Advancing Intact Protein Analysis by Top-down Mass Spectrometry: Enrichment, Separation, and Characterization

By

Bifan Chen

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Chemistry)

At the
UNIVERSITY OF WISCONSIN-MADISON
2019

Date of final oral examination: 5/06/2019

The dissertation is approved by the following members of the Final Oral Committee:

Ying Ge, Associate Professor, Cell and Regenerative Biology and Chemistry
Lloyld M. Smith, Professor, Chemistry
Lingjun Li, Professor, Pharmacy and Chemistry
Michael R. Sussman, Professor, Biochemistry

ACKNOWLEDGEMENTS

First and foremost, I am extremely grateful to my advisor Professor Ying Ge, for all her support and guidance. Five years ago, she took a leap of faith and accepted me into her group, and has been ever since giving me tremendous freedom, trust, and opportunities to grow as a scientist and as an individual. I would also like to thank all my other committee members, Professor Lingjun Li, Professor Lloyd Smith, and Professor Michael Sussman. Lingjun has always been encouraging and supportive. Lloyd gives me constructive criticism when I need it and shows me to remain scientifically curious.

It was my great pleasure working with the past and current members of the Ge group including Lynn Peng, Leekyoung Hwang, Tania Guardado, Liming Wei, Ruixiang Sun, Yu Liang, Ziqing Lin, Yanlong Zhu, Zach Gregorich, Vege Cai, Yutong Jin, Kyle Brown, Trisha Tucholski, Abe Wu, Sam Knott, Tim Tiambeng, Stanford Mitchell, David Roberts, Elizabeth Bayne, Austin Carr, Eli Larson, Jake Melby, Ben Wancewicz, Sean McILwain, Kent Wenger, Hannah Karp, and Molly Wetzel. Their cheerful and hardworking spirits have made the Ge lab a great place to work for the past few years. In fact, I have learnt something from each of them during collaboration, discussion, or simply being around. Some of them have become good friends of mine. In my second year, Ziqing Lin joined our group as a postdoc, who later became my roommate and one of my best friends. I am thankful for all the memorable moments we shared discussing science, fixing instruments, and living under the same roof. I would also like to thank Dr. Andrew Alpert, who has been a great source of knowledge and support for scientific discussions, particularly on chromatography. He is always available, patient, and supportive to talk about all the research challenges.

Finally, it was the unconditional love and support of my friends and family that made my graduate experience possible. My family has been by my side regardless of the decisions I made and is always proud of what I have accomplished along the way. Particularly, I have to thank my mom, who has sacrificed so much and gave me all the freedom to live and to pursue my dreams in my own world. I would also like to thank Karen and Jim Anway, who have known and cared for me since college. I always feel reinvigorated by the passion, love, and peace from their family. I also appreciate the company of Umi, who is so sweet and delightful to be around with. Lastly, I was extremely fortunate to meet my wife Xiaomeng, which was one of the best things that happened to me during my graduate school years. I am grateful to have her by my side and I look forward to our future together!

ABSTRACT

The study of proteins is critical for understanding cellular functions at the molecular level. Top-down mass spectrometry (MS) has emerged as a premier tool for global and comprehensive analysis of proteoforms. The top-down approach retains intact mass information, providing a "bird's-eye" view of the proteome and allowing for identification of novel proteoforms, in-depth sequence characterization, and quantification of disease associated post-translational modifications (PTMs). However, many technical challenges still exist. The research described here involves analytical development in top-down MS, particularly in the areas of enrichment, separation, and characterization of samples ranging from standard proteins and complex lysates, to large therapeutic biomolecules.

Chapter 1 provides an introduction and review of recent advances in different aspects of top-down proteomics. Chapters 2 and 3 are related to the study of intact phosphoproteins. Specifically, chapter 2 describes the use of functionalized nanoparticles for enrichment and the subsequent coupling of online liquid chromatography (LC)-MS for characterizing endogenous phosphoproteins from complex cell lysates. Chapter 3 investigates how phosphorylation moieties might influence the efficiency of electron capture dissociation (ECD). Chapters 4 and 5 focus on the development of hydrophobic interaction chromatography (HIC) that could be coupled online directly with MS and its applications to therapeutic molecules (monoclonal antibodies). Chapter 6 describes a middle-down approach to obtain multi-attribute of both cysteine and lysine conjugated antibody-drug conjugates, which overcomes some current challenges using HIC-MS and the top-down approach. Overall, these analytical developments expand the toolbox of the top-down approach and generally facilitate the analysis of intact proteins.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SUPPLEMENTAL FIGURES

LIST OF SUPPLEMENTAL TABLES

Chapter 1: Introduction and Recent Advances in Top-down Proteomics

This chapter has been published and is adapted from:

Chen, B.; Brown, K. A.; Lin, Z.; Ge, Y. Top-down Proteomics: Ready for Prime Time? *Analytical Chemistry.* 2018, 90, 110-127.

1.1 Introduction

In the postgenomic era, the study of proteins is critical for understanding cellular functions at the molecular level.[1-5] Beyond the genetic code, the human proteome is greatly diversified by genetic variations, alternative splicing of RNA transcripts, and post-translational modifications (PTMs).[6,7] In 2013, the term "proteoform" was designated to describe "all of the different molecular forms in which the protein product of a single gene can be found",[6] clearing up the confusion in nomenclature and joining research efforts to develop methodologies for proteoform characterization. Top-down proteomics, which analyzes intact proteins without digestion, has proved to be a premier mass spectrometry (MS) technology for global and comprehensive analysis of proteoforms.[4,8,9] The top-down approach retains intact protein mass information, providing a "bird's-eye" view of the proteome and allowing for identification of novel proteoforms, in-depth sequence characterization, and quantitation of disease-associated PTMs.[4,8,9]

Although some technical challenges remain, development over the past 5 years has expanded top-down proteomics from a mostly targeted approach to one capable of proteoform-profiling across multiple samples.[9] Now, thousands of proteoforms can be identified, characterized, and quantified using high throughput top-down proteomics approaches.[10,11] Moreover, developments in instrumentation and fragmentation have positioned top-down proteomics in the fast lane for future progression.[12] With the successful development of commercial high-resolution mass spectrometers such as the solariX XR Fourier transform ion cyclotron resonance (FTICR) (Bruker),[13] Orbitraps (Thermo), and quadrupole time-of-flight (Q-TOFs) (Bruker and Waters), these instruments are now widely available in academic and industrial laboratories for top-down proteomics. Moreover, the 21T FTICR mass spectrometers equipped with state-of-the-art fragmentation methods at the Pacific Northwest National Laboratory and

National High Magnetic Field Laboratory have demonstrated unprecedented resolving power, acquisition rate, and flexible tandem MS (MS/MS) capabilities, providing enormous potential for top-down proteomics practitioners to probe complicated proteomics applications.[14,15] Reciprocally, the methodological and technological gains from developing large-scale and high-throughput workflow have further empowered targeted analysis, from which top-down proteomics roots and thrives. As a result, interest in top-down MS has grown considerably and many studies have already underscored the potential of top-down proteomics for unraveling disease mechanisms and discovering novel biomarkers.[4,8,16,17]

A number of recent reviews have already given an overview of the technical requirements for top-down proteomics and delineated the history and fundamentals of the field as well as its application to biomedical research.[4,8,9,18-20] With a special emphasis on publications in the past two years (2015-2017), this review examines recent technological trends and developments in the areas of intact protein sample preparation, separation, MS/MS, data acquisition strategies, data analysis, native MS, and quantitation. We also highlight recent applications for global and targeted top-down proteomics applications, and conclude with outlooks in the field.

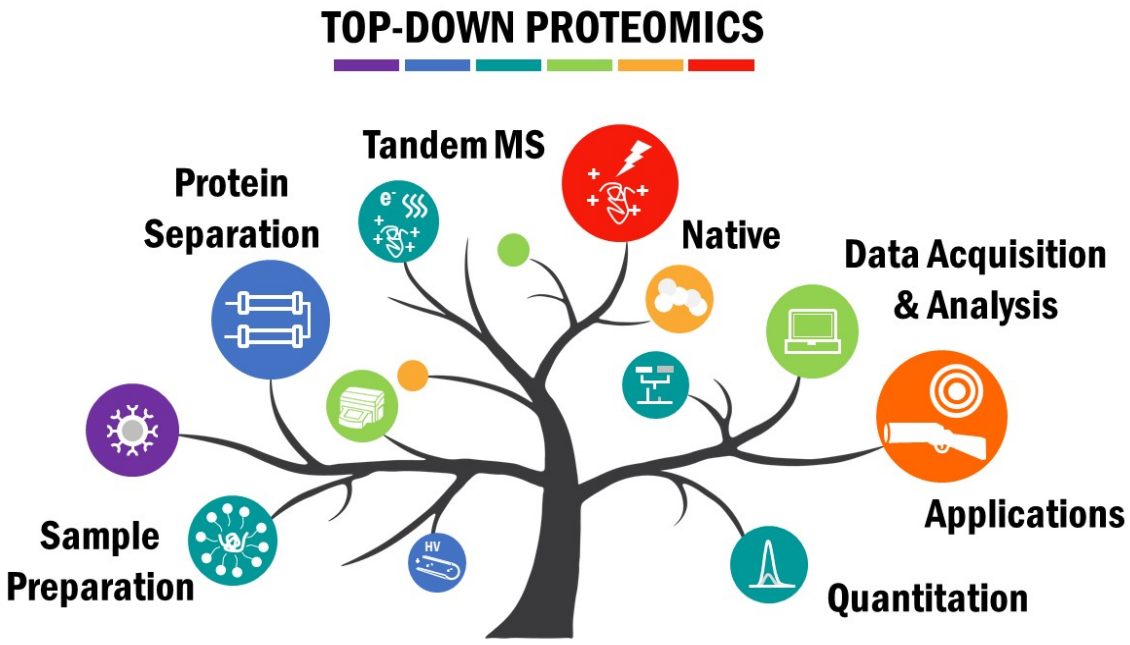


Figure 1.1. Illustration of various aspects of top-down proteomics.

1.2 Sample preparation for top-down proteomics

Often overlooked, sample preparation remains one of the most important and challenging aspects in top-down proteomics. Although MS is a sensitive analytical technique, isotope and charge state distributions of protein ions produced by electrospray ionization (ESI), spreads the signal of a single species over a large *m/z* range. Thus, signal suppression from salt adducts (e.g. Na^+, K^+), detergents, or even coexisting protein species can greatly hamper a top-down MS experiment. In the section below, common methods for extracting intact proteins, replacing/removing buffer components incompatible with MS, and approaches to decrease sample complexity and to enrich low-abundance proteins will be discussed.

Traditionally, physical methods for lysing samples (e.g. homogenization and sonication), are performed using a mixture of Good's buffers, salts, reducing agents, and protease and

phosphatase inhibitors to extract cellular components while avoiding proteoform alteration or degradation.[4] However, these conditions necessitate post-extraction work-up to remove or replace non-volatile salts that suppress MS signal by forming adducts to protein ions and increase the chemical noise.[21,22] Molecular-weight cutoff (MWCO) filters provide an easy method for exchanging protein samples into MS compatible conditions such as volatile ammonium salt buffers or low concentration, acidic solutions for downstream analysis.[23] Additionally, reversed-phase chromatography (RPC), a regular component in top-down proteomics workflows, desalts samples prior to the MS analysis.[24,25] Although proper desalting is generally critical for reliable MS data, recent developments in small emitter tips have pushed the boundaries of salt inclusion, making MS detection of protein ions produced by ESI more tolerant of non-volatile salt contamination, even with conditions mimicking physiological conditions (25 mM Tris, 150 mM KCl).[26]

Detergents are common buffer additives that facilitate cell permeabilization and aid in hydrophobic membrane proteins solublization.[27-29] In particular, anionic detergents such as sodium dodecyl sulfate (SDS) exhibit excellent protein solubility; however, they pose a challenge for downstream MS analysis by causing complete signal suppression at levels as low as 0.01%.[30] Protein precipitation, usually with chloroform/methanol or cold acetone, removes detergents and other MS incompatible contaminants.[31-33] These methods are straightforward and add little cost to the experiment, but unfortunately, loss of proteins, difficulty in re-solubilizing the protein pellet, and incomplete removal of contaminants result in low reproducibility in the down-stream analysis. In addition, hydrophobic membrane proteins generally require high levels of formic acid to dissolve the protein pellet, which can result in artificial protein formylation.[32,34] Although alternative filtration devices have been employed to deplete and remove SDS, non-specific binding and/or incomplete removal limits their widespread utility.[35,36] Recent advancements by Kachuk *et*

al. demonstrated near-complete front-end removal of SDS using a transmembrane electrophoresis.[37] Similarly an online matrix removal device developed by Kim *et al*. showed a 2- to 10-fold increase in signal after just 5 min.[37,38] Even though anionic detergent removal using cleavable linkers (e.g. acid labile detergents) between the hydrophobic tail and hydrophilic head have shown promise in bottom-up workflows, these methods have yet to be translated into top-down proteomics.[39,40] On the other hand, nonionic detergents, such as n-Dodecyl β-D-Maltoside (DDM), are also commonly used to solubilize protein, and generally exhibit lower signal suppression.[23,41,42] Often times, these mild, non-denaturing detergents do not perturb protein-protein interactions, making them indispensable tools in the area of native top-down MS.[41,43] Technological advancements in detergent removal as well as alternative extraction strategies (e.g. organic solvents[44,45]) have made membrane proteomics an intriguing area for further investigation and development.

Given the complexity and dynamic range of the proteome, enrichment strategies represent an important area of development. Organelle fractionation by differential centrifugation is widely used to isolate and enrich specific sub-proteomes.[46] Nuclear, mitochondrial, membrane, and cytosolic fractions can be collected for global protein analysis and deep proteoform coverage of organelle specific targets.[47] One common example is histone proteoform analysis, which has recently gained importance for its role in understanding epigenetic regulation.[48]

In other cases, low abundance proteoforms (e.g. phosphoproteins) are enriched by taking advantage of their unique physicochemical properties. Recently the phosphoproteomics field has benefited from technologies such as functionalized superparamagnetic nanoparticles and microspheres that enrich phosphoproteins for downstream top-down proteomics characterization.[49-52] In 2017, Chen *et al.* demonstrated a strategy coupling functionalized cobalt

ferrite nanoparticles to targeted liquid chromatography (LC)-MS/MS for phosphoprotein characterization.[52] This method demonstrated a significant enrichment of spiked-in phosphorylated β-casein from 0.5% to 94% in a complex tissue lysate, showing the high specificity and efficiency of the functionalized nanoparticles. More importantly, low-abundance endogenous phosphoproteins from a complex tissue lysate were greatly enriched, which allowed for confident identification and phosphorylation site localization.[52] Although overwhelmed by the signals of co-eluting non-phosphorylated proteins prior to enrichment, the phosphorylated ions became dominant after enrichment for subsequent MS/MS experiments.[52] Therefore, this strategy serves as a promising tool for probing novel phosphorylation sites for a more comprehensive understanding of their role in biological systems. Additionally, when targeted protein and protein complexes are of particular interest, co-immunoprecipitation or affinity purification are commonly used for comprehensive proteoform coverage and quantitation as demonstrated by a number of studies.[16,17,53-58]

1.3 Separation for top-down proteomics

The complexity of the proteome requires fractionation of intact proteins prior to top-down MS analysis.[59] The under-developed front-end separation of intact proteins has long been a major obstacle for further advancing top-down methodology.[8,60] Thus, significant efforts have recently been dedicated to the development of chromatographic and electrophoretic separation strategies to address intact protein separation challenges.[60-69] Numerically, the intact proteome appears to be a much simpler mixture than its corresponding peptide digests.[59] In practice, however, protein-level fractionation and separation are daunting tasks due to the diverse physicochemical properties (e.g. size, charge, and hydrophobicity) and the wide dynamic range of the proteome.[8,59] In this section,

we will discuss recent developments in chromatographic methods, especially in denaturing RPC and other nondenaturing techniques directly coupled to MS, electrophoretic approaches, as well as multidimensional separation strategies.

RPC is the most prevalent method for online LC-MS analysis of intact protein. Unfortunately, because of complex physicochemical properties of protein mixtures and poor recovery of intact proteins, the peak capacity and resolution of intact protein separation largely lag behind that of peptide separation in RPC.[59,60] Depending on the column and gradient length, the peak capacity obtained in the one-dimensional RPC separation of intact proteins is usually below 100.[62] Recently, Shen *et al.* reported the use of long RPC columns (>1 m) and long gradient time (>600 min) with short alkyl (C1-C4) bonded phases to achieve peak capacities greater than 400 (similar to peptide level RPC separation) for complex intact protein mixture up to 50 kDa.[68] They concluded that increasing column length results in the most dramatic improvement in their intact protein separation. Interestingly, they also noticed minimal influence of pore size between 200 Å to 400 Å,[68] although many others found 1000 Å generally provides more accessible pore volume for intact proteins.[62] Using a capillary column 4 cm long packed with sub-micron particles (0.47 μm; C18 bonded phase), Wu *et al.* demonstrated a slip-flow[65] nanocapillary RPC separation that has peak capacities up to 750 with *E. coli* lysate in a 60 min gradient.[70] However, both methods were conducted under high pressure, 14k psi for the long column approach,[68] and 9k psi for the slip-flow approach.[70] The requirement of ultrahigh pressure LC systems has limited the accessibility to these methods. Alternatively, monolithic columns have shown great promise in intact protein separations,[71] and offer several advantages such as high permeability, low backpressure, fast mass transfer, and better recovery.[72] For example, Simone *et al.* developed γ-ray-induced polymethacrylate-based monolithic capillary columns that can provide peak capacity

over 1000 in a 240 min gradient with the addition of 0.1% TFA and column heating at 60 °C.[73] Owing to the low backpressure, analysis time can be as short as 5 min using a higher flow rate, while still offering a peak capacity of 190.[73] Recent developments in longer columns, reduced stationary phase particle size, and monolithic materials have all shown potential to improve intact protein RPC.

Other than RPC, online LC-MS with non-denaturing conditions coupled to top-down analysis was also recently advanced to address pharmaceutical needs to characterize protein aggregates and conjugates.[67,74-76] Online size exclusion chromatography (SEC)-MS,[74] ion exchange chromatography (IEX)-MS,[75,76] hydrophobic interaction chromatography (HIC)-MS,[67] were developed with analytes detected by top-down native-ESI MS yielding lower charge state ions. Muneeruddin *et al.* recently demonstrated the use of IEX with an increasing concentration of salt gradient (ammonium acetate) coupled online to MS and MS/MS to obtain meaningful structural information on protein conjugates, such as PEGylated and glycosylated interferon β-1a.[75,76] In contrast to the increasing salt gradient in IEX, a decreasing salt gradient (e.g. ammonium sulfate) is often used for HIC. As a non-denaturing separation method, HIC offers high sensitivity to the hydrophobic surface of proteins. While recent work has showed that ammonium tartrate does not interfere with downstream MS analysis after a quick desalting step,[64] direct online LC-MS analysis with HIC was not realized until the use of more-hydrophobic HIC materials and ammonium acetate, as demonstrated by Chen *et al*.[67] The HIC-MS remains sensitive to minor modifications and holds a potential for analyzing antibody-drug conjugates in an online manner.[67] The SEC-MS, IEX-MS, and HIC-MS methods share the common feature of utilizing non-denaturing conditions in the form of volatile salts with proper selection of the stationary phase to maintain the efficacy of the chromatographic mode, preserving the noncovalent interactions. Thus,

the ability of these native separation methods to couple online with MS will lead to new avenues for the analysis of intact complexes.

Although LC methods are preferred for direct coupling with MS, capillary electrophoresis techniques, including capillary zone electrophoresis (CZE) and capillary isoelectric focusing (CIEF), have proved to be useful alternatives.[77,78] Recent technical advancement of the sheath-flow and sheathless interface have encouraged more effective use of CZE coupled with top-down MS.[63,79-83] Compared to traditional LC techniques, CZE offers several advantages in sensitivity, resolution, and speed.[84] However, the small sample loading amount has limited the number of total identifications from a protein mixture by CZE alone.[80]

While different chromatographic and electrophoretic methods effectively target a subset of the proteome based on specific physicochemical properties, no single separation strategy can comprehensively resolve the intact proteome. Therefore, multidimensional separation strategies combining orthogonal modes are highly desirable and necessary to achieve a deeper proteome coverage. After implementing RPC as a pre-fractionation step and optimizing fragmentation, Zhao *et al.* identified 180 proteins and 580 proteoforms from yeast in their CZE-MS/MS study in 2016.[85] Similarly, by adding an offline high pH RPC fractionation step, Wang *et al.* doubled the identifications (163 proteins and 328 proteoforms) when compared to the 1D low pH RPC-MS/MS experiment from *E. coli*.[86] The power of additional dimensions for deep proteome coverage was further showcased by the three-dimensional LC-MS approach incorporating offline IEX (first dimension), HIC (second dimension), and online RPC (third dimension) as demonstrated by Valeja *et al*.[66] A total of 640 proteins were identified from a single fraction from IEX (out of 35) as compared to 47 in the 2D approach. Moreover, only minimal improvement was achieved in the protein separation and identifications with simply prolonging the gradients in RPC in the 2D

approach. Therefore, this 3DLC (IEX-HIC-RPC) method greatly outperformed the conventional 2DLC (IEX-RPC) approach demonstrating the great potential for effective separation of intact proteins to achieve deep proteome coverage in top-down proteomics.[66]

Despite the improved resolving power of these multidimensional separation methods, high molecular weight (MW) proteins were not detected with these methods due to a variety of challenges associated with MS analysis of large proteins.[87] First, low MW proteins interfere with the detection of high MW proteins in the mass spectrometer. Second, high MW proteins inherently exhibit lower MS signal due to wider distributions of isotopes and charge states.[87] To address these challenges, practitioners of top-down MS have turned to size-based fractionation prior to MS analysis to alleviate spectral interference from low MW proteins with high MW proteins. This has been well demonstrated in the use of gel-eluted liquid fraction entrapment electrophoresis (GELFrEE).[61,62,88,89] Although effective recovery of intact protein from polyacrylamide gels has been challenging, recent attempts by Takemori *et al.* show progress in utilization of dissolvable gel for top-down analysis.[90]

Recently, without a protein precipitation step, Cai *et al*. developed a MS-compatible serial-SEC (sSEC) strategy that enabled detection of proteins up to 223 kDa using a high-resolution Q-TOF mass spectrometer.[69] By employing a two-dimensional sSEC-RPC approach, more than 4000 unique proteoforms were detected with a 15-fold increase of the proteins above 60 kDa, which greatly outperformed one-dimensional RPC, especially on high MW proteins.[69] Importantly, many of these high MW proteins undetected in one-dimensional RPC possessed important PTMs (e.g. phosphorylation). Therefore, size-based fractionation prior to other dimensions of separation will continue to play an important role in improving coverage of the intact proteome. Despite its effectiveness, the large amount of sample required and labor-intensive offline sample handling

steps have limited MDLC in practice. Although online 2DLC for top-down proteomics has been developed for histone separation,[91] wide adaptation requires further improvement to the robustness of online MDLC hardware and the proper selection of orthogonal separation methods.

1.4 MS/MS methods for top-down proteomics

One of the key merits in top-down proteomics is the prospect of comprehensively interrogating all PTMs within the whole protein sequence from a "bird's-eye" view.[4,8] Efficient and extensive MS/MS fragmentation techniques are the essential prerequisites to characterize proteoforms. Energy-based dissociation processes such as collision-induced dissociation (CID), also referred to as collisionally activated dissociation (CAD); and higher-energy collisional dissociation (HCD),[92] remain the most robust methods and have enabled many large-scale and high-throughput top-down proteomics studies.[47,88,93] These types of energy transfer processes predominantly cleave C-N amide bonds of the protein backbone, leading to abundant *b*- and *y*-type product ions. Generally, the tendency of cleaving labile PTMs (e.g. phosphorylation) has limited the utility of CID in PTM localization and structural examination. However, intact protein ions in the gas phase appear less susceptible to cleavage of labile PTMs during CID or HCD, possibly due to the high-order structure.[94]

In contrast, electron-based methods such as electron transfer dissociation (ETD) and electron capture dissociation (ECD) generally preserve labile PTMs, producing primarily *c*- and $z^{\bullet}$-type ions subsequent to the cleavage of N-C_a bonds.[95,96] In particular, targeted top-down MS with ECD and ETD represents a powerful method to comprehensively characterize biologically relevant proteins, especially those with labile PTMs, as demonstrated on multiple cardiac and skeletal muscle proteins and others.[97-101] For instance, Peng *et al*. recently identified and

characterized a previously unknown phosphoprotein known as enigma homolog isoform 2 from Z-disc by ECD, opening up future functional investigation of this novel phosphoprotein.[97] To address the need of pharmaceutical analysis, characterization of intact or reduced heterogeneous immunoglobulins G (IgG) with electron-based top-down MS/MS techniques continues to improve.[102-105] Benefiting from the advancement of the high-resolution MS instrument, Fornelli *et al.* recently showed top-down analysis of intact IgGs by ETD, achieving comparable ~30% sequence coverage with 2-fold fewer transients summed and 6-fold less acquisition time when compared to the previous study.[102,106]

Moreover, significant progress has been made in developing hybrid dissociation methods such as electron-transfer dissociation/higher-energy collisional dissociation (EThcD)[107] and activated ion electron transfer dissociation (AI-ETD),[108,109] with improved fragmentation for intact proteins. Brunner *et al.* demonstrated that with the proper setting of the collisional energy, EThcD and electron-transfer dissociation/collision-induced dissociation (ETciD) increased the appearance of *b*-/*y*- type ions, while retaining *c*-/ $z^{\bullet}$- type ions that contain phosphorylation site information.[107] Overall, EThcD and ETciD provide more extensive fragmentation than CID/HCD/ETD alone over a broad range of charge states, and retain labile phosphorylation, making these methods suitable for top-down analysis. While AI-ECD methods have been developed and yielded improved ECD product ions, it is largely limited to FT-ICR instruments.[110] By using infrared photon bombardment simultaneously with ETD reaction, Riley *et al.* showcased the improvement of AI-ETD on intact protein characterization over HCD and ETD, particularly on low charge state precursor ions, which extends the utility of electron-based fragmentation to virtually any trap-type instrument in the top-down proteomics field.[108]

Another MS/MS method for intact proteins that recently comes into the spotlight is 193 nm ultra-violet photodissociation (UVPD) pioneered by the Brodbelt group.[12,90,111-116] The high energy deposition of the photons on intact proteins results in a wide array of fragment ions: in addition to *b*-/*y*- and *c*-/$z^{\bullet}$- product ions, *a*-/*x*- from the cleavage of C_a-C bonds and even *d*-, *v*-, and *w*- ions from secondary fragmentation are produced.[12] The rich fragmentation along the protein backbone and the preservation of labile modifications lead to a promising utility of UVPD in pinpointing mutation in sequence variants,[111] unnatural amino acid incorporation,[112] and sites of modification[115] or ligand binding[113] on intact proteins. Because UVPD has minimal dependency on precursor charge states, it not only dissociates low charge state complexes produced by native mass spectrometry into subunits for probing structural topology, but also efficiently fragments monomeric subunits for secondary structural characterization.[90] For streptavidin tetramers, in contrast to CID, UVPD led to symmetric charge partitioning pathways, providing insights into subunit organization and higher order structure.[90] Interestingly, while the resulting monomeric units carried similar charges to that during CID, it remained folded during UVPD as demonstrated by the relative fragment abundances.[90] Other than targeted analysis of a single protein, large-scale proteoform profiling that compares UVPD to HCD, was recently demonstrated.[116] The results showed that while HCD still provided almost twice as much proteoform identification on a chromatographic time scale, UVPD enabled more confident proteoform characterization, indicating the complementary nature of UVPD and HCD.[116] Furthermore, proton transfer reaction (PTR)[117] that focuses multiple charge states of intact proteins into a single lower charge state can be coupled with UVPD as demonstrated by Holden *et al*.[114] The combined ion signals and the relatively efficient UVPD fragmentation on lower charge state ions allow for higher-quality tandem mass spectra of intact proteins.[114] However, applying this strategy in LC/MS analysis

remains challenging because of the time requirement to perform PTR and UVPD in a sequential manner.

As an alternative, front-end ETD, an electrical discharge-based reagent ion source that produces stable and abundant ETD reagent ions, provides flexibilities for ion-ion reactions in the later stage for intact protein characterizaiton.[118] Using sequential front-end ETD and PTR to disperse overlapping fragmentation ions over a broader *m/z* range, Anderson *et al.* showed sequence overage up to 81% for intact histone proteoforms in online LC-MS experiments.[119]

To better implement these fragmentation techniques in high-throughput top-down proteomics on chromatographic time scales, higher quality tandem mass spectra with improved signal-to-noise (*S/N*) and mass accuracy in a single scan is necessary. Under denaturing mode, intact proteins are often highly charged, and the number of charges generally increases proportionally to the molecular weight of the protein.[87] However, a trapping type cell can only accommodate a fixed number of charges during one scanning event. This implies that highly charged ions of large intact proteins or exceedingly large number of fragment ions, will result in limited number of detectable ions, and consequently lower *S/N*. Accumulation of multiple scans can mitigate the space charge effect, but the trade-off is the decrease in duty cycle during online LC-MS/MS analysis. Recognizing this compromise, Riley *et al*. presented a high capacity design of the precursor ion storage in the high pressure cell on the Orbitrap Fusion Lumos.[120] This new scheme allows nearly 3-fold more precursor charges to be accumulated (up to 1,000,000 charges or more), which ultimately leads to *S/N* improvement without sacrificing spectral acquisition speed during ETD experiments for intact proteins.[120] Instead of increasing cell capacity of storing more ions, Holden *et al.* reported a method of overfilling the ion trap with precursor ions and

subsequently implementing resonance ejection of the undissociated precursor ions after UVPD, which resulted in improved *S/N*, higher resolution, and lower mass error for the fragment ions.[121]

Various MS/MS techniques have been developed and are steadily evolving to achieve improved identification and characterization of proteoforms at a higher acquisition rate and a higher *S/N*. Now, nearly complete fragmentation at the single residue level of a 30 kDa intact protein has been demonstrated.[15,111] We envision that the arsenal of top-down MS/MS methods will continue to expand in the next few years, particularly on a chromatographic time scale with higher *S/N* and mass accuracy.

1.5 Data acquisition strategies for top-down proteomics

Top-down proteomics in early years directly adopted highly developed bottom-up data acquisition strategies implemented in commercial instruments. In general, most top-down proteomics experiments are conducted in TopN data-dependent acquisition (DDA), during which the most intense precursor ions from the MS scan are isolated in a relatively large *m/z* window (up to 15 *m/z*) for MS/MS.[47] However, data acquisition is uniquely challenging in top-down experiments. In this section, we will explain some special considerations on this subject and further discuss the strategies developed to address these.

First, ESI of intact proteins generates a wide range of charge states, and precursors of abundant proteoforms in different charge states are selected and fragmented multiple times, often yielding redundant information. Because of this, fragmentation of co-eluting and low-abundance proteoforms might be hindered or missed entirely. Autopilot, an online control system that features online spectral deconvolution, project-wise exclusion list, intelligent MS/MS (HCD and ETD) based on past fragmentation events, combination of product ions, and real-time restricted database

searching, has provided deeper and higher-confident identifications.[122] Nonetheless, in their study, DDA still provided 50 unique protein identifications that were not found during Autopilot acquisition.[122] Therefore, recently introduced and more advanced DDA that can select only the most abundant charge states while putting other redundant charge states on an exclusion list by deconvoluting and assigning charge states on-the-fly further improves intact protein identification and characterization.[48]

Second, although large isolation window (e.g. 15 *m/z*) allows for efficient precursor ion isolation for identification, it might hamper the precise characterization of individual proteoforms. Fornelli *et al*., recently demonstrated the use of 3 *m/z* isolation windows to improve specificity as well as to alleviate space charge effects, and identifed 439 proteoforms in the 30-60 kDa range from human fibroblasts.[10] Using a small isolation window, 0.6 *m/z*, Zhou *et al.* unraveled the complexity of histone PTMs in an online fashion and identified novel modifications including tyrosine bromination and glutathionylation in histone subunits from mouse brain combining a state-of-the-art high-resolution instrument and a new bioinformatics workflow.[48] With advances in mass spectrometers, narrow and efficient isolation will continue to be improved.

Finally, since many top-down studies rely heavily on Fourier transform trap-type instruments (Orbitrap and FT-ICR), the balance among ion accumulation, acquisition time, *S/N*, and mass accuracy during a scanning event critically influences the data quality. Fornelli *et al*. demonstrated improved top-down analysis of higher molecular species (up to 60 kDa) with a benchtop quadrupole-orbitrap.[10] With a shorter transient in the MS^1 scan to eliminate the accumulation of unnecessary noise before the second isotopic beat,[123] enhanced *S/N* of large protein ions was achieved while maintaining high accuracy of the average masses (e.g. 2.5 ppm for a 41 kDa protein).[10] Leveraging the high resolving power, mass accuracy, sensitivity, and

spectral acquisition rate offered by the 21 Tesla FT-ICR mass spectrometer, Anderson *et al.* recently demonstrated a large-scale top-down LC-MS/MS study on human colorectal cancer cells.[124] Instead of summing multiple transients over a longer period, the installation of a multiple storage device (MSD, external quadrupole ion trap) between the Velos ion trap and the ICR cell enables multiple fills of precursor ions or fragment ions in the MSD before high-resolution detection in the ICR cell.[124,125] This strategy not only improves *S/N*, but also boosts the duty cycle in a chromatographic time scale. As a result, the study identified 684 unique protein entries and over 3200 proteoforms in 40 LC-MS/MS runs by front-end ETD and CID, including 372 isotopically resolved proteoforms with MW above 30 kDa.[124]

1.6 Data analysis tools for top-down proteomics

Many fragment ions (in multiple charge states) are produced by top-down MS/MS, resulting in incredibly complex spectra with overlapping charge states. This, combined with the presence of unexpected and combinatorial PTMs, imposes a tremendous computational challenge for top-down data analysis. Generally, the basic analysis of top-down proteomics data consists of isotopic peak picking and deconvolution from the spectra, identification and characterization through database search, validation, and visualization. Over the years, various groups have contributed to each step, pushing the bioinformatics boundaries in top-down proteomics.[126-128] New software packages such as MASH Suite Pro[129], Informed-Proteomics[11], and ProteinGoggle 2.0,[130] as well as new search algorithms such as pTop[131], TopPIC (from MS-Align+),[132] among others,[133,134] were spotlighted in the past two years.

The resulting complexity of intact protein (MS^1 and MS^2) spectra produced by ESI, necessitates deconvolution algorithms to combine all charge states and determine monoisotopic

masses. The first algorithm for deconvolution of high-resolution mass spectra, THRASH, uses a subtractive peak finding method with charge determination and a monoisotopic mass determining strategy based on the averagine model.[135] This influential method is still widely used through adaptation and improvement, which has led to the development of Decon2LS,[136] DeconMSn,[137] Xtract,[138] MASH Suite,[139] as well as MS-Deconv,[126,127] an alternative combinatorial algorithm. In 2015, Marty *et al.* proposed UniDec, a Bayesian framework to separate mass and charge dimensions for fast deconvolution of ion mobility MS data, which could also be potentially applied to top-down data analysis.[140] In 2016, Sun *et al*. introduced a novel algorithm, pTop with a pre-processing module pParseTD, which greatly improves the accuracy of precursor detection.[131] For the human histone data set, pTop recalled 22% more correct precursors than Xtract with 30 % less exported masses.[131] The improvement was facilitated by a machine-learning module, pParseTD, which evaluates and characterizes candidate isotopic clusters through features obtained from experimental isotopic distribution, LC profiles, and the same protein ion in different charge states. Similarly, taking advantage of the isotopic clusters in different charge states and LC-MS profiles, ProMex in the recently published Informed-Proteomics package, generates a list of deconvoluted isotopic masses from LC-MS data by aggregating signals from different charge states over a LC elution time span and examining the aggregated isotopomer envelope.[11] Compared to other MS^1 feature-detection algorithm, such as ICR-2LS (http://omics.pnl.gov/software/icr-2ls) and MS-Deconv, ProMex accurately and reproducibly detected a significantly higher number of LC-MS features across all ten replicates of ovarian tumor samples, which has led to comparable coefficients of variation to label-free bottom-up studies.[11]

After deconvolution, a database search strategy is commonly used for protein identification and proteoform characterization. Given the likelihood of observing combinatorial or even

unexpected PTMs and proteolytic truncations in a top-down proteomics workflow, strategies such as extended database search (e.g. ProteinGoggle,[130,141] ProSightPC absolute mass search), blind PTM search (e.g. MS-Align+[142] (recently renamed to TopPIC[132] with improvements), PIITA[143]) or the combination of the two (ProSightPC biomarker search mode, MS-Align-E[144]) are mainly applied to interrogate proteoforms.[145]

The recent pTop uses a sequence-tag based strategy and a scoring approach to greatly accelerate the database searching process, with the introduction of indexes of both sequence tags and combinatorial modifications, leading to 10- to 100-fold decrease in time required than MS-Align+ on complex data sets.[131] Nevertheless, truncated proteoforms are not yet considered during searches in the initial version of pTop. On the other hand, TopMG, a mass graph-based approach, enables the identification of proteoforms with terminal truncations, although it requires prolonged running time.[146] With dramatically faster running speed and higher numbers of identified proteins compared to ProsightPC (V3.0) and MSAlign+, MSPathFinder, a command-line algorithm, plays an important role in the recent Informed-Proteomics package.[11] It utilizes a graph-based approach and a *de novo* sequencing algorithm based on sequence tags to improve search efficiency.[11] However, MSPathFinder requires pre-specified PTMs, and therefore does not discover unknown PTMs.

In 2016, MASH Suite Pro was introduced as a comprehensive package for large-scale top-down proteomics, which offers deconvolution (THRASH and MS-Deconv), identifications (MS-Align+), in-house developed quantitation and characterization of proteoforms.[129] It greatly simplifies and speeds up the interpretation of high-resolution MS and MS/MS data with a user-friendly and customizable interface.[129] Furthermore, with very few other softwares capable of manual validation, MASH Suite Pro provides a useful visualization tool to correct mis-assignment

of charge state and isotopic distributions, facilitating accurate and reliable proteoform characterization. In this regard, LcMsSpectator in the Informed-Proteomics package also provides an interactive results viewer to simplify the inspection of MS^1 precursor match and MS^2 fragment match.[11,48]

Statistical significance of the identification and characterization of proteoforms is evaluated by a number of parameters. *p*-value and *E*-value measure how well the fragmentation data matches the identified sequence (the lower the value, the less likely the match is random by chance), although computation of these values slightly varies in different software.[142] In general, a false-discovery-rate of 1% governs the confidence level of the identification in large-scale proteoform studies with a target/decoy approach.[10,47] Furthermore, introduction of C-score has improved identification and characterization of proteoforms, particularly in a high throughput fashion.[147] However, it is not suitable for unknown modifications.[127] Recently, also based on Bayesian models, MIScore was introduced by Kou *et al.* for proteoform characterization, providing a simple and fast method to identify and localize up to two unknown modifications in proteoforms.[127] Ongoing development and adaptation of these scoring validations among laboratories will improve workflows and provide more consistency within the community.

Over the years, a plethora of informatics tools for top-down proteomics have been developed which greatly streamlined the data analysis process. Although challenges remain (e.g. identification of larger proteins, utilization of internal fragments), [69,148] collaboration based on many open-source platforms provides tremendous opportunities for top-down bioinformatics to further advance.

1.7 Native top-down proteomics

The improvement of non-denaturing separation, fragmentation techniques, as well as instrument advances in transferring and detecting high *m/z* ions have brought native mass spectrometry and top-down proteomics together. In addition to the native separation methods described previously (e.g. SEC-MS, IEX-MS, and HIC-MS),[67,74-76] clear native GELFrEE (CN-GELFrEE) and native CZE were recently developed for separation of noncovalent assemblies of biomolecules.[148,149] Through a native and multistage (e.g. MS^3) MS approach, the protein complexes, monomer units, and monomer fragments can be readily detected from the fractions.[150] To facilitate identification and characterization of protein assemblies, a computational framework with scoring metrics was developed by Skinner *et al.*.[151] By utilizing CN-GELFrEE and the newly developed informatic framework, Melani *et al.* were able to access the large protein complexes from the king cobra venom proteome.[151] They identified two of the largest venom glycoprotein complexes, homodimeric L-amino acid oxidase (~130 kDa) with different numbers of N-glycan moieties and the multichain toxin cobra venom factor (~147 kDa) by HCD.[151] More recently, Belov *et al.* demonstrated a proof-of-principle use of native sheathless CZE coupled online to MS and MS/MS (HCD) to analyze standard protein complexes, monoclonal antibody aggregation, and *E. coli.* ribosomal extract at femtomole levels.[149] With the identification of several non-covalent protein-protein and protein-metal ion ribosomal protein complexes, this approach shows the potential to characterize protein interactions in complex biological samples with high sensitivity and low sample consumption.[149] Furthermore, the recent investigation of gas phase fragmentation propensity of HCD during denaturing and native top-down MS of by Haverland *et al*., will provide insights for developing tailored scoring metrics to improve identification of native complexes in the future.[152]

Other than identifying high MW protein assemblies from a mixture, targeted analysis of a single native complex by various fragmentation techniques can reveal non-covalent binding regions and other structural information.[152-156] Using a 15T FT-ICR mass spectrometer, Li *et al*. demonstrated that ECD of a 158 kDa aldolase tetramer under native conditions provides rich fragments in the region that ECD and CAD failed to access to under denaturing conditions.[153] Nonetheless, depending on the specific structure of the protein complex, ECD sometimes provides minimal fragmentation information. Native top-down with 30 eV electron ionization dissociation (EID), as an alternative, preserves ligand-protein or protein-protein interactions and delivers much more structural information compared to ECD in certain cases, as demonstrated by a purified apo-superoxide dismutase dimer.[155] Providing comparable fragmentation to UVPD, EID has proved to be a useful technique as it requires no additional instrument modification in FT-ICR instruments. Similarly using a targeted native top-down approach, in 2017, Schneeberger and Breuker successfully probed the binding sites of a model transactivation response RNA to a trans-activating peptides at the single–residue level and captured time-resolved information on the stoichiometry of the complex using CAD.[154] Other than using electron and collisional energy to dissociate protein complexes and fragment monomeric units, UVPD, as previously mentioned, can also provide structural characterization for protein complexes by retaining noncovalent interaction upon dissociation.[12] For instance, UVPD on native dihydrofolate reductase complexes enabled the assessment of the binding regions of cofactor NADPH and inhibitor methotrexate.[113] Besides these fragmentation techniques, surface-induced dissociation (SID) also offers insights on structural topology with symmetric charge partitioning.[155]

The native top-down approach has expanded the toolbox and will continue to enable the study of larger complexes in the proteome beyond the normal detection limit of denaturing top-

down proteomics. With the further development in non-denaturing separation strategies, we expect online analysis and characterization of native complexes to become an increasingly important area in top-down proteomics.

1.8 Quantitative top-down proteomics

To further interrogate how proteoforms are associated with complex disease phenotypes, the acquirement of quantitative measurement in addition to qualitative information becomes critical. Over the recent years, top-down quantitative proteomics has expanded from examination of PTM changes of a few targeted proteoforms to investigation of expression changes in large-scale proteoform studies.[93]

Within the same top-down mass or tandem mass spectra, relative quantitation of unmodified and modified proteoforms using intensity ratios was first demonstrated feasible by Pesavento *et al*., and used in many cardiac biomarker and others studies.[53,55,98,157,158] This method is extremely powerful to globally examine histones as shown in recent studies using high-resolution MS, providing extensive elucidation and relative quantitation of histone variants and modifications in different samples.[159-161] The relative quantitation of proteoforms abundance relies on the idea that the physicochemical properties of intact proteins in the gas phase during electrospray ionization is less affected by small modifications or variations than that of peptides;[157] however, it is worth noting that different modifications might have an impact on ionization, fragmentation, and therefore the accuracy of the quantitation.[162,163]

In the realm of quantifying proteoform changes across multiple samples in large-scale biomarker discovery experiments, measuring relative ratio of proteoforms has limited applicability. Incorporation of labeling techniques such as stable isotope labeling by amino acids (SILAC),[164]

isobaric and pseudoisobaric tags,[165-167] as well as NeuCode SILAC,[168] have shown some potential. For instance, the intact-mass strategy developed by the Smith group utilizes NeuCode SILAC to determine lysine count for the elucidation of proteoform families.[169,170] However, variation in labeling efficiency and reproducibility might lead to spreading (lower *S/N*) and asymmetry of protein ion signals.[171] In addition, dependence on precursor ion isolation, instrument resolution, and specialized data analysis software, have limited the adaptation of these labeling approaches.[172] To overcome some of these limitations, in 2016, Quijada *et al.* introduced a cost-effective partial metabolic labeling strategy coupling to top-down proteomics for living organisms, tunable intact mass protein mass increases (TIPMI), by spiking ^{13}C-glucose or ^{2}H-water into laboratory feedstocks.[173] Compared to SILAC, this method results in a more symmetric and predictable mass shift for heavy and light peak pairs, as the percentage of ^{13}C (or ^{2}H) incorporation in a protein deviates less than the percentage of lysine or arginine.[173] Although this labeling approach provides outstanding precision and accuracy, potential overlapping of the unlabeled proteoform peaks and the heavy peaks, and software requirement for picking labeled pairs remain to be addressed for large-scale quantitation. Therefore, the more accessible label-free techniques stay appealing in top-down quantitation proteomics.

The advantages of label-free quantitation such as simplicity and throughput have been demonstrated previously.[174,175] Recently, Ntai *et al.* applied and extended this idea to a larger scale platform with a hierarchical linear model for statistical analysis, which quantified over 800 proteoforms with more than 100 significantly different between two yeast samples.[176] Using this quantitation pipeline and the Autopilot acquisition strategy, recently Durbin *et al*. achieved identification and quantitation of thousands of proteoforms under 30 kDa in two human fibroblast states.[93] In addition, the new Informed-Proteomics package improves quantitation with the

enhanced detection of LC-MS features by ProMex and identification with MSPathFinder, resulting in a significant increase in the number of differentially expressed proteoforms detected when compared to a previous study on the same samples.[11,177] In short, these top-down label-free strategies and others all generally include the following steps: automatic isotopic peak picking, integration, normalization, stringent proteoform binning and filtering, and statistical analysis.[11,176,178] Another simple label-free quantitation method, spectrum counting, was recently evaluated by Geis-Asteggiante *et al.* by comparing to other label-free methods (e.g. AUC).[172] Although spectrum counting offers comparable (slightly less) accuracy, it is a simple and robust preliminary screening method for putative differentially expressed targets.[172] Moreover, to ensure the quality of the label-free quantitative experiment, large sample size, and multiple biological and technical replicates are often necessary.

It is important to recognize that quantitation is most accurate only within the linear response as implied in many quantitation works.[172,174,176] Protein signals in MS that fall close to or out of the upper and lower ends of the curve will result in over- or under-estimation of the true expression changes. Unfortunately, for intact proteins, the dynamic range and the instrument response greatly vary among instrument types; and oftentimes, sensitivity differs for each protein.[179-181] Nevertheless, with advancement of instrumentation and optimization of workflows, top-down label-free quantitation holds great potential in translational and clinical research.

1.9 Top-down proteomics applications

Top-down proteomics is uniquely equipped to identify and characterize proteoforms without prior knowledge.[4,8,9,94] Recently it has been widely applied in discovery research especially to study biological systems that have complex PTM combinations and a large number

of variants, for instance, histones and venom proteins.[151,182,183] More importantly, utilizing clinical samples (e.g. blood, tissues, cerebrospinal fluid, saliva), an increasing number of top-down applications have recently shown promise in the fields of cancer, neurodegenerative, and cardiovascular diseases.[56,97,177,178,184-192] Owing to the potential for providing a precise and comprehensive view of all proteoforms, top-down proteomics is thought to be more directly connected to complex disease phenotypes.[19] Thus, this powerful method is particularly valuable in understanding disease mechanisms and discovering potential biomarkers as well as therapeutic targets.[4,122] Here, we highlight a few recent and relevant examples in global profiling and targeted analysis of proteoforms in this fast growing field to demonstrate the value of studying proteoforms in biological and biomedical research.

1.9.1 Global profiling

As previously discussed, improved sample prefractionation methods, LC-MS/MS workflows, and label-free quantitation of top-down proteomics have led to the ability to profile samples from healthy to diseased states.[93] Global top-down proteomics in discovery mode allows for the examination of potential biomarkers at the proteoform level. Cheon *et al.* quantitatively compared the low-MW proteome (<30 kDa) of human plasma from healthy control and colorectal cancer patients.[185] In addition to identifying proteoforms containing various PTMs (e.g. acetylation, S-glutathionylation, and *O*-glycosylations), they detected previously-unreported coexisting pyroglutamylation at Gln24 and S-cysteinylation at Cys29 on Apolipoprotein A-II (8804.48 Da).[185] By using label-free quantitation, they have identified 17 low MW proteoforms that showed >1.5-fold changes from the colorectal cancer samples, some of which were verified by Western blot analysis and were consistent with the literature.[185]

Recently, the Kelleher group comparatively analyzed peripheral blood mononuclear cells from kidney and liver transplant recipients in two different studies.[186,187] For the study on kidney transplant recipients, the top-down label-free approach with GELFrEE fractionation and nanocapillary LC-MS/MS identified 2905 proteoforms corresponding to 344 proteins and preliminarily suggested changes between the two patient groups with transplant excellent and with acute rejection.[186] For the more recent liver transplant studies, among the differentially expressed proteoforms, those associated with chemokine/cytokine signaling and cytoskeletal regulation appeared most significant.[187] For instance, platelet factor 4 (PF4/CXCL4) was found more abundant in transplant excellent than acute rejection samples, which have shown potential protective effect by facilitating blockage of Th17 differentiation in other transplantation models.[187] Importantly, out of the 15 proteoforms of PF4/CXCL4, resulting mainly from N-terminal proteolytic processing, only 3 differed with statistical significance.[187] These results indicate the possibility that only specific proteoforms are related to certain disease phenotypes, and label-free top-down analysis at the proteoform level provides an indispensable tool for investigating this hypothesis.

Other than clinical blood samples, analysis of tumor tissues from patient-derived mouse xenograft models of basal and luminal B human breast cancer, WHIM2 and WHIM16, was also recently demonstrated by two independent top-down proteomics studies.[11,177] Ntai *et al.* quantified 982 proteoforms from 358 proteins (<30 kDa), and differentiated PTM changes such as phosphorylation level alternation of α-endosulfine between WHIM2 and WHIM16 samples using label-free top-down quantitation.[177] Although the bottom-up approach on the same samples yielded eight-times more identifications, it overlooked PTM changes and relative expression of heterozygous alleles in the proteoform level.[177] Using the same samples but with improved top-

down workflows, Park *et al.* demonstrated the detection of over 7000 differentially expressed proteoforms.[11] Within the 3207 identified proteoforms, 1636 proteoforms were found with adjusted *P* value < 0.01 and fold change > 2 in two subtypes of breast type tumor samples.[11] These differentially expressed entities were found about 10 times more than what Ntai *et al*. reported with much shorter instrument acquisition time.[11,177] As much as top-down proteomics has advanced in global proteoform profiling and quantitation in clinical research, these potential biomarkers found in blood and tumor tissue samples require further follow-up targeted analysis and validation, as for many other global proteomics experiments.

1.9.2 Targeted analysis

While large-scale global proteoform profiling in discovery mode is powerful, the hypothesis-driven targeted proteomics approach that detects and quantifies specific proteoforms from a sub-proteome or a purified system with high sensitivity and reproducibility has often offered insights in the understanding of underlying molecular mechanisms of normal and disease-associated cellular events, and provided potential utility in clinical diagnostics.[4,193-195] For instance, in addition to the detection of novel truncated and glycosylated proteoform, Gafvels *et al.* revealed that levels of glycosylated apolipoprotein A-I (ApoA-1) distinguished significantly from the serum sample of the diabetic to non-diabetic patients using a semi-quantitative top-down LC-MS approach, suggesting glycated ApoA-1 may serve as a potential biomarker for diabetes.[196]

To improve the selectivity and specificity of proteoform characterization, affinity purification is often coupled with top-down analysis, as demonstrated by the Ge group and others.[16,17,53-58] Elucidation of potential molecular mechanisms related to proteoforms greatly benefits from this immunoaffinity top-down MS method. Recently, Carel *et al.* used a top-down

targeted approach with other techniques to decipher the role of o-mycoloylation in targeting the outer membrane proteins (OMPs) to the mycomembrane in bacteria.[197] Subsequent to the expression and affinity purification of recombinant OMPs, PorA, ProH, PorB, and PorC of bacteria *C. glutamicum*, top-down LC-MS/MS analysis of the OMPs associated with mycoloyl-arabinogalactan-peptidoglycan (mAGP) complex and secreted in the extracellular medium, revealed well-conserved PTMs, including *O*-mycoloylatoin, pyroglutaminatoin, and N-formylation.[197] Among these modifications, in particular, *O*-mycoloylatoin was only found in the mAGP-associated proteoforms, indicating that the presence of mycoloyl residues is essential for targeting OMPs to the mycolic acid-containing lipid bilayer.[197] The top-down approach plays a particularly important role in this case, because these OMPs from *C. glutamicum* are relatively hydrophobic with the presence of PTMs containing C32-C36 mycolyl residues and lack arginine and lysine residues for generating adequate tryptic peptides. Similarly, in another microbiology study on *M. tuberculosis*, using affinity purification and top-down MS, Parra *et al.* uncovered a complex repertoire of nearly 130 proteoforms of virulence associated 19 kDa lipoglycoprotein antigen (LpqH) resulting from various combination of lipidation, glycosylation, and proteolytic truncations.[197] In addition to the examination of molecular diversity, the identification of a novel phosphorylation on unprocessed LpqH further provided insights to microbial biogenesis.

