Title

Identification and Analysis of Proteins Using Matrix Assisted Laser Desorption Ionization and Electrospray Ionization Mass Spectrometry

Author

Quebbemann, Neil Robert

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Identification and Analysis of Proteins Using Matrix Assisted Laser Desorption Ionization and Electrospray Ionization Mass Spectrometry

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy In Chemistry

by

Neil Robert Quebbemann

2020

ABSTRACT OF THE DISSERTATION

Identification and Analysis of Proteins
Using Matrix Assisted Laser Desorption Ionization
and Electrospray Ionization Mass Spectrometry

by

Neil Robert Quebbemann
Doctor of Philosophy in Chemistry
University of California, Los Angeles, 2020
Professor Joseph Ambrose Loo, Chair

Sample complexity continues to hinder the effectiveness of Top-Down mass spectrometry, which aims to become a high-throughput platform for proteomics. One possible solution to this issue is the separation and measurement of protein mixtures using virtual 2D gel electrophoresis/mass spectrometry (virtual 2D gel/MS), where intact proteins are initially separated by isoelectric focusing on immobilized pH gradient (IPG) gels followed by mass analysis using matrix-assisted laser desorption/ionization (MALDI) MS. Here, we report on improvements made to the virtual 2D gel/MS platform. With increased automation, we have reduced the time required to acquire and visualize proteins separated on a 180 mm IPG gel from several days to under 1 hour. This automation includes the implementation of a high-speed MALDI time-of-flight mass spectrometer operating with specialized MS imaging software to acquire data. Analysis of the MS data was also automated through the development of a custom program written in MATLAB. Mass spectrometry signal intensity, signal-to-noise ratio, and sensitivity were all improved with

a novel MALDI matrix application method where gels are immersed in matrix solution overnight, improving matrix crystallization. We also demonstrate for the first time the use of a 15 Tesla Fourier transform-ion cyclotron resonance mass spectrometer equipped with a MALDI source to acquire virtual 2D gel/MS data offering both an increase in resolution and accuracy of mass measurement results. Using the improved virtual 2D gel/MS technique, we identify changes to the *E. coli* proteome caused by both cold shock and antibiotic induced stress.

The aggregation and accumulation of α-synuclein in the brain have been linked to numerous neurodegenerative disorders including Parkinson's disease. To prevent these synucleinopathies much effort has been made to understand the cause of this protein aggregation and to find ways to prevent it. It has been shown that various ligands affect α-synuclein's propensity towards aggregation with the small molecule compound, CLR01, a lysine molecular tweezer, decreasing aggregation and divalent heavy metals increasing aggregation. Here, we use electrospray ionization-MS and collision induced unfolding (CIU) coupled with ion mobility spectrometry to probe the effects that CLR01, Mn(II), Co(II), and Cu(II) have on the structural stability of α-synuclein in the gas phase. Our results indicate that the binding of CLR01, Mn(II), 1x Cu(II), and 3x Cu(II) all have a stabilizing effect on the structure of the protein while Co(II) destabilizes the protein. The work presented in this thesis demonstrate new mass spectrometry-based experimental platforms to qualitatively and quantitatively profile complex protein mixtures rapidly and accurately, and to probe the structural stability of protein/ligand complexes that are complementary to other biophysical methods.

The dissertation of Neil Robert Quebbemann is approved.

Anne M. Andrews

Feng Guo

Joseph Ambrose Loo, Committee Chair

University of California, Los Angeles

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First, I would like to thank my advisors, Dr. Joseph A. Loo and Dr. Rachel R. Ogorzalek Loo for all the guidance, support, and encouragement they have provided during my time at UCLA, I would never have gotten to this point without it. Furthermore, I am especially grateful to Joe for allowing me to get a head start on my teaching career while I finished my degree. I would also like to thank Carly Daniels, Won Jung, Carter Lantz, Huilin Li, Hong Nguyen, Sean Shen, and the other Loo Lab members for providing a fun and positive working environment in the lab.

I would like to acknowledge my committee members: Dr. Anne M. Andrews, Dr. Feng Guo, and Dr. Emil Reisler for taking time out of their busy schedules to attend committee meetings and provide valuable feedback on my progress.

Chapter 1 of this dissertation is reprinted with permission from Lohnes, K.; Quebbemann, N. R.; Liu, K.; Kobzeff, F.; Loo, J. A.; Ogorzalek Loo, R. R. Combining High-Throughput MALDI-TOF Mass Spectrometry and Isoelectric Focusing Gel Electrophoresis for Virtual 2D Gel-Based Proteomics. *Methods* **2016**, *104*, 163–169. Copyright 2016 Elsevier. I would like to acknowledge my co-authors Karen Lohnes, Kate Liu, Fred Kobzeff, Joseph A. Loo, and Rachel R. Ogorzalek Loo for all their help.

The work in Chapter 2 was performed with assistance from Kate Liu who helped prepare and analyze samples and Huilin Li who provided invaluable instruction on the operation of the FT-ICR mass spectrometer. The work in Chapter 3 was performed with assistance from Hong Nguyen who guided the cultivation of bacteria and collecting in-gel digest data.

The appendix of this dissertation is reprinted with permission from Srzentić, K.; Fornelli, L.; Tsybin, Y. O.; Loo, J. A.; Seckler, H.; Agar, J. N.; Anderson, L. C.; Bai, D. L.; Beck, A.; Brodbelt, J. S.; van der Burgt, Y. E. M.; Chamot-Rooke, J.; Chatterjee, S.; Chen, Y.; Clarke, D. J.; Danis, P. O.; Diedrich, J. K.; D'Ippolito, R. A.; Dupré, M.; Gasilova, N.; Ge, Y.; Goo, Y. A.; Goodlett, D. R.; Greer, S.; Haselmann, K. F.; He, L.; Hendrickson, C. L.; Hinkle, J. D.; Holt, M. V.; Hughes, S.; Hunt, D. F.; Kelleher, N. L.; Kozhinov, A. N.; Lin, Z.; Malosse, C.; Marshall, A. G.; Menin, L.; Millikin, R. J.; Nagornov, K. O.; Nicolardi, S.; Paša-Tolić, L.; Pengelley, S.; Quebbemann, N. R.; Resemann, A.; Sandoval, W.; Sarin, R.; Schmitt, N. D.; Shabanowitz, J.; Shaw, J. B.; Shortreed, M. R.; Smith, L. M.; Sobott, F.; Suckau, D.; Toby, T.; Weisbrod, C. R.; Wildburger, N. C.; Yates, J. R.; Yoon, S. H.; Young, N. L.; Zhou, M. Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (9), 1783–1802. https://doi.org/10.1021/jasms.0c00036. Copyright 2020 American Society for Mass Spectrometry.

Finally, I would like to thank my family and friends for all the love and support they have provided during this long journey.

VITA

EDUCATION

University of California, Los Angeles, CA
Masters of Science in Chemistry — **December 2015**

University of California, Riverside, CA
Bachelors of Science in Chemistry, with High Honors — **June 2013**

AWARDS

American Society for Mass Spectrometry Travel Stipend Award — 2016
UCLA Graduate Dean's Scholar Award — 2013 – 2015
UCR Chancellor's Research Fellowship — 2012 – 2013

TEACHING EXPERIENCE

Antelope Valley College, Lancaster, CA
Assistant Professor — **2020 -**

Los Angeles Pierce College, Woodland Hills, CA
Adjunct Professor — **2018 -**

Touro College Los Angeles, West Hollywood, CA
Adjunct Professor — **2019 - 2020**

University of California, Los Angeles, CA
Teaching Assistant — **2013 - 2020**

PUBLICATIONS

1. Tao, Y.; **Quebbemann, N. R.**; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820. https://doi.org/10.1021/ac3013434.
2. Lohnes, K.; **Quebbemann, N. R.**; Liu, K.; Kobzeff, F.; Loo, J. A.; Ogorzalek Loo, R. R. Combining High-Throughput MALDI-TOF Mass Spectrometry and Isoelectric Focusing Gel Electrophoresis for Virtual 2D Gel-Based Proteomics. *Methods* **2016**, *104*, 163–169. https://doi.org/https://doi.org/10.1016/j.ymeth.2016.01.013.

3. Srzentić, K.; Fornelli, L.; Tsybin, Y. O.; Loo, J. A.; Seckler, H.; Agar, J. N.; Anderson, L. C.; Bai, D. L.; Beck, A.; Brodbelt, J. S.; van der Burgt, Y. E. M.; Chamot-Rooke, J.; Chatterjee, S.; Chen, Y.; Clarke, D. J.; Danis, P. O.; Diedrich, J. K.; D'Ippolito, R. A.; Dupré, M.; Gasilova, N.; Ge, Y.; Goo, Y. A.; Goodlett, D. R.; Greer, S.; Haselmann, K. F.; He, L.; Hendrickson, C. L.; Hinkle, J. D.; Holt, M. V.; Hughes, S.; Hunt, D. F.; Kelleher, N. L.; Kozhinov, A. N.; Lin, Z.; Malosse, C.; Marshall, A. G.; Menin, L.; Millikin, R. J.; Nagornov, K. O.; Nicolardi, S.; Paša-Tolić, L.; Pengelley, S.; **Quebbemann, N. R.**; Resemann, A.; Sandoval, W.; Sarin, R.; Schmitt, N. D.; Shabanowitz, J.; Shaw, J. B.; Shortreed, M. R.; Smith, L. M.; Sobott, F.; Suckau, D.; Toby, T.; Weisbrod, C. R.; Wildburger, N. C.; Yates, J. R.; Yoon, S. H.; Young, N. L.; Zhou, M. Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (9), 1783–1802. https://doi.org/10.1021/jasms.0c00036.

PRESENTATIONS

1. **Quebbemann N.R.**; Liu K.; Loo R.O.; Loo J.A. Optimizing Virtual 2D gel/MS through the Analysis of *E. coli* and *M. mazei* Cell Lysate. Presented at the 63rd Conference on Mass Spectrometry and Allied Topics, St. Louis, MO, June 2, 2015.

2. **Quebbemann N.R.**; Loo R.R.O.; Loo J.A. Optimizing Virtual 2D-Gel/MS for Protein Characterization. Presented at Lake Arrowhead Conference on Ion Chemistry and Mass Spectrometry, Lake Arrowhead, CA, January 15, 2016.

3. **Quebbemann N.R.**; Loo J.A.; Loo R.R.O. Differential Protein Profiling using Virtual 2D gel-Mass Spectrometry. Presented at 64th Conference on Mass Spectrometry and Allied Topics, San Antonio, TX, June 8, 2016.

4. **Quebbemann N.R.**; Loo J.A.; Loo R.R.O. Differential Protein Profiling using Virtual 2D gel-Mass Spectrometry. Presented at Asilomar Conference on Mass Spectrometry and Mass Mobility, Asilomar, CA, October 15, 2016.

5. **Quebbemann N.R.**; Loo J.A.; Loo R.R.O. Differential Protein Profiling using Virtual 2D gel-Mass Spectrometry. Presented at 65th Conference on Mass Spectrometry and Allied Topics, Indianapolis, IN, June 8, 2017.

6. **Quebbemann N.R.**; Nshanian M.; Loo J.A. Analysis of Protein-Ligand Complexes with Collision Induced Unfolding. Presented at 66th Conference on Mass Spectrometry and Allied Topics, San Diego, CA, June 5, 2018.

7. **Quebbemann N.R.**; Loo J.A. Analysis of Specific Metal Binding to Alpha-Synuclein with Collision Induced Unfolding. Presented at 67th Conference on Mass Spectrometry and Allied Topics, Atlanta, GA, June 4, 2019.

CHAPTER 1

Combining high-throughput MALDI-TOF mass spectrometry and isoelectric focusing gel electrophoresis for virtual 2D gel-based proteomics

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Combining high-throughput MALDI-TOF mass spectrometry and isoelectric focusing gel electrophoresis for virtual 2D gel-based proteomics

Karen Lohnes [a], Neil R. Quebbemann [b], Kate Liu [b], Fred Kobzeff [b], Joseph A. Loo [a,b,c,*], Rachel R. Ogorzalek Loo [a,c,*]

[a] Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA
[b] Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA
[c] DOE/UCLA Institute of Genomics and Proteomics and UCLA Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA

ARTICLE INFO

Article history:
Received 30 October 2015
Received in revised form 13 January 2016
Accepted 25 January 2016
Available online 28 January 2016

Keywords:
Mass spectrometry
Matrix-assisted laser desorption/ionization (MALDI)
Gel electrophoresis
Virtual 2D gel
Proteomics
Isoelectric focusing (IEF)

ABSTRACT

The virtual two-dimensional gel electrophoresis/mass spectrometry (virtual 2D gel/MS) technology combines the premier, high-resolution capabilities of 2D gel electrophoresis with the sensitivity and high mass accuracy of mass spectrometry (MS). Intact proteins separated by isoelectric focusing (IEF) gel electrophoresis are imaged from immobilized pH gradient (IPG) polyacrylamide gels (the first dimension of classic 2D-PAGE) by matrix-assisted laser desorption/ionization (MALDI) MS. Obtaining accurate intact masses from sub-picomole-level proteins embedded in 2D-PAGE gels or in IPG strips is desirable to elucidate how the protein of one spot identified as protein 'A' on a 2D gel differs from the protein of another spot identified as the same protein, whenever tryptic peptide maps fail to resolve the issue. This task, however, has been extremely challenging. Virtual 2D gel/MS provides access to these intact masses.

Modifications to our matrix deposition procedure improve the reliability with which IPG gels can be prepared; the new procedure is described. Development of this MALDI MS imaging (MSI) method for high-throughput MS with integrated 'top-down' MS to elucidate protein isoforms from complex biological samples is described and it is demonstrated that a 4-cm IPG gel segment can now be imaged in approximately 5 min. Gel-wide chemical and enzymatic methods with further interrogation by MALDI MS/MS provide identifications, sequence-related information, and post-translational/transcriptional modification information. The MSI-based virtual 2D gel/MS platform may potentially link the benefits of 'top-down' and 'bottom-up' proteomics.

1. Introduction

1.1. Top-down and bottom-up proteomics; top-down and bottom-up mass spectrometry

High-sensitivity, high-throughput protein analysis (*i.e.*, proteomics) is essential in the post-genome world to complement advances in DNA and RNA sequencing and genetic engineering. Great strides have been made in identifying proteins from complex mixtures, supported by two decades of developments in mass spectrometry (MS) and ultra-small-scale high performance liquid chromatography (HPLC). These achievements in protein identification rely on the free availability of stockpiled nucleic acid sequences from a multitude of organisms. Beyond DNA, RNA, and protein sequence lies an invaluable fourth dimension of information regarding the architecture of life: protein processing and modification. That fourth dimension is important because modifications to thousands of component proteins in a cell may change with the cell cycle, environmental conditions, developmental stage, and metabolic state. Markers reflecting health, environmental exposure, and disease can be encoded in this fourth dimension of information. Modifications can alter protein localization, activity, lifetime, and how that protein interacts with other molecules. Modifications are not necessarily homogeneously distributed in a protein population, making the discovery and quantification of each modification an almost insurmountable task

* Corresponding authors at: UCLA Molecular Biology Institute, Paul Boyer Hall, Los Angeles, CA 90095, USA.
E-mail addresses: JLoo@chem.ucla.edu (J.A. Loo), RLoo@mednet.ucla.edu (R.R. Ogorzalek Loo).

http://dx.doi.org/10.1016/j.ymeth.2016.01.013
1046-2023/

and one that depends on many factors including protein concentration and dynamic range of the sample under investigation [1].

For complex mixtures, we classify experimental approaches that sort (separate) intact proteins prior to analysis as 'top-down' proteomics and approaches that cleave proteins within a complex mixture, prior to analysis, as 'bottom-up' proteomics. We shall apply 'top-down' MS and 'bottom-up' MS to describe the nature of the analyte introduced to the mass spectrometer. Hence, the reversed phase HPLC delivery of intact proteins from a ribosome mixture to a mass spectrometer for mass analysis, precursor ion selection, electron capture dissociation (ECD), and analysis of the product ions (MS/MS) would constitute top-down proteomics *and* top-down MS, but collecting HPLC fractions from the ribosomal preparation and subjecting each fraction (isolated protein) to trypsin digestion followed by liquid chromatography–tandem MS (LC–MS/MS) of the peptides would, in our usage, constitute top-down proteomics and bottom-up MS.

Bottom-up proteomics integrates the cleaved products detected from all protein forms (proteoforms [2]) produced by a gene into a single peptide map of the full-length gene product (potentially revealing, but not eliminating, any splice variants that might be present) to tabulate and quantify expressed proteins efficiently. Bottom-up proteomics avoids the difficulties (*e.g.*, sample losses, extreme diversity in solubility and chromatographic properties) inherent in handling intact proteins by separating/analyzing smaller, easier to handle *peptides*, instead. In exchange for sensitivity and speed, however, bottom-up proteomic methods sever links to the protein fragments' heritage, erasing much of the means to characterize splicing variants, and co- and post-translational modifications. Thus, bottom-up proteomics tends to assume that prior to the analysis, proteins existed as predicted by the genome sequence, *i.e.*, full-length and unmodified, or with a small number of predictable modifications. Bottom-up proteomic approaches to *globally* analyze post-translational modifications usually seek evidence for a small number of modifications from the proteome, sometimes pre-enriching in advance for select modifications (*e.g.*, with antibodies). In these *global* analyses it is rarely possible to establish that *all* instances of a particular modification have been detected in a given protein, to determine how many proteoforms are present, *or even to establish whether all abundant proteoforms have been detected.* Error-tolerant or unrestricted searches have some ability to reveal unanticipated modifications, but are challenged by the extraordinarily large search space of all proteins/all modifications in concert with the finite information content of tandem mass spectra. The vagaries of undersampling (too many peptides to measure for the instrument duty cycle) *limit even the ability to establish that certain modifications are not abundantly present (failure to detect a peptide does not necessarily allow us to conclude that it is absent).* Nevertheless, valuable lists of (some) modified proteins with (some) modified sites are generated. Quantifying the levels of modified *versus* unmodified proteins is more challenging.

Top-down proteomics, particularly when pursued with polyacrylamide gel electrophoresis (PAGE)-based platforms, can complement the bottom-up proteomic approaches for addressing the problems described above. By revealing the minimum number of proteoforms present in a sample, and separating many distinct forms for individual analyses, an analyst is better able to differentiate splice variants that are present from those that are not present and to quickly estimate the number of distinct species present. Error-tolerant searches can be more sensitive than in bottom-up proteomics, because they must query only a small number of candidate proteins. Top-down proteomic methods can expose large deviations from predictions, *e.g.*, migration in the SDS-PAGE size dimension inconsistent with a given molecular weight, focusing far from the predicted isoelectric point (*pI*), or non-binding of appropriate antibodies. Deviant products can be flagged for additional analyses likely to yield new discoveries, while products found to be consistent with predictions (*e.g.*, molecular masses matching theoretical) can be considered characterized. The ability to quantitatively track one or more proteoforms (by staining intensity or autoradiography) from multiple samples and to excise these proteoforms for chemical analysis (*e.g.*, protein identification or carbohydrate composition) is a strength of top-down proteomics.

Top-down proteomics/top-down mass spectrometry has been very successfully pursued by combining Gel Elution Liquid-based Fractionation Entrapment Electrophoresis (GELFrEE) fractionation, protein precipitation, and reversed phase HPLC with electrospray ionization (ESI) and MS/MS [3]. GELFrEE fractionates proteins on a preparative scale by continuous electrophoresis in polyacrylamide tube gels. The method is compatible with hydrophilic and hydrophobic proteins, although ultimately the top-down platform's protein compatibility is limited by its reversed phase HPLC dimension downstream.

Our interest lies in linking accurate molecular masses to denaturing isoelectric focusing (IEF) and to two-dimensional (2D)-PAGE analyses (of which IEF constitutes the first dimension). With the advent of immobilized pH gradient (IPG) gels, IEF offers the highest resolution available in the proteomics toolbox, capable of separating protein isoforms whose isoelectric points differ by only 0.001 pH units [4]. IPGs, which are commercially available, stable, and highly reproducible, have been produced to span only 0.1 pH unit and can successfully separate proteins that differ by only one amino acid [5].

1.2. The virtual 2D gel

In the 'virtual 2D gel' MS platform [6–8], complex protein mixtures are separated by high resolution IEF using IPG gels, incubated with MALDI matrix, and dried. Resolved proteins are laser-desorbed and ionized directly from the 'xerogel' for mass measurement. Proteins over 100 kDa have been measured by this method, as well as small polypeptides (<10 kDa), which are often lost in the second dimension of the SDS gel [6,9]. The term 'virtual 2D gel/MS' [8] is appropriate because denaturing IEF also comprises the first dimension of traditional 2D gels, with protein size (by SDS-PAGE) as the second dimension. Indeed, the IEF separations for virtual 2D gel analysis have been performed under *identical* conditions to those employed for 2D PAGE (*i.e.*, employing urea, thiourea, CHAPS or Triton X-100 detergent, carrier ampholytes, *etc.*) to ensure that the information provided relates directly to classical 2D gels. MALDI MS replaces SDS-PAGE as the second dimension, but the mass accuracy for protein intact masses provided by MS is superior to SDS gels. The *common IEF axis* links classical and virtual 2D gels such that an intact mass and other MS-derived information can be *permanently* associated with particular 2D gel spots (Fig. 1) [7] and with other analyses linked to that spot (*e.g.*, glycan analysis, radioactivity to elucidate synthesis or degradation rate, antibody binding, lectin binding, protein identity from in-gel digestion, *etc.*). We consider the most important capability of virtual 2D gels to be revealing the intact molecular masses of proteins focusing at each isoelectric point for comparison to classical 2D gel images and the corresponding protein identities of spots (generally obtained by trypsin digestion and LC–MS/MS analyses). Combining protein identity with intact mass readily highlights modified, processed, or otherwise unexpected protein forms. It enables researchers to triage which spots of interest contain proteins in expected forms (no additional characterization effort needed) and which differ (with the mass, perhaps suggesting possibilities for those differences; *e.g.*, +42 Da suggests an acetylation or trimethylation). Protein-level quantitation, provided by staining or radiography *via* classical 2D gels, provides a useful complement

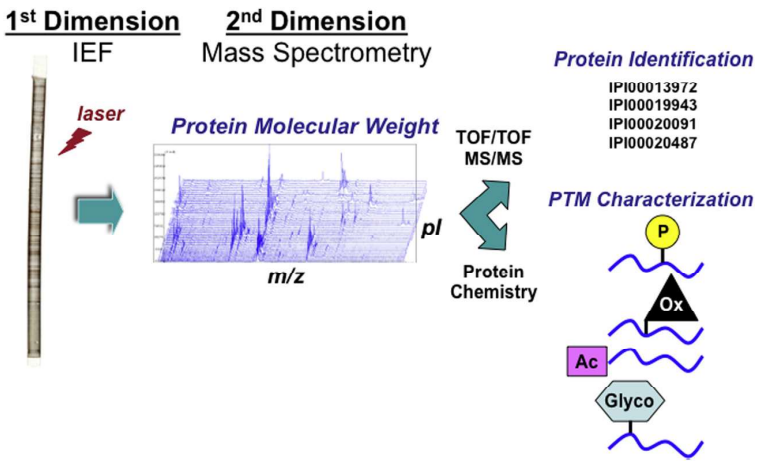


Fig. 1. Virtual 2D gel/MS scheme. Protein mixtures are separated in the first dimension by isoelectric focusing (IEF) electrophoresis on an immobilized pH gradient (IPG) gel. Matrix-deposited, dried gels are analyzed by MALDI MS to provide intact masses for proteins focused at each isoelectric point. Protein identifications can be obtained from existing information for the corresponding positions on classical 2D gels (IEF/SDS-PAGE), can be proposed from accurate MW and *pI* information, or can be obtained by in-gel digestion and MALDI-TOF/TOF MS/MS analysis. Applying alternative protein chemistries with MS/MS analysis can provide additional modification information.

to any peptide-level quantitation available from in-gel digests or bottom-up proteomics.

The early work that developed the concept of the virtual 2D gel/MS platform was performed using MALDI-time-of-flight (MALDI-TOF) instruments utilizing 20 Hz-UV lasers for desorption/ionization. Today's availability of high-speed MALDI-TOF MSI instruments equipped with kHz-repetition rate lasers along with improvements in preparing matrix-embedded dry gels (xerogels) pave the way for the virtual 2D gel/MS platform to become easier to implement. Methods for establishing this platform are discussed in this article.

2. Methods

2.1. Protein extract preparation

Sample preparation for IEF [10–18] differs from that for SDS-PAGE, primarily by avoiding ionic salts and anionic or cationic detergents. Here, pelleted cells were stored at −80 °C until lysis and extract preparation. At that time, cells were thawed and rapidly resuspended in 300 μL of 0.3% SDS (*w*/*v*), 10 mM Tris, pH 7.5, and 0.2 M dithiothreitol (DTT). Solutions were incubated in a boiling water bath for 2 min and subsequently cooled on ice. To the cooled solutions were added 30 μL of DNAse/RNAse cocktail and the mixtures were incubated on ice for 10 min. Following incubation, 1200 μL of lysis buffer, 138 mg urea, and 50 mg thiourea were added to the solutions and the mixtures were vortexed extensively. The lysis buffer contained 7 M urea, 2 M thiourea, 50 mM DTT, 1.54% Pharmalyte™ 3–10 carrier ampholytes (*v*/*v*; GE Healthcare, Piscataway, NJ), and 2.55% CHAPS (*w*/*v*). Protein extract concentrations, assessed with the NI™ (Non-Interfering™) Protein Assay (G-Biosciences), were typically 1–5 mg/mL. Extracts were aliquoted for storage at −80 °C. Microbial samples were kindly provided by Michael McInerney (University of Oklahoma) and Robert Gunsalus (UCLA).

2.2. IEF with IPG gels

Isoelectric focusing employed precast IPG strips (ReadyStrip™, Bio-Rad, Hercules, CA or Immobiline™ DryStrip, GE Healthcare) spanning pH 4–7 or 3–10 in 18 cm. Samples (20–200 μg) were loaded by passive in-gel rehydration or by cup-loading [12,19–22]. For the former method, the desired load of extract was mixed with sufficient rehydration buffer (6 M urea, 1.7 M thiourea, 12 mM DTT, 1% Pharmalyte™ 3–10 (*v*/*v*), 0.5% CHAPS (*w*/*v*), 5% glycerol, and 10% isopropanol) to yield a 350 μL volume in which to swell each dehydrated gel overnight. For cup-loading, each IPG strip was incubated overnight in 350 μL of rehydration buffer; extract was applied the next day during focusing.

Gels were focused for 65–80 kV-hs on a Multiphor II flatbed electrophoresis assembly [12,23–25] (GE Healthcare) equipped with an EPS 3501XL power supply (GE Healthcare). Following IEF, polyacrylamide gels were stored at −80 °C until processed for direct MS, whole-gel trypsin digestion, or SDS-PAGE second dimension electrophoresis.

2.3. Washing focused IPG gels and applying MALDI matrix

Focused IPG gels were removed from the freezer and allowed to rest for a few minutes at room temperature. Once thawed, the gels were removed from the storage container with tweezers and placed *gel side up* on damp paper towels. Excess mineral oil (remaining from the IEF) was drained off the gels by lifting an edge of the polyester Mylar backing with tweezers. Each gel surface was blotted once, gently, with a damp paper towel. Scissor cuts were made at the edges of each gel support/backing to provide unique identifiers. Gel strips were incubated for 15 min in a glass tray containing 200 mL of wash solution (49.9% H_2O/50% acetonitrile (ACN)/0.1% trifluoroacetic acid (TFA)), mixing slowly on a lab rocker. The large volume of liquid removes or reduces the amounts of several substances essential to high-resolution electrophoresis but detrimental to MALDI MS: detergents, urea, thiourea, ampholytes, and residual mineral oil [26]. Despite this washing, CHAPS detergent clusters often appeared in MALDI mass spectra below *m*/*z* 5000 [26]. After incubation, each gel was withdrawn from the wash solution with tweezers and placed *gel side up* atop a clean surface. Each gel surface was gently blotted with damp, water-moistened filter paper (Whatman grade 4) to remove mineral oil residue remaining on the surface. Often, fibers from the filter paper stuck to portions of the gel after blotting, but these did not hinder analysis.

Sinapinic acid solutions were prepared at least 2 days in advance to ensure saturation. Liquid solutions were withdrawn from any undissolved matrix immediately before they were

dispensed for gel incubation. Two saturated sinapinic acid formulations containing 1.5% sorbitol (*w*/*v*) were prepared in quantities sufficient to provide 15 mL of each per IPG strip. The solvent for the first solution was 25% ACN/74.9% H_2O/0.1% TFA (*v*/*v*/*v*, henceforth referred to as 1:3 matrix solution), while the second solution solvent was 33.3% ACN/66.6% H_2O/0.1% TFA (henceforth referred to as 1:2 matrix solution).

Each filter paper-blotted gel was placed, gel side up, in an individual plastic tray (220 mm × 15 mm × 5 mm, length × width × depth) atop a laboratory rotating platform. About 13–15 mL of 1:3 matrix was dispensed to each tray and gel incubation proceeded for 10 min, mixing at low speed. With tweezers, gels were removed from the matrix solutions and transferred to empty trays, covered, and left overnight (12 h or greater) at room temperature.

The next day, the dry, matrix-treated gels were placed in clean trays atop the rotating platform and ~15 mL of the 1:2 matrix solution was dispensed. Gels incubated in the solutions for 10 min, rotating at low speed, after which the gels were removed from the solutions with tweezers and transferred to empty trays, covered to prevent dust from settling on them, and left overnight for the matrix to crystallize and the gels to dry thoroughly.

Whole gel trypsin digestions were performed employing protocols described previously [7].

2.4. Mounting dry, matrix-treated IEF strips and MALDI MS

Dry IEF strips were cut into two 9 cm-long pieces and secured to the MALDI sample stage with electrically conductive, double-sided transfer tape (Adhesives Research, Glen Rock, PA), cognizant of the IEF anode/cathode orientation (Fig. 2). Positions of the crystallized gel segments (not the Mylar support) and their cathode/anode orientation were recorded. Isoelectric points may be calculated at various distances along the IPG strip from manufacturer-provided plots of pH *vs.* distance along the gel [27,28]. Dried gels trap considerable moisture and air greatly extending the time for vacuum pump-down. Thus, gels newly attached to sample targets were placed in a vacuum desiccator and exposed to vacuum for at least 20 min prior to loading them in the mass spectrometer.

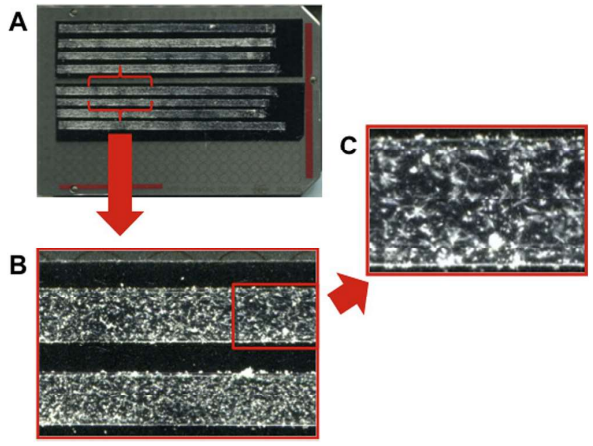


Fig. 2. Sinapinic acid-crystallized IPG gels on a MALDI target. (A) Crystallized gels adhered to the MALDI target with double sided adhesive tape. Gel segments are aligned to a row of target spots for easier navigation. The highlighted area is magnified in (B) to better display the surface-distribution of sinapinic acid crystals. Image (C) reveals that matrix deposits are present not only on the gel surface, but also embedded in the gel. The highest signal intensity and largest number of proteoforms are usually obtained from IPG gels with matrix deposited throughout the length and thickness of the dried gel.

Mass spectra were acquired with UV irradiation on Bruker Daltonics AutoFlex™ Speed and Ultraflex™ MALDI time-of-flight mass spectrometers with accompanying software (*e.g.*, FlexControl™, FlexImaging™ and FlexAnalysis™). Gels were imaged by acquiring data at raster sizes of 200–600 μm.

3. Results and discussion

3.1. Processing IEF gel strips for MALDI MS

In previous efforts [6,7,26,29,30], focused, washed IEF strips (Immobiline™ DryStrip) had been incubated in saturated matrix dissolved in 50% ACN/49.9% H_2O/0.1% TFA or 33.3% ACN/66.6% H_2O/0.1% TFA (*i.e.*, 1:1 or 1:2 matrix solution). Initial dried gels produced by this matrix-deposition method were stable under vacuum and laser irradiation for extended time periods. However, Immobiline™ DryStrip gels fabricated by the manufacturer several years later and for which matrix had been deposited by these methods were found to be unstable (as were ReadyStrip™ gels); those xerogels detached from the Mylar backing after several hours under vacuum. The source of the newly reduced stability was unclear, although we suspected that it arose from differences in the manufacturing processes. Because producing gels that remain attached to their backing (or that can be reliably detached without cracking) was essential, we explored additives for inclusion in matrix solutions to improve xerogel flexibility without degrading MS resolution and sensitivity.

Glycerol is a well-known stabilizer for polyacrylamide gels, minimizing their shrinkage and curling during drying, but high concentrations also degrade MALDI MS. We explored glycerol addition from 1% to 10% *v*/*v*, but found that amounts sufficient to maintain xerogel adherence to the Mylar backing rendered the matrix crystallization unreliable. Next, we considered saccharide additives, because they are known plasticizers and their presence in analyte solutions is often innocuous, or even helpful to protein analyses. Fucose additives have some history of benefitting MALDI MS analyses [31–34], but we selected, instead, the monosaccharide sorbitol to avoid potential covalent modifications. As a non-reducing sugar, sorbitol cannot form Schiff bases with amines. Also, sorbitol addition to a few percent does not hinder (and sometimes improves) MALDI MS analyses [30].

Experience preparing and analyzing matrix-deposited xerogels led us to consider what properties accounted for gels delivering high ion intensities at many spatial positions *versus* poorer performing gels. One property appeared to be the type and speed of matrix crystallization. Inspections of many crystallized gels yielding high and low ion intensities for ⩾20 kDa proteins suggested that clear, 'chopstick'-shaped crystals obtained by slow crystallization seemed to have a higher success rate, and that employing ACN:H_2O solvent ratios below 1:1 (*v*/*v*) provided some advantage, although matrix crystallization is notoriously hard to control. Gel incubation in 1:3 ACN:H_2O, was more likely to deliver the desired crystal form than 1:2, but crystals from the former were typically deposited sparsely (or not at all) across the gel. These observations led to the double matrix incubation described in the Methods section. Microcrystals from a first incubation in the 1:3 matrix solution (with lower sinapinic acid solubility) serve as seeds for the second crystallization from the 1:2 solution.

3.2. pI resolution and sensitivity

An optical image of a sinapinic acid-treated gel segment spanning ~1 pH unit, a spectral image generated by averaging individual mass spectra acquired across that pH span, and spectra acquired at discrete isoelectric points are illustrated in Fig. 3 for a cell lysate from the archaeon *Methanosarcina mazei*. The 26 most

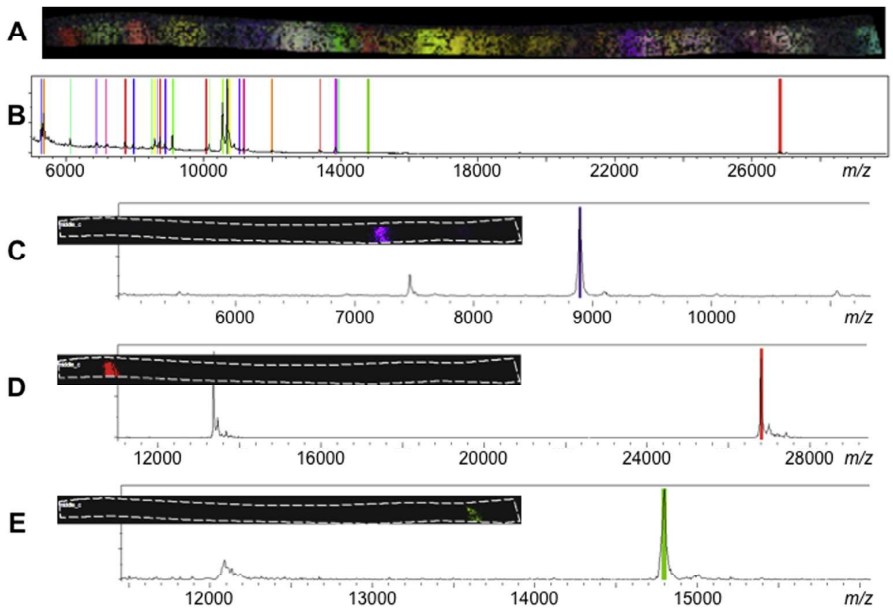


Fig. 3. (A) Image of an IPG-IEF gel section that was analyzed by MALDI MS. The gel was loaded with a whole cell lysate of *Methanosarcina mazei*. Colored highlights reflect regions where specific ion signals were detected. (B) Integrated mass spectrum obtained by summing all spectra acquired across the complete gel segment. The 26 most intense ion signals are indicated by lines, the colors of which correspond to the regions highlighted in (A). (C)-(E) MALDI mass spectra from selected regions of the gel segment.

intense ion signals over that segment are indicated by colored lines (Fig. 3**B**) with their corresponding *pIs* (gel positions) indicated by color highlights (Fig. 3**A**). From the colored regions in Fig. 3**A**, an indication of spatial resolution in the *pI* dimension can be gleaned, suggesting that the most abundant proteins were detected across regions up to 3.5 mm wide. From silver-stained IPG gels, we expect well-behaved proteins to focus to $\leqslant$1 mm; hence, we conclude that the washing and matrix treatments employed here reduced the resolution of the separation, albeit to a manageable level. Because the colored bands in Fig. 3 do not scale with intensity, the FWHM (full width at half maximum) of the MS-detected bands may be somewhat narrower.

Electrostatic spray ionization (ESTASI), an ambient ionization method [35] was previously applied to analyses from IPG gels, although that work is not directly comparable, in so far as the IEF was performed under native conditions; *i.e.*, without the urea, thiourea, surfactants, and ampholytes important to denaturing IEF and, correspondingly, to 2D-PAGE and theoretically predictable isoelectric points. For ESTASI, a spatial resolution of $\sim$3 mm was calculated, based on modeling the electric field for the ESTASI assembly; the resolution attainable experimentally was not reported.

Available high repetition rate lasers and automation packages designed for MSI simplify and speed up data acquisition compared to our previous efforts in imaging IEF gels [6,7,26,29,30]. Now, a 4 cm $\times$ 0.3 cm section of a gel can be imaged in 5 min or less. Fig. 4 illustrates the pH range of $\sim$5.4–6.1 from an IPG strip loaded with 200 μg of a whole cell lysate of *Syntrophus aciditrophicus*. Spectra were acquired with 1000 shots per position at a 500 μm step size (x and y). Five spectra were obtained at each isoelectric point (at discrete positions along the 0.3 cm strip width) and summed. The resulting 2D gel-like image is shown in Fig. 4, generated with the Bruker WARP-LC application. Data for the image was acquired in approximately 5 min.

It is clear from Fig. 4 that many protein masses in addition to the 5 highlighted ones are revealed by MS and that multiple species often appear at the same *pI*, as expected from classically stained 2D gels prepared with the same sample load of 200 μg.

Previous ESTASI analyses from IPG strips loaded with a mixture of 4 μg of myoglobin, 4 μg of cytochrome *c*, and 400 μg of *Escherichia coli* soluble proteins revealed multiply charged ions for myoglobin and cytochrome c at their migration positions, with charge state envelopes suitable for calculating their molecular mass. For the loaded *E. coli* proteins, however, it was not clear if molecular weights could be determined from the 3 spectra presented, but ion signals were observed. This comparison suggests that, at present, MALDI interrogation of IPG-IEF gels is more sensitive than ESTASI.

3.3. Protein identifications from virtual 2D gel/MS

Trypsin digestions of IEF-separated proteins (reduced and alkylated in gel after focusing) can be performed gel-wide by depositing enzyme solution onto the gel surface [7]. Diffusion of the proteolytic products is limited by scoring the wet gel with a knife or razor blade prior to depositing the enzyme. Each gel segment remains attached to the polyester strip, but diffusion is limited to the section length. Preliminary experiments demonstrated the feasibility of obtaining in-gel digestion products in this manner [7]. Here, we applied the approach to a richer region of an IEF gel, to assess whether protein identification remained feasible from mixtures with higher complexity.

Fig. 5 illustrates a classical, stained 2D gel from *Syntrophus aciditrophicus* with the pH 5.30 region boxed. Analysis of the corresponding region from a trypsin-treated, matrix-embedded IPG strip yielded the complicated peptide map (multiplex mass map) shown on the right side of Fig. 5. The proteins identified from this map were SYN0546, RSYN1909, RSYN0983, and RSYN2636 with sequence coverages of 34%, 22%, 18%, and 17% respectively. Observing these proteins together at this position on the digested IPG strip recapitulates their vertical position (*pI* 5.30) on the fluorescently stained 2D gel, and allows observed peptides to be correlated to proteoform intact masses, 2D gel staining intensities, and any additional information derived from excised spots.

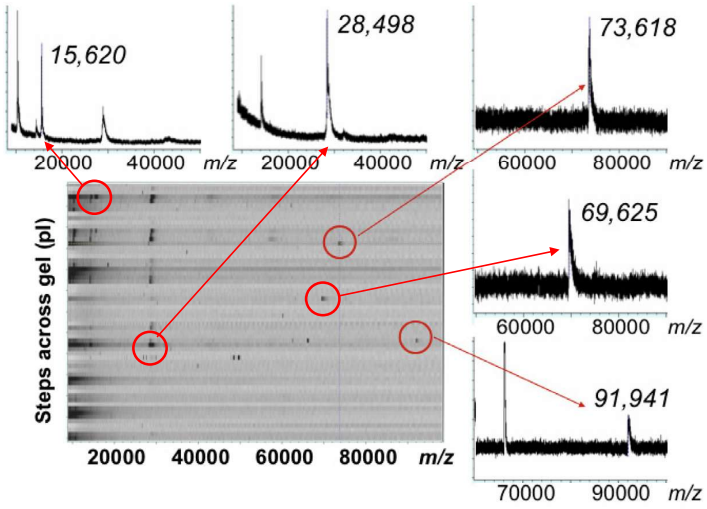


Fig. 4. From a 40-mm section of a 17-cm IPG gel strip (*pI* 4–7) of a *Syntrophus aciditrophicus* bacterial lysate, spectra were acquired (1000 shots per position) at step sizes (x and y) of 500 μm. Five spectra were obtained per isoelectric point and summed. The resulting 2D gel-like image was acquired in approximately 5 min.

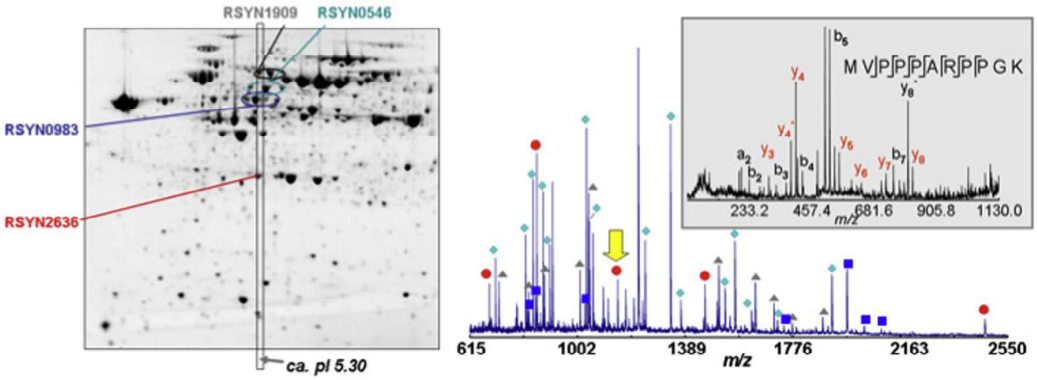


Fig. 5. (left) Sypro-Ruby-stained 2D-gel of a lysate of *Syntrophus aciditrophicus*. The region of pH 5.30 lies within the box indicated. (right) Multiplex virtual 2D gel/MS peptide map obtained from an in-gel tryptic digest of proteins at *pI* 5.30 on the IPG strip. Calibration with internal standards achieved rms mass accuracies of <25 ppm. Proteins identified include RSYN0546 (♦, 34% coverage), RSYN1909 (▲, 22% coverage), RSYN0983 (■, 18% coverage), and RSYN2636 (●, 17% coverage). (inset) MALDI MS/MS of an 1146.6 Da tryptic peptide for protein RSYN2636, the β-subunit of an electron transfer flavoprotein, obtained directly from the IPG gel.

At present, the sensitivity for analyzing peptides from these whole-gel digests trails that from in-gel digestion of excised bands by about two orders of magnitude. Nevertheless, detected peptides can be subjected to further analysis by MS/MS, as shown in the Fig. 5 inset, for an 1146.6 Da peptide from RSYN2636. Correlating the intact masses measured for selected isoelectric points to protein identities, peptide maps, and MS/MS spectra has been fruitful in revealing protein heterogeneity, as in the case of RSYN2636, which migrates as multiple 2D gel spots and appears at sizes ranging from 30 to 70 kDa. The many variations, potentially reflecting gene fusion events, are nearly invisible to analyses employing *only* bottom-up proteomics. However, linking intact protein separations to intact and proteolytic peptide mass analyses can not only begin to characterize these proteoforms, but should also provide means to monitor changes in their relative abundance with changes in cultivation conditions.

Another example of useful information accessible to virtual 2D gel electrophoresis is its elucidation of protein heterogeneity as illustrated in Fig. 6. This mass spectrum was obtained at *pI* 4.66 from an IEF separation of human saliva in which 30 μg of protein had been loaded onto the gel. The spectrum reveals several distinct proteins, including a quite heterogeneous 17 kDa species. Based on corresponding analyses from 2D gel spots and comparisons to previous 2D gel studies, the protein is identified as prolactin-inducible protein (PIP). The spacing between the 17 kDa peaks is ~144 Da, indicating the presence of repeated deoxyhexose saccharides, consistent with the rich fucosylation of this protein [36–38].

4. Conclusions and future directions

Gel-free, bottom-up proteomics is exquisitely powerful, but the information that it provides often stands alone, because most biological studies examine proteins rather than peptides. Certainly much can be inferred about proteins from their peptides, but direct protein analyses provide a means to check those inferences, and can alert us to unexpected mass discrepancies (that can lead to new discoveries). Developments in faster data acquisition and gel preparation presented here should make virtual 2D gel electrophoresis better able to complement previous, current, and

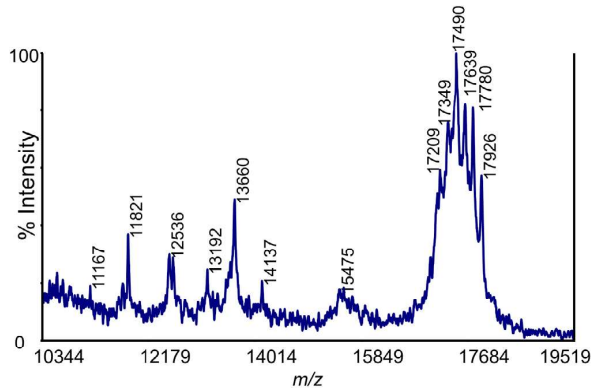


Fig. 6. Prolactin-inducible protein (17 kDa) from human saliva reveals heterogeneity due to glycosylation. Spectrum acquired from a position corresponding to *pI* 4.66 on an IPG gel (pH 4–7) loaded with 30 μg of protein.

future proteomic studies of the same or similar samples, by matching accurate protein masses to all of the other accumulated information on 2D-PAGE-separated proteins. This MALDI MS approach remains one of the most sensitive means to obtain intact masses for proteins embedded in polyacrylamide.

Acknowledgments

Support from the US National Institutes of Health (R01GM104610, S10RR025600, R01GM085402, U01DE016275) and the US Department of Energy Office of Science, Office of Biological and Environmental Research program (UCLA/DOE Institute for Genomics and Proteomics; Award Number DE-FC02-02ER63421) is acknowledged. We thank Paul Kowalski (Bruker Daltonics) for extensive help and advice.

References

[1] M. Mann, O.N. Jensen, Nat. Biotechnol. 21 (2003) 255–261.
[2] L.M. Smith, N.L. Kelleher, M. Linial, D. Goodlett, P. Langridge-Smith, Y.A. Goo, G. Safford, L. Bonilla, G. Kruppa, R. Zubarev, J. Rontree, J. Chamot-Rooke, J. Garavelli, A. Heck, J. Loo, D. Penque, M. Hornshaw, C. Hendrickson, L. Pasa-Tolic, C. Borchers, D. Chan, N. Young, J. Agar, C. Masselon, M. Gross, F. McLafferty, Y. Tsybin, Y. Ge, I. Sanders, J. Langridge, J. Whitelegge, A. Marshall, Nat. Methods 10 (2013) 186–187.
[3] A. Vellaichamy, J.C. Tran, A.D. Catherman, J.E. Lee, J.F. Kellie, S.M.M. Sweet, L. Zamdborg, P.M. Thomas, D.R. Ahlf, K.R. Durbin, G.A. Valaskovic, N.L. Kelleher, Anal. Chem. 82 (2010) 1234–1244.
[4] G.B. Smejkal, D.J. Bauer, in: D.S. Magdeldin (Ed.), Gel Electrophoresis – Principles and Basics, InTech, Croatia, 2012, pp. 157–170.
[5] G. Cossu, P.G. Righetti, J. Chromatogr. A 398 (1987) 211–216.
[6] R.R. Ogorzalek Loo, J.D. Cavalcoli, R.A. VanBogelen, C. Mitchell, J.A. Loo, B. Moldover, P.C. Andrews, Anal. Chem. 73 (2001) 4063–4070.
[7] R.R. Ogorzalek Loo, R. Hayes, Y. Yang, F. Hung, P. Ramachandran, N. Kim, R. Gunsalus, J.A. Loo, Int. J. Mass Spectrom. 240 (2005) 317–325.
[8] R.R. Ogorzalek Loo, C. Mitchell, T. Stevenson, J.A. Loo, P.C. Andrews, Techniques in Protein Chemistry VII, Academic Press, San Diego, CA, 1996. pp. 305–313.
[9] L.V. Bindschedler, L.J. McGuffin, T.A. Burgis, P.D. Spanu, R. Cramer, Methods 54 (2011) 432–441.
[10] B. Cañas, C. Piñeiro, E. Calvo, D. López-Ferrer, J.M. Gallardo, J. Chromatogr. A 1153 (2007) 235–258.
[11] Y. Chen, Z. Guo, X. Wang, C. Qiu, J. Chromatogr. A 1184 (2008) 191–219.
[12] A. Gorg, A. Klaus, C. Luck, F. Welland, W. Weiss, A Laboratory Manual, Technical University of Munich, Freising-Weihenstephan, Germany, 2007. pp. 166.
[13] L. Jiang, L. He, M. Fountoulakis, J. Chromatogr. A 1023 (2004) 317–320.
[14] T. Rabilloud, C. Lelong, J. Proteomics 74 (2011) 1829–1841.
[15] R.J. Simpson, Proteins and Proteomics: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2003.
[16] H. Wang, S. Hanash, Mass Spectrom. Rev. 24 (2005) 413–426.
[17] E.P. Wright, M.A. Partridge, M.P. Padula, V.J. Gauci, C.S. Malladi, J.R. Coorssen, Proteomics 14 (2014) 872–889.
[18] T. Rabilloud, M. Chevallet, S. Luche, C. Lelong, Proteomics 8 (2008) 3965–3973.
[19] A. Görg, O. Drews, C. Lück, F. Weiland, W. Weiss, Electrophoresis 30 (2009) S122–S132.
[20] A. Görg, W. Weiss, M.J. Dunn, Proteomics 4 (2004) 3665–3685.
[21] L.A. Kane, C.K. Yung, G. Agnetti, I. Neverova, J.E. Van Eyk, Proteomics 6 (2006) 5683–5687.
[22] C. Lamberti, E. Pessione, M.G. Giuffrida, R. Mazzoli, C. Barello, A. Conti, C. Giunta, Electrophoresis 28 (2007) 1633–1638.
[23] O. Carrette, P.R. Burkhard, J.-C. Sanchez, D.F. Hochstrasser, Nat. Protoc. 1 (2006) 812–823.
[24] G. Healthcare, Related Documents, User Manual, gelifesciences.com, 2006, pp. <http://www.gelifesciences.com/webapp/wcs/stores/servlet/productById/en/GELifeSciences-us/18101806>.
[25] P. Meleady, Meth. Mol. Biol. 784 (2011) 123–137.
[26] R.R. Ogorzalek Loo, L. Yam, J.A. Loo, V.N. Schumaker, Electrophoresis 25 (2004) 2384–2391.
[27] Immobiline DryStrip Visualization of *pH* Gradients, GE Healthcare Life Sciences, 2000, Doc.18-1140-60.
[28] ReadyStrip IPG Strip Instruction Manual, Bio-Rad Laboratories, 2011, Doc. 4006166 Rev. H.
[29] H. Chen, P. Boontheung, R.R. Ogorzalek Loo, Y. Xie, J.A. Loo, J.-y. Rao, M.D. Collins, Birth Defects Res. A Clin. Mol. Teratol. 82 (2008) 187–199.
[30] R.R. Ogorzalek Loo, J.A. Loo, Anal. Chem. 79 (2007) 1115–1125.
[31] A.M. Distler, J. Allison, Anal. Chem. 73 (2001) 5000–5003.
[32] A.I. Gusev, W.R. Wilkinson, A. Proctor, D.M. Hercules, Anal. Chem. 67 (1995) 1034–1041.
[33] C. Koester, J.A. Castoro, C.L. Wilkins, J. Am. Chem. Soc. 114 (1992) 7572–7574.
[34] M. Shahgholi, B.A. Garcia, N.H.L. Chiu, P.J. Heaney, K. Tang, Nucleic Acids Res. 29 (2001). e91/91-e91/10.
[35] L. Qiao, E. Tobolkina, B. Liu, H.H. Girault, Anal. Chem. 85 (2013) 4745–4752.
[36] C.-M. Huang, Arch. Oral Biol. 49 (2004) 951–962.
[37] J. Schaller, K. Akiyama, H. Kimura, D. Hess, M. Affolter, E.R. Rickli, FEBS J. 196 (1991) 743–750.
[38] R. Vitorino, M.J.C. Lobo, A.J. Ferrer-Correia, J.R. Dubin, K.B. Tomer, P.M. Domingues, F.M.L. Amado, Proteomics 4 (2004) 1109–1115.

