z
MAKAPIRARKRVRKQVSDGVAHIHASFNNTIVTITDRQGNALGWATAGGSGFRGSRKSTPFAAQVAAERCADAVKEYGIKNLEVMVKGPGPGRESTIRALNAAGFRITNITDVTPIPHNGCRPPKKRRV
N-terminal methylation
Deamidation
B
L16 MS2 Spectrum
N-terminal methylation
MLQPKRTKFRKMHKGRNRGLAQGTDVSFGSFGLKAVGRGRLTARQIEAARRAMTRAVKRQGKIWIRVFP
DKPITEKPLAVRMGKGKGNVEYWVALIQPGKVLYEMDGVPEELAREAFKLAAAKLPIKTTFVTKTVM
L16 precursor ions
+9
+10
Intensity
m/z
MLQPKRTKFRKMHKGRNRGLAQGTDVSFGSFGLKAVGRGRLTARQIEAARRAMTRAVKRQGKIWIRVFPDKPITEKPLAVRMGKGKGNVEYWVALIQPGKVLYEMDGVPEELAREAFKLAAAKLPIKTTFVTKTVM
N-terminal methylation
Methylation
Oxidations



Figures 4-14 A-C. Ribosomal proteins with N-terminal methylation identified by native CZE-MS^2. Examples of ribosomal proteins with N-terminal methylation detected using CZE-MS^2 under native conditions: S11 (**A**), L16 (**B**) and L33 (**C**). Assignments of *b*- and *y*- ions are shown for an example spectrum. For each protein, sequence coverage is shown above the corresponding MS^2 spectrum. Other detected PfSMs with specified PTMs are shown below each MS^2 spectrum.


A
S5 MS2 Spectrum
N-terminal acetylation
AHIEKQAGELQEKLIAVNRVSKTVKGGRIFSFTALTVVGDGNGRVGFGYGKAREVPAAIQKAMEKARRNMINVALNNGTLQHPVKGVHT
GSRVFMQPASEGTGIIAGGAMRAVLEVAGVHNVLAKAYGSTNPINVVRATIDGLENMNSPAMVAAKRGKSVEEILGK
Intact S5 ions
+8
Intensity
m/z
N-terminal acetylation
Oxidation

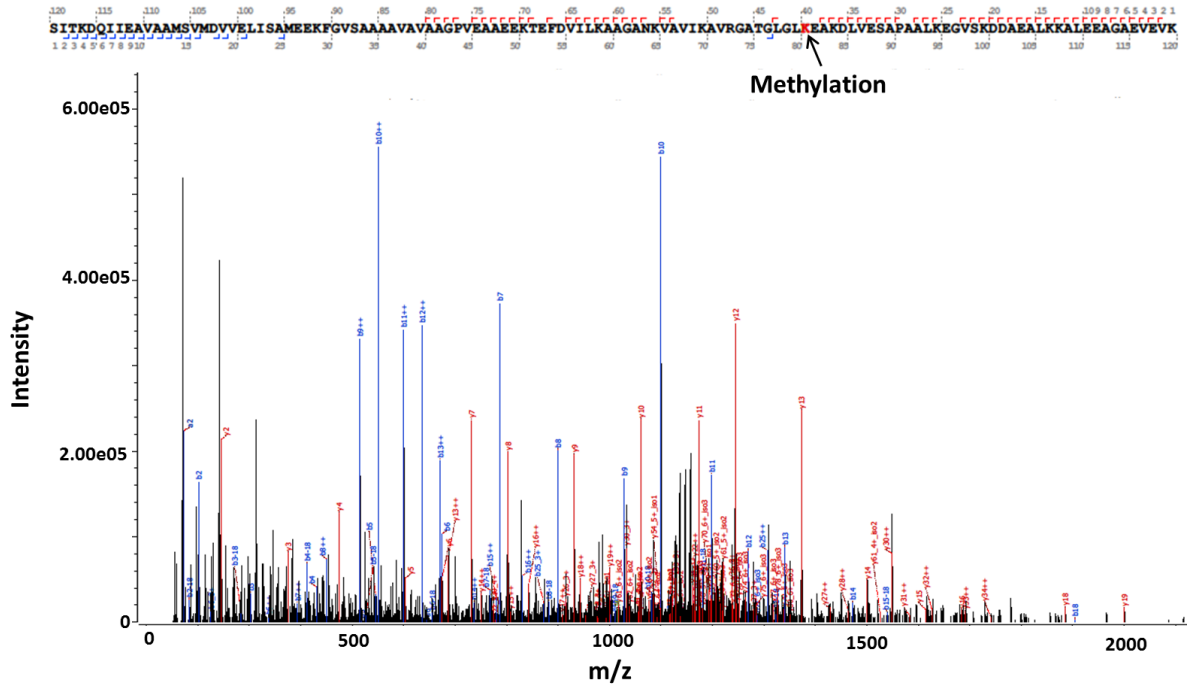

B
S18 MS2 Spectrum
ARYFRRRKFCRFTAEGVQEIDYKDIATLKNYITESGKIVPSRITGTRAKYQRQLARAIKRARYLSLLPYTDRHQ
N-terminal acetylation
S5 precursor ions
+8
Intensity
m/z
N-terminal Met loss
N-terminal acetylation
Deamidation

Figures 4-15 A-B. Ribosomal proteins with N-terminal acetylation identified by native CZE-MS^2. Native CZE-MS^2 allowed resulted in the detection of ribosomal proteins S5 (**A**), S18 (**B**) with N-terminal acetylations. Assignments of *b*- and *y*-ions were made *via* Byonic software. Similarly, annotated example MS^2 spectra are shown revealing identified *b*- and *y*- ions for each protein, in addition to MS^2 sequence coverage (top panels) and other identified PfSMs (bottom panels).

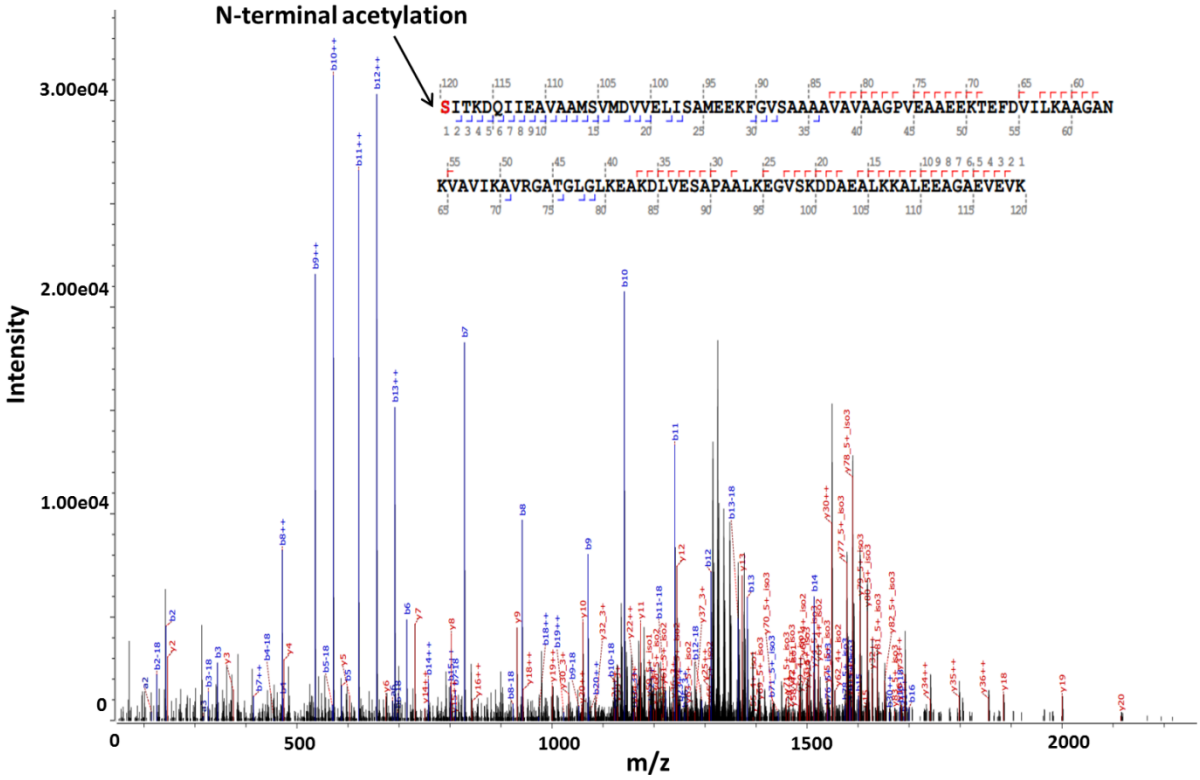
The native MS approach to protein analysis requires sophisticated and specialized MS instrumentation. For example, selection of precursor ions for MS^2 fragmentation is limited by a 2,500 Th upper selection threshold on the quadrupole of the Q Exactive Plus mass spectrometer. A considerable number of analytes were above the 2,500 Th threshold, thereby precluding their selection for MS^2 analysis. However, under denaturing acidic conditions, due to the overall increase in protein protonation and the corresponding decrease in *m/z* distributions of precursor ions, MS^2 analysis of larger ionized molecules was possible. In our case, denaturing CZE-MS^2 experiments resulted in additional identifications of RPs and their PTMs. For example, proteins S21, L7/L12, and L11 were detected in the intact CZE-MS^2 experiment using a denaturing BGE but not observed with the native BGE; presumably these proteins were present as multi-subunit RP complexes under native conditions.

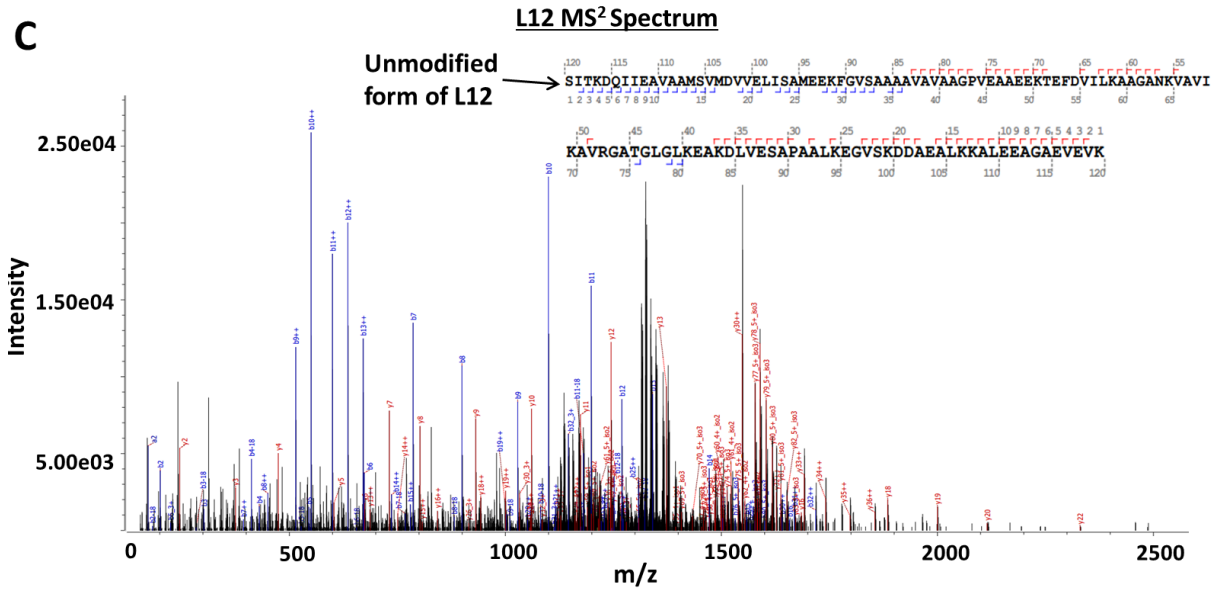
Detected exclusively under denaturing CZE-MS conditions, protein L7/L12 is an extensively-studied ribosomal protein that is biologically active as a tetramer bound to the protein L10. Proteins L7 and L12 possess identical primary structures and differ by a single PTM at their N-termini. L7 is known to be acetylated at the N-terminal Ser residue, while L12 is known to possess methylation instead of acetylation at the same residue [361, 362]. Additionally, proteins L7 and L12 are singly-methylated at their ε-amino groups of residues Lys_{81} [358]. L7/L12 forms a pentameric complex with L10 ($L10(L7/L12)_2(L7/L12)_2$) as a part of the ribosomal stalk called L8 [351, 363]. The L8 complex interacts with the translation factors IF2, EF-Tu, EF-G, and RF3, which were not identified in native or denaturing analyses; however, several of these were detected in bottom-up experiments (**Table 4-3**). The top-down CZE-MS^2 approach with the denaturing BGE employed in this work resulted in the identification of internal methylation at either Lys_{84} or Lys_{81} of L7/L12. While the exact localization of the PTM was not achieved *via* DDA-HCD MS^2, previous studies reported on internal methylation at Lys_{81} [358] (**Figure 4-16A**). Additionally, N-terminally acetylated (L7) (**Figure 4-16B**) and non-acetylated (L12) variants were also detected (**Figure 4-16C**).