As an example of the application in translational research, applying relative quantitation in a targeted approach, Gregorich *et al.* studied sarcopenia (muscle loss with aging) by assessing age-related changes in the myosin regulatory light chain (RLC) proteoforms from the fast-twitch skeletal muscles of rats.[198] RLC is a critical protein involved in the modulation of muscle contractility, and the results revealed a significant progressive decline in the RLC phosphorylation with increasing age.[198] Top-down MS/MS identified a previously unreported bis-phosphorylated

proteoform of fast skeletal RLC, and further localized the sites of decreasing phosphorylation to Ser14/15 in addition to a N-terminal trimethylation.[198] Subsequent mechanical analysis of the single muscle fiber from rats of different ages revealed that the decrease in RLC phosphorylation is responsible for altered mechanical function in aged muscle fibers, such as significant decreases in maximal force, the Ca^{2+}-sensitivity of force, loaded shortening velocity, and power output.[198] Hence, these studies have demonstrated the great potential of top-down targeted proteomics for quantifying proteoform changes across biological samples and elucidating disease mechanisms at a molecular level. [97,198,199]

Chapter 2: Online LC-MS/MS Top-down Phosphoprotomics Strategy Enabled by Functionalized Cobalt Ferrite Nanoparticles

This chapter has been published and is adapted from:

Chen, B.; Hwang, L.; Ochowicz, W.; Lin, Z.; Guardado-Alvarez, TM.; Xiu, L.; Dani, K.; Colah, C.; Jin, S.; Ge, Y. Coupling Functionalized Cobalt Ferrite Nanoparticle with Online LC/MS/MS for Top-down Phosphoproteomics. *Chemical Science.* 2017, 8, 4306-4311.

2.1 Abstract

Phosphorylation plays pivotal roles in cellular processes and dysregulated phosphorylation is considered as an underlying mechanism in many human diseases. Top-down mass spectrometry (MS) analyzes intact proteins and provides a comprehensive analysis of protein phosphorylation. However, top-down MS-based phosphoproteomics is challenging due to the difficulty in enriching low abundance intact phosphoproteins as well as separating and detecting the enriched phosphoproteins from complex mixtures. Herein, we have designed and synthesized the next generation functionalized superparamagnetic cobalt ferrite ($CoFe_2O_4$) nanoparticles (NPs), and have further developed a top-down phosphoproteomics strategy coupling phosphoprotein enrichment enabled by the functionalized $CoFe_2O_4$ NPs with online liquid chromatography (LC)/MS/MS for comprehensive characterization of phosphoproteins. We have demonstrated the highly specific enrichment of a minimal amount of spike-in β-casein from a complex tissue lysate as well as effective separation and quantification of its phosphorylated genetic variants. More importantly, this integrated top-down phosphoproteomics strategy allows for enrichment, identification, quantification, and comprehensive characterization of low abundant endogenous phosphoproteins from complex tissue extracts on a chromatographic time scale.

2.2 Introduction

Reversible protein phosphorylation orchestrates important cellular processes in eukaryotic cells including cell proliferation, protein-protein interaction, and signal propagation, among others.[200,201] Dysregulation of protein phosphorylation is known to be critically involved in the pathogenesis of many human diseases, such as cancer, cardiomyopathy, and Alzheimer's.[200,202-206] To understand these disease processes, identification and quantification of phosphoproteins are the essential steps in decoding the complex cellular signaling network regulated by phosphorylation events. Moreover, proteoforms arising from alternative splicing, amino acid polymorphisms, and other post-translational modifications (PTMs),[207] have imposed another layer of complexity to the phosphoproteome. The conventional shot-gun "bottom-up" mass-spectrometry (MS) based phosphoproteomics which analyzes peptides resulted from protein digestions,[208,209] is suboptimal because of the "protein inference problem", resulting in incomplete information on sequences or modifications.[20,210] In contrast, the top-down MS approach examines intact proteins without digestion, providing a global view and the possible merit of comprehensive analysis of all modified proteoforms.[94,204,211-213] Therefore, top-down MS allows one to study phosphoproteins and their associated proteoforms at the intact protein level, which leads to more direct connectivity to complex disease phenotypes for clinical and translational research.[20]

However, unlike the well-established bottom-up phosphoproteomics that has numerous mature phosphopeptide enrichment methods and well-developed LC/MS/MS workflows,[214-216] the top-down phosphoproteomics approach is much more challenging mainly due to the lack of phosphoprotein enrichment methods and the difficulties in separation of intact proteins from complex proteome. To address the limitation of low specificity, efficiency, and poor reproducibility on intact protein enrichment,[217-220] we recently developed the first generation

functionalized superparamagnetic Fe_3O_4-GAPT-Zn nanoparticles (NPs) for universal and effective enrichment of phosphoproteins from complex mixtures.[51] However, further improvement on the NPs and effective downstream analysis including LC separation and MS/MS characterization of enriched phosphoproteins are needed to fully interrogate the complex phosphoproteome.

Herein, we sought to develop a robust top-down proteomics workflow for comprehensive phosphoprotein characterization that integrates highly specific intact phosphoprotein enrichment by the next generation cobalt ferrite ($CoFe_2O_4$) NPs with high-resolution online LC/MS/MS analysis (**Figure 2.1**). In this top-down phosphoproteomics workflow, proteins are first extracted from tissue homogenate (or cell lysate), and subsequently incubated with functionalized superparamagnetic NPs (**Figure 2.1a**). At a neutral pH, the ligands functionalized on the NP surface bound strongly to phosphate group of the phosphoproteins, preserving their physiological condition (**Figure 2.1b**). After a magnetic pull down of the phosphoprotein-bound NPs, elution of the phosphoproteins from NPs (as well as the pre-enrichment samples) are then subjected to online LC separation and MS analysis of the intact proteins (**Figure 2.1c**). The enriched phosphorylated species are selected for tandem MS for identification and characterization. For the first time, we have shown that this integrated top-down phosphoproteomics strategy allows for highly specific enrichment, identification, quantification, and comprehensive characterization of low abundant endogenous phosphoproteins from complex tissue extracts on a chromatographic time scale.

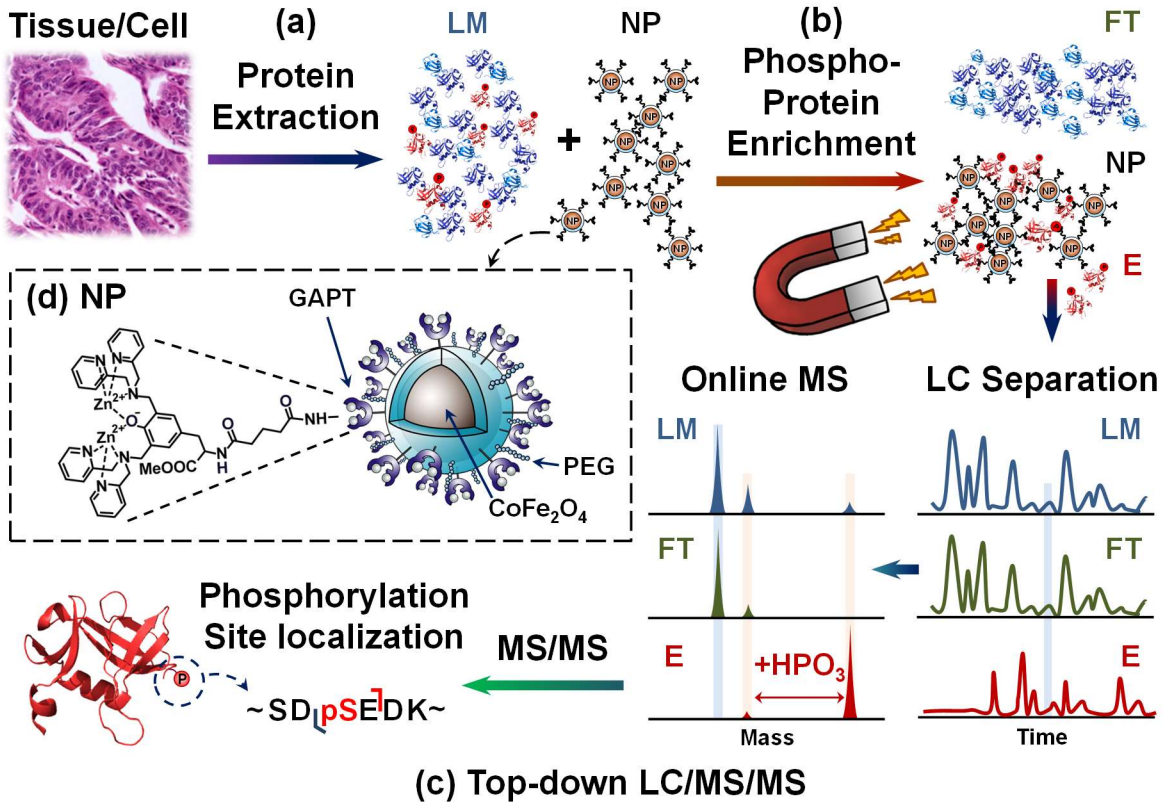


Figure 2.1. Schematic illustration of the top-down phosphoproteomics strategy integrating intact phosphoprotein enrichment using functionalized magnetic nanoparticles (NPs) with online LC/MS/MS. (a) Proteins were extracted from tissue (or cell), and subjected to enrichment by functionalized CoFe2O4 NPs. (b) Phosphoproteins bound on magnetic NPs were then pulled down by a magnet for subsequent elution. (c) The loading mixture (LM), flow through (FT), and elution (E) were further separated and analyzed by top-down LC/MS. Enriched phosphoproteins are then comprehensively characterized by online LC/MS/MS. (d) Details of the $CoFe_2O_4$ NPs functionalized with GAPT-Zn ligands that specifically bind to phosphate groups.

2.3 Material and Methods

2.3.1 Chemicals and reagents

All chemicals and reagents were used as received without further purification unless otherwise noted. N-(tert-Butoxycarbonyl)-L-tyrosine methyl ester (Boc-Tyr-OMe), di-(2-picolyl)amine (DPA), paraformaldehyde, trifluoroacetic acid (TFA), glutaric anhydride, N-

ethyldiisopropylamine (EDIPA), Iron (III) acetylacetonate (acac) (97%), benzyl ether (98%), 1,2-hexadecanediol (90%), oleic acid (OA) (99%), oleylamine (OE) (>70%), zinc chloride ($ZnCl_2$), anhydrous N,N-dimethylforamide (DMF), and anhydrous dichloromethane were purchased from Sigma-Aldrich (St. Louis, MO, USA). *N,N,N′,N′*-Tetramethyl-*O*-(1*H*-benzotriazol-1-yl) uronium hexafluorophosphate (HBTU) were obtained from TCI America (Portland, OR, USA). (3-aminopropyl)trimethoxysilane (APTMS) and 2-methoxy (polyethyleneoxy) propyl trimethoxysilane (average molecular weight =550-600, hereafter referred as Si-PEG) were purchased from Gelest (Morrisville, PA, USA). Bovine serum albumin (BSA), β-casein, and pepsin were purchased from Sigma-Aldrich. For the enrichment of β-casein for further top-down MS analysis, β-casein (MS grade) was purchased from Protea (Morgantown, WV, USA). All proteins were used without further purification or modification. 50 mM HEPES (4-(2-hydroxyethyl)-piperazineethanesulfonic acid) buffer was made by directly diluting 0.5 M HEPES buffer (pH=7.7; home-made with HEPES obtained from Fluka) with 18 MΩ nanopure de-ionized water from Milli-Q water (Millipore, Corp., Billerica, MA, USA). 100 mM sodium phosphate buffer was prepared by diluting 0.5 M sodium phosphate (pH 7.3; home-made with sodium phosphate dibasic obtained from Sigma-Aldrich). Bradford protein assay reagent was purchased from Bio-Rad (Hercules, CA, USA). 12.5% mini-gel (10 comb well, 10.0 cm × 10.0 cm) for SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) was home-made. 12.5% Criterion precast gel (18 comb well, 13.3 cm × 8.7 cm (W×L)) was purchased from Bio-Rad (Hercules, CA, USA). Pierce™ Concentrators PES (polyethersulfone)-0.5 mL centrifugal filter with a molecular-weight cutoff of 10k was purchased from ThermoFisher Scientific (Rockford, IL, USA). Sypro Ruby and Pro-Q Diamond staining solutions were purchased from ThermoFisher Scientific (Rockford, IL,

USA). Phos-tag gel staining solution was purchased from GeneCopoeia (Rockville, MD, USA). 3,4-hexanedione was perchased from Sigma-Aldrich.

2.3.2 Synthesis of functionalized $CoFe_2O_4$ nanoparticles (NPs)

$CoFe_2O_4$-OA/OE NPs. The cobalt ferrite NPs were prepared by a seed-mediated growth method following the procedure reported by Sun *et.al*.[1] Briefly, $Co(acac)_2$ (1 mmol), $Fe(acac)_3$ (2 mmol), 1,2-hexadecanediol (10 mmol), OA (6 mmol), and OE (6 mmol) were mixed in benzyl ether (20 mL) to synthesize $CoFe_2O_4$ seed NPs. The mixture was heated to 200 °C for 2 h with stirring under N_2 (g), and then heated to reflux at ~300 °C for 1 h. After the mixture was cooled to room temperature, EtOH (40 mL) was added to the mixture. A black precipitate was obtained by centrifugation, which was subsequently redissolved in hexane in the presence of OA (50 μL) and OE (50 μL). Undissolved residue was removed by centrifugation (5000 rpm, 20 min). The supernatant was precipitated with EtOH, centrifuged (5000 rpm, 20-30 min) to remove the solvent, and then redispersed in hexane. ~80 mg of the $CoFe_2O_4$ seed NPs was further grown to a bigger size of $CoFe_2O_4$ NPs by adding $Co(acac)_2$ (1 mmol), $Fe(acac)_3$ (2 mmol), 1,2-hexadecanediol (10 mmol), OA (2 mmol), and OE (2 mmol) in benzyl ether (20 mL). The mixture was heated to 100 °C for 30 min and then to 200 °C for 1h with stirring under N_2 (g), and then heated to reflux at ~300 °C for 30 min. After the mixture was cooled to room temperature, EtOH (40 mL) was added to the mixture. As described in the synthesis of $CoFe_2O_4$ seed NPs, the workup procedures were repeated to obtain $CoFe_2O_4$-OA/OE NPs with a size of ~10 nm.

$CoFe_2O_4$-NH_3/PEG NPs and $CoFe_2O_4$-GAPT NPs. The procedures of NP functionalization by trialkoxysilane molecules and GAPT ligand molecules have been described in the supporting material of our previous communication with minor modifications.[2] Briefly,

under ambient conditions, 0.5% (v/v) trialkoxy silane solution in hexane containing APTMS and Si-PEG in a ~2:1 molar ratio was added to the $CoFe_2O_4$ NPs dispersed in hexane (~ 50 mg in 120 mL). The mixture was stirred and refluxed at 70-72 °C for 24 h, during which a black-brown precipitate was formed. The precipitate was washed three times with hexane and three times with EtOH to remove excess silane molecules in solution. Finally, the NPs were dried for the next reaction.

GAPT (0.546 mmol) was dissolved in anhydrous DMF (~6 mL), and then HBTU (0.82 mmol) and EDIPA (3.28 mmol) were added to the solution. The $CoFe_2O_4$-NH_3/PEG NPs were redispersed in DMF (~2 mL). EDIPA (0.7 mmol) was added to this NP solution and it was sonicated for 1 min. Then the DMF solution containing GAPT, HBTU, and EDIPA was added, and the whole mixture was stirred at room temperature for 24 h. The resulting black-brown precipitate was washed two times with DMF, one time with DMF/MeOH (v/v 3:1), one time with DMF/MeOH (v/v 1:1), and one time with MeOH to remove unreacted GAPT, HBTU, and EDIPA. The resulting $CoFe_2O_4$-GAPT NPs were redispersed in EtOH at a concentration of ~50 mg/mL.

$CoFe_2O_4$-GAPT-Zn NPs. A 120-150 µL aliquot of $CoFe_2O_4$-GAPT NP stock solution (at ~50 mg/mL) was transferred to a 2 mL vial. The NPs (~6.0-7.5 mg) were washed with de-ionized water three times to remove EtOH. After 1.5 mL of 10 mM $ZnCl_2$ (aq) was added to the $CoFe_2O_4$-GAPT NPs, the mixture was agitated for 3.5 h at room temperature. The $CoFe_2O_4$-GAPT-Zn NPs were precipitated by a magnet (centrifugation was also used to collect the NPs if needed) and then sonicated in de-ionized water to re-disperse. The final product, $CoFe_2O_4$-GAPT-Zn NPs, was used for enriching phosphoproteins.

2.3.3 Characterization of the as-synthesized and functionalized NPs

Transmission electron microscopy (TEM) samples were prepared by pipetting one drop of as-synthesized $CoFe_2O_4$-OA/OE NP solution onto a copper TEM grid with carbon film. TEM was conducted on Philips CM200 Ultra Twin instrument operated at 200 kV and images were collected using a Gatan CCD image system with digital micrograph software program. Powder X-ray diffraction (PXRD) data were collected on as-synthesized $CoFe_2O_4$-OA/OE NPs deposited on a glass substrate using a Bruker D8 Advance using Cu Kα radiation (λ= 1.5418 Å), and the background from the glass substrate was subtracted. Transmission Fourier transform infrared (FT-IR) spectroscopy measurements were recorded on a Bruker Equinox 55 FT-IR spectrometer in the range of 4000 cm^{-1} to 600 cm^{-1} at about 2 cm^{-1} resolution by using potassium bromide (KBr) pellet. Thermogravimetric analysis (TGA) was carried out using a TGA Q500 thermal analysis system. All TGA measurements were performed under a N_2 atmosphere at a constant heating rate of 20 °C/min from 100 °C to 600 °C. All samples were first heated to 100 °C and held at that temperature for 3 min to remove adsorbed water. Data were analyzed using the Thermal Advantage (TA) universal analysis software program. Dynamic light scattering (DLS) was carried out using a Malvern Zetasizer Nano. All DLS measurements of individual samples were performed three times after 120 sec equilibration.

2.3.4 Protein sample preparation

Each of the standard proteins BSA, β-casein, and pepsin was dissolved in 50 mM HEPES buffer (pH 7.7, with 150 mM NaCl) at a concentration of about ~15-20 mg/mL to be used as stock solutions. For the test of enrichment performance of $CoFe_2O_4$-GAPT-Zn NPs, simple mixtures

containing 300 µg of BSA, 100 µg of β-casein , and 100 µg of pepsin in 1.0 mL of 50 mM HEPES buffer (pH 7.7 ± 0.1, 150 mM NaCl) were prepared.

Swine hearts were excised from healthy Yorkshire domestic pigs, snap frozen in liquid N_2, and stored under -80 °C before use. First, the frozen pig tissue samples (~1 g) were cut into small pieces and immediately washed twice in cold 25 mM HEPES buffer (pH 7.4). The tissue samples were homogenized 5-6 times using a Polytron electric homogenizer (Model PRO200, PRO Scientific Inc., Oxford, CT, USA) for 5-7 secs on ice in buffer containing 25 mM HEPES (pH 7.4), 50 mM NaF, 0.5 mM phenylmethanesulfonyl fluoride (PMSF), 2.5 mM EDTA, and protease inhibitor cocktail (Roche, Switzerland). After centrifugation, the remaining pellet was briefly homogenized two times in the same buffer followed by 30 min of centrifugation at 16,000 ×g at 4 °C. The supernatants were collected, the concentration of the tissue lysate was determined by Bradford protein assay. The samples were stored at -80 °C for later study.

The swine heart tissue extract was diluted in loading buffer consisting of 50 mM HEPES buffer (pH 7.7) with 150 mM NaCl to achieve a final protein concentration of ~0.8-0.9 mg/mL (based on Bradford protein assay). Various concentrations of β-casein solution were spiked into the complex tissue extract solution.

2.3.5 Verification of enrichment by SDS-PAGE analysis

For standard protein mixtures, the combined elution fractions were concentrated to ~100 µL and then the protein solutions before and after enrichment were separated by SDS-PAGE (12.5% mini-gel). For swine heart tissue extract containing varying amounts of the spiked-in β-casein, 10 µg of protein mixtures before and after enrichment were separated on a Criterion Tris-HCl (10.5-14%) precast gel (Bio-Rad). The gels were first stained with Phos-tag or Pro-Q diamond

solution for visualizing phosphoporteins and then Sypro Ruby solution for total proteins. The gel images were taken by a Typhoon 9200 imager (GE Healthcare, Bio-Science, Piscataway, NJ, USA) with excitation at 532 nm and emission filter of 580 nm or a ChemiDoc^{TM} MP Imaging System (Bio-Rad, Hercules, CA, USA).

2.3.6 Western blot

Equal amount of samples from loading mixture, flow through, and elution was loaded and sepeartated by SDS-PAGE, and was subsequently transferred to PVDF membranes (Millipore, Billerica, MA, USA). Membranes were then blocked with protein-Free Blocking buffer (Thermo Scientific) and bloted with antibodies against cTnI phosphorylated at Ser 22/23 (Cell Signaling Technology, Beverly, MA, USA).

2.3.7 Reverse phase chromatography (RPC)

Reverse phase chromatography was performed with an ACQUITY M-Class UPLC system (Waters; Milford, MA, USA). Mobile phase A (MPA) contained 0.1% formic acid in water, and mobile phase B (MPB) contained 0.1% formic acid in 50/50 ethanol/acetonitrile. For each injection, 5 µL of sample was loaded on a PLRP-S (5 µm, 1000Å) capillary trap column (nanoLCMS LLC; Gold River, CA, USA) for online desalting with 95% MPA at 6 µL/min for 4 minutes. A home-packed 200 mm × 250 µm PLRP-S (5 µm, 1000Å; Agilent Technology, Santa Clara, CA, USA) column was used at a constant 4 µL/min flow rate. The RPC gradient consisted of the following concentrations of MPA: 95% for 2 mins, 20% at 60 min, 15% at 61 min, and back to 95% at 70 min. Each run was 75 min long.

2.3.8 Top-down MS and MS/MS analysis

Samples eluted from RPC separation were electrosprayed into a maXis II ETD Q-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) for online LC/MS and LC/MS/MS experiments. End plate offset and capillary voltage were set at 500 and 4000 V, respectively. The nebulizer was set to 0.5 bar, and the dry gas flow rate was 4.0 L/min at 220 °C. The quadrupole low mass was set to 600 *m/z* during MS and 200 *m/z* during MS/MS. Mass range was set to 200-3000 *m/z* and spectra were acquired at 1 Hz for LC/MS runs. For the top 3 data-dependent LC/MS/MS CID runs, spectra were acquired across 200-3000 *m/z* at 2 – 6 Hz with active exclusion after 4 spectra. Targeted LC/MS/MS CID or ETD was performed at 1 Hz after determining the elution time frame from the targeted proteins. The ETD reagent was the radical anion of 3,4-hexanedione (114 *m/z*). All data were collected with OtofControl 3.4 (Bruker Daltonics). Data were analyzed and processed in DataAnalysis 4.3 (Bruker Daltonics). Maximum Entropy algorithm (Bruker Daltonics) was used to deconvolute all mass spectra with resolution set to 80000. The SNAP algorithm was applied to determine the monoisotopic mass of all detected ions.

2.3.9 Protein identification and fragmentation mapping

Fragmentation ion lists consisting of monoisotopic mass, intensity and charge were generated from DataAnalysis 4.3, and were subsequently converted to MSAlign files. Alignment-based MS-Align+ algorithm[3,4] for intact protein identification based on protein spectrum matches, was used to search against the Uniprot-Swissprot Sus scrofa (Pig) database, which was released on May 14th 2016 and contains 26167 protein sequences. Fragment mass tolerance was set to 15 ppm. All identifications were validated with statistically significant *P* and *E* values (<0.01) and satisfactory numbers of assigned fragment (>10).

2.3.10 Calculation of relative abundance of phosphorylated β-casein proteoforms

We divided the ion intensity of all β-casein proteoforms by the total ion intensity across the same elution time frame. [221]

2.4 Results and Discussion

2.4.1 Synthesis, characterization, and evaluation of $CoFe_2O_4$ NPs for intact phosphoprotein enrichment

We have designed and synthesized the cobalt ferrite ($CoFe_2O_4$) NPs (**Figure 2.1d** and **Figure 2.2a**) with stronger magnetic response[222,223] and better reproducibility for intact phosphoprotein enrichment than the first generation Fe_3O_4-GAPT-Zn NPs.[51] We synthesized the $CoFe_2O_4$ NPs capped with oleic acid (OA) and oleylamine (OE) ligands by a seed-mediated growth method (see experimental details in SI and **Figure S2.1**) following the procedure reported by Sun *et.al.*[224] Transmission electron microscopy (TEM) image analysis revealed that the diameter of the $CoFe_2O_4$ NPs was ~9.96 ± 1.03 nm (**Figure 2.2b**). Powder X-ray diffraction confirmed a typical ferrite diffraction pattern[225] (**Figure 2.2c**). The $CoFe_2O_4$ NPs have been shown to possess larger magnetic moments and magnetic anisotropy and good chemical stability.[222,223] These NPs were subsequently functionalized with dinuclear Zn (II)-dipicolylamine (Zn-DPA) ligands coupled to glutaric acid (hereafter referred to as GAPT), based on the previously reported method with minor modifications (**Figure 2.2a**).[51] These GAPT-Zn complexes provide the specific binding to the phosphate ions on phosphoproteins and are the key to the enrichment.[51,226] Notably, the step that replaces the hydrophobic OA and OE ligands on the $CoFe_2O_4$ NP surface with 3-aminopropyl trimethoxy silane (APTMS) and 2-methoxyl (polyethyleneoxy) propyl trimethoxysilane (hereafter

referred to as Si-PEG) was completed within 24 h, which was greatly improved compared to 48 h for the preparation of Fe_3O_4–NH_3/PEG NPs.[51] We then coupled GAPT ligands, the phosphate capturing moiety, to the free amino groups of the APTMS on $CoFe_2O_4$-NH_3/PEG NPs, producing $CoFe_2O_4$-GAPT NPs. Fourier transform infrared spectroscopy (FT-IR) indicating strong band around 1147-1028 cm^{-1}, peaks around 1200 cm^{-1} (Si-O-R, C-O-C, EO CH_2), and sharp peak around 1650 cm^{-1} for $CoFe_2O_4$-GAPT (amide bond), and thus confirmed the proper functionalization of the NPs with the ligand molecules (**Figure 2.2d**).[227] Additionally, thermogravimetric analysis (TGA) showed different weight loss (%) for each surface-modified $CoFe_2O_4$ NPs with the corresponding ligand molecules (**Figure 2.2e**). The small size of the $CoFe_2O_4$ NPs provides advantages such as a high surface-to-volume ratio, easy surface functionalization with multivalent ligand molecules to interact with phosphoprotein targets; better penetrations in complex mixture for higher binding rate without causing denaturation; and good solubility.[228-230]

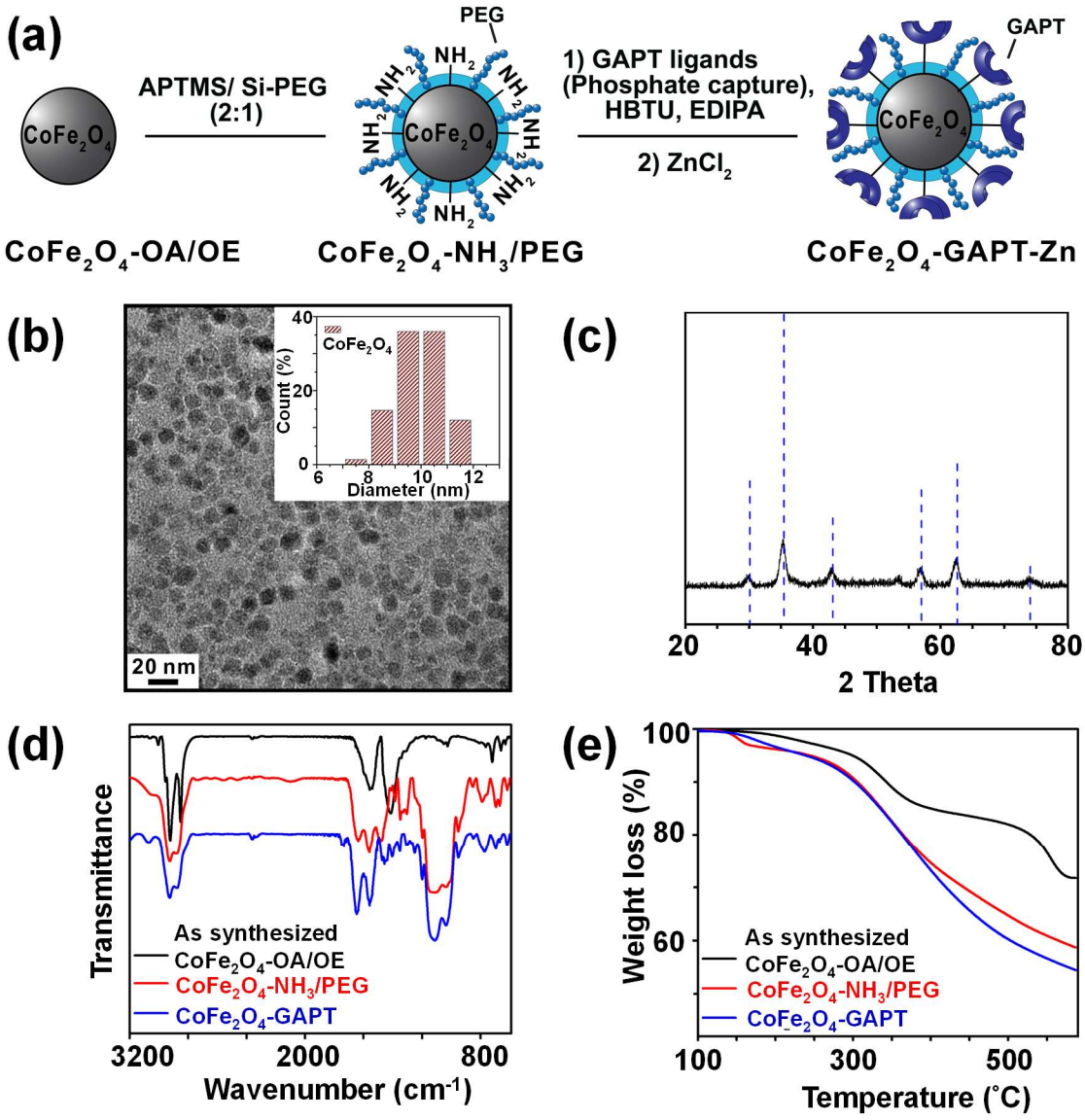


Figure 2.2. Synthesis and characterization of functionalized $CoFe_2O_4$ NPs. (a) Schematic illustration of the synthesis of functionalized $CoFe_2O_4$ NPs. The GAPT-Zn chelating groups process strong binding ability and preference for phosphate dianions at neutral pH. (b) TEM image of as-prepared $CoFe_2O_4$-OA/OE NPs. Inset shows the size distribution (9.96 ± 1.03 nm) of the NPs. (c) PXRD pattern of as-synthesized $CoFe_2O_4$-OA/OE NPs. Blue dash lines indicate a typical diffraction pattern of a ferrite structure. (d) FTIR and (e) TGA analysis of as-synthesized $CoFe_2O_4$-OA/OE (black), functionalized $CoFe_2O_4$-NH_3/PEG (red), and functionalized $CoFe_2O_4$-GAPT (blue) NPs.

After activation of $CoFe_2O_4$-GAPT NPs with 10 mM $ZnCl_2$, the resulting $CoFe_2O_4$-GAPT-Zn NPs (**Figure 2.1d**) were used for phosphoprotein enrichment. The protein samples prior to NP enrichment were termed loading mixture (LM); unbound non-phosphoproteins were collected as flow-through (FT); and enriched phosphoproteins were saved as elution (E) for further analysis (**Figure 2.1c**). The optimized enrichment procedure includes a HEPES buffer (50 mM, pH 7.7, 150 mM NaCl) for binding and subsequent washing, and a sodium phosphate (Na_2HPO_4) buffer (100 mM, pH 7.3, 50 mM NaCl) for final elution. A standard protein mixture was used to validate the enrichment efficacy of each batch of NPs (details in SI, **Figure S2.2a**, **b**). We further sought to quantitatively determine the enrichment performance by spiking in low levels of β-casein as an internal control in a highly complex swine heart tissue extract. We systematically decreased the concentration of the spike-in β-casein from 0.1 to 0.03 μg/μL. 10 μg of equal amount of proteins before and after enrichment of the tissue extract containing about 3 % (I), 5 % (II), 7 % (III), and 10 % (IV) w/w % of β-casein (Table S1) was prepared as loading mixtures. SDS-PAGE analysis was performed (**Figure 2.3a, b,** and **Figure S2.3**) paralleled to the top-down LC/MS analysis (*vide infra*).

Overall, the enrichment was consistent and reproducible across all samples with different amounts of spike-in β-casein. Abundant non-phosphoproteins in LM were effectively depleted and washed into FT, as shown in the nearly identical bands between LM and FT (**Figure 2.3a**). In contrast to LM and FT, the highly similar band patterns in the elution (E) lanes of the gel stained by Sypro Ruby and ProQ Diamond (**Figure 2.3a, b**) indicates the predominance of enriched endogenous phosphoproteins in addition to β-casein. These results demonstrated that $CoFe_2O_4$-GAPT-Zn NPs can efficiently enrich phosphoproteins from complex biological samples with high specificity and sensitivity.

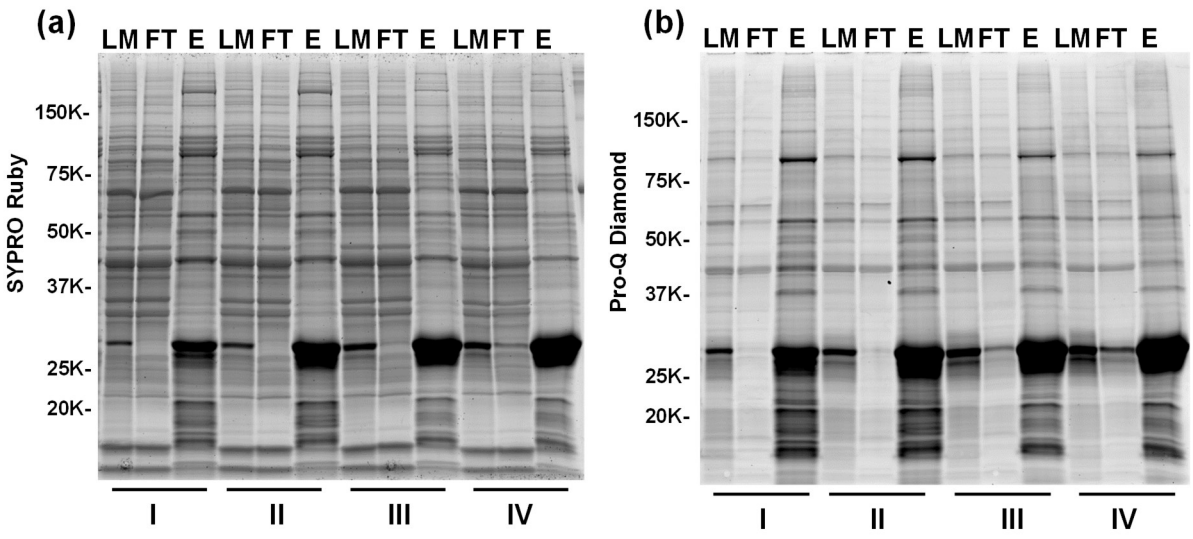


Figure 2.3. SDS-PAGE analysis of phosphoprotein enrichment by functionalized $CoFe_2O_4$ NPs from complex lysates. SDS-PAGE analysis with (a) SyproRuby and (b) Pro-Q Diamond based detection of spike-in β-casein and other endogenous phosphoproteins specifically enriched from highly complex swine heart tissue extracts (loading mixture). Each loading mixture (LM) contains different concentration of 0.03 μg/μL (I), 0.05 μg/μL (II), 0.07 μg/μL (III), and 0.1 μg/μL (IV) of spiked-in β-casein. Equal amount (10 μg) of the loading mixture, flow-through (FT), and elution after enrichment (E) was loaded on the gel. M: marker; LM: loading mixture; FT: flow through; E: elution.

2.4.2 Top-down LC/MS analysis of spike-in β-casein from tissue extract after $CoFe_2O_4$ NP enrichment

For the top-down proteomic analysis after NP enrichment, an equal amount of proteins from LM, FT, and E that were acquired from the tissue extract with 5 % of β-casein spike-in was subjected to reverse phase chromatography (RPC) LC/MS and LC/MS/MS analyses using a Waters M-class UPLC and a Bruker maXis II Q-TOF mass spectrometer (details in SI). The robust LC/MS platform allows direct comparison across the base peak chromatograms (BPC) of LM, FT, and E (**Figure 2.4a**). The overlapping separation pattern between LM and FT (**Figure 2.4a** and **Figure S2.4**) demonstrates that the majority of the proteins from LM were washed off to FT,

consistent with the SDS-PAGE band patterns (**Figure S2.3**). Subsequent MS analysis revealed that most proteins in the LM and FT solutions are highly abundant non-phosphoproteins, including blood proteins such as hemoglobin subunit α and β, which were significantly depleted after enrichment (**Figure S2.5 and S2.6**). Even though some blood proteins were still detected in MS from elution, significant depletion of highly abundant non-phosphoproteins in LM enabled us to detect low abundant phosphoproteins in E, which are reflected as low abundant peaks in the BPC for E. For example, the deconvoluted spectrum shows that the spike-in β-casein was barely detectable in LM and FT solutions because of the predominant signal from the co-eluting myoglobin (**Figure 2.4b**). However, after enrichment using $CoFe_2O_4$-GAPT-Zn NPs, β-casein (M_r 23,968.12) was significantly enriched while the intensity of myoglobin (M_r 16,942.86) was greatly decreased (**Figure 2.4b**). The enrichment factor was estimated by comparing the relative percentage of β-casein before and after the enrichment. The relative percentage of β-casein was significantly increased from 0.5 % to 94 % from the complex mixture after enrichment. Moreover, the pure spike-in β-casein standard were separated into three distinct LC peaks in E after enrichment, as seen in the BPC and extracted ion chromatogram labeled as i, ii, and iii in **Figure 2.4a**. None of these were detectable prior to enrichment in LM. The deconvoluted mass spectra revealed three genetic variants of β-casein[221], B (M_r 24,077.14; i), A1 (M_r 24,008.11; ii), and A2 (M_r 23,968.12; iii) based on their accurate mass measurement (**Figure 2.4c**, **Figure S2.7**). Although the genetic variants of β-casein differ by only one or two amino acid residues, all three proteoforms were unbiasedly enriched by the NPs and separated from the complex tissue extract using top-down LC/MS. Furthermore, the top-down approach allowed us to quantify the relative percentage of each β-casein proteoform, which would be difficult using the bottom-up MS strategy. The relative percentages among A2, A1, and B variants were determined to be 53 %, 35

% and 12 % respectively, according to the average spectrum across the same time frame (**Figure S2.8**).

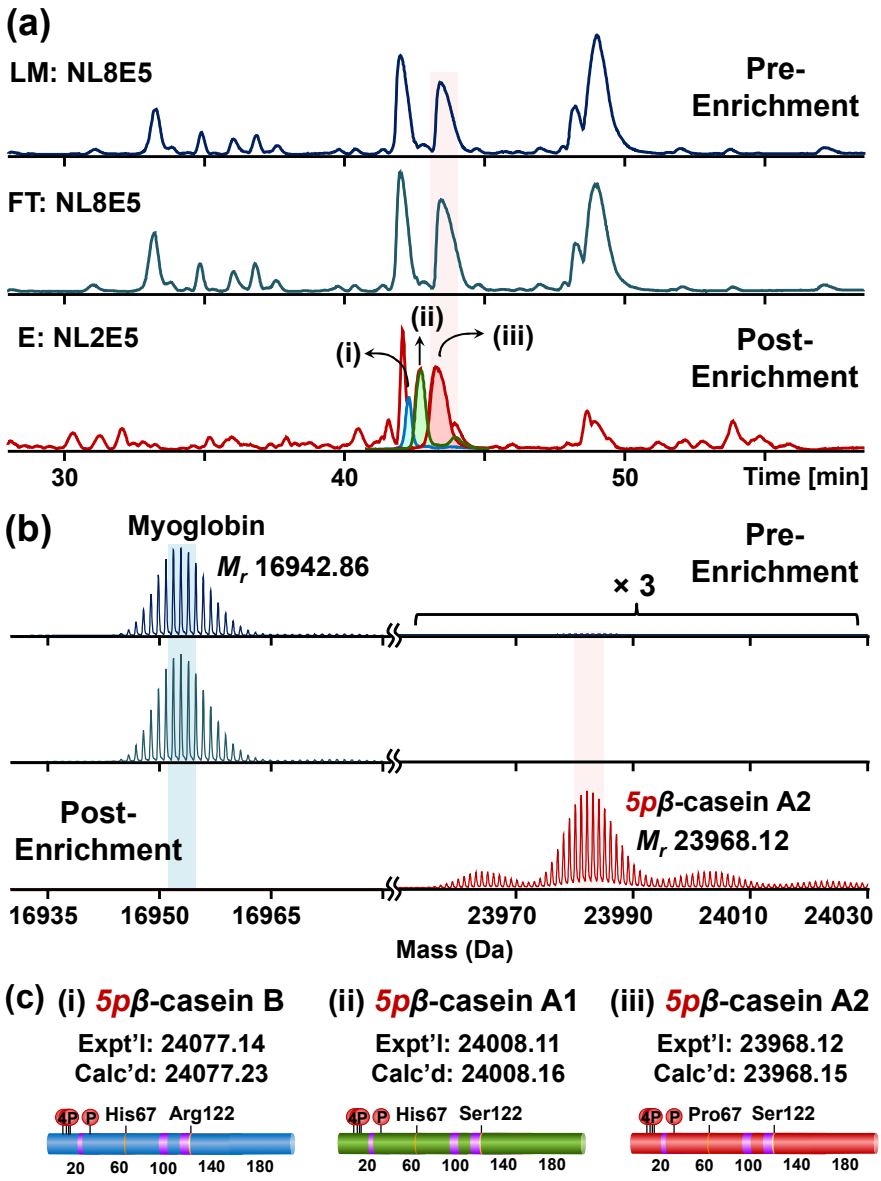


Figure 2.4. Top-down LC/MS analysis of multiple β-casein variants after phosphoprotein enrichment from a swine heart tissue extract with 0.05 μg/μL of spike-in β-casein. (a) Base peak chromatograms of LM (loading mixture; dark blue), FT (flow through; light blue), and E (elution; red). i, ii, and iii, represent extracted ion chromatograms of β-casein variants. (b) Average mass spectra and deconvoluted spectra of LM (dark blue), FT (light blue), and E (red) samples from 41.3 to 41.7 min (highlighted in grey). The intensity in the region from 23,940 Da to 24,040 Da was magnified 3 times. NL: Normalized level. (c) Cartoon illustration of the three β-casein variants B (i), A1 (ii), and A2 (iii), each carrying five phosphorylations.

2.4.3 Top-down LC/MS/MS characterization of endogenous phosphoproteins enriched from tissue extract by CoFe2O4 NPs

Furthermore, through this integrated top-down phospho-proteomics strategy, we have detected and examined multiple known cardiac phosphoproteins[206] from the eluted phosphoproteins that were previously undetected or in very low abundance in the HEPES extraction without enrichment (**Figure 2.5**). The enrichment of the phosphoproteins is apparent by comparing phosphorylation level changes across the same time frame in the LC/MS of LM, FT, and E. Notably, as shown in the normalized deconvoluted spectra in **Figure 2.5a**, phosphorylated forms of pig cardiac troponin I (24 kDa) with a possible V116A polymorphism (28 Da) were all enriched, which was consistent with what we previously found on the same protein using immunoaffinity purification.[231] Western blot analysis for phosphorylated cardiac troponin I at Ser22/23 further demonstrated the enrichment (**Figure S2.9**), which well correlates with the MS data. Nevertheless, the top-down approach in this case offers a global view of each of these highly similar phosphorylated polymorphic species. Similarly, phosphorylated tropomyosin (33 kDa) was also enriched and can be characterized after enrichment (**Figure 2.5b**).

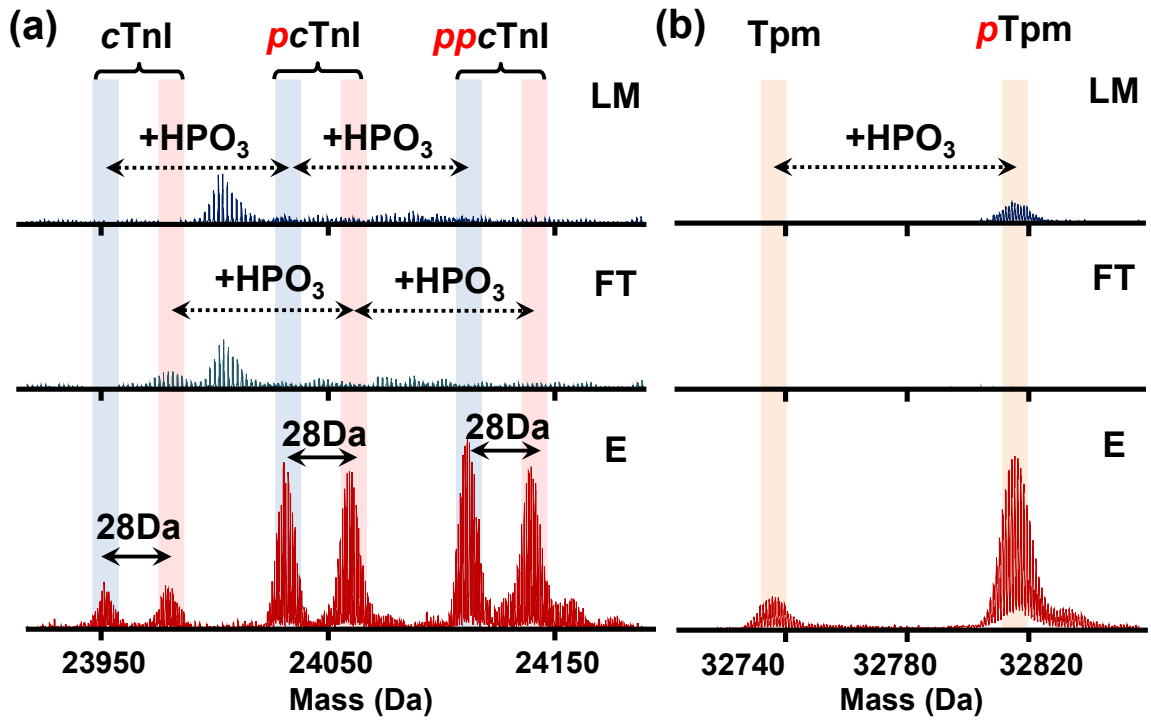


Figure 2.5. LC/MS detection of well characterized cardiac phosphoproteins enabled by highly specific NP-based enrichment from a complex tissue extract. Representative endogenous phosphoproteins, (a) cardiac troponin I (cTnI) and (b) tropomyosin (Tm) were shown by comparing normalized deconvoluted average MS spectra of loading mixture (LM), flow through (FT), and elution from 37.5-37.8 min and 41.3-41.4 min, respectively. Mass difference of 28 Da corresponding to polymorphism V116A was observed for cTnI.

More importantly, in addition to the previously known myofilament proteins, we were able to discover low abundant endogenous phosphoproteins using this integrated strategy. One example is a potential phosphoprotein (27 kDa) eluting around 33.5 min that was significantly enriched so that the triply phosphorylated form became the most intense peak/base peak (**Figure 2.6a**). The deconvoluted spectrum further shows that the protein with M_r 26,996.90 has 80 Da mass increases, corresponding to multiple phosphorylated forms of the protein (**Figure 2.6b**). As seen in both the original and the deconvoluted spectra, the triply phosphorylated form was of very low abundance in LM of the heart tissue extract, but after enrichment, its intensity improved more than 10 folds

(**Figure 2.6a**, **b**). Moreover, nearly no detection of the phosphorylated forms in FT spectra demonstrates sufficient capturing ability of the $CoFe_2O_4$-GAPT NPs. Since the triply phosphorylated form became the base peak after enrichment, subsequent LC/MS/MS was made possible through both collision-induced dissociation (CID) and electron transfer dissociation (ETD). We confidently identified this protein as hepatoma-derived growth factor (UnitProtKB/Swiss-Prot, F1RHJ2_PIG) using MS-Align+[129,232] search algorithm (**Figure S2.10**), which was previously uncharacterized and unreviewed. We further analyzed the data and mapped the sequence using MASH suite Pro.[129] Three potential phosphorylation sites were determined: Ser132, Ser133, and Ser165 (red letters in **Figure 2.6c** and **Figure S2.11**). Top-down MS/MS not only revealed an N-terminal methionine excision, but also a coexisting acetylation near the N-terminal in addition to the three phosphorylations. As another example, a protein with M_r 21,675.63 with a mass increase of 80 Da was significantly enriched and detected in E (**Figure S2.12a**), while it was nearly unobservable in LM and FT. In the normalized original mass spectra in Figure S2.12a, the non-phosphoprotein with M_r 20,827.56 was fully depleted, and the phosphorylated protein M_r 21,675.63 + 80 Da was enriched more than 10 folds, demonstrating the specificity of this NP-based enrichment and the efficacy of this integrated strategy. MS-Align+ search identified the proteins to be chromobox protein homolog 1 (UnitProtKB/Swiss-Prot, F1RWH1_PIG) based on both the CID and ETD spectra acquired from LC/MS/MS and we narrowed down the possible phosphorylation site to Ser89/Ser91 (**Figure S2.12** and **S2.13**). Similarly, we enriched and identified a 14 kDa singly phosphorylated protein, programmed cell death protein 5 (UnitProtKB/Swiss-Prot, F1RNX2_PIG) (**Figure S2.14** and **S2.15**). Interestingly, even though phosphorylation is known as a labile PTM, many of the phosphorylations were preserved during CID LC/MS/MS (**Figure 2.6d**, **Figure S2.12d**, and **S2.16**). For instance, from

the phosphoprotein with M_r 21,755.59, $_p$y121 ion was predominantly observed whereas its unphosphorylated counterpart y121 existed in low abundance (**Figure S2.16**). The globular structure of intact proteins preserves the labile phosphorylation to some degree.[94] Other enriched phosphoproteins with observable 80 Da mass difference were also detected in the comparison among LM, FT, and E (**Figure S2.17**). These results in the identification and characterization of new phosphoproteins not only validated the enrichment performance of the $CoFe_2O_4$-GAPT-Zn NPs but also demonstrated the power of this integrated top-down phosphoproteomics workflow.