CHAPTER 2

Virtual 2D Gel Electrophoresis/Mass Spectrometry Optimization and Data Analysis

2.1 Abstract

The virtual 2D gel/MS technique offers a convenient way to separate complex protein mixtures to identify novel protein isoforms, detect changes to proteomes, and more. Building on previous developments, we report on new efforts to increase MS signal intensity, signal-to-noise ratio, and improve sensitivity through the implementation of a novel matrix application method where gels are immersed in matrix solution overnight, allowing for better matrix crystal formation. We also discuss an effort to reduce the time required for data acquisition and analysis by unveiling a new data analysis program developed in MATLAB that allows for greater automation in order to reduce acquisition/analysis time from over a week to under one hour. Final, to improve the resolution and accuracy of virtual gel data we demonstrate the use of a 15 Tesla Fourier transform-ion cyclotron resonance mass spectrometer equipped with a MALDI source to acquire data using the virtual 2D gel/MS technique and compare these results to that obtained using time-of-flight mass spectrometers. All of these developments drive our goal of developing the virtual 2D gel/MS technique to be a convenient and robust method that bridges the divide between top-down and bottom-up proteomics.

2.2 Introduction

The virtual 2D gel/MS technique was initially developed to separate complex biological samples to identify and analyze protein isoforms.[1,2,3] Like traditional 2D-PAGE (polyacrylamide gel electrophoresis) techniques, the initial separation of the proteins is performed on immobilized pH gradient polyacrylamide gels (IPGs), which allows for protein separation based on their isoelectric point (pI), i.e., charge; potentially this has the ability to resolve proteins that differ by as little as 0.001 pH units.[4] The second dimension of analysis, based on molecular mass, is performed using matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS), which can record accurate mass measurements of intact proteins with relatively high throughput. Along with exceptional resolution, the use of IPG gels allows results to be linked directly back to traditional 2D-PAGE gel studies for comparison.[3]

Building on previous developments of the virtual 2D gel/MS technique, the goal of this work was to increase the MS signal intensity and signal-to-noise (S/N) and sensitivity of the method. We were also interested in developing automated methods of data acquisition and processing to create a more streamlined workflow that can produce and display the data rapidly. A primary focus for improving the S/N of the resulting mass spectra was to modify the gel preparation protocol. The sensitivity of MALDI-MS for large biomolecules is dependent on many factors, including the efficiency of the MALDI matrix for desorption/ionization of the analyte, the substrate that the analyte/matrix co-crystals reside on, the presence of contaminants in the sample that may reduce analyte S/N, and many other variables. Many avenues were pursued during this effort, including adjustments to the gel wash procedures, the matrix solution composition, and coating gels with conductive gold nanoparticles both before and after matrix

was applied. To improve data analysis and visualization many software packages were explored, including products from the instrument manufacturer, Bruker Daltonics, and third-party software vendors. Upon failing to find a software product that would meet all our needs, we decided that we would develop our own program written in MATLAB.

In this chapter, I describe efforts that were successful in increasing signal intensity and sensitivity. I also discuss the features of the MATLAB program used for post-acquisition analysis and visualization of results. Finally, I will describe efforts to improve the mass resolution and mass accuracy of results by analyzing virtual 2D gel spots of interest with an FT-ICR (Fourier transform-ion cyclotron resonance) mass spectrometer.

2.3 Methods

2.3.1 Sample preparation

Pelleted *E. coli* and *M. mazei* cells obtained from the Gunsalus lab at UCLA were stored at -80^{o}C before being lysed. These cells were lysed using methods that avoided ionic salts and anionic or cationic detergents as previously discussed.[5] Protein extract concentration was determined using the Pierce 660 nm Protein Assay on the Thermo Fisher Scientific NanoDrop 2000 Spectrophotometer. After protein concentration was determined, the cell lysate was stored at -80^{o}C.

2.3.2 Isoelectric Focusing

Precast 18-20 cm IPG strips (Readystrip, Bio-Rad or Immobiline Drystrip, GE Healthcare) with a pH range of 4-7 were loaded with between 75 and 150 µg of protein from the cell lysate using

passive in-gel rehydration as previously discussed.[5] Gels were then focused for 65 – 80 kV -hrs on a Multiphor II Electrophoresis System equipped with an EPS 3501XL power system. After being focused, the gels were stored at -80^{o}C before further processing.

2.3.3 MALDI matrix application and MS data acquisition

Focused gels were thawed to room temperature and washed for approximately 15 minutes under gentle shaking conditions in a solution of 49.9% H_2O/50% acetonitrile (ACN)/0.1% trifluoroacetic acid (TFA) to remove isoelectric focusing reagents like detergent, urea, thiourea, ampholytes, and residual mineral oil that would hinder mass spectrometry analysis. After being washed the gels were placed gel side up in a tray and blotted with a damp piece of filter paper. Washed gels were then submerged in a 1:3 ACN:H_2O, 0.15% TFA, 1.5% sorbitol solution that had been saturated with sinapinic acid MALDI matrix days in advance. The gels were allowed to soak for either two consecutive periods of 20 minutes or overnight. Overnight soaked gels would not be disturbed until solution levels had evaporated below the tops of the gel, at which point the remaining solution would be removed. In both cases, the gels were then dried at room temperature.

After drying, the gel strips were cut in half and attached to a Bruker Daltonics MTP 384 polished steel target plate with double-sided electrically conductive transfer tape (Adhesives Research, Glen Rock, PA). The gel pieces' positions and lengths were measured and recorded. The target plate was then placed in a container under a vacuum for about 30 minutes to remove additional moisture from the gels. The target plate was then removed from the vacuum and analyzed with either a Bruker Autoflex MALDI TOF/TOF or a Bruker Ultraflex MALDI

TOF/TOF both equipped with a smartbeam II laser operating at 1000 Hz, using the Bruker FlexImaging software to perform an imaging experiment with MALDI data being acquired every 200 – 600 µm of the gel strip. Upon completion of the analysis, data were exported to the universal file format mzXML using Bruker CompassXport software.

2.3.4 Data Acquisition on FT-ICR

Gels still attached to the Bruker Daltonics MTP 384 polished steel target were analyzed further using a 15 T Bruker SolariX FT-ICR MS equipped with a dual ESI/MALDI source and a smartbeam II laser firing at 20 Hz. The instrument was calibrated in ESI mode using a cesium iodide solution. After calibration, each spectrum consisted of 20 laser shots. FT-ICR data was analyzed using Bruker Compass Analysis software.

2.3.5 Virtual Gel Analysis

Using an in-house program developed within MATLAB the MALDI-TOF data was analyzed and visualized.

2.4 Results and Discussion

2.4.1 Improving signal intensity, sensitivity, and speed of the method

To improve the signal intensity and sensitivity of the method, changes were made to the matrix application procedure. IEF gels containing a total of 125 µg of protein from *E. coli* cell lysate were washed and had matrix applied using the typical method, where focused IEF strips were incubated in saturated matrix dissolved in 1:3 ACN:H_2O, 0.15% TFA, 1.5% sorbitol. The gels were soaked in the matrix solution for 20 minutes, followed by an additional 20 minute period

in fresh matrix solution. The new technique utilizes the same 1:3 ACN:H_2O, 0.15% TFA, 1.5% sorbitol matrix solution that has been prepared at least 2 days in advance. Gels are submerged in matrix solution and left undisturbed overnight until the solution has evaporated and the tops of the gels are exposed. At this point, any remaining solution is removed from the container using pipettes and kimwipes. The remaining preparation for analysis remained unchanged.

Upon analysis, a significant improvement in signal intensity and sensitivity can be seen in **Figure 2.1.B** where the matrix was applied using the overnight soaking method compared to **Figure 2.1.A** where the matrix was applied using the two 20 minute soak periods. This improvement can be seen at both low and high *m/z* ranges. It appears that the prolonged soaking and evaporation of the matrix solution allows for maximum crystal formation to occur as opposed to the crystal formation that would occur during the two twenty-minute soaking periods. MALDI-MS sensitivity is affected by matrix crystallization, as well defined and larger crystals have been thought to result in better overall sensitivity. Furthermore, the prolonged soak gives proteins a better opportunity to interact with the matrix as crystal formation occurs.

Even though the overnight soaking method improved the overall signal, there was some concern as to how much pI resolution might be lost during the process. Potentially, loss of resolution in the isoelectric focusing dimension could occur if protein migration occurred in the gel while it was submerged in the matrix solution. To test this theory, we took three gels containing 100 µg protein from *M. mazei* cell lysate that had been focused at the same time and washed them with the standard 49.9% H_2O/50% ACN/0.1% TFA solution. Afterward, one gel was chosen at random and returned to the -80^oC freezer. Another gel was subjected to the usual 2x 20-minute soak method. After the soak was completed it was allowed to dry at

ambient conditions overnight. The third gel was subjected to the new overnight soak, however, was not allowed to dry the next day. After the matrix application was complete the control gel was removed from the freeze and thawed to room temperature. The three gels were subsequently stained with SYPRO Ruby fluorescent protein stain using the typical procedure. As can be seen in **Figure 2.2**, no change in protein migration can be observed when comparing the control gel (**Figure 2.2.A**), twenty minutes soaked gel (**Figure 2.2.B**), and the overnight soaked gel (**Figure 2.2.C**). These results were also confirmed with another set of gels that were subjected to silver staining.

To further demonstrate the isoelectric resolution of gels that have been subjected to overnight soaking matrix application, a gel containing 100 µg of protein from *M. mazei* cell lysate was analyzed using the Bruker Autoflex MALDI-TOF instrument. **Figure 2.3** shows individual spectra about 0.01 pI apart from one another in the 4.91-4.95 pI range. These spectra clearly show the appearance and disappearance of proteins across this pI range and demonstrate that proteins can be separated according to their pI's and measured by MALDI-MS.

2.4.2 Visualization of Virtual 2D Gel Analysis

Along with the optimization of MALDI-MS signal, another goal has been to create an effective platform for the visualization of the virtual 2D gel results. While heat maps created by the vendor-supplied Bruker FlexImaging package had been adequate for initial results, the creation of these heat maps was tedious and involved many hours of work. Furthermore, these heat maps lacked the customizability that many researchers would desire. To improve on

FlexImaging, I investigated numerous other software packages. Finally, MATLAB by MathWorks was chosen due to its flexibility and ability to handle the large data sets generated during gel imaging that can easily be over a gigabyte in size. The program code can be found in the appendix of this chapter.

Data is imported into the processing platform using the mzXML file format, allowing for analysis to be independent of the instrument used for data acquisition. Once inside of MATLAB several data processing steps can be performed, which are outlined in **Figure 2.4** and listed here:

Spectra Summing: To reduce the computer resources required to analyze the large datasets, the spectra within set steps along the length of the gel can be summed. For example, as seen in **Figure 2.4.A**, if the gel was originally analyzed every 600 μm along the length of the gel and analyzed 3 separate times along the width of the gel, for a "step size" of 1200 μm, 6 spectra would be summed together. This setting can be adjusted by the user, with the user entering the step size (in micrometers) in line 93 of the program code. Adjacent points across a step size of moderate size typically have the same proteins present, therefore, only minimal pI resolution is lost by summing.

Invert Images: If needed, gel images can be inverted along the length of the gel. This may be necessary depending on how gels are adhered to the plate after being cut. Ideally, for Bruker instruments, gels should be attached with the low pI at the beginning of the plate increasing in pI from left to right. This feature is especially useful if the user is attempting to stitch together

multiple acquisitions into the same virtual gel. To invert the gel image, users should enter “1” under “NeedToFlip” in line 2 of the code. Otherwise, this should be set to “0”.

Smoothing: After the spectra are summed, they are smoothed using a Savitzky-Golay least-squares digital polynomial filter. Using the default settings, the frame is 15 units. This can be adjusted by the user in line 140.

Baseline Subtraction: After the data is smoothed, the baseline is subtracted with a linear fit. Baseline subtraction starts with a window size of 200 units. This window increases with *m/z* to account for the decrease of resolution with increasing *m/z* that is characteristic of TOF mass spectrometry.

Normalization: After the baseline is subtracted, the data is normalized by standardizing the area under the curve to the median intensity level.

Resampling Data: Data is resampled to homogenize the *m/z* range of all the data. Resampling is also useful for reducing the amount of data, and therefore, the computational resources needed, by removing redundant information in the high-resolution data. As a default, the program is set to cut off any data below *m/z* 4000. This setting can be adjusted in line 151 of the program. There is no high mass cutoff.

Rescaling Data: The user can amplify the signal of a particular *m/z* range if desired. This is accomplished by multiplying the intensity values by a scaling factor, which can be set in line 157 of the code. The *m/z* range of data to be rescaled can be set in line 154 and 155, with “MZ1” representing the beginning of the range and “MZ2” representing the end of the range. If the

user does not want any rescaling of data, the feature can be disabled by entering 1 as the scaling factor. At this point, only 1 *m/z* range can be rescaled at a time.

After these settings have been adjusted to suit the user's needs, the user specifies the number of data acquisitions to import in line 4 of the code. This allows for the user to acquire data in different runs and stitch together the data. The user also needs to specify the length, in millimeters, of the IPG strip that was analyzed. Once these variables are adjusted and the script is executed, MATLAB will inquire about the location of the spot lists and spectrum directories. The spot lists can be generated in the Bruker FlexImaging software under "Export" to inform the program about the X-Y coordinate of each data acquisition and the filename where that acquisition is stored. Upon successful execution, the program saves a heatmap and a 3D map of the data. Examples of the outputs obtained from an IEF gel containing 100 µg of protein from *M. mazei* cell lysate can be seen in a MATLAB generated heat map and a MATLAB generated 3D plot, **Figure 2.5**.

Along with improving the customizability and utility of the gel representation, the program also significantly reduces the number of resources required to process a gel image. Instead of requiring hours of work from a user, the program was able to process the data shown in **Figure 2.5**, which consisted of an 18 cm gel imaged every 600 µm along the length of the gel and every 900 µm along the width, in approximately 20 minutes using a modest personal computer equipped with a dual-core CPU.

2.4.3 MALDI-FT-ICR MS Analysis

To improve the mass resolution and accuracy of the virtual 2D gel analysis, IEF gels were analyzed using an FT-ICR mass spectrometer equipped with a MALDI source. All previous MALDI-MS measurements of IEF gels were acquired with TOF mass spectrometers. There are a few advantages to using an FT-ICR to analyze virtual 2D gels. First, the ionization process is completely decoupled from the mass analysis. This means that slight variations in the height of the matrix on the gels do not impact mass accuracy in the FT-ICR. Also, the FT-ICR can achieve a much greater mass resolution when compared to analysis with TOF. After analyzing an 18 cm gel containing 125 µg of protein from *E. coli* cell lysate with MALDI-TOF instruments, regions of interest were measured for further analysis. Due to the compatibility of the MALDI sources, plates could be transferred directly to the FT-ICR without further preparation. A comparison of mass spectra can be observed where **Figure 2.6.A** represents spectra acquired using MALDI-TOF and **Figure 2.6.B** represents data acquired at the same position using a MALDI-FT-ICR. The resolution from the MALDI-FT-ICR can be seen in the insert of **Figure 2.6** and additional comparisons can be seen in **Figure 2.7**.

With the advantages from using an FT-ICR MS platform comes several potential disadvantages as well. The speed of data acquisition is 5 to 10 times slower on the FT-ICR compared to the TOF instrument; this would greatly limit the throughput of the overall data acquisition strategy. There are also limitations to the *m/z* range at which the FT-ICR can be effective, given the fact that MALDI typically produces single and doubly charged ions, and the average size of proteins from cell lysates is likely to be 30-40 kDa. This may limit the usefulness of FT-ICR for some applications that are interested in species with high *m/z*. In our analysis, we

were successful in detecting ions with mass up to 10,000 Daltons in size. Finally, the FT-ICR requires tedious fine-tuning of settings to optimize the signal and these settings need to be adjusted depending on the mass range of interest. This once again slows the speed of analysis.

Taking these factors into account, it seems the best course of action is to initially screen the IEF gels using a MALDI-TOF platform to identify positions of interest for further higher mass resolution interrogation. This would take advantage of the speed and tolerance of settings for MALDI-TOF measurements. After these positions are identified, further analysis using a MALDI-FT-ICR platform can be utilized to obtain more accurate mass measurements by fine-tuning the setting on the FT-ICR mass spectrometer. It may also be possible to use the accurate mass measurements from the FT-ICR to calibrate MALDI-TOF spectra. It should be noted that the Bruker SolariX MALDI-FT-ICR platform is equipped with FlexImaging software and is capable of imaging IEF gels using the same workflow required to use FlexImaging on Bruker MALDI-TOF instruments if a user was interested in imaging a complete gel.

2.5 Conclusions

Through the course of our experiments, we were able to improve the sensitivity and signal-to-noise of the virtual 2D gel/MS platform by introducing a novel matrix application technique that employs an overnight soak of gels in matrix solution. The larger matrix crystals that were produced from this method resulted in enhanced S/N of the MS signals from the desorption/ionization of the proteins embedded in the gels. We were also able to significantly decrease the amount of time required to analyze and produce virtual 2D gel images through the development of a MATLAB program that automates data analysis. By using this program in

conjunction with a Bruker AutoFlex MALDI-TOF instrument, we were able to reduce the amount of time required to acquire and analyze virtual 2D gel/MS data to about 1 hour or less. Taking into account the gel focusing, i.e., protein separation, and MALDI matrix application procedures, it is possible to prepare and analyze a complex lysate sample in two days or less, a process that would have taken over 1 week without the automation of the data analysis. Finally, we were able to demonstrate the use of an FT-ICR high resolution mass spectrometer to analyze the virtual 2D gels. This analysis has the advantage of increasing both mass accuracy and mass resolution of the data. Overall, sensitivity, throughput, and mass accuracy/resolution of the virtual 2D gel platform was increased, allowing more researchers to apply the technique to rapidly profile complex proteomes.

2.6 Appendix

MATLAB Script:

```
%user input 1=YES 0=NO 2=SECTION DOES NOT EXIST!!!
NeedToFlip=[0; 0]; % ; represents break in file. Spaces represent different
regions within the same file.
%Import mzXML files
numberOfFiles=2; %USER INPUT REQUIRED
LengthOfGel= 180; %Report in mm ONLY
for p= 1:numberOfFiles
[SpotListFile{1,p}, SpotListPath{1,p}]= uigetfile('*.txt', 'Select Spot
List');
z{1,p}= uigetdir([],'Select Spectrum Directory');
end
for p= 1:numberOfFiles
files = fullfile(z{1,p},'0_*');
mzxmlfiles = dir(files);
numfiles(p,1)= length(mzxmlfiles);
totalspec= cell(numfiles(p,1),1);
for i= 1:numfiles(p,1)
f=fullfile(z{1,p},mzxmlfiles(i).name,'1','1SLin','Analysis.mzXML');
totalspec{i}= mzxmlread(f);
end
for k = 1:numfiles(p,1)
    peaks(k,p)= mzxml2peaks(totalspec{k,1});
end
clearvars  i k totalspec f mzxmlfiles files p
end
clear z
%Resample all spectra uniform number of sampling points
lengFINAL = 1;
for p= 1:numberOfFiles
for i= 1:numfiles(p,1)
    lengconditional = length(peaks{i,p});
    if lengconditional>lengFINAL
        lengFINAL = lengconditional;
    end
end
end
clearvars i p lengconditional
for p= 1:numberOfFiles
for i = 1:numfiles(p,1)
[reformedpeaks{i,p}(:,1),reformedpeaks{i,p}(:,2)] =
msresample(peaks{i,p}(:,1), peaks{i,p}(:,2), lengFINAL, 'Uniform', 'true');
end
end
clearvars peaks i p
MZ=reformedpeaks{1,1}(:,1);
l= length(MZ);
for p= 1:numberOfFiles
for i = 1:numfiles(p,1)
    peaks{1,p}(:,i)= reformedpeaks{i,p}(:,2);
end
```

```
end
clearvars i p reformedpeaks
for p= 1:numberOfFiles
%% Initialize variables.
filename= fullfile(SpotListPath{1,p},SpotListFile{1,p});
delimiter = ' ';
startRow = 3;
%% Format string for each line of text:
formatSpec = '%f%f%s%f%*s%*s%*s%*s%*s%*s%*s%*s%*s%[^\n\r]';
%% Open the text file.
fileID = fopen(filename,'r');
%% Read columns of data according to format string.
dataArray = textscan(fileID, formatSpec, 'Delimiter', delimiter,
'MultipleDelimsAsOne', true, 'HeaderLines' ,startRow-1, 'ReturnOnError',
false);
%% Close the text file.
fclose(fileID);
%% Create output variable
dataArray([1, 2, 4]) = cellfun(@(x) num2cell(x), dataArray([1, 2, 4]),
'UniformOutput', false);
SpotList = [dataArray{1:end-1}];
%% Clear temporary variables
clearvars filename delimiter startRow formatSpec fileID dataArray ans;
%% Sort Peaks Based on Section
sortSpotList{1,p}= sortrows(SpotList, 3);
clearvars SpotList
end
clear p
for p= 1:numberOfFiles
region= cell2mat(sortSpotList{1,p}(:,4));
num = unique(region);
leng{1,p} = length(num);
for i=1:leng{1,p}
    section{i} = find(region==i);
    sortSectSpotList{i,p}= sortSpotList{1,p}(section{i},:);
    sortpeaks{i,p}= peaks{1,p}(:,section{i});
end
clearvars section i num region
end
clear sortSpotList peaks
%% Bin Peaks Based on Section
Nfinal= cell(leng{1,1},1);
Efinal= cell(leng{1,1},1);
binFinal= cell(leng{1,1},1);
for p=1:numberOfFiles
for i=1:leng{1,p}
xPosition = cell2mat(sortSectSpotList{i,p}(:,1));
first = xPosition(1,1);
stepsize = 1200; %User Input
last = xPosition(end,1);
last1 = last+ stepsize;
edge = (first:stepsize:last1)';
[N, E, bin] = histcounts(xPosition, edge);
Nfinal{i,p} = N;
Efinal{i,p} = E;
binFinal{i,p} = bin;
end
```

```
clearvars N E bin first last last1 stepsize edge xPosition
end
clear sortSectSpotList
%% Average and Combine Spectrum
for p= 1:numberOfFiles
for y= 1:leng{1,p}
    lengN= length(Nfinal{y,p});
for z=1:lengN
    sumN(z)= sum(Nfinal{y,p}(1:z));
end
for i= 1:lengN
    splitspec{y,i}= sortpeaks{y,p}(1:end, sumN(:,i)-
Nfinal{y,p}(:,i)+1:sumN(:,i));
    avgspec{y,i}= mean(splitspec{y,i},2);

end
    combinedSpec{y,p}= horzcat(avgspec{y,:});

end
end
clearvars leng lengN avgspec splitspec Nfinal binFinal Efinal y z i sortpeaks
sumN
%% Flip Spectrum if Needed
S= size(NeedToFlip);
for p= 1:S(1,1)
    for i=1:S(1,2)
        if NeedToFlip(p,i)==1
            finalSpec{i,p}= fliplr(combinedSpec{i,p});
        end
        if NeedToFlip(p,i)==0
            finalSpec{i,p}= combinedSpec{i,p};
        end
    end
end
for p=1:numberOfFiles
    sections{1,p}= horzcat(finalSpec{:,p});
end
Y= horzcat(sections{1,:});
clearvars p i combinedSpec finalSpec NeedToFlip numberOfFiles numfiles S
sections
%%Smooth Spectra
Ysmooth = mssgolay(MZ, Y);
%Adjust Background
backfun = @(Y) 200 +Y.*(8/300);
Yback = msbackadj(MZ,Ysmooth, 'WindowSize', backfun, 'SmoothMethod',
'rlowess');
clearvars Ysmooth
%Normalize Spectra
%default span is 15
%can be adjusted by adding "mssgolay(MZ, Yback,'Span', SpanValue,)
Ynorm = msnorm(MZ, Yback,'Span', 15);
clearvars Yback
%Resample to half of original points
lre= 1./2;
lengFINAL= round(lre, 0);
```

```
[MZresample, Yresample] = msresample(MZ,Ynorm, lengFINAL , 'Range', [4000
max(MZ)]);
clearvars MZ Ynorm
%Rescale certain masses
MZ1=20001;
MZ2=52157;
h=find(MZresample>MZ1 & MZresample<MZ2);
scalefactor= 5; %SCALE FACTOR
Yrescaled= Yresample;
rSection= size(Yrescaled, 2);
for i= 1:rSection
Yrescaled(h,i)= Yrescaled(h,i).*scalefactor;
end
clearvars i h
SeparationFactor= LengthOfGel/rSection;
Position= (SeparationFactor:SeparationFactor:LengthOfGel)';
%Make 3D Plot
figure
X=MZresample;
Intensity= Yrescaled;
L=length(Intensity);
t1 =-1;
t2 =66;
h = find(Position>t1 & Position<t2);
plot3(repmat(X,1,numel(h)),repmat(Position(h)',L,1),Intensity)
%Make Heat Map
heatmap= msheatmap(MZresample, Position, Yrescaled);
print('Heatmap123', 'meta')
```

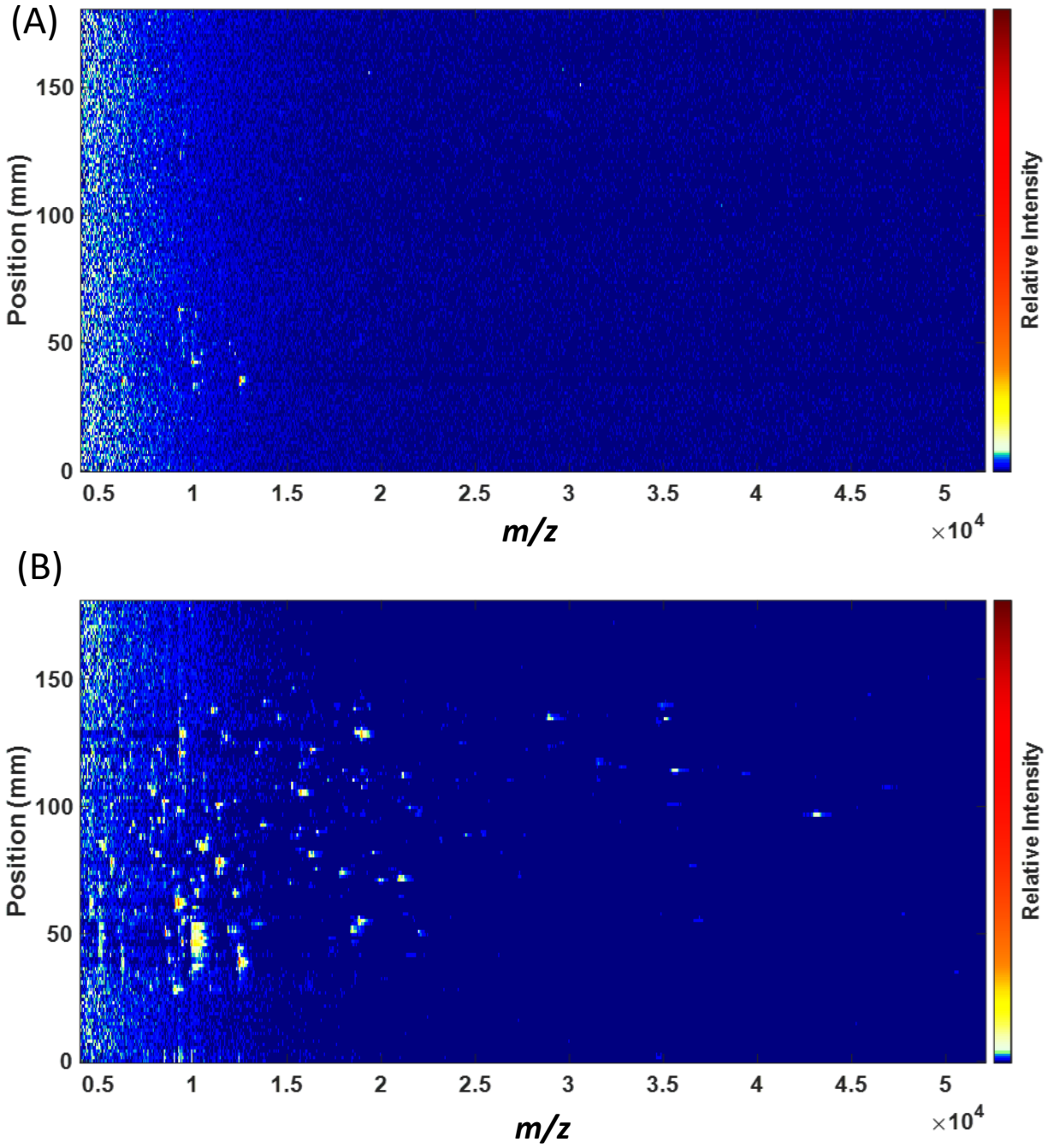


Figure 2.1: Comparison of matrix application methods. Both gels are loaded with 125 µg of protein from *E. coli* cell lysate. Gel with matrix applied using 20 minute soak method is seen in (A). Gel with matrix applied using the overnight soak method is seen in (B).

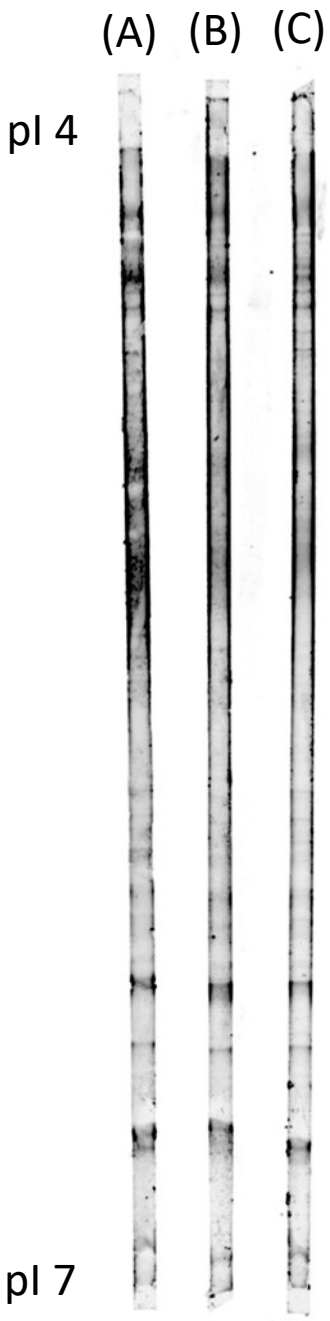


Figure 2.2: SYPRO Ruby-stained IPG gels containing *M. mazei* cell lysate demonstrating that different durations of matrix solution soaking do not cause significant protein diffusion. (A) control IPG gel that only washed in 0.1% TFA/50% acetonitrile for 15 minutes. (B) IPG gel was washed in 0.1% TFA/50% acetonitrile for 15 minutes and then soak in matrix solution for two consecutive 20 minute periods. (C) IPG gel was washed with 0.1% TFA/50% acetonitrile for 15 minutes and incubated in matrix solution overnight.

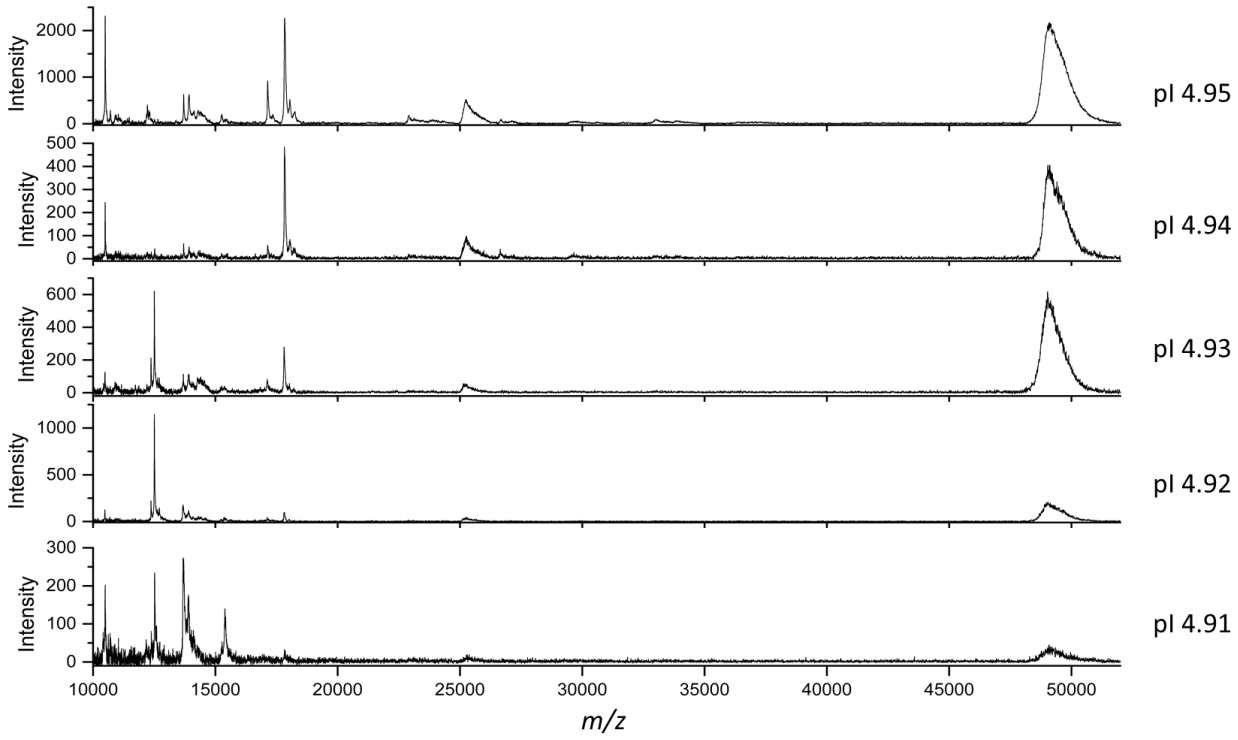


Figure 2.3: Individual spectra taken from the analysis of an IPG gel containing 100 µg of protein from *M. mazei* cell lysate prepared using the overnight matrix application protocol. The spectra represent a narrow range of isoelectric points along the gel 4.91—4.95, representing the isoelectric focusing resolution that can be obtained while using the overnight soaking method.

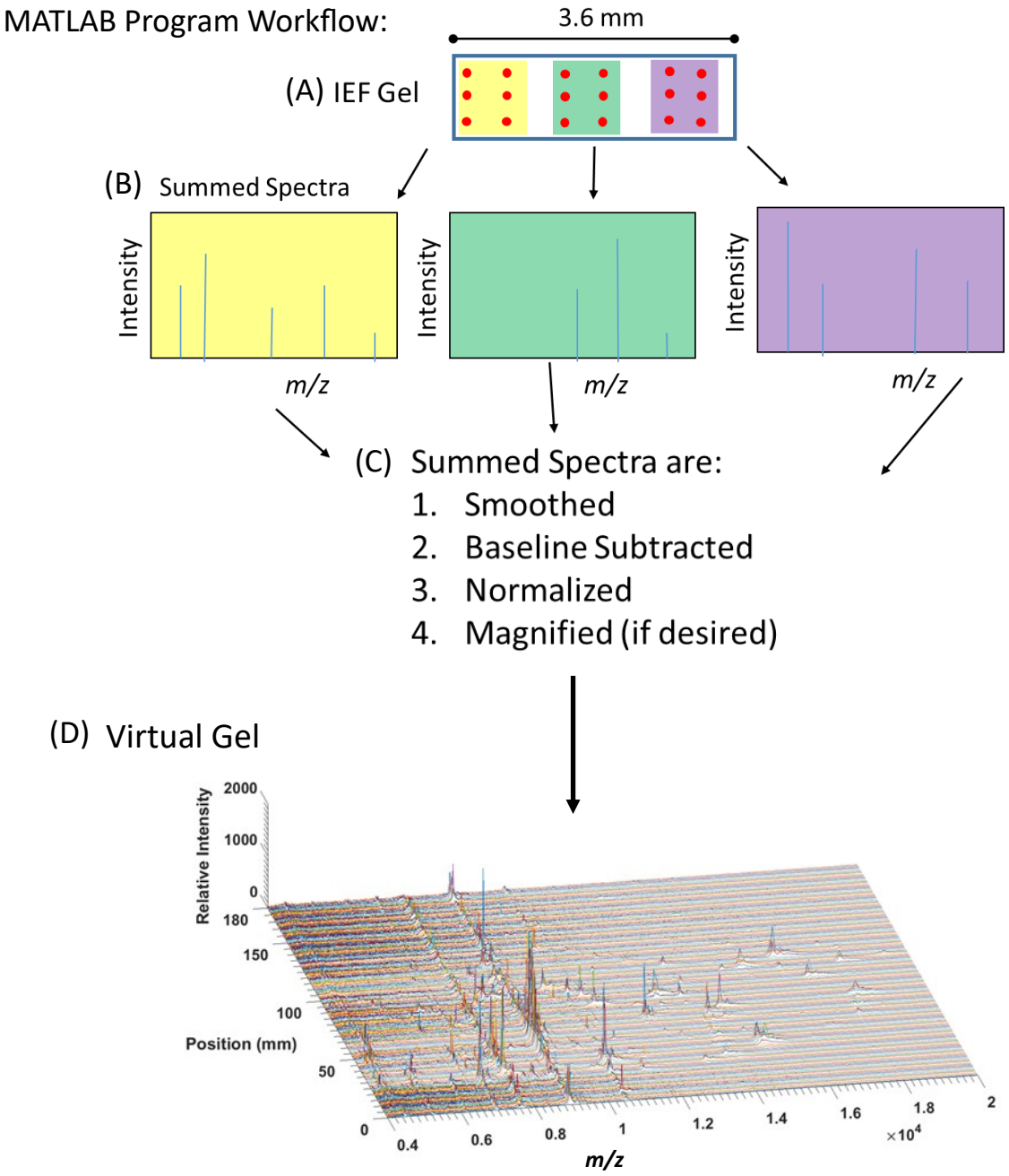
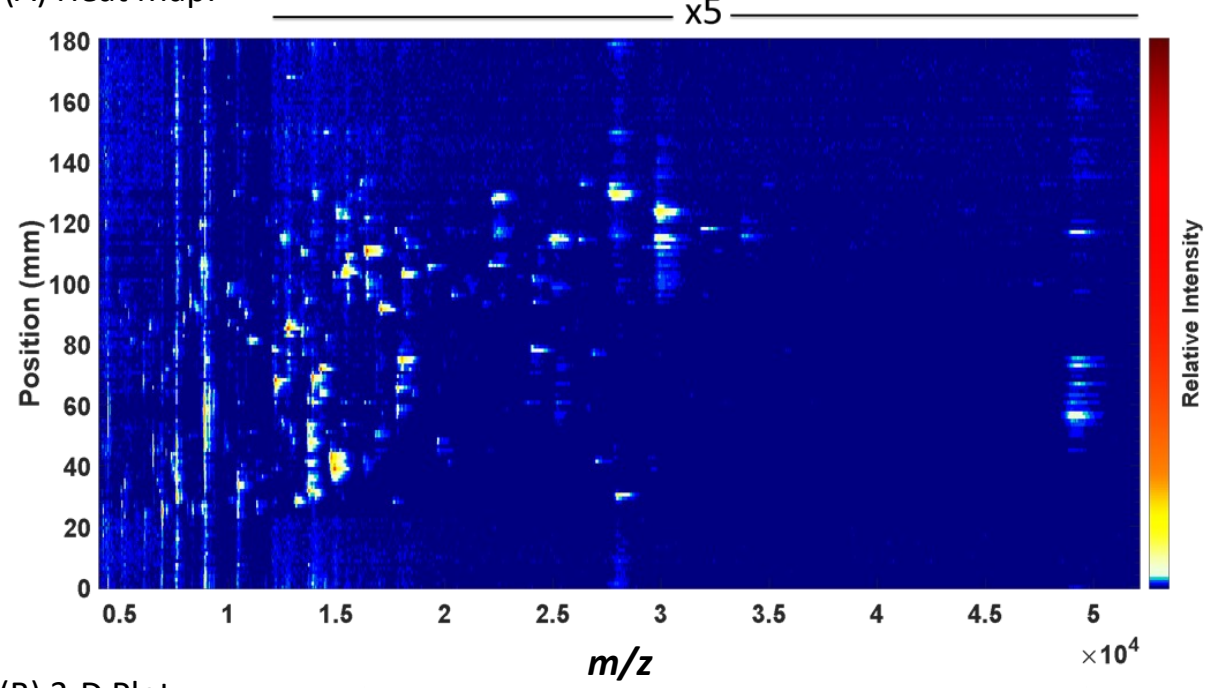


Figure 2.4: MATLAB program workflow. (A) Spectra within a defined area on the IPG gel are summed together to produced summed spectra (B). These summed spectra are smoothed, baselined subtracted, normalized, and magnified (if desired) in that order. After processing is complete a virtual gel (D) is produced. A heat map can also be produced if desired (not displayed here).

(A) Heat Map:



(B) 3-D Plot:

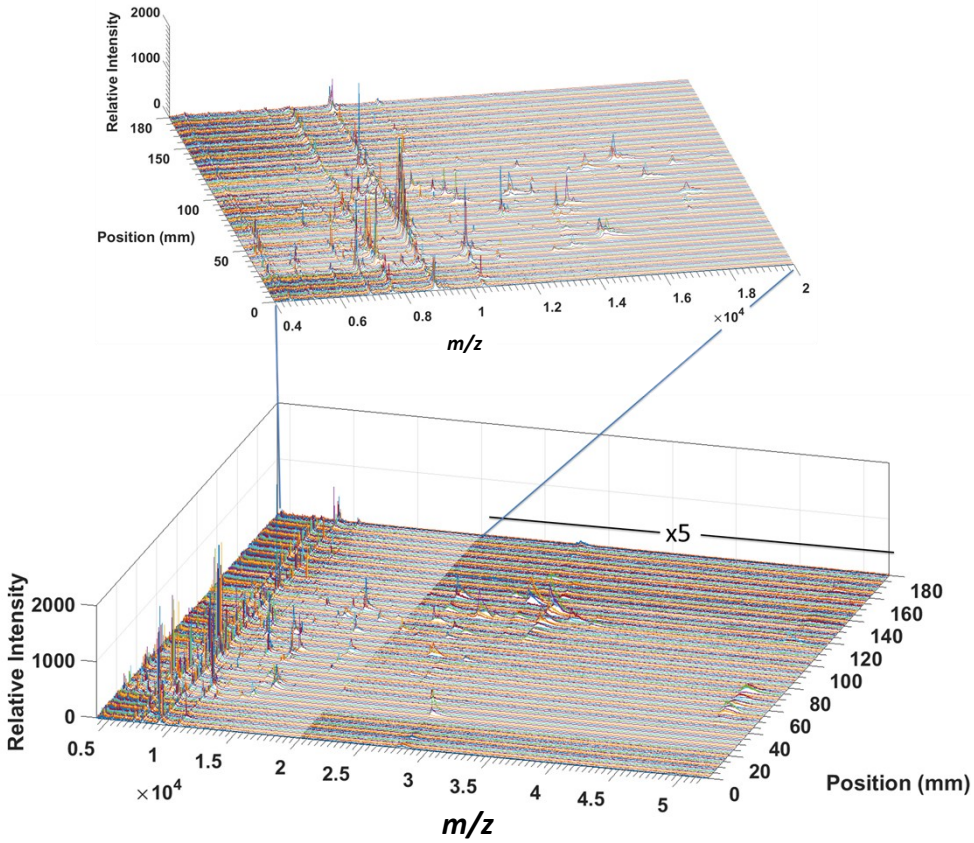


Figure 2.5: MATLAB program generated heatmap (A) and 3-D plot (B) of IPG gels containing 100 µg *M. mazei*.

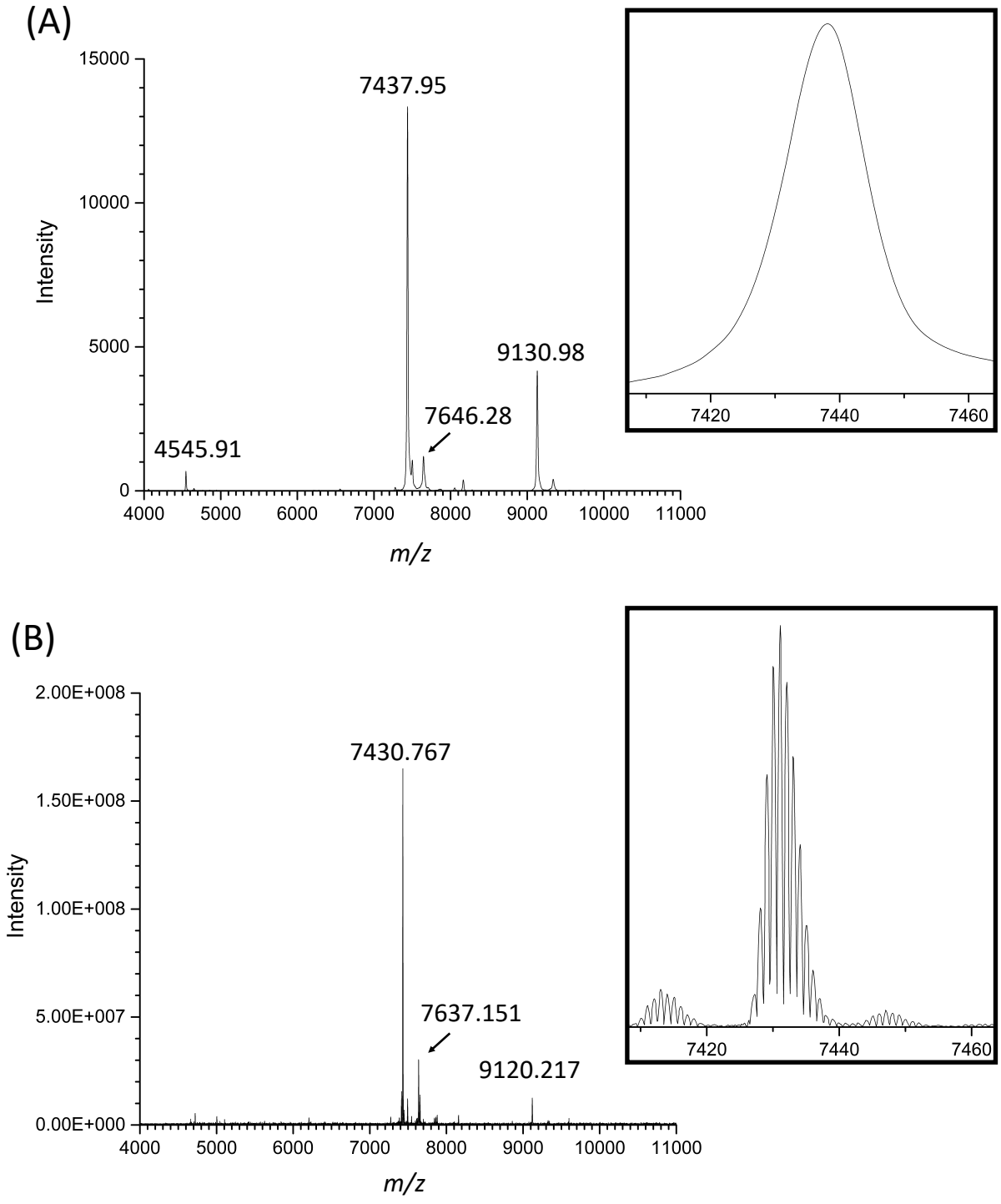


Figure 2.6: Comparison of spectra from IPG gel containing 125 µg *E. coli* cell lysate acquired using MALDI-TOF (A) and MALDI-FT-ICR (B). The base peak from both spectra have been enlarged to show resolution of results (inserts).

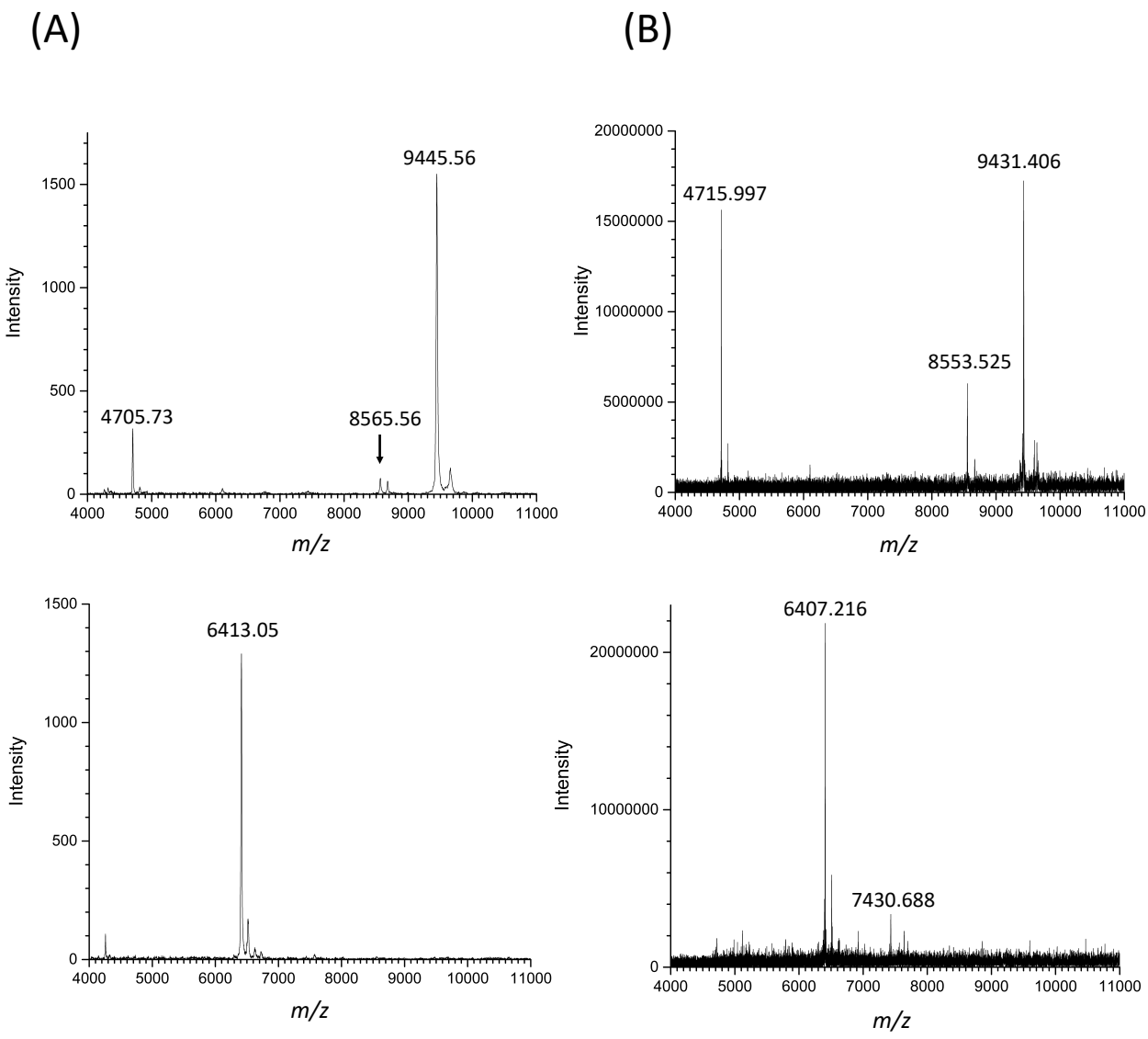


Figure 2.7: Comparison of spectra from IPG gel containing 125 µg *E. coli* cell lysate acquired using MALDI-TOF (A) and MALDI-FT-ICR (B).