L12 MS^2 Spectrum



B

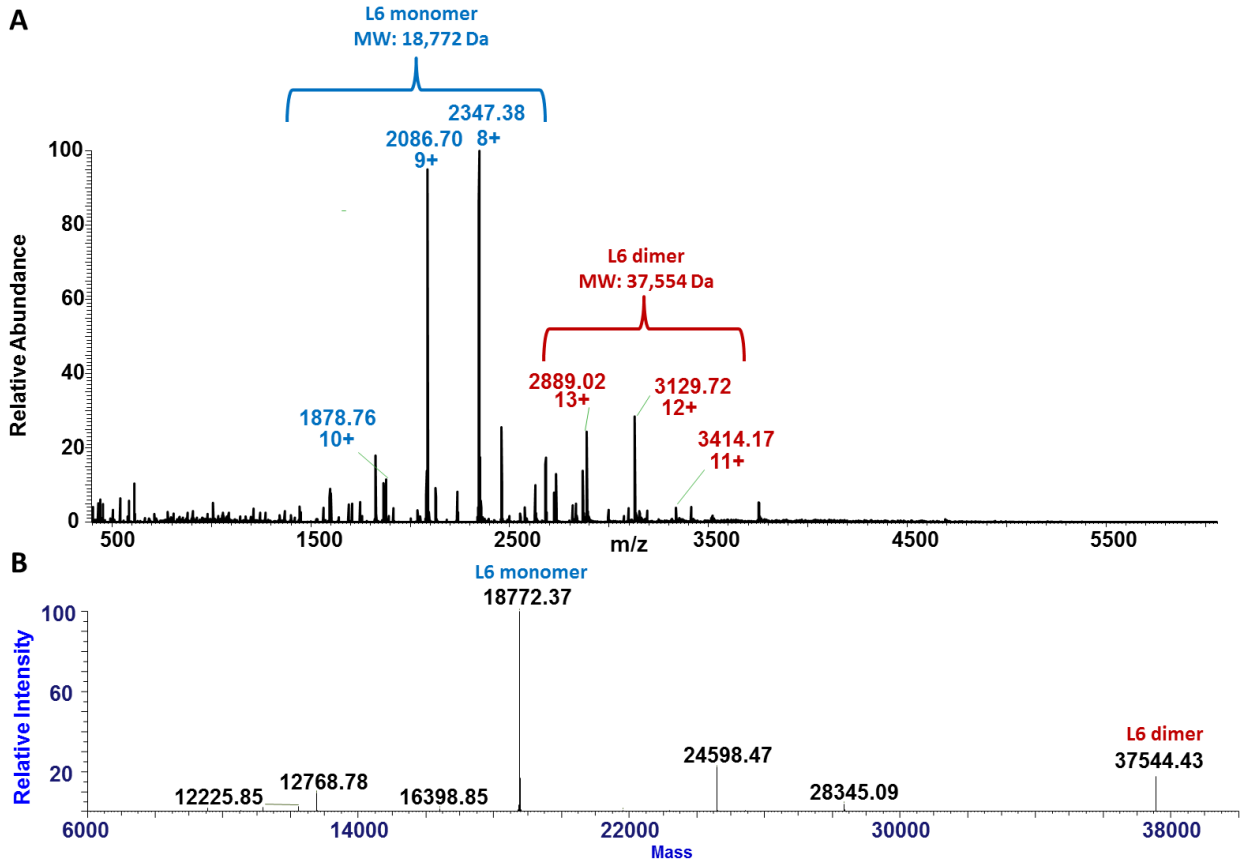




Figures 4-16 A-C: Identification of ribosomal proteins L7/L12 by denaturing CZE-MS^2. Separation and characterization of the proteins from the *E. coli* ribosomal extract under denaturing conditions resulted in the detection of several proteins not detected in CZE-MS^2 experiments conducted under native conditions. As an example, proteins L7/L12 were detected under denaturing CZE-MS^2 and not detected in native CZE-MS^2 experiments. Analysis of MS^2 spectra resulted in identification of methylation on residue Lys_{84} of protein L12 (**A**), N-terminal acetylation on protein L7 (**B**), as well as the presence of a non-acetylated form of L12 (**C**). Fragmentation coverages are shown as top panels. All identified PfSMs for L7/L12 (FDR<0.01) are shown at the bottom of the figure.

A potentially important application of native CZE-MS is the detection of protein-protein interactions. However, as stated previously, due to the sample preparation technique used in this study to extract RPs, protein-protein and protein-metal ion complexes detected after sample reconstitution may not represent native confirmations. Here, we demonstrate proof-of-principle results to show the potential of native CZE-MS in characterization of non-covalent protein complexes and interactions. For the ribosomal protein extract, several complexes were detected in native CZE-MS analyses. Several peaks detected exclusively under native conditions were interpreted as potential non-covalent complexes, including such constituents of L8 as $L10(L7/L12)_2$, $(L12)_4$ tetramer, $L10(L12)_2(L12)$ tetramer [363] (**Figure 4-11D**), while both L7 and L12 were revealed only under denaturing analysis (**Insert in Figure 4-11B and Figure 4-11C**). Also, species corresponding to the molecular weight of 37,554 Da were detected under native conditions and presumed to be a homodimer of L6, which was identified in several MS^2 spectra in its monomeric form (18,772 Da) (**Figure 4-17 A-B**). These non-covalent associations were not detected in our denaturing CZE-MS experiments, native MS direct infusion

experiments, or in the previously-reported analyses of the *E. coli* ribosomal protein extract *via* nanoLC-MS conducted under denaturing separation conditions [320].



Figures 4-17 A-B. Identification of ribosomal protein L6 and L6 dimer by native CZE-MS^1. Native CZE-MS^1 analysis of the protein extract from the *E. coli* ribosomal isolate resulted in the detection of the L6 protein and its corresponding dimer at a migration time of 29.5 minutes. The resultant averaged MS spectrum for the peak revealed the charge state profiles of both monomer (*blue*) and dimer (*red*) (**A**). Deconvolution results of the integrated spectra across the CZE-MS peak are shown in panel (**B**).

The combined advantages of the polyacrylamide-coated capillary and the Q Exactive Plus, coupled *via* the commercial sheathless CESI interface, enabled analysis of the protein isolate of the *E. coli* ribosomal extract with confirmation of protein identities with DDA-HCD-MS^2. Out of the total 55 unique *E. coli* ribosomal proteins, a single native CZE-MS^2 experiment resulted in the identification of 42 proteins of the ribosomal proteome. Additional identifications made from native CZE-MS^1 data (6) and the top-down CZE-MS^2 experiment performed under denaturing conditions (3) resulted in a total of 51 identifications, corresponding to approximately 93% of the expected proteins present in the *E. coli* ribosomal proteome. The depth of RP profiling is similar to the one reported in a recent study (46 RPs identified) that used the top-

down reversed-phase LC-MS approach [320]. Additionally, our native CZE-MS approach resulted in identification of several non-covalent protein-protein and protein-metal ion RP complexes in the reconstituted ribosomal protein extract, which demonstrates the potential of the technique to characterize protein interactions in biological and clinical samples. In our hands, direct infusion native MS experiments were far less informative and failed to provide the similar depth of RP profiling, which highlights the need for high resolution and high sensitivity native separation coupled online to MS. Separation efficiency exceeding 150,000 theoretical plates per meter were observed for selected analyte species under native CZE-MS conditions. Both high resolution and high sensitivity are among the important attributes of native CZE-MS, which were critical for enabling deep profiling of the ribosomal proteome using limited sample amounts.

4.4. Conclusions

This work represents one of the first examples of online high performance liquid-phase separation directly interfaced with MS under native conditions. The paper demonstrates the potential of CZE for high resolution separation of proteins, protein complexes and organelle proteome-level samples under native conditions coupled online to mass spectrometry. Native CZE-MS using commercially-available instrumentation enabled detailed structural characterization of such analytes in their native, biologically active states with unaltered non-covalent interactions and spatial molecular configurations. Attributes of the online CZE-MS platform used in this study such as minimal sample consumption, ultralow flow rate separation, sheathless CZE-MS without analyte dilution, the open tubular geometry, and integration with advanced MS enabled the high sensitivity (low fmol level) for characterization of intact and functionally active proteins and non-covalent protein complexes. The platform was evaluated using a high purity monoclonal antibody, mixtures of complex-forming protein standards, and a limited proteome-level isolate of *E. coli* ribosomes. Analysis of the *E. coli* ribosomal isolate on the sheathless polyacrylamide-coated capillary CZE – Q Exactive Plus system resulted in the identification of 51 intact proteins, close to 200 proteoforms, and several non-covalent protein complexes from a complementary analysis involving both native and denaturing CZE-MS, which can be instrumental in exploring ribosome specialization.

Considering that most native MS experiments have thus far been performed either with direct infusion into the MS or with a prior offline fractionation, our efforts demonstrate that the high performance separation under non-denaturing conditions can be performed online using native CZE-MS. Such an approach could serve as an effective technique in a wide variety of potential biomedical applications requiring the analysis of stable and transient protein-interactions and proteins in their native states. The high performance separation capability added to native mass spectrometry will allow one to address challenges of intrinsic heterogeneity of protein-based biomedical samples and improve the accuracy of quantitative analysis. Enabled separation of non-covalent multisubunit complexes from their monomeric and lower order

multisubunit constituents could minimize the possible ambiguity of native MS-based structural characterization and quantitation. The native CZE-MS approach seems also promising for effective characterization of whole complex proteomes if combined with up-front fractionation.

The platform could be further improved *via* modification of MS instrumentation to enable selection of higher *m/z* precursors for MS^2 fragmentation than the instruments used in this study allowed for. Alternative fragmentation techniques, such as UVPD, ETD, and EThcD and their combined use would also aid in fragmentation of internal backbone residues of larger molecular weight proteins. Lastly, availability of MS^3 or MS^n analysis on a modified Q Exactive instrument as in [135, 317], coupled with online CESI CZE with a neutral polyacrylamide-coated capillary, is expected to aid in the characterization of individual subunits of protein complexes and elucidate locations of potential PTMs down to the level of individual residues.

4.5. Acknowledgements

The authors would like to thank SCIEX for generously providing the CESI 8000 instrument and consumables used for all CZE experiments. The authors also thank Chitra Ratnayake and Anna Lou (SCIEX) for providing the prototype neutral coated OptiMS capillaries initially used in this study and Alain Beck at Centre D'Immunollogie Pierre Fabre for providing the Trastuzumab samples. We thank Thermo Fisher Scientific for providing access to an Exactive EMR instrument and Drs. Neil Kelleher and Ioanna Ntai (Northwestern University) for assistance with ProSightPC searches. We also thank Drs. Nikolai Slavov (Northeastern University) and St John Skilton (Protein Metrics) for helpful discussions. This work was supported by the ASMS 2015 Research Award (ARI), NIH 1R01GM120272 award (ARI) and the Barnett Institute of the Chemical and Biological Analysis at Northeastern University.

References

1. Bush, D.R., Zang, L., Belov, A.M., Ivanov, A.R., Karger, B.L.: High Resolution CZE-MS Quantitative Characterization of Intact Biopharmaceutical Proteins: Proteoforms of Interferon-beta1. Anal Chem. **88**, 1138-1146 (2016)
2. Belov, A.M., Viner, R., Santos, M.R., Horn, D.M., Bern, M., Karger, B.L., et al.: Analysis of Proteins, Protein Complexes, and Organellar Proteomes Using Sheathless Capillary Zone Electrophoresis - Native Mass Spectrometry. J Am Soc Mass Spectrom. (2017)
3. Pruitt, K.D., Tatusova, T., Maglott, D.R.: NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic acids research. **35**, D61-65 (2007)
4. Smith, L.M., Kelleher, N.L., Consortium for Top Down, P.: Proteoform: a single term describing protein complexity. Nat Methods. **10**, 186-187 (2013)
5. Gingras, A.C., Aebersold, R., Raught, B.: Advances in protein complex analysis using mass spectrometry. J Physiol-London. **563**, 11-21 (2005)
6. Marx, V.: Proteomics: taking on protein complexes. Nature Methods. **13**, 721-727 (2016)
7. van de Waterbeemd, M.V., Fort, K.L., Boll, D., Reinhardt-Szyba, M., Routh, A., Makarov, A., et al.: High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. Nature Methods. **14**, 283-+ (2017)
8. Manning, M.C., Patel, K., Borchardt, R.T.: Stability of Protein Pharmaceuticals. Pharm Res-Dordr. **6**, 903-918 (1989)
9. Kamionka, M.: Engineering of Therapeutic Proteins Production in Escherichia coli. Curr Pharm Biotechno. **12**, 268-274 (2011)
10. Ecker, D.M., Jones, S.D., Levine, H.L.: The therapeutic monoclonal antibody market. MAbs. **7**, 9-14 (2015)
11. Beck, A., Wurch, T., Bailly, C., Corvaia, N.: Strategies and challenges for the next generation of therapeutic antibodies. Nat Rev Immunol. **10**, 345-352 (2010)
12. Buss, N.A.P.S., Henderson, S.J., McFarlane, M., Shenton, J.M., de Haan, L.: Monoclonal antibody therapeutics: history and future. Curr Opin Pharmacol. **12**, 615-622 (2012)
13. Rosati, S., Yang, Y., Barendregt, A., Heck, A.J.: Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. Nat Protoc. **9**, 967-976 (2014)
14. Jefferis, R.: Glycosylation of recombinant antibody therapeutics. Biotechnol Prog. **21**, 11-16 (2005)
15. Chelius, D., Jing, K., Lueras, A., Rehder, D.S., Dillon, T.M., Vizel, A., et al.: Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin gamma antibodies. Anal Chem. **78**, 2370-2376 (2006)
16. Liu, H., May, K.: Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. MAbs. **4**, 17-23 (2012)
17. Brady, L.J., Martinez, T., Balland, A.: Characterization of nonenzymatic glycation on a monoclonal antibody. Anal Chem. **79**, 9403-9413 (2007)
18. Diepold, K., Bomans, K., Wiedmann, M., Zimmermann, B., Petzold, A., Schlothauer, T., et al.: Simultaneous assessment of Asp isomerization and Asn deamidation in recombinant antibodies by LC-MS following incubation at elevated temperatures. Plos One. **7**, e30295 (2012)
19. Rosati, S., Yang, Y., Barendregt, A., Heck, A.J.R.: Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. Nat. Protocols. **9**, 967-976 (2014)