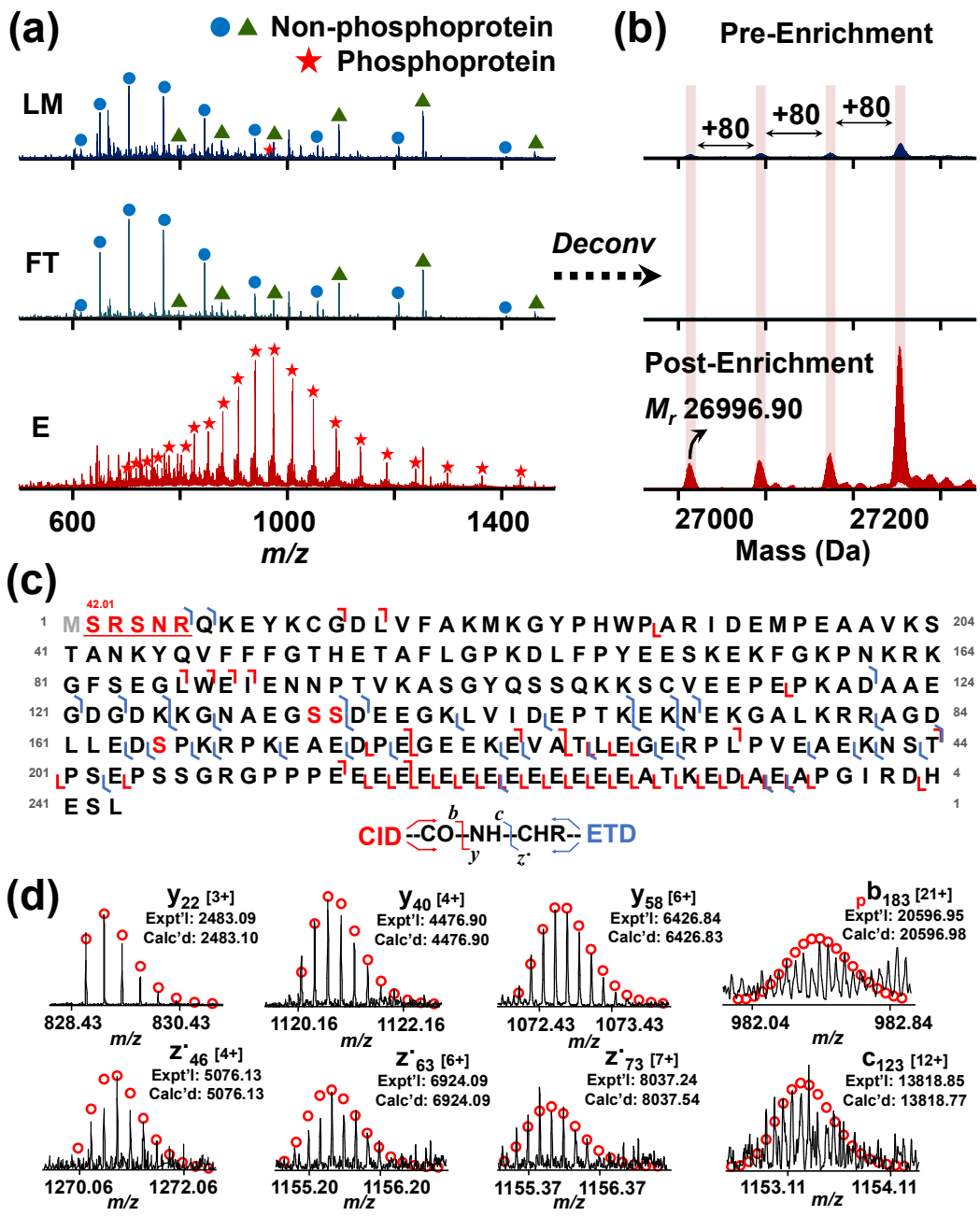


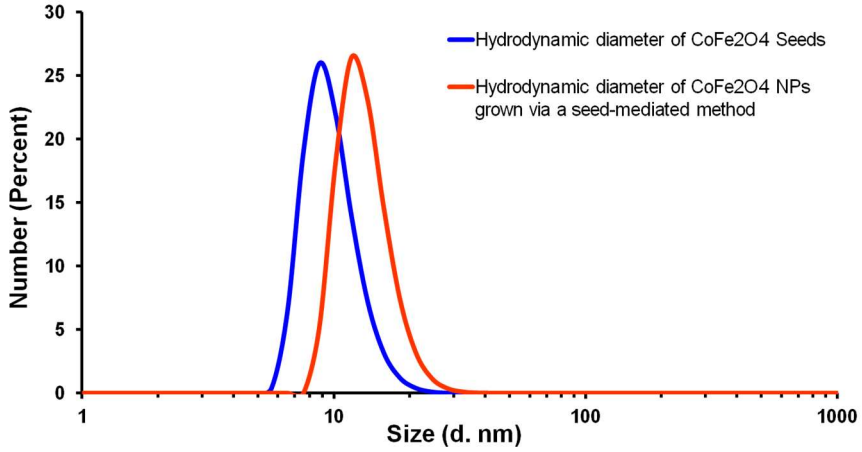
Figure 2.6. LC/MS/MS analysis of a representative previously uncharacterized phosphoprotein, hepatoma-derived growth factor, enabled by effective $CoFe_2O_4$ NP-based enrichment from a complex swine heart tissue extract. (a) MS spectra of loading mixture (LM, dark blue), flow through (FT, light blue), and elution (E, red) from 32.3 min to 32.7 min; (b) the corresponding deconvoluted spectra. (c) Fragment ion map of online MS/MS analysis with CID and ETD from triply phosphorylated precursor ion (29+). Grey "M" indicates methionine excision and red "S" indicates phosphorylation sites. Red numbers above the underlined red sequence reveal modifications with their additional mass. (d) Representative fragment ions from both online CID and ETD LC/MS/MS of the targeted protein.

2.5 Conclusions

In summary, we have developed a top-down phosphoproteomics strategy that integrates NP-based phosphoprotein enrichment and online LC/MS/MS to enrich, identify, quantify, and characterize phosphoproteins from complex protein mixtures. We have designed and synthesized functionalized magnetic $CoFe_2O_4$-GAPT-Zn NPs with stronger magnetic response and much improved robustness and reproducibility in synthetic steps. The high specificity and efficiency of phosphoprotein enrichment using these improved $CoFe_2O_4$-GAPT-Zn NPs was validated by SDS-PAGE analysis with phosphoprotein-specific gel stains using β-casein spike-in tissue extract. Top-down LC/MS results demonstrated that the spike-in phosphorylated β-casein was substantially enriched (relative percent from 0.5 % to 94 %), which allowed us to further separate, identify, and quantify all of its genetic variants. More importantly, through online top-down LC/MS/MS with CID and/or ETD, we were able to identify endogenous phosphoproteins and localize their potential phosphorylation sites. Benefiting from the unique NP-based enrichment method, this integrated online top-down LC/MS/MS workflow has the advantages of globally observing phosphorylated proteoforms; quantifying proteoform species; and identifying and characterizing phosphoproteins with co-existing PTMs and/or sequence variants. We envision that with further advances in front-end intact protein separation,[233] MS detection of large proteins (>30 kDa), and MS/MS fragmentation techniques[111,234] with improved throughput, this top-down phosphoproteomics strategy that couples functionalized NP-based phosphoprotein enrichment with online LC/MS/MS has the potential to enable a deep coverage and comprehensive characterization of the phosphoproteome for the study of molecular mechanism underlying diseases and the discovery of specific phosphoproteoforms as biomarkers.

2.6 Supplemental Information

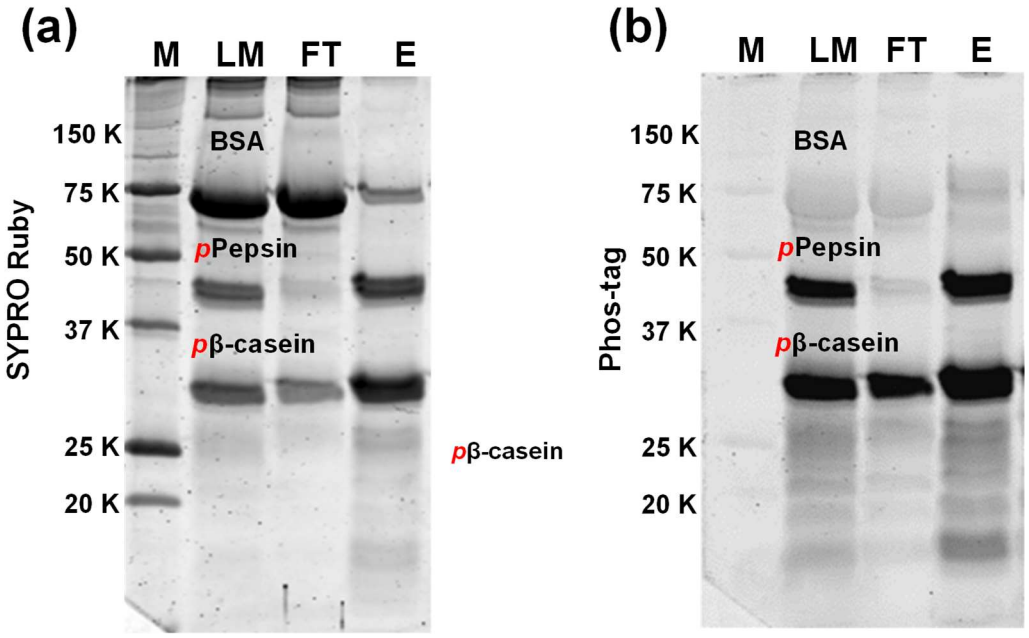
Part II. Supplementary data on the characterization of $CoFe_2O_4$ NPs



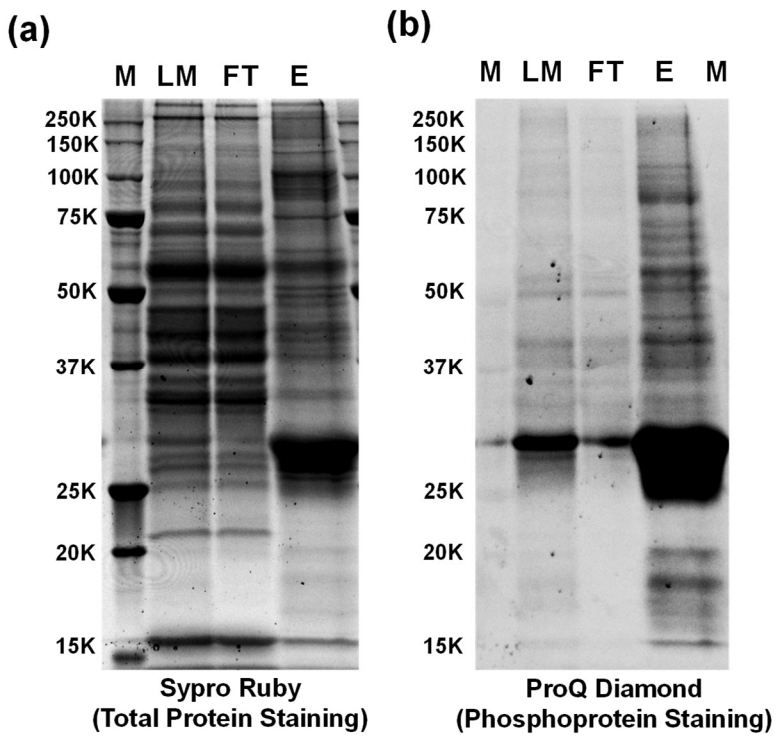
Supplemental Figure 2.1. Particle-size (average hydrodynamic diameter) distribution obtained by dynamic light scattering (DLS). DLS shows the as-synthesized $CoFe_2O_4$ seeds (blue, 9.07 ± 2.31 nm) and $CoFe_2O_4$ NPs (red, 11.77 ± 2.93 nm) produced via a seed-mediated growth method.

Part III. Supplementary data on paralleled SDS-PAGE analysis of phosphoproteins enriched using $CoFe_2O_4$ NPs

We tested the phosphoprotein enrichment from a standard protein mixture containing bovine serum albumin (BSA, a non-phosphoprotein), pepsin (phosphoprotein with one phosphorylation site), and β-casein (phosphoprotein with five phosphorylation sites) in 3:1:1 mass ratio. After SDS-PAGE analysis of the pre- and post-enrichment solutions, the gel was first stained by Phos-tag staining to visualize phosphoproteins, and then by Sypro Ruby for total protein staining. Dark bands of β-casein and pepsin which has only a single phosphoserine in elution lane indicates the effective capture and enrichment as seen in both total protein staining (**Figure S2.2a**) and phosphoprotein staining (**Figure S2.2b**). Additionally, the band of BSA is faint in the same elution lane, showing minimal non-specific binding. This result confirms the selectivity of the GAPT-Zn ligand complex on the surface of the $CoFe_2O_4$ NPs toward phosphate groups.



Supplemental Figure 2.2. SDS-PAGE analysis with (a) Sypro Ruby and (b) Phos-tag gel stain solutions confirmed the highly specific enrichment of phosphoproteins (β-casein and pepsin) from a standard protein mixture containing non-phosphoprotein (BSA). M: marker; LM: loading mixture; FT: flow through; E: elution.



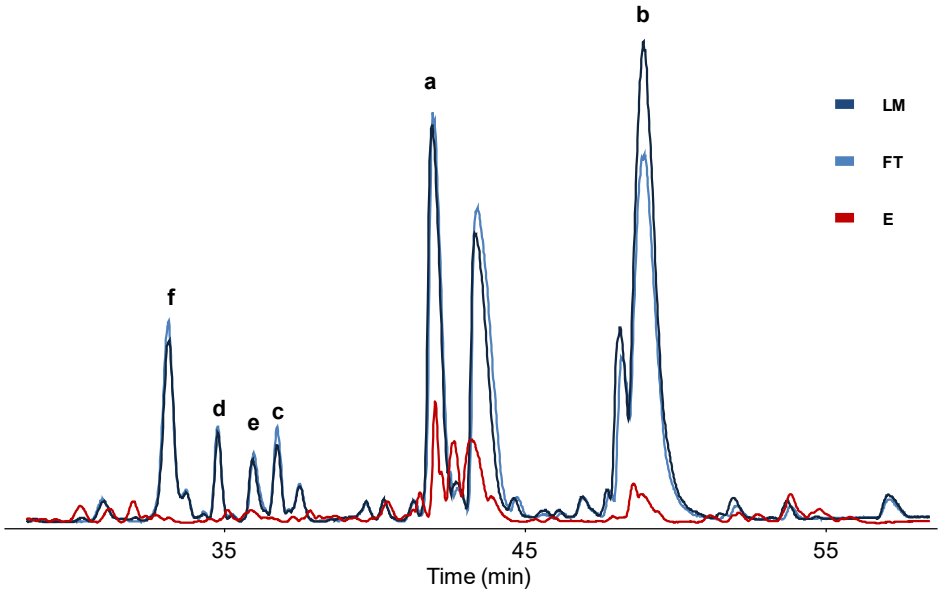
Supplemental Figure 2.3. SDS-PAGE analysis of 5% β-casein spike-in swine heart tissue extract. (a) Sypro Ruby staining and (b) ProQ Diamond staining corresponding to Figure 2.4. M: marker; LM: loading mixture; FT: flow through; E: elution.

Esimating the percentage (w/w) of the spiked-in β-casein in swine heart tissue extract. 10 mg/mL of β-casein (HPLC grade; average molecular weight 24 kDa) was prepared as a stock solution. Varying aliquots of β-casein stock solution were added into swine heart tissue extract, and then the total protein amounts was estimated by Bradford protein assay. The percentage (w/w) of the spiked-in β-casein was calculated as follows: % of spiked-in β-casein = [the amount of spiked-in β-casein /the total amount of the proteins in the mixture after adding the β-casein into swine heart tissue extract] × 100%.

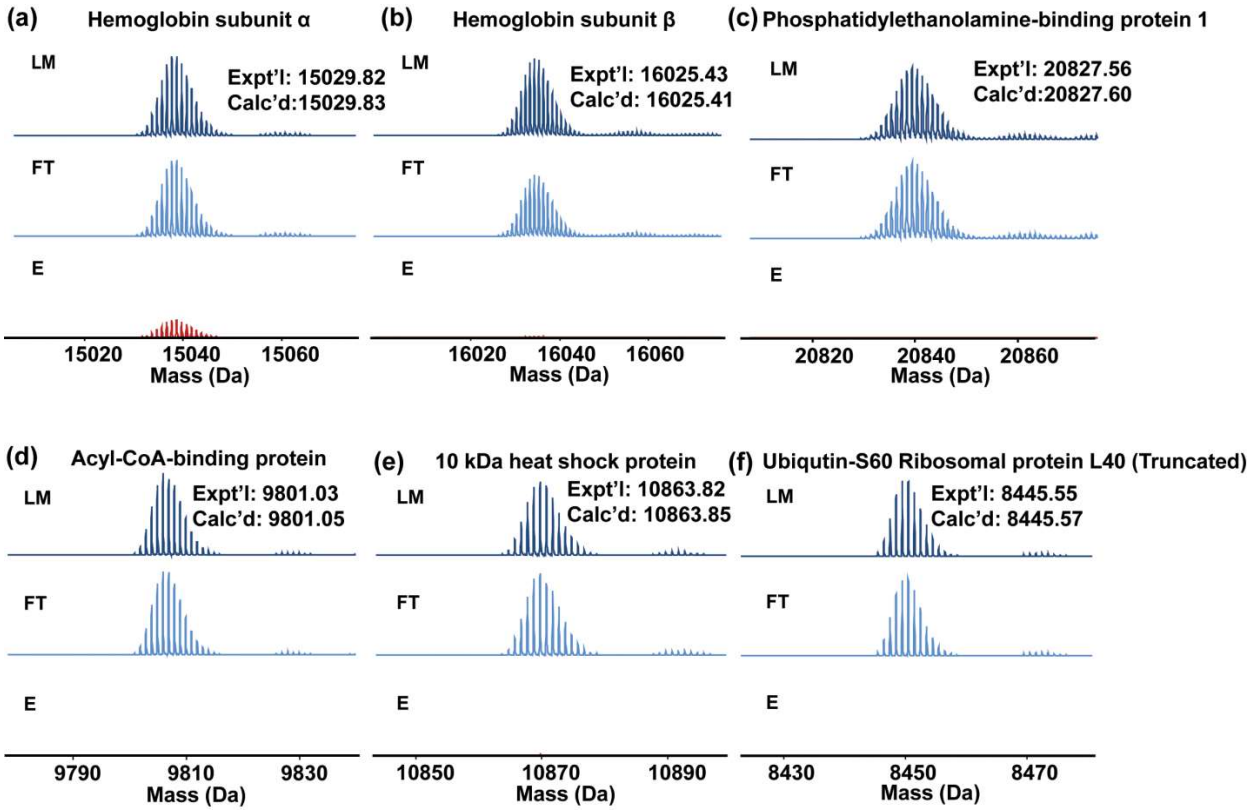
	I	II	III	IV
Concentration (µg/ µL) of spiked-in β-casein	0.03	0.05	0.07	0.1
Concentration (pmol/ µL) of spiked-in β-casein	1.3	2.1	2.9	4.2
Total protein quantity of swine heart tissue extract (LM) (µg)	885	910	764	996
% (w/w) of β-casein in LM	3	5	9	10

Supplemental Table 2.1. Final protein concentrations of the spike-in β-casein in 1 mL of the swine heart tissue extract used for phosphoprotein enrichment and the percentage (w/w) of the spiked-in β-casein in swine heart tissue extract. I, II, III, and IV corresponds to the lanes in Figure 2.3a, b.

Part IV. Supplementary data on top-down MS



Supplemental Figure 2.4. The overlapped base peak chromatograms (BPC) of LM, FT, and E. The loading mixture (LM; pre-enrichment solution), flow through (FT; washing solution), and elution (E; post-enrichment solution) of swine heart tissue extract containing 50 µg of spike-in β-casein were first separated by reverse phase chromatography. Equal amount of LM, FT, and E solutions were injected into the LC/MS. The BPC of LM (dark blue) and FT (light blue) was almost identical with each other but not with BPC of E (red). This demonstrates that the majority of proteins (presumably nonphosphoproteins) from LM were washed off into FT. The abundant a, b, c, d, e, and f peaks in LM and FT were identified and shown in Figure S2.5 and S2.6.



Supplemental Figure 2.5. Depletion of abundant non-phosphoproteins including blood proteins among others after enrichment. Normalized deconvoluated spectra from loading mixture (LM), flow through (FT), and elution (E) are shown. (a) Deconvoluted spectra of Hemoglobin subunit α from 41.8 to 42.3 min; (b) Deconvoluted spectra of Hemoglobin subunit β from 48.0-48.4 min; (c) Deconvoluted spectra of Phosphatidylethanolamine-binding protein 1 from 35.8-36.2 min; (d) Deconvoluted spectra of Acyl-CoA-binding protein from 34.7-35.0 min; (e) Deconvoluted spectra of 10 kDa heat shock protein from 36.7-37.0 min; (f) Deconvoluted spectra of Ubiquitin-S60 Ribosomal protein L40 (truncated) from 33.0-33.5 min. a, b, c, d, e, and f correspond to the labeled most abundant peaks in Figure S2.4. All proteins were identified by MSAlign+ search algorithm from LC/MS/MS with CID fragmentation (Figure S2.6).

(a) All proteins / sp|P01965|HBA_PIG Hemoglobin subunit alpha OS=Sus scrofa GN=HBA PE=1 SV=1 / Peptide #1

Protein-Spectrum-Match for Spectrum #47

PrSM ID:	32	Scan(s):	2382	Precursor charge:	17	Precursor m/z:	885.121
Precursor mass:	15029.935	Adjusted precursor mass:	15029.831	Duplicate score:	131	Unique score:	81
E-value:	1.2E-56	P-value:	5.8E-57	Spectral FDR:	n/a	Protein mass:	15029.831

No N-terminal modification(canonical)

V|L|S|A A D K A N|V K|A|A|W|G K|V|G G Q A|G A|H|G|A|E|A L E R M F L G|F|P T T K
T Y F|P H F N|L|S|H|G S D Q V K A H G Q K|V|A D|A L T K|A V G H|L|D|D|L|P|G|A|L|
S|A|L|S|D|L|H A H K L R V D|P V N F K L|L|S|H|C L|L|V|T L A A|H|H|P D D F N|P S|
V|H|A|S L D|K|F|L|A N|V|S T|V|L|T|S|K Y R

All proteins / tr|F1RII7|F1RII7_PIG Hemoglobin subunit beta OS=Sus scrofa GN=HBB PE=1 SV=1 / Peptide #5

(b) **Protein-Spectrum-Match for Spectrum #107**

PrSM ID:	75	Scan(s):	2800	Precursor charge:	22	Precursor m/z:	729.442
Precursor mass:	16025.563	Adjusted precursor mass:	16025.412	Duplicate score:	67	Unique score:	46
E-value:	2.2E-37	P-value:	2.2E-37	Spectral FDR:	n/a	Protein mass:	16156.452

N-Terminal Methionine Excision(canonical)

M]V H L S|A E E K E A|V|L|G|L|W|G K V N|V|D|E|V G|G E A L G R|L|L|V|V|Y|P W T Q
R F F E|S F G D L S N A|D|A|V|M|G N P K V K A H G K K V L Q S F S|D|G L|K H L D
N L K G T F A K L S E L H C D Q L H V D|P E N F R L L G N V I V V V L A R R L G
H D F N|P D|V|Q|A|A|F Q K|V|V|A|G|V|A|N|A|L|A|H K|Y H

All proteins / tr|F1RKG8|F1RKG8_PIG Uncharacterized protein OS=Sus scrofa GN=PEBP1 PE=1 SV=1 / Peptide #11

(c) **Protein-Spectrum-Match for Spectrum #16**

PrSM ID:	9	Scan(s):	1945	Precursor charge:	26	Precursor m/z:	802.076
Precursor mass:	20827.789	Adjusted precursor mass:	20827.597	Duplicate score:	45	Unique score:	33
E-value:	6.2E-25	P-value:	2.1E-25	Spectral FDR:	n/a	Protein mass:	20958.638

N-Terminal Methionine Excision(canonical)

M]P V D|L|G K W|S|G|P|L|S|L|Q|E|V|D E R P Q H|P L Q V K Y G G A E V D E L G K V
L T P T Q V K S R P T S I T|W D|G|L N|P D K|L|Y T|L|V L T D P D A P S R K D P K
Y R E W H H F L V V N M K G N D I S S G T V L S D Y V G S G P P K G T G L H R Y
V W L V Y E Q D G P L K C D E P I L S N R S G D H R G K F K V A S F R K K Y Q L
G A P V A|G T|C|Y|Q|A|E|W|D D|Y|V|P|K L Y E Q L S G K

(d)

All proteins / sp|P12026|ACBP_PIG Acyl-CoA-binding protein OS=Sus scrofa GN=DBI PE=1 SV=2 / Peptide #14

Protein-Spectrum-Match for Spectrum #10

PrSM ID:	6	Scan(s):	1857	Precursor charge:	13	Precursor m/z:	754.941
Precursor mass:	9801.142	Adjusted precursor mass:	9801.054	Duplicate score:	40	Unique score:	30
E-value:	2.8E-23	P-value:	1.4E-23	Spectral FDR:	n/a	Protein mass:	9890.084

N-Terminal Methionine Excision(canonical) and N-Terminal Acetylation

M]S[42.01] Q|A|E|F E K|A A E E V K N L K T K P A D D|E|M|L|F|I|Y|S H Y K Q A T|V|G D I|
N T E R P G I L D|L|K|G K A K|W D A W N G L K G T S K E D|A M|K|A Y I|N|K|V|E E|
L K K K Y G I

(e)

All proteins / tr|F1SMZ6|F1SMZ6_PIG Uncharacterized protein OS=Sus scrofa GN=HSPE1 PE=1 SV=1 / Peptide #22

Protein-Spectrum-Match for Spectrum #22

PrSM ID:	15	Scan(s):	2006	Precursor charge:	14	Precursor m/z:	777.003
Precursor mass:	10863.945	Adjusted precursor mass:	10863.847	Duplicate score:	43	Unique score:	29
E-value:	7.3E-20	P-value:	7.3E-20	Spectral FDR:	n/a	Protein mass:	10952.877

N-Terminal Methionine Excision(canonical) and N-Terminal Acetylation

M]A[42.01] G Q A F R K F L|P L F D R V L V E R S A A E T V T K G G I M|L|P E K S Q G K
V L Q|A T|V|V|A|V|G S G S K G K G G E I Q|P|V|S|V K V G D K V L|L|P E Y G G T R
V V L D|D K|D Y|F L F R D G D I L G K Y V D

(f)

All proteins / sp|P63053|RL40_PIG Ubiquitin-60S ribosomal protein L40 OS=Sus scrofa GN=UBA52 PE=1 SV=2 / Peptide #7

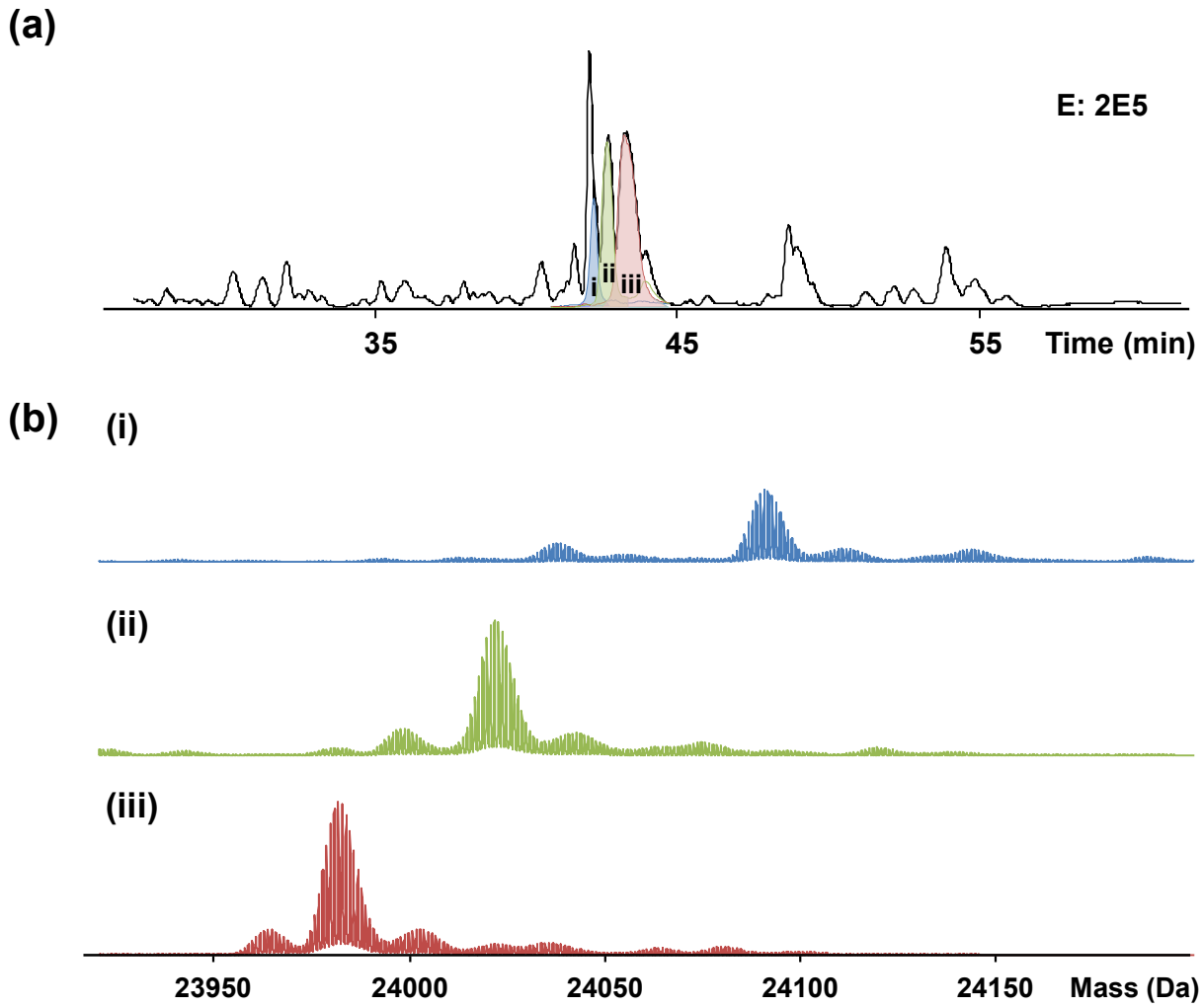
Protein-Spectrum-Match for Spectrum #3

PrSM ID:	2	Scan(s):	1763	Precursor charge:	10	Precursor m/z:	845.572
Precursor mass:	8445.646	Adjusted precursor mass:	8445.574	Duplicate score:	76	Unique score:	50
E-value:	4.9E-35	P-value:	3.5E-37	Spectral FDR:	n/a	Protein mass:	14718.958

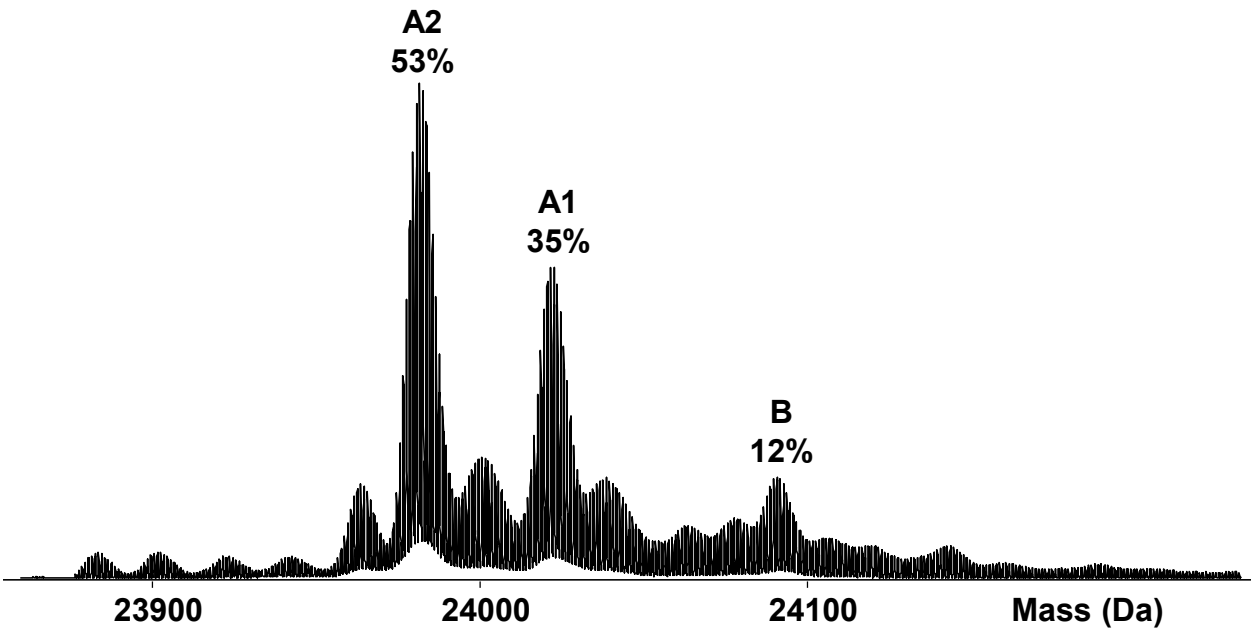
Removal of protein suffix
No N-terminal modification(canonical)

M|Q|I|F|V K T|L|T G K|T|I|T|L|E|V|E|P|S D|T|I E|N V|K|A K I Q D|K|E|G I|P|P D|Q
Q R L I F|A G|K|Q|L|E|D|G R T L S D|Y N|I|Q|K|E|S|T|L|H|L V L R L R[G G I I E P
S L R Q L A Q K Y N C D K M I C R K C Y A R L H P R A V N C R K K K C G H T N N
L R P K K K V K

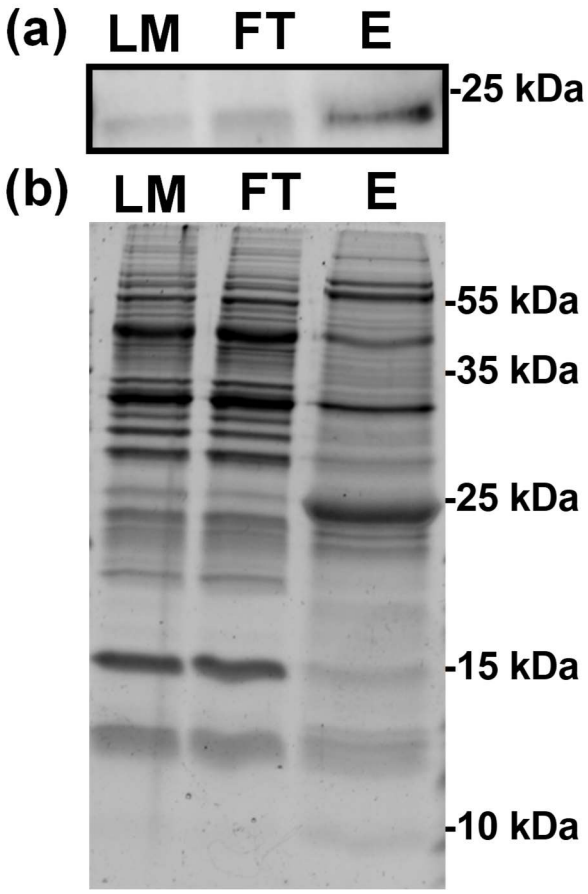
Supplemental Figure 2.6. Identification of the highly abundant non-phosphoproteins based on LC/MS/MS data with MS-Align+ search algorithm. (a) Hemoglobin subunit α (UnitProtKB/Swiss-Prot, P01965|HBA_PIG); (b) Hemoglobin subunit β (UnitProtKB/Swiss-Prot, F1RII7_PIG); (c) Phosphatidylethanolamine-binding protein 1 (UnitProtKB/Swiss-Prot, F1RKG8_PIG); (d) Acyl-CoA-binding protein (UnitProtKB/Swiss-Prot, P12026|ACBP_PIG); (e) 10 kDa heat shock protein (UnitProtKB/Swiss-Prot, F1SMZ6_PIG); (f) Ubiquitin-S60 Ribosomal protein L40 (truncated) (UnitProtKB/Swiss-Prot, P63053|RL40_PIG). The identifications correspond to figure S2.5.



Supplemental Figure 2.7. Top-down LC/MS analysis of multiple β-casein variants in elution after enrichment with swine heart tissue extract with 0.05 μg/μL of spike-in β-casein. **(a)** Base peak chromatogram of the elution fraction. (i), variant A1 (ii), and variant A2 (iii). **(b)** Normalized deconvoluted mass spectra of β-casein isoforms; Variant B (i) showed a mass difference of 109 Da from A2 variant, which corresponds with a previously reported mass difference by two amino acids at position 67 and 122, having His 67 and Ser 122 instead of Pro 67 and Arg 122 in the variant A2 . Variant A1 (ii) revealed a mass difference of 40 Da from A2 variant (iii), which corresponds with a single amino acid difference from Pro 67 (in the variant A2) to His 67.[226]



Supplemental Figure 2.8. Deconvoluted spectrum of β-casein showing multiple proteoforms.The spectrum was averaged from 41.9 – 44.5 min in the top-down LC/MS run of elution sample indicating the relative abundance among variant A2, A1, and B is 53%, 35%, and 12% respectively.



Supplemental Figure 2.9. Western blots analysis of phosphorylated cTnI demonstrating the enrichment of phosphoproteins in the elution. (a) Representative Western blot for cTnI phosphorylated at Ser22/23 with equal amount of proteins loaded in loading mixture (LM), flow through (FT), and elution (E). (b) Total proteins stained by Sypro Ruby with equal loading in LM, FT, and E.

(a) All proteins / tr|F1RHJ2|F1RHJ2_PIG Uncharacterized protein OS=Sus scrofa GN=HDGF PE=1 SV=1 / Peptide #2

Protein-Spectrum-Match for Spectrum #1

PrSM ID:	0	Scan(s):	1	Precursor charge:	29	Precursor m/z:	940.249
Precursor mass:	27238.02	Adjusted precursor mass:	27237.855	Duplicate score:	27	Unique score:	21
E-value:	4.6E-13	P-value:	8.8E-14	Spectral FDR:	n/a	Protein mass:	27088.011

N-Terminal Methionine Excision(canonical) and N-Terminal Acetylation
A modification (or multiple modifications) on internal residue

M]S(42.01) R S N R]Q]K]E Y K C G D L V F A K M K G Y P H W P A R I D E M P E A A V K S
T A N K Y Q V F F F G T H E T A F L G P K D L F P Y E E S K E K F G K P N K R K
G F S E G L W E I E N N P T V K A S G Y Q S S Q K K S C V E E P E P K A D]A A E
G]D G]D(238.87) K K G N A E G S S D E E G K L V I D E P T K E K N E K G A L K R R A G D
L L E D S P K R P K]E A E]D P]E G E E K]E V A T]L E G E R P L P V E]A E]K]N S T
P S]E P S S G R G P P P E E E E E E E E E E E E E E A T K E]D A]E]A P G I R D H
E S L

(b) All proteins / tr|F1RHJ2|F1RHJ2_PIG Uncharacterized protein OS=Sus scrofa GN=HDGF PE=1 SV=1 / Peptide #1

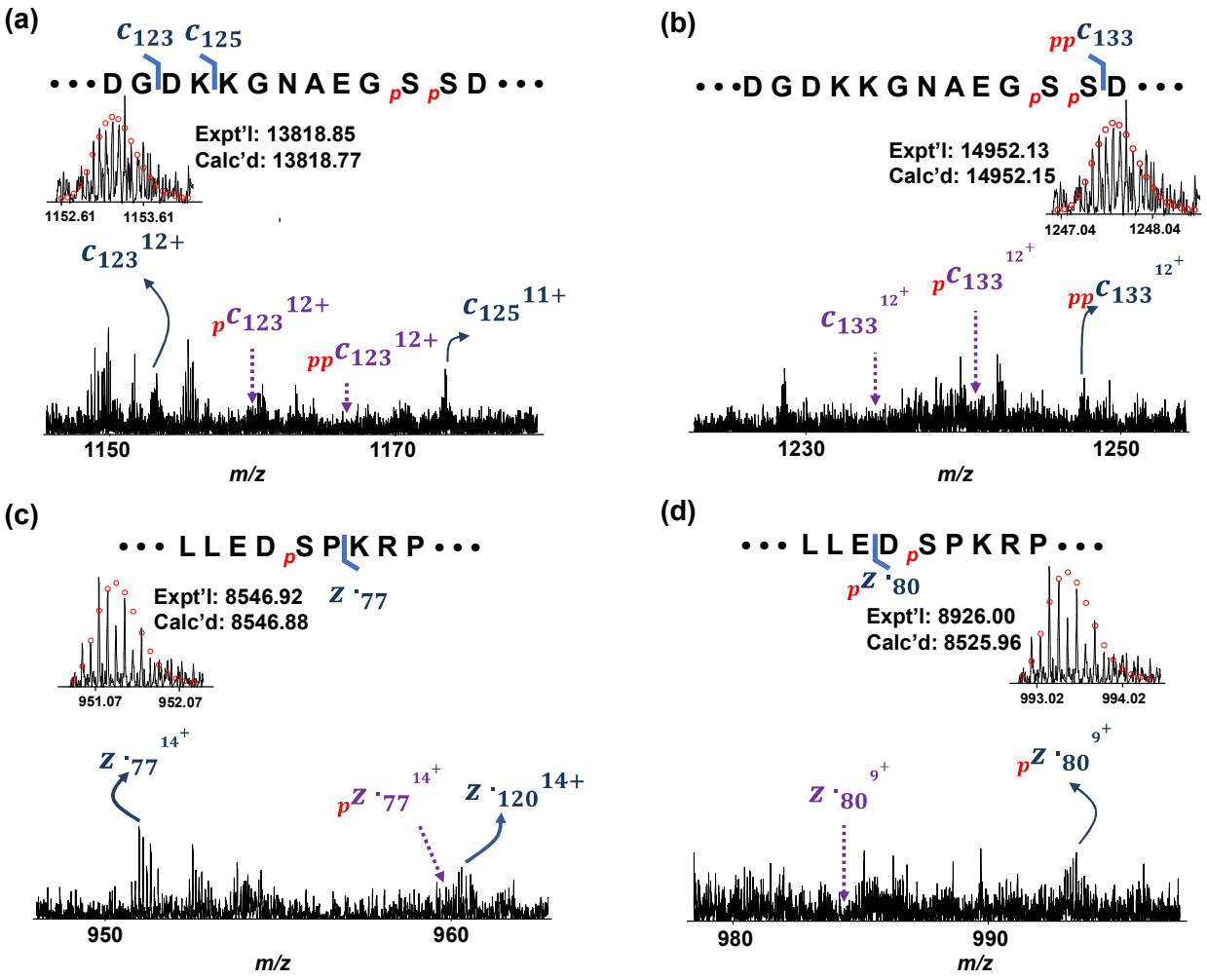
Protein-Spectrum-Match for Spectrum #1

PrSM ID:	0	Scan(s):	1	Precursor charge:	29	Precursor m/z:	940.249
Precursor mass:	27238.02	Adjusted precursor mass:	27237.915	Duplicate score:	68	Unique score:	45
E-value:	1.2E-21	P-value:	2.3E-22	Spectral FDR:	n/a	Protein mass:	27088.011

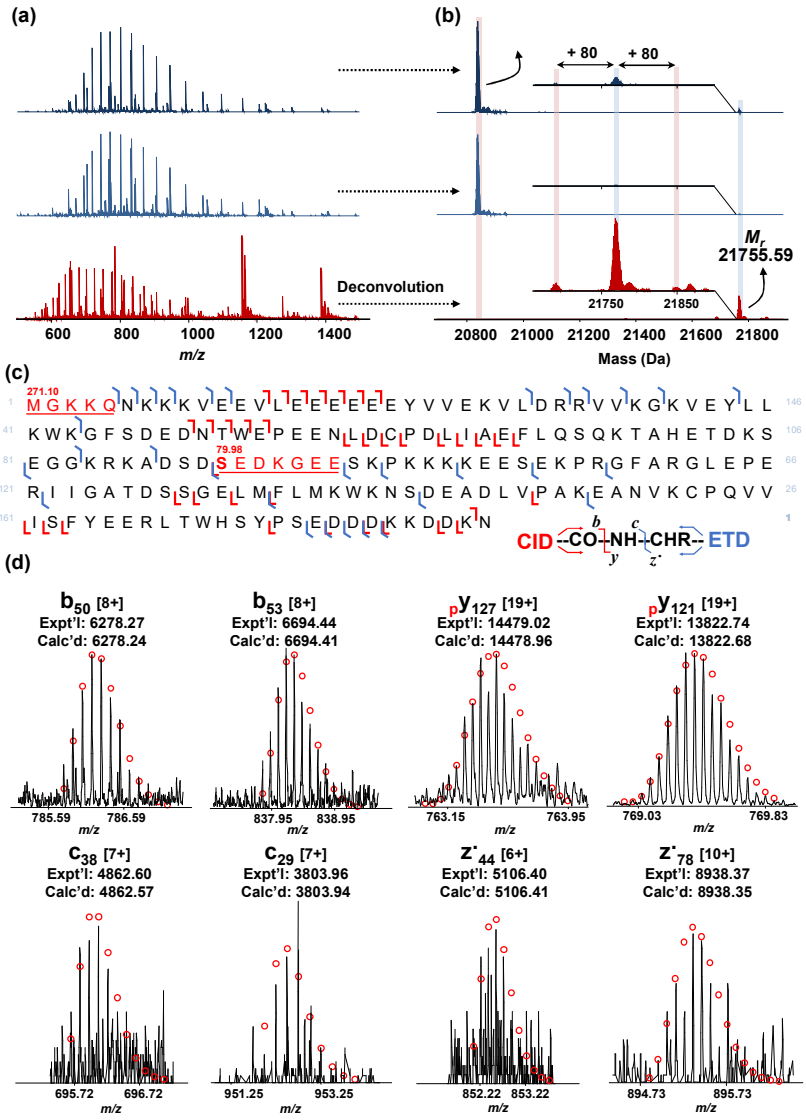
N-Terminal Methionine Excision(canonical) and N-Terminal Acetylation
A modification (or multiple modifications) on internal residue

M]S(42.01) R S N R Q K E]Y K C G D L V]F(238.93) A K M K G Y P H W P]A R I D E M P E A A V K S
T A N K Y Q V F F F G T H E T A F L G P K D L F P Y E E S K E K F G K P N K R K
G F S E G L W E I E N N P T V K A S G Y Q S S Q K K S C V E E P E P K A D A A E
G D G D K K G N A E G S S D E E G K L V I D E P T K E K N E K G A L K R R A G D
L L E D S P K R P K E A E D]P E]G E E K E]V A]T]L]E]G E R P L]P V E A E K N S T]
P S E]P S S G R G P P P E]E]E]E]E]E]E]E]E]E]E]E]A]T]K]E]D]A]E]A]P G I]R D]H
E S L

Supplemental Figure 2.10. Identification of the 27 kDa multiply phosphorylated protein by LC/MS/MS. (a) ETD and (b) CID with MS-Align+ algorithm were performed[3]. Uncharacterized protein (UnitProtKB/Swiss-Prot, F1RHJ2_PIG) was best matched and identified with N-terminal Met loss and two possible modifications (acetylation and multiple phosphorylations). The Basic Local Alignment Search Tool (BLAST) found 93.8% sequence similarity between the identified protein and Hepatoma-derived growth factor (Homo sapiens) (P51858|HDGF_HUMAN), suggesting its identity.



Supplemental Figure 2.11. MS/MS mapping of the phosphorylation sites of the identified swine hepatoma-derived growth factor after NP enrichment.(a-d) Representative MS/MS spectra of ***c*** and ***z•*** type ions from ECD spectra of triply-phosphorylated hepatoma-derived growth factor using online LC/MS/MS. The phosphorylation sites were localized to Ser132, Ser133, and Ser165. The assignment of fragment ions was made based on the sequence of uncharacterized protein (UnitProtKB/Swiss-Prot, F1RHJ2_PIG) with N-terminal Met excision and acetylation near the N-terminal. Blue lines indicate existing ions, and purple dash lines indicate non-existing ions.



Supplemental Figure 2.12. Representative MS/MS analysis of endogenous phosphoproteins (22 kDa) enriched by functionalized $CoFe_2O_4$ NPs from swine heart tissue extract. (a) Average spectra and (b) deconvoluted spectra of loading mixture (LM, dark blue), flow through (FT, light blue), and elution (E, red) from 35.7 min to 35.9 min. (c) Fragment ion map of ETD and CID product ions from triply phosphorylated precursor ion at 31+ charge state of chromobox protein homolog 1 through online MS/MS analysis with CID and ETD. Red "S" (79.98) indicates potential phosphorylation sites. Red number above the underlined red sequence reveals modifications with their additional mass.

(a) All proteins / tr|F1RWH1|F1RWH1_PIG Uncharacterized protein OS=Sus scrofa GN=CBX1 PE=1 SV=2 / Peptide #1

Protein-Spectrum-Match for Spectrum #2

PrSM ID:	1	Scan(s):	1	Precursor charge:	31	Precursor m/z:	702.802
Precursor mass:	21755.65	Adjusted precursor mass:	21755.65	Duplicate score:	26	Unique score:	22
E-value:	2.8E-15	P-value:	4.6E-18	Spectral FDR:	n/a	Protein mass:	21404.572

Unusual N-terminal modification
A modification (or multiple modifications) on internal residue

271.10
M G K K Q N K K]K V]E]E V L E E E E E E Y V V E K V L]D R]R V V]K G]K V E Y]L L
79.98
K W K]G F S D E D N T W E P E E N L D C P D L I A E F L Q S Q K T A H E T D K S
E G G K R K A D S D S E D K G E E S K P K K K K E E S E K P R]G F A R G L E P E
R]I I G A T D S]S G E L M]F L M K W K N S]D E A D L]V P A K]E A N V K C P Q V V
I S F Y E E R L T W H S Y P S]E]D]D]D K K D D K N

(b) All proteins / tr|F1RWH1|F1RWH1_PIG Uncharacterized protein OS=Sus scrofa GN=CBX1 PE=1 SV=2 / Peptide #2

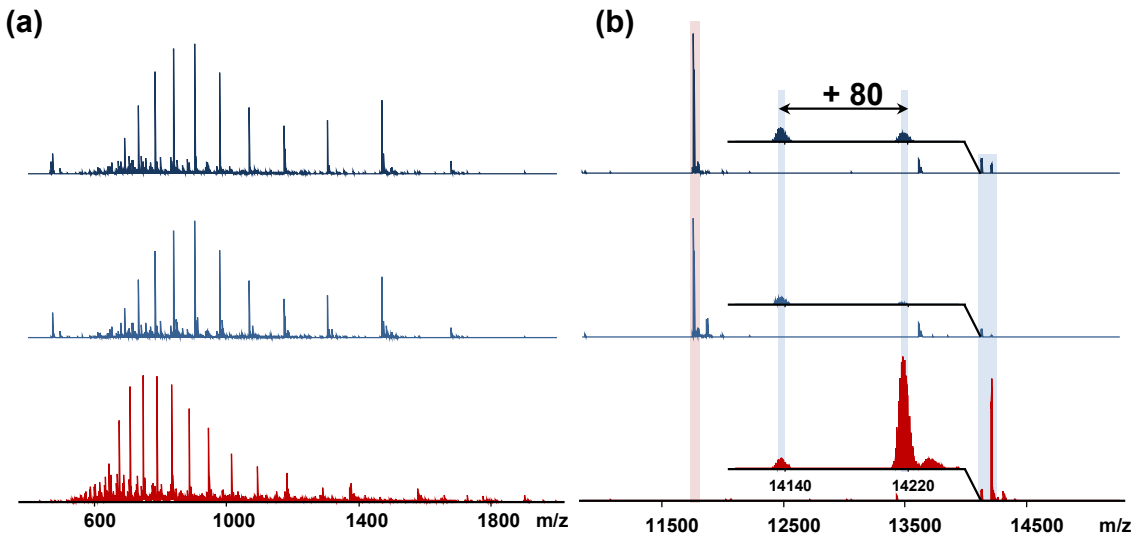
Protein-Spectrum-Match for Spectrum #2

PrSM ID:	1	Scan(s):	1	Precursor charge:	31	Precursor m/z:	702.802
Precursor mass:	21755.64	Adjusted precursor mass:	21755.64	Duplicate score:	88	Unique score:	37
E-value:	1.4E-12	P-value:	1.4E-15	Spectral FDR:	n/a	Protein mass:	21404.572

Unusual N-terminal modification
A modification (or multiple modifications) on internal residue

271.10
M G K K Q N K K K V E E V L E]E]E]E]E]E]Y V V E K V L D R R V V K G K V E Y L L
79.97
K W K G F S D E D]N]T]W]E]P E E]N]L]D]C]P D]L]I]A]E]F L Q S Q K T A H E T D K S
E G G K R K A D S D S E D K G E E S K P K K K K E E S E K P R G F A R G L E P E
R I I G A T D S]S]G E L]M]F L M K W K N S D E A D L V]P A K E A N V K C P Q V V]
I]S F Y E E R L T W H S Y]P S E]D]D]D]K K D]D]K N

Supplemental Figure 2.13. Identification of the 22 kDa phosphorylated protein by (a) ETD LC/MS/MS and (b) CID LC/MS/MS with MS-Align+ algorithm.Uncharacterized protein (UnitProtKB/Swiss-Prot, F1RWH1_PIG) was best matched and identified with N-terminal sequence variation and a possible phosphorylation. The Basic Local Alignment Search Tool (BLAST) found 100% sequence similarity between the identified protein and Chromobox protein homolog 1 (Homo sapiens) (P83916|CBX1_HUMAN), indicating its identity.



Supplemental Figure 2.14. Representative MS/(MS) analysis of endogenous phosphoproteins (14 kDa) enriched by functionalized $CoFe_2O_4$ NPs from swine heart tissue extract**.**(a) Mass spectra and (b) deconvoluted spectra of loading mixture (LM, dark blue), flow through (FT, light blue), and elution (E, red) from 37.8 min to 38.1 min.