2.7 References

(1) Ogorzalek Loo, R. R.; Mitchell, C.; Stevenson, T.; Loo, J. A.; Andrews, P. C. Interfacing Polyacrylamide Gel Electrophoresis with Mass Spectrometry; Marshak, D. R. B. T.-T. in P. C., Ed.; Academic Press, 1996; Vol. 7, pp 305–313. https://doi.org/https://doi.org/10.1016/S1080-8914(96)80034-2.

(2) Ogorzalek Loo, R. R.; Cavalcoli, J. D.; VanBogelen, R. A.; Mitchell, C.; Loo, J. A.; Moldover, B.; Andrews, P. C. Virtual 2-D Gel Electrophoresis: Visualization and Analysis of the E. Coli Proteome by Mass Spectrometry. *Anal. Chem.* **2001**, *73* (17), 4063–4070. https://doi.org/10.1021/ac0101858.

(3) Ogorzalek Loo, R. R.; Hayes, R.; Yang, Y.; Hung, F.; Ramachandran, P.; Kim, N.; Gunsalus, R.; Loo, J. A. Top-down, Bottom-up, and Side-to-Side Proteomics with Virtual 2-D Gels. *Int. J. Mass Spectrom.* **2005**, *240* (3), 317–325. https://doi.org/https://doi.org/10.1016/j.ijms.2004.10.013.

(4) Smejkal, G. B. High Speed Isoelectric Focusing of Proteins Enabling Rapid Two-Dimensional Gel Electrophoresis; Magdeldin, D. J. B. P. E.-S., Ed.; IntechOpen: Rijeka, 2012; p Ch. 10. https://doi.org/10.5772/36854.

(5) Lohnes, K.; Quebbemann, N. R.; Liu, K.; Kobzeff, F.; Loo, J. A.; Ogorzalek Loo, R. R. Combining High-Throughput MALDI-TOF Mass Spectrometry and Isoelectric Focusing Gel Electrophoresis for Virtual 2D Gel-Based Proteomics. *Methods* **2016**, *104*, 163–169. https://doi.org/https://doi.org/10.1016/j.ymeth.2016.01.013.

CHAPTER 3

Application of Virtual 2D Gel Electrophoresis/Mass Spectrometry for Profiling Proteomes

3.1 Abstract

Despite recent developments in top-down proteomics, sample complexity continues to hinder its implementation on the proteome scale. Here, we demonstrate the use of the virtual 2D gel electrophoresis/mass spectrometry (virtual 2D gel/MS) technique to effectively separate protein in *Escherichia coli* (*E. coli*) cell lysate and detect proteome changes caused by different forms of stress. We examine the effect of cold shock on the bacteria and detect the production of various cold shock related protein. These results are verified using in-gel trypsin digestion of identical IPG strips and further confirmed by comparison to previous results obtained using traditional 2D-PAGE. The antibiotic actinonin, a peptide deformylase (PDF) inhibitor, is also used to stress *E. coli*. We detect subtle changes caused by the failure of PDF to cleave an N-formyl group off the newly synthesized protein, which also prevents the removal of the initiator methionine. Future work will include the development of a robust search platform to aid in data processing.

3.2 Introduction

Because proteins, “workhorses” of the cell, take part in essentially every molecular structure and activity of life processes, studying the proteomes can help elucidate cellular pathways and processes that genomic and transcriptomic studies cannot offer. Mass spectrometry (MS) has become the leading analytical tool in proteomics since it allows global qualitative and quantitative analysis of protein composition, modifications, and dynamics under changing conditions.[1] The majority of mass spectrometry experiments have been using a Bottom-Up MS approach, where proteins are first proteolytically digested into peptides before MS interrogation.[2] This approach has some limitations such as not distinguishing peptides from proteins highly related to each other, losing the coordination of posttranslational modifications (PTMs), and having difficulties with addressing the existence of multiple protein isoforms.[3] The Top-Down MS approach, where proteins are studied in the intact form and fragmented within the mass spectrometer to gain information on sequence, identity, and the presence and sequence location of PTMs, has the promise to overcome the limitations of Bottom-Up MS.[2] With the continuing advancement of mass spectrometer sensitivity, MS fragmentation methods, and sample preparation, Top-Down MS aims to become a high-throughput platform for proteomics and compete with Bottom-Up MS. Nevertheless, the complexity of the proteome continues to be challenging for routine Top-Down MS, with complicated hours-long off- and on-line liquid chromatography separations typically being used as a solution for addressing complex samples.[4]

Virtual 2D gel/MS combines proteome-scale separation with MALDI mass spectrometry, offering another solution to the complexity issue mentioned above. Initially, proteins are

separated by isoelectric focusing (IEF) on immobilized pH gradient (IPG) gels. Subsequent size "separation" is performed using matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) to mass-measure intact proteins.[5] Along with applying a simple well-studied separation method, i.e., IEF, the virtual 2D gel/MS method offers high sensitivity, good mass measurement accuracy, minimal sample consumption, and the ability to probe membrane proteins that are often lost in SDS-PAGE or liquid chromatography separations.[6]

Here we demonstrate the applicability of virtual 2D gel/MS to profile complex samples by detecting changes in the *E. coli* proteome caused by various forms of stress. The first stress that we analyzed is the cold shock response of *E. coli*. When exposed to a dramatic reduction in temperature from 37°C to 10°C, the organism halts production of its typical proteins and instead produces so called cold shock proteins.[7] CspA was the first cold shock protein discovered in 1990 by Neidhardt and coworkers.[8] Following this discovery additional CspA homologues have been discovered in *E. coli* with CspB, CspG, and CspI being cold shock inducible.[9,10] In addition, other proteins have subsequently been identified as important to the organism's response to cold shock including the ATP-dependent RNA helicase CsdA, a 30S ribosomal binding factor, RbfA, and the transcription termination/antitermination protein NusA, among others.[11,12] Together these proteins help the organism overcome issues caused by the reduced temperature including the increased stabilization of RNA and DNA and the reduced efficiency of translation and transcription that this stabilization causes. Once the cold shock proteins have been produced, some at a level 10 times greater than observed under typical conditions, the organism is able to resume normal protein synthesis and growth.[13]

The second stress that we analyze is the response of *E. coli* to the presence of the antibiotic actinonin at low concentrations. Actinonin is a pseudo-peptide with antibiotic properties that was originally isolated from the bacteria *Streptomyces* and *Actinomyces* and works by targeting the peptide deformylase (PDF) enzyme in bacteria.[14,15] PDF is an enzyme present in bacteria and is responsible for cleaving the N-formyl group off newly synthesized polypeptides.[16] Furthermore, this failure to remove the N-formyl group also prevents the removal of the initiator methionine that is sometimes removed from the mature protein as well. Actinonin can do this by tightly binding to PDF thus inhibiting its activity.[15] PDF has been an attractive target for new antibiotics because of its widespread use in bacteria and the lack of a similar pathway in humans.[17] Wild type *E. coli* have efflux pumps that work as a defense mechanism against actinonin and other antibiotics by pumping these harmful molecules out of the cell. Here we use a *tolC* knockout strain of *E. coli* that has the efflux pump disabled leaving the cells susceptible to actinonin.[17]

3.3 Methods

3.3.1Culture Preparation: Cold Shock Experiment

A single colony of *E. coli* strain ER2738, a derivative of strain K12, was selected and cultivated using standard microbiology techniques. Once the culture reached an optical density at 600 nm (OD_{600}) of approximately 0.48 the culture was separated. One half of the culture was harvested. The other half was placed in an incubator at 10°C for 4 hours. At the end of the 4 hour period, the OD_{600} of the culture was approximately 0.52. At this point, this half of the culture was also harvested. Pelleted cells were stored at -80°C until further processing.

3.3.2 Culture Preparation: Actinonin Experiment

A single colony of *E. coli* strain *tolC* CAG 12184, a *tolC* knockout that is missing an outer membrane protein, thereby disabling the efflux pump, was selected and cultivated in 2YT medium using standard microbiology techniques. Once the culture reached the early log phase with an OD_{600} of about 0.15 it was divided, and half of the culture was exposed to actinonin at a concentration of 1 μg/mL. The separate cultures were allowed to continue to grow and were harvested when the control culture had reached an OD_{600} of approximately 2.05 and the culture exposed to actinonin reached an OD_{600} of about 1.45.

3.3.3 Intact Protein Profiling from Whole Bacteria Cells

Pelleted *E. coli* cells were thawed on ice and diluted with water to a protein concentration of approximately 200 μg/mL. This solution of cells was then diluted with saturated sinapinic acid solution at a ratio of 9:1. This solution was then spotted onto a steel target plate and analyzed using a Voyager DE-STR TOF mass spectrometer (Applied Biosystems).

3.3.4 Cell Lysate Preparation

Pelleted *E. coli* cells were thawed on ice and lysed using methods that avoided ionic salts and anionic and cationic detergents. Protein extract concentrations were determined using the Pierce 660nm Protein Assay measured with a NanoDrop 2000 Spectrophotometer (Thermo Scientific). Cell lysate was stored at -80°C after protein concentration was determined.

3.3.5 Isoelectric Focusing (IEF) Separation

Precast 180 mm IPG strips (Readystrip, Bio-Rad or Immobiline Drystrip, GE Healthcare) with a pH range of 4-7 were loaded with 150 µg of protein from the cell lysate using passive in-gel rehydration as previously discussed. Gels were then focused for 65 – 80 kV -hrs on a Multiphor II Electrophoresis System equipped with an EPS 3501XL power system. After being focused, the mineral oil was wicked away and the gels were stored at -80^{o}C before further processing.

3.3.6 MALDI Matrix Application and MALDI-MS Data Acquisition

Frozen protein-focused gels were thawed and washed in 0.1% TFA/50% acetonitrile for 15 minutes to remove isoelectric focusing reagents that would hinder mass spectrometry analysis. Washed gels were then submerged in days-old 1:3 acetonitrile:water, 0.15% TFA, 1.5% sorbitol solution with saturated sinapinic acid. The gels were allowed to soak overnight followed by room temperature drying. Dried gels were attached to a MALDI target plate using electrically conductive double-sided tape and placed in a vacuum desiccator for 30 minutes to remove additional moisture. Gels were then removed from the desiccator and MALDI-MS data was acquired with a Bruker Autoflex MALDI TOF/TOF mass spectrometer and the raw data was visualized using Bruker FlexImaging software.

3.3.7 IEF-MS Data Analysis

MS data was exported into the mzXML format for analysis using Bruker CompassXport. Using the in-house MATLAB program discussed previously, the IEF-MS data was analyzed. After the data was imported into the MATLAB environment, all spectra from each 1 mm gel segment (approx. 8-10) were summed. The resulting spectrum was then smoothed, baseline subtracted,

and normalized using MATLAB algorithms. Finally, the *m/z* range beyond 20,000 was magnified by x5 to improve visibility.

3.3.8 In-Gel Digest

Focused 180 mm IPG strips containing experimental and control cells from the *E. coli* cold shock experiment were thawed and manually cut into approximately 6 mm segments. The 30 gel segments were then placed into individual microcentrifuge tubes with the polyester backings still attached. The gel-embedded proteins were disulfide-reduced by DTT, alkylated by iodoacetamide, and digested with trypsin.

The resulting tryptic peptide extracts were desalted with an Acclaim Pepmap100 C18 nano-trap column and separated with an Acclaim Pepmap RSLC analytical column (Dionex). The tryptic peptides were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) with an EASY nLC1000 (Thermo Scientific) coupled with a Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Scientific).

Mass spectral data from each segment were searched individually in Proteome Discoverer v1.4 (Thermo Scientific) with MASCOT (Matrix Science) sequence database searching against an *Escherichia coli* strain K-12 substrain DH10B database. The following search parameters were used: 3 missed cleavages allowed, 20 ppm mass tolerance, and a minimum of 3 unique peptides.

3.4 Results and Discussion

3.4.1 Intact Protein Profiling of *E. coli*

After growing the cells, it was important to quickly confirm that the stresses we induced had caused changes in our cultures of *E. coli* before subjecting the cells to the more time-consuming lysis and virtual 2D gel/MS procedures. To do this we analyzed intact *E. coli* cells from both the control and experimental groups by simply spotting a mixture of cells and MALDI matrix on a MALDI target plate. The application of the MALDI matrix caused lysis of the cells to release the proteins. Direct MALDI-MS of the lysed cells gave us the ability to get a quick qualitative examination of the proteins from the cells.

The results of this analysis for the cold shock experiment can be seen in **Figure 3.1**. From these results, we can clearly see differences in the mass spectra with the most pronounced changes being additional protein peaks observed in the experimental (cold shock) sample between 7 and 8 kDa.

Differences in the control and experimental group in the actinionin experiment were more subtle as seen in **Figure 3.2**. Upon close inspection of the control and experimental mass spectra, there appears to be a splitting of peaks in the experimental mass spectrum as if two forms of the same protein are present. This difference may have been the result of the actinonin affecting the post-translational modification of polypeptides by inhibiting PDF from cleaving the formyl group from the newly synthesized protein. This inhibition would lead to a mass difference of 28 Da corresponding to the formyl group or 159 Da corresponding to the formyl group and initiator methionine, which cannot be removed if PDF fails to remove the

formyl group. Overall, the results from both experiments (cold shock and actinonin) presented signs that the stress induced changes in the *E. coli* proteome and that the cells should be processed for further analysis using the virtual 2D gel/MS technique.

3.4.2 Virtual 2D/gel MS of Cold Shock of *E. coli*

After lysing the control and experimental samples of *E. coli* cells from the cold shock experiment the lysate proteins were separated using 180 mm IPG strips with a pH range of 4 – 7 and analyzed with MALDI-MS using the virtual 2D gel/MS technique. Upon analyzing the results, differences in the virtual gels can be seen immediately between the control gel in **Figure 3.3.A** and the cold shock gel in **Figure 3.3.B**.

An intense band of signal can be observed in the cold shock gel around 7 kDa between pI 5.5 and pI 6.5, circled in **Figure 3.4**, which is not present in the control gel. These peaks seem to correspond to the CspA that has a calculated pI of 5.57 and a molecular weight of approximately 7.3 kDa (**Figure 3.5**). At a higher calculated pI of 5.64, CspG with a calculated molecular weight of 7.8 kDa can be seen in **Figure 3.6**. Finally, at a calculated pI of 6.53 and a calculated molecular weight of 7.7 kDa, CspB is detected in **Figure 3.7**.

These results were confirmed using a bottom-up MS approach where focused 180 mm IPG strips loaded with the same amount of control and experimental lysate were sectioned as shown in **Table 3.1** and subjected to in-gel digestion where trypsin cleaves proteins into small peptides (i.e., trypsin proteolysis cleaves after Arg and Lys residues). Upon analysis of these peptides, a number of cold shock proteins were identified as shown in **Table 3.2**. Notably, CspA, CspB, and CspG were identified in several sections of the experimental gel, while only being

found in one, or in the case of CspB, none of the control gel sections. While these in-gel digest results do not necessarily provide quantitative results as to the exact concentrations of the protein, it can be assumed that protein found in multiple sections are more widespread on the gel and would explain the appearance of the 7 - 8 kDa signal seen in the virtual 2D gel/MS results. Another protein that appears in more sections of the cold shock gel than the control gel is H-NS (histone-like nucleoid structuring protein), however, upon comparison of the experimental and control virtual gels, a signal from this protein is not easily identified. Interestingly, the other cold shock proteins identified during the in-gel digest experiment were found in approximately the same number of sections of both the control and experimental gels.

The results from the virtual gel and in-gel digest analysis also correspond with results previously reported by Inouye and coworkers, who performed cold shock experiments on *E. coli* strain SB221 where the cells were shifted from 37°C to 15°C upon reaching an OD_{600} of approximately 0.4.[18] Their conventional 2D-PAGE gels, with a pI range of 3-10, have the same broad signal around 7 kDa, corresponding to the cold shock proteins CspA, CspG, and CspB, spanning a large pI range (**Figure 3.8)**. Overall, these results confirm that the virtual 2D/gel MS technique can detect differences in the *E. coli* proteome caused by the cold shock response.

3.4.3 Virtual 2D/gel MS of *E. coli* Upon Addition of Actinonin

The virtual 2D/gel MS technique was also used to analyze the *E. coli* proteome that had been exposed to the PDF inhibitor actinonin. Upon analyzing the virtual gels, **Figure 3.9**, it became apparent that the differences were much more subtle than the results from the cold shock experiment, consistent with the results from the intact protein profile experiment (**Figure**

3.2). One general change appears to be a slight decrease in the number of peaks that are present in the experimental results. To aid in the search for differences in the proteome caused by the antibiotic, the gel images were superimposed, and differences were identified. To further narrow the interrogation of the results, only differences consistent with the effect of actinonin, e.g., small increase in mass and a decrease in pI, were considered.

One difference that conformed to these criteria was a signal observed with a pI slightly greater than 6.5 and a mass around 12 kDa in the control virtual gel. Upon inspection of the experimental (actinonin addition) virtual gel a similar signal was found to have a pI of less than 6.5 and a slight increase of mass. Upon analysis of the raw mass spectra from the control gel, **Figure 3.10.A**, a protein peak was identified to have a mass of 12704.7 Da and a pI of 6.6. From the corresponding gel from the addition of actinonin to *E. coli*, a protein peak in the mass spectrum, **Figure 3.10.B**, was determined to have a mass of 12860.9 Da and a pI of 6.4. The difference in mass between these two peaks was calculated to be 156 Da, close to the 159 Da difference expected from the failure to remove the formyl group and the initiator methionine due to the inhibition of PDF. Another example of a change that may have been caused by the addition of actinonin is seen in the control virtual gel with a mass of approximately 7 kDa and a pI between 6.5 and 7.0. Upon inspection of the experimental virtual gel, this spot is no longer visible, instead a signal closer to 6.0 pI is now visible. Closer inspection of the raw mass spectra indicates that the control gel peak, **Figure 3.11.A**, has a mass of 7300.5 Da and a pI of 6.8. The experimental peak, **Figure 3.11.B**, has a mass of 7456.7 Da at a pI of 6.1. Once again, the difference in mass is approximately 156 Da with a pI that has decreased in the experimental data. It should be noted that with the use of MALDI in-source dissociation or other

fragmentation methods the peptide sequence and identities of these proteins can be determined, unfortunately, this information is not currently available.

Overall, the difficulty of identifying the subtle differences in the actinonin virtual gels, especially in more congested areas of the gel, demonstrates the need to develop a search platform for virtual gel data to aid in analysis. Also, searching for small mass differences like the 28 Da shift caused by the failure to cleave the formyl group from the protein demonstrates the need to develop a simple and robust calibration method to improve mass accuracy of virtual gel results.

3.5 Conclusions

The experiments described here demonstrate the ability of the virtual 2D/gel MS technique to distinguish differences in proteomes caused by different stresses. In particular, the cold shock experiment displays the ease of identifying dramatic changes in protein expression caused by exposure of *E. coli* to reduced temperatures. These differences were verified by performing in-gel trypsin digestion on identical IPG strips to verify the identities of cold shock proteins CspA, CspB, and CspG. Furthermore, the results were confirmed by comparing them to past results obtained using traditional 2D-PAGE. The antibiotic experiment, on the other hand, highlighted ways to further develop and improve the virtual 2D/gel MS technique, including improving calibration and building a robust search platform to aid in processing data from future experiments. Both experimental and computational refinements are necessary to develop the platform to be competitive to conventional Bottom-Up proteomics strategies.

However, the virtual 2D/gel MS platform has the potential to profile changes in the proteome more rapidly and accurately than current methods.

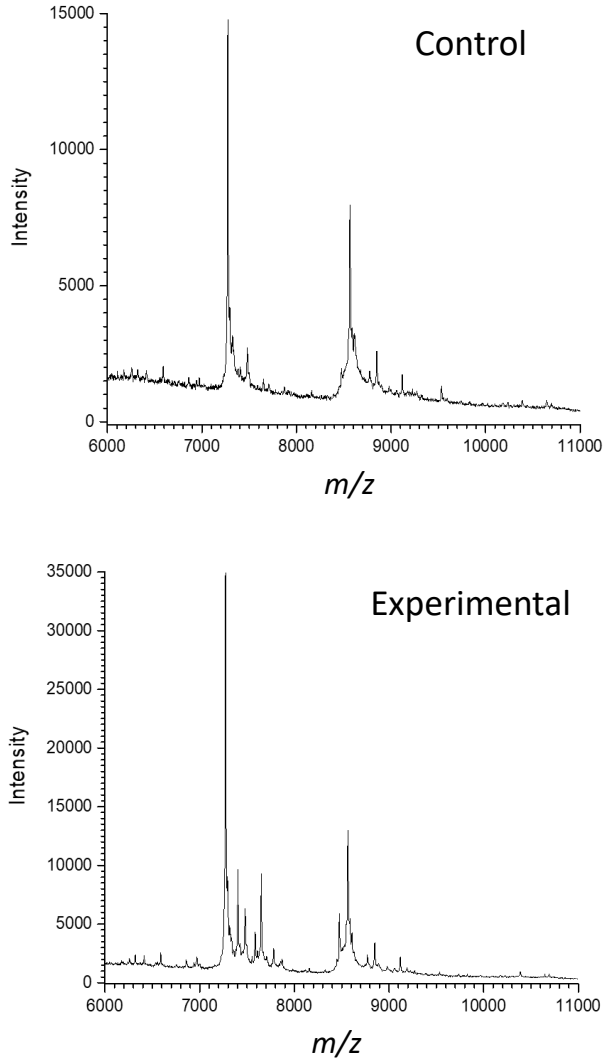


Figure 3.1: Intact protein profiles of *E. coli* grown in optimal (top) and cold shock (bottom) conditions. Differences can be observed between 7 and 8 kDa.

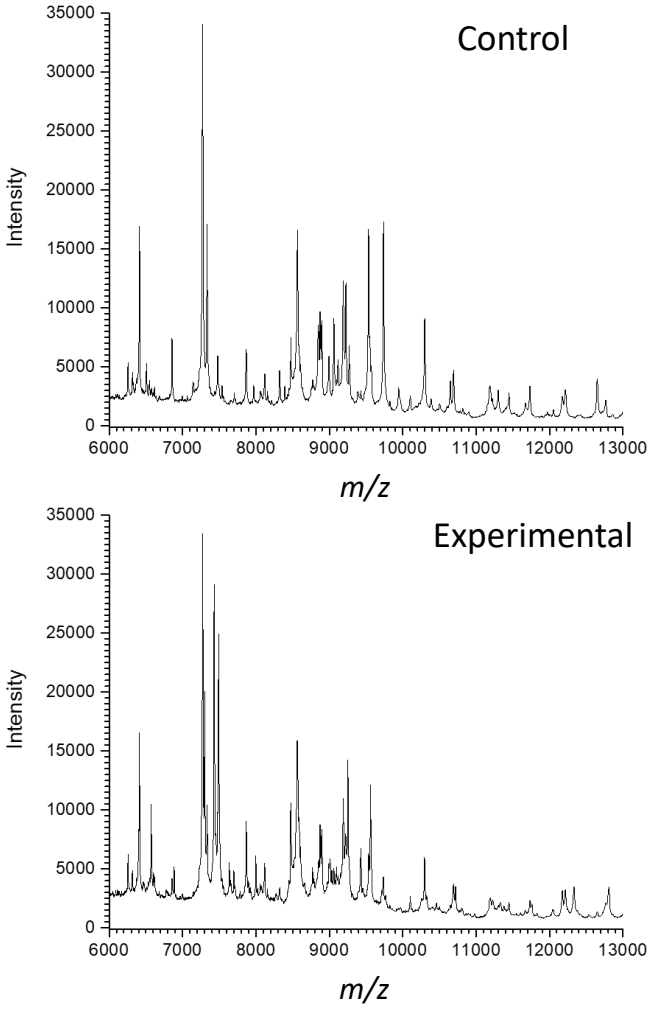


Figure 3.2: Intact protein profiles of *E. coli* grown in optimal conditions (top) and in the presence of actinonin (bottom). Subtle differences can be observed in the left spectrum with the additional peaks being observed, especially between 7 and 8 kDa.

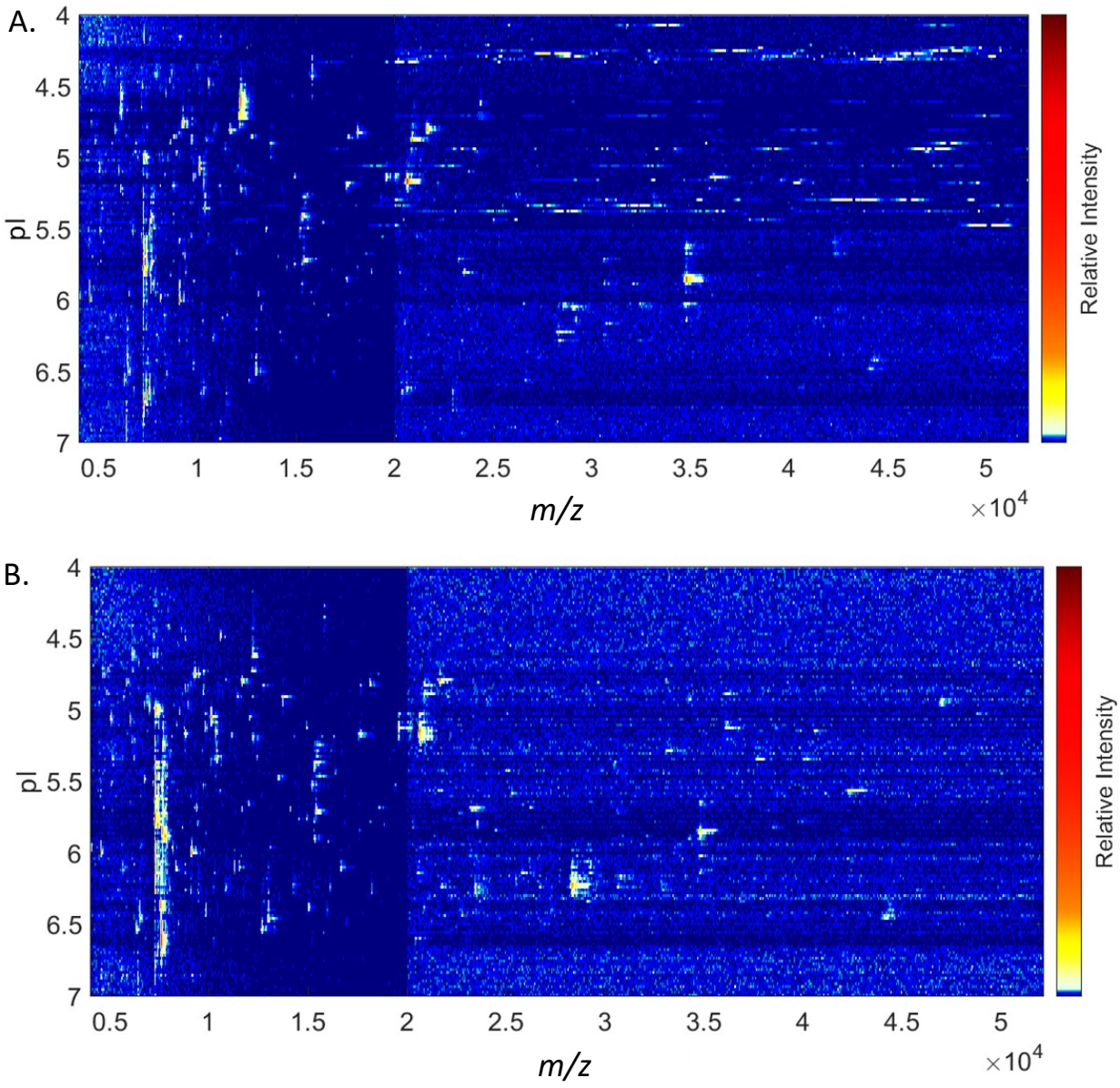


Figure 3.3: Heatmap produced from virtual 2D gel/ MS analysis of cell lysate obtained from *E. coli* grown in optimal conditions (A) and cold shock conditions (B). A major difference can be seen between 7 and 8 kDa from pI 5.5 - 6.5. Signal above 20 kDa has been amplified by a factor of 5.

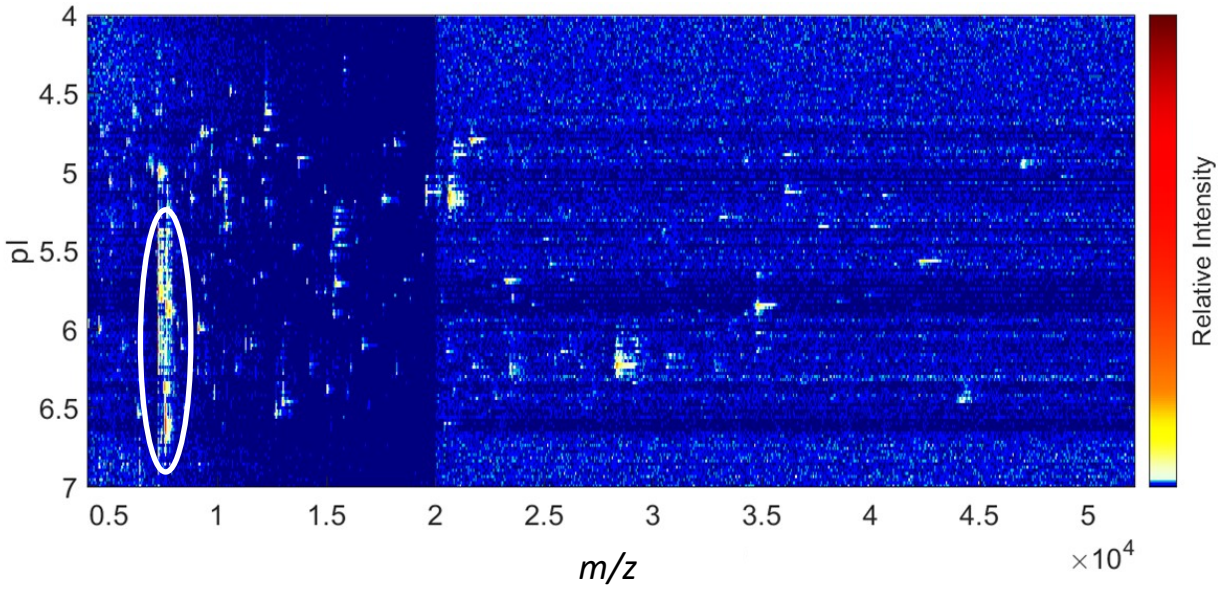


Figure 3.4: Heatmap produced from virtual 2D gel/ MS analysis of cell lysate obtained from *E. coli* grown in cold shock conditions with major difference attributed to CspA, CspB, and CspG circled. Signal above 20 kDa has been amplified by a factor of 5.

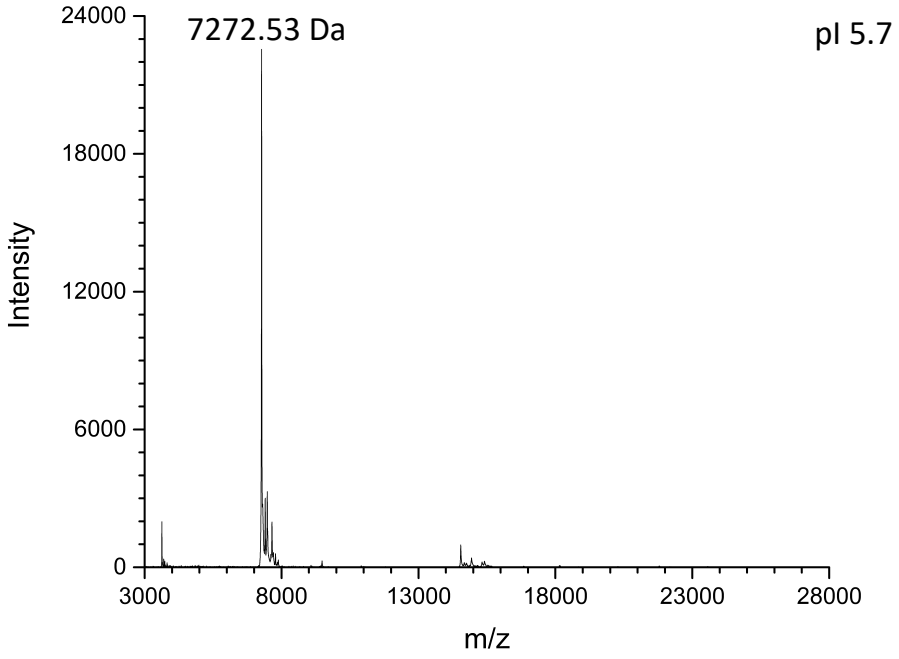


Figure 3.5: Mass spectrum of what appears to be CspA obtained from the circled region in the cold shock experimental gel heatmap (**Figure 3.4**) at a pI of 5.7.

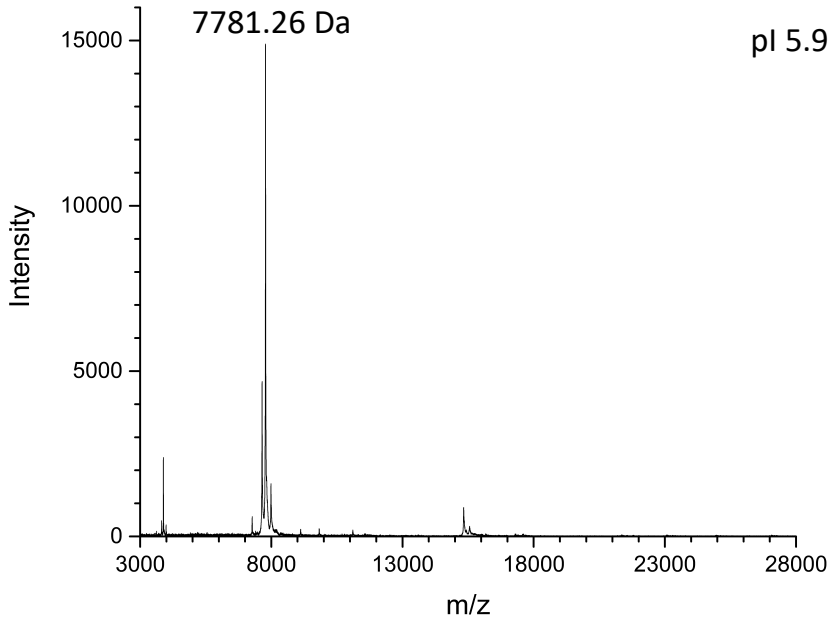


Figure 3.6: Mass spectrum of what appears to be CspG obtained from the circled region in the cold shock experimental gel heatmap (**Figure 3.4**) at a pI of 5.9.

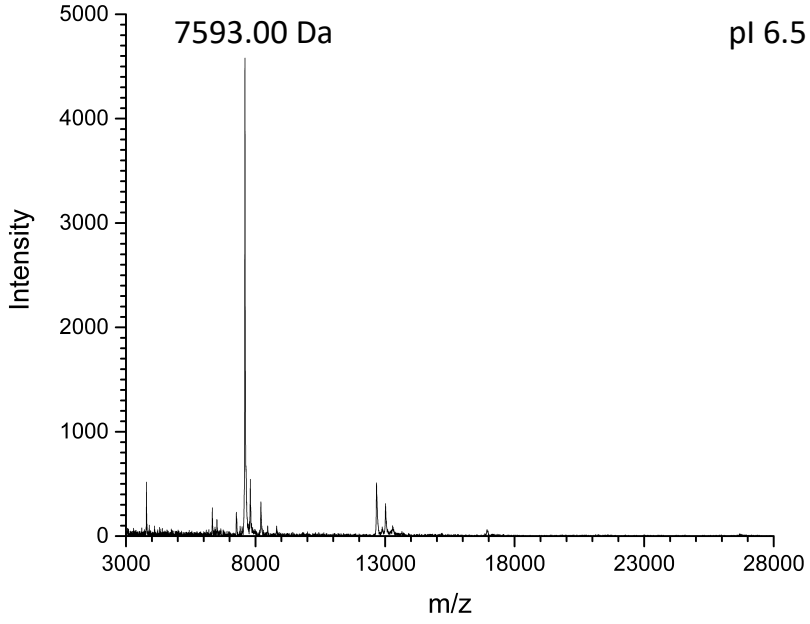


Figure 3.7: Mass spectrum of what appears to be CspB obtained from the circled region in the cold shock experimental gel heatmap (**Figure 3.4**) at a pI of 6.5.

Table 3.1: List of the gel sections obtained during the in-gel digest procedure and their approximate pI ranges for both the control and cold shock experimental IPG strips.

Experimental Gel			Control Gel		
Section #	**Length of Section (mm)**	**pI Range**	**Section #**	**Length of Section (mm)**	**pI Range**
1	7	4.00 - 4.12	1	5	4.00 - 4.08
2	5	4.12 - 4.20	2	6	4.08 - 4.18
3	5	4.20 - 4.29	3	6	4.18 - 4.28
4	6	4.29 - 4.39	4	6	4.28 - 4.38
5	6	4.39 - 4.49	5	6.5	4.38 - 4.89
6	6	4.49 - 4.59	6	6	4.89 - 4.59
7	5.5	4.59 - 4.68	7	7	4.59 - 4.70
8	6	4.68 - 4.78	8	6	4.70 - 4.80
9	6	4.78 - 4.88	9	6	4.80 - 4.90
10	6.5	4.88 - 4.99	10	6	4.90 - 5.00
11	5.5	4.99 - 5.08	11	7	5.00 - 5.12
12	5	5.08 - 5.17	12	6	5.12 - 5.22
13	6	5.17 - 5.27	13	6	5.22 - 5.32
14	6	5.27 - 5.37	14	6.5	5.32 - 5.43
15	6.5	5.37 - 5.48	15	5.5	5.43 - 5.52
16	6	5.48 - 5.58	16	6	5.52 - 5.62
17	6	5.58 - 5.68	17	5.5	5.62 - 5.71
18	6	5.68 - 5.78	18	6	5.71 - 5.81
19	6	5.78 - 5.88	19	6	5.81 - 5.91
20	6	5.88 - 5.98	20	6	5.91 - 6.01
21	6	5.98 - 6.08	21	5	6.01 - 6.09
22	6	6.08 - 6.18	22	6	6.09 - 6.19
23	6	6.18 - 6.29	23	6	6.19 - 6.29
24	6	6.29 - 6.39	24	6	6.29 - 6.39
25	6	6.39 - 6.49	25	6	6.39 - 6.49
26	6.5	6.49 - 6.60	26	6	6.49 - 6.59
27	6	6.60 - 6.70	27	5	6.59 - 6.67
28	6	6.70 - 6.80	28	8	6.67 - 6.80
29	6	6.80 - 6.90	29	6	6.80 - 6.90
30	6	6.90 - 7.00	30	6	6.90 - 7.00

Table 3.2: Sections of IPG strip that cold shock response related protein were found by searching in-gel digestion results for both experiment and control samples against a *E. coli* strain K-12 substrain DH10B database. Protein are separated based on level of induction during cold shock as reported by Inouye and coworkers. pI range of the sections can be found by referencing the previous table. Protein that did not fall within the mass range of the virtual 2D/gel MS analysis have been denoted with an asterisks.

Protein	Calculated MW (kDa)	Calculated pI	Experimental Sections	Control Sections
Greater than 10x induction				
Cold shock protein **CspA**	7.3	5.57	1, 2, 3, 12, 13, 17, 19, 20, 22, 28	19
Cold shock like protein **CspB**	7.7	6.53	1, 2, 3, 25, 27, 28	Not Found
Cold shock like protein **CspG**	7.8	5.64	2, 20, 21	20
ATP-dependent RNA helicase **DeaD***	70.4	8.76	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 28, 29, 30	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 29, 30
30S ribosome-binding factor **rbfA**	15	5.94	1, 2, 3, 9, 20, 25	2, 3, 19, 25
Transcription termination/ antitermination protein **NusA***	54.9	4.53	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14
Less than 10x induction				
Protein **RecA**	37.8	5.09	Not Found	Not Found
DNA gyrase subunit A **gyrA***	96.8	5.08	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 28, 29	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27
Translation initiation factor IF-2 **infB***	97.3	5.8	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 28, 29	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 29
DNA -binding protein **H-NS**	15.4	5.44	1, 2, 3, 10, 11, 13, 14, 15, 16, 17, 18, 19, 21, 22, 29	1, 2, 3, 27, 29, 30

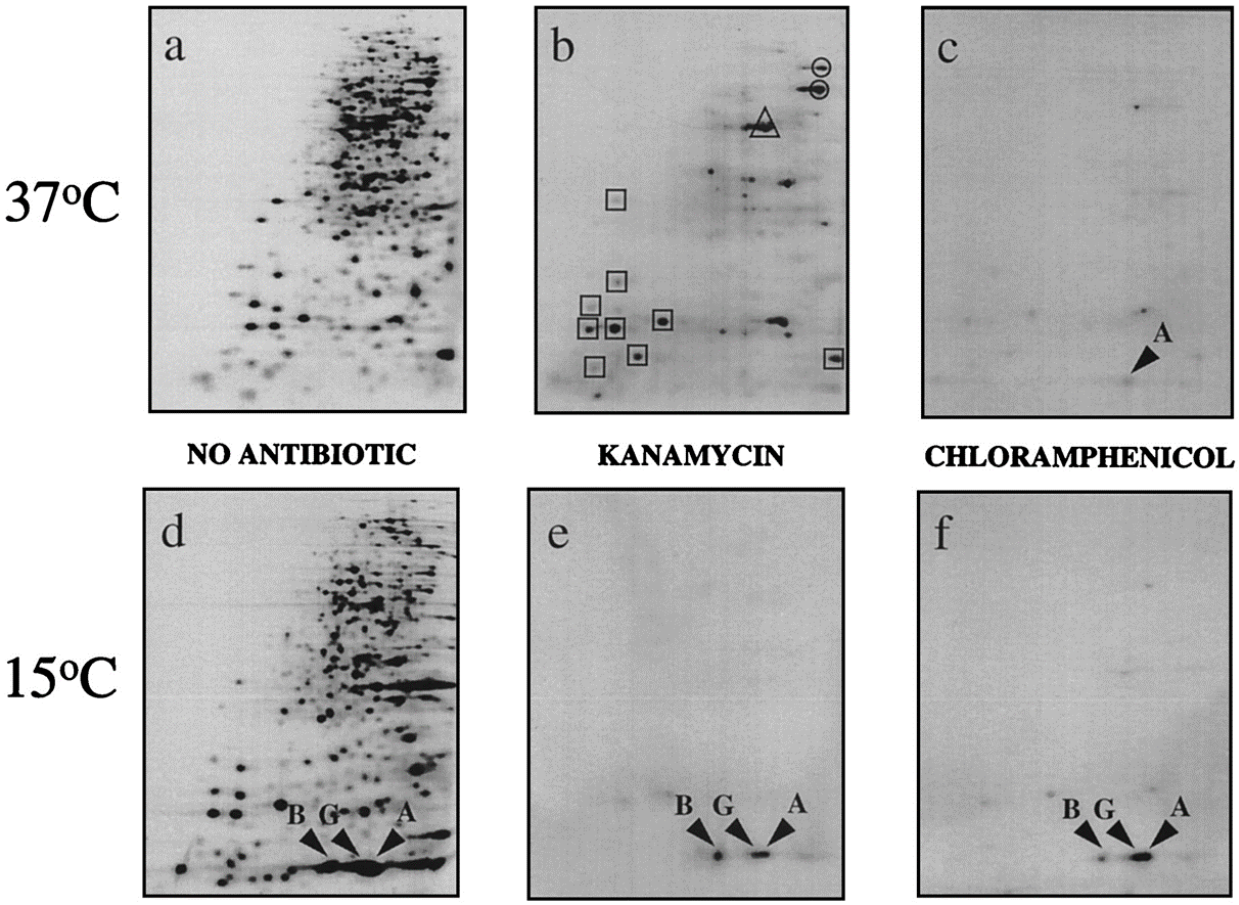


Figure 3.8: Protein expression pattens of *E. coli* SB221 grown in standard conditions at 37°C (a) and exposed to cold shock conditions at 15°C after reaching an OD_{600} of 0.4 (b). Protein were pulse labeled with 100 µCi of [trans-^{35}S]methionine before being separated by 2D-PAGE with pI range of 3—10 from left to right. The position of CspA, CspG, and CspB are denoted on the gel as A, G, and B, respectively. (Etchegaray et. al. (1999) Reproduced with permission from the American Society for Microbiology.)

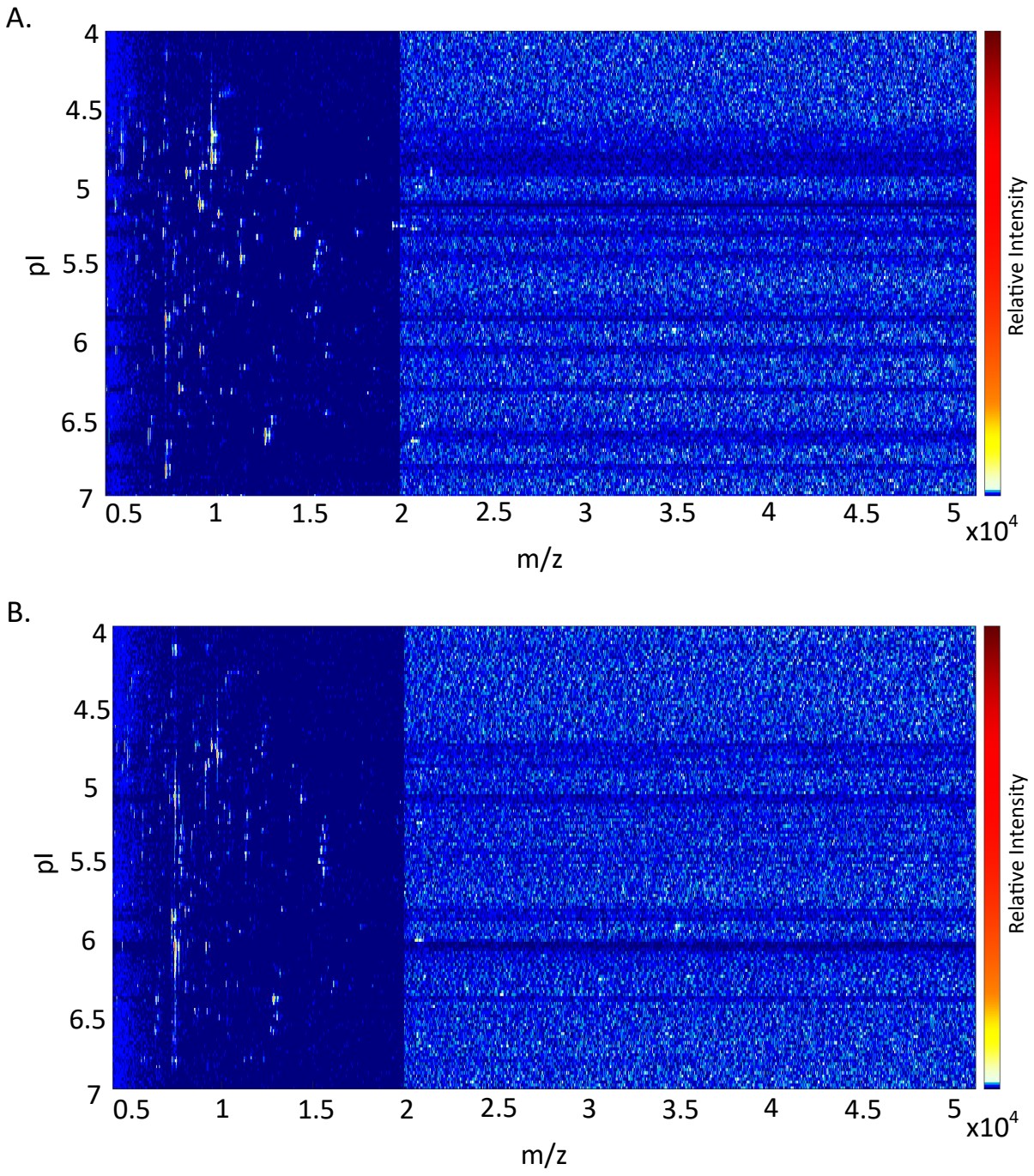


Figure 3.9: Heatmap produced from virtual 2D gel/ MS analysis of cell lysate obtained from *E. coli* grown in optimal conditions (A) and in the presence of the antibiotic actinonin (B). Some subtle difference can be noted, in particular there appears to be less spots in the experimental gel.

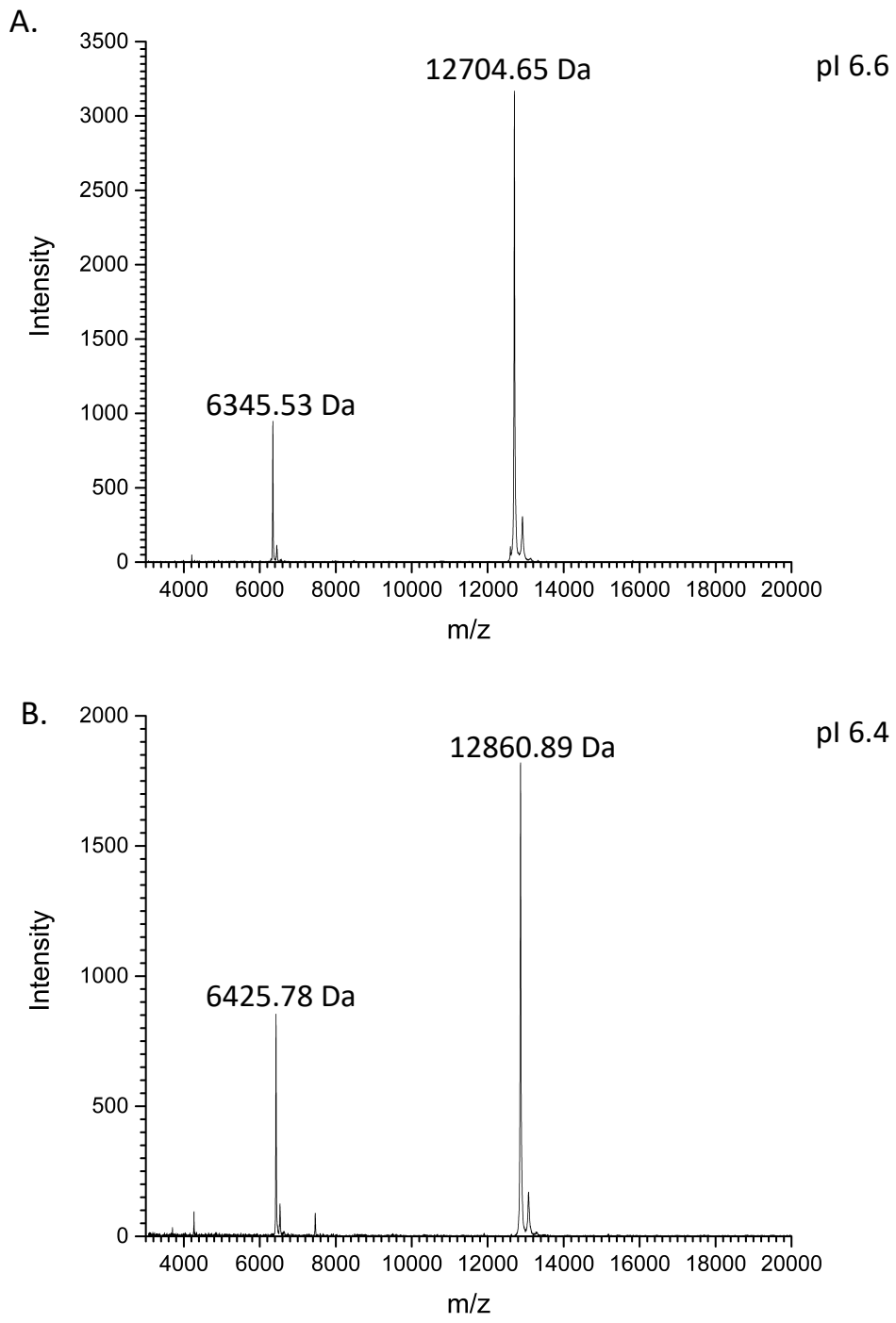


Figure 3.10: Mass spectra from virtual 2D gel/MS analysis of cell lysate from actinonin experiment. Protein at pI 6.6 with a mass of 12704.65 Da from the control cells can be seen in (A). A signal from the experimental gel at pI 6.4 and a mass difference of +156 Da can be seen in (B). This signal is suspected to be the same protein with the formyl group and initiator methionine still attached due to inhibition of PDF.