20. Fekete, S., Guillarme, D., Sandra, P., Sandra, K.: Chromatographic, Electrophoretic, and Mass Spectrometric Methods for the Analytical Characterization of Protein Biopharmaceuticals. Anal Chem. **88**, 480-507 (2016)
21. Vandermarliere, E., Stes, E., Gevaert, K., Martens, L.: Resolution of Protein Structure by Mass Spectrometry. Mass Spectrom Rev. **35**, 653-665 (2016)
22. Deutzmann, R.: Structural characterization of proteins and peptides. Methods Mol Med. **94**, 269-297 (2004)
23. Kelly, S.M., Price, N.C.: The use of circular dichroism in the investigation of protein structure and function. Curr Protein Pept Sci. **1**, 349-384 (2000)
24. Li, S., Xing, D., Li, J.: Dynamic light scattering application to study protein interactions in electrolyte solutions. J Biol Phys. **30**, 313-324 (2004)
25. Haselberg, R., de Jong, G.J., Somsen, G.W.: Low-Flow Sheathless Capillary Electrophoresis-Mass Spectrometry for Sensitive Glycoform Profiling of Intact Pharmaceutical Proteins. Anal Chem. **85**, 2289-2296 (2013)
26. Haselberg, R., Brinks, V., Hawe, A., de Jong, G.J., Somsen, G.W.: Capillary electrophoresis-mass spectrometry using noncovalently coated capillaries for the analysis of biopharmaceuticals. Anal Bioanal Chem. **400**, 295-303 (2011)
27. Chen, G.D., Warrack, B.M., Goodenough, A.K., Wei, H., Wang-Iverson, D.B., Tymiak, A.A.: Characterization of protein therapeutics by mass spectrometry: recent developments and future directions. Drug Discov Today. **16**, 58-64 (2011)
28. Heck, A.J.R.: Native mass spectrometry: a bridge between interactomics and structural biology. Nature Methods. **5**, 927-933 (2008)
29. Jonic, S., Venien-Bryan, C.: Protein structure determination by electron cryo-microscopy. Curr Opin Pharmacol. **9**, 636-642 (2009)
30. Vazquez-Rey, M., Lang, D.A.: Aggregates in monoclonal antibody manufacturing processes. Biotechnol Bioeng. **108**, 1494-1508 (2011)
31. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M.: Electrospray Ionization-Principles and Practice. Mass Spectrom Rev. **9**, 37-70 (1990)
32. Karas, M., Hillenkamp, F.: Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. Anal Chem. **60**, 2299-2301 (1988)
33. Domon, B., Aebersold, R.: Review - Mass spectrometry and protein analysis. Science. **312**, 212-217 (2006)
34. Gross, J.H. Springer Science & Business Media, (2006)
35. Kebarle, P., Tang, L.: From Ions in Solution to Ions in the Gas-Phase - the Mechanism of Electrospray Mass-Spectrometry. Anal Chem. **65**, A972-A986 (1993)
36. Banerjee, S., Mazumdar, S.: Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. Int J Anal Chem. (2012)
37. Shaw, J.B., Li, W.Z., Holden, D.D., Zhang, Y., Griep-Raming, J., Fellers, R.T., et al.: Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. J Am Chem Soc. **135**, 12646-12651 (2013)
38. Yan, B., Valliere-Douglass, J., Brady, L., Steen, S., Han, M., Pace, D., et al.: Analysis of post-translational modifications in recombinant monoclonal antibody IgG1 by reversed-phase liquid chromatography/mass spectrometry. J Chromatogr A. **1164**, 153-161 (2007)
39. Gibson, G.T.T., Mugo, S.M., Oleschuk, R.D.: Nanoelectrospray Emitters: Trends and Perspective. Mass Spectrom Rev. **28**, 918-936 (2009)
40. Gale, D.C., Smith, R.D.: Small volume and low flow-rate electrospray lonization mass spectrometry of aqueous samples. Rapid Commun Mass Sp. **7**, 1017-1021 (1993)

41. Emmett, M.R., Caprioli, R.M.: Micro-electrospray mass spectrometry: Ultra-high-sensitivity analysis of peptides and proteins. J Am Soc Mass Spectrom. **5**, 605-613 (1994)
42. Wilm, M.S., Mann, M.: Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? International Journal of Mass Spectrometry and Ion Processes. **136**, 167-180 (1994)
43. Wilm, M., Mann, M.: Analytical properties of the nanoelectrospray ion source. Anal Chem. **68**, 1-8 (1996)
44. Zhang, H., Ge, Y.: Comprehensive analysis of protein modifications by top-down mass spectrometry. Circ Cardiovasc Genet. **4**, 711 (2011)
45. Schmidt, A., Gehlenborg, N., Bodenmiller, B., Mueller, L.N., Campbell, D., Mueller, M., et al.: An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures. Mol Cell Proteomics. **7**, 2138-2150 (2008)
46. Capriotti, A.L., Cavaliere, C., Foglia, P., Samperi, R., Lagana, A.: Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics. J Chromatogr A. **1218**, 8760-8776 (2011)
47. Karger, B.L., Snyder, L.R., Horvath, C.: Introduction to Separation Science. (1973)
48. Boersema, P.J., Mohammed, S., Heck, A.J.: Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem. **391**, 151-159 (2008)
49. Haselberg, R., de Jong, G.J., Somsen, G.W.: Capillary electrophoresis-mass spectrometry for the analysis of intact proteins 2007-2010. Electrophoresis. **32**, 66-82 (2011)
50. Haselberg, R., de Jong, G.J., Somsen, G.W.: Capillary electrophoresis-mass spectrometry for the analysis of intact proteins. J Chromatogr A. **1159**, 81-109 (2007)
51. VanDeemter, J.J., Zuiderweg, F.J., Klinkenberg, A.: Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography (Reprinted from Chem Engng Sci, vol 5, pg 271-289, 1956). Chem Eng Sci. **50**, 3869-3882 (1995)
52. Grossman, P.D., Colburn, J.C. Academic Press, (2012)
53. Dass, C. John Wiley & Sons, (2007)
54. Ardrey, R.E. John Wiley & Sons, (2003)
55. Walters, J.J., Grayson, M.A., Gross, M.L., Hughes, M., Shearer, G., Kohl, D.H., et al.: Ion-exchange chromatography followed by ESI-MS for quantitative analysis of sugar monophosphates from glucose catabolism. Journal of the American Society for Mass Spectrometry. **17**, 104-107 (2006)
56. Haberger, M., Leiss, M., Heidenreich, A.K., Pester, O., Hafenmair, G., Hook, M., et al.: Rapid characterization of biotherapeutic proteins by size-exclusion chromatography coupled to native mass spectrometry. mAbs. **8**, 331-339 (2016)
57. Olivares, J.A., Nguyen, N.T., Yonker, C.R., Smith, R.D.: Online Mass-Spectrometric Detection for Capillary Zone Electrophoresis. Anal Chem. **59**, 1230-1232 (1987)
58. Schmitt-Kopplin, P., Frommberger, M.: Capillary electrophoresis-mass spectrometry: 15 years of developments and applications. Electrophoresis. **24**, 3837-3867 (2003)
59. Zhang, W., Hankemeier, T., Ramautar, R.: Next-generation capillary electrophoresis-mass spectrometry approaches in metabolomics. Curr Opin Biotechnol. **43**, 1-7 (2017)
60. Maxwell, E.J., Zhong, X., Zhang, H., van Zeijl, N., Chen, D.D.: Decoupling CE and ESI for a more robust interface with MS. Electrophoresis. **31**, 1130-1137 (2010)
61. Wojcik, R., Dada, O.O., Sadilek, M., Dovichi, N.J.: Simplified capillary electrophoresis nanospray sheath-flow interface for high efficiency and sensitive peptide analysis. Rapid Commun Mass Spectrom. **24**, 2554-2560 (2010)
62. Sun, L., Zhu, G., Zhao, Y., Yan, X., Mou, S., Dovichi, N.J.: Ultrasensitive and fast bottom-up analysis of femtogram amounts of complex proteome digests. Angew Chem Int Ed Engl. **52**, 13661-13664 (2013)

63. Sun, L., Zhu, G., Zhang, Z., Mou, S., Dovichi, N.J.: Third-generation electrokinetically pumped sheath-flow nanospray interface with improved stability and sensitivity for automated capillary zone electrophoresis-mass spectrometry analysis of complex proteome digests. J Proteome Res. **14**, 2312-2321 (2015)
64. Guo, X., Fillmore, T.L., Gao, Y., Tang, K.: Capillary Electrophoresis-Nanoelectrospray Ionization-Selected Reaction Monitoring Mass Spectrometry via a True Sheathless Metal-Coated Emitter Interface for Robust and High-Sensitivity Sample Quantification. Anal Chem. **88**, 4418-4425 (2016)
65. Moini, M.: Simplifying CE-MS operation. 2. Interfacing low-flow separation techniques to mass spectrometry using a porous tip. Analytical chemistry. **79**, 4241-4246 (2007)
66. Haselberg, R., Ratnayake, C.K., de Jong, G.J., Somsen, G.W.: Performance of a sheathless porous tip sprayer for capillary electrophoresis-electrospray ionization-mass spectrometry of intact proteins. Journal of chromatography. A. **1217**, 7605-7611 (2010)
67. Busnel, J.M., Schoenmaker, B., Ramautar, R., Carrasco-Pancorbo, A., Ratnayake, C., Feitelson, J.S., et al.: High capacity capillary electrophoresis-electrospray ionization mass spectrometry: coupling a porous sheathless interface with transient-isotachophoresis. Analytical chemistry. **82**, 9476-9483 (2010)
68. Wang, Y., Fonslow, B.R., Wong, C.C., Nakorchevsky, A., Yates, J.R., 3rd: Improving the comprehensiveness and sensitivity of sheathless capillary electrophoresis-tandem mass spectrometry for proteomic analysis. Anal Chem. **84**, 8505-8513 (2012)
69. Han, X., Wang, Y., Aslanian, A., Bern, M., Lavallee-Adam, M., Yates, J.R., 3rd: Sheathless capillary electrophoresis-tandem mass spectrometry for top-down characterization of Pyrococcus furiosus proteins on a proteome scale. Anal Chem. **86**, 11006-11012 (2014)
70. Faserl, K., Sarg, B., Kremser, L., Lindner, H.: Optimization and evaluation of a sheathless capillary electrophoresis-electrospray ionization mass spectrometry platform for peptide analysis: comparison to liquid chromatography-electrospray ionization mass spectrometry. Anal Chem. **83**, 7297-7305 (2011)
71. Batz, N.G., Mellors, J.S., Alarie, J.P., Ramsey, J.M.: Chemical vapor deposition of aminopropyl silanes in microfluidic channels for highly efficient microchip capillary electrophoresis-electrospray ionization-mass spectrometry. Analytical chemistry. **86**, 3493-3500 (2014)
72. Redman, E.A., Mellors, J.S., Starkey, J.A., Ramsey, J.M.: Characterization of Intact Antibody Drug Conjugate Variants Using Microfluidic Capillary Electrophoresis-Mass Spectrometry. Analytical chemistry. **88**, 2220-2226 (2016)
73. Redman, E.A., Batz, N.G., Mellors, J.S., Ramsey, J.M.: Integrated microfluidic capillary electrophoresis-electrospray ionization devices with online MS detection for the separation and characterization of intact monoclonal antibody variants. Analytical chemistry. **87**, 2264-2272 (2015)
74. Garcia, B.A.: What does the future hold for Top Down mass spectrometry? J Am Soc Mass Spectrom. **21**, 193-202 (2010)
75. Vanhoenacker, G., Vandenheede, I., David, F., Sandra, P., Sandra, K.: Comprehensive two-dimensional liquid chromatography of therapeutic monoclonal antibody digests. Anal Bioanal Chem. **407**, 355-366 (2015)
76. Kelstrup, C.D., Young, C., Lavallee, R., Nielsen, M.L., Olsen, J.V.: Optimized Fast and Sensitive Acquisition Methods for Shotgun Proteomics on a Quadrupole Orbitrap Mass Spectrometer. Journal of Proteome Research. **11**, 3487-3497 (2012)
77. Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., Astorga-Wells, J., Zubarev, R.A.: Rapid and deep human proteome analysis by single-dimension shotgun proteomics. Mol Cell Proteomics. **12**, 3330-3338 (2013)