All proteins / tr|F1RNX2|F1RNX2_PIG Programmed cell death protein 5 OS=Sus scrofa GN=PDCD5 PE=1 SV=1 / Peptide #2

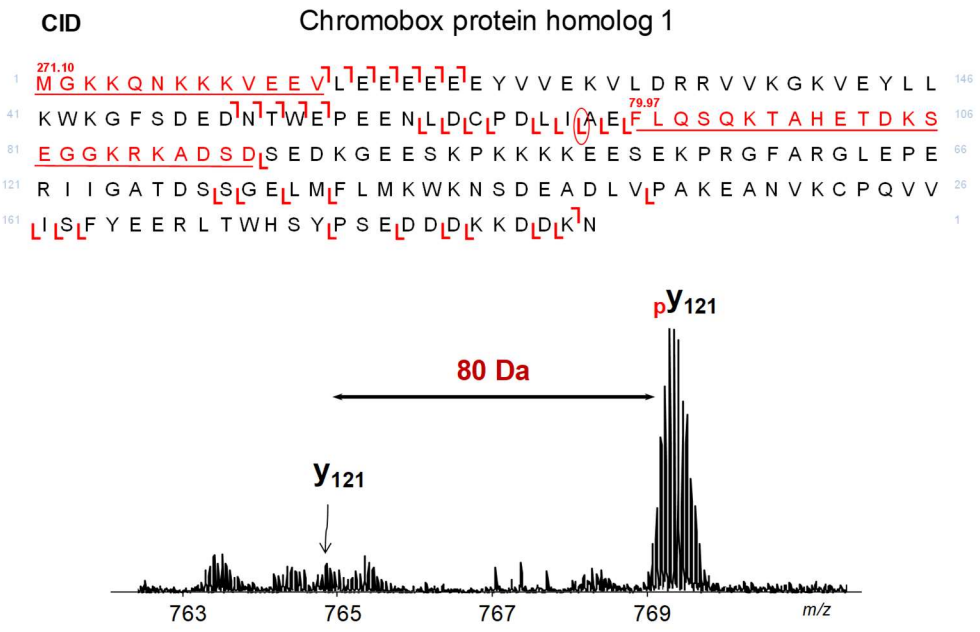
Protein-Spectrum-Match for Spectrum #1

PrSM ID:	0	Scan(s):	1	Precursor charge:	19	Precursor m/z:	748.861
Precursor mass:	14209.22	Adjusted precursor mass:	14209.22	Duplicate score:	35	Unique score:	21
E-value:	2.4E-6	P-value:	1.7E-6	Spectral FDR:	n/a	Protein mass:	14218.324

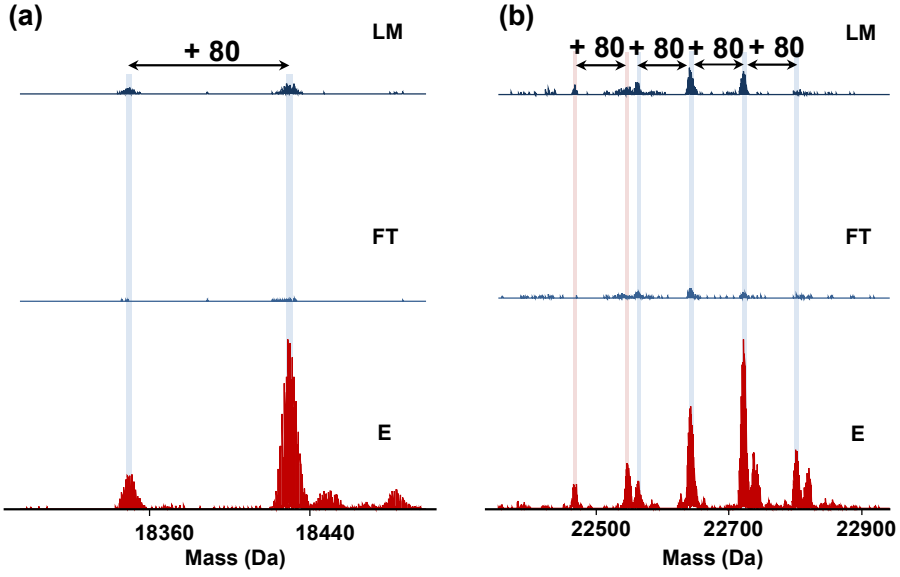
N-Terminal Methionine Excision(canonical) and N-Terminal Acetylation
A modification (or multiple modifications) on internal residue

M]A(42.01) E]E L E A L R K Q R L A E L Q A K H G D]P G D A]A]Q Q E A K H R E A E M R
N S I L A Q]V]L D Q S A R A R L S N L A L V K P E K T K A V E N Y]L]I]Q]M]A R Y
G Q L S G K V S E Q G L I E I L E K V S Q Q T E K K T T V K F N R R K V(79.93) M D S D
E D D]D]Y

Supplemental Figure 2.15. Identification of the 14 kDa phosphorylated protein by CID LC/MS/MS with MS-Align+ algorithm.Programmed cell death protein 5 (UnitProtKB/Swiss-Prot, F1RNX2_PIG) was best matched and identified with N-terminal Met excision and a possible phosphorylation.



Supplemental Figure 2.16. Fragmentation map and representing ions of the 22 kDa phosphoryalted protein chromobox protein homolog 1. The highly abundant phosphorylated y121 ion and its minimal unphosphorylated counterpart indicate preservation of phosphorylation during collisionally induced dissociation (CID).



Supplemental Figure 2.17. Additional examples of endogenously enriched phosphoproteins. Normalized deconvoluted average MS spectra of loading mixture (LM), flow through (FT), and elution (E) from (a) 37.2-37.5 min and (b) 32.7-33.0 min.

2.7 Acknowledgements

This research is supported by R01GM117058 (to SJ and YG). YG would also like to acknowledge NIH R01HL096971, R01HL109810 and S10OD018475. We thank Matt Willetts at Bruker for assistance with Data Analysis scripts. We also thank Albert Chen for preparing for heart tissue lysate.

Chapter 3: The Impact of Phosphorylation on Electron Capture Dissociation of Intact Proteins

This chapter has been published and is adapted from:

Chen, B.; Guo, X.; Tucholski, T.; Lin, Z.; Mcilwain, S.; Ge, Y. The Impact of Phosphorylation on Electron Capture Dissociation Efficiency of Proteins: A Top-down Perspective. *Journal of American Society of Mass Spectrometry.* 2017, 28(9): 1805-1814.

3.1 Abstract

Electron capture dissociation (ECD) is well suited for the characterization of phosphoproteins, with which labile phosphate groups are generally preserved during the fragmentation process. However, the impact of phosphorylation on ECD fragmentation of intact proteins remains unclear. Here, we have performed a systematic investigation of the phosphorylation effect on ECD of intact proteins by comparing the ECD cleavages of mono-phosphorylated α-casein, multi-phosphorylated β-casein, and immunoaffinity-purified phosphorylated cardiac troponin I with those of their unphosphorylated counterparts, respectively. In contrast to phosphopeptides, phosphorylation has significantly reduced deleterious effects on the fragmentation of intact proteins during ECD. On a global scale, the fragmentation patterns are highly comparable between unphosphorylated and phosphorylated precursors under the same ECD conditions, despite a slight decrease in the number of fragment ions observed for the phosphorylated forms. On a local scale, single phosphorylation of intact proteins imposes minimal effects on fragmentation near the phosphorylation sites, but multiple phosphorylations in close proximity result in a significant reduction of ECD bond cleavages.

3.2 Introduction

Since the introduction of electron capture dissociation (ECD) in 1998,[235] this powerful tandem mass spectrometry (MS/MS) technique has been widely used to elucidate amino acid sequences, structures, and, particularly, to localize post-translational modifications (PTMs) of proteins.[4,8,236-238] During ECD, ion analytes (proteins or peptides) are irradiated by low-energy electrons, cleaving N-C_α bonds and producing ***c***-type and ***z***-type ions.[235,236,239,240] Compared to traditional "slow heating" MS/MS methods such as collisionally activated dissociation (CAD) [241,242] and infrared multi-photon dissociation (IRMPD),[243] which usually cleave the weakest bonds in a given protein/peptide ion and lead to the preferred loss of labile PTMs, ECD causes fast fragmentation prior to energy randomization over the whole protein or peptide.[235,244-246] This characteristic of ECD generally results in preservation of labile PTMs,[95,236] such as phosphorylation[247] and OGlcNAc[248].

Phosphorylation, a highly prevalent PTM, plays a critical role in regulating various complex cellular processes, and altered phosphorylation is implicated in many diseases.[4,8,249] During CAD, the phosphorylated ions often experience 98 and 80 Da neutral losses, hindering the interrogation of phosphorylation sites.[250] In contrast, ECD preserves phosphorylation while providing rich fragmentation patterns, presenting a valuable option to characterize both phosphopeptides and phosphoproteins.[247] For phosphopeptides, it has been demonstrated that ECD can be applied in data dependent online LC-MS/MS experiments for large scale identification and localization of phosphorylation sites.[251] However, because of the elongated ion accumulation time, utilization of ECD in online LC-MS/MS is not widely utilized.[252] On the other hand, ECD is particularly powerful in characterizing phosphoproteins in an offline mode. Over the past years, our group has successfully applied Fourier transform ion cyclotron resonance mass spectrometry

(FT-ICR MS) equipped with ECD to comprehensively characterize biologically important phosphoproteins in the heart such as myosin binding protein C,[16] cardiac troponin I,[53,55,253] tropomyosin,[99,254] and AMPactivated protein kinase catalytic domain.[100] Furthermore, we were able to discover reduced phosphorylation of a previously unknown phosphoprotein, enigma homolog isoform 2, from swine heart tissue after acute myocardial infarction and localize the phosphorylation site with ECD,[97] which demonstrated the power of ECD for characterizing phosphoproteins.

In addition to localizing phosphorylation sites, ECD has also proven feasibility for quantitative analysis. The Kelleher group has pioneered and demonstrated the use of ECD spectra to quantify isomeric methylated and acetylated proteoforms from histone H4.[157] Our group has also developed a novel strategy to quantify isomeric phosphorylated proteins that utilized only the abundance of unphosphorylated fragment ions with intraspectrum normalization, which avoids confronting potential influence from the phosphate groups on the ECD.[255] However, the influence of phosphorylation on intact protein ECD fragmentation is still largely unknown. Although Creese and Cooper showed that the addition of phosphorylation has deleterious effect on ECD sequence coverage for doubly charged peptides,[256] and later demonstrated that the structure rather than the sequence of phosphopeptides influences ECD efficiency,[257] these results could not be generalized to proteins. Compared with peptides, the physiochemical properties for larger intact proteins tend to be much less affected by the addition of phosphorylation, thus prompting further investigation.

Here, we have systematically investigated the potential effect of phosphorylation of intact proteins on ECD behavior using mono-phosphorylated α-casein, multi-phosphorylated β- casein, and immunoaffinity-purified cardiac troponin I from swine and rhesus monkey tissues. We have localized the phosphorylation sites and statistically compared the ECD fragments of the

unphosphorylated and the phosphorylated forms of these proteins. The results collectively show that the phosphorylation has considerably less deleterious influence on ECD fragmentation of intact protein compared to peptides. Globally, the total number of fragment ions produced by ECD was slightly decreased for the phosphorylated precursor ions, but the fragmentation pattern remained highly comparable between the unphosphorylated and the phosphorylated proteins. Locally, near the amino acid residues adjacent to the single phosphorylation site, fragmentation patterns remained largely unchanged; however, multiple phosphorylations in close proximity led to a decrease in protein backbone cleavages. These findings provide a contemporary top-down perspective on ECD fragmentation of phosphoproteins.

3.3 Material and Methods

3.3.1 Chemicals and reagents

Dephosphorylated α-casein (bovine milk) was purchased from Sigma-Aldrich (St. Louis, MO, USA). β-Casein (bovine pancreas) standard was purchased from Protea Biosciences (Morgantown, WV, USA). Lambda protein phosphatase (λPP) was purchased from New England Biolabs (Ipswich, MA, USA). Pierce Protein Concentrator PES, 10K MWCO, high-performance liquid chromatography (HPLC)-grade water (H2O), and acetonitrile (ACN) were purchased from Fisher Scientific (Waltham, MA, USA). All other reagents were purchased from Sigma-Aldrich unless otherwise specified.

3.3.2 *In vitro* dephosphorylation of β-casein

For every 20 μL of dephosphorylation reaction, 10 μg of β-casein was incubated with 100 units of lambda protein phosphatase (λPP) with the supplement of 1 mM $MnCl_2$ in 1× protein metallophosphatase buffer. The mixture was incubated at 30 °C for 0.5 h. Subsequently, dephosphorylated β-casein was desalted using 0.1% formic acid in H2O wash through 10 K MWCO protein concentrators prior to top-down MS analysis. The final concentrations of the phosphorylated and the unphosphorylated β-casein were about 0.25 μg/μL in ACN: H2O (50:50) and 0.1% formic acid.

3.3.3 Immunoaffinity purification of swine and rhesus monkey cTnI

Heart tissue was obtained from juvenile Yorkshire domestic pig and rhesus monkey hearts as approved by the University of Wisconsin Animal Care and Use Committee. The excised heart tissue was immediately frozen in liquid nitrogen and stored in a – 80 °C freezer. The whole cardiac troponin complex was purified by the immunoaffinity purification method as previously described.[54,258] Briefly, 1.5 g of heart tissue was homogenized in tissue wash buffer (NaH_2PO_4 500 mM, Na_2HPO_4 100 mM, $MgCl_2$ 100 mM, EGTA 100 mM, NaCl 0.1 M, Triton X-100 1%, DTT 5 mM, protease and phosphatase inhibitor cocktail tablet, PMSF 1 mM, leupeptin 2 μg/mL, pH 7.4) using a Polytron electric homogenizer (Model PRO200, PRO Scientific Inc., Oxford, CT, USA) for 30 s on ice. The homogenate was centrifuged at 8000 rpm at 4 °C for 10 min; the pellet was collected and resuspended in 6 mL protein extraction buffer (0.7 M LiCl, 25 mM Tris, 5 mM EGTA, 0.1 mM $CaCl_2$, 5 mM DTT, 1 mM PMSF, 2 μg/mL leupeptin, and 0.75 mg/mL protease and phosphatase inhibitor cocktail, pH 8.0) and protein extraction was performed with agitation on a

nutating mixer (Fisher Scientific Inc., Pittsburgh, PA, USA) at 4 °C for 45 min. The sample was then centrifuged at 16,000 rpm (Centrifuge 5415R; Eppendorf, Hamburg, Germany) for 5 min. The supernatant was further centrifuged at 55,000 rpm (Beckman L-55 ultracentrifuge; Beckman Coulter, Fullerton, CA, USA) for 45 min to completely remove any particulate. The supernatant was incubated with 0.25 mL of CNBr-activated Sepharose CL-4B conjugated with 1.25 mg monoclonal cTnI antibody (anti-troponin I monoclonal antibody MF4 and 14G5; HyTest Ltd., Turku, Finland) for 35 min at 4 °C to ensure complete binding of troponin complex to the antibody. After washing the column with 2 mL of extraction buffer, the bound troponin complex was eluted with 100 mM glycine at pH 2 into four 0.4 mL fractions and neutralized immediately with 40 µL of 1M MOPS (pH 9). Fractions were further separated and desalted using an offline reverse phase protein microtrap (Michrom Bioresources Inc., Auburn, CA, USA) to obtain purified cTnI.

3.3.4 Top-down MS analysis

A 7T linear ion trap/Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (MS) (LTQ/FT Ultra, Thermo Scientific, Bremen, Germany) and a 12T SolariX FTICR-MS (Bruker, Bremen, Germany) equipped with ECD were used for the investigation. Triversa NanoMate (Advion, Ithaca, NY, USA), a chip-based nano-electrospray ionization (nanoESI) system delivered all samples into the mass spectrometers. For the LTQ/FT Ultra, the resolving power was set to 200,000 at 400 m/z. During ECD, individual precursor ions with the same charge state of the unphosphorylated and the phosphorylated proteins were isolated and fragmented using 2.2%–3.3% electron energy with a 50 ms duration and no delay; 1000 to 9000 transients were averaged to ensure high quality of ECD spectra. For the SolariX FTICR-MS, the mass spectra were collected over a 202.76 to 3000 m/z range with 2 mega-words (resolution

270,000 at 400 m/z) transient length (1.1534 s) and a pulse at 20% excitation power. All MS/ MS spectra were accumulated for 1000 scans unless specified otherwise. For the "standard" parameters, ECD pulse length was set to 0.05 s; ECD bias was set to 0.8 V (0.8 eV); and ECD lens was 10 V for ECD fragmentation.

3.3.5 Data Analysis

MASH Suite[139] and MASH Suite Pro[129] were used to analyze MS/MS data generated from 7T LTQ/FT Ultra and 12 T SolariX FT-ICR, respectively. All peaks were extracted with a signal-to-noise ratio set to 3 and a minimum fit of 60%, followed by manual validation. Fragment ions were assigned by matching the experimental fragment ions to the calculated ones from theoretical amino acid sequence with a 10 ppm mass tolerance. The abundance of each ion is the sum of the intensities from the top five most abundant isotopomers. Subsequently, the quantitation value was normalized to the charge of the ion. The normalized values from the same ions with different charge states were then summed. The relative abundance of each fragment ion was calculated by normalizing to the most abundant fragment ion within the same spectrum.

3.4 Results and Discussion

To evaluate the phosphorylation effect on ECD fragmentation of intact protein, it is critical to maintain phosphorylation as the only changing factor and minimize other variations arising from instrument parameters and sample introduction. ECD parameters were kept constant between ECD of unphosphorylated and phosphorylated counterparts of the same protein. Because the precursor ion signal in the FTMS instrument may introduce biased fragment ion intensity, the ion

accumulation was adjusted between the isolated unphosphorylated and phosphorylated precursor so that their intensities are comparable in a single scan. Furthermore, fragment ion intensity was normalized to the most abundant fragment ion in each spectrum to minimize the variation caused by the difference in precursor ion intensity. Here, three different protein species, including single and multiple phosphorylation sites with multiple replicates, were used to investigate the influence of phosphorylation on ECD of intact protein.

3.4.1 ECD of mono-phosphorylated vs. unphosphorylated α-casein

We first acquired MS spectra on the LTQ/FT of the commercially available "dephosphorylated" α-casein (M_r 22960.48 Da) after desalting by offline reverse-phase chromatography. The MS spectrum reveals the detection of unphosphorylated α- casein with the presence of mono-phosphorylated (major) and the bis-phosphorylated (minor) forms (**Figure 3.1a**). The monophosphorylated precursor was subsequently isolated for ECD to localize the phosphorylation site. After validation of all fragment ions, a match to the theoretical α-casein sequence (UniProtKB-P02662) provides two possible phosphorylation sites, Ser88 and Ser115 (**Figure 3.1b**). One hundred and 108 out of 198 backbone bonds were cleaved when the phosphorylation was assigned to Ser88 and Ser115, respectively. Careful manual validation of fragment ions resulting from ECD (**Figure 3.1c** and **Figure S3.1**) further confirmed the heterogeneity of the mono-phosphorylated α-casein. Specifically, representative ions $\boldsymbol{c}_{88}$, $\boldsymbol{c}_{92}$, $\boldsymbol{c}_{93}$, and $\boldsymbol{c}_{105}$ all coexisted with the monophosphorylated counter parts (**Figure 3.1c**). Prior to residue Asp85, no phosphorylated ions were observed as demonstrated by $\boldsymbol{c}_{54}$, $\boldsymbol{c}_{74}$, $\boldsymbol{c}_{83}$, and $\boldsymbol{c}_{85}$ (**Figure S3.1**). Furthermore, the presence of only the mono-phosphorylated forms of $\boldsymbol{c}_{118}$, $\boldsymbol{c}_{132}$, and $\boldsymbol{c}_{144}$ ions without the detection of the unphosphorylated counterparts proved the absence of phosphorylation

sites after Glu118. Therefore, with only two possible serine residues from Asp85 to Glu118, we confidently assigned the mono-phosphorylation site to Ser88 and Ser115, with the latter being the predominant form.

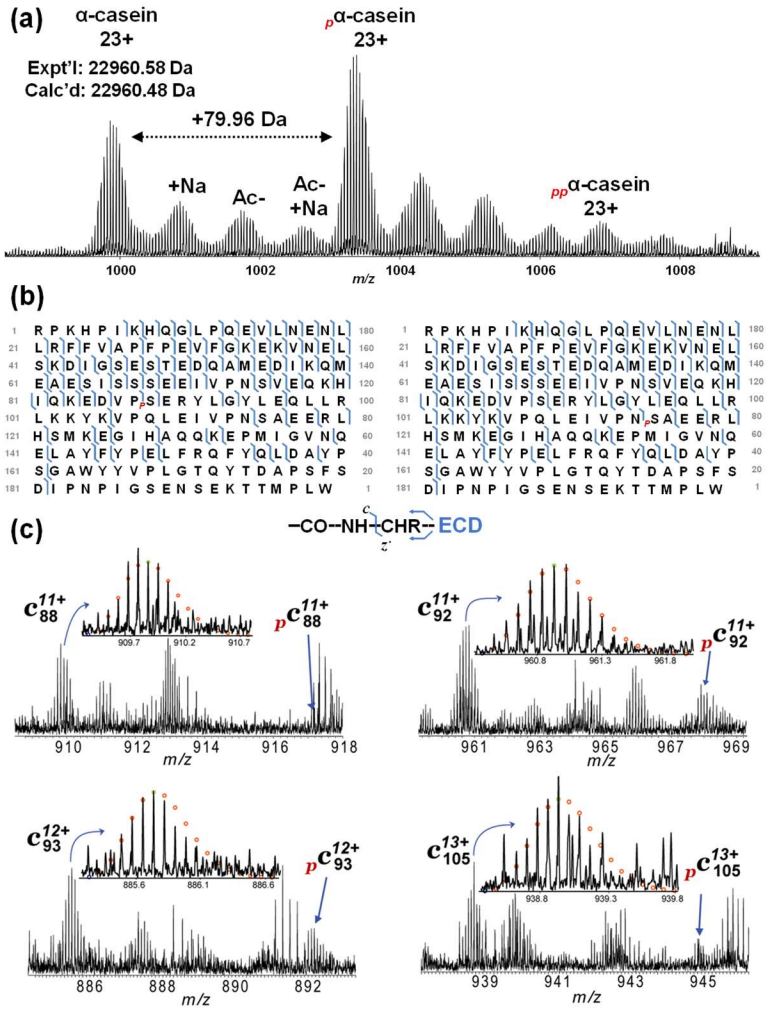


Figure 3.1. FTMS characterization of α-casein. (a) FTMS spectrum of α-casein. +22 Da, +42 Da, and +64 Da, represent possible sodium adduct, acetylation, acetylation with sodium adduct, respectively; *p*, and *pp*, mono- and bis-phosphorylation, respectively. (b) Fragmentation maps of the monophosphorylated α-casein by electron capture dissociation (ECD), which match the sequence from UniProt (UniProtKB-P02662), with phosphorylation on Ser88 (left) or Ser115 (right), respectively. (c) Representative ECD fragment ions demonstrating the coexistence of mono-phosphorylated isomers

The unphosphorylated and the mono-phosphorylated α- casein from the same charge state were then isolated and subjected to ECD. Because ECD of precursor ions with higher charge states

provide more extensive fragmentation,[259] charge state 23^+ was selected for the study. Nine different spectra of unphosphorylated and monophosphorylated α-casein with transients up to 9000 were used to ensure data quality and maximum sequence coverage. The number of observed ions increases as the number of transients accumulated increases (**Figure S3.2**), because averaging more data points resolves low abundance ions. Therefore, we analyzed spectra collected at the plateau area, where more transients do not provide as many additional observed ions. Consequently, the potential variation introduced by differences in data accumulation time and noise was minimized. Relative abundance of the $\boldsymbol{c}$ and $\boldsymbol{z}^{\bullet}$ fragment ions from the unphosphorylated (top) and monophosphorylated (bottom) precursors (23^+) normalized to the most abundant product ion were plotted against their corresponding cleavage sites (**Figure 3.2**). Globally, for the $\boldsymbol{c}$ ions, the cleavage patterns of the unphosphorylated and the monophosphorylated were similar, indicating a negligible effect of the single phosphorylation on ECD fragmentation efficiency of α-casein (**Figure 3.2a**). However, the total number of the $\boldsymbol{z}^{\bullet}$ ions was smaller compared with the c ions for both the unphosphorylated and phosphorylated forms. Furthermore, ECD of the unphosphorylated form provided more $\boldsymbol{z}^{\bullet}$ ions (77 $\boldsymbol{z}^{\bullet}$ ions) than that of the mono-phosphorylated form (61 $\boldsymbol{z}^{\bullet}$ ions), resulting in more cleavages (**Figure 3.2b**). A closer examination of cleavage sites adjacent to the phosphorylation site at Ser115 revealed similar fragmentation patterns for the unphosphorylated and the mono-phosphorylated α-casein. While the cleavage patterns remained mostly identical around the phosphorylation site Ser115, the N–C_α bond (at Leu120) that produces fragments $\boldsymbol{c}_{120}$ and $\boldsymbol{z}^{\bullet}_{89}$ was not cleaved in the unphosphorylated form (**Figure 3.2a and b insert**). However, the interpretation of these results was complicated by positional isomers. Additionally, the number and position of phosphorylation sites on an intact protein also potentially influence ECD fragmentation. Therefore, further investigation was conducted using β-casein.

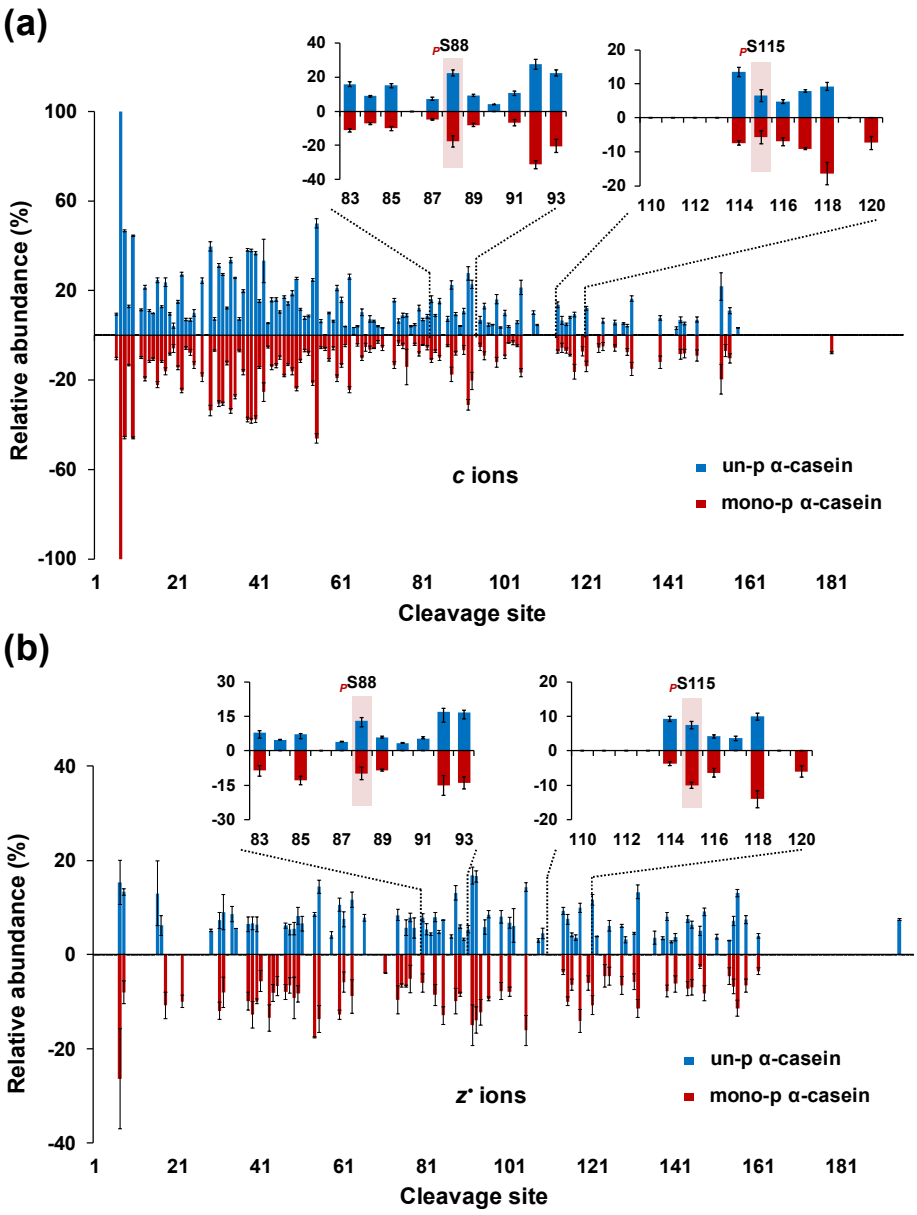


Figure 3.2. Statistical comparison of ECD fragments of un- and mono-phosphorylated α-casein (23^{+}) at Ser115. The relative intensity of ***c*** ions (a), and $z^{\bullet}$ ions (b) are plotted versus the cleavage site in the amino acid sequence. The blue bars (upper) represent unphosphorylated α-casein fragments from ECD, and the red bars (lower) represent mono-phosphorylated α-casein fragments from ECD. The standard error was calculated for each fragment based on nine spectra.

3.4.2 ECD of multi-phosphorylated vs. unphosphorylated β-casein

β-Casein (M_r 23568.32 Da) from bovine pancreas has five phosphorylation sites, with four (Ser15, Ser17, Ser18, and Ser19) clustered together and one (Ser35) further away from the N-terminal, making it a well-suited model for probing the potential influence of both the number and the location of the phosphorylation. We first assessed the reproducibility of the ECD fragmentation using identical ECD parameters (a standard default setting: 0.8 eV for electron energy and 50 ms for electron irradiation time) on 10 replicates of the fully phosphorylated β-casein. As shown in Figure 3.3a, the ECD spectra from all three replicates demonstrate nearly identical reduced precursor ion and product ion patterns. The relative abundances of representative fragment ions from all 10 replicates were plotted and grouped into unphosphorylated ***c*** ions (**Figure 3.3b**), phosphorylated ***c*** ions (**Figure 3.3c**), and unphosphorylated $z^{\bullet}$ ions (**Figure 3.3d**). Regardless of the addition of phosphorylations, the relative abundance of these ***c*** and $z^{\bullet}$ ions are highly consistent with a small standard error of the mean (SEM) within 4%. In general, smaller ions have relatively smaller SEM compared with larger ions since ions with higher molecular weight tend to exhibit lower signal-to-noise ratios and may undergo secondary fragmentations. The SEM of the larger ions can be minimized by accumulating more scans (data not shown). The highly reproducible ECD fragment patterns among multiple replicates allow us to confidently utilize one ECD experiment for each set of parameters to further evaluate its performance on phosphorylated β-casein compared with the unphosphorylated form.

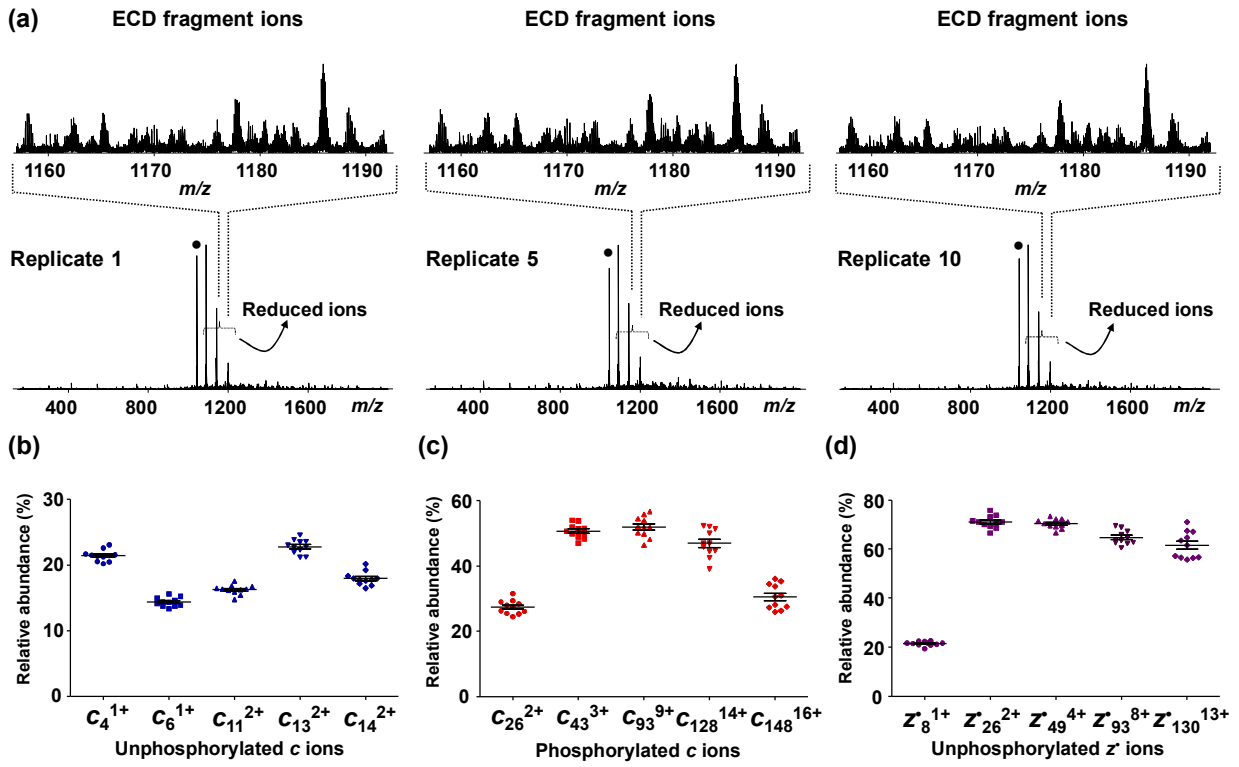


Figure 3.3. Evaluation of the reproducibility of ECD fragmentation. (a) Three representative ECD spectra from fully phosphorylated β- casein (23^+, indicated by ●) with 0.8 eV electron energy and 50 ms electron irradiation time from 10 replicates; insert, zoom-in view of the ECD fragments. (b)–(d) The normalized abundance of representative ions were plotted across all 10 replicates for unphosphorylated ***c*** ions, phosphorylated ***c*** ions, unphosphorylated $z^{\bullet}$ ions, respectively. The black bars indicate the standard error of the mean.

Using the standard parameters, the 23^+ charge state ions of the unphosphorylated and the phosphorylated β-casein were isolated individually for ECD on two different FTMS instruments, Thermo 7T LTQ/FT and Bruker 12T solariX FT-ICR MS (**Figure 3.4 and Figure S3.3**). The MS isolation of charge state 23^+ of the unphosphorylated and the phosphorylated β-casein demonstrated the existence of five phosphorylations (Δ = 399.79 Da) for the phosphorylated form and confirmed the completeness of the dephosphorylation assay (**Figure 3.4a**). Figure 3.4b and c depict the resulting fragmentation maps of the unphosphorylated and the phosphorylated β-casein

after matching to the theoretical sequence of β-casein (UniProtKB-P02666). Overall, 144 and 138 of the 208 backbone bonds were cleaved from the unphosphorylated and the phosphorylated β-casein respectively; 92 ***c*** ions and 109 $z^{\bullet}$ ions from the unphosphorylated β-casein, and 88 ***c*** ions and 106 $z^{\bullet}$ ions from the phosphorylated β-casein, were observed. Although ECD of the phosphorylated form exhibited slightly fewer ions, the cleavages were highly comparable between the phosphorylated and the unphosphorylated β-casein. Comparison of the relative abundance of ***c*** ions and $z^{\bullet}$ ions (**Figure 3.4 d and e**) further affirms the highly similar cleavage patterns, with only a few unmatched ions between the unphosphorylated (top) and phosphorylated counterparts (bottom). Most of the unmatched ions that were found in the unphosphorylated but not the phosphorylated β-casein were located near the highly phosphorylated area of the sequence surrounding phosphorylation sites at Ser15, Ser17, Ser18, and Ser19 (**Figure 3.4d**). This discrepancy could suggest that multiple phosphorylations existing in close proximity hinder ECD fragmentation. However, it is more likely that the introduction of negative charges from the multiple phosphorylated residues alters the overall positive charge of the product ions, as an extra positive charge from the basic residue Arg25 seems to alleviate this problem and resulted in the cleavage of the next ***c*** ion, $\boldsymbol{c}_{25}$. On the other hand, cleavages around the phosphorylated Ser35 were not compromised, with every N–Cα bond within nine residues before and after Ser35 cleaved ($\boldsymbol{c}_{25}$ to $\boldsymbol{c}_{45}$), suggesting that a single phosphorylation results in minimal influence on ECD fragmentation of the intact protein.

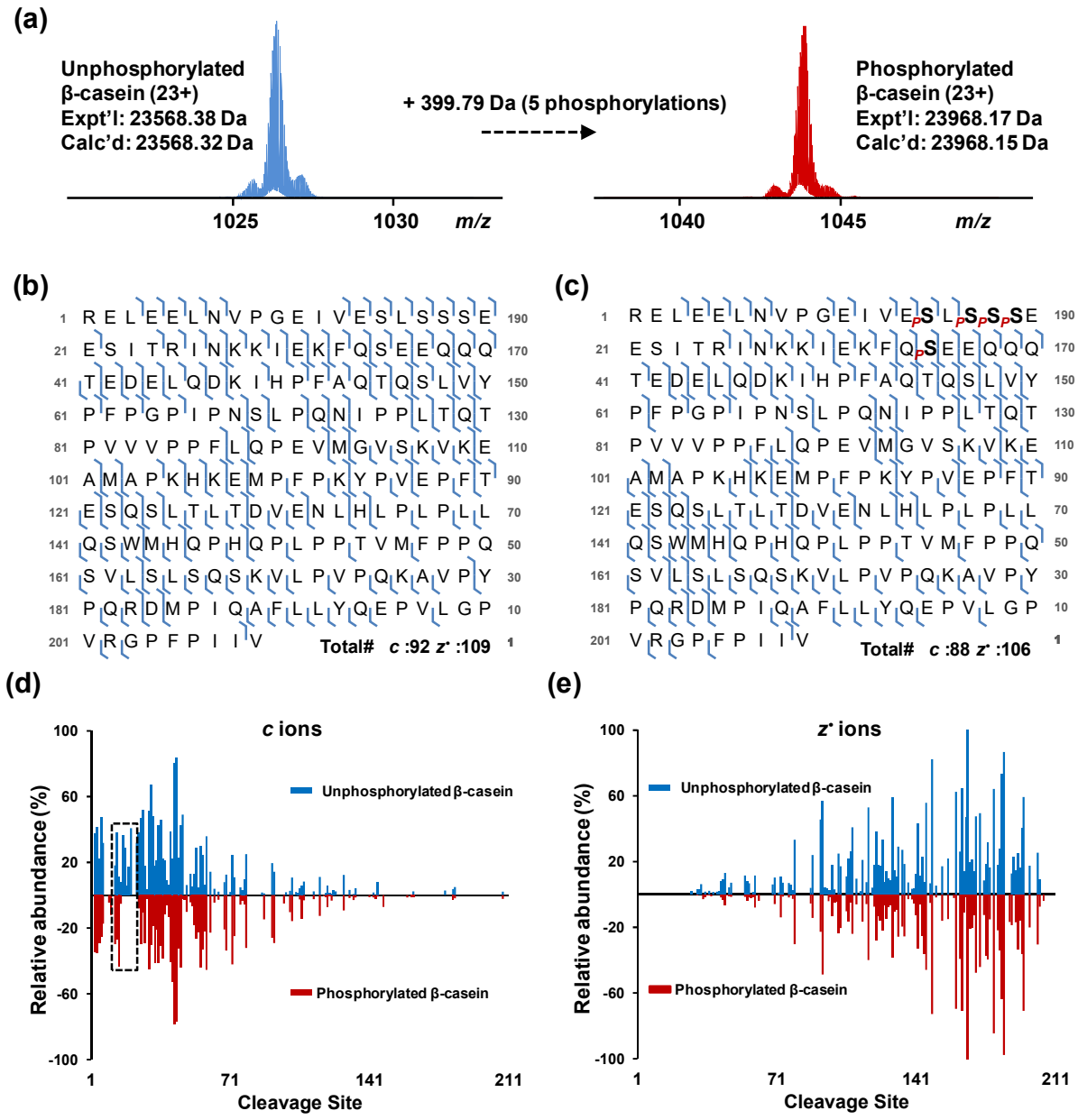


Figure 3.4. Fragmentation maps and comparison of relative abundance of the product ions from unphosphorylated and phosphorylated β-casein. (a) MS isolation of charge state 23^+ of the unphosphorylated (left) and the multi-phosphorylated β-casein (right) before ECD experiments. ECD fragmentation maps of the charge states 23^+ of the unphosphorylated (b) and the phosphorylated β- casein (c) using the "standard" ECD conditions, which match the sequence from UniProt (UniProtKB-P02666). *P* indicates the phosphorylation sites (Ser15, Ser17, Ser18, Ser19, and Ser35). The relative intensity of ***c*** ions (d), and $z^{\bullet}$ ions (e) are plotted against the cleavage sites in the amino acid sequence. The blue bars (upper) represent the unphosphorylated β-casein fragments, and the red bars (lower) represent the phosphorylated β-casein fragments.

3.4.3 ECD of mono-phosphorylated vs. unphosphorylated cardiac troponin I from swine and rhesus monkey

Cardiac troponin I (cTnI) was immunoaffinity-purified from swine (M_r 23919.72 Da) and rhesus monkey (M_r 23849.79 Da) tissue lysate followed by offline reverse phase desalting as described previously.[253,258] Similarly, we compared the relative abundance of fragment ions produced by ECD of the unphosphorylated (top) and phosphorylated (bottom) forms of endogenous cardiac troponin I from swine (phosphorylation site at Ser22), and from rhesus monkey (phosphorylation site at Ser23) (**Figure 3.5a, b**). Globally, 106 out of 209 backbone bonds were cleaved in the unphosphorylated swine cTnI, and 100 were cleaved from the mono-phosphorylated form. The total number of $\boldsymbol{c}$ ions and $\boldsymbol{z}^{\bullet}$ ions from the unphosphorylated compared with the mono-phosphorylated swine cTnI was 47 to 46 and 62 to 54, respectively. Near the phosphorylation site (Ser22), the number of $\boldsymbol{c}$ ions between $\boldsymbol{c}_{18}$ to $\boldsymbol{c}_{28}$ from the unphosphorylated compared with the mono-phosphorylated swine cTnI was 10 to 10 (**Table S3.1**). Similar findings were observed in the comparison of unphosphorylated and mono-phosphorylated cTnI from rhesus (**Figure 3.5b**), as the number of c ions between $\boldsymbol{c}_{18}$ to $\boldsymbol{c}_{28}$ from the unphosphorylated compared with the mono-phosphorylated rhesus monkey cTnI was also 10 to 10 near the phosphorylation site (Ser23). These results reiterate that although the overall number of total fragments is slightly decreased, a single phosphorylation event has a minimal influence on ECD fragmentation efficiency near the phosphorylation site of an intact protein.

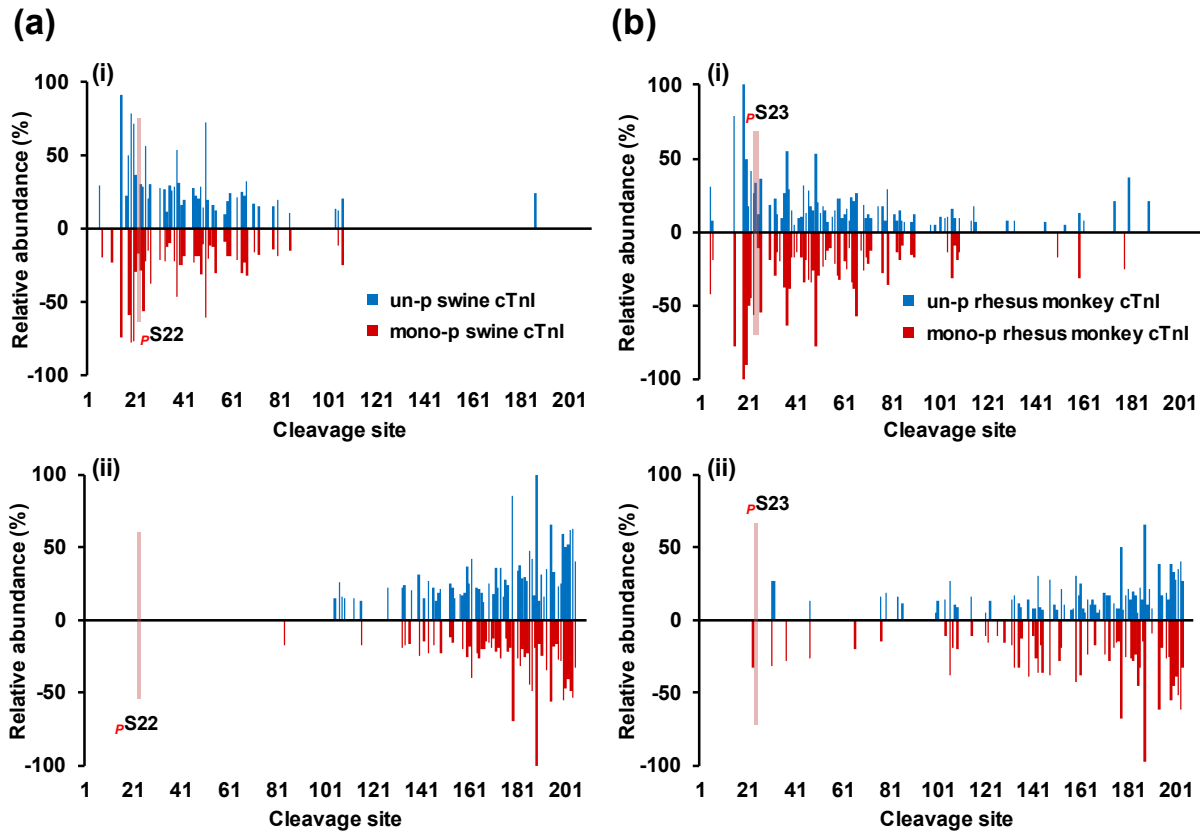


Figure 3.5. Comparison of the ECD fragments of un- and mono-phosphorylated proteins. (a) swine cardiac troponin I (cTnI) and (b) rhesus monkey cTnI. The normalized intensity of the ***c*** fragment ions (i) and the $z^{\bullet}$ fragment ions (ii) are plotted versus the cleavage site in the amino acid sequence. The matched sequences were from UniProt (UniProtKB-A5X5T5 with N-terminal acetylation and V116A variance) for swine cTnI, and (UniProtKB-A0A1D5RFH2 with N-terminal acetylation) for rhesus monkey cTnI. The upper part (blue bars) represents fragments from the unphosphorylated form; the lower part (red bars) represents fragments from the monophosphorylated form. The phosphorylation sites of the swine cTnI and the rhesus monkey cTnI are Ser22 and Ser23, respectively.

3.4.4 The effect of phosphorylation on ECD of protein vs. peptide ions

Compared with peptides, intact proteins are much larger in size and, therefore, the addition of one or more phosphorylations is less likely to perturb the overall ion behavior during ECD. We demonstrated that, locally, adjacent to the phosphorylation site, ECD cleavages remain similar

between mono-phosphorylated and unphosphorylated proteoforms. However, due to the introduction of a negatively charged moiety (deprotonated phosphate group), phosphorylated protein ions still experience different fragmentation than unphosphorylated counterparts even with selection of the same charge state for fragmentation. This becomes more apparent when multiple phosphorylations are present in close proximity on a protein. For the case of the multiply phosphorylated versus unphosphorylated β-casein, ECD fragmentation was greatly reduced at the region surrounding the multiple phosphorylation sites. In a previous study, Creese and Cooper investigated the ECD efficiency on phosphopeptides using a peptide from β- casein (Phe33 to Lys48; phosphorylated Ser35).[256] They found that the phosphorylated peptide did not fragment as well as the unphosphorylated counterpart, and postulated that the phosphorylation alters peptide folding through the formation of salt-bridges between the glutamine and the glutamic acid residues. Using a very high electron energy, they were able to cleave 13 out of the 15 N–C_α bonds on the phosphopeptide. In our case, ECD cleaved all of the previously mentioned 15 N-C_α bonds around Ser35 (**Figure 3.4b**) of the phosphorylated β- casein with low electron energy. Nevertheless, in contrast to peptide ions, protein ions carry significantly more charges in denatured conditions, which not only disturbs electrostatic interaction (salt bridges),[260,261] but also increases ECD reaction cross-section,[95] resulting in minor differences in the overall ECD cleavages between the phosphorylated and the unphosphorylated species without the requirement of additional electron energy as for phosphopeptides.[256] Our data also suggests that globally, ECD of the phosphorylated form generally produces slightly fewer fragment ions; however, it is drastically less affected by phosphorylation compared with the marked effects observed for phosphopeptides.

3.4.5 Investigation of relative ratio of individual ions between phosphorylated and unphosphorylated β-casein

Although ECD of both the phosphorylated and the unphosphorylated β-casein offers highly similar cleavages, the relative ratio of individual ions between the phosphorylated and unphosphorylated forms varies. Therefore, we further investigated the normalized relative abundance and compared the ratio between the individual ECD fragment ions from the same charge state (23+) of the phosphorylated and unphosphorylated β-casein as a function of different electron energy (**Figure S3.4 and 3.5**) and electron irradiation time (**Figure S3.6 and S3.7**). The electron energy and electron irradiation time were changed up and down systematically from the standard default parameter until nearly no product ions or no precursor ions were observed.

For both the phosphorylated and the unphosphorylated forms, the relative abundance of the small ***c*** ions (**Figure S3.4a, b**) and the $z^{\bullet}$ ions increases (**Figure S3.5a, b**); that of the intermediate ***c*** ions (**Figure S3.4d, e, g, h**) and the $z^{\bullet}$ ions (**Figure S3.5a, b**) remains relatively stable; and that of large ***c*** ions (**Figure S3.4k, l**) and the $z^{\bullet}$ ions (**Figure S3.5d, e**) decreases, as the electron energy and irradiation time increase. Consistent with the previous studies on peptides,[256,262] even a small increase in electron energy or electron irradiation time results in a reduction of the average fragmentation ion mass and a decrease of the fragmentation efficiency for ECD of intact proteins. This phenomenon occurs because highly charged large fragment ions are more susceptible to undergo secondary fragmentation and/or charge reduction.

The ratio of the relative abundance of the same fragment ions reveals that individual ion fragmentation efficiency deviates between the phosphorylated and unphosphorylated forms (**Figure 3.6**). A ratio of 1 indicates an equal fragmentation efficiency of a fragment ion from the phosphorylated and the unphosphorylated β-casein, and a ratio larger than 1 shows the specific ion

fragments better from the phosphorylated precursor. Compared to that of the larger ions, the ratio of the small ions excluding any phosphorylation sites remains relatively close to 1 (especially $\boldsymbol{c}_3$, $\boldsymbol{c}_4$, and $\boldsymbol{z}^{\bullet}_{15}$), indicating minimal variation of fragmentation efficiency of these ions from the phosphorylated the unphosphorylated β-casein (**Figure 3.6a**). For $\boldsymbol{c}_{36,}$ a $\boldsymbol{c}$ ion that includes all the phosphorylation sites, the ratio between the phosphorylated and unphosphorylated forms also remains close to 1. Similar to $\boldsymbol{c}_4$, $\boldsymbol{c}_{36}$ was the consequence of the cleavage between two adjacent highly polar glutamic acid residues (**Figure 3.6b**). Previous study show N-C_α bonds close to highly polar residues glutamic acid or aspartic acid are preferentially cleaved by ECD.[263] This is consistent with our data, where the fragment ion abundances of $\boldsymbol{c}_4$ and $\boldsymbol{c}_{36}$ were relatively high (**Figure 3.6a**), explaining the minimal difference of the relative ion abundance of $\boldsymbol{c}_4$ and $\boldsymbol{c}_{36}$ between the phosphorylated and the unphosphorylated. However, some of the ions were fragmented preferentially from the unphosphorylated precursors while others were favored from the phosphorylated species. For example, $\boldsymbol{c}_{14}$, the fragment ion prior to the first phosphorylation site S15, and $\boldsymbol{c}_{35,}$ the ion right after the last phosphorylation site S35, fragmented better from the phosphorylated β-casein (**Figure 3.6a**). Coincidentally, $\boldsymbol{c}_{14}$ and $\boldsymbol{c}_{35}$ were the products of the cleavage between a serine and glutamic acid residues. We postulated that the presence of another polar phosphorylation led to the preferential cleavage of the N-Cα bonding between the serine and the glutamic acid residues. On the other hand, $\boldsymbol{c}_{35}$ from the phosphorylated β-casein was less fragmented compared to the unphosphorylated, as the basic K32 residue possibly interacted with the phosphorylation moiety. For the larger $\boldsymbol{c}$ and $\boldsymbol{z}^{\bullet}$ ions that cover phosphorylation sites ($\boldsymbol{c}_{93}$, $\boldsymbol{c}_{128}$, $\boldsymbol{c}_{148}$, and $\boldsymbol{c}_{184}$ in **Figure S3.4d**; $\boldsymbol{z}^{\bullet}_{178}$ in **Figure S3.5b**), different from peptide studies, the ratio indicates that fragmentation favored toward the phosphorylated form. We reason that the larger ions with phosphorylations endure the deposition of the same amount of electron energy better

than the unphosphorylated ones, avoiding further dissociation. In addition, a moderately upward trend of the plot as electron energy or electron irradiation time increases hints that a higher deposition of energy improves the fragmentation of phosphoproteins, especially for smaller fragments (**Figure S3.4a, Figure S3.6a**).

More importantly, the data above provide insights on the quantitative information acquired by ECD of phosphorylated proteins: phosphorylation in fact influences ECD efficiency on the level of fragment ion abundance, and as a result, further investigation is required prior to applying the ECD fragmentation based-method for directly probing isomeric mixtures of phosphoproteins. Figure 3.6 show a distribution of ions that favors the phosphorylated or unphosphorylated forms on a case-by-case basis. For instance, for phosphorylation site S15, the ratio of $\boldsymbol{c}_{15}$ leaned toward the phosphorylated, but $\boldsymbol{c}_{16}$ toward the unphosphorylated (**Figure S3.4b**). Even with the same ions ($\boldsymbol{c}_{93}$), different charge states (10+ and 11+) resulted in different fragment ion ratios between the phosphorylated and unphosphorylated forms, though the overall patterns across electron energies and irradiation time behaved similarly (**Figure S3.4d**). If *intraspectra* normalization is needed,[212] only $\boldsymbol{c}_3$, $\boldsymbol{c}_4$, and $\boldsymbol{z}^{\bullet}_{15}$ (ratio close to 1), the small ions far away from the phosphorylation sites, meet the requirement of no change of the relative ion abundance between the phosphorylated and the unphosphorylated forms. Therefore, depending on the selection of ion pairs, bias could be introduced when determining the occupancy of phosphorylation sites. This is a similar problem demonstrated in the study of quantitative analysis of glycation using ECD described by Stefanowicz et al., in which the authors found the slope of the standard curve drastically varied depending on the selection of ion pairs [162]. Although ECD fragments modified by acetylation and methylation have shown good quantitative correlation to the composition of isomeric mixtures [157],

product ions with other PTMs such as phosphorylation and glycation might require calibration curves for accurate analysis.