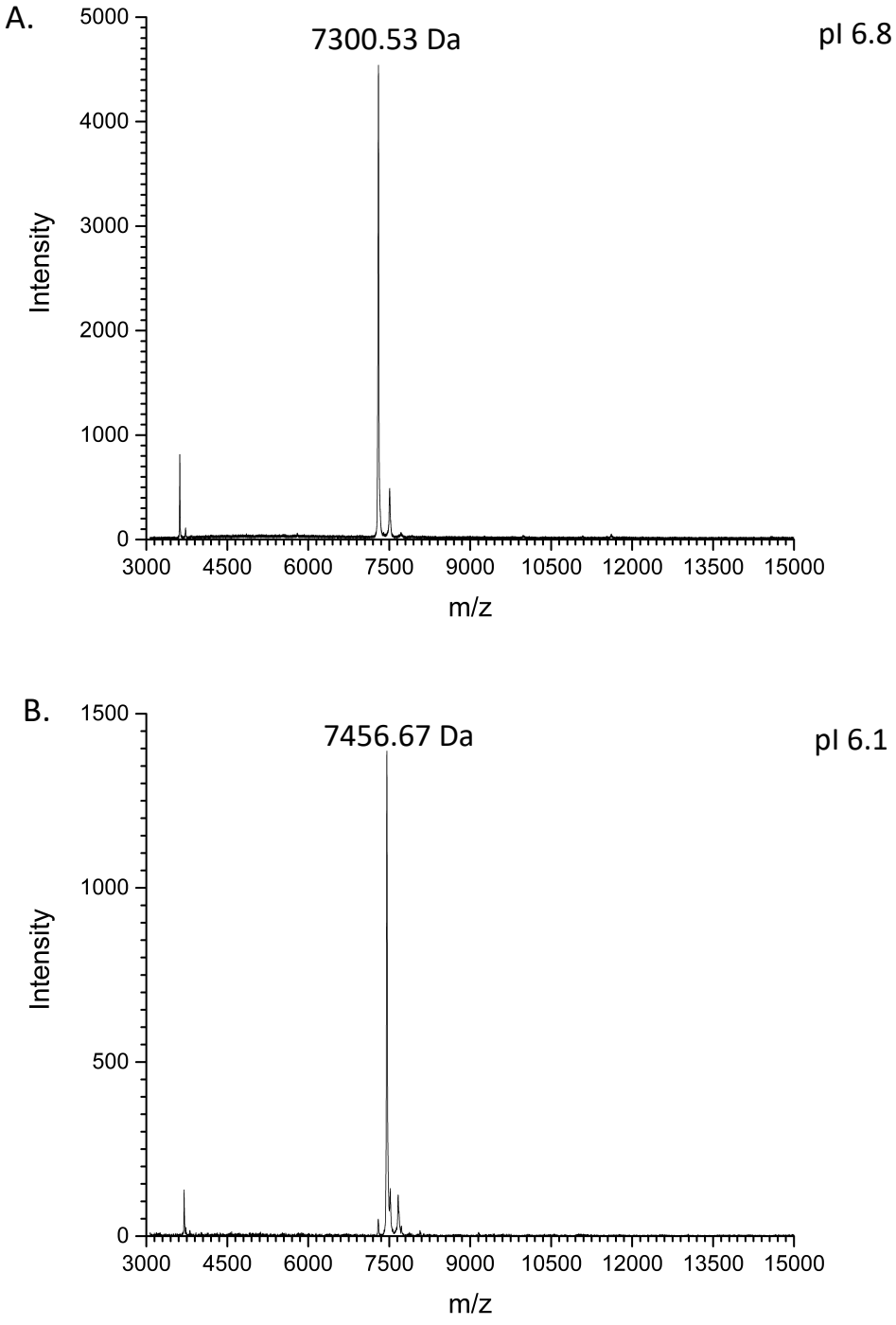


Figure 3.11: Mass spectra from virtual 2D gel/MS analysis of cell lysate from actinonin experiment. Protein at pI 6.8 with a mass of 7300.53 Da from the control cells can be seen in (A). A signal from the experimental gel at pI 6.1 and a mass difference of +156 Da can be seen in (B). This signal is suspected to be the same protein with the formyl group and initiator methionine still attached due to inhibition of PDF.

3.6 References

(1) Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science (80-.).* **2006**, *312* (5771), 212 LP – 217. https://doi.org/10.1126/science.1124619.

(2) Garcia, B. A. What Does the Future Hold for Top down Mass Spectrometry? *J. Am. Soc. Mass Spectrom.* **2010**, *21* (2), 193–202. https://doi.org/10.1016/j.jasms.2009.10.014.

(3) Wu, S.; Lourette, N. M.; Tolić, N.; Zhao, R.; Robinson, E. W.; Tolmachev, A. V.; Smith, R. D.; Paša-Tolić, L. An Integrated Top-down and Bottom-up Strategy for Broadly Characterizing Protein Isoforms and Modifications. *J. Proteome Res.* **2009**, *8* (3), 1347–1357. https://doi.org/10.1021/pr800720d.

(4) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. Mapping Intact Protein Isoforms in Discovery Mode Using Top-down Proteomics. *Nature* **2011**, *480* (7376), 254–258.

(5) Lohnes, K.; Quebbemann, N. R.; Liu, K.; Kobzeff, F.; Loo, J. A.; Ogorzalek Loo, R. R. Combining High-Throughput MALDI-TOF Mass Spectrometry and Isoelectric Focusing Gel Electrophoresis for Virtual 2D Gel-Based Proteomics. *Methods* **2016**, *104*, 163–169. https://doi.org/https://doi.org/10.1016/j.ymeth.2016.01.013.

(6) Ogorzalek Loo, R. R.; Hayes, R.; Yang, Y.; Hung, F.; Ramachandran, P.; Kim, N.; Gunsalus, R.; Loo, J. a. Top-down, Bottom-up, and Side-to-Side Proteomics with Virtual 2-D Gels. *Int. J. Mass Spectrom.* **2005**, *240* (3 SPEC. ISS.), 317–325. https://doi.org/10.1016/j.ijms.2004.10.013.

(7) Jones, P. G.; VanBogelen, R. A.; Neidhardt, F. C. Induction of Proteins in Response to Low Temperature in Escherichia Coli. *J. Bacteriol.* **1987**, *169* (5), 2092–2095. https://doi.org/10.1128/jb.169.5.2092-2095.1987.

(8) Goldstein, J.; Pollitt, N. S.; Inouye, M. Major Cold Shock Protein of Escherichia Coli. *Proc. Natl. Acad. Sci.* **1990**, *87* (1), 283 LP – 287. https://doi.org/10.1073/pnas.87.1.283.

(9) Lee, S. J.; Xie, A.; Jiang, W.; Etchegaray, J.-P.; Jones, P. G.; Inouye, M. Family of the Major Cold-Shock Protein, CspA (CS7.4), of Escherichia Coli, Whose Members Show a High Sequence Similarity with the Eukaryotic Y-Box Binding Proteins. *Mol. Microbiol.* **1994**, *11* (5), 833–839. https://doi.org/https://doi.org/10.1111/j.1365-2958.1994.tb00361.x.

(10) Wang, N.; Yamanaka, K.; Inouye, M. CspI, the Ninth Member of the CspA Family of *Escherichia coli* is Induced upon Cold Shock. *J. Bacteriol.* **1999**, *181* (5), 1603 LP – 1609. https://doi.org/10.1128/JB.181.5.1603-1609.1999.

(11) Jones, P. G.; Mitta, M.; Kim, Y.; Jiang, W.; Inouye, M. Cold Shock Induces a Major Ribosomal-Associated Protein That Unwinds Double-Stranded RNA in Escherichia Coli. *Proc. Natl. Acad. Sci.* **1996**, *93* (1), 76 LP – 80. https://doi.org/10.1073/pnas.93.1.76.

(12) Jones, P. G.; Inouye, M. RbfA, a 30S Ribosomal Binding Factor, Is a Cold-Shock Protein Whose Absence Triggers the Cold-Shock Response. *Mol. Microbiol.* **1996**, *21* (6), 1207–1218. https://doi.org/https://doi.org/10.1111/j.1365-2958.1996.tb02582.x.

(13) Thieringer, H. A.; Jones, P. G.; Inouye, M. Cold Shock and Adaptation. *BioEssays* **1998**, *20* (1), 49–57. https://doi.org/https://doi.org/10.1002/(SICI)1521-1878(199801)20:1<49::AID-BIES8>3.0.CO;2-N.

(14) GORDON, J. J.; KELLY, B. K.; MILLER, G. A. Actinonin: An Antibiotic Substance Produced by an Actinomycete. *Nature* **1962**, *195* (4842), 701–702. https://doi.org/10.1038/195701b0.

(15) Chen, D. Z.; Patel, D. V; Hackbarth, C. J.; Wang, W.; Dreyer, G.; Young, D. C.; Margolis, P. S.; Wu, C.; Ni, Z.-J.; Trias, J.; White, R. J.; Yuan, Z. Actinonin, a Naturally Occurring Antibacterial Agent, Is a Potent Deformylase Inhibitor. *Biochemistry* **2000**, *39* (6), 1256–1262. https://doi.org/10.1021/bi992245y.

(16) Adams, J. M. On the Release of the Formyl Group from Nascent Protein. *J. Mol. Biol.* **1968**, *33* (3), 571–589. https://doi.org/https://doi.org/10.1016/0022-2836(68)90307-0.

(17) Hao, B.; Gong, W.; Rajagopalan, P. T. R.; Zhou, Y.; Pei, D.; Chan, M. K. Structural Basis for the Design of Antibiotics Targeting Peptide Deformylase. *Biochemistry* **1999**, *38* (15), 4712–4719. https://doi.org/10.1021/bi982594c.

(18) Etchegaray, J.-P.; Inouye, M. CspA, CspB, and CspG, Major Cold Shock Proteins of *Escherichia coli*, Are Induced at Low Temperature under Conditions That Completely Block Protein Synthesis. *J. Bacteriol.* **1999**, *181* (6), 1827 LP – 1830. https://doi.org/10.1128/JB.181.6.1827-1830.1999.

CHAPTER 4

Application of Collision Induced Unfolding to Probe the Structures and Stabilities of Protein-Ligand Complexes

4.1 Abstract

Parkinson's disease, dementia with Lewy bodies, multiple system atrophy, and other synucleinopathies are all neurodegenerative disorders thought to be caused by the aggregation and accumulation of α-synuclein (αSyn) in the brain. To help combat these debilitating conditions research has focused on finding the cause of this aggregation and ways to prevent it from occurring. Previous studies have shown that the binding of CLR01 and other molecules to αSyn reduce its propensity for aggregation, while the binding of divalent metals promotes further aggregation of αSyn. Here, using collision induced unfolding, we probe the effects that the binding of CLR01, manganese, cobalt, and copper have on the gas phase structural stability of αSyn. Based on our results CLR01, Mn(II), and Cu(II) all stabilize the gas phase structure of the protein, while Co(II) has a destabilizing effect. These results combined with results from other biophysical techniques can help unlock the secrets of αSyn aggregation and provide information on how to treat and prevent these devastating disorders.

4.2 Introduction

Proteins have dynamic structures and form interactions with other molecules to perform biological functions. Improper folding, however, can cause proteins to lose function, or worse, become detrimental to the host organism. Therefore, it is essential to study protein structure and understand how changes in protein structure and folding can impact human health. Parkinson's disease (PD) is a neurodegenerative disorder characterized by tremors, slowed movement, impaired posture, and other neurological symptoms.[1] PD is thought to be caused by the accumulation of α-synuclein (αSyn) protein in harmful structures called Lewy bodies deposited in the brain.[2,3] In fact, six point mutations in the αSyn gene, *SNCA*, as well as gene duplication and triplication are associated with the genetic form of early onset PD.[4-7] In addition to PD, accumulation of αSyn is characteristic of other neurological diseases termed synucleinopathies, including dementia with Lewy bodies and multiple system atrophy.[8,9]

αSyn is a small protein consisting of 140 amino acids, typically found in the brain at presynaptic terminals.[10] While the purpose of αSyn is not well known, some studies report the protein promoting SNARE-complex assembly, mediating the release of dopamine, and binding to lipids.[11-13] Studies have shown that αSyn is an intrinsically disordered protein (IDP).[14] This is thought to be caused by the high number of acidic residues in its polypeptide sequence.[15] This lack of structure, i.e., high degree of disorder, and its propensity to aggregate has made it difficult to obtain an atomic resolution structure of the intact protein until recently;[16,17] much effort has been made to obtain the structure of the fibrils formed from segments of the protein.[18,19] Also, structural analysis of membrane bound αSyn and aggregated αSyn has been performed using NMR, electron paramagnetic resonance (EPR), and X-ray crystallography.[20-22]

The sequence of αSyn, **Figure 4.1**, can be broken down into three regions: a positively charged N-terminal region, a hydrophobic region in the center that has a propensity for aggregation, and a very acidic C-terminus. The central portion is of particular importance in the formation of Lewy bodies and is termed the non-amyloid component (NAC) region of the protein.[23,24] The addition of certain ligands to the protein has been shown to decrease the toxicity of αSyn. These ligands include dopamine, epigallocatechin gallate (EGCG), curcumin, rifampicin, scylloinositol, and CLR01.[25-29] These small molecule ligands are thought to alter the structure of αSyn, driving them away from their toxic form. A focus of this study, CLR01, is a lysine molecular tweezer, **Figure 4.2**.[30] While CLR01 can bind to multiple sites on αSyn, mass spectrometry (MS) studies have shown that it primarily binds to the N-terminus region.[31] It is thought that CLR01 binding to αSyn increases the protein's reconfiguration rate preventing it from forming a dimer and high order aggregates. The effectiveness of CLR01 in mediating αSyn toxicity has been demonstrated in both cell cultures and zebrafish embryos.[29]

Oppositely, the binding of divalent and trivalent heavy metals like aluminum, copper, cobalt, manganese, cadmium, and iron have been shown to promote the aggregation of αSyn.[32-34] This is thought to be caused by the metals promoting αSyn to take a more compact form that is prone to aggregation. Copper has been shown to bind to the N-terminus and His-50 with micromolar affinity.[35] Recently our group published a top-down mass spectrometry study on αSyn showing the primary and secondary binding sites of both cobalt (II) and manganese (II) to be ^{119}DPDNEAYE^{126} and ^{132}GYQDY^{136}, respectively.[36] Both metals bind in solution with relatively low millimolar affinity to αSyn.

In this study, we use a gas-phase method, Collision Induced Unfolding (CIU), to probe changes in protein stability of αSyn upon binding to CLR01 and various divalent metals. CIU has previously been shown effective for probing the stability of proteins, protein complexes, and protein-ligand interactions.[37-39] CIU is a technique that combines mass spectrometry with ion mobility (IM), two orthogonal gas-phase techniques in what can be thought of as a gas phase differential scanning calorimetry experiment.[40] By using fine-tuned solvent and MS source conditions it has been shown that proteins and protein-ligand complexes can be transmitted into the mass spectrometer such that their native solution-phase structure are largely maintained.[41,42] Once inside the mass spectrometer, the protein molecules of interest can be isolated using a quadrupole and selectively passed to a cell containing inert gas molecules, in this case, argon. Here, the protein molecules undergo collisional heating with the gas molecules. The heated proteins are then transmitted to the IM cell where it is separated based on its orientationally averaged size and charge. At this point, the mass spectrometer acting as the detector confirms that the sample remains intact and that no collision-induced dissociation (CID) products have been produced, while also determining how long it took the ions to transverse the IM cell. This ion mobility "time" is related to the size and structure of the analyte molecule; in the case of proteins, more compact or folded proteins travel faster through the IM cell than elongated or unfolded proteins. Running sequential experiments with increasing collision energy produces the CIU "fingerprint" of the sample. This fingerprint informs us as to how the protein unfolds as it is heated and the collision cross-section (CCS) of each of the intermediate structures is related to the size of the molecule.[43-45]

As an example, CIU fingerprints obtained for the +9, + 10, and +11 charge states of 29 kDa human carbonic anhydrase I (HCA-I), can be seen in **Figure 4.3**. The +11 charge state of HCA-I transitions through 4 major conformations, indicated by dark red on the heatmap, as it unfolds from its compact form, starting around a collision energy of 15 V, with a relatively small CCS value to its extended form, appearing at a collision energy of 27 V, with a larger CCS value. The +9 charge state of the protein, on the other hand, exhibits 2 major conformations and one minor conformation with an intermediate CCS value. Minor conformations are ones that do not reach the highest normalized intensity during the experiment. Thus, the +9 charged HCA-I molecule is more stable to collisional dissociation compared to the 11+ charged molecule.

Comparing CIU fingerprints of proteins with and without a bound ligand can uncover additional information, such as how ligand-binding affects a protein's stability.[39] A ligand is indicated to stabilize the protein if higher collision energy is required to transition the protein to its extended confirmations. This can be visually seen as a shift to the right when comparing CIU fingerprints. Conversely, if lower collision energy is required, it is interpreted that ligand binding destabilizes the protein. Visually this would be indicated by a shift to the left on the CIU fingerprint. We used CIU to probe the relative stability of proteins cytochrome c and αSyn upon binding to the CLR01 molecular tweezer.

4.3 Method

4.3.1 Sample Preparation

αSyn was acquired from rPeptide (Watkinsville, GA). The αSyn was resuspended in water and desalted using micro bio-spin chromatography columns (Biorad). The stock solution of

protein with a concentration of 70μM was stored at -80℃. Equine heart cytochrome c was acquired from Sigma Aldrich (C7752). The protein was resuspended to a concentration of 160 μM in 200 mM ammonium acetate at pH 6.8 and stored at -80℃. CLR01 was obtained from the Bitan Lab at UCLA at a concentration of 145μM. Cobalt(II) acetate, copper(II) acetate monohydrate, and manganese(II) acetate were all acquired from Sigma (St. Louis, MO); each was resuspended to a concentration of 10mM in 20mM ammonium acetate at pH 6.8.

4.3.1.1 CLR01 Binding Experiments

Shortly before the mass spectrometry experiments, αSyn was thawed on ice and diluted to a concentration of 10μM in 20mM ammonium acetate. The protein solution was then mixed with CLR01 to a final protein:CLR01 ratio of 1:1. Similarly, cytochrome c was also thawed and diluted to a final concentration of 10μM in 200mM ammonium acetate and then mixed with CLR01 in a 1:1 ratio.

4.3.1.2 Metal Binding Experiments

αSyn was thawed and diluted to a concentration of 20μM in 20mM ammonium acetate. The protein was then mixed in a 1:10 ratio with Mn(II) acetate, 1:8 ratio with Co(II) acetate, and 1:5 ratio with Cu(II) acetate.

4.3.2 Mass Spectrometry Data Acquisition

Data was collected using a Synapt G2-Si Q-TOF mass spectrometry/ion mobility system (Waters Corp, Manchester, U.K.) equipped with a nanospray source. Samples were sprayed with metal-coated borosilicate capillaries (Au/Pd-coated, 1μm i.d.; Thermo Fisher Scientific)

with flow rates between 10-40 nL/min. The instrument capillary voltage was set between 750V and 1000V. The sampling cone and source offset were both set to 10V to minimize the perturbation of protein structure.[46] Once inside the instrument, charged protein and protein complex molecules of interest were isolated in the quadrupole region of the mass spectrometer and subsequently excited in the trap region with the trap bias set to 37V. Trap collision energy (CE) was increased in increments of 2 volt-steps over a range of 4V – 42V or 4 – 22V, with 180 scans averaged for each step. The helium cell flow rate was set to 200mL/min, while the IMS gas flow rate was set to 90 mL/min. The IM cell wave height was set to 40V with wave velocity set to 650 m/s.

After the data acquisition, CCS data was calibrated using a linear fit calibration procedure with denatured bovine erythrocytes ubiquitin, equine heart cytochrome c, and equine heart apo myoglobin acting as standards.[47,48] Calibrated data was exported using TWIMExtract, a program developed by Ruotolo and coworkers that allows for automated extraction of ion mobility data from Waters' proprietary .raw file format.[49] CIU fingerprints were created using the CIUSuite 2.1 software package, also developed by Ruotolo and coworkers.[50] Along with producing CIU fingerprints the software also performed feature detection and determined the midpoint between conformer transitions by fitting a sigmoid function to features that have been detected in what is referred to as CIU50.

4.4 Results and Discussion

4.4.1 CIU of Cytochrome C-CLR01 Complexes

Initial CIU experiments were performed on 12 kDa equine heart cytochrome c to validate our results with data reported by other labs.[44] Upon analysis of cytochrome c using native electrospray ionization-MS, **Figure 4.4.A**, the +7 and +6 charge states were identified as the most abundant, with the +7 charge state being the dominant form. Upon binding of CLR01 to cytochrome c, sizeable peaks representing cytochrome c bound to one molecule of CLR01 are observed in the mass spectrum, **Figure 4.4.B**, while the ratio of the +7 to +6 charge states remains unchanged.

CIU experiments were performed on both the +6 and +7 charge states of the protein. Upon isolation of each charge state, the CE was ramped from 4V to a final energy of 42V in 2V intervals. Through the experiment, two confirmations of the +6-cytochrome c are observed, **Figure 4.5.A-C**. The compact form has a CCS of 15.41 nm^2. An extended form is also observed with a CCS of 17.11 nm^2. The protein makes a transition from the compact form to extended form at a CE of 15.5V based on CIU50 fitting. The +7-charge state of the protein provides an even more interesting CIU fingerprint with 3 distinct forms, **Figure 4.6.A-C**. Initially, the protein starts in a compact form with a CCS averaging 17.26 nm^2. This form of the protein can be seen becoming less compact as the CE is increased from 4V to about 12V. At this energy, the protein unfolds to a distinct form that has a CCS of 20.68 nm^2. This form of the protein persists in a narrow range of energies, with the CIU50 transitioning to an even less compact structure at 15.7V, at which point the protein has a CCS of 22.48 nm^2. Results from both the +6 and +7

charge state of the cytochrome c match well with results obtained previously by Ruotolo and coworkers.[44]

Identical CIU experiments were performed on equine heart cytochrome c bound to CLR01. The +6-charge state of the protein complex produced interesting results with the same compact form with a CCS of 15.87 nm^2 being observed at low CE, **Figure 4.5.D-F**. Upon increasing the CE to 27.5V, the more expanded form of the complex with a CCS of 17.72 nm^2 is observed, however, this population appears to decrease as the CE is increased further. Unlike the unbound cytochrome c, the more compact population of the protein never completely disappears. The increase in CIU50 from 14.7V to 27.5V to observe the expanded form of the protein is attributed to the binding of CLR01 stabilizing the protein. This stabilization is also seen as the more compact form of the protein is observed throughout the whole range of CE investigated.

The CIU fingerprint of the +7-charge state of cytochrome c was also significantly altered with the addition of one CLR01 molecule, **Figure 4.6.D-F**. At a low CE of 4V, the same compact form is observed with a CCS of about 16.72 nm^2. The protein then expands to a form with a CCS of 18.16 nm^2 when a CE of about 19V is applied. At this point, the signal intensity decreases to the level of noise as the protein deteriorates, **Figure 4.7**. It appears that the CLR01 initially stabilizes the protein, however, the complex is not able to reach the expanded form that is observed when the protein is by itself. This may be because CLR01 binding disrupts an interaction that is present in the expanded form that would typically hold the protein together. Overall, the contradictory results from the +6 and +7 charge state of the protein is interesting as it appears

CLR01 stabilizes the +6 charge state while having the opposite effect on the +7 charge state of the protein.

4.4.2 CIU of α-Synuclein with CLR01

After initial results were obtained for cytochrome c bound to CLR01, the focus shifted to the IDP αSyn. Upon analysis with native electrospray ionization-MS in 20 mM ammonium acetate, **Figure 4.8.A**, a range of charge states from +16 to +7 are easily observed, with the +10 charge state being dominant. Next, CLR01 was added to αSyn in a 1:1 ratio. Native electrospray ionization-MS was performed on the complex and revealed a similar charge state distribution that was displayed previously by unbound αSyn with the addition of peaks corresponding to αSyn bound to one CLR01 for each charge state, **Figure 4.8.B**. CIU experiments were carried out on the +7, +10, +11, and +12 charge states of the complex.

Upon isolation of the +12 charge state for apo-αSyn with the lowest CE of 4V, two forms of the protein are seen, **Figure 4.9.A-B**. The more compact form of the protein had a CCS of approximately 28 nm^2 and the expanded form has a CCS of 32.08 nm^2. The expanded form of the protein is dominant throughout the experiment, with the compact form, which has a low relative abundance, completely disappearing at about 10V. The CIU experiment of the +12-charge state of the αSyn-CLR01 complex reveals two forms of the protein, a compact form with a CCS of 28.53 nm^2 and an expanded form with a CCS of 32.70 nm^2, **Figure 4.9.C-D**. Initially, the compact form of the complex is dominant. As CE is increased a shift occurs at 5.7V to the expanded form being dominant. The addition of CLR01 to the +12-charge state of αSyn appears

to stabilize the protein with the compact form to the protein now being dominant until a CE of 5.7V, whereas it was not dominant in the unbound protein CIU result.

CIU experiments on the +11 charge state of αSyn also reveal two distinct conformations, **Figure 4.10.A-C**. A short-lived compact state that has a CCS of 26.58 nm^2 and an expanded state with a CCS value of 30.54 nm^2. Initially, the compact state is dominant, however this shifts at a CE of 5.2V. The +11-charge state of the αSyn-CLR01 complex showed similar results having a compact state with a CCS of 26.29 nm^2 being dominant initially and shifting to an expanded state with a CCS of 31.10 nm^2 at 8.7V, **Figure 4.10.D-F**. The addition of CLR01 to the protein stabilized the compact conformation of the complex with a shift in CIU50 of 3.5V.

The +10 charge state of the apo-protein has three forms, **Figure 4.11.A-C**. The most compact form has a CCS value of 25.18 nm^2. There is also a minor form that has a CCS of about 28 nm^2. Finally, an extended form can be seen that has a CCS of 30.07 nm^2. The protein initially starts in the compact form, as the CE is increased the intermediate conformation is briefly observed but never becomes dominant. The protein transitions from the compact to the expanded state at a CE of 7.5V. In comparison, the +10-charge state of the complex exhibited 2 dominant features with CCS values of 24.41 nm^2 and 30.84 nm^2 and a minor intermediate feature with a CCS of about 29 nm^2, **Figure 4.11.D-F**. The protein complex shifts from the compact state to an expanded state at 13.5V. The addition of CLR01 to the protein once again stabilizes the protein with a shift in CIU50 of 6.0V.

The +7 charge state of the apo-protein exists in 4 different forms, **Figure 4.12.A-C**. The compact form with a CCS of 17.97 nm^2 is dominant at low CE. The protein then expands to a

CCS of 19.05 nm^2 at 12.7V. At a CE of 16.2V, the protein then transitions from this state to an expanded form that has a CCS of 22.47 nm^2. During this last transition, the protein has a minor conformation with a CCS of about 21 nm^2 that is briefly exhibited. Finally, the +7-charge state of the complex displayed 3 forms, **Figure 4.12.D-F**. A compact form that has a CCS of 18.15 nm^2 which expands to a form with a CCS of 19.59 nm^2 at 16.1V. The complex then expands again at 20.8V to an expanded form with a CCS of 22.83 nm^2. All forms of the +7 charge state protein are stabilized by the addition CLR01, with the first transition having a shift in CIU50 of 3.4V and the second transition having a shift of 4.6V. Overall, all charge states of αSyn analyzed show noticeable stabilization upon binding of CLR01. The greatest shift in CIU50 is seen in the +10 charge state where a shift of 6V is observed.

4.4.3 CIU of αSyn Bound to Cobalt(II) or Manganese(II)

αSyn was bound to cobalt(II) or manganese(II), both of which bind to αSyn with millimolar affinity in solution. When analyzed with native electrospray ionization-MS the protein bound to Co(II), **Figure 4.13.A**, and Mn(II), **Figure 4.14.A**; both protein-metal complexes showed the same charge state distribution that was observed for apo-αSyn, **Figure 4.8.A**. Upon closer examination, the peak corresponding to the complex with one metal bound for both Co(II), **Figure 4.13.B**, and Mn(II), **Figure 4.14.B**, are about 50-75% the intensity of the unbound protein peak with a protein:metal ratio of 1:8 and 1:10 in solution, respectively. This is consistent for the millimolar binding affinity of these metals to αSyn.

Next, CIU experiments were performed on αSyn bound to either Co(II) or Mn(II) for charge states +7 through +14. Charge states +12, +13, and +14 yielded little information due to

exhibiting only one conformation and were left out of the further analysis. Both αSyn with Mn(II), **Figure 4.15.D-F,** and Co(II), **Figure 4.15.G-H**, revealed two confirmations upon analysis of the +11 charge state. A compact form with a CCS of 26.58 nm^2 and an expanded form with a CCS of 30.54 nm^2. The protein-bound to Mn(II) initially had the compact form as dominant. The complex then transitioned to the expanded form being dominant at a CE of 6.7V. αSyn bound to Co(II) on the other hand has the expanded form as dominant for the whole experiment, with only a minor portion of the complex existing in the compact form at low CE. These results indicate that binding Mn(II) to the protein stabilizes the protein, shifting the CIU50 by 1.5V. On the other hand, Co(II) binding to the protein has a slightly destabilizing effect.

Upon analysis of the +10 charge state of the complexes, two major forms and two minor forms are observed when Mn(II), **Figure 4.16.D-F**, or Co(II), **Figure 4.16.G-I**, bind to αSyn. A minor compact form with a CCS of about 24 nm^2 is seen at low CE. A major compact form is seen with a CCS of 25.18 nm^2. Another minor feature with a CCS of 28 nm^2 is observed briefly between the major compact and major expanded form at a CE of 8-10V for the complex containing Mn(II) and 6-8V for the complex containing Co(II). Finally, a dominant expanded form with a CCS of 30.07 nm^2 exists. The complex containing Mn(II) has a transition from the major compact to the major expanded feature at a CE of 8.5V, while the Co(II) complex makes the same transition at a CE of 6.5V. The binding of Mn(II) to the protein increases the CIU50 by 1.0V indicating the metal stabilizes the protein slightly. The binding of Co(II), on the other hand, again slightly destabilizes the protein and decreases CIU50 by 1.0V.

Two forms of the +9 charge state complexes are observed. A compact form that is dominant at low CE has a CCS of 23.35 nm^2 and an expanded form with a CCS of 30.07 nm^2 is observed

when the αSyn is bound to Mn(II), **Figure 4.17.D-F**, or 30.50 nm^2 when αSyn is bound to Co(II), **Figure 4.17.G-I**, is observed at higher CE. The complex containing Mn(II) transitions from the compact to an expanded form at a CE of 21.0V, while the Co(II) complex makes the same transition at 17.0V. As seen with the +10 charge state of the complex, the +9 charge state is also stabilized with the addition of Mn(II) with a 2.0V increase in CIU50, while the Co(II) complex destabilizes the protein with a 2.0V decrease in CIU50.

The +8 charge state of the complexes revealed 4 forms of the Mn(II) complex, **Figure 4.18.D-F**, and 3 forms of the Co(II) complex, **Figure 4.18.G-I,** similar to the 3 forms observed with apo-αSyn. The Co(II) complex has 2 major forms with CCS values of 19.73 nm^2 and 22.81 nm^2 and one minor form with a CCS of approximately 21 nm^2. The Mn(II) complex contains these forms with the addition of a compact form existing at low CE with a CCS of 18.90 nm^2. In the Co(II) complex the protein transitions from its major compact to a major extended form at 8.5V. The same transition for the Mn(II) complex is observed at a higher CE of 10.5V. Compared to the apo-protein the CIU50 of this transition has increased by 1.0V for the Mn(II) complex and decreased by 1.0V for the Co(II) complex. The Mn(II) complex also stabilizes the protein enough that an additional compact conformation is now distinguishable.

Finally, the +7 charge state of both complexes has two major forms. The complex with Mn(II) has CCS values of 17.97 nm^2 and 22.29 nm^2, **Figure 4.19.D-F**. The complex with Co(II) has CCS values of 18.15 nm^2 and 22.65 nm^2, **Figure 4.19.G-I**. They both also have a minor form between the other forms with a CCS of about 21 nm^2. The complex with Mn(II) transitions from the major compact to major unfolded form at a CE of 17.2V, which is a 1.0V increase of CIU50 compared to unbound αSyn, indicating that once again Mn(II) has a stabilizing effect. The Co(II)

complex makes the same transition at 16.3V, representing a 0.1V increase in CIU50 compared to the unbound protein.

In all charge states analyzed, the addition of Mn(II) appears to have a stabilizing effect on the protein. In each, the protein bound to Mn(II) required more CE to unfold from its compact to expanded form. The stabilizing shift averages to 1.34V for all of the charge states analyzed, with the greatest shift of 2V for the +9 charge state. Co(II), on the other hand, appears to have a destabilizing effect on the protein, with all charge states unfolding at lower CE than the apo-protein, except for the +7 charge state. The average shift for all of the charge states measured is about 1V, with the greatest shift of 2V also being observed for the +9 charge state of the Co(II) complex. In general, all charge states of the Mn(II) and Co(II) complexes display the same forms that are seen in the apo-protein. The one exception is the +8 charge state of the protein complexed with Mn(II); here a compact form of the complex exists at low CE and disappears at 6.4V.

4.4.4 CIU of αSyn Bound to copper(II)

Along with Mn(II) and Co(II), αSyn was also analyzed while bound to Cu(II), which is thought to promote the aggregation of the protein. The protein binds with a micromolar affinity to Cu(II) and can bind to more than one metal ion. Upon analysis of the complex with native electrospray ionization-MS, a charge state distribution similar to that seen previously with the +10 charge state being dominant, **Figure 4.20.A.** Upon closer examination the unbound protein can no longer be observed, instead, peaks corresponding to the protein bound to one, two,

three, and four copper ions are seen, **Figure 4.20.B**. The protein bound to two copper ions has the greatest relative abundance.

The protein bound to one or three copper ions were analyzed using CIU. The +11 through +7 charge states were analyzed with only the +11, +10, +9, and +7 charge states yielding useful information. The +11 charge state shows the same forms of the protein that were measured for the apo-αSyn form, a compact form with a CCS value of 26.58 nm^2 (1 copper ion), 26.86nm^2 (3 copper ions), and an expanded form with a CCS of 30.54 nm^2, **Figure 4.21**. The protein with one copper bound shifts from the compact to an expanded form at a CE of 6.2V, while the complex with 3 copper bound makes the same shift at a CE of 8.2V.

The +9 charge state of the copper complexes shows the greatest differences when compared to the unbound protein, which has a major compact and major expanded form. In the bound form, the compact form with a CCS value of 23.35 nm^2 is seen in the complex with one copper ion and 23.59 nm^2 in the complex with three copper ions. Also, there is a short-lived minor form that appears in the complexes with a CCS of around 21.97 nm^2. This minor complex disappears at about 6V in the one copper complex and around 7V in the complex with three coppers.

Upon analysis of the +8 charged complexes, the same forms found in the previous analysis of the protein are observed. The unfolding of the major compact form to a major expanded form occurs at 11.0V and 12.8V for the complex with one and three coppers, respectively.

Finally, the +7 charge state of the complexes was analyzed. Once again, the same features were observed in the protein and complexes. There was a change in CE required to unfold the

protein, with 17.2V and 19.8V being required by the complex with one and three coppers, respectively.

Overall, all charge states show some amount of stabilization upon the addition of Cu(II). Upon closer examination, the addition of 3 metal ions provides more stabilization in all cases compared to one metal ion. The greatest amount of stabilization can be seen in the results from the +9 charge state where the extended form of the protein does not appear.

4.5 Conclusions

Based on the CIU results obtained during this study the binding of CLR01, Mn(II), 1x Cu(II), and 3x Cu(II) all stabilize the structure of αSyn. On the other hand, the binding of Co(II) destabilizes the protein. These results indicate that simply stabilizing or destabilizing protein structure in the gas phase does not adequately explain why the binding of CLR01 hinders the aggregation of αSyn, while the binding of divalent metals promotes aggregation. Furthermore, the CCS values of the conformations detected in the CIU fingerprints are the same, indicating that the binding of a ligand is not causing the protein to form a unique conformation that would explain differences in aggregation. One unexpected result is Co(II) having the opposite effect on the protein compared to the other divalent metals. This is especially puzzling as both Mn(II) and Co(II) bind to the protein in similar locations. Further investigation into the structures of αSyn is needed to determine the cause of its aggregation and ways to prevent it.

The application of collision induced unfolding provides information on the structure and stability of macromolecules that is complementary to data obtained from other biophysical tools. CIU probes the relative thermodynamic stability of a gas phase molecule. One of the best

applications of CIU is to probe the relative stability of protein-ligand interactions, as we examined here. In many cases, the binding of a small molecule, e.g., drug, to a protein or other large macromolecule imparts stability to prevent unfolding. CIU gives clues on how ligands interact with proteins, and the pharmaceutical industry and biotechnology labs are applying CIU to drug discovery efforts.

1 MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH

51 GVATVAEKTK EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL

101 GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

Figure 4.1: Sequence of human αSyn. Positively charged N-terminal region is highlighted in green, non-amyloid component (NAC) region that is responsible for aggregation is highlighted in red, and acidic C-terminus is highlighted in yellow.

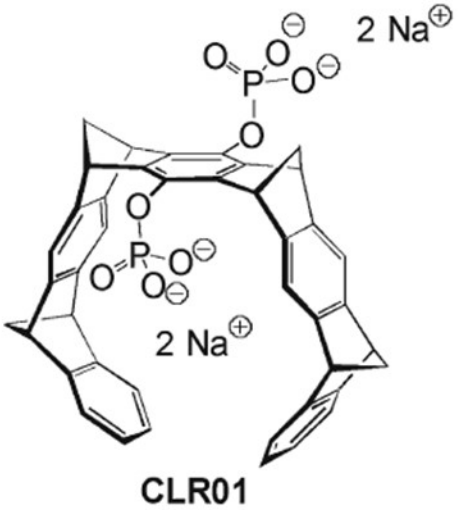


Figure 4.2: Structure of the molecule tweezer CLR01.

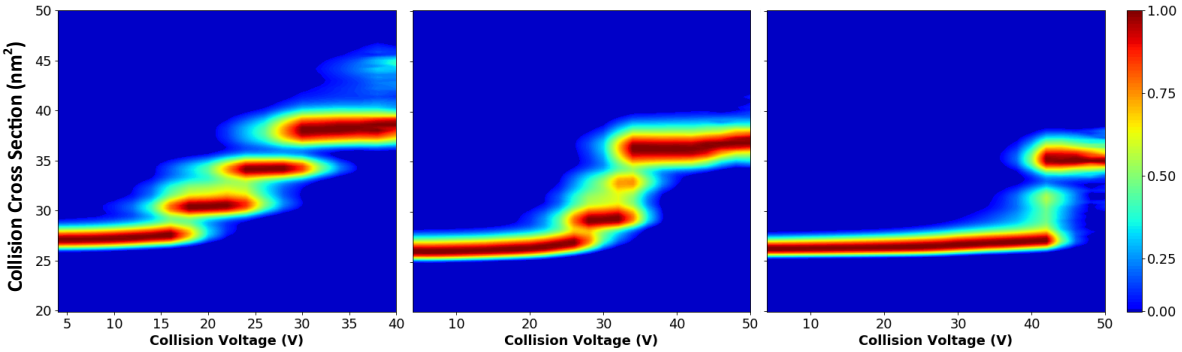


Figure 4.3: CIU fingerprints of human carbonic anhydrase I from left to right +11, +10, +9.

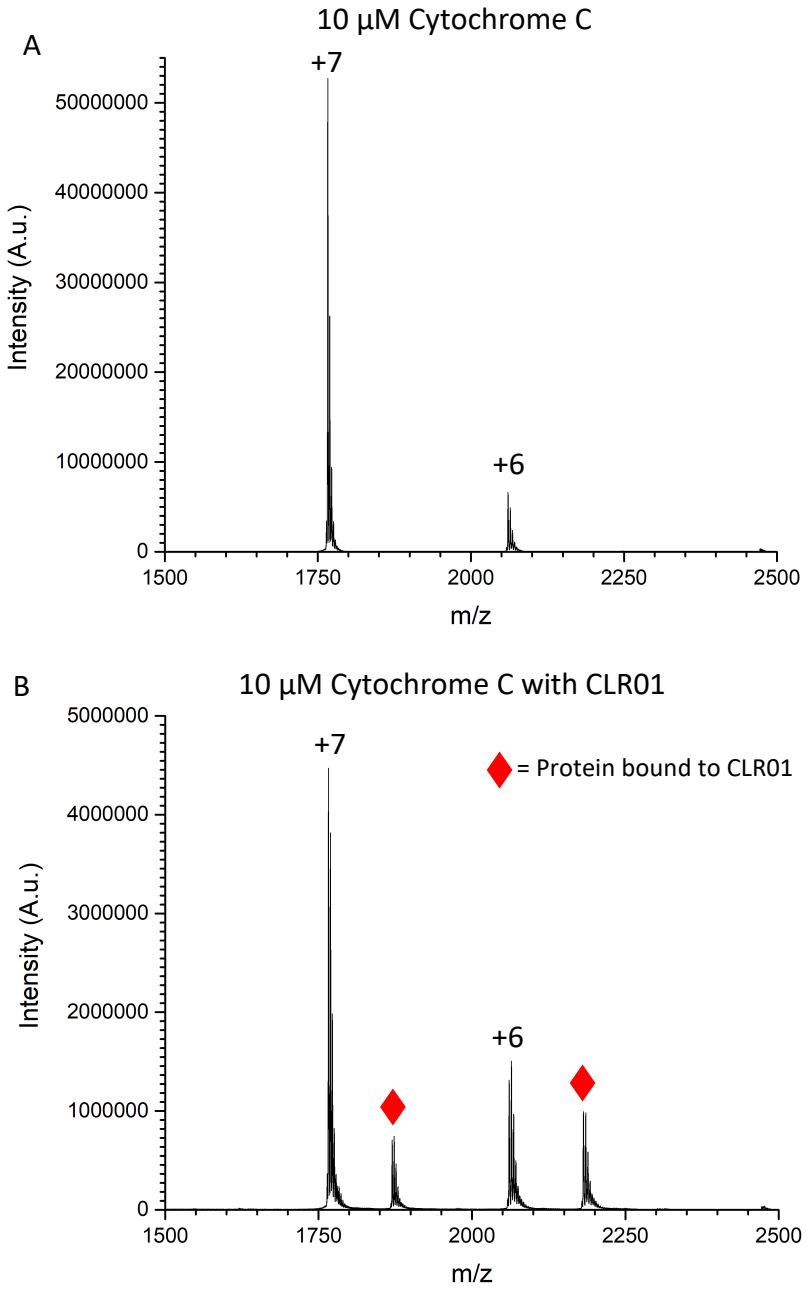


Figure 4.4: Native electrospray ionization mass spectrum of 10 µM equine heart cytochrome c in 200 mM ammonium acetate (A) and 10 µM equine heart cytrochrome c in a 1:1 ratio with CLR01 in 200 mM ammonium acetate (B). Peaks corresponding to protein bound to CLR01 are marked with a red diamond.

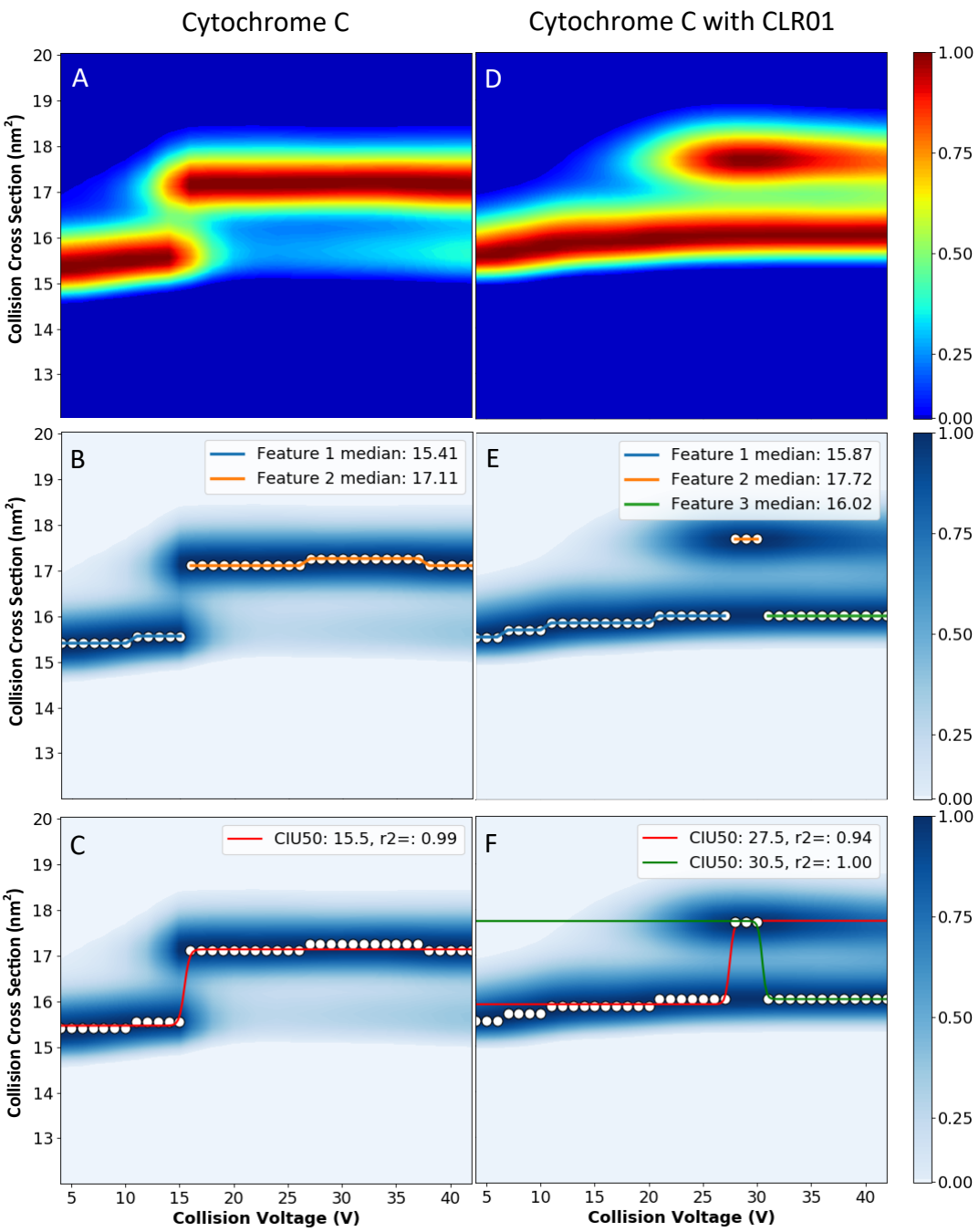


Figure 4.5: CIU fingerprints of +6 charge state of equine heart cytochrome c (A) and equine heart cytochrome c bound to CLR01 (D) with collision voltage ranging from 4V to 42V. Feature detected in fingerprints for unbound (B) and bound (E) protein. CIU50 fit analysis shows midpoint of transitions for unbound (C) and bound (F) protein. CLR01 appears to have stabilized the protein with the compact conformation being present for all CE.

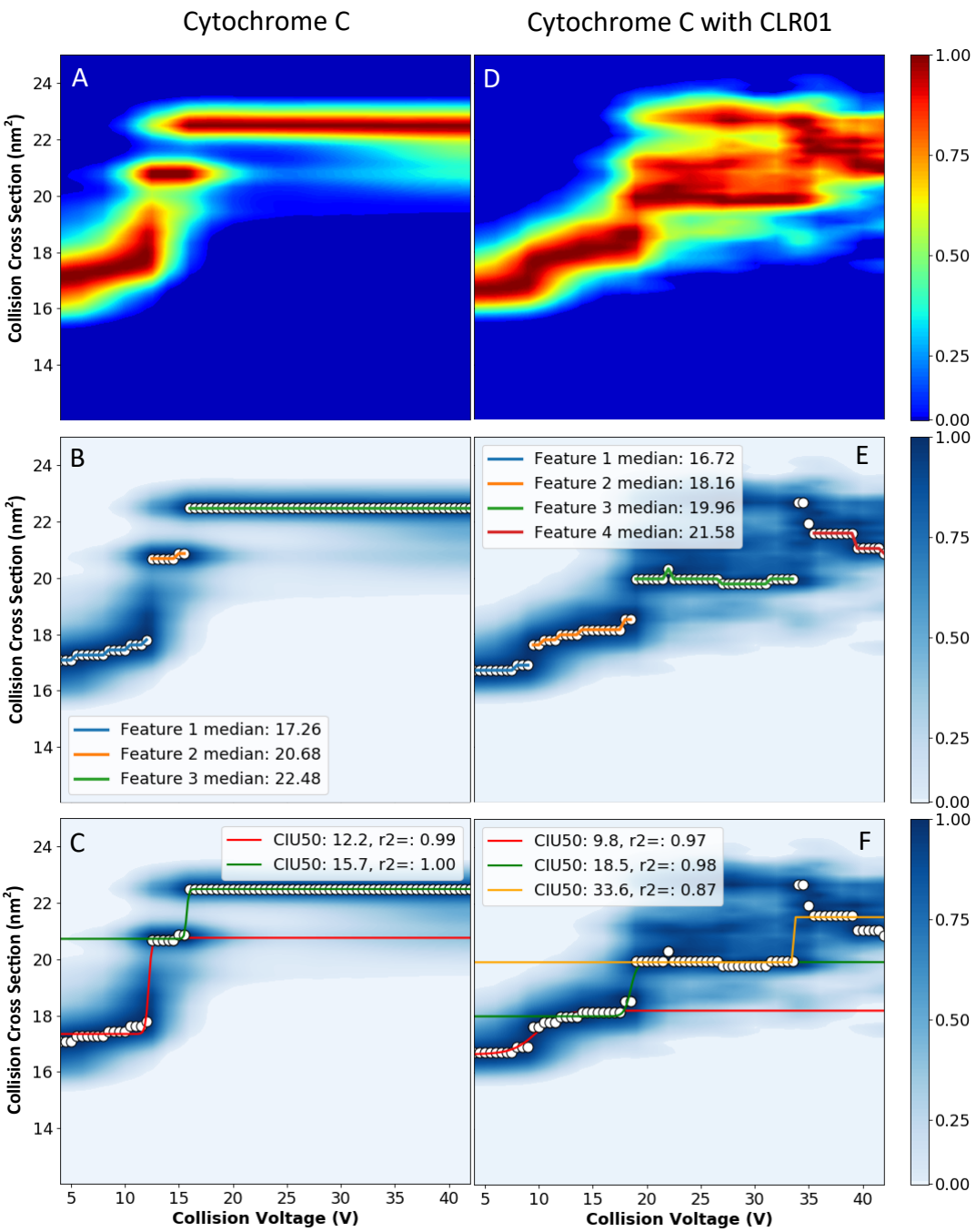


Figure 4.6: CIU fingerprints of +7 charge state equine heart cytochrome c (A) and equine heart cytochrome c bound to CLR01 (D) with collision voltage ranging from 4V to 42V. Feature detected in fingerprints for unbound (B) and bound (E) protein. CIU50 fit analysis shows midpoint of transitions for unbound (C) and bound (F) protein. CLR01 appears to destabilize the expanded conformation of the protein as signal rapidly deteriorates at approximately 20V.