78. Nagaraj, N., Kulak, N.A., Cox, J., Neuhauser, N., Mayr, K., Hoerning, O., et al.: System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. Mol Cell Proteomics. **11**, M111 013722 (2012)
79. Hebert, A.S., Richards, A.L., Bailey, D.J., Ulbrich, A., Coughlin, E.E., Westphall, M.S., et al.: The one hour yeast proteome. Mol Cell Proteomics. **13**, 339-347 (2014)
80. Brunner, E., Ahrens, C.H., Mohanty, S., Baetschmann, H., Loevenich, S., Potthast, F., et al.: A high-quality catalog of the Drosophila melanogaster proteome. Nat Biotechnol. **25**, 576-583 (2007)
81. Sun, L., Hebert, A.S., Yan, X., Zhao, Y., Westphall, M.S., Rush, M.J., et al.: Over 10,000 peptide identifications from the HeLa proteome by using single-shot capillary zone electrophoresis combined with tandem mass spectrometry. Angew Chem Int Ed Engl. **53**, 13931-13933 (2014)
82. Catherman, A.D., Skinner, O.S., Kelleher, N.L.: Top Down proteomics: facts and perspectives. Biochemical and biophysical research communications. **445**, 683-693 (2014)
83. Carroll, J., Fearnley, I.M., Walker, J.E.: Definition of the mitochondrial proteome by measurement of molecular masses of membrane proteins. Proc Natl Acad Sci U S A. **103**, 16170-16175 (2006)
84. Catherman, A.D., Li, M., Tran, J.C., Durbin, K.R., Compton, P.D., Early, B.P., et al.: Top down proteomics of human membrane proteins from enriched mitochondrial fractions. Anal Chem. **85**, 1880-1888 (2013)
85. Gregorich, Z.R., Ge, Y.: Top-down proteomics in health and disease: challenges and opportunities. Proteomics. **14**, 1195-1210 (2014)
86. Gault, J., Donlan, J.A., Liko, I., Hopper, J.T., Gupta, K., Housden, N.G., et al.: High-resolution mass spectrometry of small molecules bound to membrane proteins. Nat Methods. **13**, 333-336 (2016)
87. Barrera, N.P., Isaacson, S.C., Zhou, M., Bavro, V.N., Welch, A., Schaedler, T.A., et al.: Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. Nat Methods. **6**, 585-587 (2009)
88. Landreh, M., Robinson, C.V.: A new window into the molecular physiology of membrane proteins. J Physiol. **593**, 355-362 (2015)
89. Marshall, A.G., Hendrickson, C.L., Jackson, G.S.: Fourier transform ion cyclotron resonance mass spectrometry: A primer. Mass Spectrom Rev. **17**, 1-35 (1998)
90. Zubarev, R.A., Makarov, A.: Orbitrap Mass Spectrometry. Anal Chem. **85**, 5288-5296 (2013)
91. Leney, A.C., Heck, A.J.R.: Native Mass Spectrometry: What is in the Name? J Am Soc Mass Spectr. **28**, 5-13 (2017)
92. Sharon, M.: How Far Can We Go with Structural Mass Spectrometry of Protein Complexes? J Am Soc Mass Spectr. **21**, 487-500 (2010)
93. Boeri Erba, E., Petosa, C.: The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. Protein science : a publication of the Protein Society. **24**, 1176-1192 (2015)
94. de la Mora, J.F.: Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. Anal Chim Acta. **406**, 93-104 (2000)
95. Hewitt, D., Marklund, E., Scott, D.J., Robinson, C.V., Borysik, A.J.: A Hydrodynamic Comparison of Solution and Gas Phase Proteins and Their Complexes. J Phys Chem B. **118**, 8489-8495 (2014)
96. Siuzdak, G., Bothner, B., Yeager, M., Brugidou, C., Fauquet, C.M., Hoey, K., et al.: Mass spectrometry and viral analysis. Chem Biol. **3**, 45-48 (1996)
97. Lanucara, F., Holman, S.W., Gray, C.J., Eyers, C.E.: The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. Nat Chem. **6**, 281-294 (2014)

98. Jurneczko, E., Barran, P.E.: How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. Analyst. **136**, 20-28 (2011)
99. Covey, T., Douglas, D.J.: Collision cross sections for protein ions. J Am Soc Mass Spectrom. **4**, 616-623 (1993)
100. van Duijn, E., Barendregt, A., Synowsky, S., Versluis, C., Heck, A.J.: Chaperonin complexes monitored by ion mobility mass spectrometry. J Am Chem Soc. **131**, 1452-1459 (2009)
101. Wyttenbach, T., Pierson, N.A., Clemmer, D.E., Bowers, M.T.: Ion mobility analysis of molecular dynamics. Annu Rev Phys Chem. **65**, 175-196 (2014)
102. Smith, D.P., Giles, K., Bateman, R.H., Radford, S.E., Ashcroft, A.E.: Monitoring copopulated conformational states during protein folding events using electrospray ionization-ion mobility spectrometry-mass spectrometry. J Am Soc Mass Spectrom. **18**, 2180-2190 (2007)
103. Scarff, C.A., Ashcroft, A.E., Radford, S.E.: Characterization of Amyloid Oligomers by Electrospray Ionization-Ion Mobility Spectrometry-Mass Spectrometry (ESI-IMS-MS). Methods Mol Biol. **1345**, 115-132 (2016)
104. Uetrecht, C., Barbu, I.M., Shoemaker, G.K., van Duijn, E., Heck, A.J.: Interrogating viral capsid assembly with ion mobility-mass spectrometry. Nat Chem. **3**, 126-132 (2011)
105. Henry, K.D., Williams, E.R., Wang, B.H., McLafferty, F.W., Shabanowitz, J., Hunt, D.F.: Fourier-transform mass spectrometry of large molecules by electrospray ionization. Proc Natl Acad Sci U S A. **86**, 9075-9078 (1989)
106. Beu, S.C., Senko, M.W., Quinn, J.P., McLafferty, F.W.: Improved fourier-transform ion-cyclotron-resonance mass spectrometry of large biomolecules. J Am Soc Mass Spectrom. **4**, 190-192 (1993)
107. Senko, M.W., Hendrickson, C.L., Pasa-Tolic, L., Marto, J.A., White, F.M., Guan, S., et al.: Electrospray ionization Fourier transform ion cyclotron resonance at 9.4 T. Rapid Commun Mass Spectrom. **10**, 1824-1828 (1996)
108. Shaw, J.B., Lin, T.Y., Leach, F.E., 3rd, Tolmachev, A.V., Tolic, N., Robinson, E.W., et al.: 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer Greatly Expands Mass Spectrometry Toolbox. J Am Soc Mass Spectrom. **27**, 1929-1936 (2016)
109. Belov, M.E., Gorshkov, M.V., Udseth, H.R., Anderson, G.A., Smith, R.D.: Zeptomole-sensitivity electrospray ionization--Fourier transform ion cyclotron resonance mass spectrometry of proteins. Anal Chem. **72**, 2271-2279 (2000)
110. Belov, M.E., Nikolaev, E.N., Anderson, G.A., Auberry, K.J., Harkewicz, R., Smith, R.D.: Electrospray ionization-Fourier transform ion cyclotron mass spectrometry using ion preselection and external accumulation for ultrahigh sensitivity. J Am Soc Mass Spectrom. **12**, 38-48 (2001)
111. Wilcox, B.E., Hendrickson, C.L., Marshall, A.G.: Improved ion extraction from a linear octopole ion trap: SIMION analysis and experimental demonstration. J Am Soc Mass Spectrom. **13**, 1304-1312 (2002)
112. Senko, M.W., Hendrickson, C.L., Emmett, M.R., Shi, S.D.H., Marshall, A.G.: External accumulation of ions for enhanced electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. J Am Soc Mass Spectr. **8**, 970-976 (1997)
113. Belov, M.E., Nikolaev, E.N., Anderson, G.A., Udseth, H.R., Conrads, T.P., Veenstra, T.D., et al.: Design and performance of an ESI interface for selective external ion accumulation coupled to a Fourier transform ion cyclotron mass spectrometer. Anal Chem. **73**, 253-261 (2001)
114. Belov, M.E., Anderson, G.A., Wingerd, M.A., Udseth, H.R., Tang, K., Prior, D.C., et al.: An automated high performance capillary liquid chromatography-Fourier transform ion cyclotron resonance mass spectrometer for high-throughput proteomics. J Am Soc Mass Spectrom. **15**, 212-232 (2004)

115. Meng, F., Cargile, B.J., Patrie, S.M., Johnson, J.R., McLoughlin, S.M., Kelleher, N.L.: Processing complex mixtures of intact proteins for direct analysis by mass spectrometry. Anal Chem. **74**, 2923-2929 (2002)
116. Du, Y., Meng, F., Patrie, S.M., Miller, L.M., Kelleher, N.L.: Improved molecular weight-based processing of intact proteins for interrogation by quadrupole-enhanced FT MS/MS. J Proteome Res. **3**, 801-806 (2004)
117. Roth, M.J., Forbes, A.J., Boyne, M.T., 2nd, Kim, Y.B., Robinson, D.E., Kelleher, N.L.: Precise and parallel characterization of coding polymorphisms, alternative splicing, and modifications in human proteins by mass spectrometry. Mol Cell Proteomics. **4**, 1002-1008 (2005)
118. Syka, J.E., Marto, J.A., Bai, D.L., Horning, S., Senko, M.W., Schwartz, J.C., et al.: Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. J Proteome Res. **3**, 621-626 (2004)
119. Parks, B.A., Jiang, L., Thomas, P.M., Wenger, C.D., Roth, M.J., Boyne, M.T., 2nd, et al.: Top-down proteomics on a chromatographic time scale using linear ion trap fourier transform hybrid mass spectrometers. Anal Chem. **79**, 7984-7991 (2007)
120. Ryan, C.M., Souda, P., Bassilian, S., Ujwal, R., Zhang, J., Abramson, J., et al.: Post-translational modifications of integral membrane proteins resolved by top-down Fourier transform mass spectrometry with collisionally activated dissociation. Mol Cell Proteomics. **9**, 791-803 (2010)
121. Thangaraj, B., Ryan, C.M., Souda, P., Krause, K., Faull, K.F., Weber, A.P., et al.: Data-directed top-down Fourier-transform mass spectrometry of a large integral membrane protein complex: photosystem II from Galdieria sulphuraria. Proteomics. **10**, 3644-3656 (2010)
122. Wenger, C.D., Boyne, M.T., 2nd, Ferguson, J.T., Robinson, D.E., Kelleher, N.L.: Versatile online-offline engine for automated acquisition of high-resolution tandem mass spectra. Anal Chem. **80**, 8055-8063 (2008)
123. Kellie, J.F., Catherman, A.D., Durbin, K.R., Tran, J.C., Tipton, J.D., Norris, J.L., et al.: Robust analysis of the yeast proteome under 50 kDa by molecular-mass-based fractionation and top-down mass spectrometry. Anal Chem. **84**, 209-215 (2012)
124. Tran, J.C., Zamdborg, L., Ahlf, D.R., Lee, J.E., Catherman, A.D., Durbin, K.R., et al.: Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature. **480**, 254-258 (2011)
125. Schaub, T.M., Hendrickson, C.L., Horning, S., Quinn, J.P., Senko, M.W., Marshall, A.G.: High-performance mass spectrometry: Fourier transform ion cyclotron resonance at 14.5 Tesla. Anal Chem. **80**, 3985-3990 (2008)
126. Makarov, A.: Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal Chem. **72**, 1156-1162 (2000)
127. Hu, Q.Z., Noll, R.J., Li, H.Y., Makarov, A., Hardman, M., Cooks, R.G.: The Orbitrap: a new mass spectrometer. J Mass Spectrom. **40**, 430-443 (2005)
128. Lange, O., Damoc, E., Wieghaus, A., Makarov, A.: Enhanced Fourier transform for Orbitrap mass spectrometry. Int J Mass Spectrom. **369**, 16-22 (2014)
129. Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., et al.: Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. Anal Chem. **78**, 2113-2120 (2006)
130. Michalski, A., Damoc, E., Hauschild, J.P., Lange, O., Wieghaus, A., Makarov, A., et al.: Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol Cell Proteomics. **10**, M111 011015 (2011)
131. Senko, M.W., Remes, P.M., Canterbury, J.D., Mathur, R., Song, Q., Eliuk, S.M., et al.: Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. Anal Chem. **85**, 11710-11714 (2013)