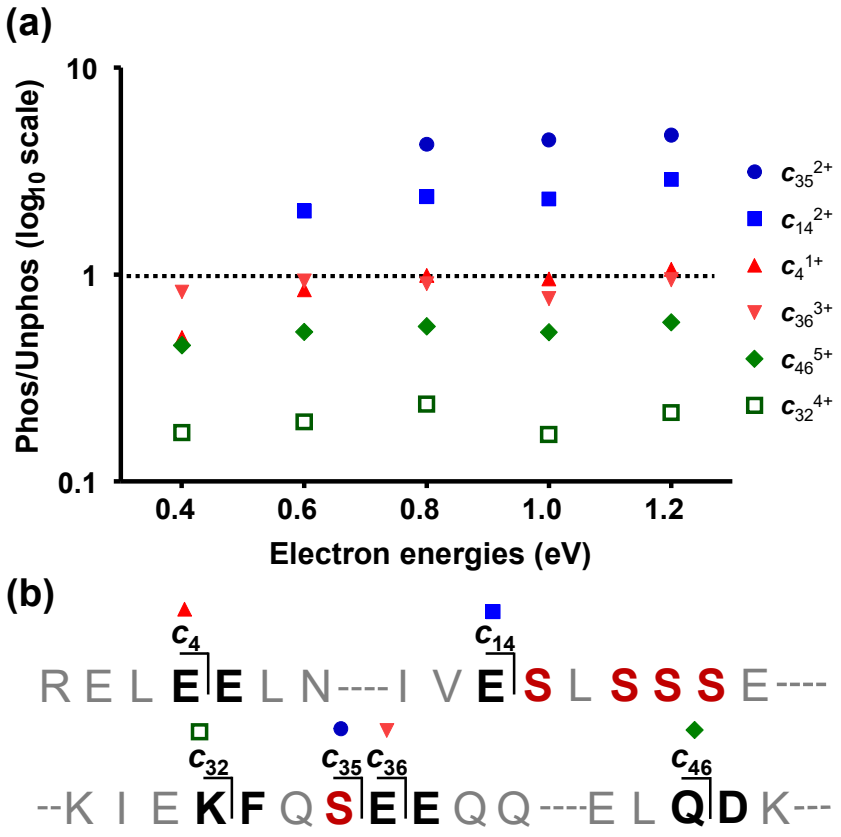
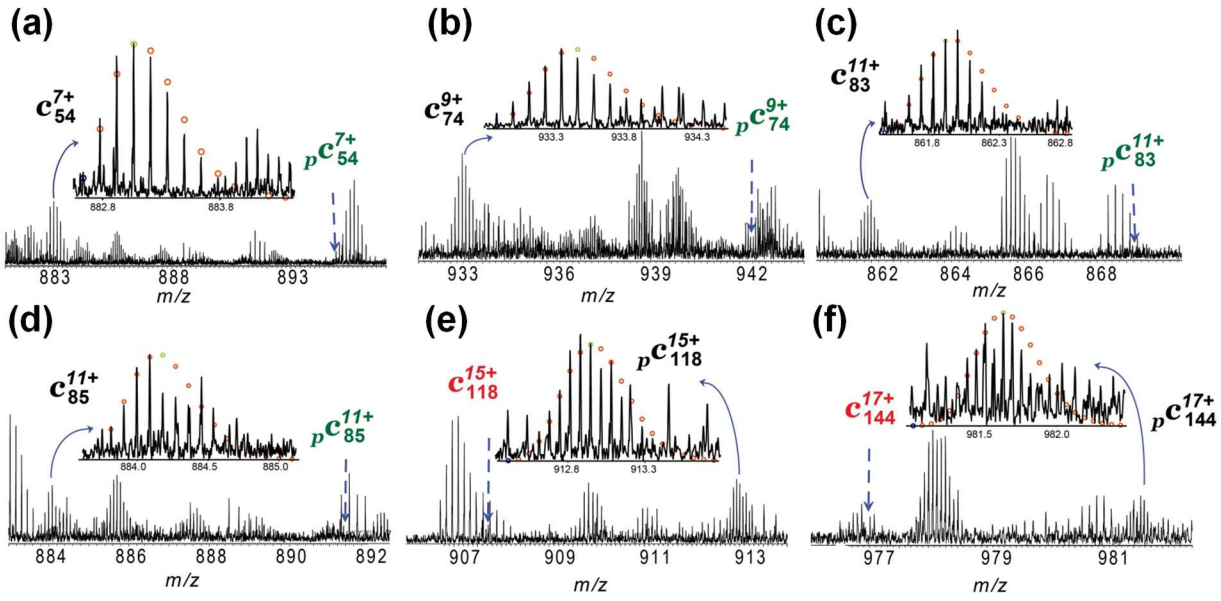


Figure 3.6. ECD results in varying relative abundance of fragment ions formed from the same cleavage between the phosphorylated and the unphosphorylated β-casein. (a) The ratio (phosphorylated to unphosphorylated) of the relative abundance of the representative ***c*** ions with different electron energies. Dots above the dash line at 1 indicates favorable ion fragmentation in the phosphorylated form; dots below the dash line indicates the opposite. (b) Illustration of the cleavages and sequence corresponding to the representative ions. The phosphorylation sites are S15, S17, S18, S19, and S35.

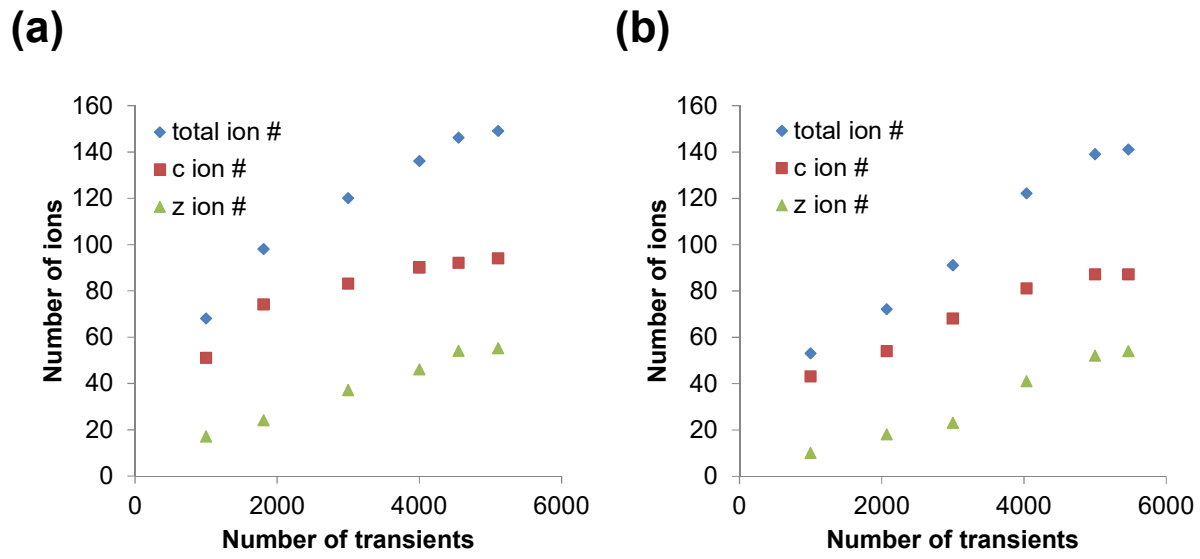
3.5 Conclusions

In this study, we systematically investigated the influence of phosphorylation on ECD fragmentation of intact proteins, including mono-phosphorylated α-casein, multiply phosphorylated β-casein, and immunoaffinity-purified cardiac troponin I from swine and rhesus monkey, by analyzing the relative ion abundance from the unphosphorylated and the phosphorylated forms under the same ECD conditions. The results suggest that compared with phosphopeptides, phosphorylation has little deleterious effect on ECD fragmentation of intact proteins. Globally, phosphorylation on intact proteins slightly decreases the total number of fragment ions, but the pattern of fragmentation remains highly similar. Locally, phosphorylation on intact proteins tends not to significantly affect the fragmentation and cleavages of the amino acid residues adjacent to the phosphorylation sites; however, when multiple phosphorylations are present in close proximity, cleavages near the phosphorylation sites decrease.

3.6 Supplemental Information

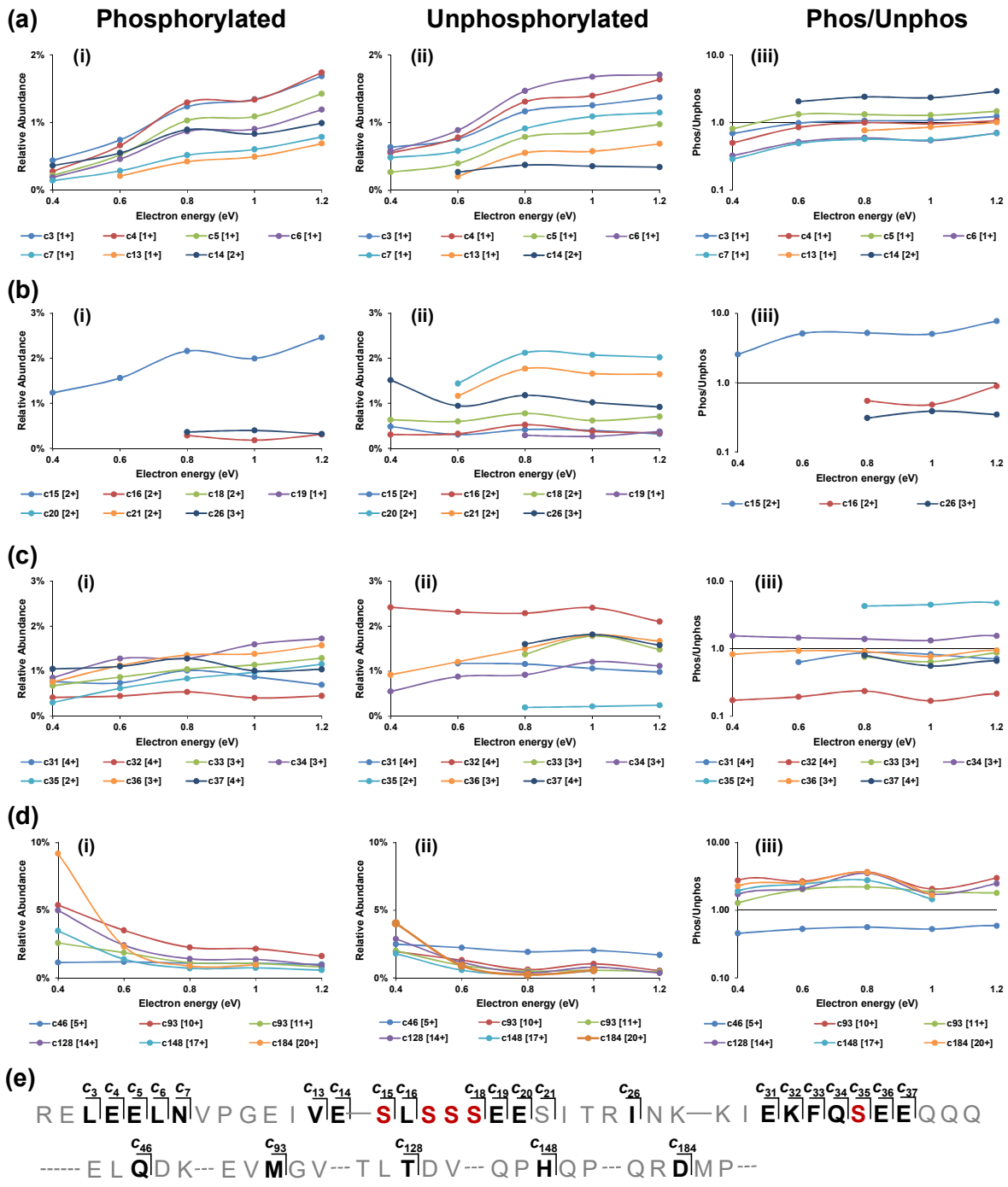


Supplemental Figure 3.1. Other representative ECD fragment ions for phosphorylation site localization in mono-phosphorylated α-casein.The arrows with solid line indicate observed ions whereas arrows with dash line point to the position of the undetected phosphorylated or unphosphorylated ions. The insert shows the isotopic distribution and the theoretical match of the ions.

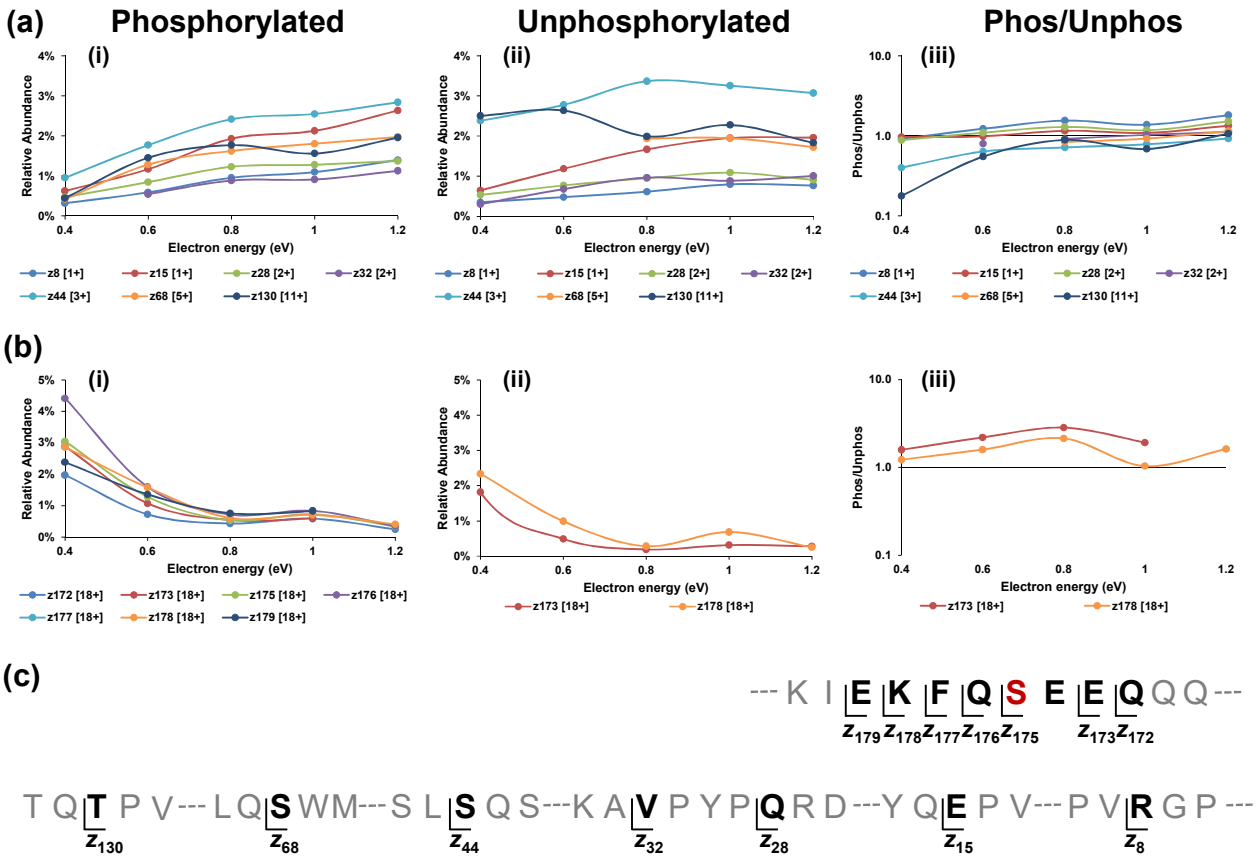


Supplemental Figure 3.2. Number of detected product ions *versus* number of transients. (a) Un- and (b) mono-phosphorylated α-casein. As the number of transients accumulates, the numbers of observed ions (***c***, $z^{\bullet}$, and total fragment ions) increase and plateau eventually.

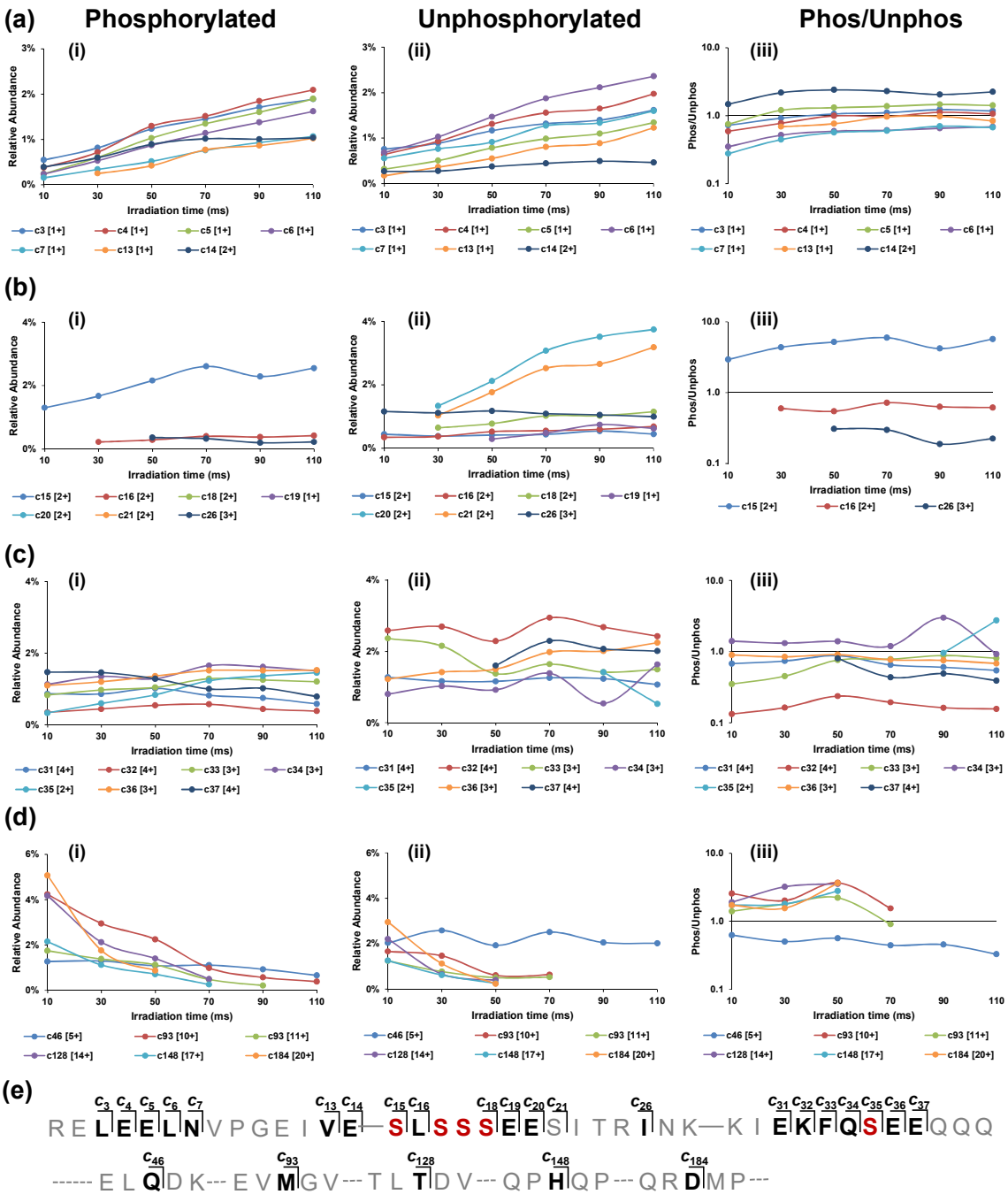
Supplemental Figure 3.3. ECD fragmentation maps of the (a) un- and (b) multi-phosphorylated β-casein acquired on the 7T LTQ/FTMS.Highlighted letter $_P$S indicates the location of the phosphorylation sites (Ser15, Ser17, Ser18, Ser19, and Ser35).



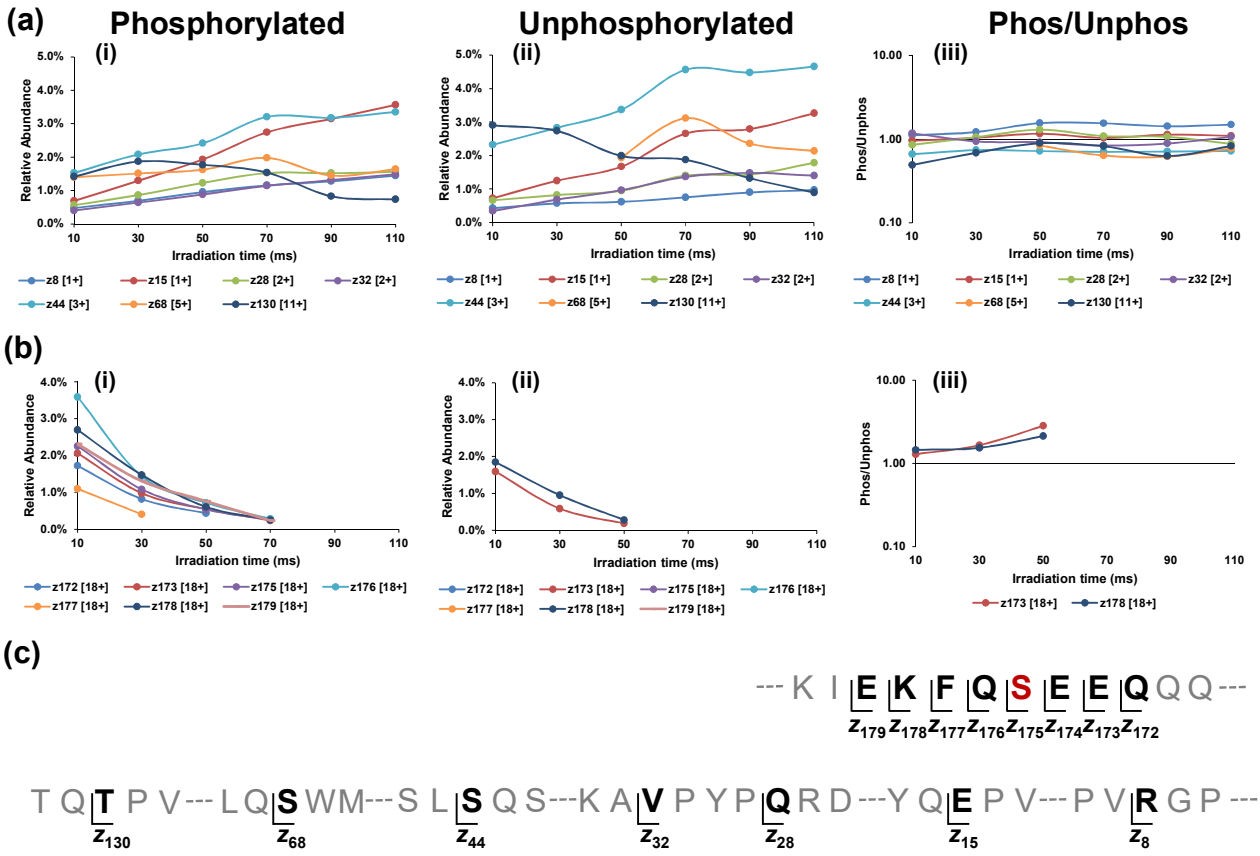
Supplemental Figure 3.4. Relative ion abundance of ECD fragment ***c*** ion from the phosphoryated and the unphosphorylated β-casein at different electron energies. Normalized relative abundance of representative ***c*** ions from the ECD fragmentation of the phosphorylated β-casein (i) and unphosphorylated β-casein (ii); and the ratio (log scale) of the relative abundance of the ***c*** ions from the phosphorylated to the unphosphorylated (iii), versus ECD electron energy. The representative ions covers the ***c*** ions near the N-terminal prior to the first phosphorylation site S15 (a); the ***c*** ions around phosphorylation sites S15, S17, S18, and S19 (b); the ***c*** ions around phosphorylation site S35 (c); and the larger ***c*** ions (d).



Supplemental Figure 3.5. Relative ion abundance of ECD fragment $z^{\bullet}$ ion from the phosphoryated and the unphosphorylated β-casein at different electron energies. Normalized relative abundance of representative $z^{\bullet}$ ions from ECD fragmentation of the phosphorylated β-casein (i) and unphosphorylated β-casein (ii); and the ratio (log scale) of the relative abundance of the $z^{\bullet}$ ions from the phosphorylated to the unphosphorylated (iii), versus ECD electron energy. The representative ions covers the $z^{\bullet}$ ions prior to the phosphorylation sites (a) and the larger $z^{\bullet}$ ions around phosphorylation site S35 (b). The cleavages and sequence corresponding to the representative ions are shown in panel c.



Supplemental Figure 3.6. Relative ion abundance of ECD fragment ***c*** ion from the phosphoryated and the unphosphorylated β-casein with different electron irradiation times. Normalized relative abundance of representative ***c*** ions from the ECD fragmentation of the phosphorylated (i) and unphosphorylated β-casein (ii); and the ratio (log scale) of the relative abundance of the ***c*** ions from the phosphorylated to the unphosphorylated (iii), versus ECD irradiation time. The representative ions covers the small ***c*** ions near the N terminal prior to the first phosphorylation site S15 (a); the early ***c*** ions around phosphorylation sites S15, S17, S18, and S19 (b); the middle c ions around phosphorylation site S35 (c); and the larger c ions (d). The cleavages and sequence corresponding to the representative ions are shown in panel e.



Supplemental Figure 3.7. Relative ion abundance of ECD fragment $z^{\bullet}$ ion from the phosphoryated and the unphosphorylated β-casein with different electron irradiation times. Normalized relative abundance of representative $z^{\bullet}$ ions from the ECD fragmentation of the phosphorylated β-casein (i) and the unphosphorylated β-casein (ii); and the ratio (log scale) of the relative abundance of the $z^{\bullet}$ ions from the phosphorylated to the unphosphorylated (iii), versus ECD radiation time. The representative ions covers the $z^{\bullet}$ ions prior to the phosphorylation sites (a) and the larger $z^{\bullet}$ ions around phosphorylation site S35 (b). The cleavages and sequence corresponding to the representative ions are shown in panel c.

	un-p swine cTnI	mono-p swine cTnI	un-p primate cTnI	mono-p primate cTnI
Total number of ***c*** ion	47	46	86	61
Total number of ***z•*** ion	62	54	82	57
Number of ***c***18-***c***28	10	10	10	10
Number of ***z•***181-***z•***191	0	0	0	1

Supplemental Table 3.1. The number of total and local ***c*** and ***z•*** ions adjacent to the phosphorylation sites of cardiac troponin I. Ions from unphosphorylated and mono-phosphorylated cardiac troponin I (cTnI) from swine and rhesus monkey (primate) heart tissues in one ECD spectrum were summarized.

3.7 Acknowledgements

This work is dedicated to Professor Kristina Håkansson, the recipient of the 2016 Biemann Medal of the ASMS, for her outstanding achievement in the field of mass spectrometry related to electron-based activation methods. The authors thank Fangmin Xu, Han Zhang, and Huseyin Guner on the initial studies on the LTQ/FT. They thank Jeremy Wolff from Bruker for technical assistance on the Solarix FTMS. They also thank Dr. Barbara Lewis and Zachery Gregorich for critical reading of the manuscript. Financial support was kindly provided by NIH R01HL109810 and HL096971 (to Y.G.) and GM117058 (to Song Jin and Y.G.). They also acknowledge NIH high-end instrument grant S10OD018475 (to Y.G.). T.T. acknowledges support of the NIH Chemical Biology Interface Training grant NIH T32GM008505.

Chapter 4: Online Hydrophobic Interaction Chromatography-Mass Spectrometry for Top-down proteomics

This chapter has been published and is adapted from:

Chen, B.; Peng, Y.; Valeja, S. G.; Xiu, L.; Alpert, A. J.; Ge, Y. Online Hydrophobic Interaction Chromatography-Mass spectrometry for Top-down Proteomics. *Analytical Chemistry.* 2016, 88, 1885-1891.

4.1 Abstract

Recent progress in top-down proteomics has led to a demand for mass spectrometry (MS)-compatible chromatography techniques to separate intact proteins using volatile mobile phases. Conventional hydrophobic interaction chromatography (HIC) provides high-resolution separation of proteins under non-denaturing conditions but requires high concentrations of nonvolatile salts. Herein, we introduce a series of more hydrophobic HIC materials that can retain proteins using MS-compatible concentrations of ammonium acetate. The new HIC materials appear to function as a hybrid form of conventional HIC and reverse phase chromatography. The function of the salt seems to be preserving protein structure rather than promoting retention. Online HIC-MS is feasible for both qualitative and quantitative analysis. This is demonstrated with standard proteins and a complex cell lysate. The mass spectra of proteins from the online HIC-MS exhibit low charge state distributions, consistent with those commonly observed in native mass spectrometry. Furthermore, HIC-MS can chromatographically separate proteoforms differing by minor modifications. Hence, this new HIC-MS combination is promising for top-down proteomics.

4.2 Introduction

Increasing evidence has implicated protein post-translational modifications (PTMs) in the regulation of protein function in health and disease.[4,17,55,57,97,158,205,264-266] Although bottom-up proteomics offers high sensitivity, high throughput, and deep proteome coverage, it is suboptimal for analysis of PTMs.[210] Top-down mass spectrometry (MS)-based proteomics is arguably the most powerful method for the analysis of proteoforms[207], including those arising as a result of PTMs, genetic variations (e.g., polymorphisms and mutations), and alternative splicing of RNA transcripts.[4,19,94,267] Consequently, there is a significant interest in the top-down approach.[4,18-20,94,268,269]

Nevertheless, many challenges remain for top-down proteomics, including difficulty in separating intact proteins.[4,59] Liquid chromatography (LC) with volatile mobile phases is the preferred method for separation of complex mixtures because it affords online separation with direct coupling to a mass spectrometer and is amenable to automation, thus eliminating tedious fraction collection procedures and allowing for high throughput analysis.[270] However, the number of chromatography methods that are compatible with this arrangement is limited. Reversed-phase chromatography (RPC) is the most popular such method.[91,97,158,211,271,272] However, the conditions used in RPC denature many intact proteins, exposing numerous hydrophobic residues in the protein core.[273-275] This can lead to peak broadening, low protein recovery, and failure to elute at all, with the severity depending on the RPC material used.[273] Hydrophilic interaction chromatography (HILIC) has been utilized for direct coupling with MS, but HILIC can only be used to separate proteins that remain soluble in solutions containing high concentrations of organic solvent.[44,276,277] Recent studies have demonstrated online coupling of either size exclusion chromatography (SEC) or ion exchange chromatography (IEC) with MS using a volatile buffer

containing ammonium acetate (NH_4OAc).[74,75] However, SEC is not yet a high-resolution method.[59,278] While IEC is a high-resolution method, complex protein mixtures often require more than one dimension of fractionation.[233] Therefore, there is a need for additional methods for online LC-MS that can provide high-resolution separation of a wide range of proteins with minimal denaturation.

Hydrophobic interaction chromatography[279] (HIC) is a technique that separates proteins based on hydrophobicity but with complementary selectivity to RPC and with high-resolution under non-denaturing conditions.[233,280,281] The stationary phases used for HIC generally feature low-density and moderately hydrophobic ligands (propyl and phenyl) attached to a hydrophilic underlayer.[280] Being a non-denaturing mode, HIC only interacts with a limited set of hydrophobic residues on the surface of a protein's tertiary structure. Proteins then elute in order of increasing surface hydrophobicity, with high recovery and unusually high sensitivity to conformational variation.[279] In HIC, a decreasing gradient of a salt high in the Hofmeister series[279] is used, typically a sulfate, phosphate, or citrate. Such salts are not compatible with MS analysis. We recently reported that ammonium tartrate provides effective separation of intact proteins when used in HIC and minimizes protein adduction and MS signal suppression as compared to the common sulfate salts.[233,281] Nevertheless, ammonium tartrate is not volatile and, thus, a desalting step is necessary prior to MS analysis of the HIC eluates. Ammonium acetate (NH_4OAc), on the other hand, is volatile and features reduced nonspecific sodium adduction.[282,283] Unfortunately, it provides inadequate protein retention when used with conventional HIC materials, such as PolyPROPYL A.[281] Using a HIC column with pentyl-ligands, Gooding *et al.* obtained adequate retention of proteins only with 4 M NH_4OAc.[283] Hence, we propose to increase the hydrophobic character of the HIC stationary phase to the point that proteins are retained and eluted using

NH_4OAc concentrations of 1 M or less, a range which has been previously demonstrated to be compatible with MS analysis.[43] Accordingly, in this study, a series of more hydrophobic HIC materials were synthesized and characterized in search of a balance between retention and denaturation with a MS-compatible salt concentration. We have further demonstrated the feasibility of online HIC-MS for top-down proteomics using these new HIC materials.

4.3 Material and Methods

4.3.1 Chemicals and reagents

All reagents were acquired from Sigma-Aldrich Inc. (St. Louis, MO, USA) unless noted otherwise. HPLC grade water, acetonitrile (ACN), and NH_4OAc were purchased from Fisher Scientific (Fair Lawn, NJ, USA).

4.3.2 Sample preparation

Standard protein samples (RiA, ribonuclease A from bovine pancreas; Cyt, cytochrome C from equine heart; ChA, α-chymotrypsinogen A from bovine pancreas; Lys, lysozyme from chicken egg white; Apr, aprotinin from bovine lung; Trp, trpsinogen from bovine pancreas; Lac, β-lactoglobulin B from bovine milk) were used without additional purification. HPLC-grade water was used to prepare protein standards at 10 mg/mL. Subsequent dilution using 1 M NH_4OAc (at least 3 vols. per vol. protein solution) brought the final concentration to 0.1-0.5 mg/mL unless noted otherwise. The six-protein mixture contained RiA, Apr, Trp, Lys, ChA, and Lac, in which each protein had a final concentration to be 0.4 mg/mL, 0.1 mg/mL, 0.5 mg/mL, 0.1 mg/mL, 0.4

mg/mL, and 0.5 mg/mL respectively. *Escherichia coli* (*E. coli*) cell pellet from the BL21 strain grown in-house was suspended in lysis buffer containing 50 mM HEPES, 150 mM NaCl, 0.1% NP-40, and 1x protease and phosphatase inhibitor cocktail tablets (Roche, Penzburg, Germany) with the ratio of pellet to buffer being 1:10 (mg/μL). The suspension was sonicated to lyse the cells, and was subsequently centrifuged at 10,400 *g*, 4 °C for 20 min. The resulting lysate (supernatant) was centrifuged 5 times with 10kDa ultracentrifugal filters at 4 °C for 5 min at 16,100 *g* with 1 M NH_4OAc. The supernatant solution containing soluble proteins from *E. coli* cell lysate was utilized directly for online HIC-MS analysis.

4.3.3 HIC materials

In addition to the existing PolyPROPYL A and PolyBUTYL A materials,[280] PolyPENTYL A, PolyHEXYL A, PolyHEPTYL A, PolyOCTYL A, PolyNONYL A, PolyDECYL A, and PolyHYDROXYDECYL A were synthesized at PolyLC Inc. (Columbia, MD, USA). The resulting coatings are polyasparagines with functional side chains of various lengths as indicated by the names. The stationary phase materials were slurry-packed into 100 mm × 4.6 mm i.d. stainless steel columns for initial experiments with HPLC-UV, and 100 mm × 0.2 mm i.d. capillary columns for online HIC-MS analysis. Initial studies and optimization were performed using an Essence LC system from Scientific Systems Inc. (State College, PA, USA). The flow rate was 1.0 mL/min and UV absorbance detection was at 280 nm.

4.3.4 HIC-MS

100 mm × 0.2 mm i.d. PolyPENTYL A, PolyHEXYL A and PolyHEPTYL A capillaries were used on a Bruker Nano-Advance HPLC system (Fremont, CA, USA). Mobile phase A (MPA) contained 1M NH_4OAc and mobile phase B (MPB) contained 20 mM NH_4OAc (overall) in 50:50 water/ACN without any adjustment of pH. A 15 min linear gradient (from 100% MPA to 100% MPB) followed by 100% MPB for 5 min and 100% MPA for another 5 min was used at a flow rate of 3.0 or 2.4 µL/min. 2 µL of sample were injected for all the HIC-MS runs. Samples eluted from HIC columns were electrosprayed into a MaXis Plus Q-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany). End plate offset and capillary voltage were optimized at 500 V and 4000 V, respectively. The nebulizer was set to 1.5 bar, and the dry gas flow rate was 4.0 L/min at 200 °C. Mass range was set to 200 – 2800 *m/z* initially and 100-6000 *m/z* for subsequence runs to see lower charge states. The quadruple low mass was set to 1600 *m/z*. All data were collected using Hystar 3.2 and otofControl 3.4 (Bruker Daltonics). The total HIC-MS data run was 25 min per sample. All data were analyzed and chromatograms were smoothed using Compass DataAnalysis 4.3. Maximum Entropy was used as the deconvolution algorithm (resolution was set below 40,000).

4.4 Results

4.4.1 Conventional HIC columns with the volatile salt NH4OAc

The performance of NH_4OAc was first evaluated with columns of the existing HIC materials PolyPROPYL A and PolyBUTYL A to estimate the concentration of NH_4OAc necessary for

retention and to assess the effect of ligand length. With PolyPROPYL A, a gradient starting with concentrations of NH_4OAc ranging from 0.5-1 M resulted in elution of RiA and ChA in the void volume (**Figure 4.1a**). ChA was slightly retained when starting with 2.5 M NH_4OAc (**Figure 4.1a**), but such a high concentration of salt is impractical for online LC-MS analysis since it could overwhelmingly suppress protein signals. With the slightly more hydrophobic PolyBUTYL A, retention and separation of RiA and ChA was improved but still inadequate (**Figure 4.1b**). Retention decreased between 2.5 and 0.75 M salt, as expected, but retention of ChA increased again with 0.5 M salt concentration. At this low concentration of salt, the hydrophobic residues in the core may be exposed to the stationary phase,[284] leading to increased retention and peak broadening. We speculated that a rather high initial salt concentration would be needed to preserve the tertiary structure.

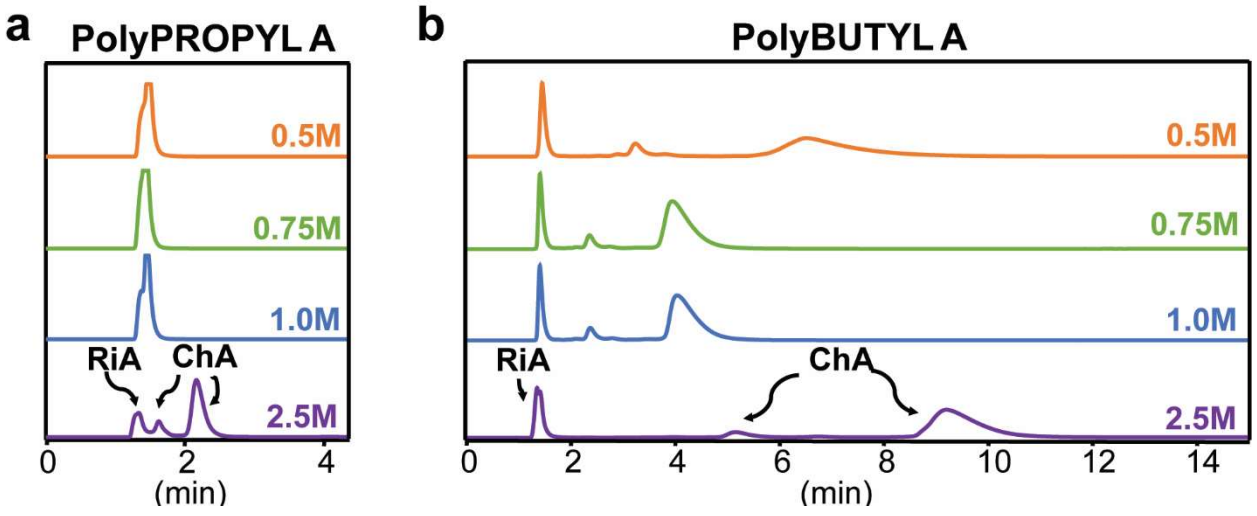


Figure 4.1. Conventional HIC columns with NH_4OAc. UV chromatograms of RiA and ChA mixture using (a) PolyPROPYL A and (b) PolyBUTYL A from the indicated concentration of NH_4OAc to 20 mM NH_4OAc over a 15 min gradient.

4.4.2 Characterization of a series of new HIC materials

We first tested solvent conditions for PolyPENTYL A (**Figure 4.2a**). A typical HIC gradient from 2.5 M to 20 mM NH_4OAc resulted in no elution of Lys even after a subsequent gradient to 20 mM NH_4OAc with 50% ACN (**Figure 4.2a i**; an A→B→C solvent sequence). We next used a reversed-phase-type gradient, from 20 mM NH_4OAc to 20 mM NH_4OAc with 50% ACN (a B→C solvent sequence), which also failed to elute Lys (**Figure 4.2a ii**). Elution of Lys was only observed in the 15-min time frame in a well-shaped peak when the organic solvent was introduced at the same time that the salt concentration was decreased (**Figure 4.2a iii**; an A→C solvent sequence). Therefore, with the new HIC materials some organic solvent seems to be required in any mobile phase used for a gradient for the elution of some proteins in a reasonable time frame. With a higher concentration of ACN in the final mobile phase, Lys eluted earlier, as expected (**Figure S4.1**). A comparison across all new HIC materials shows no appreciable increase in retention time of Lys in its native structure (**Figure 4.2b**). However, with the more hydrophobic materials, an increasing percentage of the Lys sample is seen in a broad and later-eluting peak which probably corresponds to a denatured form. With a RiA and ChA mixture, retention increased significantly when going from PolyBUTYL A to PolyPENTYL A (**Figure 4.2c**); but retention times did not change significantly with a further increase in ligand length (**Figure 4.2b**). Additionally, high concentrations of NH_4OAc did not increase retention of these protein standards by the more hydrophobic materials in the series (**Figure S4.2**). These properties are usually associated with RPC rather than conventional HIC. Our data imply that the function of a high salt concentration is most likely to preserve protein structure, since starting at low salt concentration resulted in partial or no elution. At least 0.75 M of NH_4OAc was required for this with the standards RiA and ChA (**Figure S4.2**).

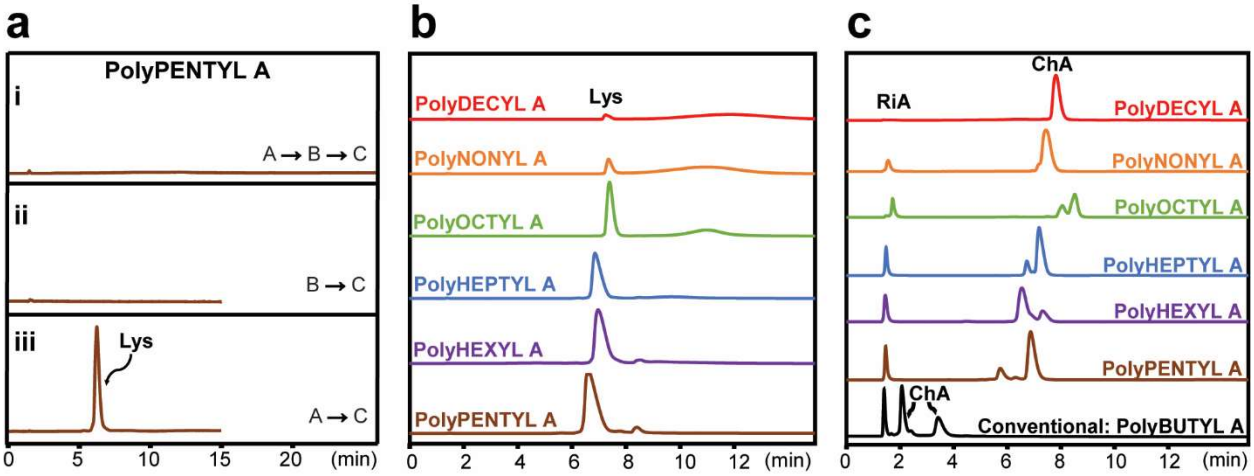


Figure 4.2. Characterization of the new HIC materials. (a) Lys with PolyPENTYL A in three different elution gradient sequences. (i) 10 min gradient from A (2.5 M NH_4OAc) to B (20 mM NH_4OAc), then 10 min of gradient from B to C (20 mM NH_4OAc with 50% ACN) followed by 5 min isocratic elution with C. (ii) 10 min gradient from B to C followed by 5 min isocratic elution with C. (iii) 10 min gradient from A to C followed by 5 min elution with C. (b) Lys and (c) RiA and ChA mixture with different HIC columns; 1 M NH_4OAc to 20 mM NH_4OAc with 50 % ACN over a 15 min gradient. Detection: 280 nm.

As noted previously, with increasing hydrophobicity of the stationary phase there is a tendency for some proteins to be denatured upon contact despite the protective effect of the NH_4OAc.[285] Denaturation typically results in later elution or no elution at all. Our data indicate that susceptibility to this phenomenon varies with different proteins. Cyt eluted as a well-shaped peak from PolyPENTYL A but was retained much more strongly on PolyHEXYL A and PolyHEPTYL A (**Figure S4.3**). Lys and RiA exhibited a similar trend starting with PolyOCTYL A, whereas ChA seems to have withstood contact even with PolyDECYL A (**Figure 4.2c**). Thus, the choice of HIC material to use can vary depending on the proteins. For example, an antibody-drug conjugate containing several molecules of a hydrophobic drug should probably be run on a

less hydrophobic column (cf. PolyPENTYL A) than for a hydrophilic protein such as cytochrome C. However, the mass spectra of these proteins suggest that the later elution does not necessarily reflect denaturation (**Figure S4.4**). The performance of the PolyHYDROXYDECYL A material was similar to that of PolyBUTYL A [*data not shown*], and it was not considered further.

4.4.3 Online HIC-MS of standard proteins

Based on these initial studies, to minimize denaturation and ensure adequate separation, we chose capillaries of PolyHEXYL A, PolyHEPTYL A, and PolyPENTYL A, 100 mm x 0.2-mm, for online HIC-MS experiments. We first examined a single protein, Lys, in a 15 min gradient from 1 M NH_4OAc to 20 mM NH_4OAc containing 50% ACN. The base peak chromatogram featured a well-retained peak with minimal denaturation (**Figure 4.3a**). The average mass spectrum of the peak demonstrated the detection of Lys and the distribution of the charge states fell predominantly at low charge states (8+, 7+, and 6+), characteristic of a native mass spectrum of non-denatured proteins.[43,286] The zoom-in mass spectrum of highlighted charge state 8+ showed the high-resolution isotopomer envelope of Lys and minimal adduction of Na^+. This result demonstrates the feasibility of the online HIC-MS method and the minimization of denaturation. Subsequently we subjected the Lys sample to serial dilution and assessed reproducibility and quantitation of HIC-MS. The area under the peak of each chromatogram in all runs was plotted against the corresponding Lys concentration. Overlaid chromatograms exhibited consistent retention time (**Figure 4.3b**), and the coefficient of determination (R^2) of the working curve in regression analysis, 99.9%, which demonstrated that HIC-MS can be used for protein quantitation (**Figure 4.3c**).

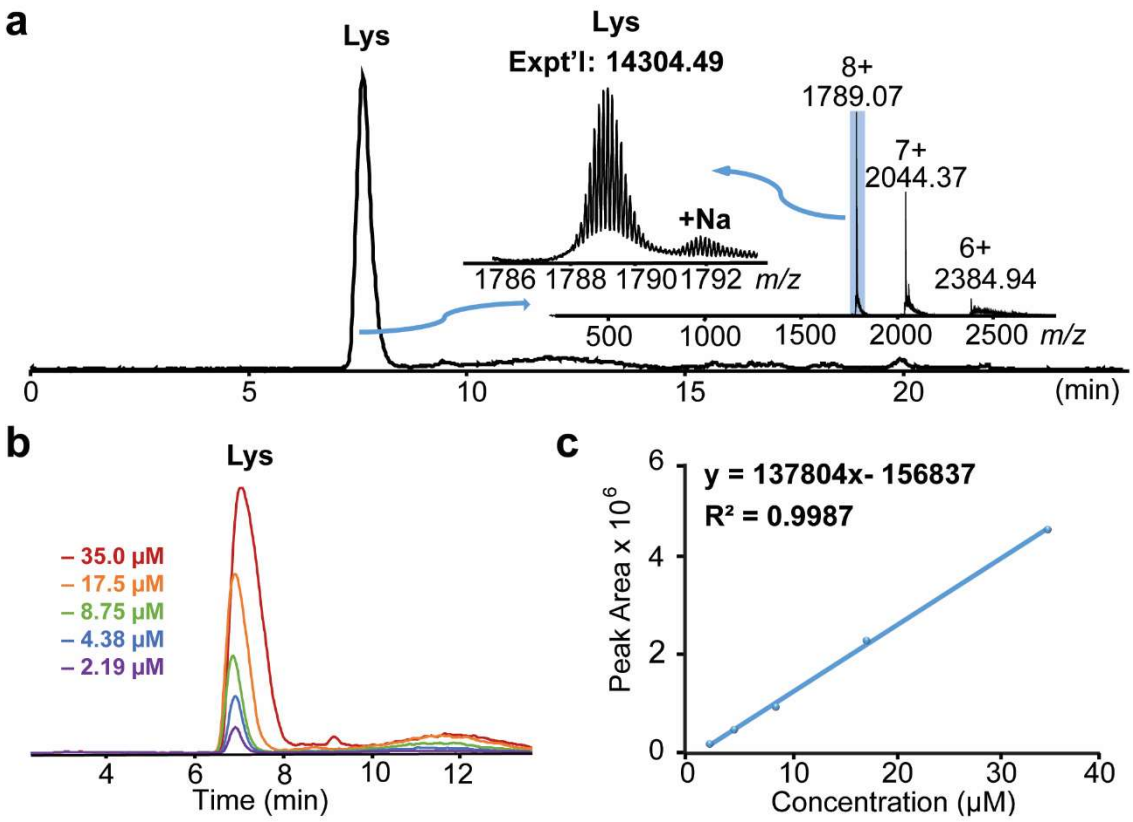


Figure 4.3. Feasibility of online HIC-MS of a single protein for qualitative and quantative analysis. (a) Base peak chromatogram of 2 µL of 0.06 mg/mL Lys; (b) overlaid HIC-MS chromatograms of serially diluted Lys in different concentrations and (c) working curve of serially diluted Lys with a PolyHEPTYL A capillary running from 1 M NH_4OAc to 20 mM NH_4OAc with 50% ACN in a 15 min gradient followed by 20 mM NH_4OAc with 50% ACN for 5 min at 3 µL/min. Insert in (a): Mass spectrum of Lys and the zoom-in spectrum of the highlighted charge state.

The practicality of this online HIC-MS method was further established with a mixture of six protein standards: RiA, Apr, Trp, Lys, ChA, and Lac. The six-protein standards were well separated in clean peaks with a PolyHEPTYL A capillary (**Figure 4.4**). Comparison of the overlaid HIC-MS chromatograms of individual standard proteins (**Figure 4.4a**) and the HIC-MS chromatogram of the six-protein mixture (**Figure 4.4b**) confirmed the identities of the peaks as well as the consistency of their retention times. The mass spectrum of each protein possessed a distinctive low charge state pattern with minimal adduction, typical of native MS.[43,286] They were

deconvoluted to calculate the theoretical intact masses as shown (**Figure 4.4c**). A comparison of PolyPENTYL A, PolyHEXYL A, and PolyHEPTYL A using the same six-protein mixture confirmed the earlier observation that the ligand length of the new HIC materials does not affect retention time significantly for most of these standards, but Lac proved to be an exception (**Figure S4.5**).

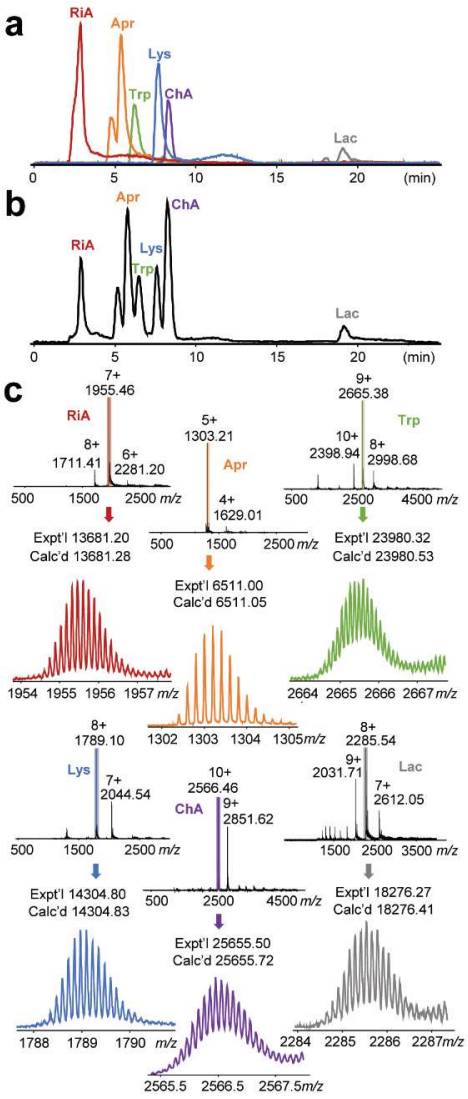


Figure 4.4. HIC-MS run of a six-protein mixture on PolyHEPTYL A at 2.4 µL/min; gradient as described in Figure 4.3. (a) Overlaid HIC-MS base peak chromatograms of individual proteins. (b) Base peak chromatogram of a six-protein mixture. (c) Charge state pattern and isotopic resolution of each protein on a chromatographic time scale.