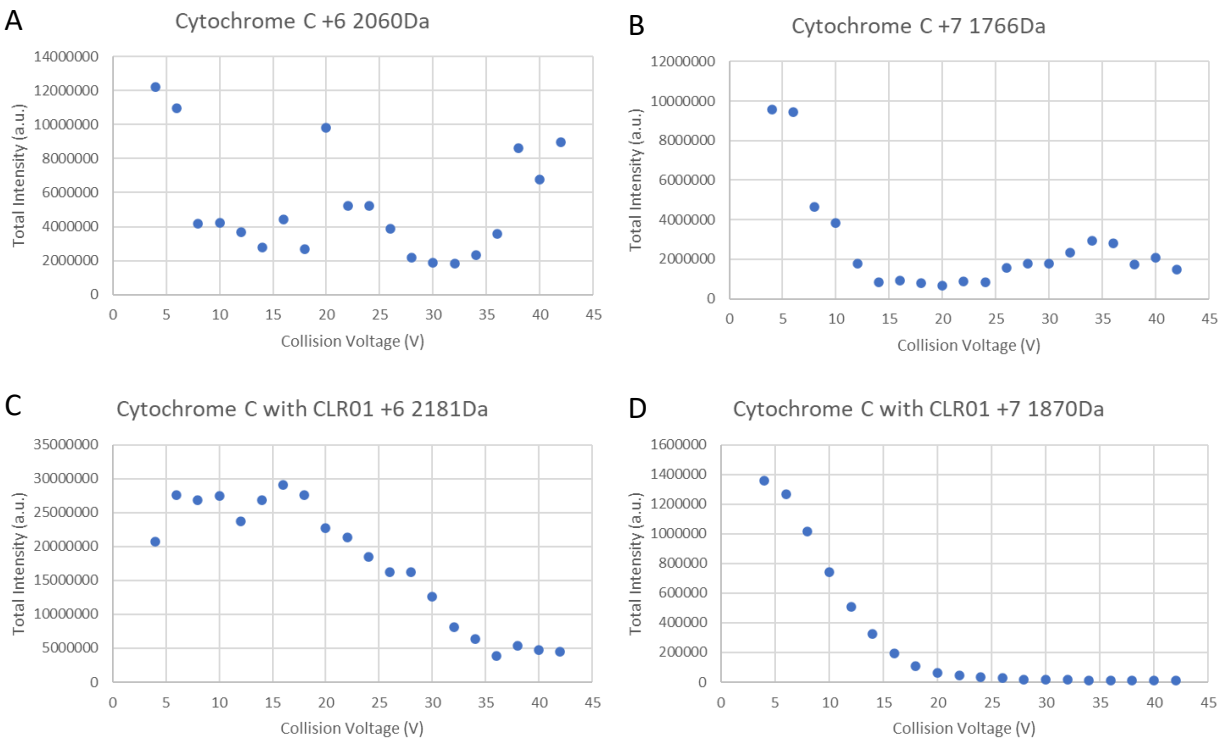


Figure 4.7: Comparison of total signal intensity during CIU experiments of +6 and +7 charge states equine heart cytochrome c unbound (A,B) and bound to CLR01 (C,D). The +7 charge state of the protein bound to CLR01 appears to disintegrate as CE reaches 20V, indicating the ligand may destabilize the protein.

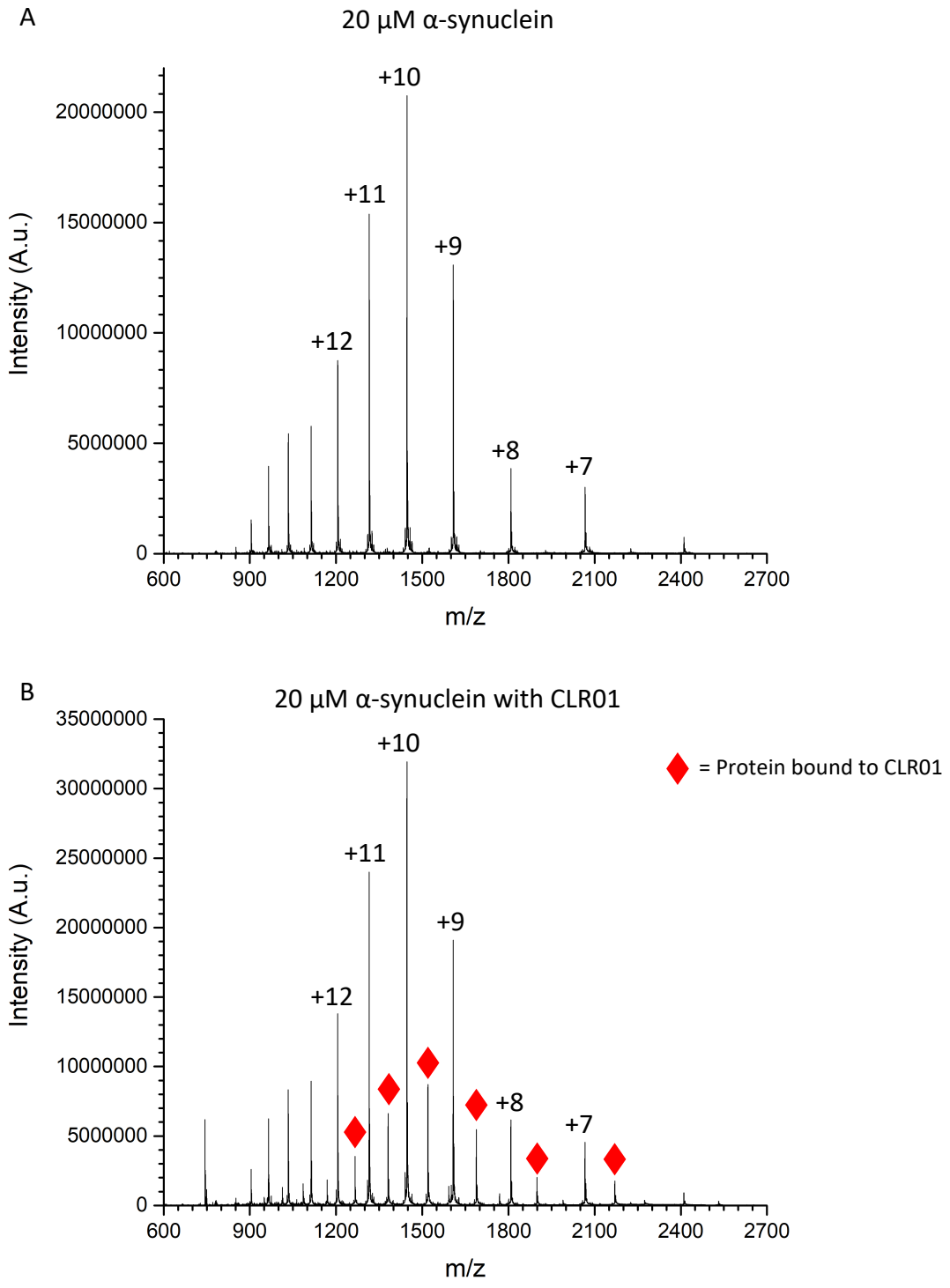


Figure 4.8: Native electrospray ionization mass spectrum of 10 µM αSyn in 20 mM ammonium acetate (A) and 10 µM αSyn in a 1:1 with CLR01 in 20 mM ammonium acetate (B). Peaks corresponding to protein bound to CLR01 are marked with a red diamond.

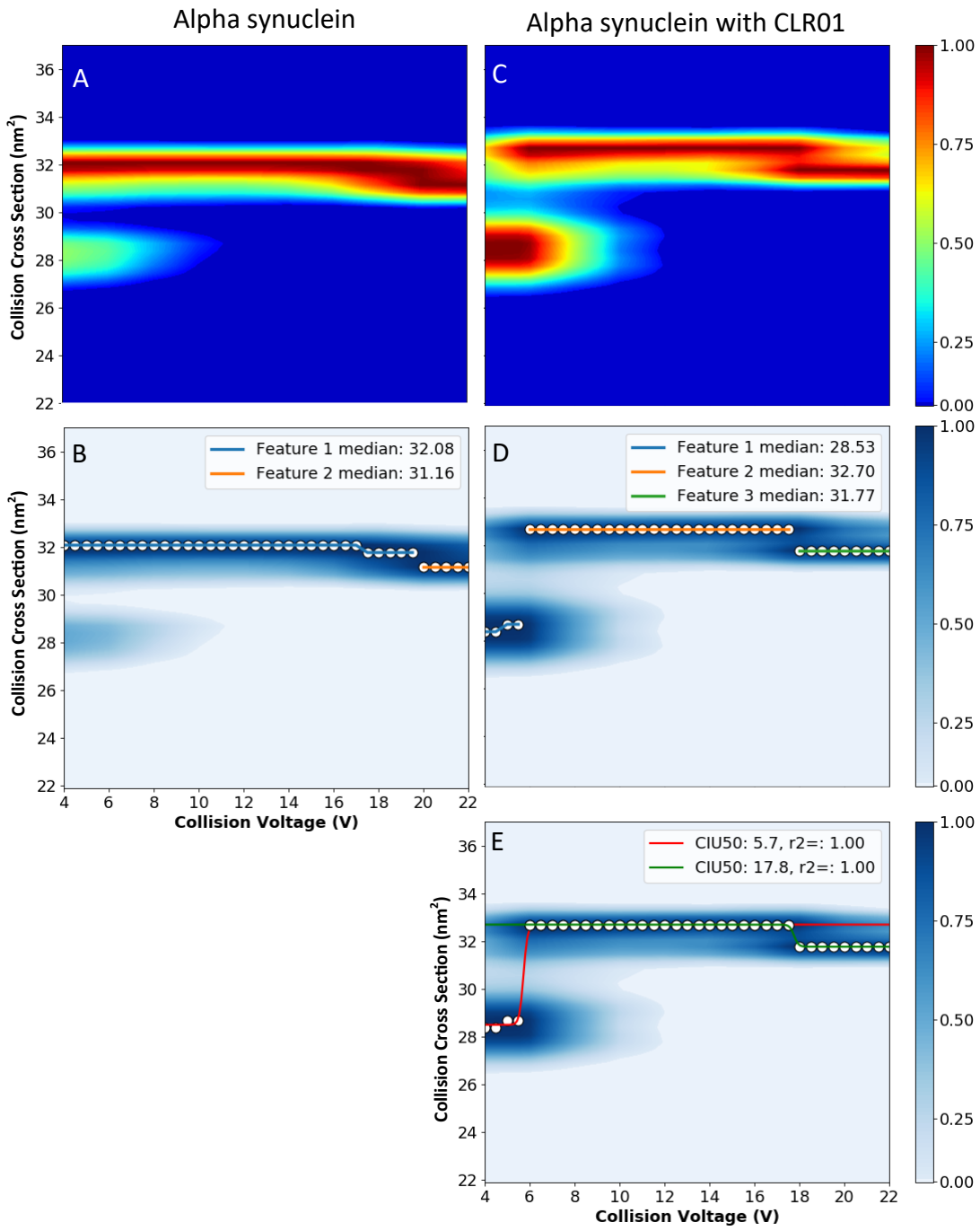


Figure 4.9: CIU fingerprints of +12 charge state αSyn (A) and αSyn bound to CLR01 (C) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound (B) and bound (D) protein. CIU50 fit analysis shows midpoint of transitions for bound protein (E). CIU50 fit analysis not available for unbound protein because compact form is never dominant feature in unbound protein CIU fingerprint. The binding of CLR01 to the protein appears to have a stabilizing effect.

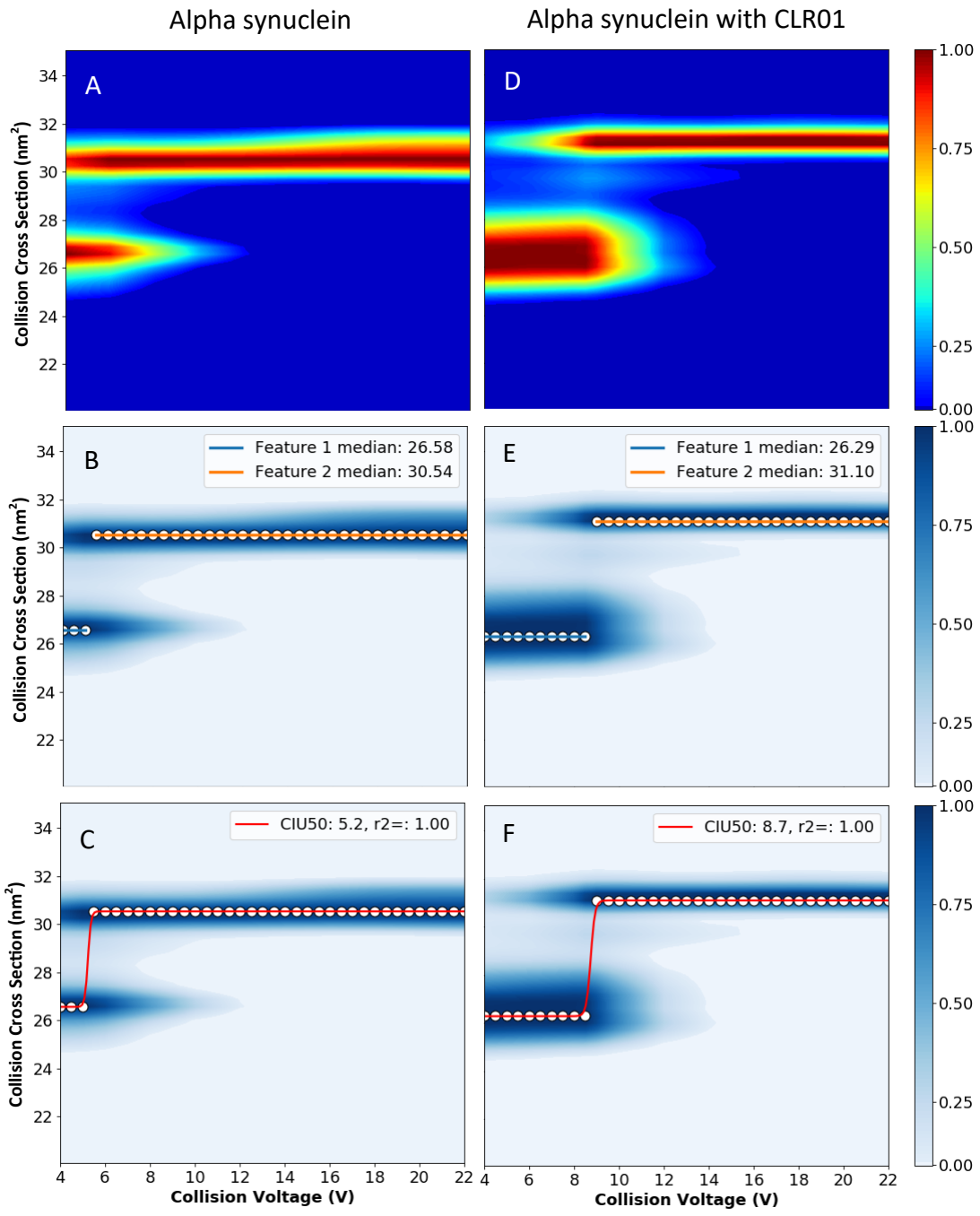


Figure 4.10: CIU fingerprints of +11 charge state αSyn (A) and αSyn bound to CLR01 (D) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound (B) and bound (E) protein. CIU50 fit analysis shows midpoint of transitions for unbound (C) and bound (F) protein. The binding of CLR01 to the protein appears to have a stabilizing effect with a shift in CIU50 of 3.5V.

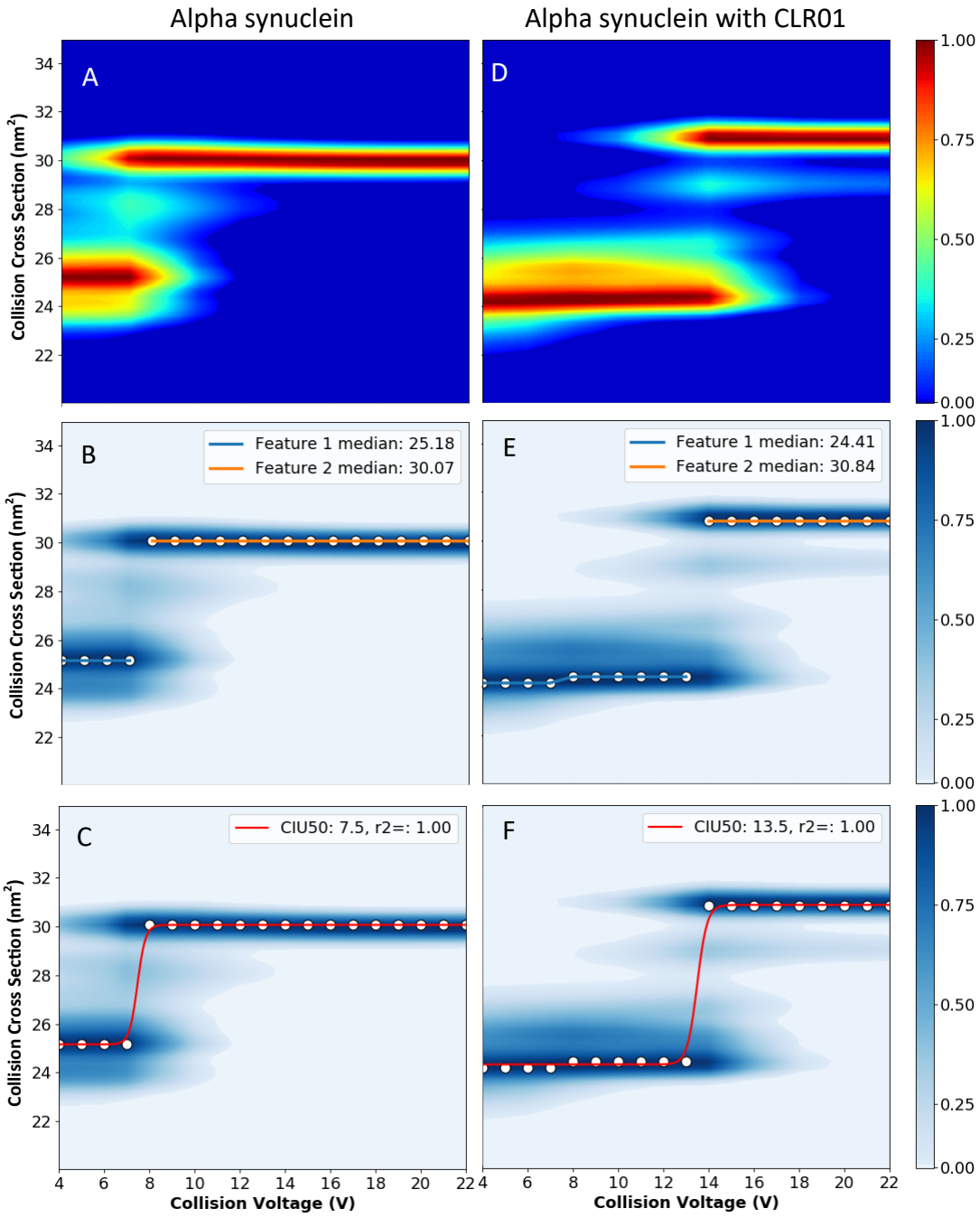


Figure 4.11: CIU fingerprints of +10 charge state αSyn (A) and αSyn bound to CLR01 (D) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound (B) and bound (E) protein. CIU50 fit analysis shows midpoint of transitions for unbound (C) and bound (F) protein. The binding of CLR01 to the protein appears to have a stabilizing effect, shifting CIU50 higher by 6.0V.

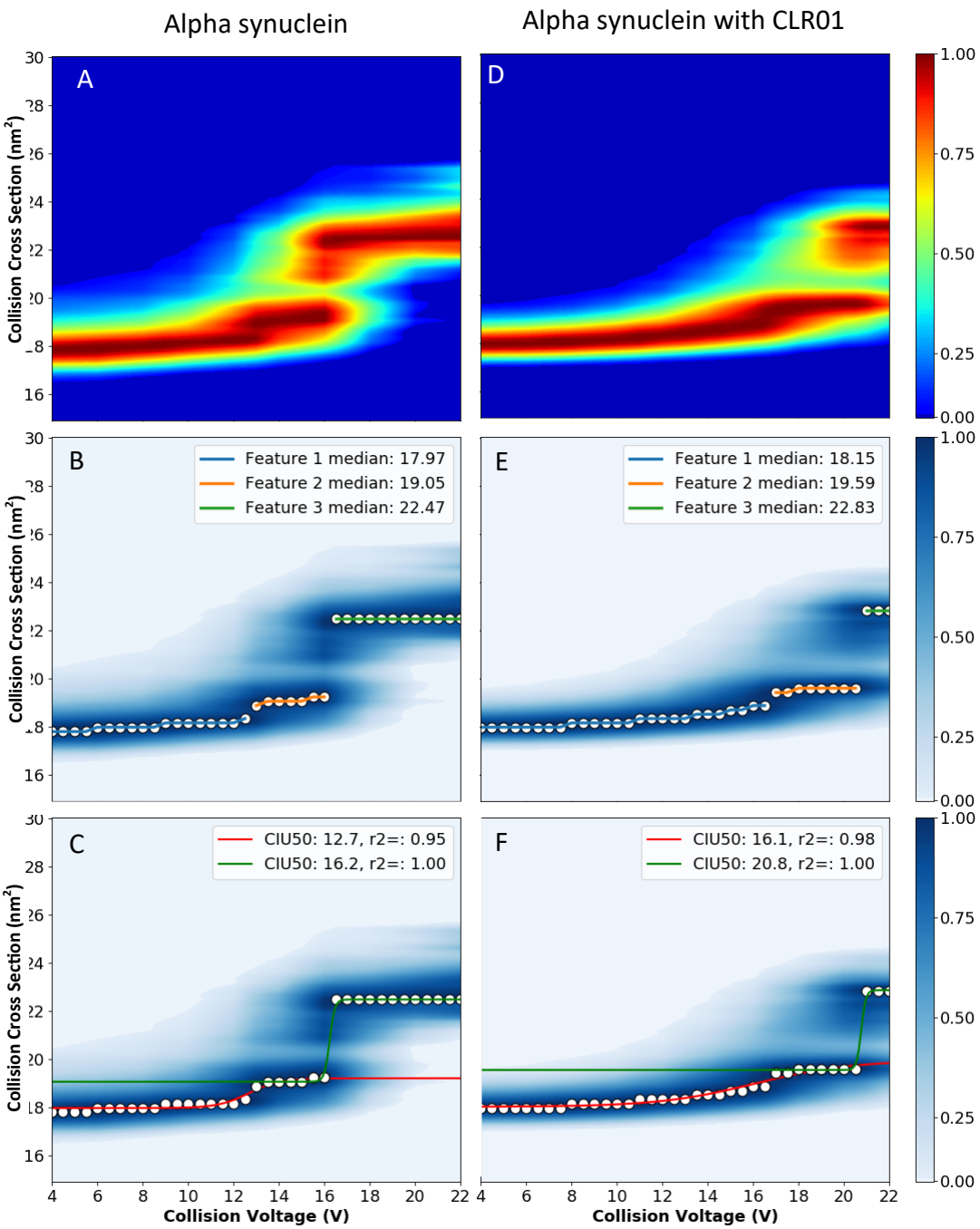


Figure 4.12: CIU fingerprints of +7 charge state αSyn (A) and αSyn bound to CLR01 (D) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound (B) and bound (E) protein. CIU50 fit analysis shows midpoint of transitions for unbound (C) and bound (F) protein. The binding of CLR01 to the protein appears to have a stabilizing effect with the fist transition CIU50 increasing by 3.4V and the second transition increasing by 4.6V.

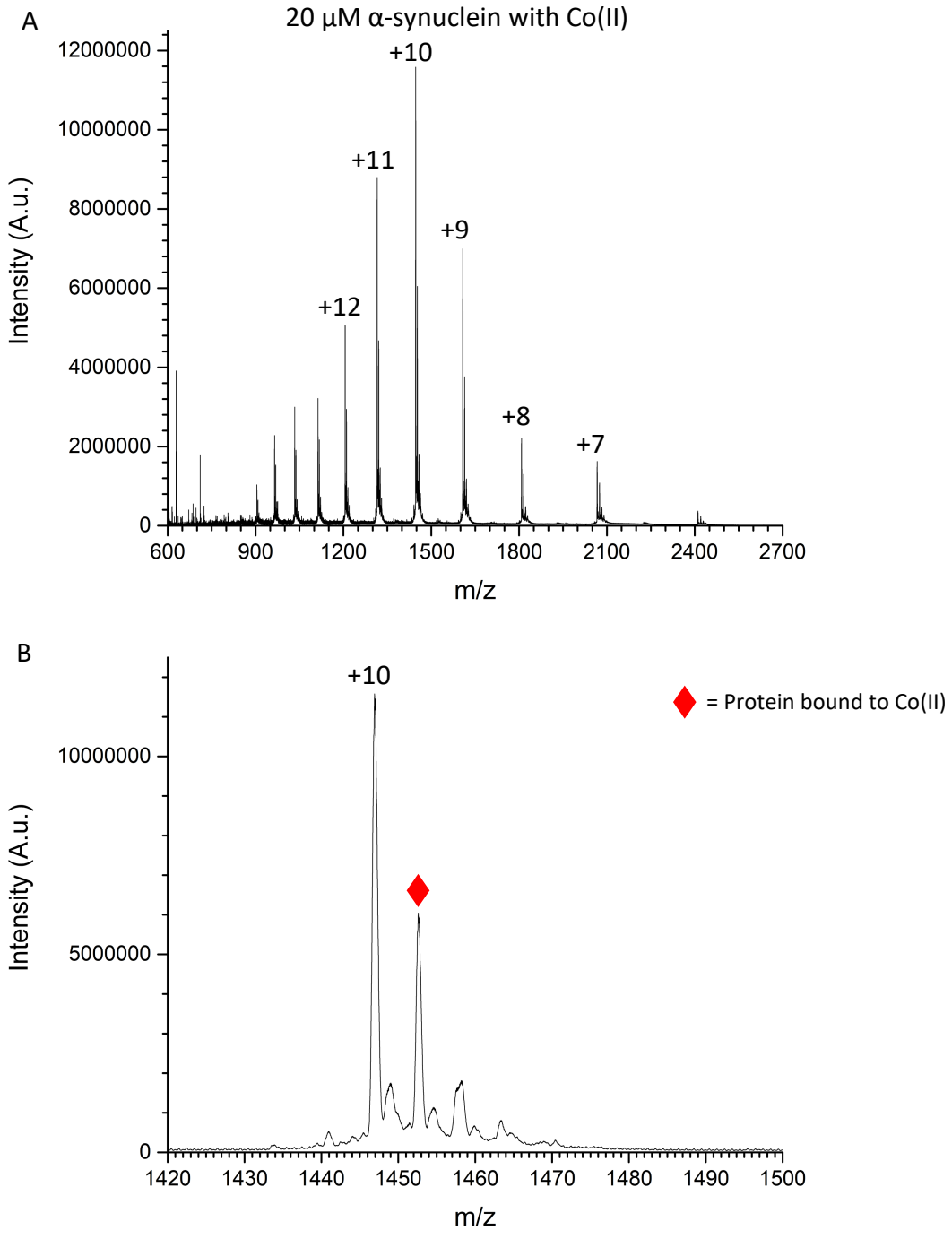


Figure 4.13: Native electrospray ionization mass spectrum of 20 µM αSyn in a 1:8 ratio with Co(II) acetate in 20 mM ammonium acetate (A). +10 charge state of αSyn with Co(II) (B). Peak corresponding to protein bound to Co(II) is marked with a red diamond.

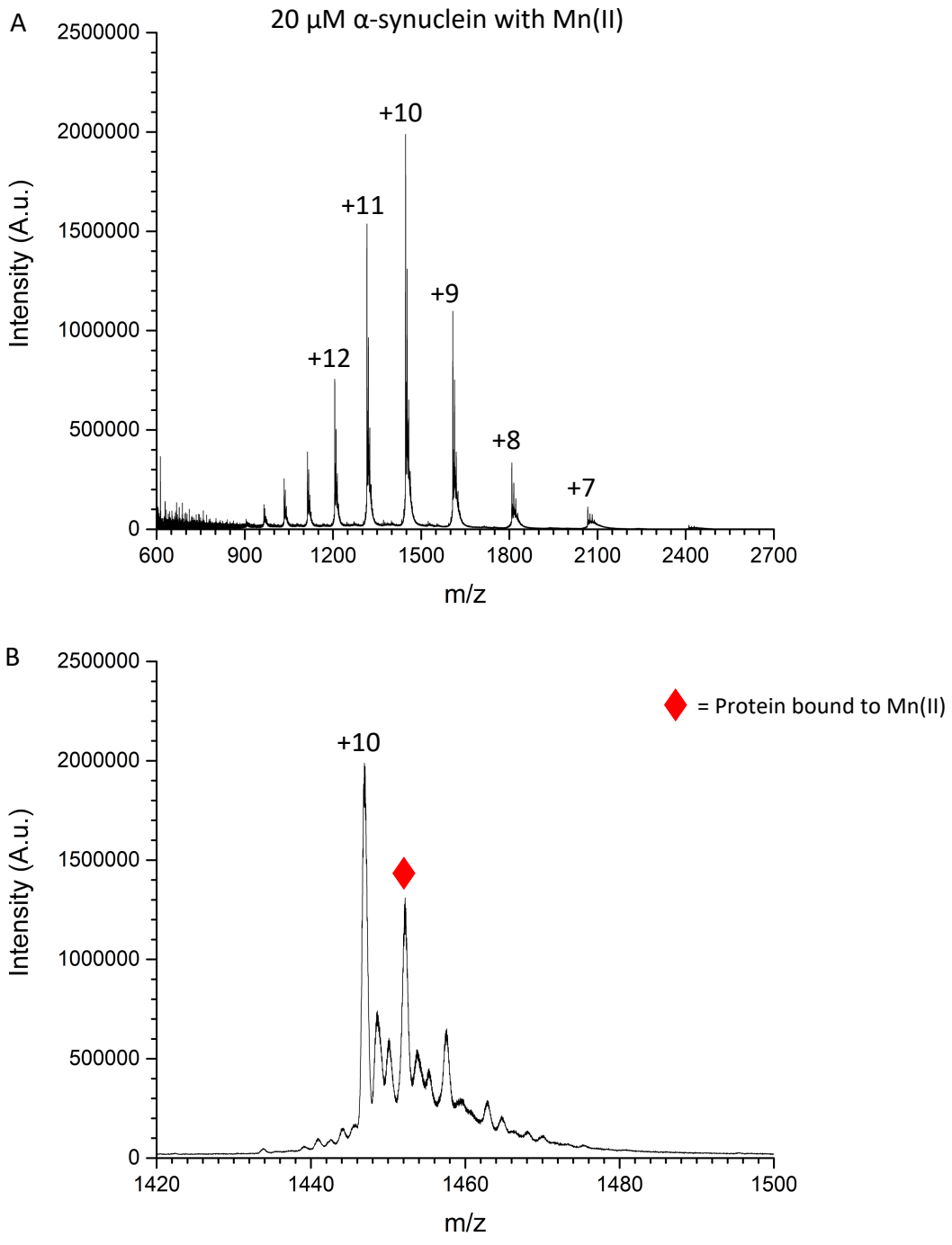


Figure 4.14: Native electrospray ionization mass spectrum of 20 µM αSyn in a 1:10 ratio with Mn(II) acetate in 20 mM ammonium acetate (A). +10 charge state of αSyn with Mn(II) (B). Peak corresponding to protein bound to Mn(II) is marked with a red diamond.

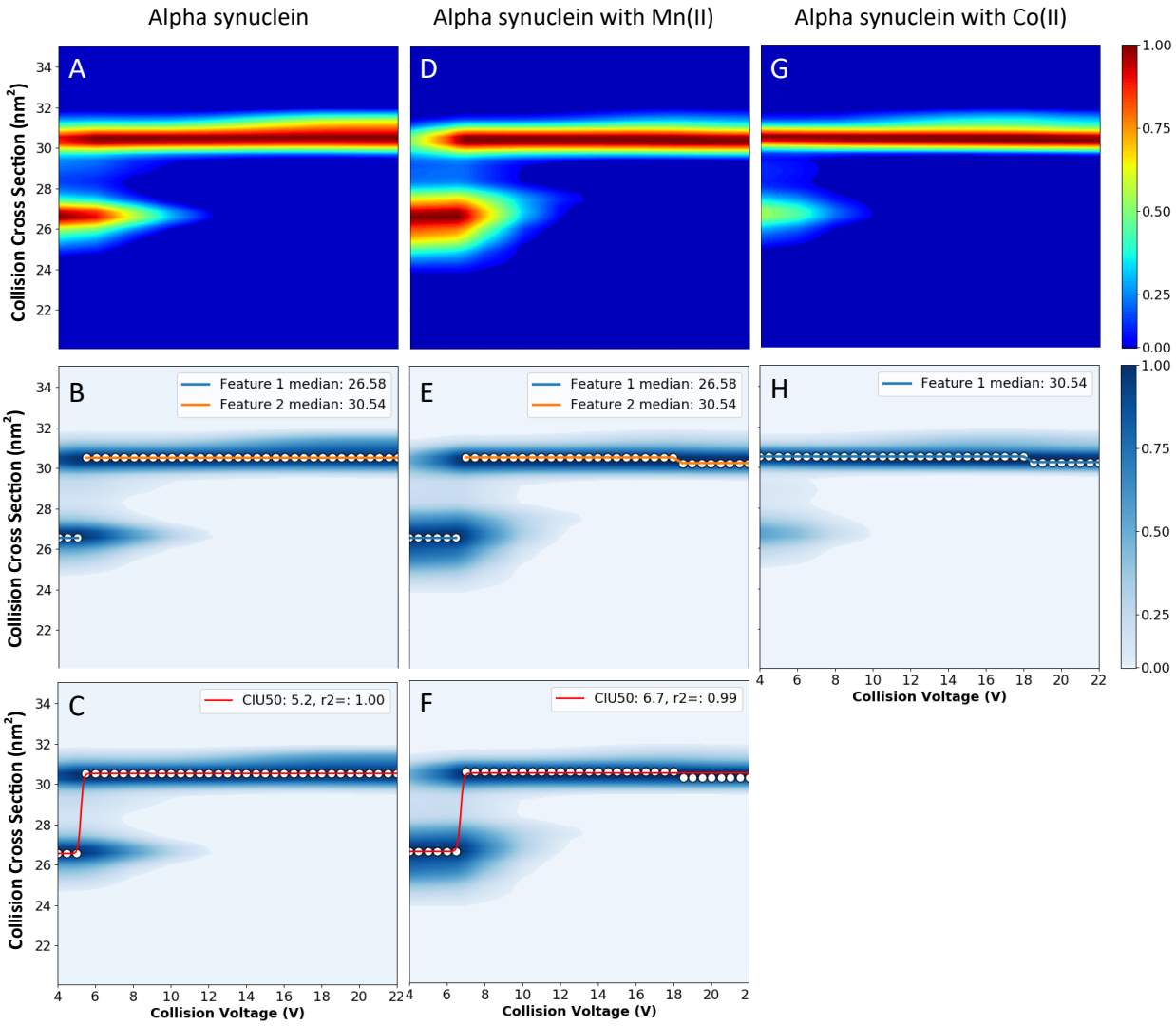


Figure 4.15: CIU fingerprints of +11 charge state αSyn (A), αSyn bound to Mn(II) (D), and αSyn bound to Co(II) (G) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound protein (B), protein with Mn(II) (E), and protein with Co(II) (H). CIU50 fit analysis shows midpoint of transitions for unbound protein (C) and protein bound to Mn(II) (F). CIU50 fit analysis not available for protein bound to Co(II) because the compact form is never the dominant feature CIU fingerprint. The binding of Mn(II) to the protein has a stabilizing effect, while Co(II) has the opposite effect.

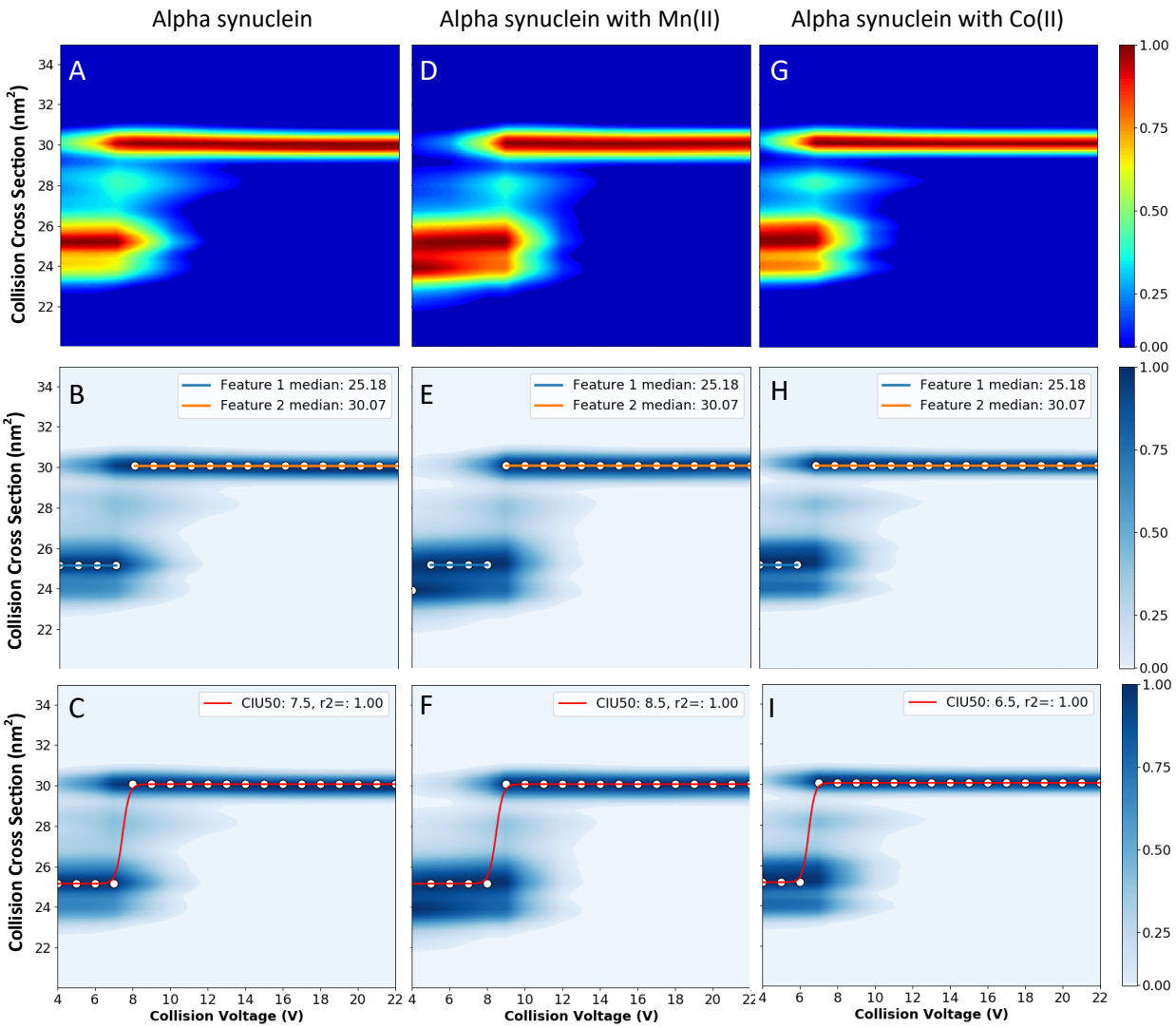


Figure 4.16: CIU fingerprints of +10 charge state αSyn (A), αSyn bound to Mn(II) (D), and αSyn bound to Co(II) (G) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound protein (B), protein with Mn(II) (E), and protein with Co(II) (H). CIU50 fit analysis shows midpoint of transitions for unbound protein (C), protein bound to Mn(II) (F), and protein bound to Co(II) (I). The binding of Mn(II) to the protein has a stabilizing effect, increasing CIU50 by 1.0V, while Co(II) has the opposite effect decreasing CIU50 by 1.0V.

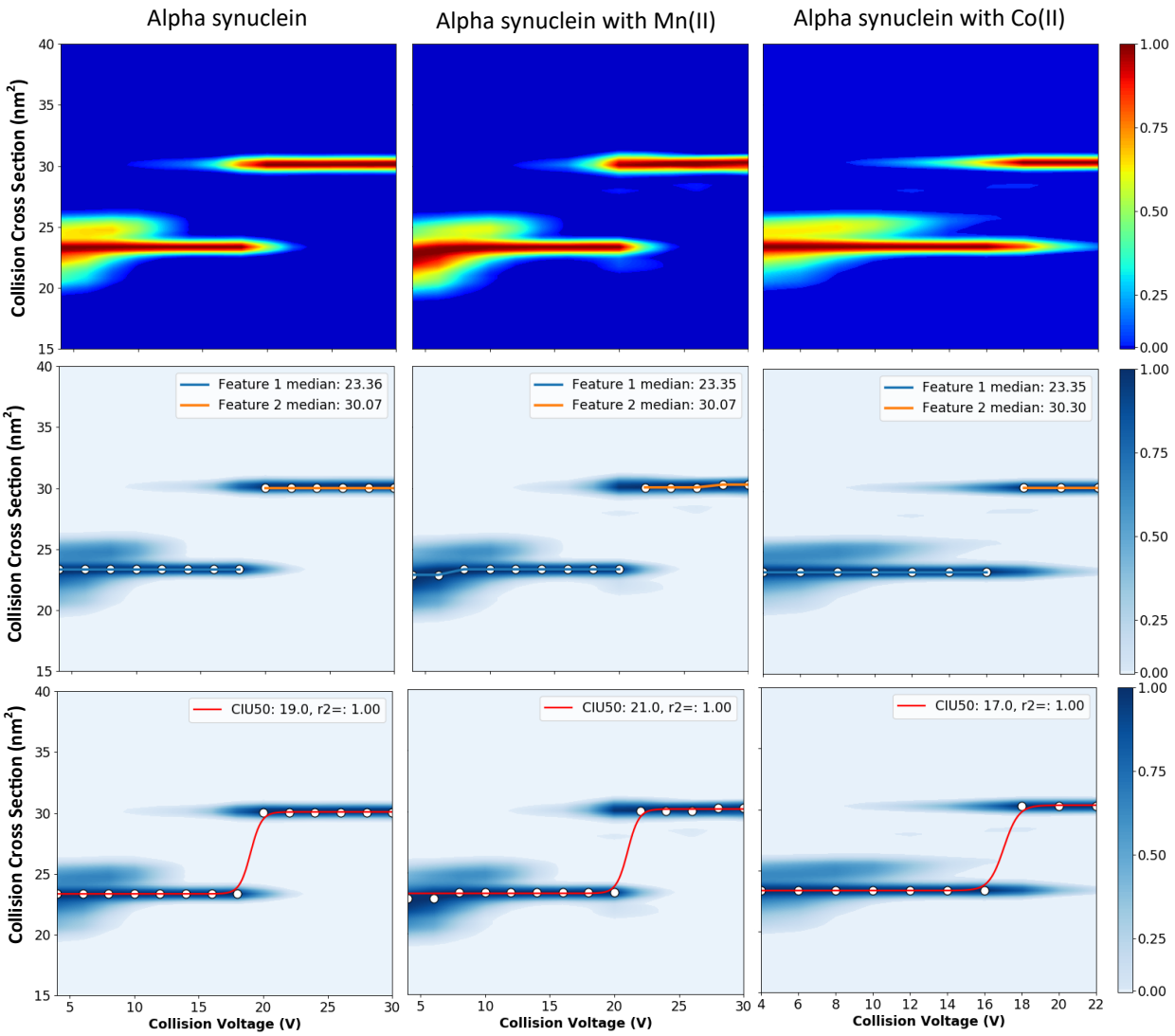


Figure 4.17: CIU fingerprints of +9 charge state αSyn (A), αSyn bound to Mn(II) (D), and αSyn bound to Co(II) (G). Feature detected in fingerprints for unbound protein (B), protein with Mn (II) (E), and protein with Co(II) (H). CIU50 fit analysis shows midpoint of transitions for unbound protein (C), protein bound to Mn(II) (F), and protein bound to Co(II) (I). The binding of Mn(II) to the protein has a stabilizing effect, increasing CIU50 by 2.0V, while Co(II) has the opposite effect decreasing CIU50 by 2.0V.

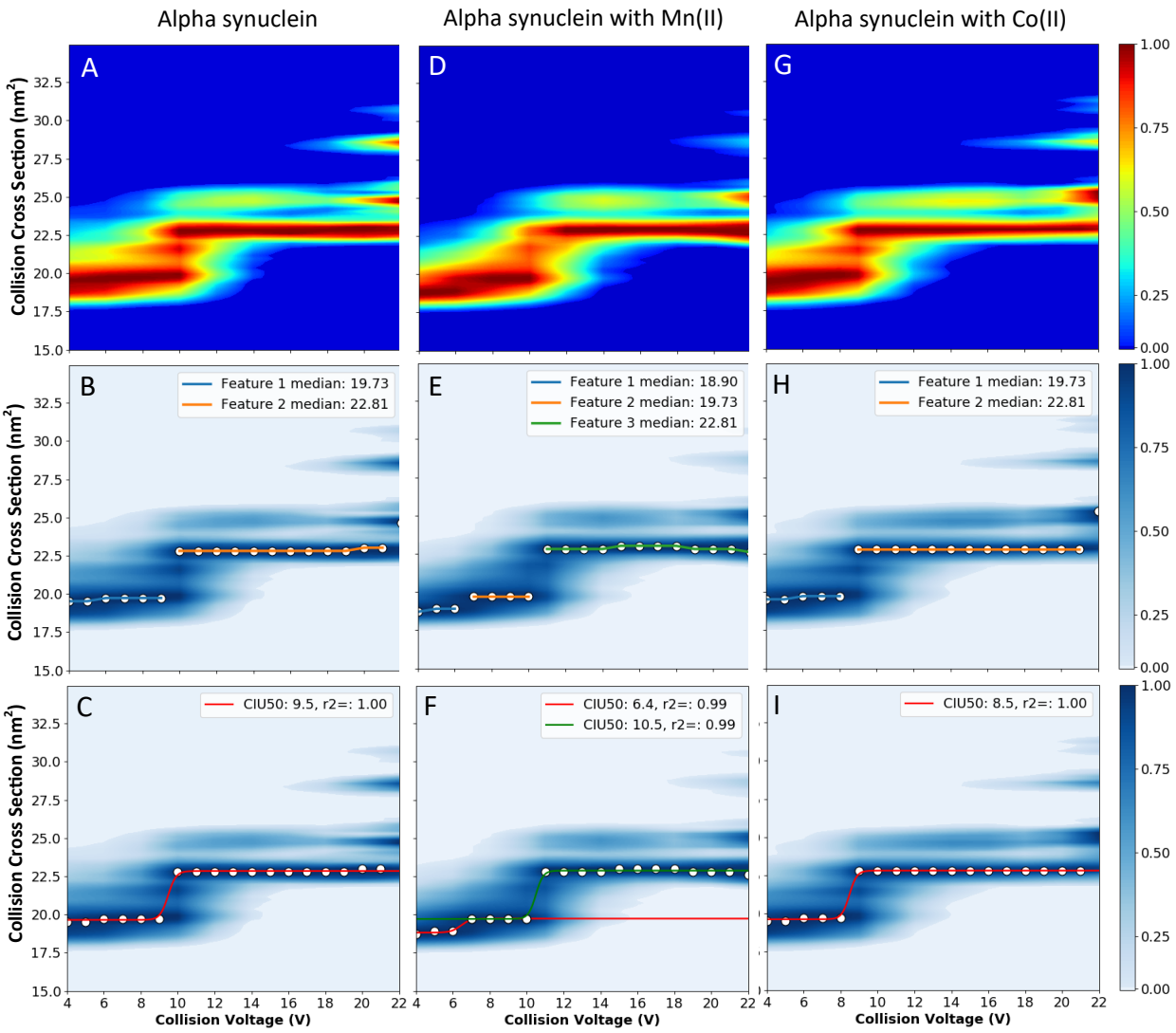


Figure 4.18: CIU fingerprints of +8 charge state αSyn (A), αSyn bound to Mn(II) (D), and αSyn bound to Co(II) (G) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound protein (B), protein with Mn(II) (E), and protein with Co(II) (H). CIU50 fit analysis shows midpoint of transitions for unbound protein (C), protein bound to Mn(II) (F), and protein bound to Co(II) (I). The binding of Mn(II) to the protein has a stabilizing effect, increasing CIU50 by 1.0V while revealing an additional compact form of the protein that was not previously observed. The Co(II) has a destabilizing effect, decreasing CIU50 by 1.0V.

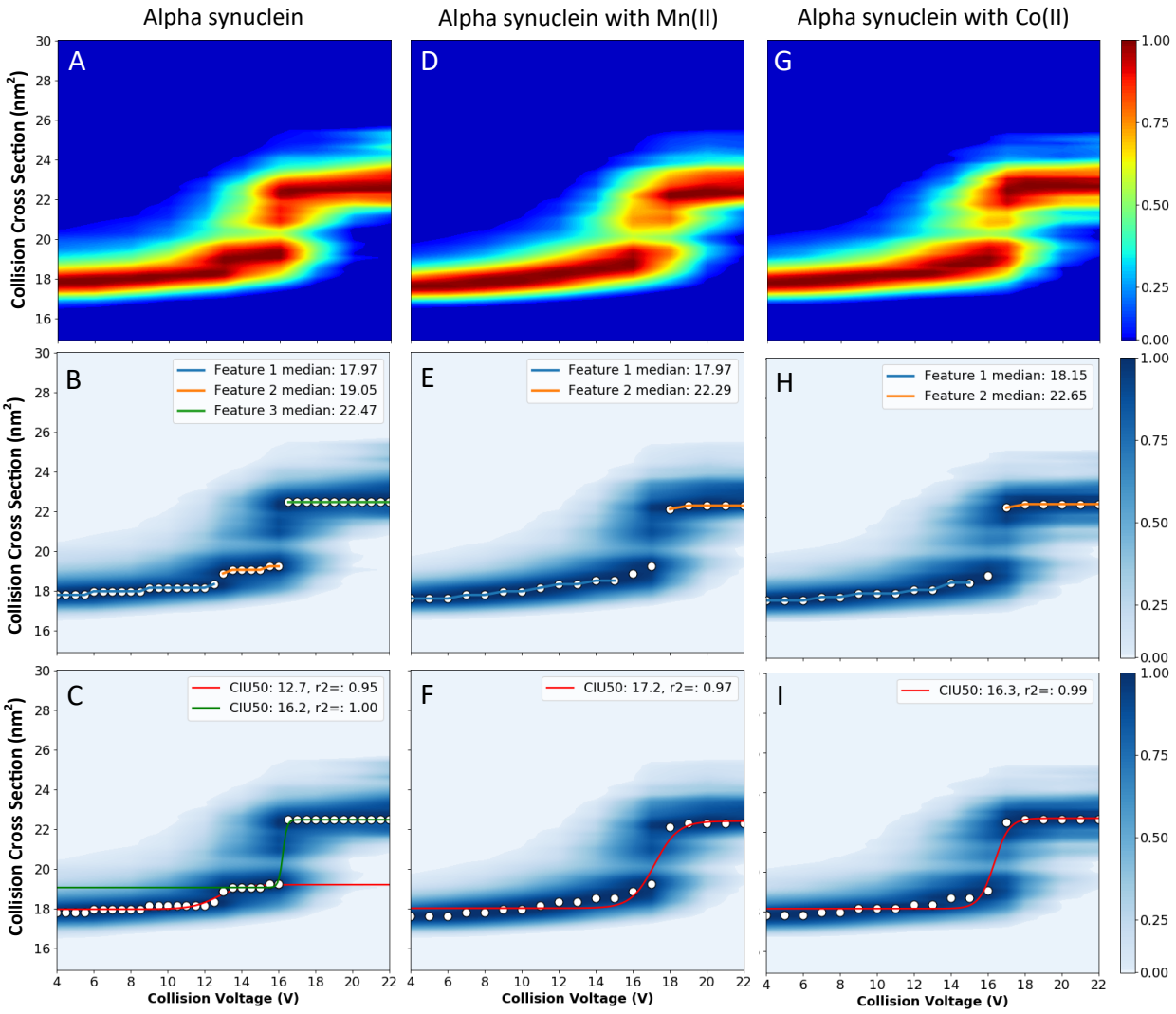


Figure 4.19: CIU fingerprints of +7 charge state αSyn (A), αSyn bound to Mn(II) (D), and αSyn bound to Co(II) (G) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound protein (B), protein with Mn(II) (E), and protein with Co(II) (H). CIU50 fit analysis shows midpoint of transitions for unbound protein (C), protein bound to Mn(II) (F), and protein bound to Co(II) (I). The binding of Mn(II) to the protein has a stabilizing effect, increasing CIU50 by 1.0V. Co(II) has a negligible effect on the protein slightly increasing CIU50 by 0.1V.