132. Scigelova, M., Hornshaw, M., Giannakopulos, A., Makarov, A.: Fourier Transform Mass Spectrometry. Molecular & Cellular Proteomics. **10**, (2011)
133. Dyachenko, A., Wang, G., Belov, M., Makarov, A., de Jong, R.N., van den Bremer, E.T., et al.: Tandem Native Mass-Spectrometry on Antibody-Drug Conjugates and Submillion Da Antibody-Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. Anal Chem. **87**, 6095-6102 (2015)
134. Rose, R.J., Damoc, E., Denisov, E., Makarov, A., Heck, A.J.: High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. Nat Methods. **9**, 1084-1086 (2012)
135. Belov, M.E., Damoc, E., Denisov, E., Compton, P.D., Horning, S., Makarov, A.A., et al.: From protein complexes to subunit backbone fragments: a multi-stage approach to native mass spectrometry. Anal Chem. **85**, 11163-11173 (2013)
136. Ben-Nissan, G., Belov, M.E., Morgenstern, D., Levin, Y., Dym, O., Arkind, G., et al.: Triple-Stage Mass Spectrometry Unravels the Heterogeneity of an Endogenous Protein Complex. Anal Chem. **89**, 4708-4715 (2017)
137. Chernushevich, I.V., Loboda, A.V., Thomson, B.A.: An introduction to quadrupole-time-of-flight mass spectrometry. J Mass Spectrom. **36**, 849-865 (2001)
138. Heck, A.J.R., van den Heuvel, R.H.H.: Investigation of intact protein complexes by mass spectrometry. Mass Spectrom Rev. **23**, 368-389 (2004)
139. Christ, P., Rutzinger, S., Seidel, W., Uchaikin, S., Pro, F., Koy, C., et al.: High detection sensitivity achieved with cryogenic detectors in combination with matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Eur J Mass Spectrom (Chichester). **10**, 469-476 (2004)
140. Chen, X.Y., Westphall, M.S., Smith, L.M.: Mass spectrometric analysis of DNA mixtures: Instrumental effects responsible for decreased sensitivity with increasing mass. Anal Chem. **75**, 5944-5952 (2003)
141. van Remoortere, A., van Zeijl, R.J.M., van den Oever, N., Franck, J., Longuespee, R., Wisztorski, M., et al.: MALDI Imaging and Profiling MS of Higher Mass Proteins from Tissue. J Am Soc Mass Spectr. **21**, 1922-1929 (2010)
142. Sobott, F., Benesch, J.L., Vierling, E., Robinson, C.V.: Subunit exchange of multimeric protein complexes. Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. J Biol Chem. **277**, 38921-38929 (2002)
143. Wells, J.M., McLuckey, S.A.: Collision-induced dissociation (CID) of peptides and proteins. Methods Enzymol. **402**, 148-185 (2005)
144. Sobott, F., Watt, S.J., Smith, J., Edelmann, M.J., Kramer, H.B., Kessler, B.M.: Comparison of CID versus ETD based MS/MS fragmentation for the analysis of protein ubiquitination. J Am Soc Mass Spectrom. **20**, 1652-1659 (2009)
145. Hopfgartner, G., Husser, C., Zell, M.: Rapid screening and characterization of drug metabolites using a new quadrupole-linear ion trap mass spectrometer. J Mass Spectrom. **38**, 138-150 (2003)
146. Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S., Mann, M.: Higher-energy C-trap dissociation for peptide modification analysis. Nature Methods. **4**, 709-712 (2007)
147. Beu, S.C., Senko, M.W., Quinn, J.P., Wampler, F.M., McLafferty, F.W.: Fourier-transform electrospray instrumentation for tandem high-resolution mass spectrometry of large molecules. J Am Soc Mass Spectrom. **4**, 557-565 (1993)
148. Zubarev, R.A., Kelleher, N.L., McLafferty, F.W.: Electron capture dissociation of multiply charged protein cations. A nonergodic process. J Am Chem Soc. **120**, 3265-3266 (1998)
149. Cooper, H.J., Hakansson, K., Marshall, A.G.: The role of electron capture dissociation in biomolecular analysis. Mass Spectrom Rev. **24**, 201-222 (2005)

150. Zubarev, R.A., Horn, D.M., Fridriksson, E.K., Kelleher, N.L., Kruger, N.A., Lewis, M.A., et al.: Electron capture dissociation for structural characterization of multiply charged protein cations. Anal Chem. **72**, 563-573 (2000)
151. Breuker, K., Oh, H.B., Lin, C., Carpenter, B.K., McLafferty, F.W.: Nonergodic and conformational control of the electron capture dissociation of protein cations. P Natl Acad Sci USA. **101**, 14011-14016 (2004)
152. Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., Hunt, D.F.: Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. P Natl Acad Sci USA. **101**, 9528-9533 (2004)
153. Brunner, A.M., Lossl, P., Liu, F., Huguet, R., Mullen, C., Yamashita, M., et al.: Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. Anal Chem. **87**, 4152-4158 (2015)
154. Frese, C.K., Altelaar, A.F.M., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A.J.R., et al.: Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. Anal Chem. **84**, 9668-9673 (2012)
155. Frese, C.K., Zhou, H.J., Taus, T., Altelaar, A.F.M., Mechter, K., Heck, A.J.R., et al.: Unambiguous Phosphosite Localization using Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). Journal of Proteome Research. **12**, 1520-1525 (2013)
156. O'Brien, J.P., Li, W.Z., Zhang, Y., Brodbelt, J.S.: Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry. J Am Chem Soc. **136**, 12920-12928 (2014)
157. Wysocki, V.H., Joyce, K.E., Jones, C.M., Beardsley, R.L.: Surface-induced dissociation of small molecules, peptides,and non-covalent protein complexes. J Am Soc Mass Spectr. **19**, 190-208 (2008)
158. Cooks, R.G., Ast, T., Beynon, J.H.: Anomalous Metastable Peaks. International Journal of Mass Spectrometry and Ion Processes. **16**, 348-352 (1975)
159. Cooks, R.G., Terwilliger, D.T., Ast, T., Beynon, J.H., Keough, T.: Surface Modified Mass-Spectrometry. J Am Chem Soc. **97**, 1583-1585 (1975)
160. Mabud, M.D.A., Dekrey, M.J., Cooks, R.G.: Surface-Induced Dissociation of Molecular-Ions. International Journal of Mass Spectrometry and Ion Processes. **67**, 285-294 (1985)
161. Wysocki, V.H., Jones, C.M., Galhena, A.S., Blackwell, A.E.: Surface-induced dissociation shows potential to be more informative than collision-induced dissociation for structural studies of large systems. J Am Soc Mass Spectr. **19**, 903-913 (2008)
162. Rogers, B.J., Birdsall, R.E., Wu, Z., Wirth, M.J.: RPLC of intact proteins using sub-0.5 mum particles and commercial instrumentation. Anal Chem. **85**, 6820-6825 (2013)
163. Muneeruddin, K., Nazzaro, M., Kaltashov, I.A.: Characterization of intact protein conjugates and biopharmaceuticals using ion-exchange chromatography with online detection by native electrospray ionization mass spectrometry and top-down tandem mass spectrometry. Analytical chemistry. **87**, 10138-10145 (2015)
164. Wagner, K., Racaityte, K., Unger, K.K., Miliotis, T., Edholm, L.E., Bischoff, R., et al.: Protein mapping by two-dimensional high performance liquid chromatography. J Chromatogr A. **893**, 293-305 (2000)
165. Eubel, H., Braun, H.P., Millar, A.H.: Blue-native PAGE in plants: a tool in analysis of protein-protein interactions. Plant Methods. **1**, (2005)
166. Gundry, R.L., White, M.Y., Murray, C.I., Kane, L.A., Fu, Q., Stanley, B.A., et al.: Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. Curr Protoc Mol Biol. **Chapter 10**, Unit10 25 (2009)

167. Tran, J.C., Doucette, A.A.: Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. Analytical chemistry. **80**, 1568-1573 (2008)
168. Skinner, O.S., Do Vale, L.H.F., Catherman, A.D., Havugimana, P.C., de Sousa, M.V., Compton, P.D., et al.: Native GELFrEE: A New Separation Technique for Biomolecular Assemblies. Anal Chem. **87**, 3032-3038 (2015)
169. Konijnenberg, A., Butterer, A., Sobott, F.: Native ion mobility-mass spectrometry and related methods in structural biology. Biochim Biophys Acta. **1834**, 1239-1256 (2013)
170. Harvey, S.R., Macphee, C.E., Barran, P.E.: Ion mobility mass spectrometry for peptide analysis. Methods. **54**, 454-461 (2011)
171. Lanucara, F., Holman, S.W., Gray, C.J., Eyers, C.E.: The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. Nat Chem. **6**, 281-294 (2014)
172. Ibrahim, Y.M., Garimella, S.V., Prost, S.A., Wojcik, R., Norheim, R.V., Baker, E.S., et al.: Development of an Ion Mobility Spectrometry-Orbitrap Mass Spectrometer Platform. Anal Chem. **88**, 12152-12160 (2016)
173.
174. Creese, A.J., Cooper, H.J.: Separation and identification of isomeric glycopeptides by high field asymmetric waveform ion mobility spectrometry. Anal Chem. **84**, 2597-2601 (2012)
175. Borysik, A.J., Read, P., Little, D.R., Bateman, R.H., Radford, S.E., Ashcroft, A.E.: Separation of beta2-microglobulin conformers by high-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled to electrospray ionisation mass spectrometry. Rapid Commun Mass Spectrom. **18**, 2229-2234 (2004)
176. Purves, R.W., Barnett, D.A., Guevremont, R.: Separation of protein conformers using electrospray-high field asymmetric waveform ion mobility spectrometry-mass spectrometry. Int J Mass Spectrom. **197**, 163-177 (2000)
177. Huang, X.H., Coleman, W.F., Zare, R.N.: Analysis of Factors Causing Peak Broadening in Capillary Zone Electrophoresis. Journal of Chromatography. **480**, 95-110 (1989)
178. Creamer, J.S., Oborny, N.J., Lunte, S.M.: Recent advances in the analysis of therapeutic proteins by capillary and microchip electrophoresis. Anal Methods-Uk. **6**, 5427-5449 (2014)
179. Gilges, M., Kleemiss, M.H., Schomburg, G.: Capillary Zone Electrophoresis Separations of Basic and Acidic Proteins Using Poly(Vinyl Alcohol) Coatings in Fused-Silica Capillaries. Anal Chem. **66**, 2038-2046 (1994)
180. Horvath, J., Dolnik, V.: Polymer wall coatings for capillary electrophoresis. Electrophoresis. **22**, 644-655 (2001)
181. Righetti, P.G., Gelfi, C., Verzola, B., Castelletti, L.: The state of the art of dynamic coatings. Electrophoresis. **22**, 603-611 (2001)
182. Thompson, T.J., Foret, F., Vouros, P., Karger, B.L.: Capillary Electrophoresis Electrospray Ionization Mass-Spectrometry - Improvement of Protein-Detection Limits Using on-Column Transient Isotachophoretic Sample Preconcentration. Anal Chem. **65**, 900-906 (1993)
183. Osbourn, D.M., Weiss, D.J., Lunte, C.E.: On-line preconcentration methods for capillary electrophoresis. Electrophoresis. **21**, 2768-2779 (2000)
184. Vanderschans, M.J., Beckers, J.L., Molling, M.C., Everaerts, F.M.: Intrinsic Isotachophoretic Preconcentration in Capillary Gel-Electrophoresis of DNA Restriction Fragments. J Chromatogr A. **717**, 139-147 (1995)
185. Zhang, C.X., Thormann, W.: Head-column field-amplified sample stacking in binary system capillary electrophoresis. 2. Optimization with a preinjection plug and application to micellar electrokinetic chromatography. Anal Chem. **70**, 540-548 (1998)

186. Zhang, C.X., Thormann, W.: Head-Column Field-Amplified Sample Stacking in Binary System Capillary Electrophoresis: A Robust Approach Providing over 1000-Fold Sensitivity Enhancement. Anal Chem. **68**, 2523-2532 (1996)
187. Zhao, Y., Lunte, C.E.: pH-mediated field amplification on-column preconcentration of anions in physiological samples for capillary electrophoresis. Anal Chem. **71**, 3985-3991 (1999)
188. Zhao, Y., McLaughlin, K., Lunte, C.E.: On-column sample preconcentration using sample matrix switching and field amplification for increased sensitivity of capillary electrophoretic analysis of physiological samples. Anal Chem. **70**, 4578-4585 (1998)
189. Palmarsdottir, S., Thordarson, E., Edholm, L.E., Jonsson, J.A., Mathiasson, L.: Miniaturized supported liquid membrane device for selective on-line enrichment of basic drugs in plasma combined with capillary zone electrophoresis. Anal Chem. **69**, 1732-1737 (1997)
190. Cui, W., Rohrs, H.W., Gross, M.L.: Top-down mass spectrometry: recent developments, applications and perspectives. Analyst. **136**, 3854-3864 (2011)
191. Horn, D.M., Zubarev, R.A., McLafferty, F.W.: Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. J Am Soc Mass Spectrom. **11**, 320-332 (2000)
192. Mayampurath, A.M., Jaitly, N., Purvine, S.O., Monroe, M.E., Auberry, K.J., Adkins, J.N., et al.: DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. Bioinformatics. **24**, 1021-1023 (2008)
193. Jaitly, N., Mayampurath, A., Littlefield, K., Adkins, J.N., Anderson, G.A., Smith, R.D.: Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. BMC Bioinformatics. **10**, 87 (2009)
194. Zabrouskov, V., Senko, M.W., Du, Y., Leduc, R.D., Kelleher, N.L.: New and automated MSn approaches for top-down identification of modified proteins. J Am Soc Mass Spectrom. **16**, 2027-2038 (2005)
195.
196. Kinter, M., Sherman, N.E. John Wiley & Sons, (2005)
197. LeDuc, R.D., Taylor, G.K., Kim, Y.B., Januszyk, T.E., Bynum, L.H., Sola, J.V., et al.: ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. Nucleic Acids Res. **32**, W340-345 (2004)
198. Bern, M., Kil, Y.J., Becker, C.: Byonic: advanced peptide and protein identification software. Curr Protoc Bioinformatics. **Chapter 13**, Unit13 20 (2012)
199. Olsen, J.V., Schwartz, J.C., Griep-Raming, J., Nielsen, M.L., Damoc, E., Denisov, E., et al.: A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. Mol Cell Proteomics. **8**, 2759-2769 (2009)
200. Kim, T., Tolmachev, A.V., Harkewicz, R., Prior, D.C., Anderson, G., Udseth, H.R., et al.: Design and implementation of a new electrodynamic ion funnel. Anal Chem. **72**, 2247-2255 (2000)
201. Rosati, S., Rose, R.J., Thompson, N.J., van Duijn, E., Damoc, E., Denisov, E., et al.: Exploring an Orbitrap Analyzer for the Characterization of Intact Antibodies by Native Mass Spectrometry. Angew Chem Int Edit. **51**, 12992-12996 (2012)
202. Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Muller, M., et al.: Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. Molecular & cellular proteomics : MCP. **11**, O111 013698 (2012)
203. Lindsley, C.W.: The Top Prescription Drugs of 2012 Globally: Biologics Dominate, But Small Molecule CNS Drugs Hold on to Top Spots. ACS Chemical Neuroscience. **4**, 905-907 (2013)
204. Smith, L.M., Kelleher, N.L.: Proteoform: a single term describing protein complexity. Nat Meth. **10**, 186-187 (2013)