Notably, the standard Apr eluted in two separate peaks from PolyHEPTYL A. The earlier-eluting one was likely to be an oxidized form of Apr because the species from the two peaks had a 15.98 Da mass difference after deconvolution (**Figure 4.5**). This well exemplified that HIC is sensitive to small conformational variations.[287,288] While the HIC-MS data confirm the significantly later elution of Cyt from PolyHEPTYL A than from PolyPENTYL A, the mass spectra exhibited the same charge state distribution (**Figure S4.4**). This suggests that Cyt elutes from both materials predominantly in its native conformation. In addition, we have demonstrated orthogonal selectivity between HIC-MS and RPC-MS using the six-protein mixture. The eluting order from RPC-MS (Apr, Lys, RiA, Trp, ChA, Lac) differs from that we observed in HIC-MS (RiA, Apr, Trp, Lys, ChA, Lac) (**Figure S4.6**).

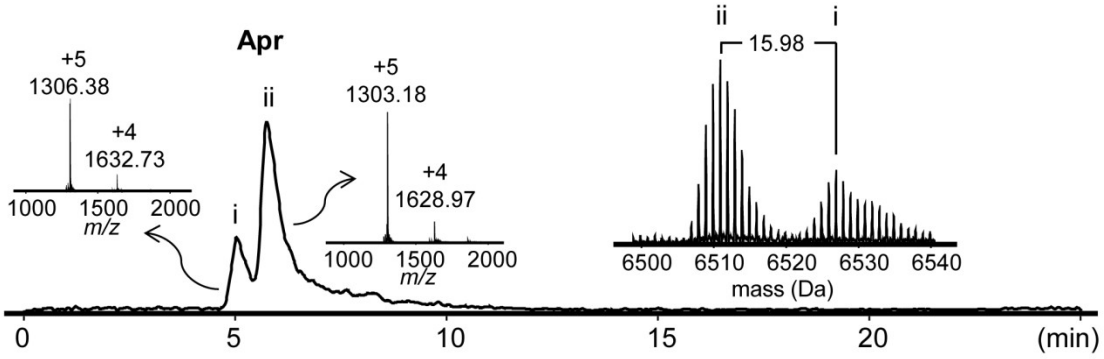


Figure 4.5. HIC-MS of Apr with modified forms. (a) Base peak chromatogram of Apr on PolyHEPTYL A with the same chromatographic conditions as described in Figure 4.4. Insert: mass spectra of two peaks (i and ii) of Apr and deconvoluted spectrum showing a 15.98 Da difference.

4.4.4 Online HIC-MS on *E.coli* lysate

We further evaluated online HIC-MS with a complex *E.coli* cell lysate. Figure 4.6a shows the ten-fold magnified total ion chromatogram (TIC) of HIC-MS of the lysate, in which the most hydrophilic proteins eluted immediately in a broad peak (2.5 - 5 min). The overall pattern of the TIC resembles the HIC UV chromatogram of *E. coli* lysate using ammonium tartrate and a PolyPROPYL A column as published previously.[281] Representative mass spectra (**Figure 4.6 b-e**) display proteins in lower and fewer charge states with color-coded charge-state clusters in contrast with denaturing conditions. Using only a single dimension of LC separation, co-elution of some proteins is inevitable (**Figure 4.6c and d**) and, therefore, fractionation and multi-dimensional separation are often employed.[233] Here, for proof-of-concept purposes, we used only one-dimension HIC-MS to demonstrate the online feasibility even with a complex cell lysate. Moreover, since the spectra contain fewer charge states than under denaturing conditions, the MS analysis becomes less complex. For larger size proteins in denaturing conditions, higher charge states fall into a low *m/z* range, in which the distance between each charge state becomes narrow. This hinders deconvolution of the charge states at low intensity and with limited resolution. In contrast, non-denaturing conditions typically result in lower charge distribution of protein ions and a shift to a higher *m/z* range, which enables easier assignment of the charge states for deconvolution (**Figure 6.6c**). Representative data for proteins ranging from 7.3 kDa to 206 kDa are shown in the insertions of Figure 4.6 b-e.

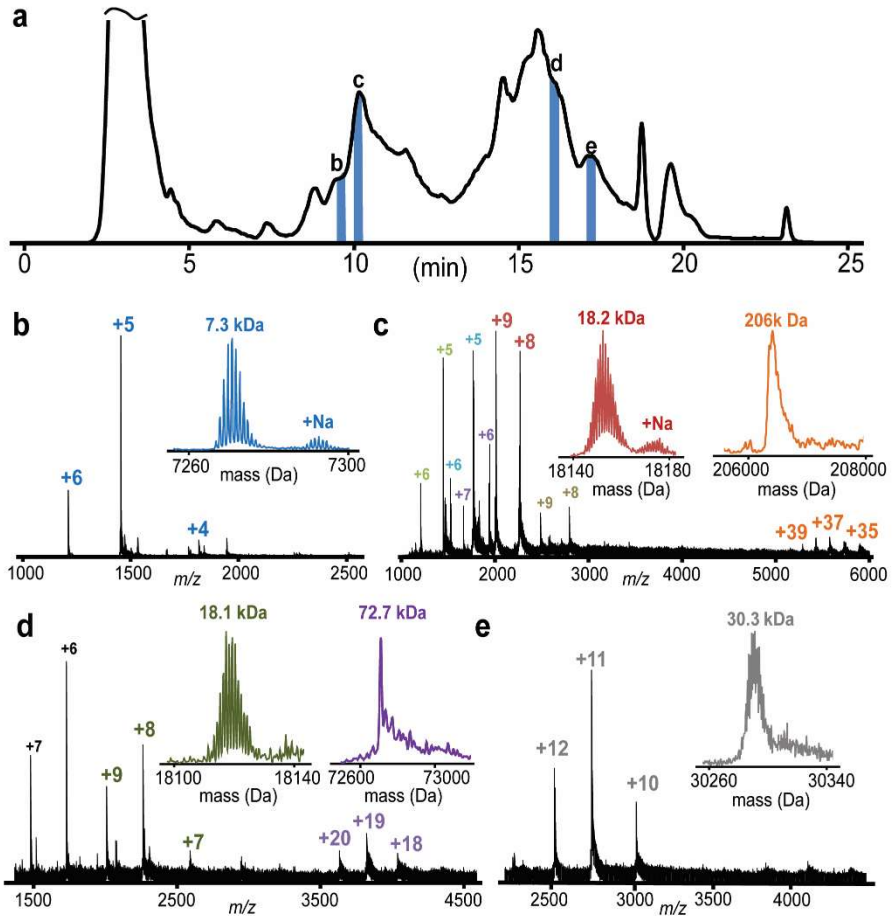


Figure 4.6. HIC-MS run of *E. coli* cell lysate sample on PolyHEXYL A. Representative mass spectra are shown. (a) TIC (ten-fold zoom-in) of *E. coli* lysate by HIC separation with conditions as described in Figure 4.4. Average spectra and deconvoluted spectra from respective color-coded charge states from (b) 9.4 – 9.8 min, (c) 10.0 – 10.4 min, (d) 16.0 – 16.2 min, (e) 17.0 – 17.3 min.

4.5 Discussion

In this study, we have synthesized a new series of unusually hydrophobic HIC materials and demonstrated their successful online implementation of HIC-MS with NH_4OAc as the salt. Moreover, we have demonstrated the qualitative and quantitative potential of this HIC-MS method. HIC was known to be an orthogonal method that complements other separation techniques,[233,281]

and we have also demonstrated the orthogonal selectivity between the new HIC materials and RPC (**Figure S4.6**). The unique advantages of conventional HIC, including its non-denaturing character and remarkable sensitivity to structure variance, were retained online with the new HIC materials (**Figure 4.5**), expanding the online LC-MS toolbox for top-down proteomics and providing additional multi-dimensional options. The mass spectra of HIC-MS feature low charge state distribution, implying retention of native protein structures under these conditions. As demonstrated, this HIC-MS method provides alternative high-resolution approach with minimal denaturation for cases where non-denaturing condition is preferred or cases where proteins cannot be eluted from RPC.

HIC is generally considered to be a non-denaturing mode of chromatography. However, a search of the literature discloses that this property is not absolute.[284,289] The degree of protein denaturation is directly proportional to the time of contact with the hydrophobic stationary phase and inversely proportional to the starting concentration of salt in the mobile phase.[289] The same observations pertain to proteins on reversed-phase surfaces.[284] Success in HIC-MS, then, requires separation and elution of the proteins faster than the kinetics of denaturation using a volatile solvent. That seems to have been accomplished with these new hydrophobic HIC materials, provided that any solvent used for a gradient from starting conditions be one that suffices for elution. That probably means one that contains some organic solvent, since a gradient to a low-salt buffer without organic solvent might lead to denaturation of proteins caused by prolonged contact with the hydrophobic ligands without the structure-protecting effect of the high initial salt concentration. As a result, proteins can fail to elute (**Figure 4.2a**). The PolyPENTYL A, PolyHEXYL A, and PolyHEPTYL A materials introduced here appear to afford kinetics of elution that are faster than those of denaturation for most proteins using MS-compatible mobile phases.

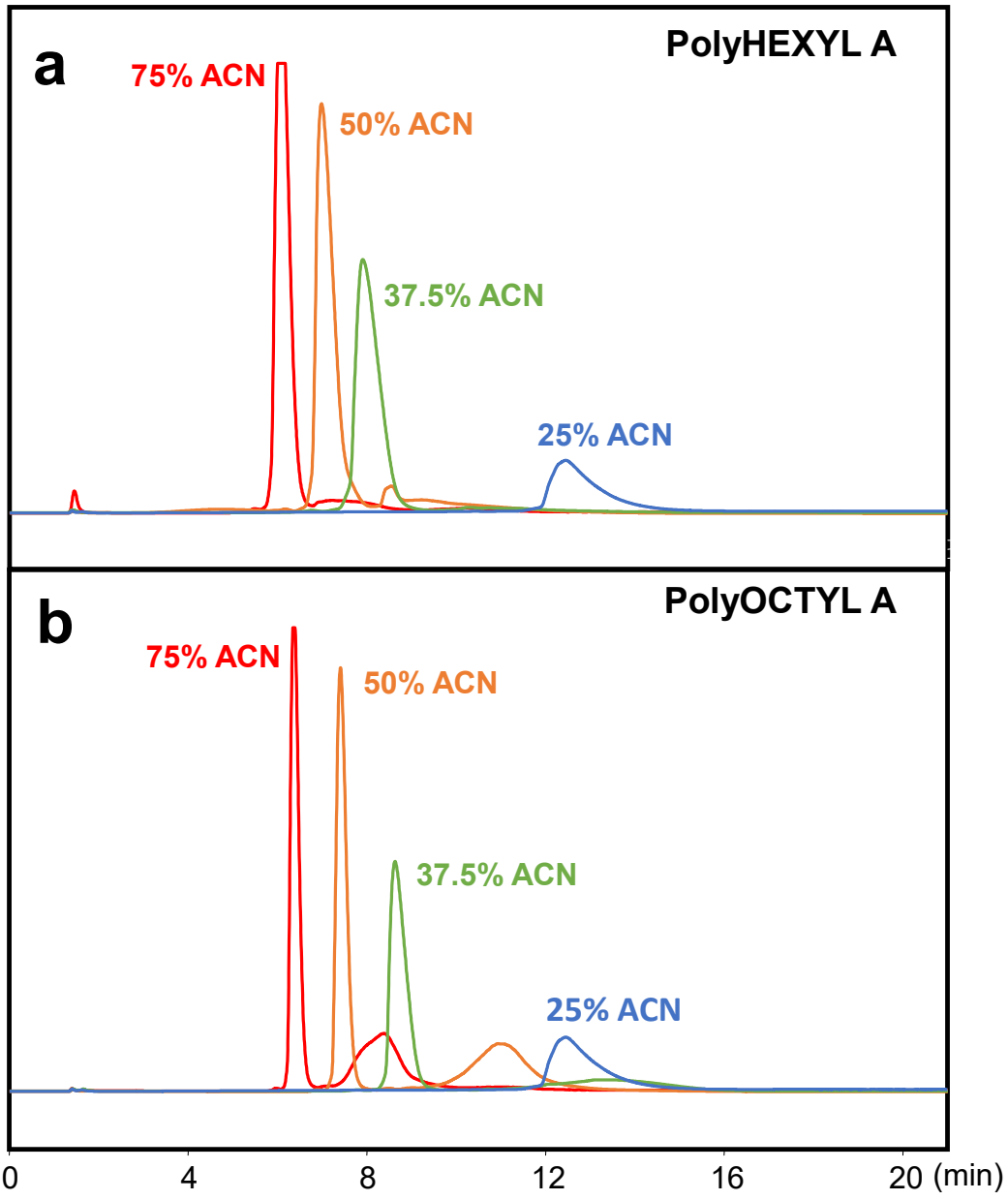
The involvement of high salt concentration also seems to further protect the tertiary structure from denaturation, but the ability of the salt to promote retention and protect structure do not necessarily correlate.[284] In effect, at least from the observed mass spectra, proteins were in low charge states indicating predominantly native structures (**Figure 4.4 and Figure 4.6**).

With the MS-compatible salt NH_4OAc and the presence of organic solvent in the final mobile phase, the operating mechanisms of the new materials remain incompletely understood and need further investigation. In conventional HIC, retention is strongly affected by ligand length and salt concentration. That was true here of Cyt and Lac but much less so for the other protein standards with the materials more hydrophobic than PolyBUTYL A, a behavior more characteristic of RPC. An interesting calorimetric study by Lin *et al*. suggests that retention of some protein standards by CM-butyl-Sepharose in 4 M NaCl is an adsorption process while retention on CM-octyl-Sepharose involves both adsorption and partitioning.[290] Lin *et al*. saw the same trend with an increase in ligand density.[291] An observation more relevant to this study is that retention of isomorphs of apolipoprotein E were the same on columns with ethyl- and propyl- ligands.[280] These apolipoproteins lack a well-defined tertiary structure, and so any hydrophobic domains in the proteins that are accessible to a propyl- ligand are also available to an ethyl- ligand. With proteins that do have well-defined tertiary structures, access to sequestered hydrophobic domains seems to increase with ligand length up to the pentyl- ligand. At that point, every hydrophobic domain of some proteins is accessible and so retention no longer increases significantly with the length of the ligands. This provides a possible explanation for the phenomenon we observed above. In the absence of a more detailed study of the mechanism, we tentatively call our conditions a HIC-RPC hybrid.

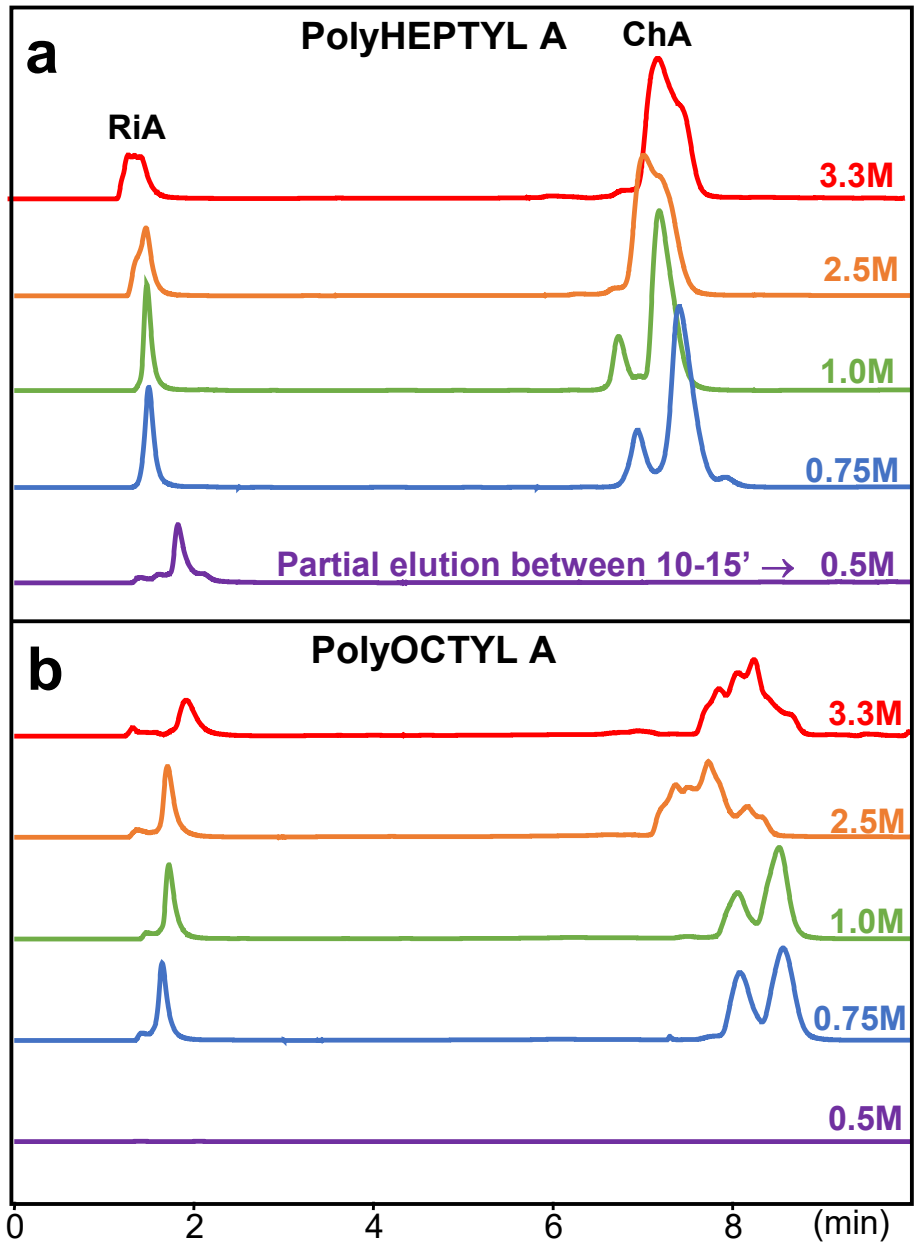
4.6 Conclusions

We have reported here the development of new HIC materials for online HIC-MS analysis of intact proteins. For the first time HIC and MS were directly coupled in an online fashion using volatile NH_4OAc as the salt in the mobile phase. The standards examined eluted cleanly in well-shaped peaks, and the mass spectra of the proteins eluted from HIC resembled those of native mass spectra, implying minimal denaturation. Furthermore, we have demonstrated the use of HIC-MS for online separation and analysis of a complex cell lysate. This new HIC has some aspects of both RPC and conventional HIC: Although it requires some organic solvent for elution, the salt in the mobile phase minimizes denaturation of proteins. Rather than promoting retention, the function of the salt primarily serves to protect the tertiary structure of the protein. Implementation of this method allows separation of proteins under generally non-denaturing conditions and adds options for multidimensional LC. This LC-MS combination could potentially be applied to the study of antibody-drug conjugates and larger macromolecular protein complexes in the future.

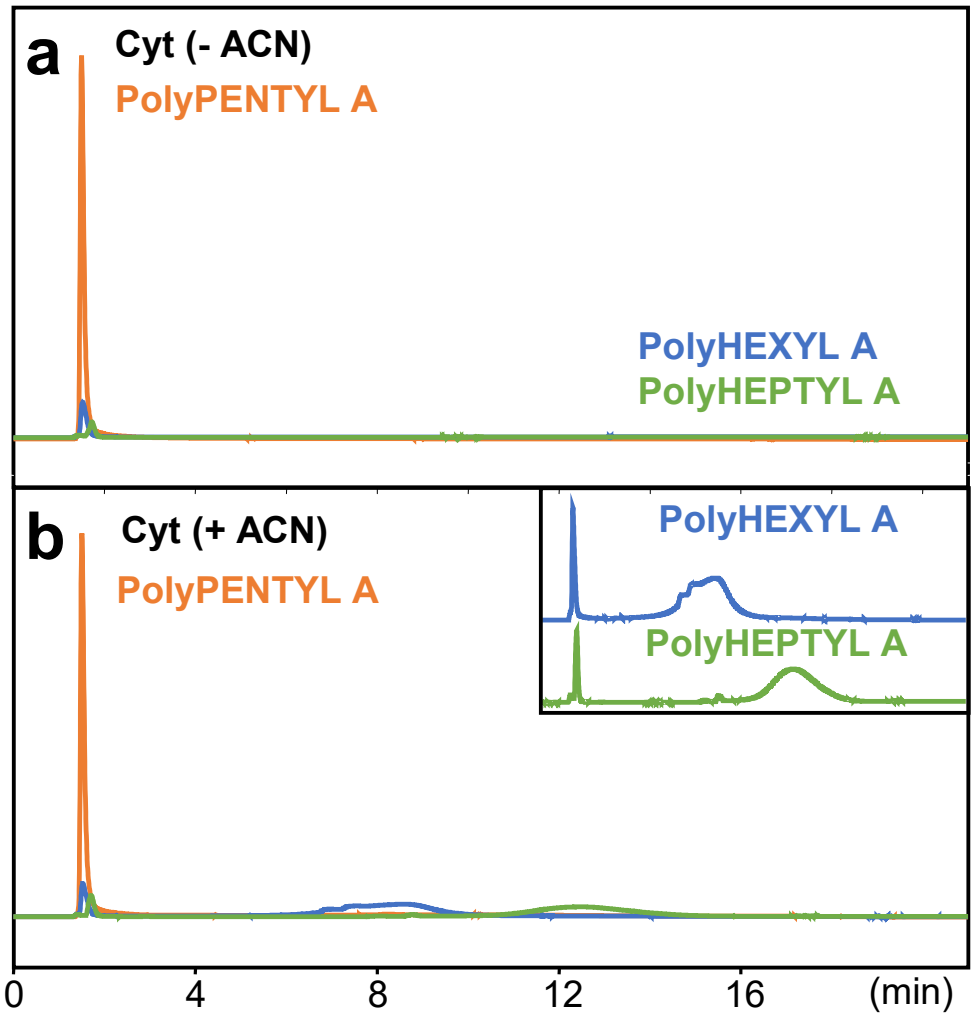
4.7 Supplemental Information



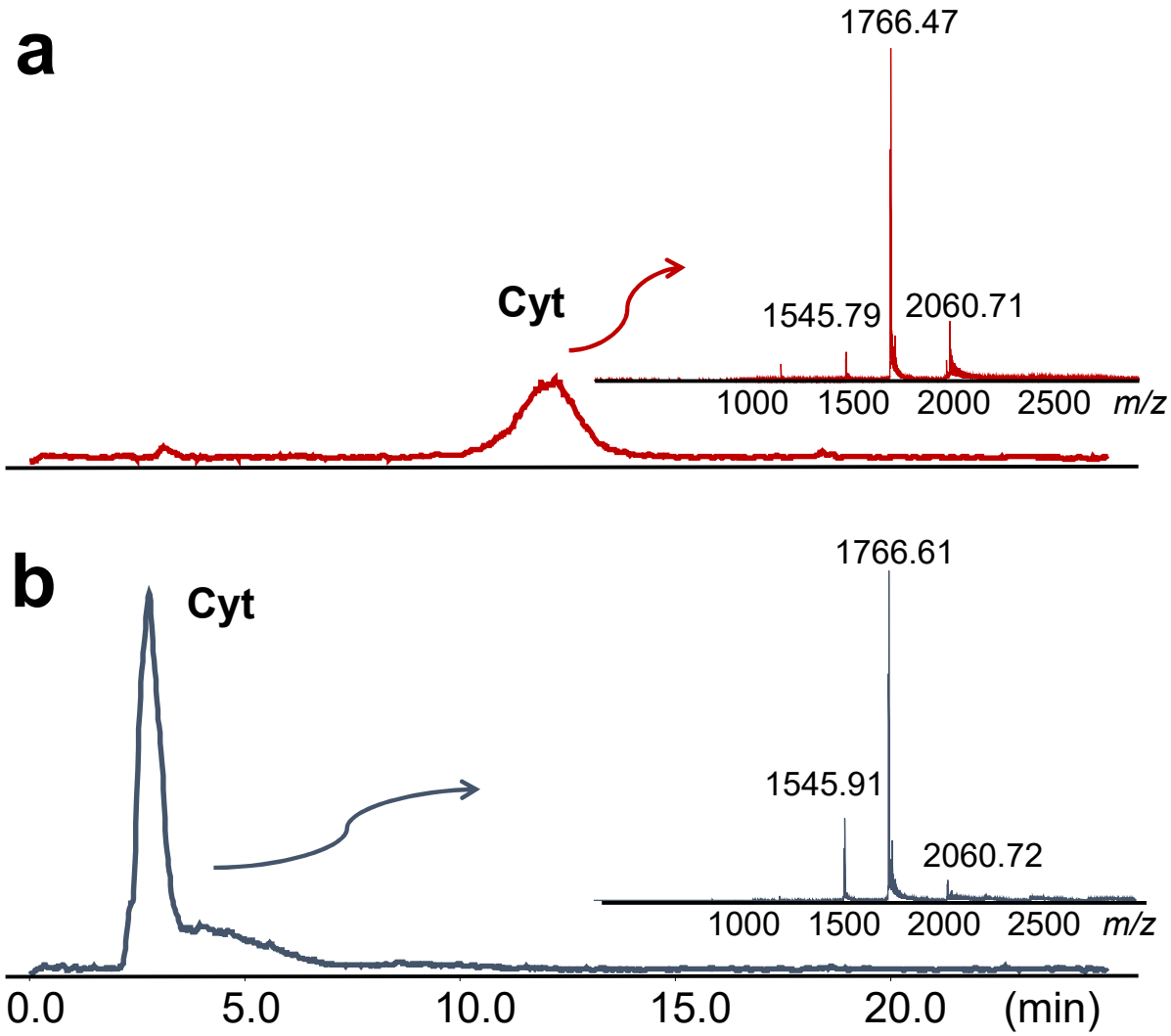
Supplemental Figure 4.1. Effect of ACN percentage in mobile phase B on retention of Lys. Overlaid chromatograms of Lys on (a) PolyHEXYL A and (b) PolyOCTYL A running from 1 M NH_4OAc to 20 mM NH_4OAc containing the indicated percentage of ACN over a 15 min gradient and remaining at the final mobile phase for 5 min. Detection: 280 nm.



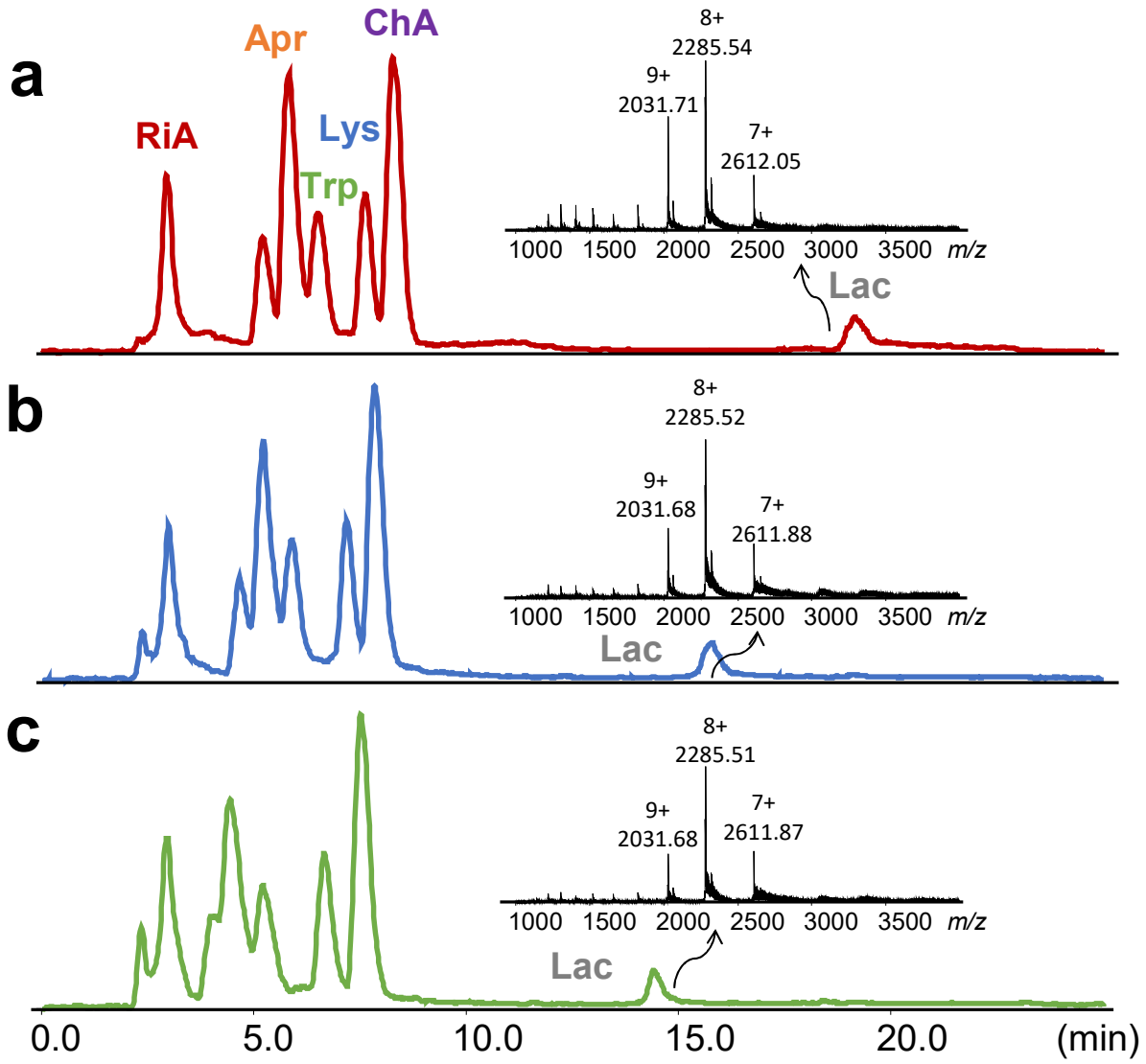
Supplemental Figure 4.2. Effect of initial salt concentration on the chromatography. Chromatograms of a RiA and ChA mixture on (a) PolyHEPTYL A and (b) PolyOCTYL A running from the indicated concentration of NH_4OAc to 20 mM NH_4OAc with 50% ACN over a 15 min gradient and remaining at the final mobile phase for 5 min. Detection: 280 nm.



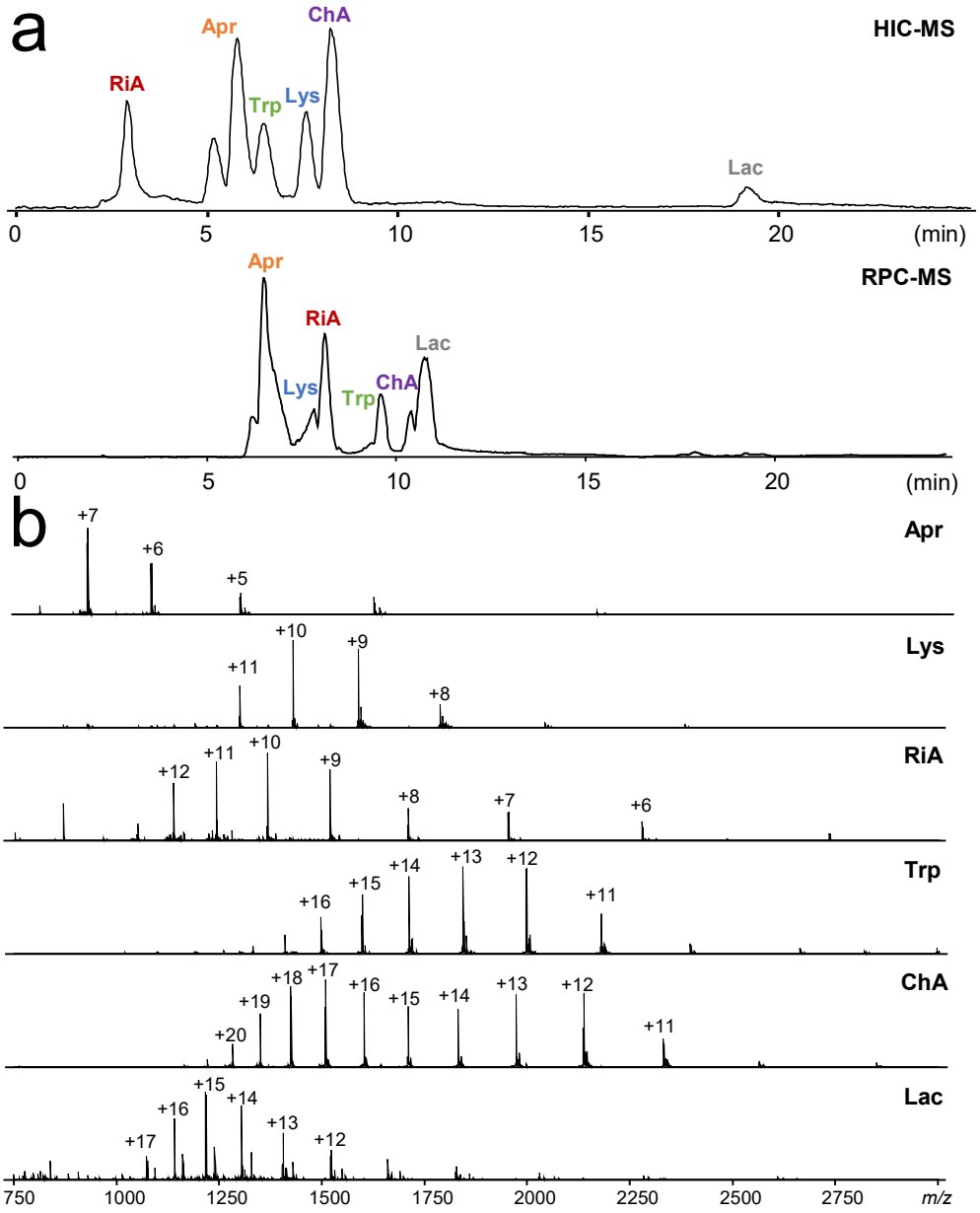
Supplemental Figure 4.3. Overlaid chromatograms of Cyt on new HIC columns PolyPENTYL A, PolyHEXYL A, and PolyHEPTYL A. (a) Running from 1 M NH_4OAc to 20 mM NH_4OAc (no ACN) in a 10 min gradient and remaining at 20 mM NH_4OAc for 5 min. (b) Same conditions, but the final mobile phase contains 37.5% ACN. Detection: 280 nm.



Supplemental Figure 4.4. Base peak chromatogram of Cyt on PolyHEPTYL A (a) and PolyPENTYL A (b). Insert: mass spectra of the labeled peaks. Both mass spectra show the same charge state pattern, implying that the tertiary structure has been retained in both cases.



Supplemental Figure 4.5. Base peak chromatogram of six-protein mixture on (a) PolyHEPTYL A, (b) PolyHEXYL A, and (c) PolyPENTYL A with the same chromatographic conditions as described in Figure 4.5. Retention does not improve significantly going from PolyPENTYL A to PolyHEPTYL A except for the strongly retained Lac. Insert: mass spectrum of Lac.



Supplemental Figure 4.6. Orthogonality between HIC-MS an RPC-MS. (a) Base peak chromatograms of the six-protein mixture on a (top) PolyHEPTYL A (Figure 4.4) and on a (bottom) home-packed PLRP-S (8 µm, 4000 Å) column at 2 µL/min running from 95% MPA to 5% MPA in a 15 min linear gradient. MPA: H_2O with 0.1% formic acid; MPB: ACN with 0.1% formic acid. Column length was 15 cm and i.d. was 150 µm. (b) From top to bottom, mass spectrum of each peak corresponding to the eluting order in RPC-MS. The base peak chromatogram of the mixture and the mass spectra confirmed that the eluting order of the six-protein mixture (Apr, Lys, RiA, Trp, ChA, and Lac) for reverse phase chromatography is different from that we observed for HIC-MS (RiA, Apr, Trp, Lys, ChA, and Lac). Therefore, we have shown orthogonal selectivity between RPC and the new HIC.

4.8 Acknowledgements

We thank Zachery Gregorich, Ziqing Lin, Tania Guardado, and Zhijie Wu for critical reading of the manuscript, Deyang Yu for growing *E. coli*, and Nathan Miklos for technical assistance with the chromatography. We also thank Jeremy Wolff, Lori Ann Upton, Allan Martinez, Mark Marispini, and Vladimir Ondruska at Bruker for technical assistance. We would like to acknowledge NIH grants, R01HL096971, R01HL109810, R01GM11705 and S10 OD01847 (to Y.G.).

Chapter 5: Online Hydrophobic Interaction Chromatography-Mass Spectrometry for Intact Monoclonal Antibodies

This chapter has been published and is adapted from:

Chen, B.; Lin, Z.; Alpert, AJ.; Fu, C.; Zhang, Q.; Wayne, P.; Ge, Y. Online Hydrophobic Interaction Chromatography-Mass Spectrometry for the Analysis of Intact Monoclonal Antibodies. *Analytical Chemistry.* 2018, 90, 7135-7138.

5.1 Abstract

Therapeutic monoclonal antibodies (mAbs) are an important class of drugs for a wide spectrum of human diseases. Liquid chromatography (LC) coupled to mass spectrometry (MS) is one of techniques in the forefront for comprehensive characterization of analytical attributes of mAbs. Among various protein chromatography modes, hydrophobic interaction chromatography (HIC) is a popular offline non-denaturing separation technique utilized to purify and analyze mAbs, typically with the use of non-MS-compatible mobile phases. Herein we demonstrate for the first time, the application of direct HIC-MS and HIC-tandem MS (MS/MS) with electron capture dissociation (ECD) for analyzing intact mAbs on Q-TOF and FT-ICR mass spectrometers, respectively. Our method allows for rapid determination of relative hydrophobicity, intact masses, and glycosylation profiles of mAbs, as well as sequence and structural characterization of the complementarity-determining regions in an online configuration.

5.2 Introduction

Since the first approval of monoclonal antibody drug in 1986, therapeutic monoclonal antibodies (mAbs) and their derivatives have arguably become the most important and promising class of therapeutics for many human diseases such as cancer, autoimmunity, metabolic disorders, and infection.[292-294] The unique pharmacological advantages of mAbs (e.g. target specificity, selectivity, long half-life, and excellent safety profile) and the evolving protein engineering (bispecific antibody, fusion protein, antibody-drug conjugate and nanobody) continuously propel the development of new mAb-based therapeutics, resulting in hundreds of candidates in clinical trials.[295] However, with complex conformational and structural dynamics, large molecular sizes, and micro-heterogeneity due to various post-translational modifications, mAbs and their derivatives are among the most complex biologics ever produced, imposing tremendous analytical challenges for comprehensive characterization and control to ensure product quality.[296,297]

Liquid chromatography (LC) and mass spectrometry (MS) play essential roles, and are the dominant techniques for characterizing mAb-based therapeutics.[297,298] Reversed-phase chromatography (RPC), size exclusion chromatography (SEC), hydrophobic interaction chromatography (HIC), and affinity chromatography are often employed for assessing heterogeneity and impurity; top-down, middle-down, bottom-up, and native MS have been used to identify and characterize primary sequences and study higher order structures.[297-299] Coupling LC and MS in an online fashion (LC-MS) not only integrates the advantages from separation and detection of these techniques,[287,299,300] but also simplifies the workflow, eliminating offline-handling steps.

Recently we reported the first online coupling of HIC to MS by using columns of greater hydrophobic character and ammonium acetate with some organic solvent in the eluting mobile

phase for top-down proteomics.[8,9,67,301] Importantly, HIC remains the only non-denaturing separation method that exploits hydrophobicity of the native analytes in their native structures, and has been adopted as a powerful approach to analyze mAbs.[279,287] Therefore, online coupling of HIC to native MS holds great promise for determining relative hydrophobicity, assessing heterogeneity, and sequence and structural characterization of mAbs when applying tandem MS (MS/MS) techniques. Herein, for the first time, we demonstrate online HIC-MS analysis of intact mAbs using Q-TOF and FT-ICR mass spectrometers, revealing variation in hydrophobicity and glycosylation of proteoforms of the mAbs. Non-covalent interactions that produce higher order dimers and trimers were preserved and detected upon elution condition. Moreover, we have demonstrated that online HIC-MS/MS with electron capture dissociation (ECD) enabled sequence and structural characterization of mAbs, particularly on the complementarity-determining regions (CDRs) outside of the confinement of disulfide bridges.

5.3 Material and Methods

5.3.1 Reagents

HPLC grade water, acetonitrile (ACN), and NH_4OAc were purchased from Fisher Scientific (Fair Lawn, NJ, USA).

5.3.2 Sample preparation

mAb1 is a model antibody AB004 in 7 mg/mL (Abbvie), and mAb2 is IgG1κ purchased from NIST (Reference material 8671) in 10 mg/mL. A mixture was prepared containing equal amount

of mAb1 and mAb2. Subsequent dilution using 1 M NH_4OAc (at least 3:1 volume ratio) brought the final concentration to 0.5-1 mg/mL for a single mAb as well as for the mixture unless noted otherwise. Deglycosylation was performed using IgGZERO (EndoS) (GENOVIS). 1 unit of enzymes and 1 µg of mAbs were incubated in 15 mM Tris buffer for 30 min incubation time at 37 °C.

5.3.3 HIC-MS and -MS/MS

100 mm × 0.2 mm i.d. PolyPENTYL A capillaries (PolyLC, Columbia, MD, USA) with 1.9 µm particle were used on a Waters ACQUITY UPLC M-class system (Milford, MA, USA). Mobile phase A (MPA) contained 1M NH_4OAc and mobile phase B (MPB) contained 20 mM NH_4OAc (overall) in 50:50 water/ACN without any adjustment of pH. For Q-TOF MS experiments, a total of 45 min run consisting of a 30 min linear gradient (0 min 99% MPA; 5 min 99% MPA; 35 min 1% MPA; 40 min 1% MPA; 42 min 99% MPA; 45 min 99% MPA) was used at a flow rate of 2.6 µL/min. Later on for FT-ICR MS experiments, a faster 30 min run consisting of a 20 min linear gradient (0 min 99% MPA; 2 min 99% MPA; 22 min 1% MPA; 27 min 1% MPA; 28 min 99% MPA; 30 min 99% MPA) was used at a flow rate of 2.6 µL/min. Samples eluted from HIC columns were electrosprayed into a maXis II ETD Q-TOF or a 12 T solariX FT-ICR mass spectrometer (Bruker Daltonics, Bremen, Germany). For maXis II ETD, the end plate offset and capillary voltage were optimized at 500 V and 3800 V, respectively. The nebulizer was set to 1.5 bar, and the dry gas flow rate was 5.0 L/min at 150 °C. In-source CID was set to 165-170 eV, collision cell RF to 4000 V_{pp}, transfer time to 180 µs, and pre pulse storage to 60 µs to improve desolvation and transmission of large ions. The quadruple low mass was set to 1600 *m/z*. For solariX, dry gas flow rate was set to 3 L/min at 106 °C. Skimmer 1 was set to 120 V and 25 V collision energy was set

for the quadrupole to improve desolvation. Lowest frequency values for octopole (2 MHz), quadrupole (1.4 MHz), and transfer hexapole (1 MHz) were used to improve ion transition. Time of flight was set to 2 ms. Ions were accumulated for 1.3 s and 1 M words of data were collected. For ECD parameters, pulse length, ECD bias, and ECD lens were set to 50 ms, 0.6 V, and 15 V, respectively. Mass range was set to 200 or 500 -10,000 *m/z* calibrated by 10 mM cesium iodide salt clusters.

5.3.4 Data Analysis

All data were processed and analyzed by Compass DataAnalysis 4.3, 4.4, and MASH Suite Pro. Chromatograms and spectra were smoothed. Maximum Entropy was used for spectrum deconvolution with resolution was set to 10,000. SNAP algorithm were used for peak picking, in which quality threshold was set to 0.3. All fragment ions were manually validated. The crystal structure of the Fab fragment of mAb2 (5K8A) was obtained and modified using Protein Workshop from RCSB protein data bank.

5.4 Results and Discussion

We first applied HIC-MS on a Q-TOF mass spectrometer that has extended *m/z* range. A mixture of equal amounts of two different immunoglobulin G 1's (IgG1), namely mAb1 and mAb2 (refer to Experimental procedures in SI), are well separated in the HIC-MS experiments, in which a linear gradient from 99% MPA (1M ammonium acetate) to 99% MPB (20 mM ammonium acetate in 50% ACN) was applied. mAb2 appears to be relatively more hydrophobic than mAb1,

and elutes in a sharper peak in the total ion chromatogram (**Figure 5.1a**). The two mAb peaks were calculated to have a roughly 1:1 area under the curve (AUC), indicating that they had similar ionization efficiency between the two mAbs (**Table S5.1**). Additionally, using mAb2 as an example, the AUCs showed good linear response as a function of injection amounts (**Figure S5.1**). mAb2 can be detected with the injection amount as low as 50 ng on a 10 cm, 200 μm i.d. PolyPENTYL A capillary column (**Figure S5.1**).

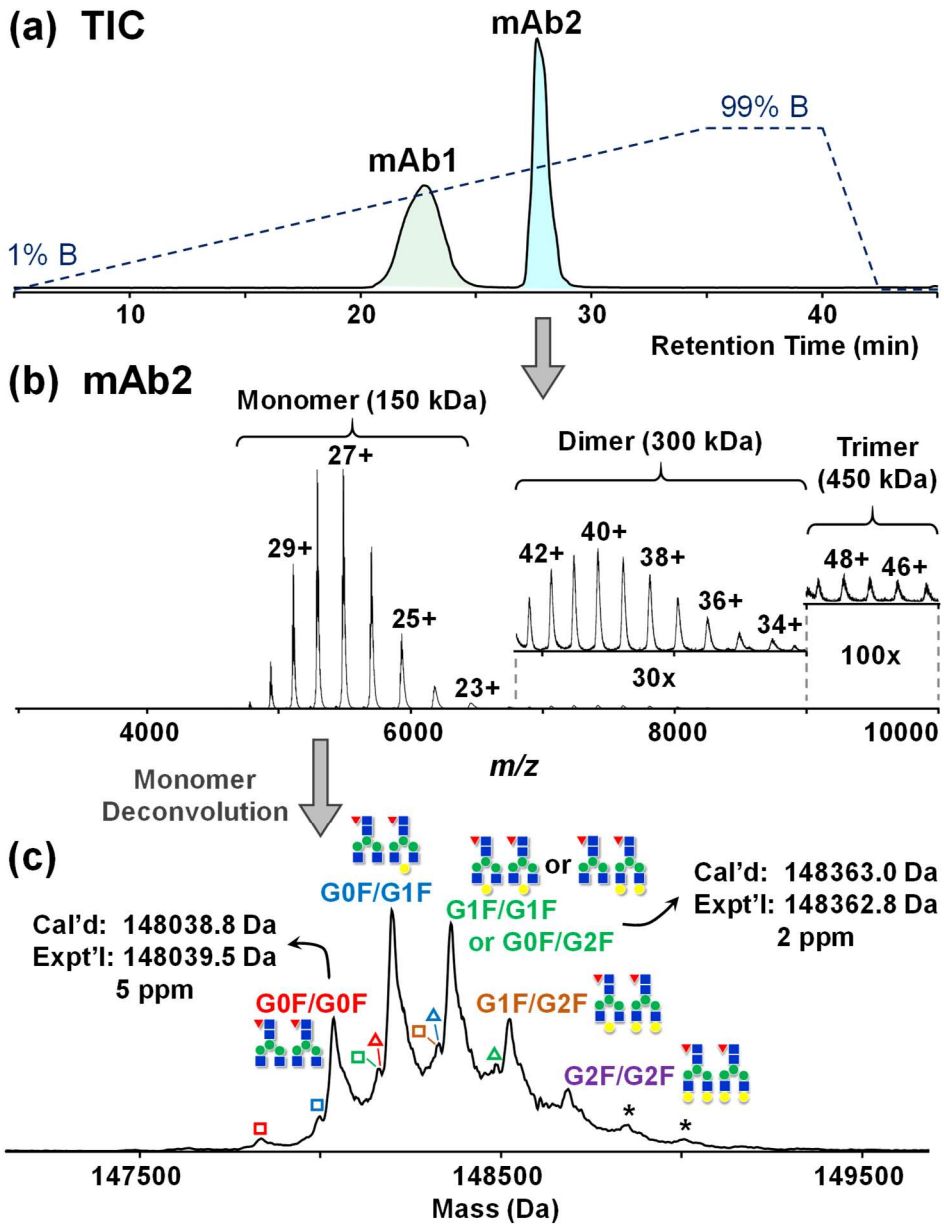


Figure 5.1. Online HIC-MS of mAb mixtures on a maXis II Q-TOF mass spectrometer. (a) Total ion chromatogram demonstrating the separation of mAb1 and mAb2. The dash line indicates the gradient changes of mobile phase B. (b) Mass spectrum of mAb2 showing the detection of monomers, dimers (30x zoom-in), and trimers (100x zoom-in). (c) Deconvoluted mass spectrum of mAb2 monomer with annotated glycosylation forms (red triangle: Fucose; blue square: GlcNAc; green circle: Mannose; yellow circle: Galactose); hollow square represents the loss of one GlcNAc (-203 Da), hollow triangle represents the preservation of C-terminal Lys on heavy chain (+128 Da) and asterisk * represents the addition of a hexose (+162 Da). GxF indicates Fc-oligosaccharides terminated by x number of galactoses.

The average mass spectrum from the Q-TOF mass spectrometer further demonstrates that the low-charge-state envelopes of both mAbs fell into the 5,000 to 7,000 *m/*z range (**Figure 5.1b and Figure S5.2**), which accords with the characteristics of native MS as shown in the previous studies[302,303]. The deconvoluted spectra further reveal heterogeneity due to different glycosylations (**Figure 5.1c and Figure S5.2**).[304] For instance, multiple addition of 162 Da were observed on mAb2, corresponding to mAb2 with glycan structure G0F/G0F, G0F/G1F, G1F/G1F or G0F/G2F, G1F/G2F, G2F/G2F, with G0F/G1F being the most abundant proteoform (**Figure 5.1c**), consistent with previous studies.[304,305] The average masses are in good agreement with the theoretical values within 5 ppm of the identified proteoforms, except for the low abundance G2F/G2F proteoform (**Figure 5.1c and Table S5.2**). Minor proteoforms with 203 Da truncation and 128 Da addition were also detected, which correspond to the removal of one terminal GlcNAc and the preservation of C-terminal lysine on one of the heavy chains, respectively.[304] On the other hand, mAb1 exhibited a significantly different glycosylation profile from those of mAb2 (**Figure S5.2 and Figure 5.1c**), most likely due to different manufacturing procedures. More interestingly, under the HIC-MS conditions, we also observed high-molecular weight aggregates for both mAbs.[306] In particular, dimers (300 kDa) and trimers (450 kDa) of mAb2 were detected within the same elution time frame as the monomers (150 kDa) (**Figure 5.1b**). The presence of high molecular-weight dimers and trimers is consistent with the SEC-UV analysis from the mAb2 investigation report (~3.2%),[304] consisting of about 2.6 % of the total amount from our HIC-MS result based on the peak areas of the deconvoluted spectrum. Dimers and trimers of mAb2 were eluted in about 33% ACN and 370 mM ammonium acetate, under which the rather strong non-covalent interaction was preserved. We speculated that the kinetics of the chromatography was faster than the kinetics of the denaturation under the HIC-MS conditions used here, and the high

concentration of ammonium acetate salt seemed to prevent excessive exposure to ACN, which is consistent with our previous observation.[67]

To further apply the HIC-MS method to gain more sequence and structural information, we coupled HIC online to a 12T FT-ICR mass spectrometer to analyze mAb2 with a faster gradient. Deglycosylation was performed with an endoglycosidase IgGZERO to reduce sample heterogeneity. Removal of glycosylation also increased mAb2 hydrophobicity, leading to slightly longer retention time and the revelation of preserved heavy chain C-terminal lysine (+128 Da) and glycated (+162 Da) proteoforms and their relative intensities (**Figure S5.3**).[304] Without inducing dissociation, the signal of mAb2 can be greatly improved by more than 10-fold after increasing the direct current (DC) bias between the quadrupole and the collision cell in the FT-ICR mass spectrometer (**Figure S5.4**).[303] This observation could result from further desolvation of the small charged droplets during ion transfer by applying additional collisional activation.