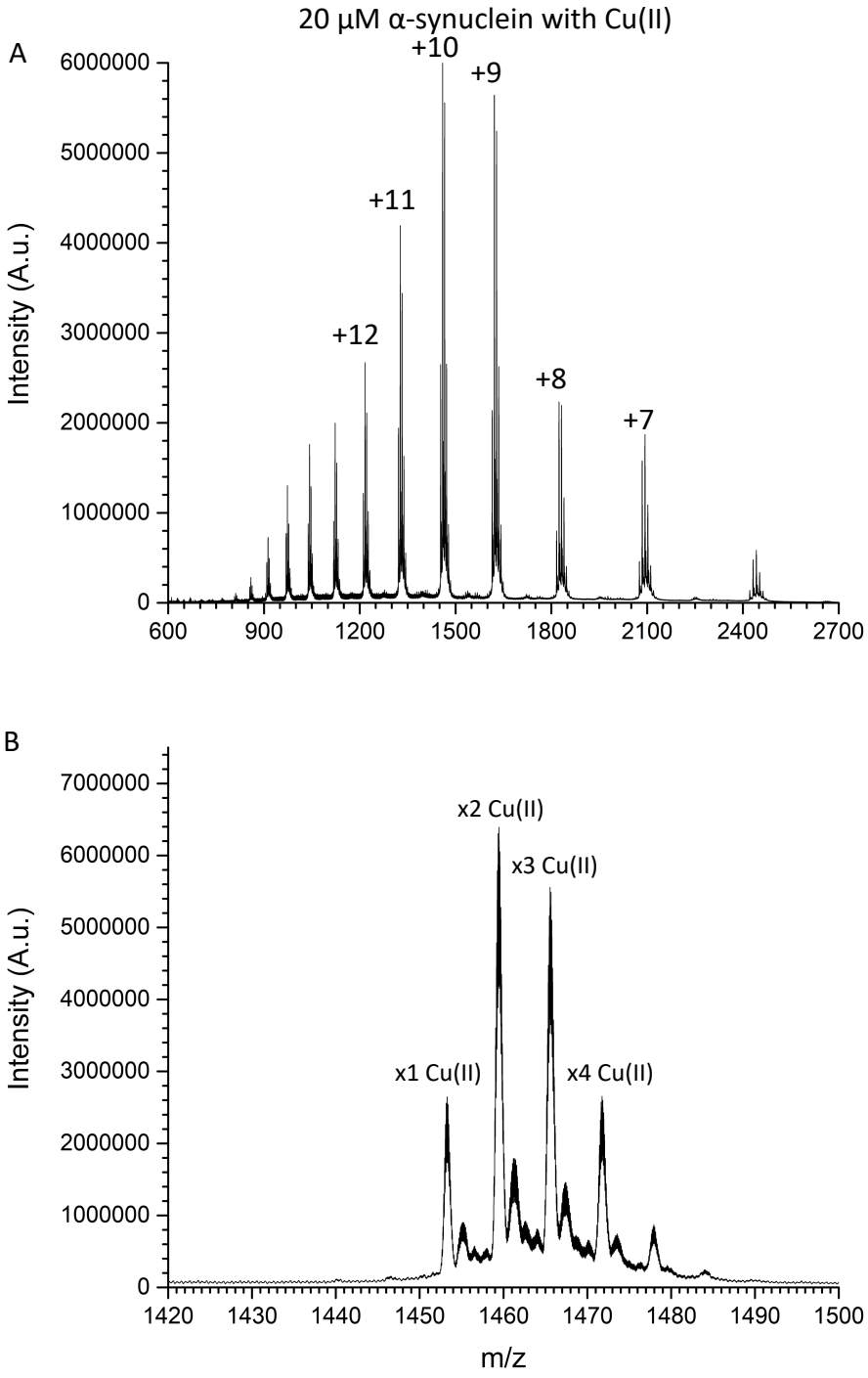


Figure 4.20: Native electrospray ionization mass spectrum of 20 µM αSyn in a 1:5 ratio with Cu(II) acetate in 20 mM ammonium acetate (A). +10 charge state of αSyn with Cu(II) (B). All of the αsyn is bound to one or more Cu(II) with most of the protein being bound to either 2 or 3 Cu(II) ions.

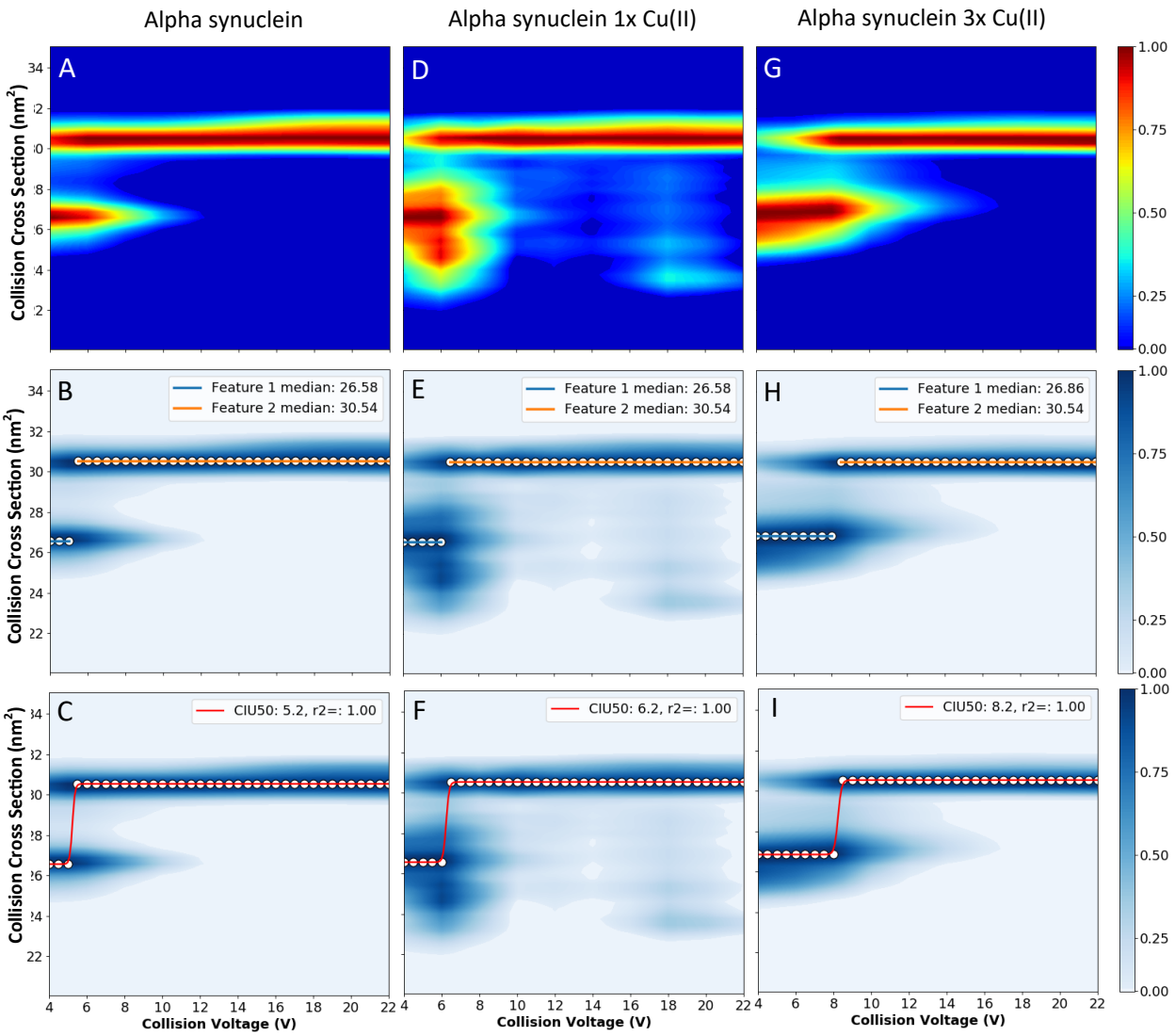


Figure 4.21: CIU fingerprints of +11 charge state αSyn (A), αSyn bound to 1x Cu(II) (D), and αSyn bound to 3x Cu(II) (G) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound protein (B), protein with 1x Cu(II) (E), and protein with 3x Cu(II) (H). CIU50 fit analysis shows midpoint of transitions for unbound protein (C), protein bound to 1x Cu(II) (F), and protein bound to 3x Cu(II) (I). The binding of 1x Cu(II) to the protein has a stabilizing effect, increasing CIU50 by 1.0V. 3x Cu(II) has an even greater stabilizing effect on the protein increasing CIU50 by 3.0V.

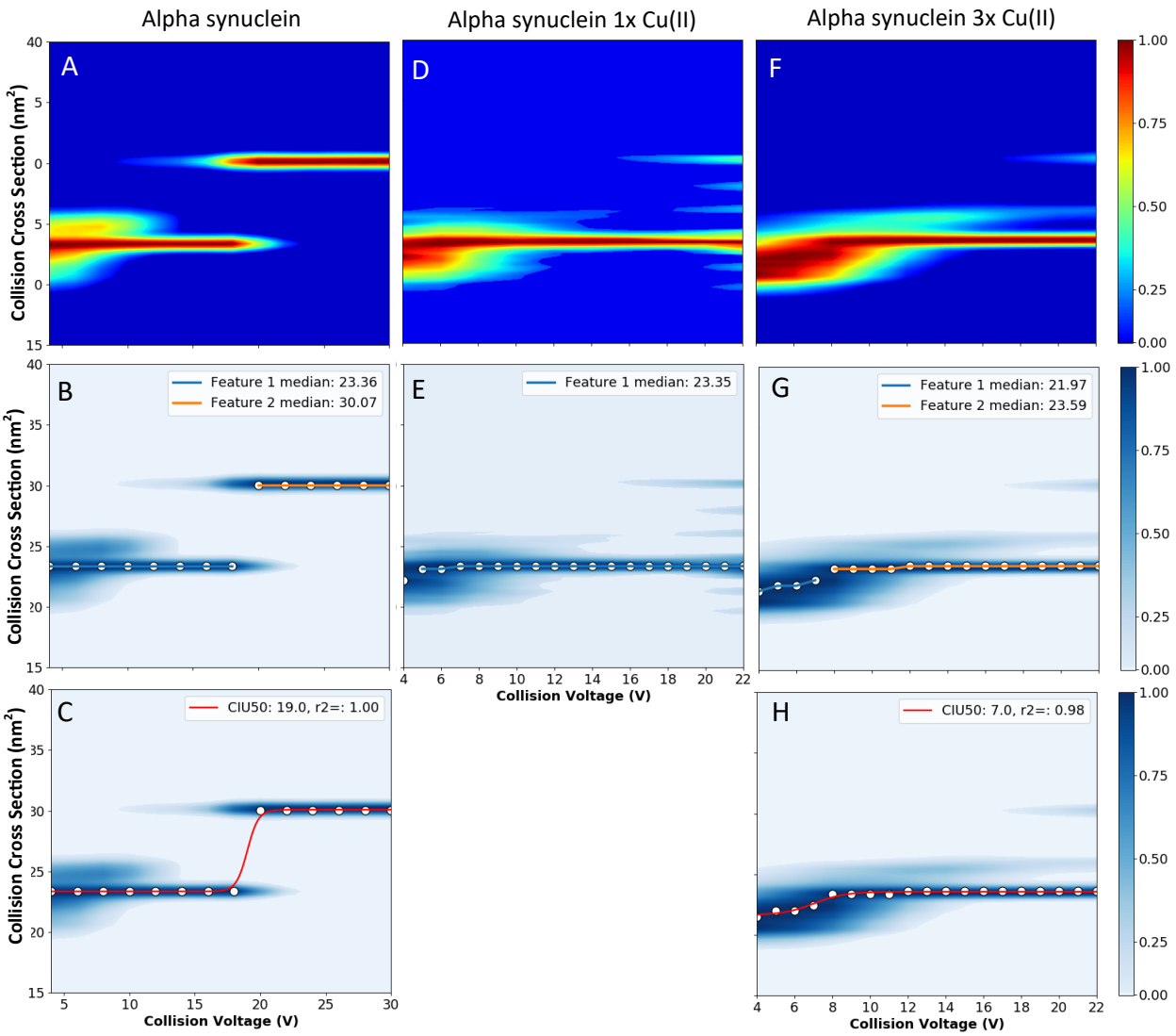


Figure 4.22: CIU fingerprints of +9 charge state αSyn (A), αSyn bound to 1x Cu(II) (D), and αSyn bound to 3x Cu(II) (F). Feature detected in fingerprints for unbound protein (B), protein with 1x Cu(II) (E), and protein with 3x Cu(II) (G). CIU50 fit analysis shows midpoint of transitions for unbound protein (C) and protein bound to 3x Cu(II) (H). CIU50 fit analysis not available for protein bound to 1x Cu(II) because the expanded form is never the dominant feature CIU fingerprint. The binding of 1x and 3x Cu(II) to the protein has a stabilizing effect with the expanded form not appearing in the range analyzed.

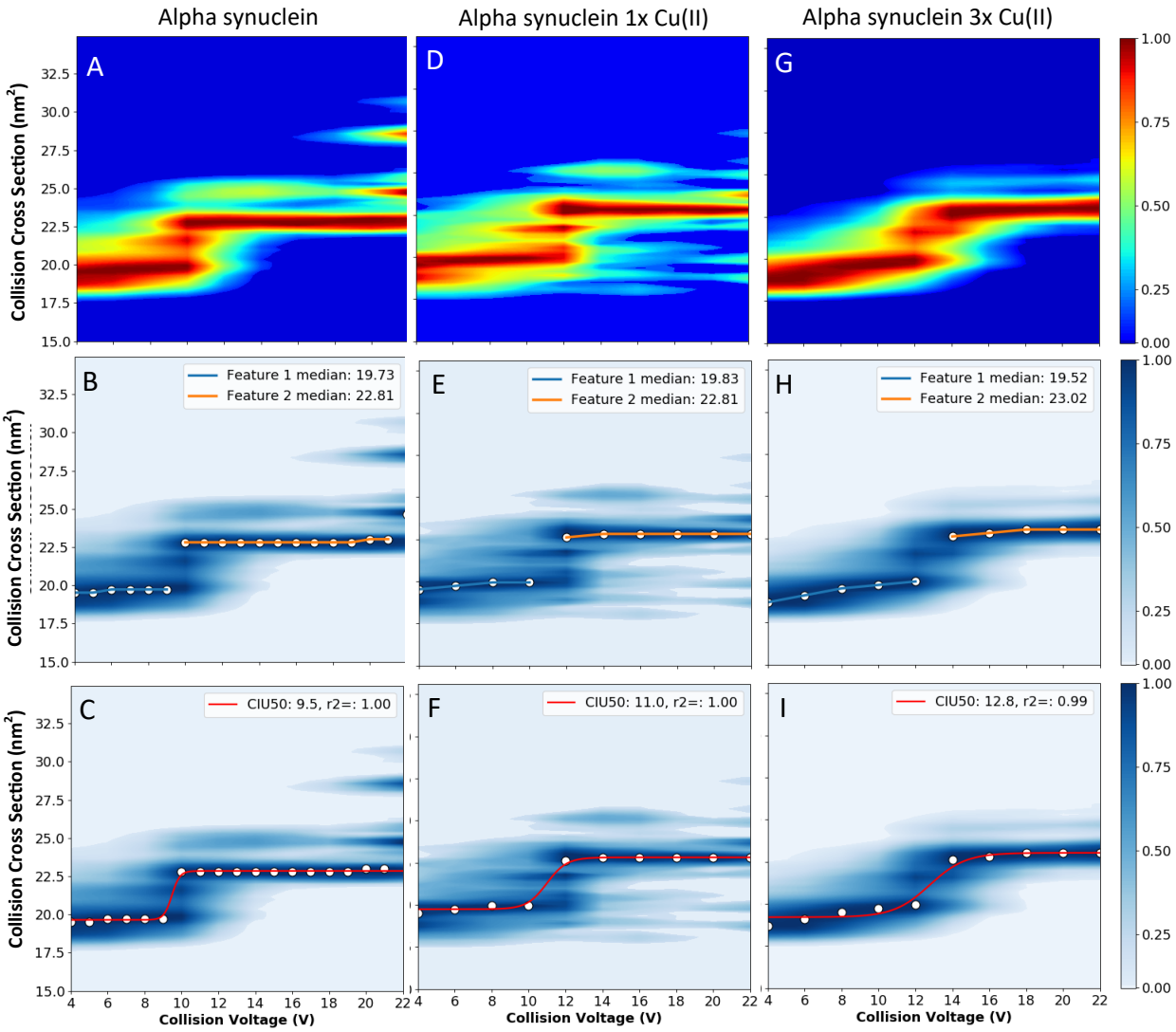


Figure 4.23: CIU fingerprints of +8 charge state αSyn (A), αSyn bound to 1x Cu(II) (D), and αSyn bound to 3x Cu(II) (G) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound protein (B), protein with 1x Cu(II) (E), and protein with 3x Cu(II) (H). CIU50 fit analysis shows midpoint of transitions for unbound protein (C), protein bound to 1x Cu(II) (F), and protein bound to 3x Cu(II) (I). The binding of 1x Cu(II) to the protein has a stabilizing effect, increasing CIU50 by 1.5V. 3x Cu(II) has an even greater stabilizing effect on the protein increasing CIU50 by 3.3V.

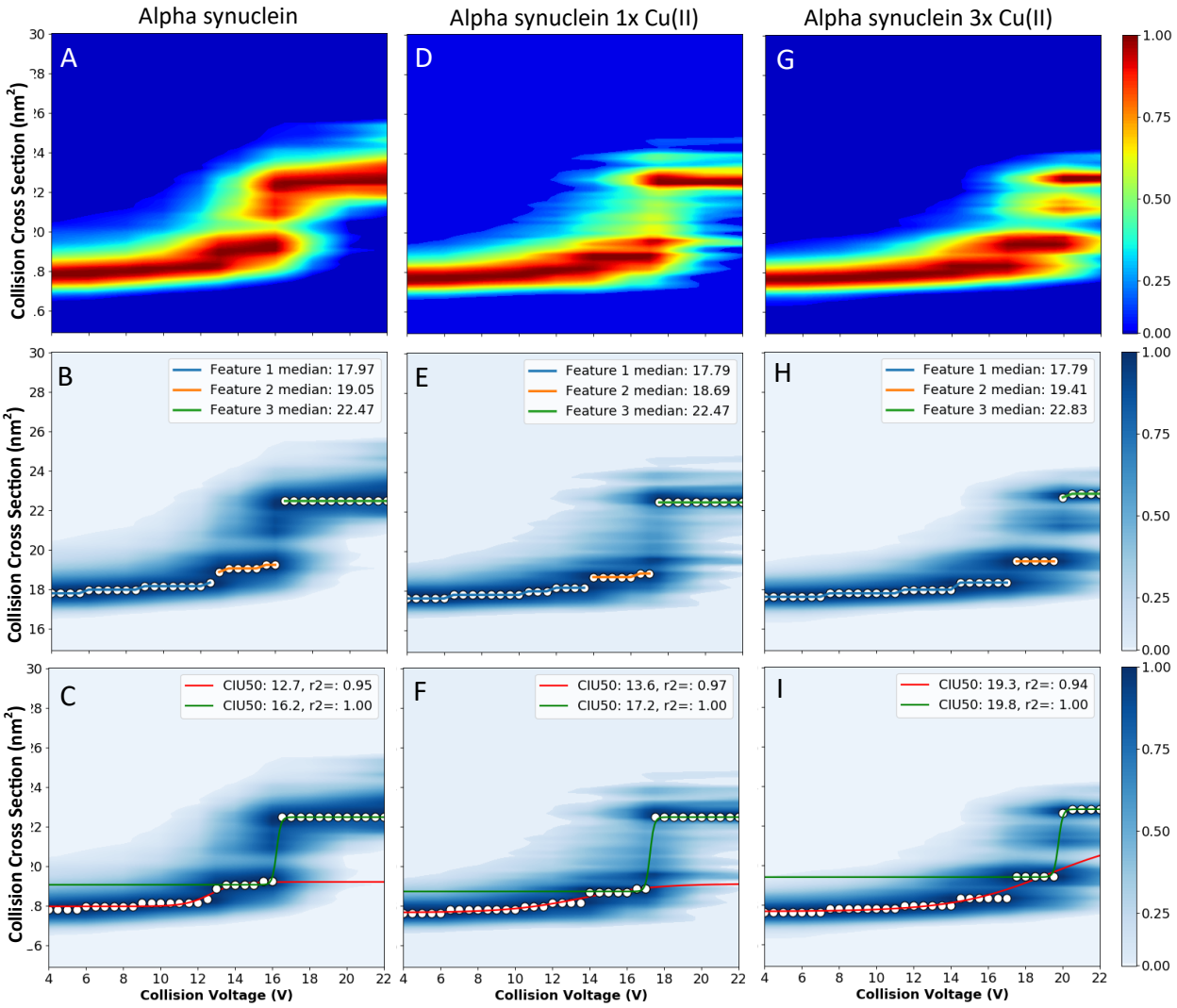


Figure 4.24: CIU fingerprints of +7 charge state αSyn (A), αSyn bound to 1x Cu(II) (D), and αSyn bound to 3x Cu(II) (G) with collision voltage ranging from 4V to 22V. Feature detected in fingerprints for unbound protein (B), protein with 1x Cu(II) (E), and protein with 3x Cu(II) (H). CIU50 fit analysis shows midpoint of transitions for unbound protein (C), protein bound to 1x Cu(II) (F), and protein bound to 3x Cu(II) (I). The binding of 1x Cu(II) to the protein has a stabilizing effect, increasing CIU50 by 1.0V. 3x Cu(II) has an even greater stabilizing effect on the protein increasing CIU50 by 3.6V.

4.6 References

(1) Parkinson, J. An Essay on the Shaking Palsy. *J. Neuropsychiatry Clin. Neurosci.* **2002**, *14* (2), 223–236. https://doi.org/10.1176/jnp.14.2.223.

(2) Spillantini, M. G.; Schmidt, M. L.; Lee, V. M.-Y.; Trojanowski, J. Q.; Jakes, R.; Goedert, M. α-Synuclein in Lewy Bodies. *Nature* **1997**, *388* (6645), 839–840. https://doi.org/10.1038/42166.

(3) Mukaetova-Ladinska, E. B.; McKeith, I. G. Pathophysiology of Synuclein Aggregation in Lewy Body Disease. *Mech. Ageing Dev.* **2006**, *127* (2), 188–202. https://doi.org/https://doi.org/10.1016/j.mad.2005.09.014.

(4) Krüger, R.; Kuhn, W.; Müller, T.; Woitalla, D.; Graeber, M.; Kösel, S.; Przuntek, H.; Epplen, J. T.; Schols, L.; Riess, O. AlaSOPro Mutation in the Gene Encoding α-Synuclein in Parkinson's Disease. *Nat. Genet.* **1998**, *18* (2), 106–108. https://doi.org/10.1038/ng0298-106.

(5) Nussbaum, R. L.; Polymeropoulos, M. H. Genetics of Parkinson's Disease. *Hum. Mol. Genet.* **1997**, *6* (10), 1687–1691. https://doi.org/10.1093/hmg/6.10.1687.

(6) Chartier-Harlin, M.-C.; Kachergus, J.; Roumier, C.; Mouroux, V.; Douay, X.; Lincoln, S.; Levecque, C.; Larvor, L.; Andrieux, J.; Hulihan, M.; Waucquier, N.; Defebvre, L.; Amouyel, P.; Farrer, M.; Destée, A. α-Synuclein Locus Duplication as a Cause of Familial Parkinson's Disease. *Lancet* **2004**, *364* (9440), 1167–1169. https://doi.org/https://doi.org/10.1016/S0140-6736(04)17103-1.

(7) Singleton, A. B.; Farrer, M.; Johnson, J.; Singleton, A.; Hague, S.; Kachergus, J.; Hulihan, M.; Peuralinna, T.; Dutra, A.; Nussbaum, R.; Lincoln, S.; Crawley, A.; Hanson, M.; Maraganore, D.; Adler, C.; Cookson, M. R.; Muenter, M.; Baptista, M.; Miller, D.; Blancato, J.; Hardy, J.; Gwinn-Hardy, K. α-Synuclein Locus Triplication Causes Parkinson's Disease. *Science (80-.).* **2003**, *302* (5646), 841. https://doi.org/10.1126/science.1090278.

(8) Grazia Spillantini, M.; Anthony Crowther, R.; Jakes, R.; Cairns, N. J.; Lantos, P. L.; Goedert, M. Filamentous α-Synuclein Inclusions Link Multiple System Atrophy with Parkinson's Disease and Dementia with Lewy Bodies. *Neurosci. Lett.* **1998**, *251* (3), 205–208.

https://doi.org/https://doi.org/10.1016/S0304-3940(98)00504-7.

(9) Wakabayashi, K.; Yoshimoto, M.; Tsuji, S.; Takahashi, H. α-Synuclein Immunoreactivity in Glial Cytoplasmic Inclusions in Multiple System Atrophy. *Neurosci. Lett.* **1998**, *249* (2), 180–182. https://doi.org/https://doi.org/10.1016/S0304-3940(98)00407-8.

(10) Maroteaux, L.; Campanelli, J. T.; Scheller, R. H. Synuclein: A Neuron-Specific Protein Localized to the Nucleus and Presynaptic Nerve Terminal. *J. Neurosci.* **1988**, *8* (8), 2804 LP – 2815. https://doi.org/10.1523/JNEUROSCI.08-08-02804.1988.

(11) Burré, J.; Sharma, M.; Tsetsenis, T.; Buchman, V.; Etherton, M. R.; Südhof, T. C. α-Synuclein Promotes SNARE-Complex Assembly in Vivo and in Vitro. *Science (80-.).* **2010**, *329* (5999), 1663 LP – 1667. https://doi.org/10.1126/science.1195227.

(12) Masliah, E.; Rockenstein, E.; Veinbergs, I.; Mallory, M.; Hashimoto, M.; Takeda, A.; Sagara, Y.; Sisk, A.; Mucke, L. Dopaminergic Loss and Inclusion Body Formation in α-Synuclein Mice: Implications for Neurodegenerative Disorders. *Science (80-.).* **2000**, *287* (5456), 1265 LP – 1269. https://doi.org/10.1126/science.287.5456.1265.

(13) Lee, S. J. C.; Lee, J. W.; Choi, T. S.; Jin, K. S.; Lee, S.; Ban, C.; Kim, H. I. Probing Conformational Change of Intrinsically Disordered α-Synuclein to Helical Structures by Distinctive Regional Interactions with Lipid Membranes. *Anal. Chem.* **2014**, *86* (3), 1909–1916. https://doi.org/10.1021/ac404132g.

(14) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Mapping Long-Range Interactions in α-Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005**, *127* (2), 476–477. https://doi.org/10.1021/ja044834j.

(15) Uversky, V. N. A Protein-Chameleon: Conformational Plasticity of α-Synuclein, a Disordered Protein Involved in Neurodegenerative Disorders. *J. Biomol. Struct. Dyn.* **2003**, *21* (2), 211–234. https://doi.org/10.1080/07391102.2003.10506918.

(16) Li, B.; Ge, P.; Murray, K. A.; Sheth, P.; Zhang, M.; Nair, G.; Sawaya, M. R.; Shin, W. S.; Boyer, D. R.; Ye, S.; Eisenberg, D. S.; Zhou, Z. H.; Jiang, L. Cryo-EM of Full-Length α-

Synuclein Reveals Fibril Polymorphs with a Common Structural Kernel. *Nat. Commun.* **2018**, *9* (1), 3609. https://doi.org/10.1038/s41467-018-05971-2.

(17) Schweighauser, M.; Shi, Y.; Tarutani, A.; Kametani, F.; Murzin, A. G.; Ghetti, B.; Matsubara, T.; Tomita, T.; Ando, T.; Hasegawa, K.; Murayama, S.; Yoshida, M.; Hasegawa, M.; Scheres, S. H. W.; Goedert, M. Structures of α-Synuclein Filaments from Multiple System Atrophy. *Nature* **2020**, *585* (7825), 464–469. https://doi.org/10.1038/s41586-020-2317-6.

(18) Rodriguez, J. A.; Ivanova, M. I.; Sawaya, M. R.; Cascio, D.; Reyes, F. E.; Shi, D.; Sangwan, S.; Guenther, E. L.; Johnson, L. M.; Zhang, M.; Jiang, L.; Arbing, M. A.; Nannenga, B. L.; Hattne, J.; Whitelegge, J.; Brewster, A. S.; Messerschmidt, M.; Boutet, S.; Sauter, N. K.; Gonen, T.; Eisenberg, D. S. Structure of the Toxic Core of α-Synuclein from Invisible Crystals. *Nature* **2015**, *525* (7570), 486–490. https://doi.org/10.1038/nature15368.

(19) Zhao, M.; Cascio, D.; Sawaya, M. R.; Eisenberg, D. Structures of Segments of α-Synuclein Fused to Maltose-Binding Protein Suggest Intermediate States during Amyloid Formation. *Protein Sci.* **2011**, *20* (6), 996–1004. https://doi.org/10.1002/pro.630.

(20) Heise, H.; Hoyer, W.; Becker, S.; Andronesi, O. C.; Riedel, D.; Baldus, M. Molecular-Level Secondary Structure, Polymorphism, and Dynamics of Full-Length α-Synuclein Fibrils Studied by Solid-State NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (44), 15871 LP – 15876. https://doi.org/10.1073/pnas.0506109102.

(21) Jao, C. C.; Der-Sarkissian, A.; Chen, J.; Langen, R. Structure of Membrane-Bound α-Synuclein Studied by Site-Directed Spin Labeling. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (22), 8331 LP – 8336. https://doi.org/10.1073/pnas.0400553101.

(22) Li, J.; Uversky, V. N.; Fink, A. L. Conformational Behavior of Human α-Synuclein Is Modulated by Familial Parkinson's Disease Point Mutations A30P and A53T. *Neurotoxicology* **2002**, *23* (4), 553–567. https://doi.org/https://doi.org/10.1016/S0161-813X(02)00066-9.

(23) Uéda, K.; Fukushima, H.; Masliah, E.; Xia, Y.; Iwai, A.; Yoshimoto, M.; Otero, D. A.; Kondo, J.; Ihara, Y.; Saitoh, T. Molecular Cloning of CDNA Encoding an Unrecognized Component of Amyloid in Alzheimer Disease. *Proc. Natl. Acad. Sci.* **1993**, *90* (23), 11282 LP – 11286.

https://doi.org/10.1073/pnas.90.23.11282.

(24) Giasson, B. I.; Murray, I. V. J.; Trojanowski, J. Q.; Lee, V. M.-Y. A Hydrophobic Stretch of 12 Amino Acid Residues in the Middle of α-Synuclein Is Essential for Filament Assembly. *J. Biol. Chem.* **2001**, *276* (4), 2380–2386. https://doi.org/10.1074/jbc.M008919200.

(25) Ehrnhoefer, D. E.; Bieschke, J.; Boeddrich, A.; Herbst, M.; Masino, L.; Lurz, R.; Engemann, S.; Pastore, A.; Wanker, E. E. EGCG Redirects Amyloidogenic Polypeptides into Unstructured, off-Pathway Oligomers. *Nat. Struct. Mol. Biol.* **2008**, *15* (6), 558–566. https://doi.org/10.1038/nsmb.1437.

(26) Singh, P. K.; Kotia, V.; Ghosh, D.; Mohite, G. M.; Kumar, A.; Maji, S. K. Curcumin Modulates α-Synuclein Aggregation and Toxicity. *ACS Chem. Neurosci.* **2013**, *4* (3), 393–407. https://doi.org/10.1021/cn3001203.

(27) Tucker, W. C.; Edwardson, J. M.; Bai, J.; Kim, H.-J.; Martin, T. F. J.; Chapman, E. R. Identification of Synaptotagmin Effectors via Acute Inhibition of Secretion from Cracked PC12 Cells . *J. Cell Biol.* **2003**, *162* (2), 199–209. https://doi.org/10.1083/jcb.200302060.

(28) Li, J.; Zhu, M.; Rajamani, S.; Uversky, V. N.; Fink, A. L. Rifampicin Inhibits α-Synuclein Fibrillation and Disaggregates Fibrils. *Chem. Biol.* **2004**, *11* (11), 1513–1521. https://doi.org/https://doi.org/10.1016/j.chembiol.2004.08.025.

(29) Prabhudesai, S.; Sinha, S.; Attar, A.; Kotagiri, A.; Fitzmaurice, A. G.; Lakshmanan, R.; Ivanova, M. I.; Loo, J. A.; Klärner, F.-G.; Schrader, T.; Stahl, M.; Bitan, G.; Bronstein, J. M. A Novel "Molecular Tweezer" Inhibitor of α-Synuclein Neurotoxicity in Vitro and in Vivo. *Neurotherapeutics* **2012**, *9* (2), 464–476. https://doi.org/10.1007/s13311-012-0105-1.

(30) Fokkens, M.; Schrader, T.; Klärner, F.-G. A Molecular Tweezer for Lysine and Arginine. *J. Am. Chem. Soc.* **2005**, *127* (41), 14415–14421. https://doi.org/10.1021/ja052806a.

(31) Acharya, S.; Safaie, B. M.; Wongkongkathep, P.; Ivanova, M. I.; Attar, A.; Klärner, F.-G.; Schrader, T.; Loo, J. A.; Bitan, G.; Lapidus, L. J. Molecular Basis for Preventing α-Synuclein Aggregation by a Molecular Tweezer. *J. Biol. Chem.* **2014**, *289* (15), 10727–10737. https://doi.org/10.1074/jbc.M113.524520.

(32) Uversky, V. N.; Li, J.; Fink, A. L. Metal-Triggered Structural Transformations, Aggregation, and Fibrillation of Human α-Synuclein: A POSSIBLE MOLECULAR LINK BETWEEN PARKINSON'S DISEASE AND HEAVY METAL EXPOSURE. *J. Biol. Chem.* **2001**, *276* (47), 44284–44296. https://doi.org/10.1074/jbc.M105343200.

(33) Binolfi, A.; Rasia, R. M.; Bertoncini, C. W.; Ceolin, M.; Zweckstetter, M.; Griesinger, C.; Jovin, T. M.; Fernández, C. O. Interaction of α-Synuclein with Divalent Metal Ions Reveals Key Differences: A Link between Structure, Binding Specificity and Fibrillation Enhancement. *J. Am. Chem. Soc.* **2006**, *128* (30), 9893–9901. https://doi.org/10.1021/ja0618649.

(34) Brown, D. R. Metal Binding to Alpha-Synuclein Peptides and Its Contribution to Toxicity. *Biochem. Biophys. Res. Commun.* **2009**, *380* (2), 377–381. https://doi.org/https://doi.org/10.1016/j.bbrc.2009.01.103.

(35) Moriarty, G. M.; Minetti, C. A. S. A.; Remeta, D. P.; Baum, J. A Revised Picture of the Cu(II)−α-Synuclein Complex: The Role of N-Terminal Acetylation. *Biochemistry* **2014**, *53* (17), 2815–2817. https://doi.org/10.1021/bi5003025.

(36) Wongkongkathep, P.; Han, J. Y.; Choi, T. S.; Yin, S.; Kim, H. I.; Loo, J. A. Native Top-Down Mass Spectrometry and Ion Mobility MS for Characterizing the Cobalt and Manganese Metal Binding of α-Synuclein Protein. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (9), 1870–1880. https://doi.org/10.1021/jasms.8b05906.

(37) Benesch, J. L. P.; Aquilina, J. A.; Ruotolo, B. T.; Sobott, F.; Robinson, C. V. Tandem Mass Spectrometry Reveals the Quaternary Organization of Macromolecular Assemblies. *Chem. Biol.* **2006**, *13* (6), 597–605. https://doi.org/https://doi.org/10.1016/j.chembiol.2006.04.006.

(38) Hyung, S.-J.; Robinson, C. V; Ruotolo, B. T. Gas-Phase Unfolding and Disassembly Reveals Stability Differences in Ligand-Bound Multiprotein Complexes. *Chem. Biol.* **2009**, *16* (4), 382–390. https://doi.org/https://doi.org/10.1016/j.chembiol.2009.02.008.

(39) Allison, T. M.; Reading, E.; Liko, I.; Baldwin, A. J.; Laganowsky, A.; Robinson, C. V. Quantifying the Stabilizing Effects of Protein–Ligand Interactions in the Gas Phase. *Nat. Commun.* **2015**, *6* (1), 8551. https://doi.org/10.1038/ncomms9551.

(40) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome C. *J. Am. Chem. Soc.* **1997**, *119* (9), 2240–2248. https://doi.org/10.1021/ja9619059.

(41) Han, X.; Jin, M.; Breuker, K.; McLafferty, F. W. Extending Top-Down Mass Spectrometry to Proteins with Masses Greater Than 200 Kilodaltons. *Science (80-.).* **2006**, *314* (5796), 109 LP – 112. https://doi.org/10.1126/science.1128868.

(42) Xie, Y.; Zhang, J.; Yin, S.; Loo, J. A. Top-Down ESI-ECD-FT-ICR Mass Spectrometry Localizes Noncovalent Protein-Ligand Binding Sites. *J. Am. Chem. Soc.* **2006**, *128* (45), 14432–14433. https://doi.org/10.1021/ja063197p.

(43) Eschweiler, J. D.; Martini, R. M.; Ruotolo, B. T. Chemical Probes and Engineered Constructs Reveal a Detailed Unfolding Mechanism for a Solvent-Free Multidomain Protein. *J. Am. Chem. Soc.* **2017**, *139* (1), 534–540. https://doi.org/10.1021/jacs.6b11678.

(44) Zhong, Y.; Han, L.; Ruotolo, B. T. Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution. *Angew. Chemie Int. Ed.* **2014**, *53* (35), 9209–9212. https://doi.org/https://doi.org/10.1002/anie.201403784.

(45) Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T. Collision Induced Unfolding of Isolated Proteins in the Gas Phase: Past, Present, and Future. *Curr. Opin. Chem. Biol.* **2018**, *42*, 93–100. https://doi.org/https://doi.org/10.1016/j.cbpa.2017.11.010.

(46) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. Ion Mobility–Mass Spectrometry Analysis of Large Protein Complexes. *Nat. Protoc.* **2008**, *3* (7), 1139–1152. https://doi.org/10.1038/nprot.2008.78.

(47) Thalassinos, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H. Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry. *Anal. Chem.* **2009**, *81* (1), 248–254. https://doi.org/10.1021/ac801916h.

(48) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V; Ruotolo, B. T. Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for

Gas-Phase Structural Biology. *Anal. Chem.* **2010**, *82* (22), 9557–9565. https://doi.org/10.1021/ac1022953.

(49) Haynes, S. E.; Polasky, D. A.; Dixit, S. M.; Majmudar, J. D.; Neeson, K.; Ruotolo, B. T.; Martin, B. R. Variable-Velocity Traveling-Wave Ion Mobility Separation Enhancing Peak Capacity for Data-Independent Acquisition Proteomics. *Anal. Chem.* **2017**, *89* (11), 5669–5672. https://doi.org/10.1021/acs.analchem.7b00112.

(50) Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T. CIUSuite 2: Next-Generation Software for the Analysis of Gas-Phase Protein Unfolding Data. *Anal. Chem.* **2019**, *91* (4), 3147–3155. https://doi.org/10.1021/acs.analchem.8b05762.

APPENDIX TO THE DISSERTATION

ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms Research Article

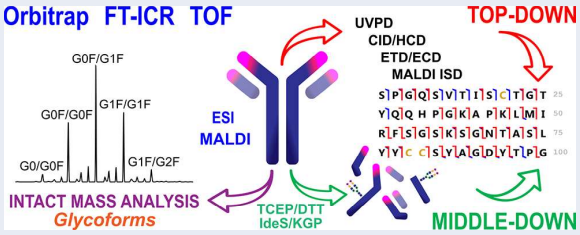
Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry

Kristina Srzentić,[†] Luca Fornelli,[†] Yury O. Tsybin,* Joseph A. Loo, Henrique Seckler, Jeffrey N. Agar, Lissa C. Anderson, Dina L. Bai, Alain Beck, Jennifer S. Brodbelt, Yuri E. M. van der Burgt, Julia Chamot-Rooke, Sneha Chatterjee, Yunqiu Chen, David J. Clarke, Paul O. Danis, Jolene K. Diedrich, Robert A. D'Ippolito, Mathieu Dupré, Natalia Gasilova, Ying Ge, Young Ah Goo, David R. Goodlett, Sylvester Greer, Kim F. Haselmann, Lidong He, Christopher L. Hendrickson, Joshua D. Hinkle, Matthew V. Holt, Sam Hughes, Donald F. Hunt, Neil L. Kelleher, Anton N. Kozhinov, Ziqing Lin, Christian Malosse, Alan G. Marshall, Laure Menin, Robert J. Millikin, Konstantin O. Nagornov, Simone Nicolardi, Ljiljana Paša-Tolić, Stuart Pengelley, Neil R. Quebbemann, Anja Resemann, Wendy Sandoval, Richa Sarin, Nicholas D. Schmitt, Jeffrey Shabanowitz, Jared B. Shaw, Michael R. Shortreed, Lloyd M. Smith, Frank Sobott, Detlev Suckau, Timothy Toby, Chad R. Weisbrod, Norelle C. Wildburger, John R. Yates, III, Sung Hwan Yoon, Nicolas L. Young, and Mowei Zhou

Cite This: *J. Am. Soc. Mass Spectrom.* 2020, 31, 1783−1802 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The Consortium for Top-Down Proteomics (www.topdownproteomics.org) launched the present study to assess the current state of top-down mass spectrometry (TD MS) and middle-down mass spectrometry (MD MS) for characterizing monoclonal antibody (mAb) primary structures, including their modifications. To meet the needs of the rapidly growing therapeutic antibody market, it is important to develop analytical strategies to characterize the heterogeneity of a therapeutic product's primary structure accurately and reproducibly. The major objective of the present study is to determine whether current TD/MD MS technologies and protocols can add value to the more commonly employed bottom-up (BU) approaches with regard to confirming protein integrity, sequencing variable domains, avoiding artifacts, and revealing modifications and their locations. We also aim to gather information on the common TD/MD MS methods and practices in the field. A panel of three mAbs was selected and centrally provided to 20 laboratories worldwide for the analysis: Sigma mAb standard (SiLuLite), NIST mAb standard, and the therapeutic mAb Herceptin (trastuzumab). Various MS instrument platforms and ion dissociation techniques were employed. The present study confirms that TD/MD MS tools are available in laboratories worldwide and provide complementary information to the BU approach that can be crucial for comprehensive mAb characterization. The current limitations, as well as possible solutions to overcome them, are also outlined. A primary limitation revealed by the results of the present study is that the expert knowledge in both experiment and data analysis is indispensable to practice TD/MD MS.

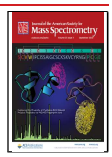


KEYWORDS: *Therapeutic protein, glycoform, intact mass measurement, tandem mass spectrometry, MS/MS, Fourier transform mass spectrometry, FTMS*

INTRODUCTION

The complete characterization of protein therapeutics (amino acid sequence, clipping or truncation, glycosylation profiling, disulfide bonding patterns, secondary and higher order structure, etc.)[1] and associated impurities (e.g., host cell proteins) are a major concern for the pharmaceutical and biotechnology industries.[2] The structural characterization of monoclonal antibodies (mAbs), both as therapeutics and as reagents, is an essential part of their production and regulatory approval.[3] It could be envisioned that the in-depth characterization of a protein drug in the future will be as complete and rapid as it is for small molecule drugs today, but there are no

Received: February 1, 2020
Revised: July 31, 2020
Accepted: July 31, 2020
Published: July 31, 2020



https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

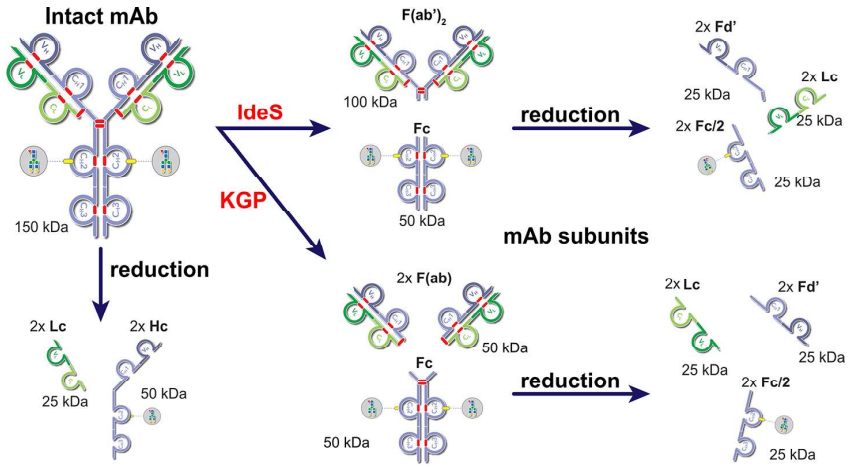


Figure 1. Structural organization of monoclonal antibodies (mAbs) of immunoglobulins G (IgG1) type and enzymatically/chemically assisted structure-specific generation of mAb subunits (25, 50, and 100 kDa). Highly specific enzymes considered here are IdeS (FabRICATOR) and KGP (GingisKHAN).

standard protocols or technologies (yet). Considering its importance in small molecule characterization, high-performance mass spectrometry is likely to play a prominent role in this endeavor. The recent progress in development, industrial, and regulatory acceptance of the multiattribute method (MAM) that includes MS procedures for mAb characterization is an example.[4−8] MAM has been developed to monitor and quantify the array of post-translational modifications (PTMs) found on biotherapeutic molecules during product characterization, in-process control, Good Manufacturing Practice (GMP) release and stability testing steps.[4]

The market for therapeutic proteins, specifically the monoclonal antibody (mAb) market, has significantly increased in the past decade (7−18% growth each year).[9] mAb sales exceeded 98 billion USD as of December 2017 and are to grow to 130−200 billion USD by 2022.[9,10] In 2019 alone, the US Food and Drug Administration (FDA) Center for Drug Evaluation and Research approved three mAbs, ten biosimilars, one nanobody, one single chain Fv (scFv), and three antibody-drug conjugates (ADCs) as drugs (www.antibodysociety.org).[11,12] At least 79 novel antibodies are under investigation in late-stage clinical studies.[13]

Regulatory agencies (e.g., FDA and Pharmacopeia) require thorough characterization of all new drug products before their approval. However, compared to characterizing small molecules, therapeutic mAbs present distinct challenges for the analytical laboratory. In addition to their large size—on the order of 150 kDa for a full-sized mAb—the presence of post-translational modifications (PTMs), such as glycosylation, oxidation, and deamidation, add to the structural complexity and heterogeneity of mAbs.[14] Such PTMs can be referred to as critical quality attributes (CQAs) and may occur during production, purification, or storage.[15] Because biotherapeutics are primarily produced in recombinant expression systems (i.e., Chinese hamster ovary (CHO) cells), the final product is heterogeneous. Changes in these systems may manifest as variation in number and distribution of proteoforms,[16] alter efficacy and binding characteristics, or be immunogenic.[17] As a result, comprehensive structural characterization, from primary structure to proteoform elucidation of therapeutic mAbs, is essential.[2,3,18,19]

Mass spectrometry (MS) is a powerful tool for protein characterization.[2,20,21] Two major methodologies that are commonly employed in MS-based protein/proteome analysis are top-down (TD) and bottom-up (BU). The BU approach utilizes proteases or chemical means to cleave proteins at backbone sites to generate smaller, more readily ionizable peptides that are then further fragmented in the gas phase via tandem mass spectrometry (MS/MS) to elucidate sequence and delineate the presence of modifications.[22−26] TD MS strategies eschew the use of proteolytic enzymes, providing the intact masses of the molecules being studied, which in turn can determine the presence of multiple proteoforms.[27] MS/MS fragmentation of the intact proteins and large protein subunits can provide amino acid resolution for interpretation of sequence heterogeneity and presence of PTMs.[21,27,28]

Currently, standard workflows for mAb structural characterization are centered around peptide mapping—analysis of trypsin-derived peptides of 0.6−2 kDa size.[4,5] However, this trypsin-based BU approach does not provide 100% sequence coverage for all mAbs. For example, the blockbuster biotherapeutic, Humira (adalimumab, the world's number 1 drug in sales value), has a very long, more than 50 amino acids, sequence stretch without lysine or arginine residues, which are the specific cleavage sites of trypsin. Therefore, this region of adalimumab may remain unaccounted for (invisible) in a trypsin-based BU approach. To achieve complete sequence coverage and improve PTM characterization, a BU approach employing multiple enzymes is usually required. The additional enzymes are selected to provide cleavage specificity or cleavage frequency complementary to that of trypsin. Examples include chymotrypsin, Lys-C, Glu-C, Asp-N, and Sap9.[23,26,29,30] However, alkylation, followed by digestion with trypsin and other enzymes, entails reaction conditions that can increase the likelihood of artifacts, such as amino acid isomerization, deamidation, or oxidation.[26,31]

TD MS-associated technologies have advanced greatly over the past few decades.[32,33] Since the development of electrospray ionization (ESI), advances in TD MS have focused on two areas—instrumentation and fragmentation approaches.[21,34] Advances, such as improved mass accuracy and resolving power, came with the development of mass analyzers,

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

such as the time-of-flight (TOF), Fourier transform ion cyclotron resonance (FT-ICR), and Orbitrap.[35−37] Newer ion activation/dissociation methods such as electron capture/transfer dissociation (ECD, ETD),[38,39] ultraviolet photodissociation (UVPD),[40] and more efficient matrix-assisted laser desorption ionization (MALDI) matrices for in-source decay (ISD) fragmentation[41,42] drastically increase the sequence and PTM information that can be obtained in TD MS experiments. These technological advances have been supported by the corresponding TD/MD-specific developments in data analysis tools, such as deconvolution of highly convoluted product ion distributions and product ion annotation approaches.[43−47] As a result, TD MS/proteomics has emerged as a powerful tool in basic, translational, and clinical research, not only for protein identification but also for large-scale proteoform elucidation.

Unlike BU approaches, TD MS offers extensive, if not complete, sequence coverage and proteoform mapping in a single experiment, and relies on sample handling protocols minimizing the introduction of artifactual modifications (e.g., oxidation and deamidation). Among other virtues, proteoform mapping provides valuable information on the integrity of a mAb, informing about structural integrity of a whole mAb in solution, which is inherently lost in the BU approach.[48] However, for proteins of 150 kDa size, overall sequence coverage based on current TD MS technologies remains incomplete.[49−53] Therefore, reducing the size of the target proteins through reduction of mAbs into their heavy and light chain (Hc and Lc) subunits is often required to provide more detailed information, Figure 1.[33,54−57]

Furthermore, artifact-minimizing enzymatic processing of intact mAbs into 25−50 kDa subunits by use of structure-specific proteases, for example, IdeS and KGP, has recently received attention.[58−60] Separating mAbs by backbone cleavage above (KGP) or below (IdeS) the hinge region yields a monovalent antigen-binding fragment (Fab) or a bivalent F(ab′)2 subunit, respectively, and a fragment crystallizable (Fc) glycosylated subunit, Figure 1. Further size reduction of large F(ab) and $F(ab')_2$ subunits is typically accomplished with the use of disulfide bond chemical reduction, to yield smaller subunits: Lc, Fd′ (the N-terminal part of an Hc), and Fc/2 (the C-terminal part of an Hc). Chemical reduction of S−S bonds can be performed by use of, for example, TCEP or DTT. Naturally, the number of amino acids in the Fd′ and Fc/2 subunits formed with IdeS or KGP enzymes will be different.

Despite the relatively large size of typical mAb subunits (Lc, Hc, F(ab), Fd′, and Fc/2), enzymatic or chemical separation of an intact mAb into these smaller components for MS-based analysis should be referred to as middle-up (MU) MS, and their MS/MS analysis as middle-down (MD) MS.[27] For example, the combination of MU and MD measurements were previously applied to detect the reference sequence errors and curation of the cetuximab and natalizumab sequences.[33,57] Certain MD proteomics approaches may utilize even more extensive protein cleavage specificity, yielding proteolytic peptides in the range of 3−7 kDa,[25,61,62] as demonstrated for mAbs analysis.[26]

To assess the added-value of the current technologies and protocols offered by TD/MD MS for characterizing mAbs compared to BU MS, the present study was launched and supervised by the Consortium for Top-Down Proteomics (www.topdownproteomics.org). This study engaged 20 laboratories worldwide. All of the groups who volunteered have some level of expertise in TD/MD MS, but several had limited or no experience with 150 kDa mAb proteins. Three commercially available mAbs (SiLuLite mAb, NIST mAb, and trastuzumab, vide infra) were centrally provided to participants. Each group performed TD/MD MS experiments following their own best practices and approaches on all or some of the provided mAbs. A broad range of MS instruments and ion activation/dissociation techniques was employed. To provide a detailed description of the employed techniques and the results obtained, the current report is supported by Supporting Information. It presents results of data analysis in the form of box-plots (Figures S1−S26) and experimental results (Figures S27−S66), as well as tables with sample information, description of the employed methods and techniques, as well as selected results overview (Tables S1−S16).

Ultimately, the study highlights the current state of TD/MD MS to address the challenges for ensuring the quality of biotechnology medicinal products, their limitations, and where future development is needed. This report and perspective should be of value not only to protein mass spectrometrists who are interested in TD/MD MS and biopharmaceutical scientists currently engaged in characterization of therapeutic proteins but also to students and early career researchers who wish to be educated in the important field of mAb structure characterization.