205. Eon-Duval, A., Broly, H., Gleixner, R.: Quality attributes of recombinant therapeutic proteins: An assessment of impact on safety and efficacy as part of a quality by design development approach. Biotechnology Progress. **28**, 608-622 (2012)
206. van Beers, M.C., Sauerborn, M., Gilli, F., Brinks, V., Schellekens, H., Jiskoot, W.: Oxidized and Aggregated Recombinant Human Interferon Beta is Immunogenic in Human Interferon Beta Transgenic Mice. Pharm Res. **28**, 2393-2402 (2011)
207. Vázquez-Rey, M., Lang, D.A.: Aggregates in monoclonal antibody manufacturing processes. Biotechnology and Bioengineering. **108**, 1494-1508 (2011)
208. Dissing-Olesen, L., Thaysen-Andersen, M., Meldgaard, M., Højrup, P., Finsen, B.: The Function of the Human Interferon-β1a Glycan Determined in Vivo. Journal of Pharmacology and Experimental Therapeutics. **326**, 338-347 (2008)
209. Takeuchi, M., Inoue, N., Strickland, T.W., Kubota, M., Wada, M., Shimizu, R., et al.: Relationship between sugar chain structure and biological activity of recombinant human erythropoietin produced in Chinese hamster ovary cells. Proceedings of the National Academy of Sciences. **86**, 7819-7822 (1989)
210. Shinkawa, T., Nakamura, K., Yamane, N., Shoji-Hosaka, E., Kanda, Y., Sakurada, M., et al.: The Absence of Fucose but Not the Presence of Galactose or Bisecting N-Acetylglucosamine of Human IgG1 Complex-type Oligosaccharides Shows the Critical Role of Enhancing Antibody-dependent Cellular Cytotoxicity. Journal of Biological Chemistry. **278**, 3466-3473 (2003)
211. Cacia, J., Keck, R., Presta, L.G., Frenz, J.: Isomerization of an Aspartic Acid Residue in the Complementarity-Determining Regions of a Recombinant Antibody to Human IgE: Identification and Effect on Binding Affinity. Biochemistry. **35**, 1897-1903 (1996)
212. Vlasak, J., Bussat, M.C., Wang, S., Wagner-Rousset, E., Schaefer, M., Klinguer-Hamour, C., et al.: Identification and characterization of asparagine deamidation in the light chain CDR1 of a humanized IgG1 antibody. Analytical Biochemistry. **392**, 145-154 (2009)
213. Huang, L., Lu, J., Wroblewski, V.J., Beals, J.M., Riggin, R.M.: In Vivo Deamidation Characterization of Monoclonal Antibody by LC/MS/MS. Analytical Chemistry. **77**, 1432-1439 (2005)
214. Boja, E.S., Fales, H.M.: Overalkylation of a protein digest with iodoacetamide. Anal Chem. **73**, 3576-3582 (2001)
215. Li, X., Cournoyer, J., Lin, C., O'Connor, P.: Use of 18O labels to monitor deamidation during protein and peptide sample processing. J. Am. Soc. Mass Spectrom. **19**, 855-864 (2008)
216. Perdivara, I., Deterding, L., Przybylski, M., Tomer, K.: Mass spectrometric identification of oxidative modifications of tryptophan residues in proteins: Chemical artifact or post-translational modification? J. Am. Soc. Mass Spectrom. **21**, 1114-1117 (2010)
217. Hao, P., Ren, Y., Alpert, A.J., Sze, S.K.: Detection, Evaluation and Minimization of Nonenzymatic Deamidation in Proteomic Sample Preparation. Molecular & Cellular Proteomics. **10**, (2011)
218. Du, Y., Wang, F., May, K., Xu, W., Liu, H.: Determination of Deamidation Artifacts Introduced by Sample Preparation Using 18O-Labeling and Tandem Mass Spectrometry Analysis. Analytical Chemistry. **84**, 6355-6360 (2012)
219. Nielsen, M.L., Vermeulen, M., Bonaldi, T., Cox, J., Moroder, L., Mann, M.: Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. Nat Meth. **5**, 459-460 (2008)
220. Bischoff, R., Bronsema, K.J., van de Merbel, N.C.: Analysis of biopharmaceutical proteins in biological matrices by LC-MS/MS I. Sample preparation. TrAC Trends in Analytical Chemistry. **48**, 41-51 (2013)
221. Zabrouskov, V., Han, X., Welker, E., Zhai, H., Lin, C., van Wijk, K.J., et al.: Stepwise Deamidation of Ribonuclease A at Five Sites Determined by Top Down Mass Spectrometry†. Biochemistry. **45**, 987-992 (2005)

222. Pesavento, J.J., Kim, Y.-B., Taylor, G.K., Kelleher, N.L.: Shotgun Annotation of Histone Modifications: A New Approach for Streamlined Characterization of Proteins by Top Down Mass Spectrometry. Journal of the American Chemical Society. **126**, 3386-3387 (2004)
223. Boyne, M.T., Pesavento, J.J., Mizzen, C.A., Kelleher, N.L.: Precise Characterization of Human Histones in the H2A Gene Family by Top Down Mass Spectrometry. Journal of Proteome Research. **5**, 248-253 (2006)
224. Thomas, C.E., Kelleher, N.L., Mizzen, C.A.: Mass Spectrometric Characterization of Human Histone H3: A Bird's Eye View. Journal of Proteome Research. **5**, 240-247 (2006)
225. Garcia, B.A., Pesavento, J.J., Mizzen, C.A., Kelleher, N.L.: Pervasive combinatorial modification of histone H3 in human cells. Nat Meth. **4**, 487-489 (2007)
226. Pesavento, J.J., Bullock, C.R., LeDuc, R.D., Mizzen, C.A., Kelleher, N.L.: Combinatorial Modification of Human Histone H4 Quantitated by Two-dimensional Liquid Chromatography Coupled with Top Down Mass Spectrometry. Journal of Biological Chemistry. **283**, 14927-14937 (2008)
227. Pesavento, J.J., Yang, H., Kelleher, N.L., Mizzen, C.A.: Certain and Progressive Methylation of Histone H4 at Lysine 20 during the Cell Cycle. Molecular and Cellular Biology. **28**, 468-486 (2008)
228. Sandra, K., Vandenheede, I., Sandra, P.: Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization. Journal of Chromatography A. **1335**, 81-103 (2014)
229. Haselberg, R., de Jong, G.J., Somsen, G.W.: Low-Flow Sheathless Capillary Electrophoresis–Mass Spectrometry for Sensitive Glycoform Profiling of Intact Pharmaceutical Proteins. Analytical Chemistry. **85**, 2289-2296 (2013)
230. Lyubarskaya, Y., Houde, D., Woodard, J., Murphy, D., Mhatre, R.: Analysis of recombinant monoclonal antibody isoforms by electrospray ionization mass spectrometry as a strategy for streamlining characterization of recombinant monoclonal antibody charge heterogeneity. Analytical Biochemistry. **348**, 24-39 (2006)
231. Wheeler, T.D., Sun, J.L., Pleiner, S., Geier, H., Dobberthien, P., Studts, J., et al.: Microchip Zone Electrophoresis for High-Throughput Analysis of Monoclonal Antibody Charge Variants. Analytical Chemistry. (2014)
232. Wojcik, R., Dada, O.O., Sadilek, M., Dovichi, N.J.: Simplified capillary electrophoresis nanospray sheath-flow interface for high efficiency and sensitive peptide analysis. Rapid Communications in Mass Spectrometry. **24**, 2554-2560 (2010)
233. Sun, L., Knierman, M.D., Zhu, G., Dovichi, N.J.: Fast Top-Down Intact Protein Characterization with Capillary Zone Electrophoresis–Electrospray Ionization Tandem Mass Spectrometry. Analytical Chemistry. **85**, 5989-5995 (2013)
234. Moini, M.: Simplifying CE−MS Operation. 2. Interfacing Low-Flow Separation Techniques to Mass Spectrometry Using a Porous Tip. Analytical Chemistry. **79**, 4241-4246 (2007)
235. Busnel, J.-M., Schoenmaker, B., Ramautar, R., Carrasco-Pancorbo, A., Ratnayake, C., Feitelson, J.S., et al.: High Capacity Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry: Coupling a Porous Sheathless Interface with Transient-Isotachophoresis. Analytical Chemistry. **82**, 9476-9483 (2010)
236. Heemskerk, A.A.M., Busnel, J.-M., Schoenmaker, B., Derks, R.J.E., Klychnikov, O., Hensbergen, P.J., et al.: Ultra-Low Flow Electrospray Ionization-Mass Spectrometry for Improved Ionization Efficiency in Phosphoproteomics. Analytical Chemistry. **84**, 4552-4559 (2012)
237. Haselberg, R., Ratnayake, C.K., de Jong, G.J., Somsen, G.W.: Performance of a sheathless porous tip sprayer for capillary electrophoresis–electrospray ionization-mass spectrometry of intact proteins. Journal of Chromatography A. **1217**, 7605-7611 (2010)

238. Thakur, D., Rejtar, T., Karger, B.L., Washburn, N.J., Bosques, C.J., Gunay, N.S., et al.: Profiling the Glycoforms of the Intact α Subunit of Recombinant Human Chorionic Gonadotropin by High-Resolution Capillary Electrophoresis−Mass Spectrometry. Analytical Chemistry. **81**, 8900-8907 (2009)
239. Taichrib, A., Pioch, M., Neususs, C.: Toward a screening method for the analysis of small intact proteins by CE-ESI-TOF MS. Electrophoresis. **33**, 1356-1366 (2012)
240. Han, X., Wang, Y., Aslanian, A., Bern, M., Lavallée-Adam, M., Yates, J.R.: Sheathless Capillary Electrophoresis-Tandem Mass Spectrometry for Top-Down Characterization of Pyrococcus furiosus Proteins on a Proteome Scale. Analytical Chemistry. (2014)
241. Li, Y., Compton, P.D., Tran, J.C., Ntai, I., Kelleher, N.L.: Optimizing capillary electrophoresis for top-down proteomics of 30–80 kDa proteins. PROTEOMICS. **14**, 1158-1164 (2014)
242. Zhao, Y., Sun, L., Champion, M.M., Knierman, M.D., Dovichi, N.J.: Capillary Zone Electrophoresis–Electrospray Ionization-Tandem Mass Spectrometry for Top-Down Characterization of the Mycobacterium marinum Secretome. Analytical Chemistry. **86**, 4873-4878 (2014)
243. Rudick, R.A., Goodkin, D.E., Jacobs, L.D., Cookfair, D.L., Herndon, R.M., Richert, J.R., et al.: Impact of interferon beta-1a on neurologic disability in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG). Neurology. **49**, 358-363 (1997)
244. Orru, S., Amoresano, A., Siciliano, R., Napoleoni, R., Finocchiaro, O., Datola, A., et al.: Structural analysis of modified forms of recombinant IFN-beta produced under stress-simulating conditions. Biol Chem. **381**, 7-17 (2000)
245. Conradt, H.S., Egge, H., Peter-Katalinic, J., Reiser, W., Siklosi, T., Schaper, K.: Structure of the carbohydrate moiety of human interferon-beta secreted by a recombinant Chinese hamster ovary cell line. Journal of Biological Chemistry. **262**, 14600-14605 (1987)
246.
247. Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., et al.: A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotech. **30**, 918-920 (2012)
248. Córdova, E., Gao, J., Whitesides, G.M.: Noncovalent Polycationic Coatings for Capillaries in Capillary Electrophoresis of Proteins. Analytical Chemistry. **69**, 1370-1379 (1997)
249. Haselberg, R., de Jong, G.J., Somsen, G.W.: Capillary electrophoresis of intact basic proteins using noncovalently triple-layer coated capillaries. Journal of Separation Science. **32**, 2408-2415 (2009)
250. Haselberg, R., de Jong, G.J., Somsen, G.W.: Capillary electrophoresis–mass spectrometry of intact basic proteins using Polybrene–dextran sulfate–Polybrene-coated capillaries: System optimization and performance. Analytica Chimica Acta. **678**, 128-134 (2010)
251. Horvath, J., Dolník, V.: Polymer wall coatings for capillary electrophoresis. Electrophoresis. **22**, 644-655 (2001)
252. Jefferis, R.: Glycosylation of Recombinant Antibody Therapeutics. Biotechnology Progress. **21**, 11-16 (2005)
253. North, S.J., Huang, H.-H., Sundaram, S., Jang-Lee, J., Etienne, A.T., Trollope, A., et al.: Glycomics Profiling of Chinese Hamster Ovary Cell Glycosylation Mutants Reveals N-Glycans of a Novel Size and Complexity. Journal of Biological Chemistry. **285**, 5759-5775 (2010)
254. Li, Y., Compton, P.D., Tran, J.C., Ntai, I., Kelleher, N.L.: Optimizing capillary electrophoresis for top-down proteomics of 30-80 kDa proteins. Proteomics. **14**, 1158-1164 (2014)
255. Haselberg, R., de Jong, G.J., Somsen, G.W.: Capillary electrophoresis of intact basic proteins using noncovalently triple-layer coated capillaries. J Sep Sci. **32**, 2408-2415 (2009)
256. Song, E., Pyreddy, S., Mechref, Y.: Quantification of glycopeptides by multiple reaction monitoring liquid chromatography/tandem mass spectrometry. Rapid Communications in Mass Spectrometry. **26**, 1941-1954 (2012)