We then performed online broadband ECD on mAb2 in the Para cell with the improved signals, resulting in around 80 MS/MS scans on the LC time scale (**Figure 5.2a**). Low electron energy yielded higher signal to noise ratio (*S/N*), as higher electron energy is likely to cause secondary dissociation leading to low *S/N* (**Figure S5.5**). The averaged spectrum from a single experiment with 0.6 eV electron energy and 50 ms irradiation time shows charge-reduced species from 5,000 to 10,000 *m/z*. Fragment ions are readily distinguishable from 1,000 to 3,000 *m/z* (**Figure 5.2a**). The residues of the fragment ions identified within a short 30 min LC-MS/MS run represent 66% of the sequence for the light chain, and 57% for the heavy chain (**Figure S5.6**). However, ECD was inefficient in cleaving disulfide bonds of the 150 kDa mAb2, yielding no fragment ions within disulfide bridges, except $\boldsymbol{c}_{141}$ from the heavy chain.[307,308] Similarly, because of the disulfide linkage between light chain and heavy chain, no $\boldsymbol{z}\bullet$ ions were observed from the

C-terminus of the light chain. Therefore, only part of the CDRs, namely DMIFNFYFDV from the heavy chain (H3) and FQGSGYPFT from the light chain (L3), had cleavages. For the heavy chain, $\boldsymbol{c}_{100}$, $\boldsymbol{c}_{103}$, $\boldsymbol{c}_{104}$, $\boldsymbol{c}_{105}$, $\boldsymbol{c}_{106}$, and $\boldsymbol{c}_{107}$ consisted of the 6 out of 10 possible N-Cα bonds fragmented from H3 (**Figure 5.2b**). The CDR L3 from light chain were well fragmented, where the 8 possible N-Cα bonds were all cleaved ($\boldsymbol{c}_{88}$, $\boldsymbol{c}_{89}$, $\boldsymbol{c}_{90}$, $\boldsymbol{c}_{91}$, $\boldsymbol{c}_{92}$, $\boldsymbol{c}_{94}$, $\boldsymbol{c}_{95}$, and $\boldsymbol{c}_{96}$), except the N-terminal side of proline (**Figure 5.2b**). Representative fragment ions around 10 kDa were detected in low charge states with mass accuracy below 3 ppm (**Figure 5.2b**). ECD has previously been shown to provide tertiary and even quaternary information of native protein and complexes.[309,310] We then compared the ECD fragments obtained from the native-like mAb2 MS spectrum to the known crystal structure of mAb2. The crystal structure of the Fab fragment of mAb2 (PDB 5K8A)[311] shows that all CDR loops are facing outward and well-exposed, and therefore susceptible to ECD fragmentation (**Figure 5.2c**). Indeed, we observed cleavages around the exposed loops of CDRs (H3 and L3) highlighted in red (**Figure 5.2c**). Since ECD preserves non-covalent interactions, antigen binding sites on the H3 and L3 CDRs could potentially be localized. We speculate that ECD also cleaved at other CDRs (H1, H2, L1, and L2), however, because of the disulfide bond linkages, no fragment ions could be observed. To our surprise, ECD yielded no cleavages near the N- and C- termini of both the light chain and heavy chain, which indicated their potential roles in the interaction interface that might have prevented electron migration during the fragmentation event. Overall, 106 bond cleavages were observed from the full mAb2 sequence (**Figure S5.6**).

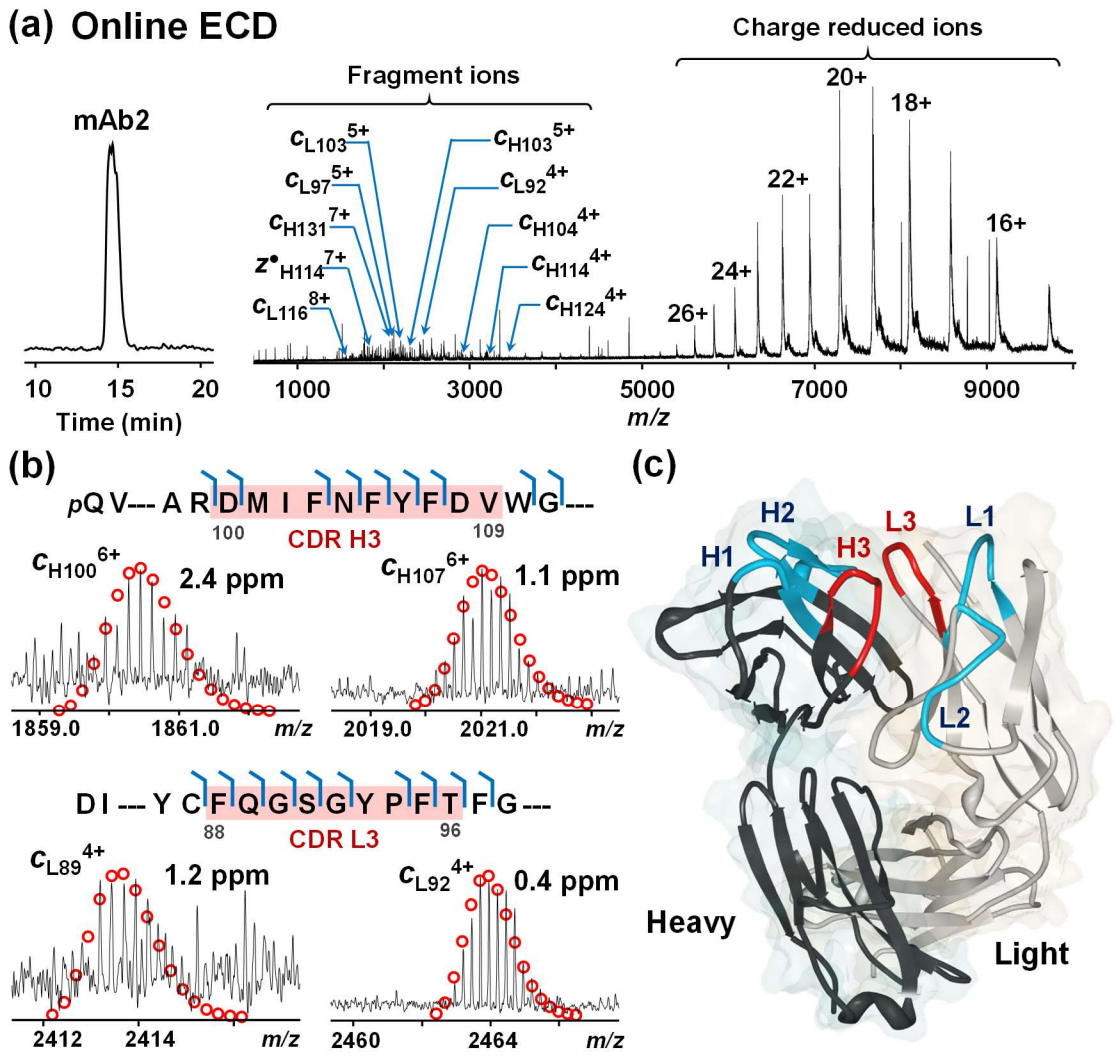


Figure 5.2. Online HIC-MS/MS with broadband ECD of deglycosylated mAb2 on a 12T solariX XR FT-ICR mass spectrometer. (a) Chromatogram and ECD spectrum of mAb2 showing charge reduced species and fragment ions. (b) Representative fragment ions from the CDRs H3 and L3 (highlighted in red) of heavy chain (top) and light chain (bottom). (c) Crystal structure of the Fab fragment of mAb2 (PDB 5K8A). The CDRs fragmented by ECD (H3 and L3) are highlighted in red, and the other CDRs are highlighted in blue (H1, H2, L1, and L2).

With a gradient that includes increasing organic solvent concentration, it is reasonable to question the use of the term HIC. In our previous work, we speculate that this combination of stationary and mobile phase obliterates the boundary between HIC and RPC. Here we propose that the term HIC continue to be applied in its traditional usage, because it describes a mode that

separates proteins based on their hydrophobic character while preserving their three-dimensional structure. The results we have demonstrated above fit in this circumstance.

5.5 Conclusions

To recapitulate, we have demonstrated the use of online HIC-MS for the analysis of mAbs in two different types of mass spectrometers. In a high-throughput online manner, HIC separates mAbs based on their relative hydrophobicity; and MS reveals the intact masses and proteoform heterogeneity. Sequence and structural characterization have been further obtained by applying MS/MS techniques such as ECD. As described above, HIC separates proteins based on hydrophobicity under native conditions. When coupled online with MS, it has high potential and value for efficient separation and comprehensive characterization of complex mAbs. We envision growing interest in this method, and continued improvement of the HIC material will further refine HIC-MS, expanding the toolbox for characterizing mAb-based therapeutics.

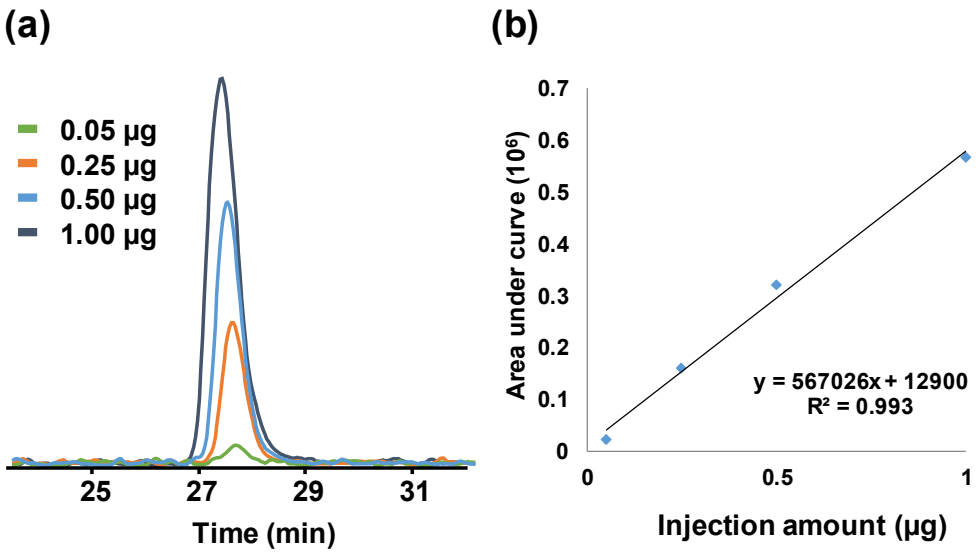
5.6 Supplemental Information

Compound	Retention Time	Area Under Curve (AUC)	Normalized AUC
mAb1	22.8	218910352	0.994
mAb2	27.7	220319488	1

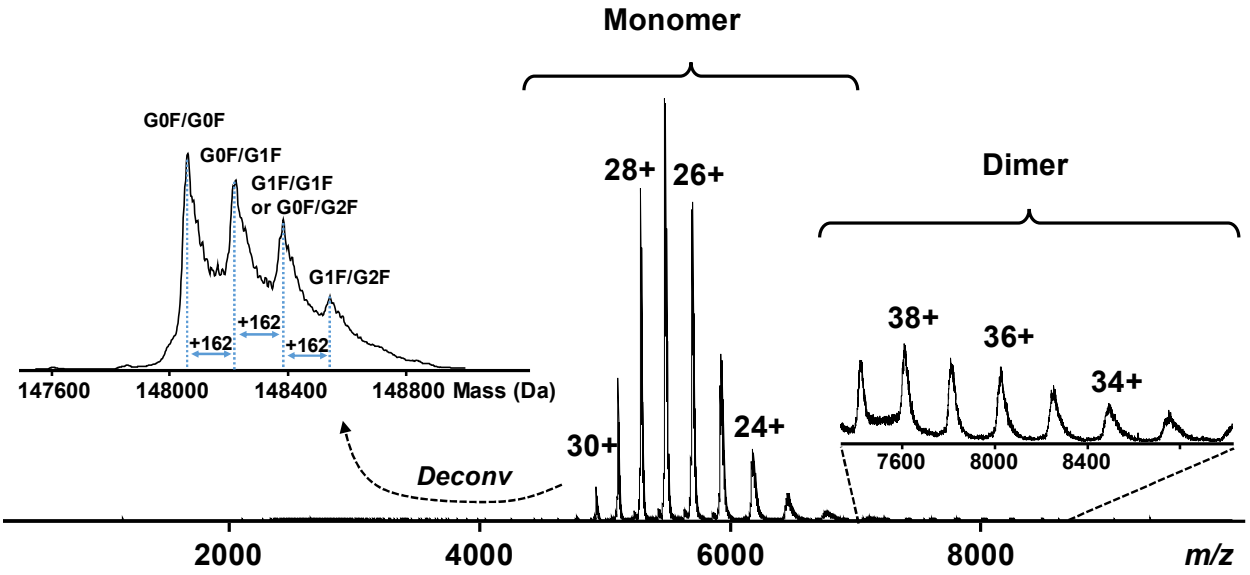
Supplemental Table 5.1. Retention time and area under curve (AUC) of mAb1 and mAb2 eluted from HIC-MS on a maXis II Q-TOF mass spectrometer.

Glycoproteoforms	Experimental Mass (Da)	Theoretical Mass (Da)	Mass Error (ppm)
G0F/G0F	148038.8	148039.5	4.8
G0F/G1F	148200.9	148200.6	2.0
G1F/G1F or G0F/G2F	148363.0	148362.8	1.7
G1F/G2F	148525.2	148524.5	4.2
G2F/G2F	148687.3	148684.8	17

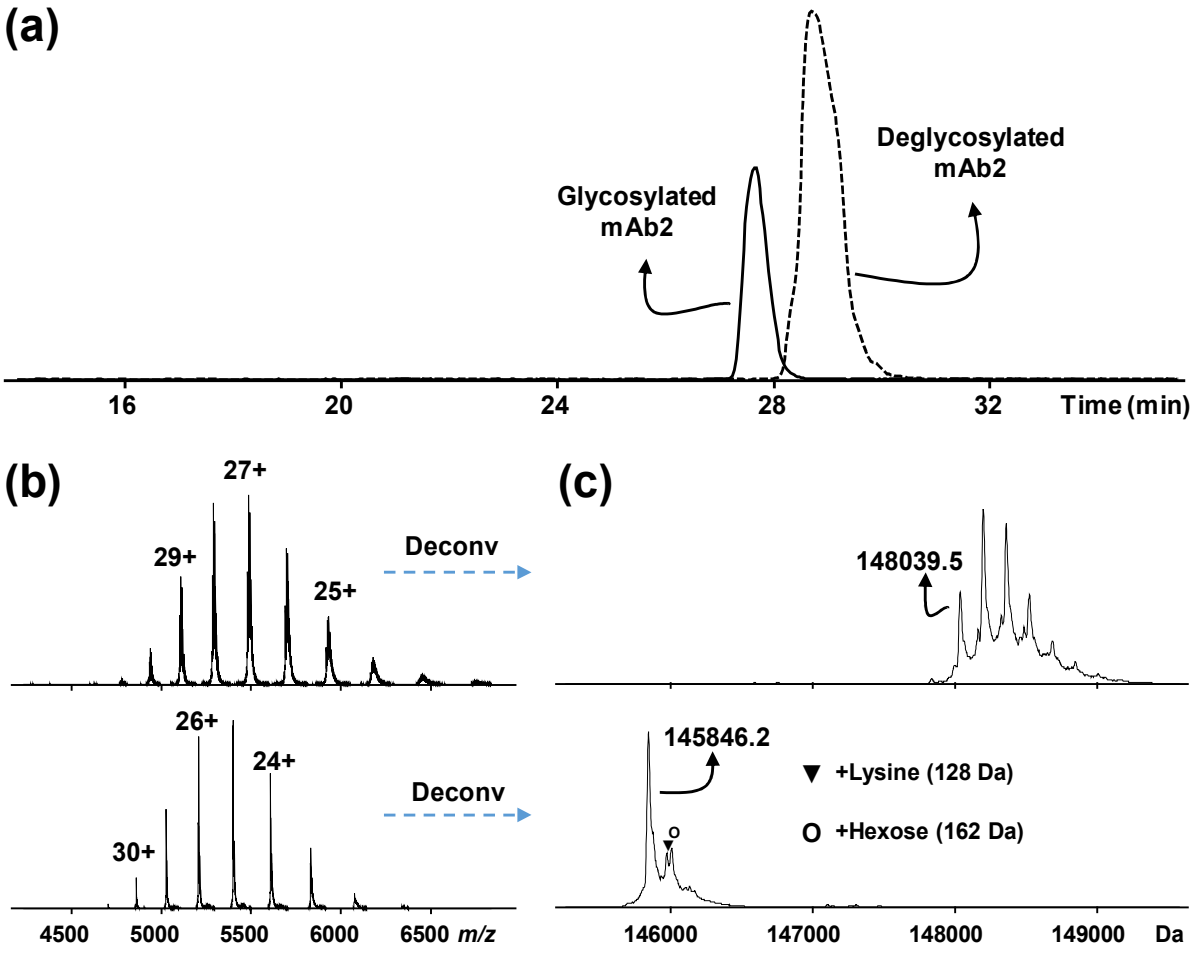
Supplemental Table 5.2. Experimental vs. theoretical masses of mAb2 glycoproteoforms[304] detected from HIC-MS on a maXis II Q-TOF mass spectrometer.



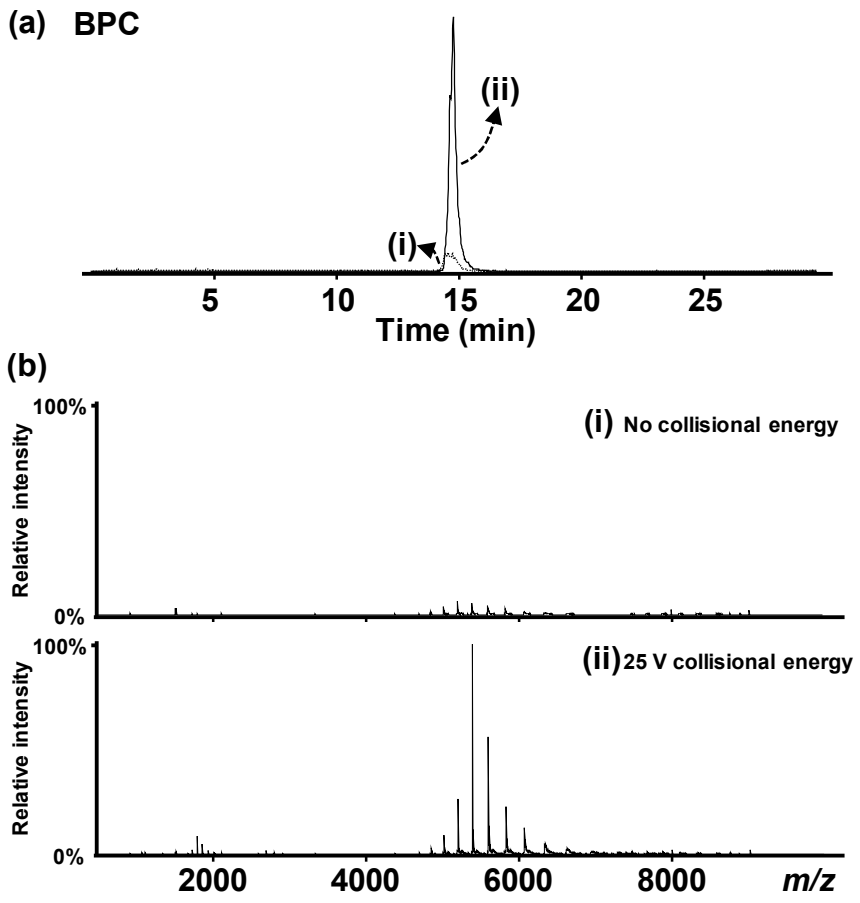
Supplemental Figure 5.1. Quantitative HIC-MS analysis of mAb2 on a maXis II QTOF mass spectrometer. (a) Overlaid chromatograms of mAb2 in different injection amounts; (b) working curve of peak area versus the amount of antibody from a single injection in (a) showing an R^2 of 0.993.



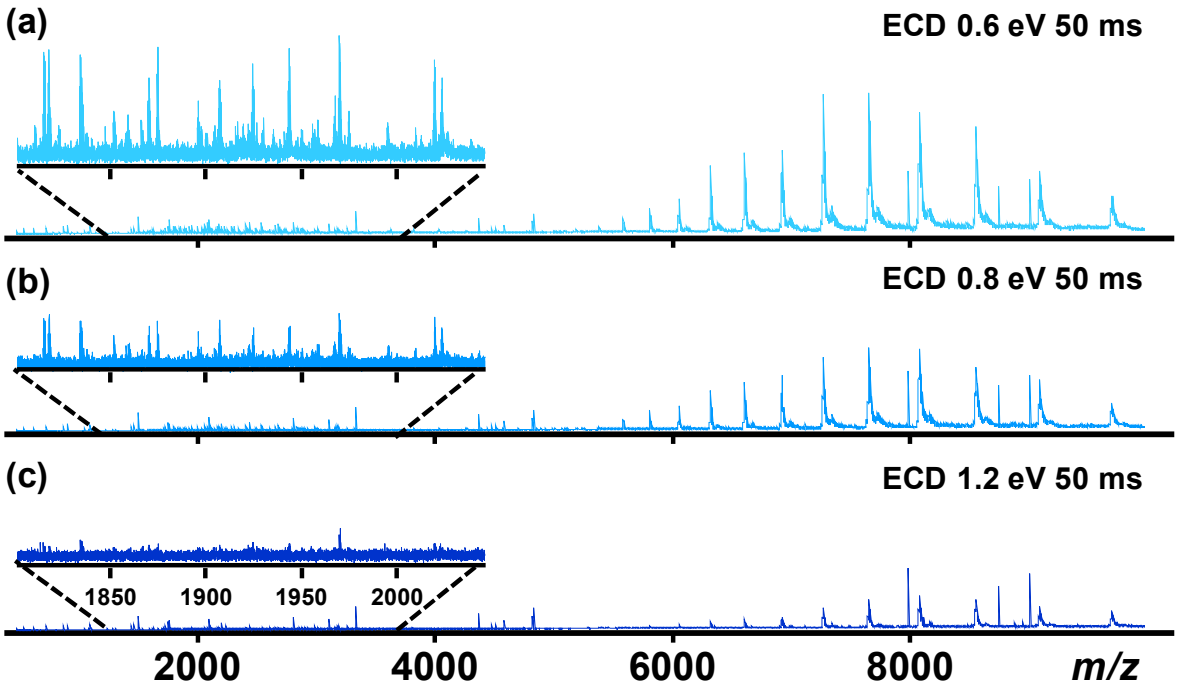
Supplemental Figure 5.2. Average mass spectrum of mAb1 showing the detection of monomer and dimer from a maXis II Q-TOF mass spectrometer. The call-out shows the deconvoluted mass spectrum of mAb1 monomer with multiple glycosylated proteoforms.



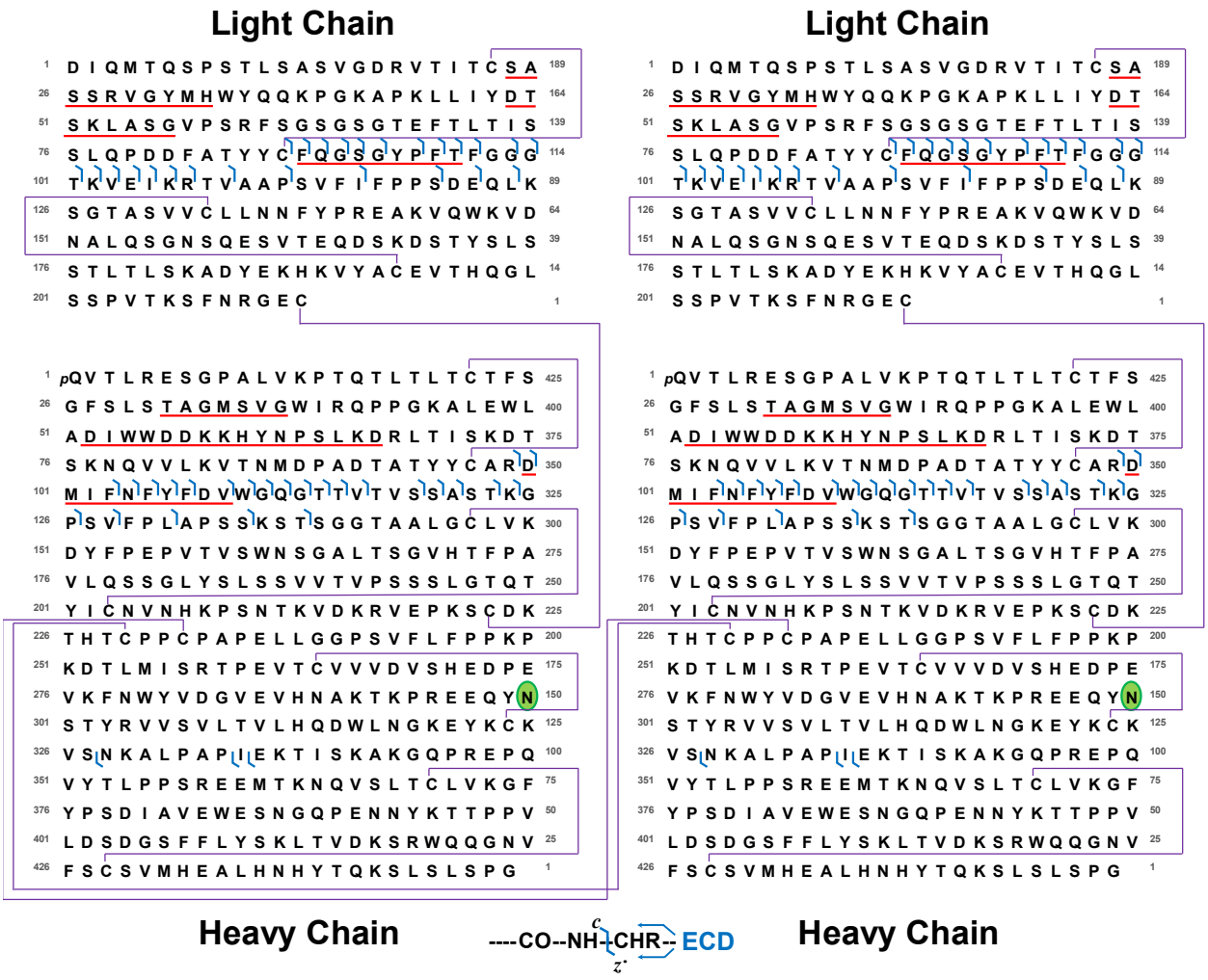
Supplemental Figure 5.3. Comparison of HIC-MS analysis of glycosylated and deglycosylated mAb2 on a maXis II QTOF mass spectrometer. (a) Overlaid total ion chromatograms of glycosylated (solid) and deglycosylated (dash) mAb2; (b) mass spectra of glycosylated (top) and deglycosylated (bottom) mAb2; (c) Deconvoluted spectra of glycosylated (top) and deglycosylated (bottom) from b. Solid triangle represents a +Lysine (+128) modification and hollow circle represents a + Hexose (+162 Da) modification.



Supplemental Figure 5.4. Improvement of mAb2 signals in the 12T solariX XR FT-ICR mass spectrometer by applying collisional energy. (a) Overlaid HIC-MS base peak chromatograms of deglycosylated mAb2 in a shorter gradient with no (i) or 25 V (ii) collisional energy applied; (b) Average mass spectra of deglycosylated mAb2 when applying no (top) and 25 V (bottom) collisional energy with normalized intensity.



Supplemental Figure 5.5. Online broadband ECD fragmentation with different electron energies: (a) 0.6 eV (b) 0.8 eV (c) 1.2 eV and 50 ms irradiation time using a 12T solariX FT-ICR mass spectrometer. The intensities of all the spectra are normalized to the most abundant ion (9.4E6) in (a).



Supplemental Figure 5.6. Broadband ECD (0.6 eV and 50 ms) fragmentation map of intact deglycosylated mAb2 from the online HIC-MS experiment on the 12T solariX FT-ICR mass spectrometer. The solid purple lines represent known disulfide linkages. For heavy chains, *p*Q indicates pyroglutamic acid, and the highlighted N300 has +349.14 Da after deglycosylation by IgGZERO. Complementarity-determining regions are underlined in red.

5.7 Acknowledgements

We would like to acknowledge the financial support from Abbvie. We also would like to acknowledge the high-end instrument grant S10OD018475 (to Y.G.) for the acquisition of the high-resolution mass spectrometers that were used in this study. Furthermore, Y.G. would like to acknowledge NIH R01 grants, GM117058, HL109810, and HL096971.

Chapter 6: Middle-down Multi-Attribute Analysis of Antibody-Drug Conjugates with Electron Transfer Dissociation

This chapter is a manuscript in preparation:

Chen, B.; Lin, Z. Zhu, Y.; Jin, Y.; Xu, L.; Fu, C.; Zhang, Z.; Zhang, Q.; Wayne, P.; Ge, Y.

6.1 Abstract

Antibody-drug conjugates (ADCs) are designed to combine the target specificity of monoclonal antibodies and potent cytotoxin drugs to achieve better therapeutic outcome. Comprehensive evaluation of the quality attributes of ADCs is critical for drug development but remains challenging due to heterogeneity of the construct. Currently, peptide mapping with reversed-phase liquid chromatography (RPLC) coupled to mass spectrometry (MS), has been the predominant approach to characterize ADCs. However, it is suboptimal for sequence characterization and quantification of ADCs because it lacks a comprehensive view of co-existing variants and suffers from varying ionization effects of drug-conjugated peptides compared to unconjugated counterparts. Here, we present the first middle-down RPLC-MS analysis of both cysteine (Adcetris®; BV) and lysine (Kadcyla®; T-DM1) conjugated ADCs at the subunit level (~25 kDa) with electron transfer dissociation (ETD). We successfully achieved high-resolution separation of subunit isomers arising from different drug conjugation and subsequently localized the conjugation sites. Moreover, we obtained a comprehensive overview of the micro-variants associated with each subunits and characterized them such as oxidized variants with different sites. Furthermore, we observed relatively high levels of conjugation near complimentarity-determining regions (CDRs) from the heavy chain but no drug conjugation near CDRs of light chain (Lc) from lysine conjugated T-DM1. Based on the extracted ion chromatograms, we measured accurate average drug to antibody ratio (DAR) values and relative occupancy of drug-conjugated subunits. Overall, the middle-down MS approach enables the evaluation of multiple quality attributes including DAR, positional isomers, conjugation sites, occupancy, and micro-variants, which potentially open up a new avenue to characterize ADCs.

6.2 Introduction

Over the past two decades, antibody-drug conjugates (ADCs) have emerged as a novel class of bioconjugates consisting of cytotoxic agents and recombinant monoclonal antibodies (mAbs) interconnected by stable chemical linkers.[312-314] The concept of ADCs exploits the specificity and stability of mAbs as well as the potency of the small molecule drugs against antigen-expressing targets, for instance, cancer cells.[312,315,316] Although the idea of ADC being the "magic bullet" appears highly promising and attractive, the translation of ADCs into clinical practice and success has been enormously challenging. Currently, only 4 FDA-approved ADCs are on the market: brentuximab vedotin (BV; Adcetris®), ado-trastuzumab emtansine (T-DM1; Kadcyla®), inotuzumab ozogamicin (Besponsa®), and gemtuzumab ozogamicin (Mylotarg®) whereas more than 65 ADC candidates are under clinical evaluation.[315]

The majority of these ADCs on the market and under clinical trials are based on lysine or cysteine conjugation. Site specific conjugation technologies are promising but still under further development.[317] Lysine conjugation occurs mainly on solvent accessible lysine residues (e.g., T-DM1), whereas cysteine conjugation takes place at the cysteine residues after the reduction of interchain disulfide bonds (e.g., BV).[313] Consequently, both conjugation schemes manifest various positional isomers for a given drug-to-antibody ratio (DAR) species. The heterogeneity generated by the inherently random lysine or cysteine conjugation further complicates the analysis and characterization of ADCs, in addition to the high hydrophobicity introduced by the small molecule conjugation and the readily presence of post-translational modification (PTM) variants (e.g., amino acid clipping, oxidation, and glycosylation, etc.) on the mAbs. As a result, the efficacy, stability, and safety of the ADCs could be compromised.[318] Therefore, a comprehensive

characterization of the quality attributes is critically important during ADC development and quality control.

Among various analytical techniques used for analyzing ADCs, liquid chromatography coupled to mass spectrometry (LC-MS) remains one of the most versatile and powerful methods to obtain quality attributes.[298,319] In particular, peptide mapping with reversed phase liquid chromatography (RPLC)-MS, is a well-established and sensitive approach to identify and confirm amino acid sequences, variants, and modifications of ADCs.[320] However, the bottom-up approach is prone to the introduction of artifacts due to prolonged sample preparation steps, and lacks of a comprehensive view on ADCs with different variants due to sequence variation, drug conjugation, and PTMs.[321,322] Moreover, because of the drastic differences in size, hydrophobicity, and protonation propensity, the attachment of small molecule drugs to peptides often results in strong retention on the column resin, poor recovery, and subdued ionization efficiency when compared to the unconjugated peptide counterparts. Hence, it makes the relative quantitation with the drug-conjugated and unconjugated peptides based solely on LC-MS peak intensity less reliable without full consideration of potential interferences.[323]

Alternatively, top-down[4,8,9,324] and middle-down[325,326] approaches from the intact (~150 kDa) or subunits (~25 kDa) levels can provide a shorter sample preparation time, a more global and comprehensive view on the heterogeneity and variants,[303,327] and conceivably a more reliable quantitation (less dominant impact from drug-linker than the case of conjugated peptides).[157,328] Although mAbs have been extensively investigated in the past few years by top-down or middle-down MS,[102,103,324,329-332] there are very limited studies on ADCs with these approaches.[327] Recently, Dyachenko *et al.* successfully demonstrated native tandem MS (MS/MS) of a cysteine conjugated ADC, BV.[327] In addition to profiling the overall heterogeneity, they were able to isolate

individual DAR species and performed MS/MS to obtain information on drug localization.[327] However, because of the large molecular weight of ADCs, MS/MS from an intact level results in limited sequence coverage and therefore remain challenging to precisely localize conjugation site to a single amino acid residue. Furthermore, positional isomers and low abundance micro-variants are difficult to separate and characterize from a top-down approach. Alternatively, subunits at a size around 25 kDa produced by enzymes such as IdeS appear to be more feasible and promising.[333] Nevertheless, middle-down LC-MS methods with MS/MS to characterize ADCs, particularly to pinpoint conjugation sites of isomers and to characterize small variants, remain unexplored.

Here, we developed a middle-down RPLC-MS strategy with ETD to investigate both cysteine and lysine conjugated ADCs (i.e., BV and T-DM1) at the subunit levels. The middle-down method evaluates multiple analytical attributes including average DAR, positional isomers, conjugation sites, occupancy, and micro-variants. For BV, we explored both IdeS[334] (FabRICATOR) and KGP[335] (GingisKHAN) enzymes as well as partial (only inter-chain reduced) and full reduction of all disulfide bonds in order to maximize the fragmentation around the hinge region for confident assignment of the conjugation sites (Scheme S1). For T-DM1, IdeS digestion and full reduction was used to achieve high yields of subunits for RPLC separation and efficient fragmentation. Overall, the described middle-down RPLC-MS approach with ETD shows a great potential for examining multiple quality attributes of ADCs during product development and quality control.

6.3 Material and Methods

6.3.1 Reagents

HPLC grade water and isopropyl alcohol (IPA) were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Formic acid (FA), 3,3,3-Trifluoropropionic acid (TFPA), 2-amino-2-(hydroxymethyl)-1,3-propanediol hydrochloride (Tris-HCl), dithiothreitol (DTT), tris(2-carboxyehtyl)phosphine hydrochloride (TCEP) and guanidine hydrochloride (HCl) were purchased from Sigma-Aldrich (St. Louis, MO, USA).

6.3.2 Sample preparation

Brentuximab vedotin (BV; Adcetris®), Ado-trastuzumab emtansine (T-DM1; Kadcyla®) in their standard formulation buffer were provided by AbbVie (North Chicago, IL, USA). FabRICATOR (IdeS), GingisKHAN (KGP), and GlyCINATOR (EndoS2) were purchased from Genovis (Sweden). IdeS and KGP proteolysis was carried out based on manufacturer's recommendation. Additional details are provided in the Supporting Information.

6.3.3 LC-MS/(MS)

A Waters ACQUITY UPLC M-class system (Milford, MA, USA) was coupled to a maXis II ETD Q-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany). 150 mm × 0.5 mm BIOshell A400 Protein C4 columns with 3.4 μm particle size (Sigma-Aldrich, St. Louis, MO, USA) was used for the reversed-phase separation. On column desalting was performed and bypassed the

buffer salts into waste. For the mass spectrometer, the end plate offset and capillary voltage were optimized at 500 V and 4500 V, respectively. The nebulizer was set to 0.5 bar, and the dry gas flow rate was 4.0 L/min at 220 °C. Collision cell RF was set to 2000 V_{pp}, transfer time to 110 µs, and pre pulse storage to 10 µs for ion desolvation and transmission. The quadruple low mass was set to 500 *m/z* and the scan range was from 200-3500 *m/z* at 1Hz. ETD reagent was 3, 4-hexanedione (114 *m/z* negative ion). Precursor accumulation was ranging from 400 to 1000 ms, the reagent time ranging from 5-22 ms, and the reaction time ranging from 0-2 ms. Additional details are provided in the Supporting Information.

6.3.4 Data Analysis

All data were processed and analyzed by Compass DataAnalysis 4.3 and MASH Suite Pro[129]. Extracted ion chromatograms were created based on top-5 most abundant charge state ions with an isolation window of 2 *m/z* for DAR calculation and 0.5 *m/z* for oxidation investigation. Chromatograms were smoothed for illustration. Area under curve was subsequently integrated for DAR calculation. Maximum Entropy was used for spectrum deconvolution with resolution set 80,000 for the high-resolution spectra. SNAP algorithm were used for peak picking, in which quality threshold was set to 0.4. All fragment ions were manually validated.

6.4 Results and Discussion

6.4.1 RPLC-MS/(MS) for the cysteine-conjugated ADC

IdeS cleaves BV below the hinge region followed by a reduction step to produce subunits around 25 kDa (**Figure S7.1**). Sample treatment with IdeS, GlycINATOR, and a disulfide reduction step (partial or full) resulted in a total of seven separated subunits on a RPLC-MS run including Fc/2, Lc without drug (Lc0), Lc with 1 drug (Lc1), Fd without drug (Fd0), Fd with 1-3 drugs (Fd1-3, respectively) (**Figure 7.1 and Figure S7.2**). Among all the subunits, each of Fd1 and Fd2 contained two isomeric peaks, namely, Fd1a, Fd1b, and Fd2a, Fd2b. We assigned these putative identifications based on accurate masses of the subunits with multiple injection and loading replicates (**Figure S7.**3).

While we observed all seven subunits from the partial and full reduction results, chromatogram with better separation and higher recovery of Fd2 and Fd3 were achieved when all disulfide bonds were reduced (**Figure 7.1**). Moreover, full reduction is preferred, because precise control of the partial reduction was difficult, which led to the observation of small and stronger retaining peaks corresponding to subunits with one more intra-chain disulfide bond reduced (**Figure S7.2**). Because of the small size of the drug with respect to the large subunits, we expect similar ionization efficiency among unconjugated and conjugated subunits based on previous studies.[16,97,157,212,253] By extracting the top-5 most abundant charge states and calculating the AUC of the EIC of each subunit from the fully reduced samples (**Figure 7.1**), we found 50% of Lc were conjugated with 1 drug and 32%, 27%, and 22% of Fd were conjugated with 1, 2, and 3 drugs respectively. Importantly, when fully reduced, the drug-linker portion of Fd2 and Fd3 subunits were more susceptible to fragmentation (loss of 761 Da) during ionization and ion transfer process

in the mass spectrometer (**Figure S7.4**). Therefore, we have included the top-5 charge states of the Fd2 and Fd3 subunits with a loss of 761 Da (corresponding to the peak at 762 *m/z*) in addition to the intact Fd2 and Fd3 for the average DAR calculation. An average DAR of 4.0, which was in good agreement with the reference value, were obtained for BV based on the EIC (**Table S7.1**). Our data demonstrated that calculation of average DAR value using EIC of subunits was consistent with UV results.

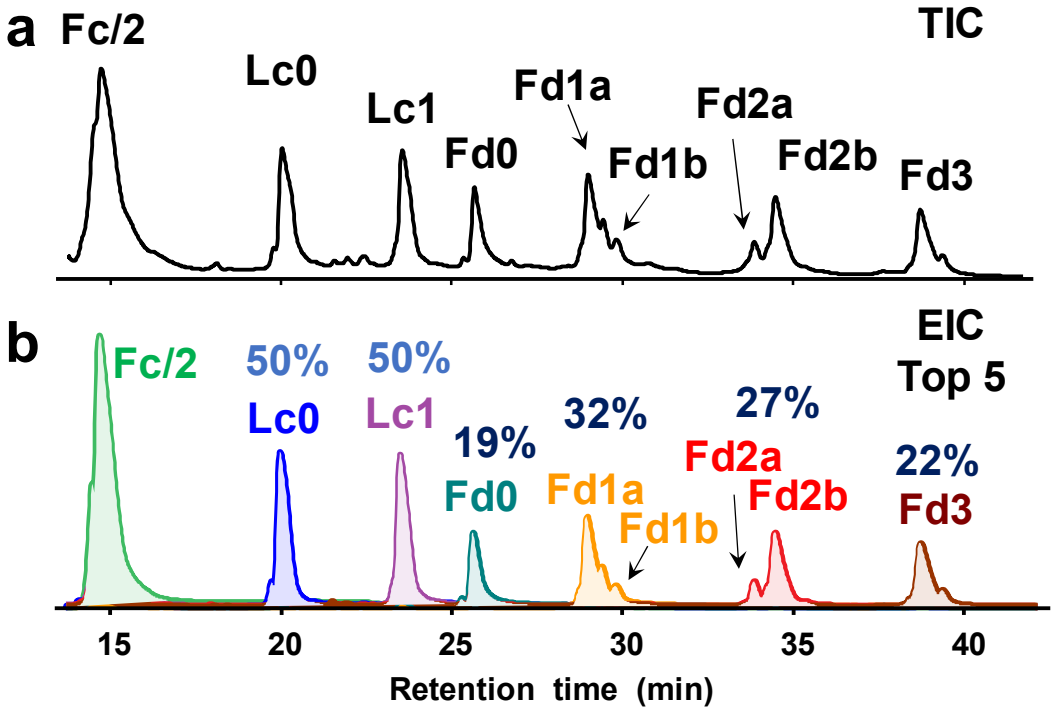


Figure 6.1. Online RPLC-MS of fully reduced BV digested by IdeS. (a) Total ion chromatogram (TIC) and (b) Extracted ion chromatograms of top-5 charge states of each subunits. Relative percentage of unconjugated and conjugated species for Lc and Fd were shown.

The MS1 spectra of the middle-down approach also provided a comprehensive view of each subunit and its associated micro-variants. For example, the EIC of Fc/2 showed a minor peak eluting slightly earlier than the major peak (**Figure 7.2a**). The deconvoluted spectrum of peak (i) reveals the oxidized (+16 Da) and the unclipped C-terminal lysine (+128 Da) variants of Fc/2

(**Figure 7.2b**). Both of these two species are more hydrophilic, leading to an earlier elution than the main peak (ii). The deconvoluted spectrum of the main peak (ii) corresponds to Fc/2 with all intra-chain disulfide bonds reduced (27093.00 Da) and the variants with the core GlcNAc and fucose after deglycosylation (GlycINATOR), conforming the site of N-glycan attachment of Fc chain. Minor variants of water loss (-18 Da) and fucose loss (-164 Da) were observed (**Figure 7.2b**). [336] We also observed a prevalent +18 Da variants of drug conjugated subunits (i.e., Lc1, Fd1, Fd2, and Fd3), likely due to the hydrolysis of succinimide-thioether ring. The EIC of Fd3 shows the separation of the minor peak (ii) from main peak (i) (**Figure 7.2c**). Deconvoluted mass spectra reveal the later eluting peak (ii) corresponds to Fd3 with one succinimide ring being hydrolyzed, with an additional mass shift of 18 Da (**Figure 7.2d**).

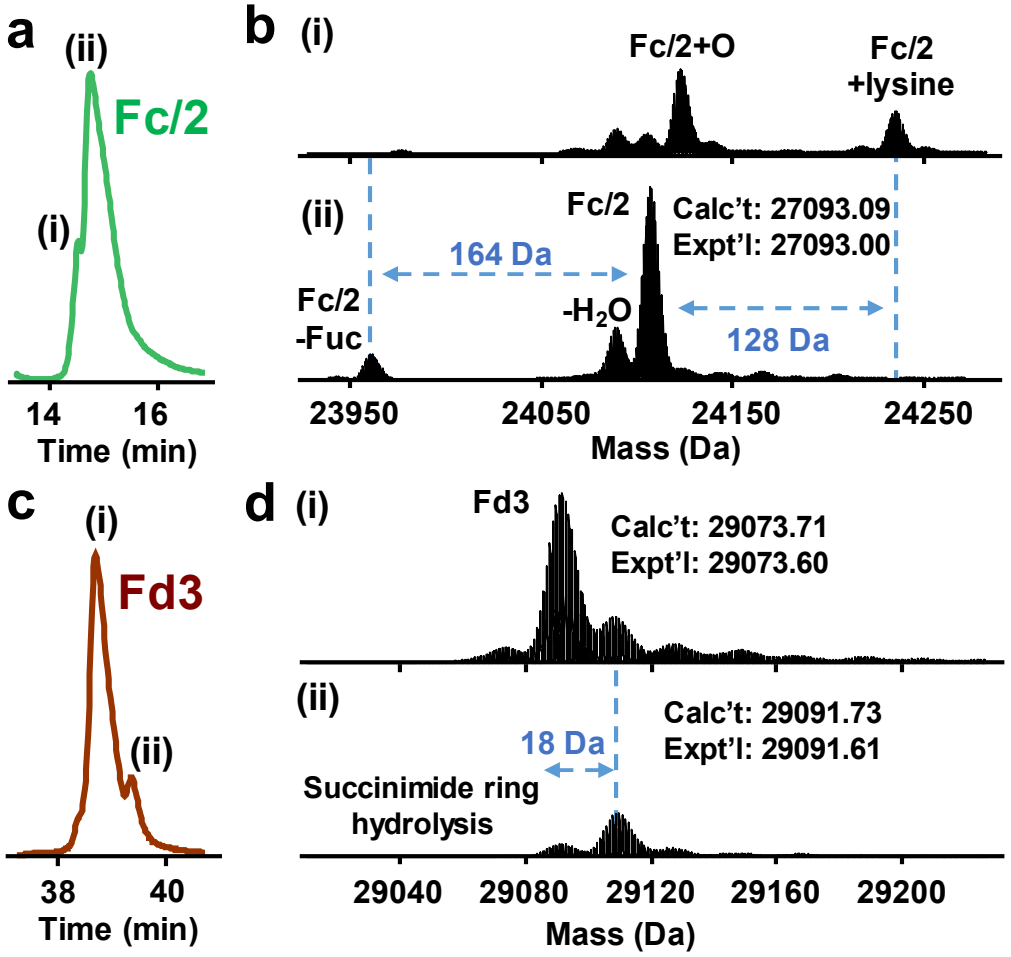


Figure 6.2. Micro-variants of BV subunits revealed by middle-down RPLC-MS. (a) Extracted ion chromatogram of Fc/2 showing two peaks (i) and (ii). (b) Deconvoluted spectra of peak (i) and (ii). (c) Extracted ion chromatogram of Fd3 showing two peaks (i) and (ii). (d) Deconvoluted spectra of peak (i) and (ii).

Next, we developed a middle-down approach with MS/MS aiming to characterize micro-variants and localize conjugation sites. When first applying collision-induced dissociation (CID) during MS/MS experiments, we found CID preferentially triggers the immolation of p-aminobenzyl alcohol fragmentation around the carbamate bonds of the drug-linker portion of BV, leading to predominantly drug fragments and no subunit backbone cleavages (**Figure S7.5**).

Therefore, we subsequently applied electron-transfer dissociation (ETD) to all the subunits, which was known to provide rich sequence information with the preservation of labile bonds.[96]

Among all the observed micro-variants, oxidation in the Fc/2 subunit of BV is particularly interesting, as methionine or tryptophan oxidation on the Fc region can impact the binding of Fc receptors, stability, and half-life of the mAb.[337,338] By extracting the top-5 most abundant charge states of oxidized Fc/2 and Fc/2 with a narrower isolation window (0.1 *m/z*), we observed the separation of more hydrophilic oxidized Fc/2 from un-oxidized Fc/2 (**Figure 7.3a**). Two distinct peaks were observed from the EIC of oxidized Fc/2, indicating the potential existence of isomers and more than one oxidation sites. Subsequent ETD MS/MS comparison between the oxidized Fc/2 isomers (i and ii) and unoxidized Fc/2 (iii) confirmed our speculation (**Figure 7.3b**). Oxidized $\boldsymbol{z}^{\bullet}_{57}$ was detected with no observation of oxidized $\boldsymbol{c}_{57}$ from oxidized Fc/2 isomer i, whereas oxidized $\boldsymbol{c}_{57}$ was observed with no detection of oxidized $\boldsymbol{z}^{\bullet}_{57}$ from oxidized Fc/2 isomer ii. The finding suggested that there are at least two oxidation sites, one near the C-terminus (Met16 or Tyr41), and the other near N-terminus of Fc/2 (Tyr181 or Met192). The locations of the oxidation sites were in good agreement a previous mAb study in which forced degradation were performed to induce two oxidation sites.[330] Overall, 11% of the Fc/2 were oxidized without force degradation; and a higher level of oxidation was found near the N-terminus (ii) than C-terminus (i) of Fc/2 based on the AUC of the EICs. Therefore, the developed middle-down method is capable to investigate oxidation at a relatively low level.

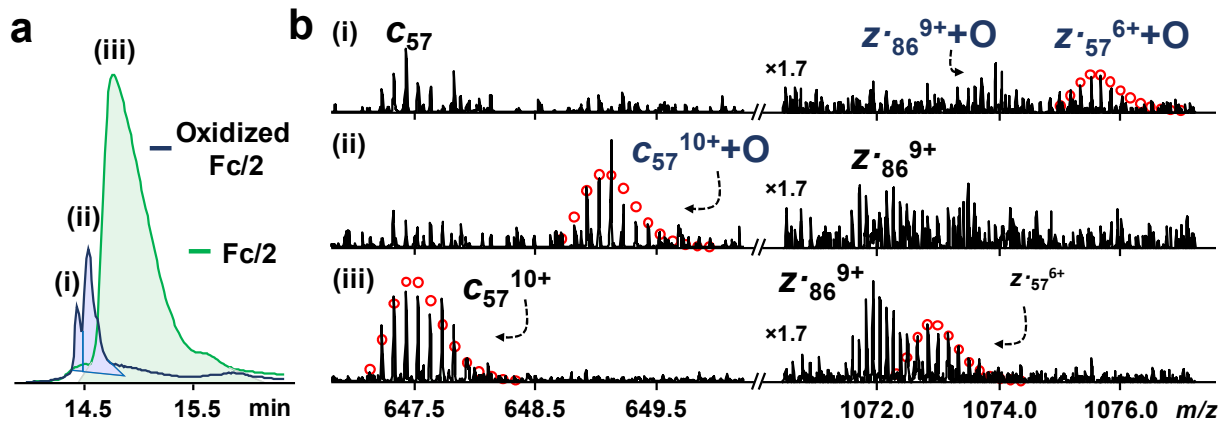


Figure 6.3. ETD LC-MS/MS characterization of oxidation of BV subunit Fc/2. (a) Extracted ion chromatogram of oxidized Fc/2 (i and ii isomers) and un-oxidized Fc/2 (iii) (b) ETD spectra validating the present of oxidized Fc/2 isomers. The intensities were normalized and 1071-1077 *m/z* was zoomed in 1.7 times. Theoretical ion distributions are indicated by the red dots.

For the analysis of different subunits, most subunits with or without conjugated drugs yielded a sequence coverage about 35% with ETD when all the intra-chain disulfide bonds were reduced (**Figure 7.4a, S7.6 and S7.7**). The N-glycosylation site on the Fc/2 was therefore confidently assigned with many fragment ions containing the core GlcNAc with fucose after deglycosylation (GlyCINATOR). When intra-chain disulfide bonds were not reduced, the overall sequence coverage was about 15% for most of the subunits (**Figure 7.4b, S7.8 and S7.9**). The seemingly low fragmentation efficiency and sequence coverage was caused by the presence of disulfide bonds, which led to nearly no cleavages of the residues within the disulfide bridges.[339] However, as a result, more extensive ETD fragmentation were observed in the middle of the sequence outside of the disulfide bonds confinement, particularly in the middle of the sequence.

Although the fully reduced ETD spectra provided higher overall sequence coverage and confidently localize the drug conjugation sites in Lc1 and Fd3, it has yet to unambiguously localize

the conjugation sites in Fd1a compared to the partially reduced ETD spectra. In the partially reduced ETD spectrum (**Figure 7.4c**), we were able to pinpoint the conjugation site of isomer Fd1a to the cysteine above the hinge region based on the detection of unconjugated ($z^{\bullet}_{15}$ and $z^{\bullet}_{16}$) and drug conjugated ($z^{\bullet}_{24}$) diagnostic fragment ions from the partially reduced conditions (**Figure 7.4d**). In contrast, even though the fully reduced ETD spectrum provided high sequence coverage (**Figure 7.4a and b**), key diagnostic ions (e.g., $z^{\bullet}_{15}$ and $z^{\bullet}_{16}$) to preclude possible conjugation on hinge region cysteines (CPPC) were undetectable. Therefore, partial reduction may work better from a site localization perspective, because the disulfide linkages limit ETD cleavages to only certain regions, resulting in simpler spectra with higher signal to noise.

For Fd1b, Fd2a, and Fd2b, confident assignment of conjugation sites from both partially and fully reduced ETD spectra remained challenging. Although the fully reduced subunits bear more charges and therefore a high reaction cross section during ETD event, diagnostic ions around the hinge regions of Fd were still not detected, leaving ambiguous localization of conjugation sites (**Figure S7.7**). For the partially reduced subunits, the $z^{\bullet}_{23}$ ion from Fd2a and Fd2b was only able to confirm that there are two drugs located near to the C-terminus hinge regions out of three possible conjugation cysteines. (**Figure S7.9**). We speculated that this is due to the lack of charged residues on the C-terminus (around the hinge region) of Fd and potential steric hindrance of the drug and proline residues blocking electron/radical transfer on the protein backbone.