EXPERIMENTAL SECTION

Initial Antibody Sample Preparation. Three commercially available mAbs (immunoglobulin G or IgG, isotype 1) were provided to participants in equal amounts (from 50 to 100 μg/sample): Sigma mAb standard (SiLuLite, IgG1 lambda, CHO, Sigma); NIST mAb standard (HzIgG1 kappa, NS0, NIST) and Herceptin (trastuzumab, HzIgG1 kappa, CHO, Roche); see Table S1 for mAb sequence information, Table S2 for details about sample preparation and handling, and Table S3 for the results of quality control measurements. Each mAb was provided in three forms for different experimental workflows, Figure 1: (i) intact form for mass measurements and TD analysis of the 150 kDa mAbs; (ii) mAbs digested by use of the highly specific protease KGP (GingisKHAN, Genovis, Lund, Sweden);[63] and (iii) mAbs digested by the IdeS protease (Fabricator, Genovis).[64] In the latter case, the Fc subunit integrity is due to the noncovalent bonds between the two Fc/2 parts. The enzyme/protein ratio for IdeS and KGP digestion corresponded to the manufacturer's suggestions (Genovis). Briefly, one unit of enzyme was added to each microgram of a mAb for IdeS digestion,[58] and two units of enzyme were added to each microgram of a mAb for KGP digestion.[59] Sequence information for each mAb, with the CDR sequences outlined and IdeS/KGP digestion sites highlighted can be found in Table S1. The molecular formulas and calculated masses of modified and unmodified intact mAbs involved in the study and their subunits are presented, together with the details on the atomic masses and abundances employed for the calculations, in Tables S4−S7.

Quality control measurements were performed prior to sample shipping (Table S3 and Figures S27−S29). The samples were shipped on dry ice, after one freeze−thaw cycle, with intact mAbs and their subunits dissolved at high concentration (from 1 to 10 μg/μL in either formulation buffers (intact mAbs) or buffers used for proteolytic digestions

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

(in which case the pH was reduced by addition of formic acid to quench the enzymatic reaction). Note that no reduction of the disulfide bonds was performed before distributing mAbs to participants. The recommended storage conditions were +4 °C for intact mAbs and −20 °C for digests. The participating groups were responsible for sample storage and handling according to their best practices.

Sample Fractionation and Ionization. Participants were provided the opportunity to decide on the best workflow to purify or separate the intact mAbs and their subunits prior to ionization and TD/MD MS analysis.[34] In most cases, each group used a separation method that they had already established in that laboratory for the types of samples with which they were familiar (Table S8). Groups performing online liquid chromatography (LC) or capillary electrophoresis (CE) coupled with ESI did not typically perform sample cleanup prior to sample fractionation/separation. Conversely, off-line desalting and adduct removal was performed by most groups that were either directly infusing mAb samples (with a syringe or nanoESI needles, or by use of a TriVersa Nanomate robot from Advion BioSciences, Ithaca, NY) or spotting them onto a MALDI plate. The most commonly employed desalting techniques were reversed-phase LC followed by sample collection and solid-phase extraction. Details on the experimental sample preparation (including approaches to disulfide bonds reduction) and fractionation (separation) methods and parameters used by each participant can be found in Table S8.

Mass Spectrometry. The groups performed TD/MD MS measurements with different instruments, both commercially available and customized (Table S9). The mass analyzers in the present study are TOF, ICR, and Orbitrap FTMS. TOF MS and FT-ICR MS instruments were coupled with either ESI or MALDI ion sources, whereas Orbitrap FTMS instruments used ESI only. Ion activation and dissociation methods coupled to ESI source included: ETD/ECD, higher energy collisional dissociation/collision induced dissociation (HCD/CID), UVPD, MALDI ISD, and a hybrid electron transfer higher-energy collision induced dissociation (EThcD).[65] Instruments were calibrated and maintained by the use of the best laboratory practices specific to each participating group (data not provided). Details on the experimental MS and MS/MS parameters used by participants can be found in Table S9.

Data Analysis. Complete freedom was left to the participants regarding deconvolution of both nonisotopically and isotopically resolved TD/MD MS spectra, as well as calculation of masses based on known elemental composition and product ion assignment (Table S10). The list of deconvolution algorithms specified by participating laboratories included: Xtract, THRASH, ReSpect (found in both a commercial package from Thermo Scientific and in an open resource as implemented, for example in the MASH Suite), SNAP and MaxEnt (Bruker Daltonics),[66] Intact Mass (Protein Metrics), and UniDec (Oxford University, UK).[67] The list of software used for spectral processing, product ion assignment and validation included the commercially available BioPharma Compass (Bruker Daltonics), ProSight PC and BioPharma Finder (Thermo Scientific), TDValidator (Proteinaceous),[68] Peak-by-Peak and AutoVectis (both from Spectroswiss); and the freeware packages MASH Suite Pro,[45] Informed Proteomics suite,[47] ProSight Lite,[44] and ISDetect.[69] Software tools that bypass deconvolution procedures and perform direct matching of isotopic envelopes of multiply charged product ions to simulated profiles included TDValidator, LcMsSpectator (data viewer in Informed Proteomics suite), and AutoVectis. A detailed description of the data processing and analysis tools and parameters employed by participants can be found in Table S10. Notably, mass tolerance for product ion annotation in MS/MS experiments was typically 10 ppm, and sometimes lowered to 5 ppm. Precursor mass tolerances of up to 50 ppm and 10 Da were reported (Table S10).

For the determination of mass accuracy or mass measurement errors (expressed in parts per million, ppm) from deconvolved full MS measurements, results submitted by the participating laboratories were compared to a single set of calculated masses (determined for both intact mAbs and their subunits). Such masses were calculated by the use of monoisotopic and average atomic masses, as well as abundances, as employed in ChemCalc isotopic calculator algorithm and described elsewhere (Table S4).[70] The respective mAb sequences (Table S1) were used to generate molecular formulas further employed for monoisotopic and average mass calculations of mAbs and their subunits (see examples provided by Tables S5−S7, for data on SiLuLite mAb, NIST mAb, and trastuzumab, correspondingly). These computational results were achieved either with the web-version of ChemCalc (www.chemcalc.org) or via the desktop version of the FTMS Isotopic Simulator (part of Peak-by-Peak software, Spectroswiss). Note that Tables S5−S7 include all major proteoforms identified by participants along with unmodified and deglycosylated proteoforms, some of which were not present in the samples but are given here for didactic and self-controlling reasons. The latter is justified by the errors in molecular weight calculations and proteoform misassignment demonstrated in some reports.

Results of the statistical analysis, in the form of box-and-whisker plots, containing errors (accuracy) for MU and intact mass measurements (Figures S1−S13), as well as sequence coverages (Figure S14−S26), are presented in the Supporting Information. It is important to note that, in all box-and-whisker plots, the box indicates the interquartile range, whereas the horizontal line in the box is the median. All plots were generated by use of R. To perform annotation of glycoforms, this report follows the standard nomenclature, as outlined, for example in a comprehensive report on NIST mAb glycosylation profiling.[71] For tandem MS results, sequence coverage values were obtained directly from the participants' reports, without data reprocessing.

Aggregated Analysis of Results. A compilation of sample introduction methods, MS and MS/MS instrumentation and allied approaches employed for the study, as reported by the participants, is depicted in Figure 2 and in allied Tables S11 and S12.

The following brief observations could be made (Figure 2 and Tables S11 and S12): (i) most groups employed LC for online sample purification and separation; (ii) about half of the employed instruments were Orbitraps; (iii) most of the currently available MS/MS methods were employed, from a traditional CID to the less frequent but up-and-coming UVPD and MALDI ISD; (iv) most data processing approaches utilized instrument vendor provided software (with some packages requiring additional licenses); and (v) in most cases sequence maps were visualized by use of additional freeware such as the ProSight Lite tool. In general, most groups relied on multiple instruments, activation methods, or data analysis software to characterize the provided mAb samples. We shall

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

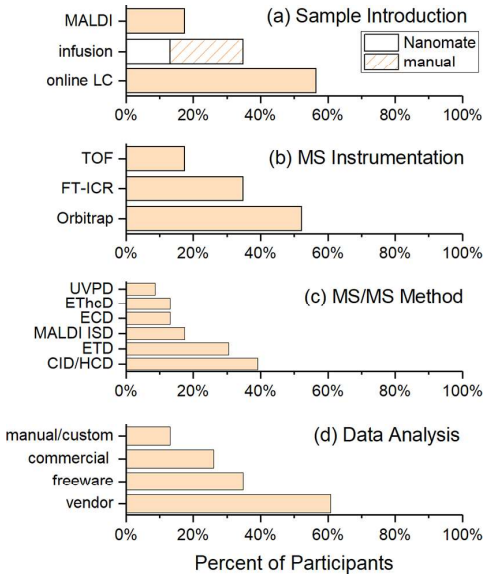


Figure 2. Summary of the experimental methods used by the participants for TD/MD (a total of 20 participants). The percent of participants who have used specific methods for (a) sample introduction; infusion and online LC refer to ESI approaches, (b) MS instrumentation, (c) MS/MS, and (d) data analysis (where "commercial" refers to nonvendor third-party software, and "vendor" are software tools provided by the instrument manufacturers) are shown in bar graphs. Many groups used more than one method.

note that in the current description we do not separate TD/MD MS results generated with CID (resonant collisional activation and dissociation performed in ion traps) and HCD (beam-type higher energy collisional activation and dissociation). Further participating group-specific details of the experimental methods are provided in Tables S8–S10.

Bottom-Up Mass Spectrometry. Standard BU MS approaches, that is, sample handling, mass spectrometry, and data analysis have been employed by several groups, as for example detailed by Smith and co-workers.[72] Details on the employed BU MS approaches, as performed by groups 1, 16, 17, and 24, are given in Tables S8–S10. The corresponding results are summarized in Tables S13 and S15.

RESULTS

The following analytical criteria of interest for mAb development, production, quality control, and release have been considered: (i) protein integrity, structure completeness and heterogeneity; (ii) glycoforms, identity and relative quantitation; (iii) protein sequence coverage; (iv) complementarity-determining regions (CDRs) sequencing degree; and (v) other PTM identity and location. These results are reported separately since each of these criteria represents a different level of mAb structural characterization, that is based on different MS methods to achieve the required information (e.g., MS-only versus MS/MS). Another objective of the present study was to evaluate the ease of use and reproducibility of TD/MD MS technologies in current laboratory practice. In this respect, the study aimed to evaluate the maturity of these novel MS approaches in the analytical laboratories performing mAb structure characterization, and potentially widen their acceptance by contract research organizations (CROs), biopharma companies, and proteomics facilities of research institutions.

Protein Structural Integrity. The structural integrity of mAbs is one of the CQAs in mAb characterization that can reveal potential sources of heterogeneity, including amino acid clipping (truncation) from the C-terminus, modification of N-terminal amino acids, glycosylation, the potential presence of remaining signaling peptides at the N-termini of the Lc and Hc, and overall completeness and stability of mAb's primary structure. Decomposition of an intact mAb into its subunits by chemical and enzymatic digestion, as depicted in Figure 1, facilitates and extends the overall analysis of structural integrity. For example, disulfide bond reduction of a properly assembled intact mAb should result in release of Lc and Hc subunits. The latter may not be the case when mAb structural integrity is not present, which will become apparent by mass measurements of reaction products.

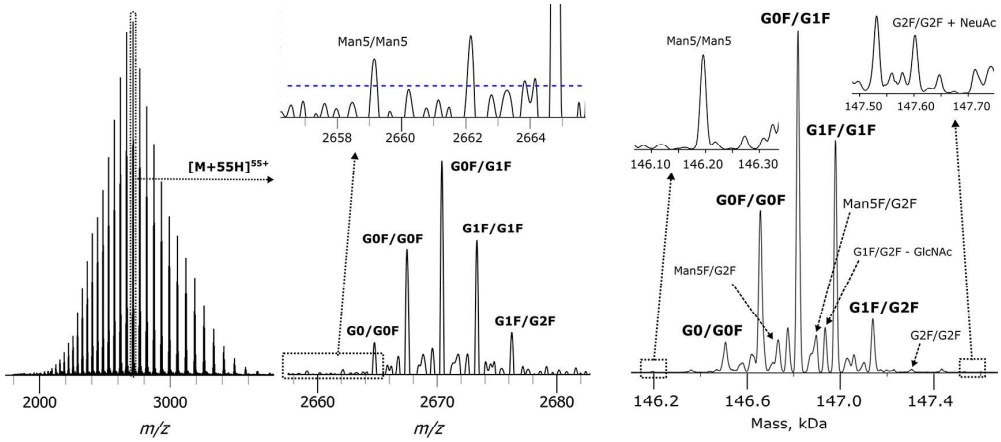


Figure 3. Example of intact mAb average mass measurements: ESI Q Exactive HF Orbitrap FTMS of intact SiLuLite mAb (group 19). Shown are the charge state distribution (left panel), expanded view into a selected charge state (middle panel), and a deconvolved mass spectrum (right panel). Deconvolution performed by use of UniDec software. Glycoform annotation follows standard rules.[71,76] For more details and examples of experimental results, see Figures S30–S34 and Tables S8–S10.

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783–1802

Table 1. Examples of the Absolute Mass Measurement Errors (Mean Value ± a Single Standard Deviation, in ppm) for Analysis of Intact mAbs and Their Subunits Generated via IdeS Digestion with and without Disulfide Bond Reduction[a]

		absolute mass measurement error (ppm)			ref
chain	experimental details	FT-ICR MS	orbitrap FTMS	TOF MS	SI Figure
intact	3 major glycoforms	14.9 ± 11.4	17.4 ± 19.5	35.5 ± 33.8	S1–S3
intact	G0F/G0F	11.3 ± 8.0	20.2 ± 23.2	46.1 ± 48.6	S4–S6
	G0F/G1F	17.2 ± 16.0	16.0 ± 21.7	31.9 ± 29.8	
	G1F/G1F	15.8 ± 8.6	16.0 ± 14.6	28.6 ± 26.7	
subunits (IdeS)	$F(ab')_2$ Fc	6.5 ± 7.8	4.8 ± 3.6	8.2	S13
subunits (IdeS)	all 25 kDa subunits	5.6 ± 6.7	3.3 ± 2.4	1.2 ± 0.6	S7
subunits (IdeS)	Lc/Lc pQ	6.7 ± 7.9	3.0 ± 1.8	1.1 ± 0.4	S8–S12
	Fd′/Fd′ pQ	4.1 ± 4.5	3.5 ± 2.8	1.0 ± 0.5	
	Fc/2 G0F	8.2 ± 9.3	2.7 ± 2.3	1.3 ± 0.4	
	Fc/2 G1F	4.8 ± 5.3	2.2 ± 1.7	1.3 ± 0.8	
	Fc/2 G2F	1.7 ± 0.3	6.1 ± 2.3	1.1 ± 1.0	

[a]Results are grouped by the MS platform employed: FT-ICR MS, Orbitrap FTMS, and TOF MS. Note: Table 1 reports mean values of mass measurement errors, whereas the corresponding reference figures (Figures S1–S13) report median errors.

The structural integrity analysis typically starts with mass measurements at the intact mAb level, which is an important part of TD MS workflows.[27,34] Achieving isotopic resolution of intact mAbs today is a challenge that requires the exceptional performance of ultrahigh resolution mass spectrometers.[73,74] It is thus a common practice to perform mass measurements of intact mAbs at a (moderate) resolution sufficient to determine average masses of mAbs' proteoforms.[75] Attention should be paid to the isotopic mass and abundance table employed for calculation of the mAb's average and monoisotopic masses, which should be listed in corresponding reports (Table S4). For example, average masses for SiLuLite mAb and its subunits as calculated in the current study by use of ChemCalc resource (listed in Table S5) are consistent with the values reported in a follow-up study by Ge and co-workers.[76] A particular benefit of isotopically resolved, and thus higher resolution, mass spectra is disentangling multiple contributions when more than one PTM is present.[76]

Mass measurements of intact mAbs were performed by more than half of the groups participating in this study (Table S12). Overall, the three mAbs analyzed demonstrated excellent protein integrity at the intact mass level, as expected for these samples. Notably, two groups using high-performance FT-ICR MS instruments isotopically resolved the intact mAbs and reported monoisotopic masses. Examples of the FT-ICR mass spectra of the three intact mAbs showing isotopic resolution of the 53+ charge state are shown in Figure S34. To obtain isotopic resolution of these heavy molecules, resolving power exceeding 300 000 at the target m/z was achieved by use of a 12 T FT-ICR MS, which correlates with the published results generated with FT-ICR MS or Orbitrap FTMS instruments.[73,74,76] Other groups relied on lower resolution mass spectra that are not isotopically resolved (Figures S30–S33). The average masses of mAbs' proteoforms were obtained from the charge state envelopes generated by ESI MS (Figure 3) or singly or doubly charged components in MALDI MS (data not shown).[77]

To highlight the variation of reported results, the mass measurement errors (expressed in ppm) are presented in this report by both their mean values (Table 1) and their median values. The latter are represented via box-and-whisker plots showing the mass measurement error distributions (Figures S1–S13).

The mass measurement errors achieved through the determination of the monoisotopic mass were not substantially different from those observed for the groups reporting the average intact mAb mass (Figures S1–S3). Indeed, large protein size is known to result in wide isotopic distributions and difficulties in accurately defining monoisotopic masses. In practice, the monoisotopic mass has to be calculated based on statistical methods (e.g., using the averagine approach) that can lead to the associated errors for such large biomolecules.[77] In addition, post-translational modifications (particularly disulfide bonds), incomplete desolvation, and salt adducts (such as sodium) can introduce overlapping isotopic profiles that further complicate the deconvolution and limit the mass accuracy that can be achieved.[78]

To improve the accuracy of their mass measurement, intact mAbs can be broken down into large, 25–100 kDa, subunits by use of enzymatic or chemical sample processing as depicted in Figure 1. The mass measurements of these subunits usually provide results from which Lc (~25 kDa molecular weight) can be isotopically resolved and the monoisotopic mass determined (Figures S7–S13). Isotopic distributions of the larger Hc (~50 kDa) would often remain unresolved and average mass would be reported (Figures S35–S37). Notably, Group 1 reported TOF MS data with monoisotopic masses calculated for both chains, not only for the Lc (see mass assignment method description in Table S10).

The 25 kDa mAb subunits obtained with structure-specific enzymes, similar to the Lc analysis discussed above, are well suited for high-resolution mass measurements.[54,55,58] Examples of MU MS results obtained for NIST mAb analysis with a TOF MS are presented in Figure 4.

Similarly, examples of MU MS results obtained for SiLuLite mAb analysis with Orbitrap FTMS can be consulted in Figure S40. For more details and other examples of experimental results, see Figures S38–S46.

The KGP/IdeS-digested samples were analyzed as provided (unreduced 100 kDa and 50 kDa subunits), or after reduction (and possibly alkylation) of the cysteine residues involved in disulfide bonds to produce three ~25 kDa subunits per mAb, Figure 1. Most of the groups analyzed the reduced subunits, and only one group alkylated the mAb chains, Table S12. As for the Lc and Hc mass measurements, when 25 kDa mAb subunits (Fd′, Fc/2, and Lc) are generated, monoisotopic masses can be derived (Figures S38–S46), but for the 50 kDa

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783–1802

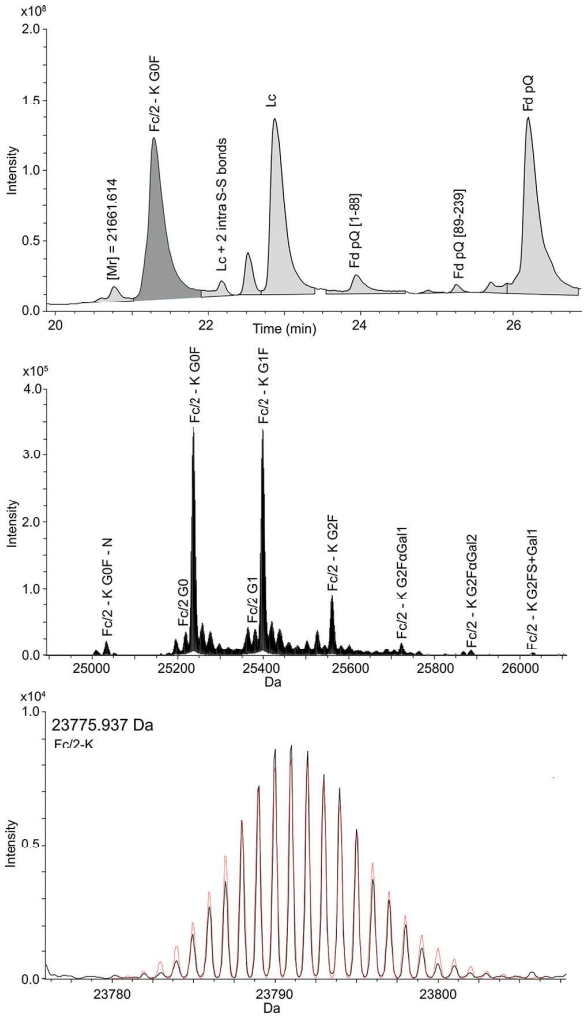


Figure 4. MU approach examples for subunit mAb isotopically resolved mass measurements: ESI maXis II QTOF MS of 25 kDa subunits (Lc, Fd′, and Fc/2) of NIST mAb obtained by IdeS digestion and TCEP reduction of S−S bonds (group 1). Shown are (top panel) LC-MS elution profiles, with the most abundant glycoforms labeled on top of each elution peak; (middle panel) deconvolved mass measurements of glycoforms of Fc/2 subunit; and (bottom panels) example of a deconvolved baseline-resolved isotopic envelope of a glycoform Fc/2-Lys obtained with accurate isotopic distributions (calculated isotopic pattern is overlaid in red) and dynamic range spanning 2 orders of magnitude. The calculated monoisotopic mass value is given in Da. All monoisotopic neutral mass assignments were obtained by the SNAP algorithm following MaxEnt deconvolution. Proteo/glycoforms annotations are as defined in the BioPharma Compass method.[71]

subunits both the average and the monoisotopic masses were deduced (Figures S47−S52).

Box-plots showing the mass measurement errors (expressed in ppm) for the subunits derived from mAb digestion with IdeS are shown in Figures S7−S13. Results are summarized for both reduced subunits (25 kDa) and nonreduced subunits (100 and 50 kDa). For instance, Figure S7 shows the box plots for the 25 kDa subunits (derived from all three mAbs) generated by reduction of the IdeS digestion products, aggregated according to the mass analyzer employed. Note that the mass measurement error was typically less than 1−2 ppm for all three mass analyzers when there was no misassignment of the monoisotopic peaks (vide infra), as exemplified by the median values reported in the box-plots. However, the mean mass measurement error calculated from all the reported values, still shows a signification variation, Table 1.

The average mass determination for the nonreduced (50 and 100 kDa, see Figure 1) subunits produced by IdeS digestion (and particularly the ∼100 kDa F(ab)2 subunit, which was analyzed only by several groups), was achieved with less than 10 ppm mass measurement error (6.5 ppm for SiLuLite mAb, Figure S13), Table 1. Reduction of intramolecular S−S bonds, if it occurs during IdeS digestion, would result in mass shifts that should be incorporated when average masses are calculated for these large subunits. These mass shifts were not considered in this study for results reported in Figure S13.

Notably, the number of participants that analyzed the ∼25 kDa reduced subunits (N = 9 for each antibody, Figures S7−S12) largely exceeds that of the groups that measured the larger, nonreduced subunits and specifically the ∼100 kDa F(ab′)2 (N = 4−5 for each antibody, Figure S13), also see Table S12. On one hand, this difference may indicate that the mass measurements of >100 kDa proteins may not be routinely achievable. On the other hand, mass measurements of intact 150 kDa mAbs have been performed by more groups.

Glycoforms: Identity and Relative Quantitation. N-linked glycans typically represent the most abundant PTMs present in mAbs.[54,76] Because of their important biological and structural role,[79] these complex moieties should be studied at multiple levels of mAb structural organization.[54] Mass measurements of intact mAbs allow the verification of the pairing of N-linked glycans (i.e., by weighing intact mAb glycoforms or proteoforms), as well as the determination of their relative abundances, Figure 3 and box plots in Figures S4−S13. MU MS is usually more sensitive and accurate than TD MS thanks to a decrease both in protein mass and structural complexity because the sources of heterogeneity (Fd′ and Fc/2 subunits in multiple modified forms) can be decoupled. This approach widens the list of glycans identified (Figures 3 and 4), especially if glycosylation occurs at noncanonical sites (e.g., lysine glycation).[80] Glycation in recombinant mAbs refers to the nonenzymatic addition of monosaccharide (typically a hexose) at free amine groups.[81] Unlike the canonical N-glycosylation, glycation is believed to have minimal effect on target binding, but it contributes to sample heterogeneity and is a necessary target in quality control. Coupled with appropriate molecular mass deconvolution procedures, intact (TD) and MU MS are capable of producing extensive lists of glycoforms (Table S14). Here, we differentiate glycoforms similarly to the standard proteoform notation, for example, the same glycosylation modification on two proteins that differ only by the presence of the C-terminal Lys residue would result in two distinct glycoforms (proteoforms).[16]

Figures S4−S6 show the mass measurement errors for each of the top three glycoforms of the intact mAbs, grouped by mass analyzer employed, namely Orbitrap FTMS, FT-ICR MS, and TOF MS. For example, for SiLuLite mAb, the vast majority of the mass measurements showed accuracy of ≤50 ppm (i.e., ≤8 Da) for each of the three most abundant glycoforms, namely G0F/G0F, G0F/G1F, and G1F/G1F

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

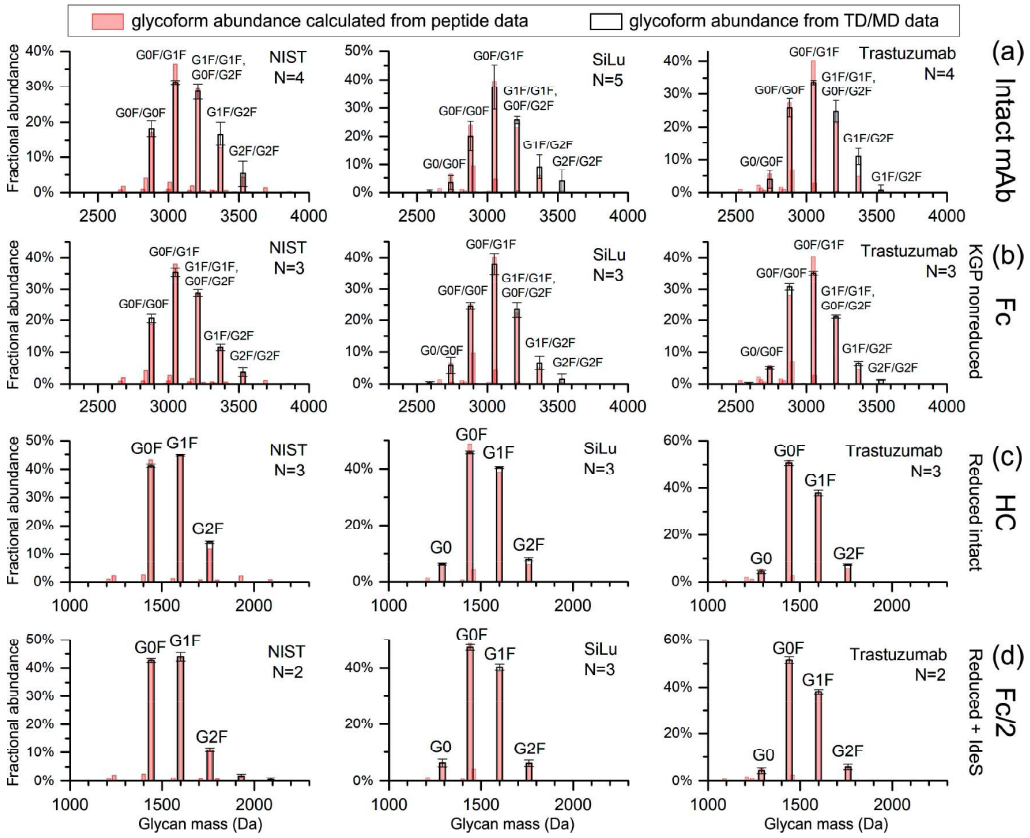


Figure 5. Relative quantitation of major glycoforms at the (a) intact mAb level, (b) Fc from KGP digestion, (c) intact Hc after mAb reduction, and (d) Fc/2 from IdeS digestion with disulfide bond reduction for the 3 mAbs (open bars with black border). The "N" in the legend indicates the number of groups with corresponding TD/MD MS data, and the standard deviations are shown as error bars. The relative abundances were normalized to the selected major glycoforms labeled in the plots. The relative abundances from peptide mapping were overlaid as red bars, which are calculated based on the glycopeptide relative abundances of the conserved Asn300 in Fc and glycation (assuming only one glycation). For data comparison with the intact protein and the Fc region (contains two glycosylation sites) the relative abundances at the peptide level (group 16) were paired to generate a simulated profile of glycoforms assuming random combination.

(Figures S5 and S30). In line with these results, MU MS shows that the Hc subunit (Figure S35) and Fc/2 subunit (Figure S41) of SiLuLite mAb are predominantly G0F and G1F modified, with G0 and G2F present also in significant abundance, and with sialylated forms detected at the <1% relative abundance level. These findings are in line with the results of an in-depth TD/MD MS study reported in a recent paper.[76] Similarly, the most abundant glycoforms for NIST mAb and trastuzumab were confirmed in a follow-up TD/MD MS study based on MALDI FT-ICR MS.[82]

Owing to the lower number of participants who reported mass measurement errors for the top three glycoforms of NIST mAb (Figure S4) and trastuzumab (Figure S6), a slightly broader variation in the mass accuracies was reported for the SiLuLite mAb (Figure S5). The latter observation, once again, highlights the importance of advanced training in experimental TD/MD MS. Box-plots showing the errors (expressed in ppm) of glycoform-specific mass measurements for the IdeS-digested and disulfide bond reduced NIST mAb (Figure S9), SiLuLite mAb (Figure S10), and trastuzumab (Figure S11) are displayed by instrument and by proteoform (Supporting Information). A distribution of mass measurement errors for all subunits considered together reported for the most abundant proteoforms, is shown in Figure S12. The reported results indicate that mass measurements errors reported for TOF MS are consistently lower than for the FTMS instruments, Table 1. This rather unexpected result can be rationalized by, presumably, a substantially larger number of participants reporting results for FTMS measurements and in their broader experience level, compared to TOF MS results that were obtained only by high-level experts.

The glycosylation profile revealed by TD/MD MS (Table S14) can be compared with the results generated by the BU MS analysis of short (tryptic) glycopeptides (Table S15) or by the analysis of isolated glycans prepared by enzymatic removal.[83] A detailed comparison of Tables S14 (TD/MD MS data) and S15 (BU MS data) indicates that differences may be observed between minor glycoforms. For example, BU MS reports a minor, <1% relative abundance, G3F glycosylation for the NIST mAb, which is not reported by TD/MD MS. The latter may indicate that the glycoforms for intact mAbs reported as G1F/G2F, could be instead represented as G0F/G3F glycoforms. Furthermore, BU MS reports Man5 glycosylation for all three mAbs at 1−2% relative abundance. Only a few TD/MD MS reports reported this glycosylation, for example, see Figure 3. Resolving minor

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

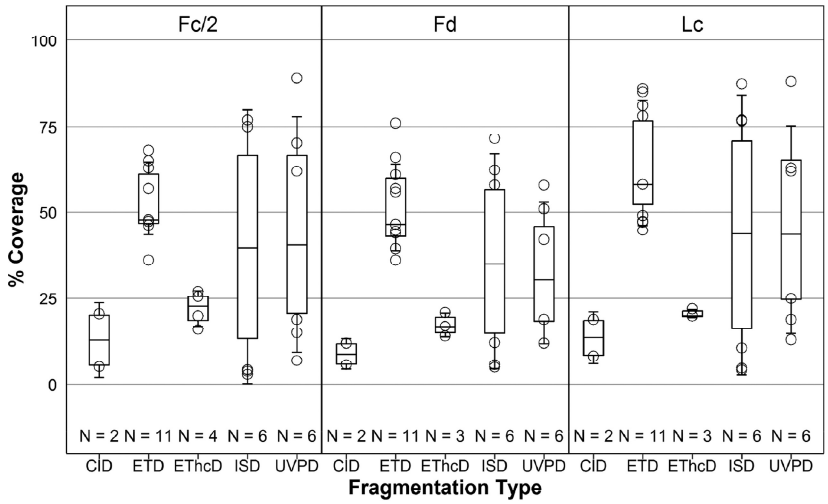


Figure 6. Fragmentation method-classified sequence coverage in % for IdeS-digested NIST, SiLuLite, and trastuzumab mAbs obtained for disulfide bond reduced mAbs (data from all mAbs are shown together, grouped by MS/MS method and mAb subunit type). The corresponding mAb-specific data are shown in Figures S21–S23.

glycoforms by TD/MD MS requires optimal operation parameters (e.g., desalting and desolvation), mass resolution, and data processing pipelines. Insufficient resolution/high spectral baseline noise may have rendered the minor glycoforms unreported in many of the TD/MD MS results. For example, results reported in Figure 3 that allowed detection of Man5/Man5 glycoform, have been generated by averaging of unprocessed FTMS data (time-domain transients) from multiple LC-MS technical replicates, see Tables S8–S10 for experimental details.[59]

To evaluate the relative quantitation of major glycoforms, we compared the TD/MD results at the intact mAb, Fc subunit, reduced heavy chain, and Fc/2 levels for the three mAbs. These data are further compared with those obtained at the peptide (BU) level, Figure 5. Overall, the distributions obtained for intact and MU mass measurements match well with those predicted from BU data, in particular for the reduced heavy chain and Fc/2. However, some differences between the intact protein and Fc measurements are observed. In Figure 5a, the simulated profile was overall biased toward lower mass of total glycans compared to the profile measured at the intact protein level (i.e., average fractional abundances of G0F/G1F in NIST, G0F/G0F in SiLuLite, and G0F/G1F in trastuzumab are higher in BU measurements than in intact, and vice versa for G1F/G2F in the three mAbs). The differences between the simulated profile and the experimental data were smaller for Fc (Figure 5b). The same KGP-digested mAb sample also showed that the glycation level on the F(ab) (complementary subunit to Fc as shown in Figure 1) was shown to be 4–7% (standard deviation 1–2%, Table S16), which is higher than the expected level of 1–2% calculated from BU experiments (Table S15).

Assuming no significant loss of glycans during sample processing and ion manipulations, discrepancy in glycosylation levels among different methods may be attributed to two factors. First, ionization efficiencies of glycoforms (large glycoproteins) differ from those of corresponding glycopeptides, and, furthermore, glycosylation may induce changes in digestion efficiency that impacts BU but not intact/MU methods. Second, any potential preferred combinations of glycosylations would invalidate simulated distributions by randomly combining BU data. For example, the G2F/G2F glycoform in intact SiLuLite and G1F/G2F glycoform in trastuzumab (Figure 5a) appeared to be higher than in the simulated values from BU data. It is possible that the chain pairing is preferred between similar glycoforms instead of being purely random (for example, preferentially paring two Hc subunits both with G2F will favor formation of G2F/G2F). Because of limited data in the current study, we could not confirm if the discrepancy seen in Figure 5a is statistically significant. Overall, the distributions from intact/MU measurements were consistent with the simulated values from BU data (assuming random combination), suggesting that chain pairing is largely nonselective. Nonetheless, it is anticipated that such intact and MU analyses will aid in characterizing chain pairing of bispecific antibodies. The information on PTM combinations cannot be easily retrieved from BU data.[59,84,85]

Amino Acid Sequence Coverage. Obtaining 100% sequence coverage of mAb primary structure is desired to provide unambiguous and complete characterization.[33,54] In practice, users may define a "100% sequence coverage" for BU and TD/MD approaches differently. It is widely accepted that the BU approach may deliver a 100% sequence coverage of mAbs (often through the use of multiple enzymes and MS/MS methods).[26,29,30,86] That claim assumes that a complete mAb sequence will be confirmed by (overlapping) enzymatically derived peptides. However, not all of these peptides have tandem mass spectra with product ions covering the entire peptide backbone. In the TD/MD MS terminology, a 100% sequence coverage is achieved when protein backbone bonds between *each* pair of amino acids in a protein sequence are cleaved and at least one of the corresponding product ions is detected.[46] In this respect, a complete sequence coverage of mAbs solely by TD or MD approaches has not yet been reported.[53,68] Nevertheless, TD/MD MS provides extensive *sequence information*, which can be instrumental in a mAb characterization strategy when combined with intact mass or BU data or with known mAb structural features (e.g.,

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783–1802

homologous series and constant regions).[57,87,88] The combination of MS/MS data with accurate mass measurements in TD/MD MS was demonstrated to achieve the required sequence confirmation and even curation in particular cases.[33]

As expected, the highest sequence coverage in this study was obtained by fragmenting the disulfide bond reduced 25 kDa mAb subunits, Figures 6 and S14–S26. Figure 6 shows that, on average, the sequence coverage obtained was 53%, 34%, and 51% for the Lc subunit; 44%, 30%, and 40% for the Fd′ subunit; and 48%, 40%, and 45% for the Fc/2 subunit for NIST, SiLuLite, and trastuzumab mAbs, respectively. The wide range in the reported sequence coverages is due to a few reports that described a sequence coverage as low as 5% from, for example, MALDI ISD. On the other hand, other groups reported substantially higher sequence coverages by use of MALDI ISD, with the maximum sequence coverages among all MS/MS techniques reported for the SiLuLite's Lc subunit (77%), Fd′ subunit (62%), and Fc/2 subunit (87%). Similarly, MALDI ISD yielded the highest sequence coverage for the Fc/2 subunit of trastuzumab (75%). Reporting such a broad distribution for the same MS/MS method may indicate a strong influence of experimental procedures (see Tables S8, S9, and S11). For example, detailed comparison of experimental details provided by groups 1, 15, and 21, which reported MALDI ISD results, suggests that higher sequence coverage of mAb subunits is provided when (i) IdeS digestion is employed together with a complete reduction of intrachain disulfide bonds, (ii) LC is employed to separate the subunits; and, potentially, and (iii) superdihydroxybenzoic acid (sDHB) is employed as a matrix. These conclusions are indicated by results presented in Figures 6 and S22–S24. The added value of point i is further supported by Figures S25 and S26.

Apart from MALDI ISD, the most comprehensive sequence coverage was provided by UVPD (88% for LC subunit of NIST mAb and 89% for the Fc/2 subunit of NIST), and a combination of CID with ETD enhanced by ion–ion proton transfer reactions or PTR (86% for the Lc subunit and 66% for the Fd' subunit of trastuzumab and 76% for the Fd′ subunit of the NIST mAb). Both MS/MS methods, CID and ETD, were performed in the linear ion trap (LTQ) of the 21 T FT-ICR MS platform.[88] Figure 7 displays the rich tandem mass spectra obtained across a wide m/z range generated by the combination of ETD and PTR.

Despite the pronounced simplification of product ion distributions achieved by the PTR approach, which reduces the charge of the product ions to enhance their m/z separation, Figure 7 demonstrates a well-resolved but still complex pattern of overlapping isotopic distributions of product ions. Disentangling these product ion contributions can be particularly difficult, as exemplified in the related studies on TD/MD MS of mAbs for diverse MS/MS methods.[49–52,59,68,76,89]

Examples of the sequence maps generated with TD/MD MS approaches in the current study are provided in Figures S60–S66. The total sequence coverage obtained for the ETD/PTR MS/MS data depicted in Figure 7 is one of the highest in the study, Figure 8.

Figure 8 further indicates a certain degree of complementarity in the sequence information obtained between ESI ETD MS/MS and ESI CID MS/MS. Namely, these results report 30 common cleavage sites represented by c/b product ions and 25 common sites represented by z/y product ions, whereas the number of product ions specific to ETD MS/MS are 81 for c-

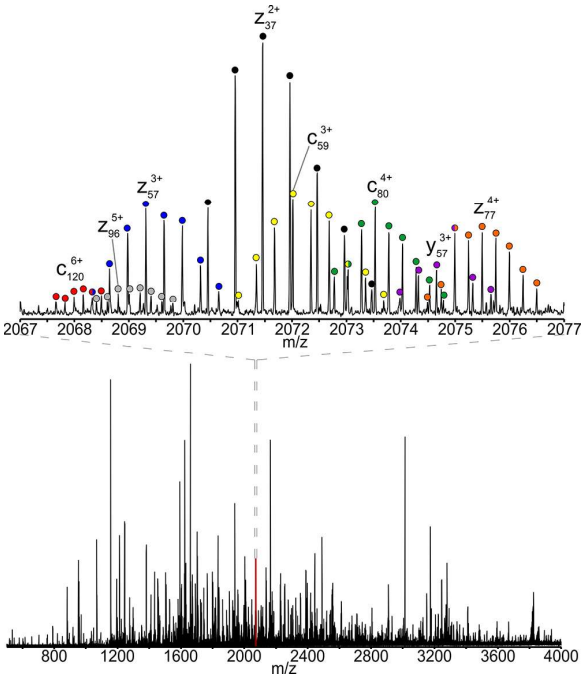


Figure 7. Example of an MD MS application to mAb analysis: sequencing of a light chain of SiLuLite mAb with a 21 T ESI FT-ICR MS employing ETD/PTR MS/MS (group 8). The inset shows an expanded view of a tandem mass spectrum with isotopic envelopes of product ions assigned and color coded for facile visualization.

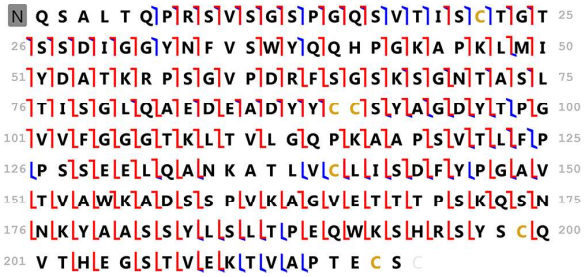


Figure 8. Total sequence coverage of 85% achieved for the analysis of the disulfide bond-reduced light chain of SiLuLite mAb with a 21 T ESI FT-ICR MS (group 8), based on middle-down MS/MS (combination of results from two tandem mass spectra). Included are product ions identified from CID/PTR MS/MS (10 transients averaged, b/y-ions, cleavage sites shown in blue) and of ETD/PTR MS/MS (10 transients averaged, c/z-ions, cleavage sites shown in red).

ions and 75 for z-ions and those specific to CID MS/MS are 11 for b-ions and 11 for y-ions. Similarly, integrating data from ESI UVPD MS/MS, MALDI ISD MS/MS, and ESI ETD MS/MS spectra leads to an increased sequence coverage and more confident characterization, Figure 9.

Note that a narrow (a single charge state) or wide (multiple charge states) precursor ion isolation could be performed for ETD MS/MS experiments, whereas there is no precursor ion isolation in MALDI ISD. The latter renders LC separation of mAb subunits prior to MALDI ISD analysis essential. Notably, ETD MS/MS analysis was complementary to MALDI ISD because it frequently extends the coverage of product ions

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783–1802

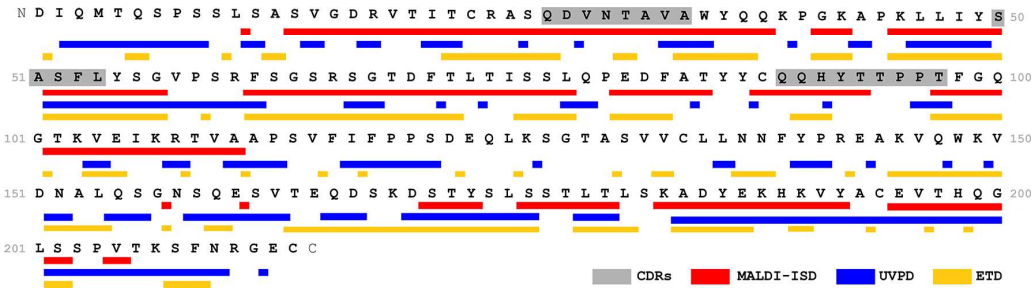


Figure 9. Comparison of sequence coverage (represented by sequence tags) and individual backbone cleavage sites obtained with ESI UVPD, ESI ETD, and MALDI ISD MS/MS for the Lc of trastuzumab. Included data are from groups 9 (ETD), 14 (UVPD), and 21 (MALDI ISD). See Figures S35, S39, and S41 for more details.

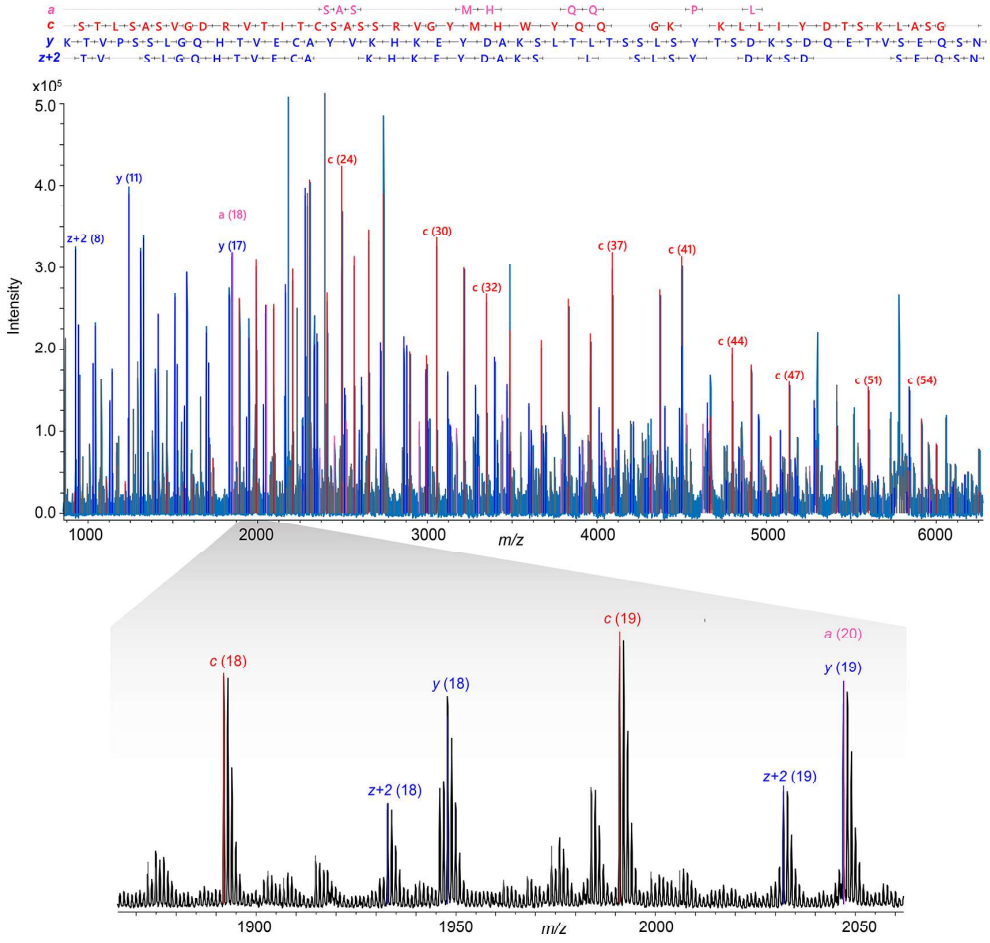


Figure 10. Example of an MD MS mAb analysis: sequencing of the NIST mAb Lc after chromatographic separation with a Bruker rapifleX MALDI-TOF MS based on MALDI ISD MS/MS. (Top panel) MALDI ISD mass spectrum. Product ion types *a*, *c*, *y*, and $z + 2$ were assigned. Monoisotopic peak list was obtained by usage of the SNAP algorithm for singly charged ions. (Bottom panel) Expanded view of the full range mass spectrum exhibiting baseline resolution of isotopic distributions. Additional expanded views are in Figure S59.

toward the termini. For example, the MALDI ISD product ion ladder from a TOF MS would typically begin with c_8/y_8 or larger product ions, whereas the ETD TOF MS/MS sequence coverage reported here shows the following pattern: (i) Fc/2 subunit, readout from c_2 and $(z_4 + 1)$; (ii) Fd′ subunit, readout from c_2 and $(z_2 + 1)$; and (iii) Lc subunit, readout from c_4 and $(z_9 + 1)$, Figure 10 and Figure 11.

Here, notation "$z_n + 1$" refers to one of the possible product ion types in ETD MS/MS. Similar observations were reported for MALDI ISD ($z_n + 2$ type ions) and for ETD MS/MS measurements performed by FT-ICR MS.[82] The "sequence validation percentage", or SVP, approach employed here by group 1 was originally introduced to improve the interpretation of TD/MD sequencing data compared to sequence

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

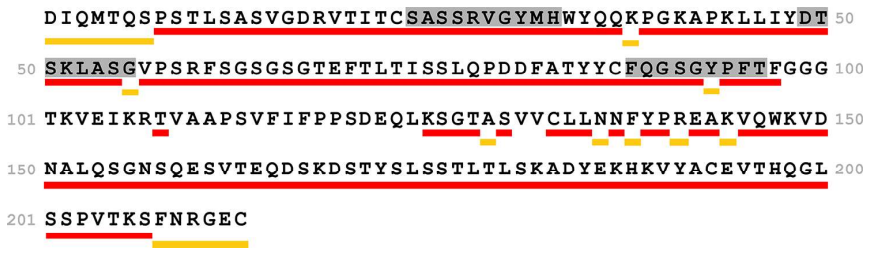


Figure 11. Example of an MD MS application for mAb analysis: sequencing of NIST mAb Lc with a Bruker rapifleX MALDI-TOF MS based on MALDI ISD MS/MS (see Figure 10). 77% sequence coverage with matching N- and C-terminal product ions (red bricks) was obtained (CDRs are shown in gray). Yellow bricks indicate accepted gaps when sequence calculations are performed by use of the "sequence validation percentage" approach.[33]

coverage alone.[33] It uses parametrized conditions to obtain a metric for the percentage of the analyzed sequence that could be validated based on the available data. For example, the SVP method considers sequence gaps as validated if (i) sequence readout starts before or at the 10th (or another user-defined number) amino acid residue from the N- or C- terminus of a given chain, (ii) internal gaps are attributed to Pro-containing moieties, such as ···PX··· or ···PPX···; or (iii) other internal gaps, say, two amino acid residues long, are immediately followed by an at least equally long sequence tag.[33] Below we describe results from MALDI ISD sequencing, which, although is less common than ETD/CID approaches, has demonstrated promise for mAb analysis.[33,57,82,84]

The LC-free MALDI ISD approach applied to an intact mAb leads to a tandem mass spectrum integrating ISD mass spectra from both mAb Hc and Lc chains. Typically, amino acid readouts are shorter compared to the off-line LC-separated IdeS/KGP-derived subunits and no Fc glycosylation is covered by this approach. In the current study, MALDI ISD performed by off-line LC separation in conjunction with MALDI-TOF MS yielded sequence coverage typically exceeding 70% for all subunits, with the highest value for the Lc of SiLuLite mAb being ∼87%; Figures 10 and S62.

The analysis of the trastuzumab, SiLuLite, and NIST mAb Fc/2 subunits yielded 75/77/77% sequence coverage, respectively, in addition to Fd′ 58/62/71%, and Lc 77/87/77%. Thus, an average sequence coverage of 74 ± 8% was observed in the three analyses from LC-MALDI ISD as reported by group 1. For example, as a result of LC-MALDI ISD sequencing of the NIST mAb IdeS-derived subunits, up to 80−100 residues from both the N- and C-termini toward the subunit center were confirmed (Figure 10). Generally, ∼7−10 terminal residues were not directly observed in the low-mass background region in MALDI ISD mass spectra. The match was considered as valid (match rules were parametrized in BioPharma Compass software)[33] if the downstream product ions were consistently observed, as detailed above for the rules employed by the SVP approach. In the case of SiLuLite mAb, the Lc subunit was largely sequenced, with the only gap between residues 109−113. Similarly, for the Fd′ subunit, sequence was largely confirmed, except for residues between positions 89−147. However, for both Lc and Fd′, additional modifications or sequence variations can be excluded based on the mass measurements, which all agree with the calculated masses of the fully reduced subunits. In the case of the Fc/2 subunit, the sequence was confirmed for the C-terminus lysine-loss proteoform (lysine clipping). Residues 98−124 were not directly covered, and thus modifications/sequence variations in this region could not be ruled out. However, as for Lc and Fd′ subunits, the complementary intact and MU MS data are in agreement with the absence of such variations. These MALDI ISD TOF MS sequencing results can be compared with the follow-up report that employed a higher-resolution MS instrument, namely, 12 T FT-ICR MS.[82]

CDRs Sequencing. CDRs represent the variable domains of mAbs. They contribute to the unique antigen-binding properties of mAbs, distinguishing various mAbs from each other.[54] As a result, complete sequencing of CDRs is one of the CQAs for mAb structural analysis. The mAbs considered here are of the IgG1 isotype and thus each of them has three CDRs in both the Lc and Hc subunits (Table S1).