257. Song, T., Ozcan, S., Becker, A., Lebrilla, C.B.: In-Depth Method for the Characterization of Glycosylation in Manufactured Recombinant Monoclonal Antibody Drugs. Analytical Chemistry. **86**, 5661-5666 (2014)
258. Karpusas, M., Nolte, M., Benton, C.B., Meier, W., Lipscomb, W.N., Goelz, S.: The crystal structure of human interferon β at 2.2-Å resolution. Proceedings of the National Academy of Sciences. **94**, 11813-11818 (1997)
259. Rudd, P.M., Morgan, B.P., Wormald, M.R., Harvey, D.J., van den Berg, C.W., Davis, S.J., et al.: The Glycosylation of the Complement Regulatory Protein, Human Erythrocyte CD59. Journal of Biological Chemistry. **272**, 7229-7244 (1997)
260. Mittermayr, S., Bones, J., Doherty, M., Guttman, A., Rudd, P.M.: Multiplexed Analytical Glycomics: Rapid and Confident IgG N-Glycan Structural Elucidation. Journal of Proteome Research. **10**, 3820-3829 (2011)
261. Yang, J.M., Ai, J., Bao, Y., Yuan, Z., Qin, Y., Xie, Y.W., et al.: Investigation of the correlation between charge and glycosylation of IgG1 variants by liquid chromatography-mass spectrometry. Anal Biochem. **448**, 82-91 (2014)
262. Chudasama, V., Maruani, A., Caddick, S.: Recent advances in the construction of antibody-drug conjugates. Nat Chem. **8**, 113-118 (2016)
263. Khawli, L.A., Goswami, S., Hutchinson, R., Kwong, Z.W., Yang, J., Wang, X., et al.: Charge variants in IgG1: Isolation, characterization, in vitro binding properties and pharmacokinetics in rats. mAbs. **2**, 613-624 (2010)
264. Siuti, N., Kelleher, N.L.: Decoding protein modifications using top-down mass spectrometry. Nat Methods. **4**, 817-821 (2007)
265. Fornelli, L., Ayoub, D., Aizikov, K., Beck, A., Tsybin, Y.O.: Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. Anal Chem. **86**, 3005-3012 (2014)
266. Han, M., Rock, B.M., Pearson, J.T., Rock, D.A.: Intact mass analysis of monoclonal antibodies by capillary electrophoresis-Mass spectrometry. J Chromatogr B. **1011**, 24-32 (2016)
267. Bondarenko, P.V., Second, T.P., Zabrouskov, V., Makarov, A.A., Zhang, Z.: Mass measurement and top-down HPLC/MS analysis of intact monoclonal antibodies on a hybrid linear quadrupole ion trap-Orbitrap mass spectrometer. J Am Soc Mass Spectrom. **20**, 1415-1424 (2009)
268. Gahoual, R., Beck, A., Leize-Wagner, E., Francois, Y.N.: Cutting-edge capillary electrophoresis characterization of monoclonal antibodies and related products. J Chromatogr B Analyt Technol Biomed Life Sci. **1032**, 61-78 (2016)
269. Tran, B.Q., Barton, C., Feng, J., Sandjong, A., Yoon, S.H., Awasthi, S., et al.: Glycosylation characterization of therapeutic mAbs by top- and middle-down mass spectrometry. Data Brief. **6**, 68-76 (2016)
270. Dyachenko, A., Wang, G., Belov, M., Makarov, A., de Jong, R.N., van den Bremer, E.T., et al.: Tandem Native Mass-Spectrometry on Antibody-Drug Conjugates and Submillion Da Antibody-Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. Analytical Chemistry. **87**, 6095-6102 (2015)
271. Kukrer, B., Filipe, V., van Duijn, E., Kasper, P.T., Vreeken, R.J., Heck, A.J.R., et al.: Mass Spectrometric Analysis of Intact Human Monoclonal Antibody Aggregates Fractionated by Size-Exclusion Chromatography. Pharm Res-Dordr. **27**, 2197-2204 (2010)
272. Haverick, M., Mengisen, S., Shameem, M., Ambrogelly, A.: Separation of mAbs molecular variants by analytical hydrophobic interaction chromatography HPLC Overview and applications. Mabs. **6**, 852-858 (2014)
273. D'Atri, V., Fekete, S., Beck, A., Lauber, M.A., Guillarme, D.: Hydrophilic Interaction Chromatography (HILIC) hyphenated with Mass Spectrometry: a powerful analytical tool for the

comparison of originator and biosimilar therapeutic monoclonal antibodies at the middle-up level of analysis. (2017)
274. Haselberg, R., de Jong, G.J., Somsen, G.W.: Capillary electrophoresis-mass spectrometry for the analysis of intact proteins. Journal of Chromatography A. **1159**, 81-109 (2007)
275. Haselberg, R., Ratnayake, C.K., de Jong, G.J., Somsen, G.W.: Performance of a sheathless porous tip sprayer for capillary electrophoresis-electrospray ionization-mass spectrometry of intact proteins. Journal of Chromatography A. **1217**, 7605-7611 (2010)
276. Redman, E.A., Mellors, J.S., Starkey, J.A., Ramsey, J.M.: Characterization of Intact Antibody Drug Conjugate Variants Using Microfluidic Capillary Electrophoresis-Mass Spectrometry. Anal Chem. **88**, 2220-2226 (2016)
277. Verzola, B., Sebastiano, R., Righetti, P.G., Gelfi, C., Lapadula, M., Citterio, A.: Mechanism of action of quaternary diamino quenchers in capillary zone electrophoresis. Electrophoresis. **24**, 121-129 (2003)
278. Sarg, B., Faserl, K., Kremser, L., Halfinger, B., Sebastiano, R., Lindner, H.H.: Comparing and combining capillary electrophoresis electrospray ionization mass spectrometry and nano-liquid chromatography electrospray ionization mass spectrometry for the characterization of post-translationally modified histones. Molecular & cellular proteomics : MCP. **12**, 2640-2656 (2013)
279. Elhamili, A., Wetterhall, M., Arvidsson, B., Sebastiano, R., Righetti, P.G., Bergquist, J.: Rapid capillary electrophoresis time-of-flight mass spectrometry separations of peptides and proteins using a monoquaternarized piperazine compound (M7C4I) for capillary coatings. Electrophoresis. **29**, 1619-1625 (2008)
280. Sebastiano, R., Gelfi, C., Righetti, P.G., Citterio, A.: Omega-iodoalkylammonium salts as permanent capillary silica wall modifiers. Comparative analysis of their structural parameters and substituent effects. Journal of Chromatography A. **924**, 71-81 (2001)
281. Pontoglio, A., Vigano, A., Sebastiano, R., Citterio, A., Maragnoli, L., Righetti, P.G., et al.: Peptide and protein separations by capillary electrophoresis in the presence of mono- and diquaternarized diamines. Electrophoresis. **25**, 1065-1070 (2004)
282. Batz, N.G., Mellors, J.S., Alarie, J.P., Ramsey, J.M.: Chemical vapor deposition of aminopropyl silanes in microfluidic channels for highly efficient microchip capillary electrophoresis-electrospray ionization-mass spectrometry. Analytical Chemistry. **86**, 3493-3500 (2014)
283. Liu, H., Gaza-Bulseco, G., Faldu, D., Chumsae, C., Sun, J.: Heterogeneity of monoclonal antibodies. J Pharm Sci. **97**, 2426-2447 (2008)
284. Fu, M.X., Requena, J.R., Jenkins, A.J., Lyons, T.J., Baynes, J.W., Thorpe, S.R.: The advanced glycation end product, Nepsilon-(carboxymethyl)lysine, is a product of both lipid peroxidation and glycoxidation reactions. J Biol Chem. **271**, 9982-9986 (1996)
285. Liu, Y.D., Goetze, A.M., Bass, R.B., Flynn, G.C.: N-terminal glutamate to pyroglutamate conversion in vivo for human IgG2 antibodies. J Biol Chem. **286**, 11211-11217 (2011)
286. Alexandru, I., Kells, D.I.C., Dorrington, K.J., Klein, M.: Non-Covalent Association of Heavy and Light-Chains of Human Immunoglobulin-G - Studies Using Light Chain Labeled with a Fluorescent-Probe. Mol Immunol. **17**, 1351-1363 (1980)
287. Horne, C., Klein, M., Polidoulis, I., Dorrington, K.J.: Noncovalent association of heavy and light chains of human immunoglobulins. III. Specific interactions between VH and VL. J Immunol. **129**, 660-664 (1982)
288. Han, M., Rock, B.M., Pearson, J.T., Rock, D.A.: Intact mass analysis of monoclonal antibodies by capillary electrophoresis-Mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci. **1011**, 24-32 (2016)

289. An, Y., Zhang, Y., Mueller, H.M., Shameem, M., Chen, X.: A new tool for monoclonal antibody analysis: application of IdeS proteolysis in IgG domain-specific characterization. MAbs. **6**, 879-893 (2014)
290. Leblanc, Y., Ramon, C., Bihoreau, N., Chevreux, G.: Charge variants characterization of a monoclonal antibody by ion exchange chromatography coupled on-line to native mass spectrometry: Case study after a long-term storage at +5 degrees C. J Chromatogr B Analyt Technol Biomed Life Sci. **1048**, 130-139 (2017)
291. Heck, A.J.: Native mass spectrometry: a bridge between interactomics and structural biology. Nat Methods. **5**, 927-933 (2008)
292. Weiss, V.U., Bereszczazk, J.Z., Havlik, M., Kallinger, P., Gosler, I., Kumar, M., et al.: Analysis of a common cold virus and its subviral particles by gas-phase electrophoretic mobility molecular analysis and native mass spectrometry. Analytical Chemistry. **87**, 8709-8717 (2015)
293. van de Waterbeemd, M., Snijder, J., Tsvetkova, I.B., Dragnea, B.G., Cornelissen, J.J., Heck, A.J.: Examining the Heterogeneous Genome Content of Multipartite Viruses BMV and CCMV by Native Mass Spectrometry. J Am Soc Mass Spectrom. **27**, 1000-1009 (2016)
294. Saliou, J.M., Manival, X., Tillault, A.S., Atmanene, C., Bobo, C., Branlant, C., et al.: Combining native MS approaches to decipher archaeal box H/ACA ribonucleoprotein particle structure and activity. Proteomics. **15**, 2851-2861 (2015)
295. Blackwell, A.E., Dodds, E.D., Bandarian, V., Wysocki, V.H.: Revealing the quaternary structure of a heterogeneous noncovalent protein complex through surface-induced dissociation. Anal Chem. **83**, 2862-2865 (2011)
296. Zhang, H., Cui, W.D., Wen, J.Z., Blankenship, R.E., Gross, M.L.: Native Electrospray and Electron-Capture Dissociation in FTICR Mass Spectrometry Provide Top-Down Sequencing of a Protein Component in an Intact Protein Assembly. J Am Soc Mass Spectr. **21**, 1966-1968 (2010)
297. Haselberg, R., Brinks, V., Hawe, A., de Jong, G.J., Somsen, G.W.: Capillary electrophoresis-mass spectrometry using noncovalently coated capillaries for the analysis of biopharmaceuticals. Analytical and bioanalytical chemistry. **400**, 295-303 (2011)
298. Haselberg, R., de Jong, G.J., Somsen, G.W.: Low-flow sheathless capillary electrophoresis-mass spectrometry for sensitive glycoform profiling of intact pharmaceutical proteins. Analytical chemistry. **85**, 2289-2296 (2013)
299. Bush, D.R., Zang, L., Belov, A.M., Ivanov, A.R., Karger, B.L.: High Resolution CZE-MS Quantitative Characterization of Intact Biopharmaceutical Proteins: Proteoforms of Interferon-beta1. Analytical chemistry. (2015)
300. Li, S., Plouffe, B.D., Belov, A.M., Ray, S., Wang, X., Murthy, S.K., et al.: An Integrated Platform for Isolation, Processing and Mass Spectrometry-based Proteomic Profiling of Rare Cells in Whole Blood. Mol Cell Proteomics. (2015)
301. Ivanov, A.R., Zang, L., Karger, B.L.: Low-attomole electrospray ionization MS and MS/MS analysis of protein tryptic digests using 20-microm-i.d. polystyrene-divinylbenzene monolithic capillary columns. Anal Chem. **75**, 5306-5316. (2003)
302. Kukrer, B., Filipe, V., van Duijn, E., Kasper, P.T., Vreeken, R.J., Heck, A.J., et al.: Mass spectrometric analysis of intact human monoclonal antibody aggregates fractionated by size-exclusion chromatography. Pharmaceutical research. **27**, 2197-2204 (2010)
303. Debaene, F., Boeuf, A., Wagner-Rousset, E., Colas, O., Ayoub, D., Corvaia, N., et al.: Innovative native MS methodologies for antibody drug conjugate characterization: High resolution native MS and IM-MS for average DAR and DAR distribution assessment. Anal Chem. **86**, 10674-10683 (2014)