BU MS is currently the method of choice for CDR sequencing.[29] It can provide 100% sequence coverage of all CDRs, especially if de novo sequencing is not needed, as was the case with the present study. To obtain benchmarking results, three groups participating in the current study were tasked to perform BU MS in addition to TD/MD MS approaches. For the three mAbs employed in this study, a conventional BU approach provided 100% sequence coverage for all CDRs. For example, group 16 obtained 100% sequence coverage for all CDRs in all three mAbs by employing Lys-C for digestion. Group 24 employed a mixture of 3 enzymes (trypsin, Glu-C, and chymotrypsin) to digest the mAbs and obtained 100% sequence coverage for all CDRs except one (CDR3 in the NIST mAb Hc).

TD/MD MS sequencing of CDRs follows the same pattern described for general sequencing: the Lc CDRs were on average better covered than those belonging to the Fd′ (or the Hc). In particular, CDR3 of the Fd′ was sometimes poorly covered; for example, group 10 reported zero backbone bonds cleaved for trastuzumab's CDR3 and only one backbone bond cleaved for the NIST mAb's CDR3 based on the MD MS/MS of mAb subunits with an Orbitrap FTMS platform. Conversely, the CDR1 and CDR2 of trastuzumab and NIST mAb were on average characterized with ∼50% of the bonds cleaved, as reported by the same group. Group 8 employed a 21 T FT-ICR MS to extensively sequence mAb subunits, with mass spectral data and sequence coverage map examples shown in Figures 6 and 7, respectively. With regard to the CDR coverage reported in Figure 8 for the Lc of SiLuLite mAb, a total of 11 bonds out of 13 were cleaved for CDR1, 5 out of 6 for CDR2 and all bonds were cleaved for CDR3. LC-MALDI ISD TOF MS of the reduced NIST mAb (Lc + Hc mixture) sequenced four out of six CDRs (group 1). When the same approach was applied to the IdeS-derived subunits of NIST mAb, five out of six CDRs were confirmed with relatively high, >85%, sequence

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

coverage for 3 CDRs on the Lc and 2 CDRs on the Fd' domain (Figure 11). In summary, together with the achievable low ppm mass accuracy for mass measurements of mAb subunits (IdeS digestion followed by disulfide bond reduction), the combination of both MD and BU approaches leaves little uncertainty about the correctness of the given sequence in general and of CDRs in particular.

The CDRs of the Hc and Lc subunits were similarly covered and highly sequenced with MALDI ISD implemented on a 15 T FT-ICR MS, Figures 9 and S66. Briefly, the following results were reported: for the Hc subunit, 100% (8 out of 8 bonds cleaved) sequence coverage of CDR1, 88% (7 out of 8 bonds cleaved) sequence coverage of CDR2, and 86% (12 out of 14 bonds cleaved) sequence coverage of CDR3; and for the Lc subunit, 100% (8 out of 8 bonds cleaved) for CDR1, 83% (5 out of 6 bonds cleaved) for CDR2, and 78% (7 out of 9 bonds cleaved) for CDR3.

Overall, sequence analysis of the 50 kDa subunits of trastuzumab based on MALDI ISD MS/MS shows that the CDR coverage follows what was also observed for the disulfide-protected Lc and Fd′ subunits obtained by IdeS proteolysis. The MS/MS analysis of the larger subunits, for example, 100 kDa F(ab′)2 subunits generated by IdeS proteolysis without disulfide bond reduction, shows the importance of high order structure (retained in the gas phase mainly due to the presence of disulfide bridges) for any MS/MS approach employed.[51] These results are rationalized by considering that the CDR3 is located in the disulfide-free loop approximately at the center of each F(ab′)2 chain, and correlates with the prior reports on TD MS of intact mAbs.[49−51]

DISCUSSION

Performance Evaluation of Middle/Top-Down Approaches. The results reported above demonstrate an overall correlation between the analytical performance and mAb structural information obtained here and the TD/MD MS reports published to date (see Introduction). The mass measurements of intact mAbs and their subunits are critical in providing knowledge on the mAb structural integrity, a feature which cannot be directly revealed by BU approaches.

This study further emphasizes that care should be taken when calculating molecular weights of mAbs and their subunits. In particular, these calculations have to take into account the state (oxidized or reduced) of all disulfide bonds (Tables S5−S7). The risk of misassignments of the monoisotopic peak for 50−100 kDa subunits increases compared to 25 kDa subunits. Misassignment can increase the mass errors by multiples of about ±1 Da.[45] Furthermore, peak interferences due to partially reduced mAb subunits are common and should be prevented by thorough reduction methods or considered during data analysis.[33] The molecular weight determination of 25 kDa mAb subunits can drastically reduce this shortcoming, as seen with the IdeS or KGP approach (with disulfide bond reduction) and the disulfide-bond reduced mAb data set analysis. However, even monoisotopic mass measurements of 25 kDa based on state-of-the-art MS instruments can provide mass measurement errors exceeding 10 ppm (Figure S7). Among the origins for these errors are (i) incomplete reduction of disulfide bonds of intact mAbs or mAb subunits after IdeS/KGP digestion, (ii) misassignment of monoisotopic mass upon deconvolution,[45] and (iii) reporting of average mass instead of monoisotopic mass. Let us consider these sources of errors in more detail.

Several proteoforms of mAb subunits exhibiting a different number of oxidized/reduced disulfide bonds can often be observed in the same MU/MD MS experiment. Therefore, the appropriate precursor ion mass needs to be considered for peak annotation (Tables S5−S7). In the current study, most of the groups reported monoisotopic masses of 25 kDa mAb subunits in which all disulfide bonds were reduced. Fewer groups reported results for precursor ions with all (two) disulfide bonds intact (mass difference of about 4 Da). One group reported results for subunits with a single intact disulfide bond (mass difference of 2 Da). Once the precursor mass was corrected for an appropriate number of disulfide bonds reduced/oxidized, the mass measurement errors were generally within 1−2 ppm (Figure S7). However, the reported results demonstrate high variation of the mass measurement errors, significantly increasing the mean mass measurement errors even for high-performance FTMS instruments, Table 1. The misassignment of a monoisotopic mass as a result of the deconvolution procedure remains a common problem. The submitted results exhibited both 1 and 2 Da shifts toward lighter or heavier experimentally obtained masses, which resulted in significant mass measurement errors. These errors could be due to erroneous outcomes of deconvolution algorithms applied to statistically poorly represented isotopic envelopes of multiply charged protein precursor ions.

For glycan profiling, the general approach was to measure the masses of the intact mAb or IdeS/KGP-digested mAbs, with or without disulfide bond reduction. The difference between the expertise levels of participants was apparent: a few groups were able to provide more information on intact proteoforms than just the typical N-linked glycosylation, C-terminal Lys clipping, or N-terminal formation of pyroglutamate (pyro-Q or pyro-E). Subunit-level glycoform profiling brings an additional benefit of separating the Fc/2 subunit with the expected N-linked glycosylation from the Fd′ and Lc subunits. This profiling aids in confirming the expected modifications or revealing unexpected glycosylation modifications on other than the Fc/2 subunits.[57] In the mAb samples employed in the current study, most of the glycosylation modifications are located in the conserved site on Fc/2, whereas glycation sites (Hex) were detected in the Lc and Fd′ at a few percent by MU and BU experiments (Tables S15 and S16). The latter is normally related to the use of enzymes, for example, EndoS or IgGZERO, for removing the glycans during sample preparation.[90] Similarly, it is also known that the integrated BU and MU approach can be useful in detecting O-HexNAc modification in the hinge region of some mAbs, which was not the case here.[91,92]

The reported maximum sequence coverages for diverse MS/MS methods for both TD and MD MS/MS approaches correlate well with the current literature on this subject. For example, in addition to ETD with PTR and multiple fills of the C-trap (Figure S57), one of the higher sequence coverages was obtained by 193 nm UVPD for the reduced Fc/2 subunit of the NIST mAb (89%) performed with an Orbitrap FTMS instrument (Figure S61). Different ion activation methods have shown a good degree of complementarity for the characterization of mAb subunits (see fragmentation maps integrating CID and ETD in Figure 8, as well as UVPD, ETD, and MALDI ISD in Figure 9).

Overall, TD/MD MS results are reliable if a significant length of sequence tag, more than 2−3 amino acids, is obtained, and they are further strengthened by complementary

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

product ions (from both ends of a protein sequence). Individually matching product ions, ones that do not form sequence tags with other product ions, are questionable and may not be suitable to confirm or reject sequence assignments. It is apparent that achieving maximum sequence coverages necessitates a high level of expertise in TD/MD MS/MS practice, which was reflected by the wide spread of sequence coverages reported here (Figures S14–S26). Nevertheless, the results of this study show that achieving complete sequence coverage for all CDRs for the same mAb by use of TD/MD MS even by the expert users was not always possible (Figures 7–9).

The strength of the 50–100 kDa subunit mass measurement approach perhaps lies not so much in the best possible accuracy of the mass determination but in its ability to provide chain pairing information via linking Lc and Fd′ subunits to each other, among other uses.[59] Clearly, MS/MS data obtained from these larger 50–100 kDa disulfide bond-linked subunits are less suitable for full sequence confirmation but can be useful to assess subunit extremities. On the other hand, these conclusions could be a consequence of the fact that, as shown by this study, TD/MD MS analysis of these larger subunits is not as widely practiced by the participating laboratories as intact mAb or smaller subunit analysis. Nevertheless, the large mAb subunits may provide information complementary to intact mass measurements and to MS/MS data, for example when obtaining smaller subunits is prohibited because of sample stability not supporting additional sample processing.

An equally formidable challenge is the identification and assignment of the anticipated dozens or hundreds of internal ions (i.e., fragment ions that do not contain the known N- or C-terminus of the protein) in the MS/MS spectra. Not only are these ions difficult to assign, but they also have the potential to increase the false discovery rate and congest the already dense mass spectra created upon fragmentation of subunits or intact proteins. Despite these hurdles, internal ions could offer an additional rich source of structural information if their data content could be mined.[93,94]

For PTM analysis, MD and TD MS may complement BU MS results and increase the confidence in their identification by deciphering their origins (i.e., sample preparation vs naturally occurring) and their relative stoichiometry. Overall, the modifications observed for the three mAbs were among the most commonly detected in all mAbs-Lys clipping, N-terminal pyro-Glu, and N-linked glycosylation. TD/MD MS revealed Lys clipping on the C-termini of both trastuzumab and NIST mAb, with about 3.1–3.4% glycation of the Lc of SiLuLite mAb. The Fc/2 subunits of all mAbs expectedly contained the truncated glycans resulting from the IgGZERO digestion (deglycosylation) at Asn300 (Asn61 after IdeS digestion, see Table S1).

The selected mAbs represent the three possible cases for pyroglutamate formation at the N-terminus. The two amino acids that can spontaneously cyclize when present at the N-terminus are glutamic acid (Glu) and glutamine (Gln). In the mAbs targeted in the present study, SiLuLite mAb contains both Gln on the N-terminus of the Lc and Glu on the N-terminus of the Hc; whereas NIST mAb has a Gln on the N-terminus of the Hc and trastuzumab has a Glu on its Hc (Table S1). The results reported here allowed for the assignment of all three mAbs to their corresponding incidence of pyro-Glu/Gln modification. Clearly, pyro-Gln (or pyro-Q) formation was present in significantly higher amounts compared to pyro-Glu (pyro-E) modification. For example, the Lc of SiLuLite mAb was found to be about 94% pyroglutamylated (pyro-Q). The calculated masses with pyro-Glu and pyro-Gln modifications for intact mAbs and their subunits are listed in Tables S5–S7.

FURTHER CONSIDERATIONS AND OUTLOOK

On the basis of the reports obtained in the present study and from the literature, the current set of conventional approaches for mAb characterization entails (Figure 2 and Tables S11 and S12): (i) sample introduction and ionization, direct ESI infusion or MALDI sample deposition, online LC-MS and LC-MS/MS with ESI or LC-MALDI ISD, or off-line LC-MS and LC-MALDI ISD; (ii) mass spectrometry, MS and MS/MS following the best practices in the corresponding methods and techniques, with use of a single MS/MS method in its conventional implementation (vide infra); and (iii) data processing, deconvolution of MS and MS/MS data for further data analysis. As a few reports have demonstrated, this set of approaches can be efficiently complemented with advanced methods to deliver increased performance, including greater sequence coverage.

The following advanced methods appear to be of particular interest, and could be more widespread in the future: (i) sample introduction and ionization, providing complementary solution-phase separation of proteins using capillary electrophoresis (CE); (ii) mass spectrometry, increasing population of product ions by multiple fills of the external ion traps (such as the C-trap in Orbitraps), gas-phase fractionating of overlapping ion populations using ion mobility, spreading the condensed product ion distribution in a wider mass range by use of gas-phase ion–ion PTR to reduce the average charge state of product ions (see Figure 7),[95,96] unfolding precursor ions to facilitate product ion separation via precursor ion activation,[97] and considering complementary and chimeric product ion data obtained from multiple MS/MS methods or from the same MS/MS method performed with different experimental parameters; and (iii) data processing, assigning isotopic envelopes of product ions without mass spectra deconvolution and increasing sensitivity and dynamic range via averaging unreduced data from LC-MS/MS technical replicates.[51,59] These advanced approaches aim to reduce the complexity of both mAb precursor and tandem mass spectra by an improved solution-phase and gas-phase fractionation, accompanied by increased sensitivity via enhancing the number of precursor and product ions or by acquiring more data. All of the approaches mentioned above have been employed by the participants of the present study but only by 1–2 groups in each case. Therefore, wider distribution and acceptance of these advanced methods has yet to be achieved.

A more viable combination of MS methods for mAb analysis today in many laboratories is afforded by a conventional BU CID/HCD-based approach (with trypsin, complemented by chymotrypsin digestion, if needed) integrated with intact mAb measurements and MU MS of mAb subunits produced by IdeS digestion and disulfide bond reduction.[98] To provide sequencing and PTM information on mAbs, the complementarity of MS/MS methods suggests usage of a CID-type method, for example, HCD, with a radical-chemistry driven MS/MS approach, for example, ETD, EThcD, or UVPD. Adding MALDI ISD sequencing appears to be attractive for mAb analysis. Potentially, a single MS platform should be able to perform efficient BU, MD (for example with IdeS and other

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783–1802

enzymes) and intact mass measurement on whole mAbs (low resolution) and MU on subunits (high resolution). However, based on the experience of this project, this strategy is not trivial. Standardization of reporting and data analysis software is necessary. If BU data and TD/MD MS data can be more effectively integrated, some of the lower magnitude peaks in TD/MD mass spectra will be more proficiently assigned. In turn, this integration will inform how the pieces detected in BU assemble as a whole. Therefore, further software development is needed to enable data integration. For example, group 17 performed extensive BU MS studies of SiLuLite, employing four enzymes and two different MS/MS methods, HCD and ETD (Table S13). Nevertheless, they were not able to provide 100% sequence coverage for the Hc of SiLuLite, even by combining sequencing information from all four employed enzymes and despite applying advanced bioinformatics approaches.[72] A completely unsequenced region was found to be part of a Hc above the hinge region with the sequence DYFPEPVTVSW. On the other hand, MD MS data for the same sequence region of SiLuLite performed on the Fd′ subunit show complete sequence coverage of this region (see, for example, Figure S5 in a report by Ge and co-workers).[76] Therefore, integration of BU MS and MD MS data could provide 100% sequence coverage for SiLuLite mAb. Interestingly, BU MS-derived 100% sequence coverage for the Hc subunit was reported only by one group and only for trastuzumab (Tables S13 and S15).

In the present study, many of the figures submitted by the participants displayed manually annotated mass spectra, which indicates that an extensive effort is still required to derive structural conclusions from raw data. Automation of data analysis with no loss in accuracy is evolving with the recent developments of both open source and commercial software workflows for biopharma applications.[33] Recent advancements in algorithms for data analysis of protein mass spectra aim at more accurate peak assignment (for accurate monoisotopic or average mass calculation),[99] as well as improved deconvolution of complex product ion mass spectra and attempts to perform de novo TD MS sequencing.[100]

In principle, TD/MD MS could be integrated into the next generation MAM workflows, which presently employ BU-derived methods. TD/MD MS could be beneficial for MAM by reducing labor involved in sample preparation and minimizing the risk of artifacts. In fact, intact mass measurements are already a part of MAM approaches accepted in biopharmaceutical industry, with MU approaches being under evaluation.[7] For these approaches to be fully accepted, mass tolerance settings (e.g., <0.3 Da or <12 ppm at 25 kDa) could be used as analytical criteria to reliably distinguish target mAb structures from, for example, deamidated species. Nevertheless, extensive efforts to increase sample throughput of TD/MD MS, automate data acquisition, and develop specialized software are needed to integrate the full capabilities of TD/MD MS into MAM workflows. Standard reference materials, such as the commercially available and analyzed here NIST and SiLuLite mAbs, can be used to establish system suitability that is similarly required for regulatory approval of BU analysis.[101]

CONCLUSIONS

This comparative interlaboratory study highlights the value of readily available TD/MD MS tools for mAb structural analysis. The wide diversity in the instrumentation and methods employed by the 20 partner laboratories of this project captures the current state of TD/MD MS, that is, there is no "one size fits all." Compared to the techniques and instruments used for BU, TD/MD, approaches are still very much in development. However, the potential for higher performance protein analysis in the future is clearly apparent as instrumentation capabilities improve, and advanced TD/MD methods become more widely available. Already today, many of the TD/MD tools are accepted by analytical departments in various CROs, pharma/biotech companies, and service facilities in academic institutions worldwide.

Although mAb sequence coverage from MS/MS data is currently less than 100% for either MD or TD MS, information complementary to BU approaches is obtained in a more rapid fashion, which may be crucial or beneficial for comprehensive and unambiguous mAb characterization. Particularly, mass measurements of intact mAbs (part of TD MS) and structural subunits of mAbs (MU MS) are an important complement to BU approaches. On the other hand, the present study also uncovered the wide range of expertise levels found among the participants for TD/MD MS protein characterization. Clearly, there is a need for advanced training of the analytical scientists performing complex MS-based experiments, such as denaturing and native TD/MD MS of mAbs.[27]

Interestingly, less popular compared to the ESI-based approaches nowadays, the MALDI-based ISD fragmentation approach performed on either TOF or FT-ICR MS instruments has demonstrated a particularly attractive efficiency for mAb sequencing.[82] These results may establish the groundwork for the use of MALDI-based methods for mAb structural analysis and lead to the wider acceptance of TD and MD methods in the industrial environment. That would be in-line with the success of another MALDI-based technology, where MALDI TOF MS is routinely used for microorganism identification via intact mass measurements in many hospitals and healthcare organizations worldwide.[102]

As demonstrated by several groups in this study, there is a strong potential for further development of TD/MD MS techniques that would result in their improved sequencing efficiency and wider acceptance. Reported here are extensions of the standard TD and MD methods, including ETD with multiple fills, CID, ETD and UVPD coupled with PTR, and interexperiment averaging of unreduced (raw) multiple LC-MS/MS technical replicates,[59] that can be among the methods available for implementation in routine workflows. Ion activation techniques, such as UVPD and in-beam ECD,[53] could gain wider use for mAb characterization. The arsenal of MS/MS technologies available today may be further extended by the development of novel methods specifically targeting TD MS of large proteins and protein complexes, such as surface induced dissociation (SID)[103] or hydrogen atom attachment to precursor protein ions in the gas phase.[104]

The original wide scope of the present interlaboratory study was to provide a broad overview of the field as currently practiced. Future studies could monitor the progress of TD/MD MS as the technologies advance. Also, follow-up studies could focus on more targeted topics related to TD/MD MS of mAbs, for example ADC analysis.[105−107]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jasms.0c00036.

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

Details and examples of the experimental results, including tables and figures with information about sample preparation, mass spectrometry, and data analysis methods and techniques employed by the study participants, and the results of the statistical analysis of the measurements performed by the study participants (PDF)

AUTHOR INFORMATION

Corresponding Author

Yury O. Tsybin − *Spectroswiss, 1015 Lausanne, Switzerland;* orcid.org/0000-0001-7533-0774; Email: tsybin@spectroswiss.ch

Authors

Kristina Srzentić − *Northwestern University, Evanston, Illinois 60208-0001, United States*

Luca Fornelli − *Northwestern University, Evanston, Illinois 60208-0001, United States*

Joseph A. Loo − *University of California-Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0001-9989-1437

Henrique Seckler − *Northwestern University, Evanston, Illinois 60208-0001, United States*

Jeffrey N. Agar − *Northeastern University, Boston, Massachusetts 02115, United States*

Lissa C. Anderson − *National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States;* orcid.org/0000-0001-8633-0251

Dina L. Bai − *University of Virginia, Charlottesville, Virginia 22901, United States*

Alain Beck − *Centre d'immunologie Pierre Fabre, 74160 Saint-Julien-en-Genevois, France;* orcid.org/0000-0002-4725-1777

Jennifer S. Brodbelt − *University of Texas at Austin, Austin, Texas 78712-1224, United States;* orcid.org/0000-0003-3207-0217

Yuri E. M. van der Burgt − *Leiden University Medical Centre, 2300 RC Leiden, The Netherlands;* orcid.org/0000-0003-0556-5564

Julia Chamot-Rooke − *Institute Pasteur, 75015 Paris, France;* orcid.org/0000-0002-9427-543X

Sneha Chatterjee − *University of Antwerp, 2000 Antwerp, Belgium*

Yunqiu Chen − *Biogen, Inc., Cambridge, Massachusetts 02142-1031, United States*

David J. Clarke − *The University of Edinburgh, EH9 3FJ Edinburgh, United Kingdom;* orcid.org/0000-0002-3741-2952

Paul O. Danis − *Consortium for Top-Down Proteomics, Cambridge, Massachusetts 02142, United States*

Jolene K. Diedrich − *The Scripps Research Institute, La Jolla, California 92037, United States*

Robert A. D'Ippolito − *University of Virginia, Charlottesville, Virginia 22901, United States;* orcid.org/0000-0001-7542-6629

Mathieu Dupré − *Institute Pasteur, 75015 Paris, France;* orcid.org/0000-0002-1845-0048

Natalia Gasilova − *Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*

Ying Ge − *University of Wisconsin-Madison, Madison, Wisconsin 53706, United States;* orcid.org/0000-0001-5211-6812

Young Ah Goo − *University of Maryland, Baltimore, Maryland 21201, United States*

David R. Goodlett − *University of Maryland, Baltimore, Maryland 21201, United States*

Sylvester Greer − *University of Texas at Austin, Austin, Texas 78712-1224, United States*

Kim F. Haselmann − *Novo Nordisk, DK-2760 Malov, Denmark*

Lidong He − *National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States*

Christopher L. Hendrickson − *National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States*

Joshua D. Hinkle − *University of Virginia, Charlottesville, Virginia 22901, United States;* orcid.org/0000-0002-7626-8957

Matthew V. Holt − *Baylor College of Medicine, Houston, Texas 77030-3411, United States*

Sam Hughes − *The University of Edinburgh, EH9 3FJ Edinburgh, United Kingdom*

Donald F. Hunt − *University of Virginia, Charlottesville, Virginia 22901, United States*

Neil L. Kelleher − *Northwestern University, Evanston, Illinois 60208-0001, United States;* orcid.org/0000-0002-8815-3372

Anton N. Kozhinov − *Spectroswiss, 1015 Lausanne, Switzerland*

Ziqing Lin − *University of Wisconsin-Madison, Madison, Wisconsin 53706, United States*

Christian Malosse − *Institute Pasteur, 75015 Paris, France*

Alan G. Marshall − *National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States; Florida State University, Tallahassee, Florida 32310-4005, United States;* orcid.org/0000-0001-9375-2532

Laure Menin − *Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*

Robert J. Millikin − *University of Wisconsin-Madison, Madison, Wisconsin 53706, United States;* orcid.org/0000-0001-7440-3695

Konstantin O. Nagornov − *Spectroswiss, 1015 Lausanne, Switzerland*

Simone Nicolardi − *Leiden University Medical Centre, 2300 RC Leiden, The Netherlands;* orcid.org/0000-0001-8393-1625

Ljiljana Paša-Tolić − *Pacific Northwest National Laboratory, Richland, Washington 99354, United States*

Stuart Pengelley − *Bruker Daltonik GmbH, 28359 Bremen, Germany;* orcid.org/0000-0003-2676-8096

Neil R. Quebbemann − *University of California-Los Angeles, Los Angeles, California 90095, United States*

Anja Resemann − *Bruker Daltonik GmbH, 28359 Bremen, Germany*

Wendy Sandoval − *Genentech, Inc., South San Francisco, California 94080-4990, United States*

Richa Sarin − *Biogen, Inc., Cambridge, Massachusetts 02142-1031, United States*

Nicholas D. Schmitt − *Northeastern University, Boston, Massachusetts 02115, United States*

Jeffrey Shabanowitz − *University of Virginia, Charlottesville, Virginia 22901, United States*

Jared B. Shaw − *Pacific Northwest National Laboratory, Richland, Washington 99354, United States;* orcid.org/0000-0002-1130-1728

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

Michael R. Shortreed − *University of Wisconsin-Madison, Madison, Wisconsin 53706, United States*
Lloyd M. Smith − *University of Wisconsin-Madison, Madison, Wisconsin 53706, United States*
Frank Sobott − *University of Antwerp, 2000 Antwerp, Belgium; University of Leeds, LS2 9JT Leeds, United Kingdom;* orcid.org/0000-0001-9029-1865
Detlev Suckau − *Bruker Daltonik GmbH, 28359 Bremen, Germany*
Timothy Toby − *Northwestern University, Evanston, Illinois 60208-0001, United States*
Chad R. Weisbrod − *National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States;* orcid.org/0000-0001-5324-4525
Norelle C. Wildburger − *Washington University School of Medicine, St. Louis, Missouri 63110, United States*
John R. Yates, III − *The Scripps Research Institute, La Jolla, California 92037, United States;* orcid.org/0000-0001-5267-1672
Sung Hwan Yoon − *University of Maryland, Baltimore, Maryland 21201, United States*
Nicolas L. Young − *Baylor College of Medicine, Houston, Texas 77030-3411, United States;* orcid.org/0000-0002-3323-2815
Mowei Zhou − *Pacific Northwest National Laboratory, Richland, Washington 99354, United States;* orcid.org/0000-0003-3575-3224

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.0c00036

Author Contributions

†K.S. and L.F. contributed equally.

Notes

The authors declare the following competing financial interest(s): Anja Resemann, Stuart Pengelley, and Detlev Suckau are employees of Bruker Daltonics, which manufactures TOF MS and FT-ICR MS instruments and commercializes allied software. Anton Kozhinov, Konstantin Nagornov, and Yury Tsybin are employees of Spectroswiss, which develops and commercializes FTMS data processing and data analysis software. Neil Kelleher is involved with commercialization of ProSight software for data analysis. Paul Danis is also the Founder and Principal of Eastwoods Consulting, providing business advisory services to life science companies.

■ ACKNOWLEDGMENTS

We greatly appreciate the support of the Pilot Project by the Consortium for Top-Down Proteomics and its supporting organizations, Thermo Fisher Scientific, Inc., Bruker Corp., and Genovis AB. We also acknowledge Genovis for providing enzymes (IdeS and KGP) and Millipore Sigma for donation of their mAb standard (SiLuLite mAb). J.A.L. acknowledges support from the US National Institutes of Health (R01GM103479, S10RR028893, S10 OD018504); Y.O.T. and J.C.R. from European Horizon 2020 research and innovation program under Grant Agreement No. 829157 (TopSpec); J.S.B. from the Welch Foundation (F-1155) and the National Science Foundation (CHE1402753); YG from NIH GM125085 and S10OD018475; and D.F.H. from the NIH Grant GM037537. L.P.T., J.B.S., and M.Z. acknowledge the support from the Intramural program at Environmental Molecular Sciences Laboratory (EMSL), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and operated under Contract No. DE-AC05-76RL01830. C.L.H., L.C.A., L.H., and A.G.M. acknowledge support for work performed at the National High Magnetic Field Laboratory, which is supported by National Science Foundation Cooperative Agreement No. DMR-1644779 and the State of Florida. This research was carried out in collaboration with the National Resource for Translational and Developmental Proteomics under Grant P41 GM108569 from the National Institute of General Medical Sciences, National Institutes of Health.

■ REFERENCES

(1) Beck, A.; Liu, H. Macro- and Micro-Heterogeneity of Natural and Recombinant IgG Antibodies. *Antibodies* **2019**, *8*, 18.
(2) Rathore, D.; Faustino, A.; Schiel, J.; Pang, E.; Boyne, M.; Rogstad, S. The role of mass spectrometry in the characterization of biologic protein products. *Expert Rev. Proteomics* **2018**, *15*, 431−449.
(3) Taussig, M. J.; Fonseca, C.; Trimmer, J. S. Antibody validation: a view from the mountains. *New Biotechnol.* **2018**, *45*, 1−8.
(4) Rogers, R. S.; Nightlinger, N. S.; Livingston, B.; Campbell, P.; Bailey, R.; Balland, A. Development of a quantitative mass spectrometry multi-attribute method for characterization, quality control testing and disposition of biologics. *mAbs.* **2015**, *7*, 881−890.
(5) Rogers, R. S.; Abernathy, M.; Richardson, D. D.; Rouse, J. C.; Sperry, J. B.; Swann, P.; et al. A View on the Importance of "Multi-Attribute Method" for Measuring Purity of Biopharmaceuticals and Improving Overall Control Strategy. *AAPS J.* **2018**, *20*, 7.
(6) Xu, W.; Jimenez, R. B.; Mowery, R.; Luo, H.; Cao, M.; Agarwal, N.; et al. A Quadrupole Dalton-based multi-attribute method for product characterization, process development, and quality control of therapeutic proteins. *mAbs.* **2017**, *9*, 1186−1196.
(7) Chen, B.; Lin, Z.; Zhu, Y.; Jin, Y.; Larson, E.; Xu, Q.; et al. Middle-Down Multi-Attribute Analysis of Antibody-Drug Conjugates with Electron Transfer Dissociation. *Anal. Chem.* **2019**, *91*, 11661.
(8) Rogstad, S.; Yan, H.; Wang, X.; Powers, D.; Brorson, K.; Damdinsuren, B.; et al. Multi-Attribute Method for Quality Control of Therapeutic Proteins. *Anal. Chem.* **2019**, *91*, 14170−14177.
(9) Grilo, A. L.; Mantalaris, A. The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends Biotechnol.* **2019**, *37*, 9−16.
(10) Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **2018**, *36*, 1136.
(11) *Advancing Health Through Innovation: 2019 New Drug Therapy Approvals by the FDA's Center for Drug Evaluation and Research*, 2020. https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/new-drug-therapy-approvals-2019.
(12) Duivelshof, B. L.; Jiskoot, W.; Beck, A.; Veuthey, J.-L.; Guillarme, D.; D'Atri, V. Glycosylation of biosimilars: Recent advances in analytical characterization and clinical implications. *Anal. Chim. Acta* **2019**, *1089*, 1−18.
(13) Kaplon, H.; Muralidharan, M.; Schneider, Z.; Reichert, J. M. Antibodies to watch in 2020. *mAbs.* **2020**, *12*, 1703531.
(14) Xu, Y.; Wang, D.; Mason, B.; Rossomando, T.; Li, N.; Liu, D.; et al. Structure, heterogeneity and developability assessment of therapeutic antibodies. *mAbs.* **2019**, *11*, 239−264.
(15) Nowak, C. K.; Cheung, J. M.; Dellatore, S.; Katiyar, A.; Bhat, R.; Sun, J.; et al. Forced degradation of recombinant monoclonal antibodies: A practical guide. *mAbs* **2017**, *9*, 1217−1230.
(16) Smith, L. M.; Kelleher, N. L.; The Consortium for Top Down Proteomics. Proteoform: a single term describing protein complexity. *Nat. Methods* **2013**, *10*, 186.
(17) Ebbers, H. C.; Crow, S. A.; Vulto, A. G.; Schellekens, H. Interchangeability, immunogenicity and biosimilars. *Nat. Biotechnol.* **2012**, *30*, 1186.
(18) Rogstad, S.; Faustino, A.; Ruth, A.; Keire, D.; Boyne, M.; Park, J. A Retrospective Evaluation of the Use of Mass Spectrometry in

FDA Biologics License Applications. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 786−794.
(19) Ambrogelly, A.; Gozo, S.; Katiyar, A.; Dellatore, S.; Kune, Y.; Bhat, R.; et al. Analytical comparability study of recombinant monoclonal antibody therapeutics. *mAbs.* **2018**, *10*, 513−538.
(20) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; et al. How many human proteoforms are there? *Nat. Chem. Biol.* **2018**, *14*, 206−214.
(21) Fornelli, L.; Toby, T. K.; Schachner, L. F.; Doubleday, P. F.; Srzentić, K.; DeHart, C. J.; et al. Top-down proteomics: Where we are, where we are going? *J. Proteomics* **2018**, *175*, 3−4.
(22) Boutz, D. R.; Horton, A. P.; Wine, Y.; Lavinder, J. J.; Georgiou, G.; Marcotte, E. M. Proteomic Identification of Monoclonal Antibodies from Serum. *Anal. Chem.* **2014**, *86*, 4758−4766.
(23) Miller, R. M.; Millikin, R. J.; Hoffmann, C. V.; Solntsev, S. K.; Sheynkman, G. M.; Shortreed, M. R.; et al. Improved Protein Inference from Multiple Protease Bottom-Up Mass Spectrometry Data. *Journal of Proteome Research.* **2019**, *18*, 3429−3438.
(24) Cheung, W. C.; Beausoleil, S. A.; Zhang, X.; Sato, S.; Schieferl, S. M.; Wieler, J. S.; et al. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat. Biotechnol.* **2012**, *30*, 447−452.
(25) Srzentić, K.; Zhurov, K. O.; Lobas, A. A.; Nikitin, G.; Fornelli, L.; Gorshkov, M. V.; et al. Chemical-Mediated Digestion: An Alternative Realm for Middle-down Proteomics? *Journal of Proteome Research.* **2018**, *17*, 2005−2016.
(26) Srzentić, K.; Fornelli, L.; Laskay, Ü.A.; Monod, M.; Beck, A.; Ayoub, D.; et al. Advantages of extended bottom-up proteomics using Sap9 for analysis of monoclonal antibodies. *Anal. Chem.* **2014**, *86*, 9945−9953.
(27) Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A. Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 1149.
(28) McLafferty, F. W.; Breuker, K.; Jin, M.; Han, X.; Infusini, G.; Jiang, H.; et al. Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* **2007**, *274*, 6256−6268.
(29) Sen, K. I.; Tang, W. H.; Nayak, S.; Kil, Y. J.; Bern, M.; Ozoglu, B.; et al. Automated antibody de novo sequencing and its utility in biopharmaceutical discovery. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 803−810.
(30) Morsa, D.; Baiwir, D.; La Rocca, R.; Zimmerman, T. A.; Hanozin, E.; Grifnée, E.; et al. Multi-Enzymatic Limited Digestion: The Next-Generation Sequencing for Proteomics? *Journal of Proteome Research.* **2019**, *18*, 2501.
(31) Müller, T.; Winter, D. Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents. *Mol. Cell. Proteomics* **2017**, *16*, 1173−1187.
(32) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in top-down proteomics and the analysis of proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499−519.
(33) Resemann, A.; Jabs, W.; Wiechmann, A.; Wagner, E.; Colas, O.; Evers, W.; et al. Full validation of therapeutic antibody sequences by middle-up mass measurements and middle-down protein sequencing. *mAbs.* **2016**, *8*, 318−330.
(34) Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat. Methods* **2019**, *16*, 587−594.
(35) Loo, J.; Edmonds, C.; Smith, R. Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. *Science* **1990**, *248*, 201−204.
(36) Loo, J. A.; Edmonds, C. G.; Smith, R. D. Tandem mass spectrometry of very large molecules: serum albumin sequence information from multiply charged ions formed by electrospray ionization. *Anal. Chem.* **1991**, *63*, 2488−2499.
(37) Han, X.; Jin, M.; Breuker, K.; McLafferty, F. W. Extending top-down mass spectrometry to proteins with masses greater than 200 kDas. *Science* **2006**, *314*, 109−112.
(38) McLafferty, F. W.; Horn, D. M.; Breuker, K.; Ge, Y.; Lewis, M. A.; Cerda, B.; et al. Electron capture dissociation of gaseous multiply charged ions by Fourier transform ion cyclotron resonance. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 245−249.
(39) 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. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528−9533.
(40) Brodbelt, J. S. Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem. Soc. Rev.* **2014**, *43*, 2757−2783.
(41) Demeure, K.; Quinton, L.; Gabelica, V.; De Pauw, E. Rational Selection of the Optimum MALDI Matrix for Top-Down Proteomics by In-Source Decay. *Anal. Chem.* **2007**, *79*, 8678−8685.
(42) Fukuyama, Y.; Iwamoto, S.; Tanaka, K. Rapid sequencing and disulfide mapping of peptides containing disulfide bonds by using 1,5-diaminonaphthalene as a reductive matrix. *J. Mass Spectrom.* **2006**, *41*, 191−201.
(43) LeDuc, R. D.; Kelleher, N. L. Using ProSight PTM and Related Tools for Targeted Protein Identification and Characterization with High Mass Accuracy Tandem MS Data. *Current Protocols in Bioinformatics.* **2007**, *19*, 13.16.11−13.16.28.
(44) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; et al. ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *Proteomics* **2015**, *15*, 1235−1238.
(45) Cai, W.; Guner, H.; Gregorich, Z. R.; Chen, A. J.; Ayaz-Guner, S.; Peng, Y.; et al. MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. *Mol. Cell. Proteomics* **2016**, *15*, 703−714.
(46) Smith, L. M.; Thomas, P. M.; Shortreed, M. R.; Schaffer, L. V.; Fellers, R. T.; LeDuc, R. D.; et al. A five-level classification system for proteoform identifications. *Nat. Methods* **2019**, *16*, 939−940.
(47) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; et al. Informed-Proteomics: open-source software package for top-down proteomics. *Nat. Methods* **2017**, *14*, 909.
(48) Nesvizhskii, A. I.; Aebersold, R. Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. *Mol. Cell. Proteomics* **2005**, *4*, 1419−1440.
(49) Tsybin, Y. O.; Fornelli, L.; Stoermer, C.; Luebeck, M.; Parra, J.; Nallet, S.; et al. Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Anal. Chem.* **2011**, *83*, 8919−8927.
(50) Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. Top-down structural analysis of an intact monoclonal antibody by electron capture dissociation Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2013**, *85*, 4239−4246.
(51) Fornelli, L.; Ayoub, D.; Aizikov, K.; Liu, X.; Damoc, E.; Pevzner, P. A.; et al. Top-down analysis of immunoglobulin G isotypes 1 and 2 with electron transfer dissociation on a high-field Orbitrap mass spectrometer. *J. Proteomics* **2017**, *159*, 67−76.
(52) Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; et al. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation Orbitrap FTMS. *Mol. Cell. Proteomics* **2012**, *11*, 1758−1767.
(53) Shaw, J. B.; Malhan, N.; Vasil'ev, Y. V.; Lopez, N. I.; Makarov, A. A.; Beckman, J. S.; et al. Sequencing grade tandem mass spectrometry for top-down proteomics using hybrid electron capture dissociation methods in a benchtop Orbitrap mass spectrometer. *Anal. Chem.* **2018**, *90*, 10819.
(54) Zhang, Z.; Pan, H.; Chen, X. Mass spectrometry for structural characterization of therapeutic antibodies. *Mass Spectrom. Rev.* **2009**, *28*, 147−176.
(55) Wang, B.; Gucinski, A. C.; Keire, D. A.; Buhse, L. F.; Boyne, M. T., II Structural comparison of two anti-CD20 monoclonal antibody drug products using middle-down mass spectrometry. *Analyst* **2013**, *138*, 3058−3065.

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802

(56) 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.* **2009**, *20*, 1415–1424.
(57) Ayoub, D.; Jabs, W.; Resemann, A.; Evers, W.; Evans, C.; Main, L.; et al. Correct primary structure assessment and extensive glycoprofiling of cetuximab by a combination of intact, middle-up, middle-down and bottom-up ESI and MALDI mass spectrometry techniques. *mAbs.* **2013**, *5*, 699–710.
(58) 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.* **2014**, *86*, 3005–3012.
(59) Srzentić, K.; Nagornov, K. O.; Fornelli, L.; Lobas, A. A.; Ayoub, D.; Kozhinov, A. N.; et al. Multiplexed Middle-Down Mass Spectrometry as a Method for Revealing Light and Heavy Chain Connectivity in a Monoclonal Antibody. *Anal. Chem.* **2018**, *90*, 12527–12535.
(60) Cotham, V. C.; Brodbelt, J. S. Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level using Middle-Down 193 nm Ultraviolet Photodissociation. *Anal. Chem.* **2016**, *88*, 4004–4013.
(61) Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W. P.; Mohammed, S.; Heck, A. J. R. Toward an Optimized Workflow for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89*, 3318–3325.
(62) Laskay, Ü.A.; Srzentić, K.; Monod, M.; Tsybin, Y. O. Extended bottom-up proteomics with secreted aspartic protease Sap9. *J. Proteomics* **2014**, *110*, 20–31.
(63) Sjögren, J.; Andersson, L.; Mejàre, M.; Olsson, F. In *Generating and Purifying Fab Fragments from Human and Mouse IgG Using the Bacterial Enzymes IdeS, SpeB and Kgp*; Nordenfelt, P, Collin, M, Eds.; Springer: New York, NY, 2017.
(64) Faid, V.; Leblanc, Y.; Bihoreau, N.; Chevreux, G. Middle-up analysis of monoclonal antibodies after combined IgdE and IdeS hinge proteolysis: Investigation of free sulfhydryls. *J. Pharm. Biomed. Anal.* **2018**, *149*, 541–546.
(65) Brunner, A. M.; Lössl, 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.* **2015**, *87*, 4152–4158.
(66) Fornelli, L.; Parra, J.; Hartmer, R.; Stoermer, C.; Lubeck, M.; Tsybin, Y. O. Top-down analysis of 30–80 kDa proteins by electron transfer dissociation time-of-flight mass spectrometry. *Anal. Bioanal. Chem.* **2013**, *405*, 8505–8514.
(67) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V. Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **2015**, *87*, 4370–4376.
(68) Fornelli, L.; Srzentić, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; et al. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **2018**, *90*, 8421–8429.
(69) Bakalarski, C. E.; Gan, Y.; Wertz, I.; Lill, J. R.; Sandoval, W. Rapid, semi-automated protein terminal characterization using ISDetect. *Nat. Biotechnol.* **2016**, *34*, 811–813.
(70) Patiny, L.; Borel, A. ChemCalc: A Building Block for Tomorrow's Chemical Infrastructure. *J. Chem. Inf. Model.* **2013**, *53*, 1223–1228.
(71) De Leoz, M. L. A.; Duewer, D. L.; Stein, S. E. *NIST Interlaboratory Study on the Glycosylation of NIST mAb, a Monoclonal Antibody Reference Material, June 2015 to February 2016*, NISTIR 8186; NIST, 2017. DOI: 10.6028/NIST.IR.8186.
(72) Li, Q.; Shortreed, M. R.; Wenger, C. D.; Frey, B. L.; Schaffer, L. V.; Scalf, M.; et al. Global Post-Translational Modification Discovery. *Journal of Proteome Research.* **2017**, *16*, 1383–1390.
(73) Valeja, S. G.; Kaiser, N. K.; Xian, F.; Hendrickson, C. L.; Rouse, J. C.; Marshall, A. G. Unit Mass Baseline Resolution for an Intact 148 kDa Therapeutic Monoclonal Antibody by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2011**, *83*, 8391–8395.
(74) Shaw, J. B.; Brodbelt, J. S. Extending the Isotopically Resolved Mass Range of Orbitrap Mass Spectrometers. *Anal. Chem.* **2013**, *85*, 8313–8318.
(75) Farrell, A.; Carillo, S.; Scheffler, K.; Cook, K.; Bones, J. Monoclonal antibody sequence assessment using a hybrid quadrupole-Orbitrap mass spectrometer. *Anal. Methods* **2018**, *10*, 3100–3109.
(76) Jin, Y.; Lin, Z.; Xu, Q.; Fu, C.; Zhang, Z.; Zhang, Q.; et al. Comprehensive characterization of monoclonal antibody by Fourier transform ion cyclotron resonance mass spectrometry. *mAbs.* **2019**, *11*, 106–115.
(77) Fornelli, L.; Durbin, K. R.; Fellers, R. T.; Early, B. P.; Greer, J. B.; LeDuc, R. D.; et al. Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *Journal of Proteome Research.* **2017**, *16*, 609–618.
(78) Lössl, P.; Snijder, J.; Heck, A. J. R. Boundaries of Mass Resolution in Native Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 906–917.
(79) Jennewein, M. F.; Alter, G. The Immunoregulatory Roles of Antibody Glycosylation. *Trends Immunol.* **2017**, *38*, 358–372.
(80) Wei, B.; Berning, K.; Quan, C.; Zhang, Y. T. Glycation of antibodies: Modification, methods and potential effects on biological functions. *mAbs.* **2017**, *9*, 586–594.
(81) Li, W.; Kerwin, J. L.; Schiel, J.; Formolo, T.; Davis, D.; Mahan, A. et al. *Structural Elucidation of Post-Translational Modifications in Monoclonal Antibodies*; American Chemical Society, 2015.
(82) van der Burgt, Y. E. M.; Kilgour, D. P. A.; Tsybin, Y. O.; Srzentić, K.; Fornelli, L.; Beck, A.; et al. Structural Analysis of Monoclonal Antibodies by Ultrahigh Resolution MALDI In-Source Decay FT-ICR Mass Spectrometry. *Anal. Chem.* **2019**, *91*, 2079–2085.
(83) Nallet, S.; Fornelli, L.; Schmitt, S.; Parra, J.; Baldi, L.; Tsybin, Y. O.; et al. Glycan variability on a recombinant IgG antibody transiently produced in HEK-293E cells. *New Biotechnol.* **2012**, *29*, 471–476.
(84) Gstöttner, C.; Reusch, D.; Haberger, M.; Dragan, I.; Van Veelen, P.; Kilgour, D. P. A.; et al. Monitoring glycation levels of a bispecific monoclonal antibody at subunit level by ultrahigh-resolution MALDI FT-ICR mass spectrometry. *mAbs.* **2020**, *12*, 1682403.
(85) Tyshchuk, O.; Gstöttner, C.; Funk, D.; Nicolardi, S.; Frost, S.; Klostermann, S.; et al. Characterization and prediction of positional 4-hydroxyproline and sulfotyrosine, two post-translational modifications that can occur at substantial levels in CHO cells-expressed biotherapeutics. *mAbs* **2019**, *11*, 1219–1232.
(86) Tran, N. H.; Rahman, M. Z.; He, L.; Xin, L.; Shan, B.; Li, M. Complete De Novo Assembly of Monoclonal Antibody Sequences. *Sci. Rep.* **2016**, *6*, 31730.
(87) Wu, S.; Lourette, N. M.; Tolić, N.; Zhao, R.; Robinson, E. W.; Tolmachev, A. V.; et al. An Integrated Top-Down and Bottom-Up Strategy for Broadly Characterizing Protein Isoforms and Modifications. *Journal of Proteome Research.* **2009**, *8*, 1347–1357.
(88) He, L.; Anderson, L. C.; Barnidge, D. R.; Murray, D. L.; Dasari, S.; Dispenzieri, A.; et al. Classification of Plasma Cell Disorders by 21 T Fourier Transform Ion Cyclotron Resonance Top-Down and Middle-Down MS/MS Analysis of Monoclonal Immunoglobulin Light Chains in Human Serum. *Anal. Chem.* **2019**, *91*, 3263–3269.
(89) Shaw, J. B.; Liu, W.; Vasil'ev, Y. V.; Bracken, C. C.; Malhan, N.; Guthals, A.; et al. Direct Determination of Antibody Chain Pairing by Top-down and Middle-down Mass Spectrometry Using Electron Capture Dissociation and Ultraviolet Photodissociation. *Anal. Chem.* **2020**, *92*, 766–773.
(90) Sjögren, J.; Cosgrave, E. F. J.; Allhorn, M.; Nordgren, M.; Björk, S.; Olsson, F.; et al. EndoS and EndoS2 hydrolyze Fc-glycans on therapeutic antibodies with different glycoform selectivity and can be used for rapid quantification of high-mannose glycans. *Glycobiology* **2015**, *25*, 1053–1063.

(91) Plomp, R.; Dekkers, G.; Rombouts, Y.; Visser, R.; Koeleman, C. A. M.; Kammeijer, G. S. M.; et al. Hinge-Region *O*-Glycosylation of Human Immunoglobulin G3 (IgG3). *Mol. Cell. Proteomics* **2015**, *14*, 1373−1384.
(92) Zauner, G.; Selman, M. H. J.; Bondt, A.; Rombouts, Y.; Blank, D.; Deelder, A. M.; et al. Glycoproteomic Analysis of Antibodies. *Mol. Cell. Proteomics* **2013**, *12*, 856−865.
(93) Lyon, Y. A.; Riggs, D.; Fornelli, L.; Compton, P. D.; Julian, R. R. The Ups and Downs of Repeated Cleavage and Internal Fragment Production in Top-Down Proteomics. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 150−157.
(94) Zenaidee, M. A.; Lantz, C.; Perkins, T.; Jung, W.; Ogorzalek Loo, R. R.; Loo, J. A. Internal Fragments Generated by Electron Ionization Dissociation Enhance Protein Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.*, 2020 in press (https://doi.org/10.1021/jasms.0c00160).
(95) Huang, T.-Y.; McLuckey, S. A. Top-down protein characterization facilitated by ion/ion reactions on a quadrupole/time of flight platform. *Proteomics* **2010**, *10*, 3577−3588.
(96) Ugrin, S. A.; English, A. M.; Syka, J. E. P.; Bai, D. L.; Anderson, L. C.; Shabanowitz, J.; et al. Ion-Ion Proton Transfer and Parallel Ion Parking for the Analysis of Mixtures of Intact Proteins on a Modified Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2163.
(97) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal. Chem.* **2015**, *87*, 7109−7116.
(98) Giorgetti, J.; Beck, A.; Leize-Wagner, E.; François, Y.-N Combination of intact, middle-up and bottom-up levels to characterize 7 therapeutic monoclonal antibodies by capillary electrophoresis − mass spectrometry. *J. Pharm. Biomed. Anal.* **2020**, *182*, 113107.
(99) Lermyte, F.; Dittwald, P.; Claesen, J.; Baggerman, G.; Sobott, F.; O'Connor, P. B.; et al. MIND: A Double-Linear Model To Accurately Determine Monoisotopic Precursor Mass in High-Resolution Top-Down Proteomics. *Anal. Chem.* **2019**, *91*, 10310−10319.
(100) Vyatkina, K.; Wu, S.; Dekker, L. J. M.; VanDuijn, M. M.; Liu, X.; Tolić, N.; et al. Top-down analysis of protein samples by de novo sequencing techniques. *Bioinformatics* **2016**, *32*, 2753−2759.
(101) Zhou, M.; Gucinski, A. C.; Boyne, M. T. Performance metrics for evaluating system suitability in liquid chromatography—Mass spectrometry peptide mass mapping of protein therapeutics and monoclonal antibodies. *mAbs.* **2015**, *7*, 1104−1117.
(102) Florio, W.; Tavanti, A.; Barnini, S.; Ghelardi, E.; Lupetti, A. Recent Advances and Ongoing Challenges in the Diagnosis of Microbial Infections by MALDI-TOF Mass Spectrometry. *Front. Microbiol.* **2018**, DOI: 10.3389/fmicb.2018.01097.
(103) Zhou, M.; Wysocki, V. H. Surface Induced Dissociation: Dissecting Noncovalent Protein Complexes in the Gas phase. *Acc. Chem. Res.* **2014**, *47*, 1010−1018.
(104) Di Stefano, L. H.; Papanastasiou, D.; Zubarev, R. A. Size-Dependent Hydrogen Atom Attachment to Gas-Phase Hydrogen-Deficient Polypeptide Radical Cations. *J. Am. Chem. Soc.* **2018**, *140*, 531−533.
(105) Beck, A.; D'Atri, V.; Ehkirch, A.; Fekete, S.; Hernandez-Alba, O.; Gahoual, R.; et al. Cutting-edge multi-level analytical and structural characterization of antibody-drug conjugates: present and future. *Expert Rev. Proteomics* **2019**, *16*, 337−362.
(106) Hernandez-Alba, O.; Houel, S.; Hessmann, S.; Erb, S.; Rabuka, D.; Huguet, R.; et al. A Case Study to Identify the Drug Conjugation Site of a Site-Specific Antibody-Drug-Conjugate Using Middle-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2419−2429.
(107) Watts, E.; Williams, J. D.; Miesbauer, L. J.; Bruncko, M.; Brodbelt, J. S. Comprehensive Middle-Down Mass Spectrometry Characterization of an Antibody−Drug Conjugate by Combined Ion Activation Methods. *Anal. Chem.* **2020**, *92*, 9790−9798.

https://dx.doi.org/10.1021/jasms.0c00036
J. Am. Soc. Mass Spectrom. 2020, 31, 1783−1802