304. Gahoual, R., Heidenreich, A.K., Somsen, G.W., Bulau, P., Reusch, D., Wuhrer, M., et al.: Detailed characterization of monoclonal antibody receptor interaction using affinity liquid chromatography hyphenated to native mass spectrometry. Anal Chem. (2017)
305. Martinovic, S., Berger, S.J., Pasa-Tolic, L., Smith, R.D.: Separation and detection of intact noncovalent protein complexes from mixtures by on-line capillary isoelectric focusing-mass spectrometry. Anal Chem. **72**, 5356-5360 (2000)
306. Fonslow, B.R., Kang, S.A., Gestaut, D.R., Graczyk, B., Davis, T.N., Sabatini, D.M., et al.: Native capillary isoelectric focusing for the separation of protein complex isoforms and subcomplexes. Anal Chem. **82**, 6643-6651 (2010)
307. Skinner, O.S., Do Vale, L.H., Catherman, A.D., Havugimana, P.C., de Sousa, M.V., Compton, P.D., et al.: Native GELFrEE: a new separation technique for biomolecular assemblies. Analytical chemistry. **87**, 3032-3038 (2015)
308. Zheng, J., Wei, C., Zhao, L., Liu, L., Leng, W., Li, W., et al.: Combining blue native polyacrylamide gel electrophoresis with liquid chromatography tandem mass spectrometry as an effective strategy for analyzing potential membrane protein complexes of Mycobacterium bovis bacillus Calmette-Guerin. BMC genomics. **12**, 40 (2011)
309. Han, X., Wang, Y., Aslanian, A., Fonslow, B., Graczyk, B., Davis, T.N., 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. **13**, 6078-6086 (2014)
310. Rambaldi, D.C., Reschiglian, P., Zattoni, A.: Flow field-flow fractionation: recent trends in protein analysis. Anal Bioanal Chem. **399**, 1439-1447 (2011)
311. Reschiglian, P., Zattoni, A., Roda, B., Roda, A., Parisi, D., Moon, M.H., et al.: Hollow-fiber flow field-flow fractionation: a gentle separation method for mass spectrometry of native proteins. Annali di chimica. **96**, 253-257 (2006)
312. Issaq, H.J., Janini, G.M., Atamna, I.Z., Muschik, G.M.: Separations by High-Performance Liquid-Chromatography and Capillary Zone Electrophoresis - a Comparative-Study. Journal of Liquid Chromatography. **14**, 817-845 (1991)
313. Maxwell, E.J., Chen, D.D.: Twenty years of interface development for capillary electrophoresis-electrospray ionization-mass spectrometry. Analytica chimica acta. **627**, 25-33 (2008)
314. Breuker, K., McLafferty, F.W.: Stepwise evolution of protein native structure with electrospray into the gas phase, 10(-12) to 10(2) s. Proc Natl Acad Sci U S A. **105**, 18145-18152 (2008)
315. Hewitt, D., Marklund, E., Scott, D.J., Robinson, C.V., Borysik, A.J.: A hydrodynamic comparison of solution and gas phase proteins and their complexes. The journal of physical chemistry. B. **118**, 8489-8495 (2014)
316. Meyer, T., Gabelica, V., Grubmüler, H., Orozco, M.: Proteins in the gas phase. WIREs Comput Mol Sci. 408–425 (2013)
317. Ben-Nissan, G., Belov, M.E., Morgenstern, D., Levin, Y., Dym, O., Arkind, G., et al.: Triple-Stage Mass Spectrometry Unravels the Heterogeneity of an Endogenous Protein Complex. Analytical Chemistry. **89**, 4708-4715 (2017)
318. van de Waterbeemd, M., Fort, K.L., Boll, D., Reinhardt-Szyba, M., Routh, A., Makarov, A., et al.: High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. Nat Methods. (2017)
319. Hardy, S.J., Kurland, C.G., Voynow, P., Mora, G.: The ribosomal proteins of Escherichia coli. I. Purification of the 30S ribosomal proteins. Biochemistry. **8**, 2897-2905 (1969)
320. Cannon, J.R., Cammarata, M.B., Robotham, S.A., Cotham, V.C., Shaw, J.B., Fellers, R.T., et al.: Ultraviolet photodissociation for characterization of whole proteins on a chromatographic time scale. Analytical Chemistry. **86**, 2185-2192 (2014)

321. Kreimer, S., Belov, M.E., Danielson, W.F., Levitsky, L.I., Gorshkov, M.V., Karger, B.L., et al.: Advanced Precursor Ion Selection Algorithms for Increased Depth of Bottom-Up Proteomic Profiling. J Proteome Res. **15**, 3563-3573 (2016)
322. Kall, L., Canterbury, J.D., Weston, J., Noble, W.S., MacCoss, M.J.: Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods. **4**, 923-925 (2007)
323. Whitt, J.T., Moini, M.: Capillary electrophoresis to mass spectrometry interface using a porous junction. Anal Chem. **75**, 2188-2191 (2003)
324. Valaskovic, G.A., Kelleher, N.L., McLafferty, F.W.: Attomole protein characterization by capillary electrophoresis-mass spectrometry. Science. **273**, 1199-1202 (1996)
325. Holland, M.J., Holland, J.P., Thill, G.P., Jackson, K.A.: The primary structures of two yeast enolase genes. Homology between the 5' noncoding flanking regions of yeast enolase and glyceraldehyde-3-phosphate dehydrogenase genes. J Biol Chem. **256**, 1385-1395 (1981)
326. Chin, C.C., Brewer, J.M., Wold, F.: The amino acid sequence of yeast enolase. J Biol Chem. **256**, 1377-1384 (1981)
327. Poyner, R.R., Laughlin, L.T., Sowa, G.A., Reed, G.H.: Toward identification of acid/base catalysts in the active site of enolase: comparison of the properties of K345A, E168Q, and E211Q variants. Biochemistry. **35**, 1692-1699 (1996)
328. Brewer, J.M.: Specificity and Mechanism of Action of Metal-Ions in Yeast Enolase. FEBS Lett. **182**, 8-14 (1985)
329. Lebioda, L., Stec, B.: Crystal structure of enolase indicates that enolase and pyruvate kinase evolved from a common ancestor. Nature. **333**, 683-686 (1988)
330. Leskovac, V., Trivic, S., Pericin, D.: The three zinc-containing alcohol dehydrogenases from baker's yeast, Saccharomyces cerevisiae. FEMS Yeast Res. **2**, 481-494 (2002)
331. Jornvall, H.: The primary structure of yeast alcohol dehydrogenase. Eur J Biochem. **72**, 425-442 (1977)
332. Li, H., Wongkongkathep, P., Van Orden, S.L., Ogorzalek Loo, R.R., Loo, J.A.: Revealing ligand binding sites and quantifying subunit variants of noncovalent protein complexes in a single native top-down FTICR MS experiment. J Am Soc Mass Spectrom. **25**, 2060-2068 (2014)
333. Saito, R., Sato, T., Ikai, A., Tanaka, N.: Structure of bovine carbonic anhydrase II at 1.95 A resolution. Acta Crystallogr D Biol Crystallogr. **60**, 792-795 (2004)
334. Lauer, H.H., Mcmanigill, D.: Capillary Zone Electrophoresis of Proteins in Untreated Fused-Silica Tubing. Anal Chem. **58**, 166-170 (1986)
335. Chen, F.T.A.: Rapid Protein-Analysis by Capillary Electrophoresis. Journal of Chromatography. **559**, 445-453 (1991)
336. Schwartz, H., Pritchett, T. Beckman, (1994)
337. Beck, A., Sanglier-Cianferani, S., Van Dorsselaer, A.: Biosimilar, biobetter, and next generation antibody characterization by mass spectrometry. Analytical Chemistry. **84**, 4637-4646 (2012)
338. Harris, R.J., Kabakoff, B., Macchi, F.D., Shen, F.J., Kwong, M., Andya, J.D., et al.: Identification of multiple sources of charge heterogeneity in a recombinant antibody. J Chromatogr B Biomed Sci Appl. **752**, 233-245 (2001)
339. Xie, H., Chakraborty, A., Ahn, J., Yu, Y.Q., Dakshinamoorthy, D.P., Gilar, M., et al.: Rapid comparison of a candidate biosimilar to an innovator monoclonal antibody with advanced liquid chromatography and mass spectrometry technologies. mAbs. **2**, 379-394 (2010)
340. Ha, S., Ou, Y., Vlasak, J., Li, Y., Wang, S., Vo, K., et al.: Isolation and characterization of IgG1 with asymmetrical Fc glycosylation. Glycobiology. **21**, 1087-1096 (2011)
341. Sahin, E., Roberts, C.J.: Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanisms. Methods Mol Biol. **899**, 403-423 (2012)

342. Indrawati, L., Stroshine, R.L., Narsimhan, G.: Low-field NMR: A tool for studying protein aggregation. J Sci Food Agr. **87**, 2207-2216 (2007)
343. Brandt, J.P., Patapoff, T.W., Aragon, S.R.: Construction, MD simulation, and hydrodynamic validation of an all-atom model of a monoclonal IgG antibody. Biophys J. **99**, 905-913 (2010)
344. Wu, H., Kroe-Barrett, R., Singh, S., Robinson, A.S., Roberts, C.J.: Competing aggregation pathways for monoclonal antibodies. FEBS Lett. **588**, 936-941 (2014)
345. Yin, L.S., Chen, X.Y., Tiwari, A., Vicini, P., Hickling, T.P.: The Role of Aggregates of Therapeutic Protein Products in Immunogenicity: An Evaluation by Mathematical Modeling. J Immunol Res. (2015)
346. Xue, S., Barna, M.: Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat Rev Mol Cell Biol. **13**, 355-369 (2012)
347. Slavov, N., Semrau, S., Airoldi, E., Budnik, B., van Oudenaarden, A.: Differential Stoichiometry among Core Ribosomal Proteins. Cell reports. **13**, 865-873 (2015)
348. Gilbert, W.V.: Functional specialization of ribosomes? Trends Biochem Sci. **36**, 127-132 (2011)
349. Filipovska, A., Rackham, O.: Specialization from synthesis: how ribosome diversity can customize protein function. FEBS Lett. **587**, 1189-1197 (2013)
350. Byrgazov, K., Vesper, O., Moll, I.: Ribosome heterogeneity: another level of complexity in bacterial translation regulation. Current Opinion in Microbiology. **16**, 133-139 (2013)
351. Ilag, L.L., Videler, H., McKay, A.R., Sobott, F., Fucini, P., Nierhaus, K.H., et al.: Heptameric (L12)6/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria. Proc Natl Acad Sci U S A. **102**, 8192-8197 (2005)
352. Senko, M.W., Beu, S.C., McLaffertycor, F.W.: Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. J Am Soc Mass Spectrom. **6**, 229-233 (1995)
353. Schmitt, N.D., Agar, J.N.: Parsing disease-relevant protein modifications from epiphenomena: perspective on the structural basis of SOD1-mediated ALS. J Mass Spectrom. **52**, 480-491 (2017)
354. Gervais, D., King, D.: Capillary isoelectric focusing of a difficult-to-denature tetrameric enzyme using alkylurea-urea mixtures. Anal Biochem. **465**, 90-95 (2014)
355. Ngounou Wetie, A.G., Sokolowska, I., Woods, A.G., Roy, U., Loo, J.A., Darie, C.C.: Investigation of stable and transient protein-protein interactions: Past, present, and future. Proteomics. **13**, 538-557 (2013)
356. Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., et al.: Structure of the 30S ribosomal subunit. Nature. **407**, 327-339 (2000)
357. Stock, A., Clarke, S., Clarke, C., Stock, J.: N-terminal methylation of proteins: structure, function and specificity. FEBS Lett. **220**, 8-14 (1987)
358. Nesterchuk, M.V., Sergiev, P.V., Dontsova, O.A.: Posttranslational Modifications of Ribosomal Proteins in Escherichia coli. Acta Naturae. **3**, 22-33 (2011)
359. Wittmann-Liebold, B., Greuer, B.: The primary structure of protein S5 from the small subunit of the Escherichia coli ribosome. FEBS Lett. **95**, 91-98 (1978)
360. Yaguchi, M.: Primary structure of protein S18 from the small Escherichia coli ribosomal subunit. FEBS Lett. **59**, 217-220 (1975)
361. Terhorst, C., Moller, W., Laursen, R., Wittmann-Liebold, B.: The primary structure of an acidic protein from 50-S ribosomes of Escherichia coli which is involved in GTP hydrolysis dependent on elongation factors G and T. Eur J Biochem. **34**, 138-152 (1973)
362. Tanaka, S., Matsushita, Y., Yoshikawa, A., Isono, K.: Cloning and molecular characterization of the gene rimL which encodes an enzyme acetylating ribosomal protein L12 of Escherichia coli K12. Mol Gen Genet. **217**, 289-293 (1989)

363. Diaconu, M., Kothe, U., Schlunzen, F., Fischer, N., Harms, J.M., Tonevitsky, A.G., et al.: Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. Cell. **121**, 991-1004 (2005)