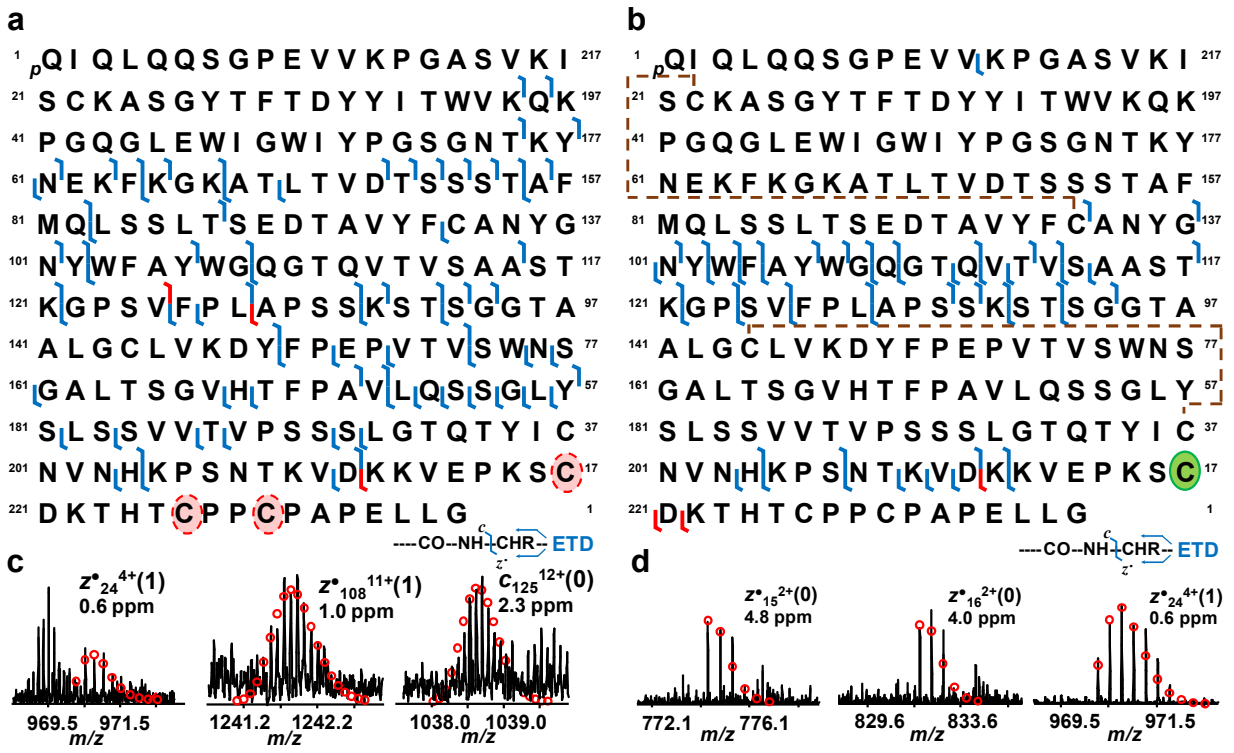


Figure 6.4. ETD LC-MS/MS characterization of partially and fully reduced Fd1a isomer of BV digested by IdeS. (a) ETD Fragmentation maps of fully reduced Fd1a with (b) representative ETD fragment ion with drug ($z^{\bullet}_{24}{}^{4+}$ and $z^{\bullet}_{108}{}^{11+}$) and without drug ($c_{125}{}^{12+}$). The red circles with a dash line indicate possible localization of the single conjugation site. (c) ETD Fragmentation maps of partially reduced Fd1a with (d) diagnostic ETD fragment ion with drug ($z^{\bullet}_{24}{}^{4+}$) and without drug ($z^{\bullet}_{15}{}^{2+}$ and $z^{\bullet}_{16}{}^{2+}$), localizing the conjugation site to Cys220. The green circle with a solid line indicates confident localization of the conjugation site. Theoretical ion distributions are indicated by the red dots.

In order to differentiate positional isomer and localize conjugation sites around the hinge region, we have further investigated the use of enzyme KGP to cleave BV above the hinge region into new subunits below 30 kDa (**Figure S7.1 and Figure S7.10**). With KGP, the hinge region locates near the N-terminus of Fc'/2 (near C-terminus of Fd from IdeS). *c*-type fragment ions therefore includes a charge from N-terminal amine. The c_6 ion (THTCPP|CPAP) between two potential conjugated cysteine residues were observed in the Fc'/2 as well as Fc'2 with 1 drug (Fc'2_1) during ETD experiments (**Figure S7.11 and S7.12a**). However, drug conjugated c_6 ion

were not detected neither in Fc'2_1 nor in Fc'2_2 (**Figure S7.12**). Therefore, drug conjugated isomers around the hinge region (CPPC) remain indistinguishable with our current middle down method (see extended discussion in supplementary information).

Overall, all the subunits of BV and some positional isomers were well-separated in the newly developed middle-down RPLC-MS/(MS) method. We obtained an average DAR value of 4.0 and assessed relative conjugation levels of each subunits for BV from AUC of EICs of top-5 most abundant charge states. From the MS1 spectra, micro-variants associated with the subunits were separated and revealed. Moreover, we examined the relatively low level of oxidation of Fc/2, showing the presence of more than one oxidation sites by using ETD. We further characterized partially and fully reduced BV subunits. Although the overall sequence coverage was higher when subunits were fully reduced, more extensive fragmentation around the middle and the C-terminus of drug conjugated Fd were obtained when subunits were partially reduced with intra-chain disulfide bonds intact. Nevertheless, our middle-down with IdeS and KGP as well as the current peptide mapping approaches were unable to differentiate the two cysteine residues around the hinge regions (THT**C**PP**C**PAP),[328] possibly because of the lack of charged residues and the steric hindrance of the small molecule drug during fragmentation events.

6.4.2 RPLC-MS/(MS) for the lysine-conjugated ADC

T-DM1 is an ADC that could have drug-linker attachment to any of the 88 Lysine residues and the N-terminal amines of Lc and Hc. Because of the complexity arising from all the possible conjugation sites, peptide mapping approach has been exclusively used to localize conjugation sites and to quantify their relative abundance. However, potential drawbacks include prolonged

sample preparation, complicated inference of the whole ADC, and ionization suppression of drug conjugated peptides during quantification. Here, we developed and evaluated a RPLC middle-down approach with ETD, aiming to provide a more global view of the lysine-conjugated, localization of the abundant conjugation sites, and reliable assessment of relative abundance.

To facilitate MS/MS analysis, IdeS and strong reducing conditions were used to produce high yield of subunits with complete reduction of inter-chain and intra-chain disulfide bonds. From the TIC of T-DM1 (**Figure 7.5a**), we observed subunits with an elution order of Fc/2_0, Fc/2_1, Lc0, Lc1, Lc1, Fc/2_2, Fd0, Lc2, Fd1, Fd2, Fd3, along with some possible low abundance truncated species (Mr 20404.23 Da) at the beginning of the gradient. We next extracted the top-5 most abundant ions of each subunits, revealing the existence of multiple isomers of drug conjugated subunits in the EIC (**Figure 7.5b, c, d**). Only EIC traces of the corresponding subunits at the right elution time were shown due to the complexity of overlapping charge states of other subunits (**Figure S7.13**). In particular, in Fd1, multiple isomeric peaks were observed (**Figure 7.5d**). For Fc/2 and Lc, subunits with up to 2 drug conjugation were detected. Although the abundance was relatively low, Fd subunit with up to 3 drug conjugation was observed. By integrating AUC of the EIC of each subunit, we found that about 50% of the subunits were conjugated with small molecule drugs, and the average DAR of 3.5 can be therefore calculated (**Table S7.2**), in good correspondence to the reference value of 3.5.[340]

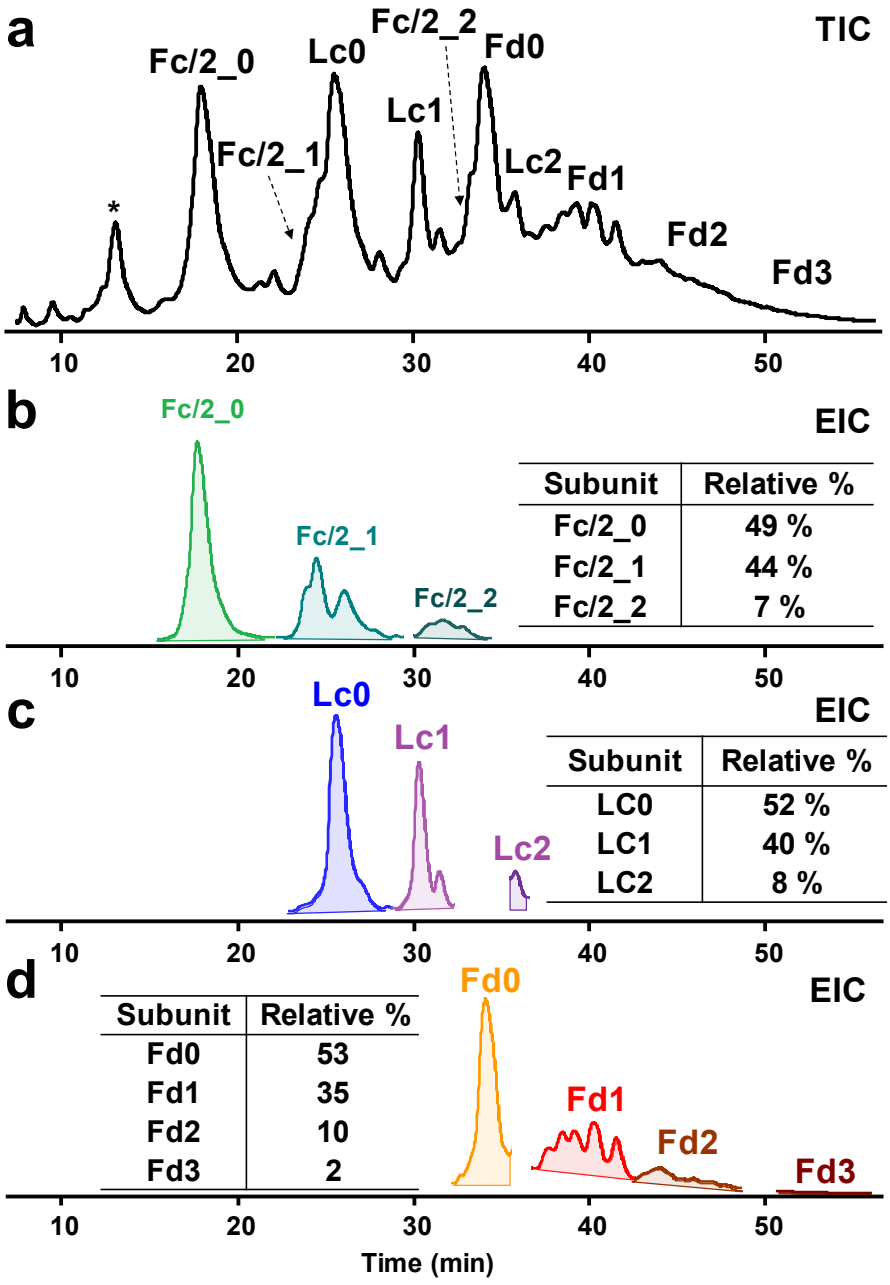


Figure 6.5. LC-MS analysis of subunits of T-DM1 digested by IdeS. (a) Total ion chromatogram (TIC).* corresponds to a possible truncation with a mass of 20404.23 Da. (b) Extracted ion chromatogram (EIC) of Fc/2 with 0, 1, and 2 drugs (Fc/2_0, Fc/2_1, and Fc/2_2). (c) EIC of Lc with 0, 1, and 2 drugs (Lc0, Lc1, and Lc2). (d) EIC of Fd with 0, 1, 2, and 3 drugs (Fd0, Fd1, Fd2, and Fd3). The inset tables show the relative percentage of each subunit based on the area under curve. Intensities are normalized in b, c, and d.

Next, we further investigated conjugation sites and their relative abundance, particularly for Lc1 and Fd1. Lc and Fd subunits are important targets to study because they contain complementarity determining regions (CDRs) and drug conjugation on certain sites can potentially influence antigen binding. Furthermore, subunits with a single drug conjugation sites consisted of majority of drug conjugated species and more feasible for further investigation. We therefore applied ETD to localize conjugation sites of the relatively abundant Lc1 and Fd1.

We further optimized the gradient to improve separation particularly for Fd1 subunits. As shown in Figure 7.6a, 2 peaks were observed for Lc1 isomers, and 11 peaks were separated for Fd1 isomers from the EICs. For Lc1, although isomers were not readily resolved from the EIC, we examined the targeted ETD MS/MS spectra along the corresponding elution window. Interestingly, after examining the fragment ions, no drug-conjugated *c*-ions were observed. In fact, we found that the conjugation sites were all localized near the C-terminus, away from the CDRs regions of the Lc (**Figure 7.6b**). The close proximity of these potential conjugated sites (K169, K183, K188, K190, and K207) near the N-terminus explains the inefficiency of separating the Lc1 isomers (**Figure 7.6a and Figure S7.14**).

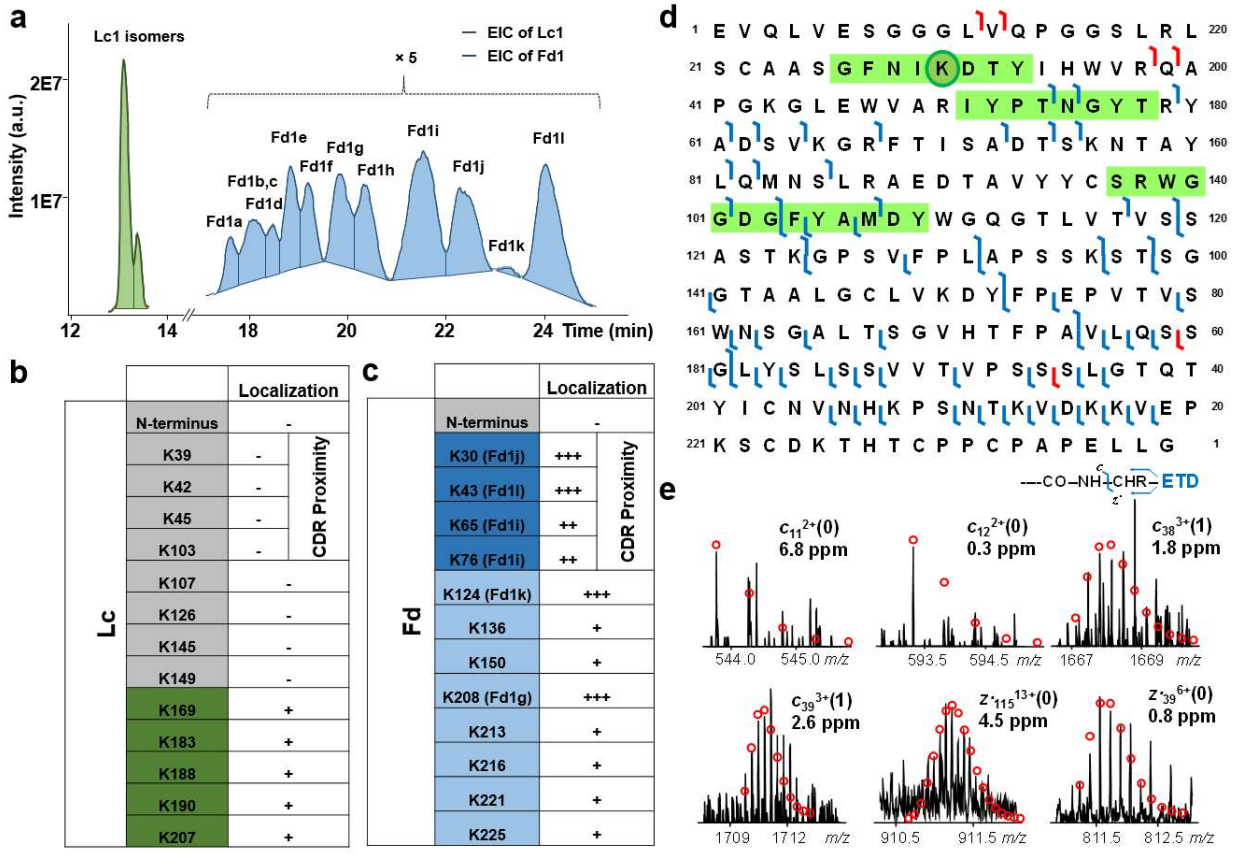


Figure 6.6. ETD LC-MS/MS characterization of T-DM1 Lc1 and Fd1 subunits digested by IdeS. (a) Extracted ion chromatograms (EICs) of Lc1 (green) and Fd1 (blue) (5x zoom-in). Summary of potential and confirmed conjugation sites of (b) Lc1 and (c) Fd1 species of T-DM1 digested by IdeS with ETD LC-MS/MS. +++: confidently localized conjugation sites; ++: highly possible conjugation sites; +: possible conjugation sites; and –: no conjugation observed. The darker blue shades indicates relatively higher abundance (d) Fragmentation map of Fdj localizing the conjugation sites of K30 within CDR1-VH. The highlighted regions in the sequence maps correspond to the CDR regions in the heavy chain. (e) Representative ETD fragment ions with one drug (c_{38}^{3+} and c_{39}^{3+}) and without drug (c_{11}^{2+}, c_{12}^{2+}, z_{39}^{6+} and z_{115}^{115+}), corresponding to the red cleavages in d. Theoretical ion distributions are indicated by the red dots.

For Fd1, at least 11 isomers can be observed from the EIC (**Figure 7.6a**). We were able to unambiguously localize the conjugation sites of four of these isomers, and narrowed down the location of the conjugation sites for the others (**Figure 7.6c, S7.15 and S6.16**). We also found that Fd1 isomers with conjugation sites near the C-terminus tend to elute earlier than those isomers

with conjugation sites near the N-terminus. Among these, high level of conjugation was found in the middle of CDR1-VH (K30) as well as near the other two CDR regions (CDR2-VH K43 and CDR3-VH K65/K76) (**Figure 7.6c**). Specifically, the conjugation site of K30 corresponds to Fd1j. Figure 7.6d shows the representative fragmentation map of Fdj from an online RPLC-ETD MS/MS analysis. Confident localization of the conjugation site to K30 was based on accurate detection of diagnostic fragment ions (**Figure 7.6d**).

Our analysis demonstrated the potential of applying middle down approaches to characterize lysine-conjugated ADCs. Because of the complexity of lysine conjugation, analysis of subunits with more than one conjugated drug remains challenging. However, our data revealed that majority of the drug-containing subunits contains only one drug. Subunits with more than one drugs would likely to be a combinatorial arrangement of the conjugation sites detected on the subunits with one drug conjugated. Therefore, the analysis of subunits containing one drug should shed light on the location of conjugation sites and their relative abundance as demonstrated for Fd1 isomers from our data.

Although peptide mapping is prevalent and widely-used for T-DM1, various results of characterizing T-DM1 were reported. Recently, Chen *et al.* used a peptide mapping approach to characterize the conjugation sites and compared their relative abundance.[341] They quantified over 35 conjugation sites with various abundance. Sang *et al.* also investigated the conjugation sites of T-DM1 in their more recent work, in which they were only able to localize and quantify less than a half of the conjugation sites found in Chen *et al.*'s work.[323] However, instead of directly comparing the intensity of the unconjugated and the conjugated peptides, they performed serial dilution to create a calibration curve to account for the ionization and detection difference of each drug conjugated peptides. Therefore, other than the localization of conjugation sites, relative

abundance of the same sites in these two works are vastly different. Our results revealed that for Lc1, nearly all drug conjugation is located far away from the CDRs (Lc K169, K183, K188, K190, and K207), which is consistent with the data Sang *et al.* observed. On the other hand, high level of conjugation was found on and around the CDR regions of the Fd subunit from the Hc, in particular, K65/K76 (Fd1i), K30 (Fd1j), and K43 (Fd1l), which provides an alternative perspective to the peptide mapping approach. All of these were in relatively high abundance compared to the other Fd1 isomers and are in close proximity with the CDRs of the heavy chain, which could be a potential risk impacting the specificity of antigen binding. Furthermore, by using the AUC of the EIC of each subunits, comparable average DAR value (3.5) to reference value (3.5) can be obtained. Therefore, we reason that the conjugation of a small molecule drug (1 kDa) to a relatively large size subunit (25 kDa) appears not to severely alter the ionization efficiency. Overall, our results showed that middle-down analysis with ETD is a viable way to localize drug conjugation sites, and provides a better relative quantification considering the significant influence of small molecule drug on peptide ionization.

6.5 Conclusions

In this work, we present a comprehensive middle-down investigation on both cysteine (BV) and lysine (T-DM1) conjugated ADCs with ETD to probe multiple quality attributes. With this middle-down approach, we not only detected variants associated with each subunit from a comprehensive view, but also demonstrated that relatively low level of oxidation can be investigated and quantified. Because of the small size of the drug molecule and the similar molecular weight of each subunit, we were able to obtain accurate average DAR values (4.0 and 3.5 for BV and T-DM1 respectively) and relative percentage of drug-conjugated subunits from the

EICs. In particular, we applied ETD on the separated positional isomers of drug-conjugated subunits from BV and T-DM1, demonstrating the feasibility of separating and localizing conjugation sites of positional isomers in addition to assessing their relative abundance. With the other emerging effective fragmentation techniques such as UVPD[111] and activated ion ETD (AI-ETD),[108] middle-down MS characterization of ADCs will become more effective and powerful. Taken together, we have shown the middle-down methods with high-resolution MS allows for the determination of average DAR, positional isomers, conjugations sites, occupancies, and variants from the subunit levels, providing a useful and complementary tool for characterizing multiple analytical attributes of ADCs during the development of novel therapeutics.

6.6 Supplemental Information

6.6.1 Extended method for sample preparation

For IdeS digestion of BV, ADC samples were diluted in 50 mM Tris-HCl buffer (pH=8), followed by digestion and deglysocylation by IdeS and GlycINATOR (1 unit per μg of ADC) respectively at 37 °C for 30 min. Partially reduced BV was obtained by adding 20 mM DTT to reduce inter-chain disulfide bonds at 37 °C for 30 min (pH=8). For fully reduced BV, 20 mM DTT was used with 8M Guanidine HCl for complete disulfide bond reduction (inter- and intra-chain) at 37 °C for 60 min. The resulting sample solution was desalted using PD SpinTrap G-25 (GE Healthcare) according to the recommended procedures. Because BV were found particularly susceptible to disulfide scrambling after full reduction (data not shown), 0.1% FA were added to solution immediately after the desalting step and prior to LC-MS analysis. For KGP digestion, BV was diluted in 100 mM Tris-HCl buffer (pH=8) to 2.0 μg/μL with or without the deglycosylation step

by GlycINATOR (1 unit per µg ADC) at 37 °C for 30 min. The KCP enzyme (3 unit per µg ADC) was then added to the sample with 2 mM freshly made cysteine (mild reducing agent) to a final concentration 0.5 µg/µL at 37 °C for 120 min. 20 mM DTT to reduce inter-chain disulfide bonds at 37 °C for 30 min (pH=8). For T-DM1, ADC samples were diluted in 50 mM Tris-HCl buffer (pH=8), followed by digestion and deglysocylation by IdeS and GlycINATOR (1 unit per µg ADC) respectively at 37 °C for 30 min. Complete reduction of inter- and intra-chain disulfide bonds was initially carried out with 50 mM TCEP at pH3 at 37 °C for 90 min and was later changed to 20 mM DTT with 8M Guanidine HCl at 37 °C for 60 min. The resulting sample solution from Guanidine HCl treatment was desalted using PD SpinTrap G-25 (GE Healthcare) according to the recommended procedures. No disulfide scrambling was observed from LC-MS analysis of T-DM1. All ADC samples had a final concentration of 0.5 -1.0 µg/µL.

6.6.2 Extended method for LC-MS/(MS)

150 mm × 0.5 mm BIOshell A400 Protein C4 columns with 3.4 µm particle size (Sigma-Aldrich, St. Louis, MO, USA) was used at a flow rate of 30 µL/min. The column was heated at 70 or 80 °C. MPA contained 0.2 % FA or (0.1 % FA and 0.1% TFPA) in water and MPB contained 0.2 % FA or (0.1 % FA and 0.1% TFPA) in 50/50 ACN/IPA. For the partially reduced BV digested by IdeS, a 45 min gradient with 70 °C column heating was used: 20% MPB initially; 20% MPB at 5 min; 23% MPB at 6 min; 40% MPB at 40 min; 90% MPB at 40.1 min; 90% MPB at 42 min; 20% MPB at 42.1 min; and 20% B at 45 min (corresponding to Figure S7.2). For the partially reduced BV digested by KGP, a 45 min gradient with 70 °C column heating was used: 20% MPB initially; 20% MPB at 5 min; 25% MPB at 6 min; 40% MPB at 40 min; 90% MPB at 40.1 min; 90% MPB at 42 min; 20% MPB at 42.1 min; and 20% B at 45 min (corresponding to Figure S7.10). For the fully

reduced BV digested by IdeS, a 60 min gradient with 80 °C column heating was used: 20% MPB initially; 20% MPB at 10 min; 25% MPB at 11 min; 45% MPB at 52 min; 90% MPB at 53 min; 90% MPB at 56 min; 20% MPB at 56.1 min; and 20% B at 60 min (corresponding to Figure 7.1).

T-DM1 was run initially in an 80 min gradient with column heating at 70 °C: 20% MPB initially; 20% MPB at 5 min; 24% MPB at 6 min; 36% MPB at 72 min; 90% MPB at 73 min; 90% MPB at 76 min; 20% MPB at 76.1 min; and 20% B at 80 min (corresponding to Figure 7.5). The flowing 60 min gradient was later optimized with column heating at 80 °C to separate Fd1 isomers: 20% MPB initially; 20% MPB at 6 min; 29% MPB at 11 min; 34% MPB at 52 min; 90% MPB at 53 min; 90% MPB at 56 min; 20% MPB at 56.1 min; and 20% B at 60 min (corresponding to Figure 7.6).

6.6.3 Extended method for Data Analysis

Relative abundance (%) was calculated based on the area under curve (AUC) of the specific drug conjugated or unconjugated subunit divided by the total AUC of the specific subunits. The integrated AUCs were highlighted in Figure 7.1, 7.3a, 7.5, and S7.13. For BV, average DAR was calculated as the following: $DAR = 2 \times (\frac{AUC_{Lc1}}{AUC_{Lc0}+AUC_{Lc}} + \frac{AUC_{Fd1}+2AUC_{Fd}\ +3AUC_{Fd}}{AUC_{Fd0}+AUC_{Fd}\ +AUC_{Fd2}+AUC_{Fd3}})$. For T-DM1, average DAR was calculated as the following: $DAR = 2 \times (\frac{AUC_{Fc/2\_1}+2AUC_{Fc/2\_2}}{AUC_{Fc/2\_0}+AUC_{Fc/2\_1}+AUC_{Fc/2\_2}} + \frac{AUC_{Lc1}+2AUC_{Lc2}}{AUC_{Lc0}+AUC_{Lc1}+AUC_{Lc}} + \frac{AUC_{Fd1}+2AUC_{Fd}\ +3AUC_{Fd}}{AUC_{Fd0}+AUC_{Fd}\ +AUC_{Fd}\ +AUC_{Fd}})$.

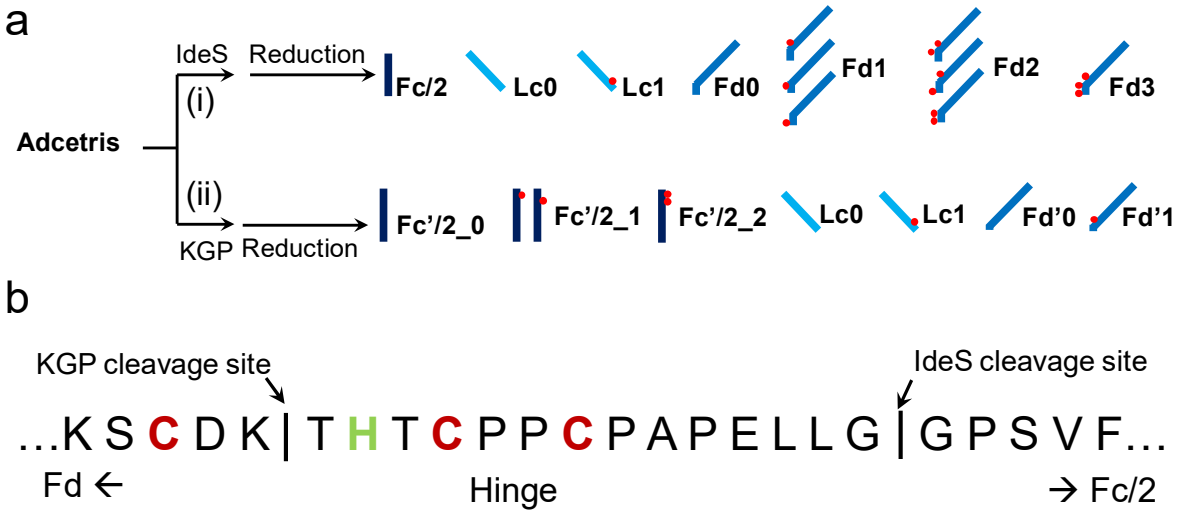
6.6.4 Extended discussion

A total of seven subunits were observed in KGP-digested BV: Fc'/2 with no drug (Fc'/2_0), Fc'/2 with 1 drug (Fc'/2_1), Fc'/2 with 2 drugs (Fc'/2_2), Fd' with no drug (Fd'0), Fd' with 1 drug (Fd'1), as well as Lc with no drug (Lc0) and Lc with 1 drug (Lc1) (identical to IdeS digestion) (Figure S7.10). Although previous studies have shown that KGP can digest mAbs with high efficiency,[335] the observation of low level of heavy chain (Hc) suggested that KGP digestion was not fully completed for BV (data now shown). Notably, the intensity of Fc'/2_1 and Fc'2_2 were particularly low. We speculate that potential aggregation and precipitation of drug conjugated Fc'/2 subunits might have caused the loss of recovery due to the increased hydrophobicity around the hinge region. The recovery of Fc'/2_1 as well Fc'/2_2 can be slightly improved by keeping the glycosylation (data not shown). Furthermore, drug conjugation around the hinge region (steric hindrance) might have negatively impacted the KGP digestion efficiency. Overall, the MS1 spectrum revealed consistent findings of different micro-variants of each subunit when compared to the IdeS data set.

With the hinge region present at the N-terminus, Fc'/2_0 was not only cleaved extensively between two disulfide linkages similar to Lc0 and Fd'0, but also near the C-terminus, covering the hinge region (THT**C**PP**C**PAPELLG) (Figure S7.11). This region was located in Fd of BV after IdeS digestions, and no cleavages were observed due to lack of charged residues (no $\boldsymbol{z}$ ions observed). With KGP, the hinge regions locates near the C-terminus with an N-terminal amine that can carry a charge if $\boldsymbol{c}$ ions were produced during ETD fragmentation. Indeed, high abundance $\boldsymbol{c}_6$ ion between two potential conjugation sites was observed. For the drug-linked subunits, besides Lc1, Fd'1 and Fc'2_2 were well fragmented, leading to confident localization of the conjugation sites (Figure S7.12 a and b). In order to differentiate the single conjugation site on Fc'/2_1, the

only diagnostic ion was $\boldsymbol{c}_6$, which was resulted from the only possible cleavage site between two cysteine residues of hinge region during the ETD event. Although in low abundance, $\boldsymbol{c}_6$ without a drug was detected (Figure S7.12a), implying the presence of isomer with a drug conjugation on the second cysteine of the hinge sequence (THTCPPCPAP). However, no drug-linked $\boldsymbol{c}_6$ was observed from Fc'/2_1, suggesting that the drug either hindered the fragmentation, or the isomer with drug at the top of the hinge is of low abundance. Examination of the Fc'2_2 data, further demonstrated that the lack of drug-linked $\boldsymbol{c}_6$ ion from the spectrum was likely due to the drug-linkage preventing effective ETD fragmentation (Figure S7.12b).

Therefore, our data suggests that the lack of charged residues on the C-terminus (around the hinge region) and potential steric hindrance of the drug blocking electron/radical transfer on the peptide backbone, might have led to no observation of diagnostic ions for site localization of certain drug-linked Fd isomers. Even though the alternative enzyme KGP cleavages BV samples above the hinge region, leaving 2 possible drug-linked sites on Fc'/2 with more a N-terminal amine, steric hindrance of the drug still prevented effective fragmentation around the conjugation sites as demonstrated by Figure S7.12. The observation of $\boldsymbol{c}_6$ ion in Fc'2_0 and Fc'2_1 but not in the drug-conjugated counterpart exemplified the influence of the small molecule drug on fragmentation.



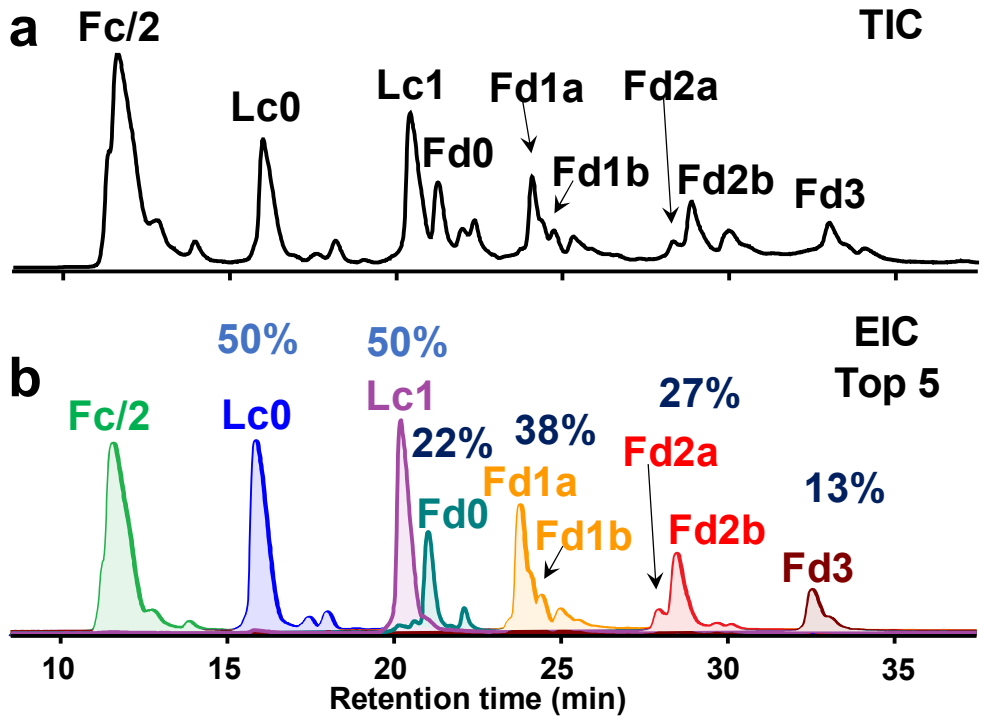
Supplemental Figure 6.1. Illustration of BV digested by IdeS and KGP. (a) Subunits generated from BV by IdeS and KGP enzymes. (b) Cleavage sites of IdeS or KGP enzymes near the hinge region.

RT	Subunit	Exp'tl mono-mass	Calc't mono-mass	Δm ppm	AUC
14.8	Fc/2	24093.00	24093.09	3.7	1669750656
20.0	Lc0	23713.40	23713.50	4.2	597638720
23.6	Lc1	25029.18	25029.14	1.6	596227584
25.7	Fd0	25126.29	25126.38	3.6	268483168
29.0	Fd1	26442.05	26442.16	4.2	466425600
34.5	Fd2	27757.83	27757.94	4.0	391996192
38.7	Fd3	29073.60	29073.71	3.8	308670048

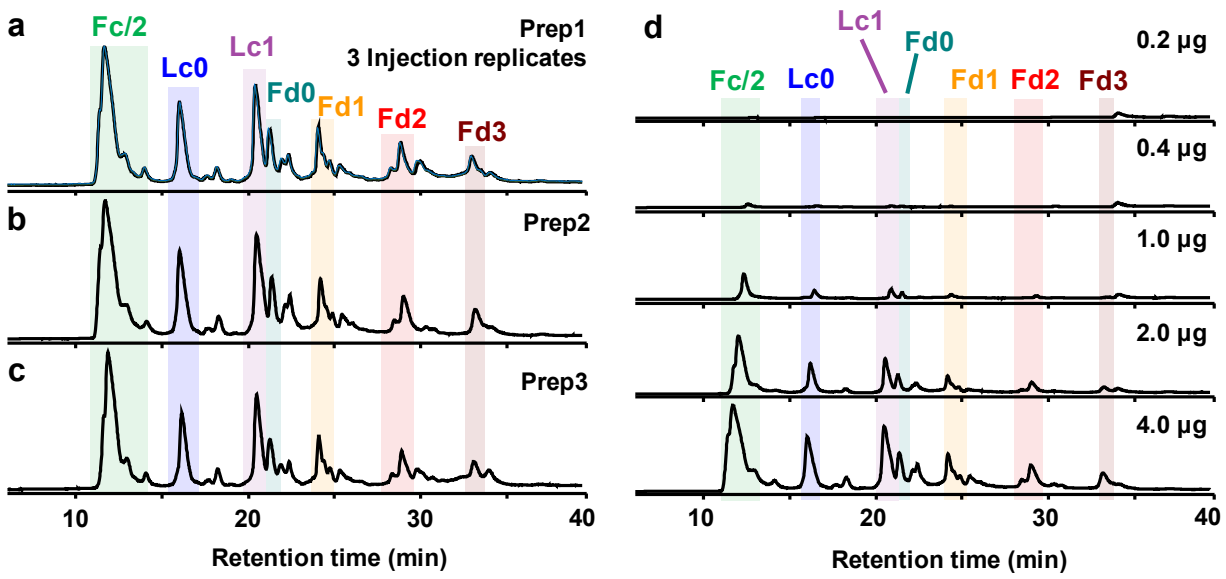
Supplemental Table 6.1. RPLC-MS of fully reduced BV subunits digested by IdeS. Area under curve (AUC) was obtained from the extracted ion chromatogram of each subunit based on top-5 most abundant charge states.

RT	Subunit	Exp'tl mono-mass	Calc't mono-mass	Δm ppm	AUC
18.1	FC/2_0	24125.13	24125.07	2.5	1463745920
24.8	LC0	23428.56	23428.52	1.7	890468736
25.8	Fc/2_1	25081.46	25081.43	1.2	1307905920
30.4	LC_1	24384.91	24384.88	1.2	678866432
32.0	Fc/2_2	26038.80	26037.80	38	206543456
34.2	Fd0	25367.54	25367.52	0.7	1086681088
35.9	LC_2	25341.29	25341.25	1.5	132241912
40.3	Fd_1	26324.88	26323.88	38	716603520
43.9	Fd_2	27280.23	27280.24	36	175125424
50.8	Fd_3	28238.57	28236.61	69	15800084

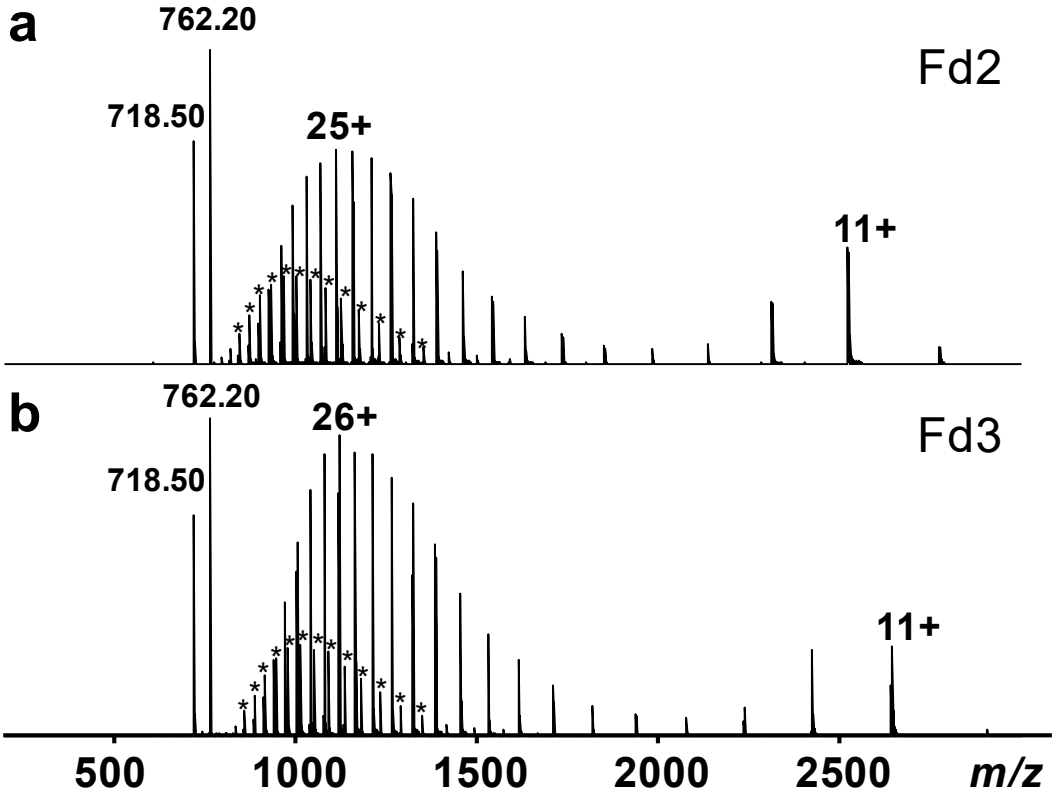
Supplemental Table 6.2. RPLC-MS of fully reduced T-DM1 subunits digested by IdeS. Area under curve (AUC) was obtained from the extracted ion chromatogram of each subunit based on top-5 most abundant charge states. 1 or 2 Da error for some of the low abundance subunits were caused by the error associated with the averagine model.



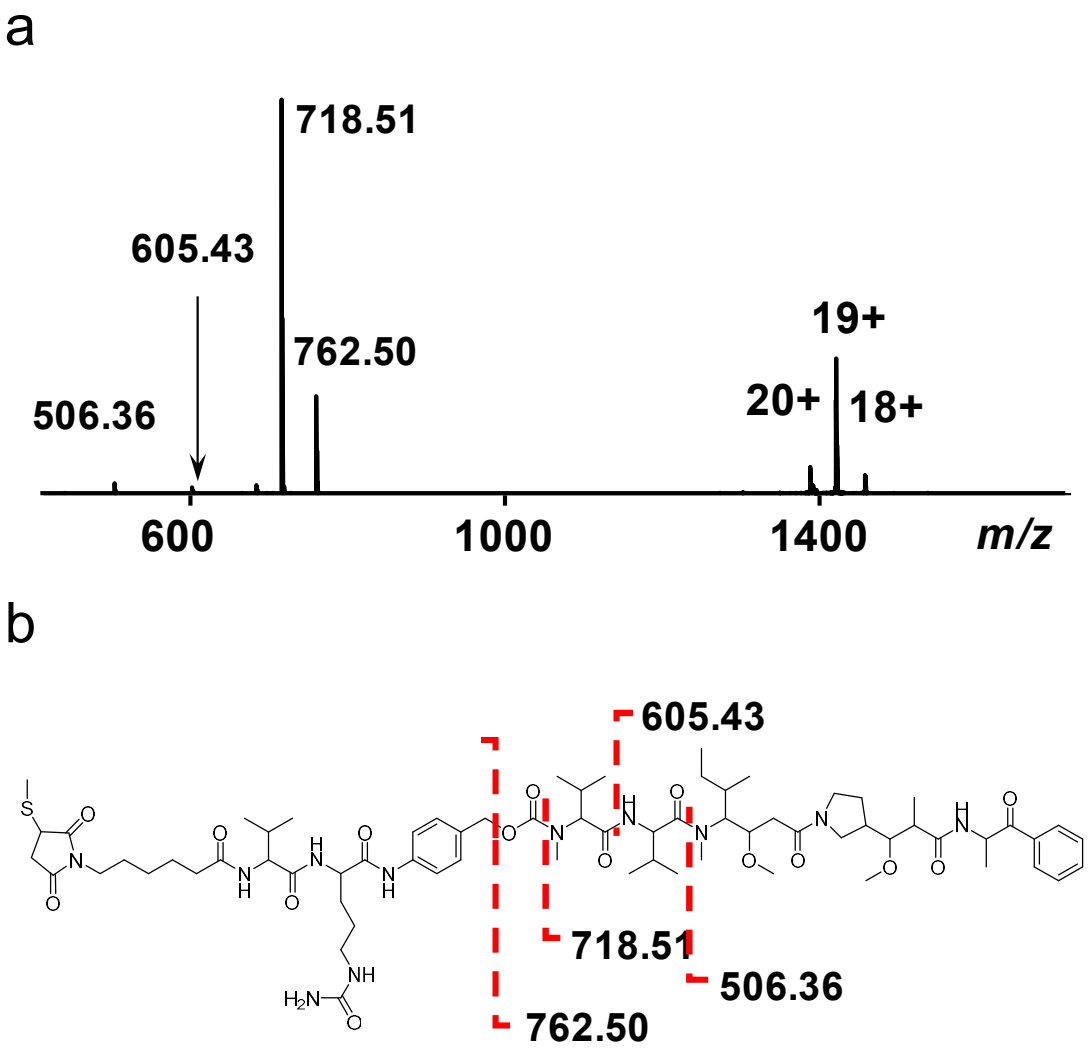
Supplemental Figure 6.2. Online RPLC-MS of partially reduced (intra-chain disulfide bonds intact) BV digested with IdeS. (a) Total ion chromatogram (TIC) and (b) Extracted ion chromatograms of top-5 charge states of each subunits. Relative percentage of unconjugated and conjugated species for Lc and Fd were shown.



Supplemental Figure 6.3. Assessment of the reproducibility of LC-MS injection, sample preparation, and injection amount using partially reduced (intra-chain disulfide bonds intact) BV. (a) Overlaid total ion chromatograms from three injection replicates. (b) and (c) Total ion chromatograms of the second and third sample preparation replicates, respectively. (d) Total ion chromatograms of from various injection amount.

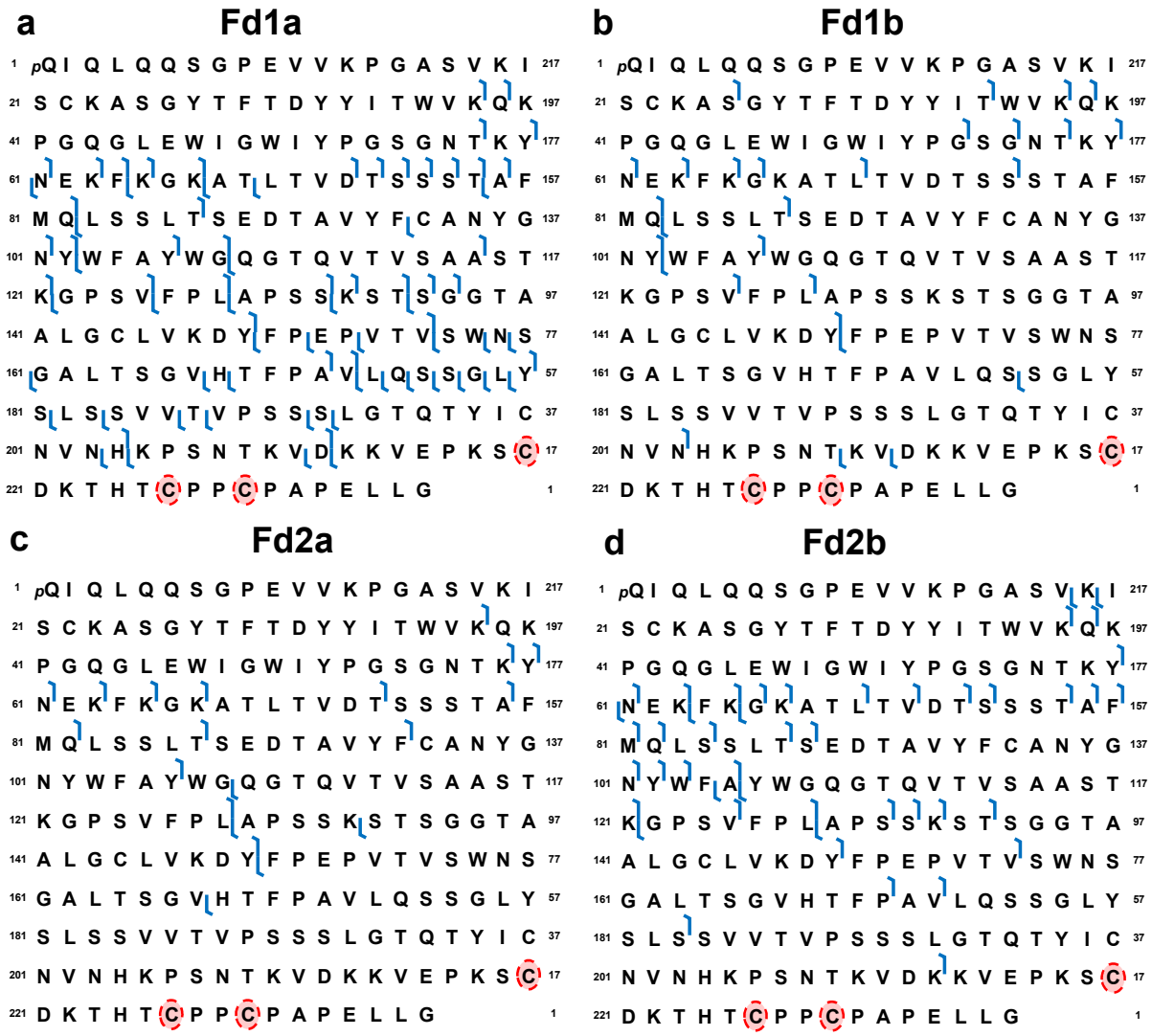


Supplemental Figure 6.4. Raw spectra of fully reduced BV Fd2 (a) and Fd3 (b) showing the fragmentation of drug-linker portion. * represents the Fd2 and Fd3 subunits

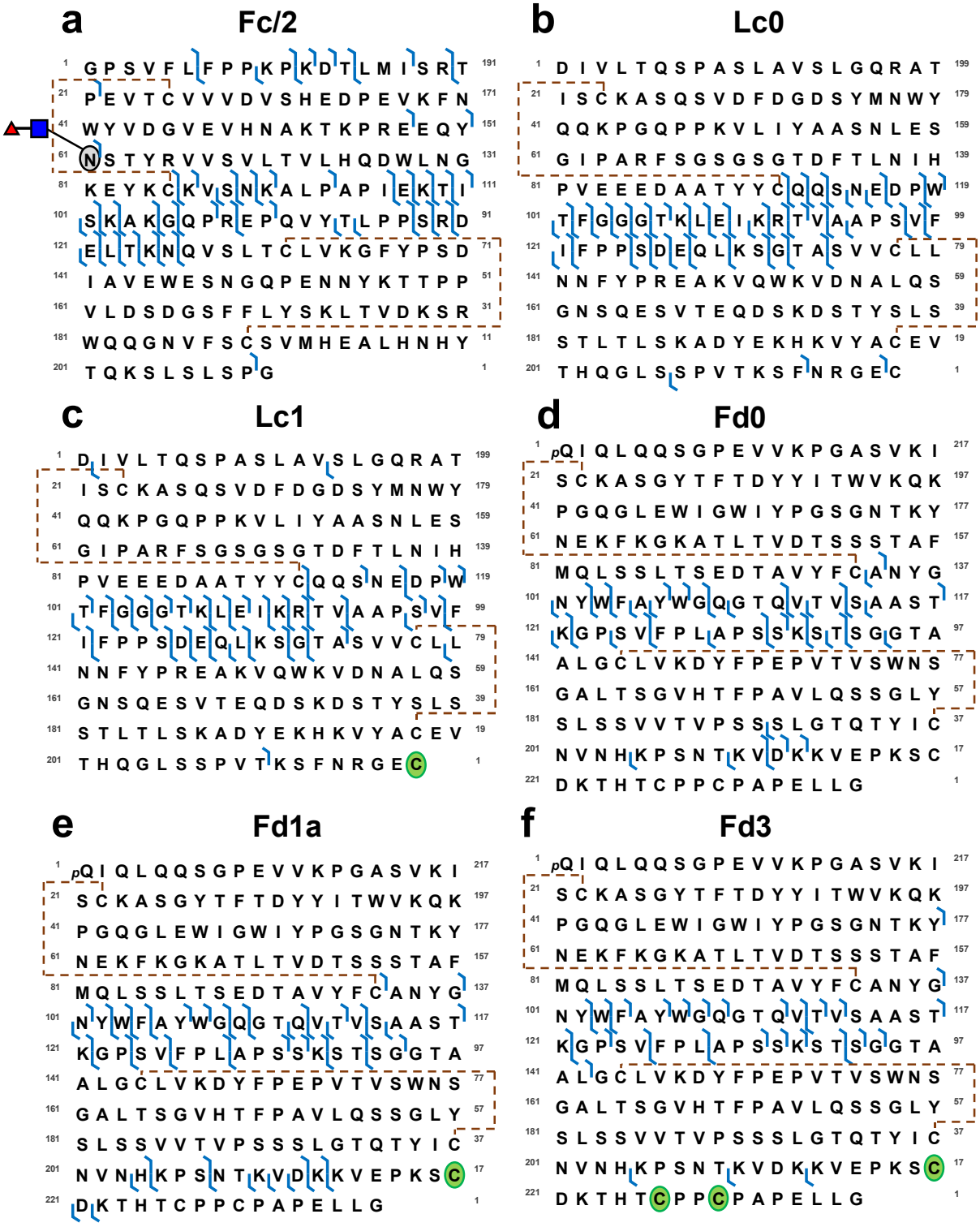


Supplemental Figure 6.5. Collision-induced dissociation (CID) of a drug conjugated subunit from BV. (a) CID MS/MS spectrum of the drug conjugated subunit. (b) Fragments from the small molecule drug of BV corresponding to the dominant fragmentation peaks at the low m/z from (a).

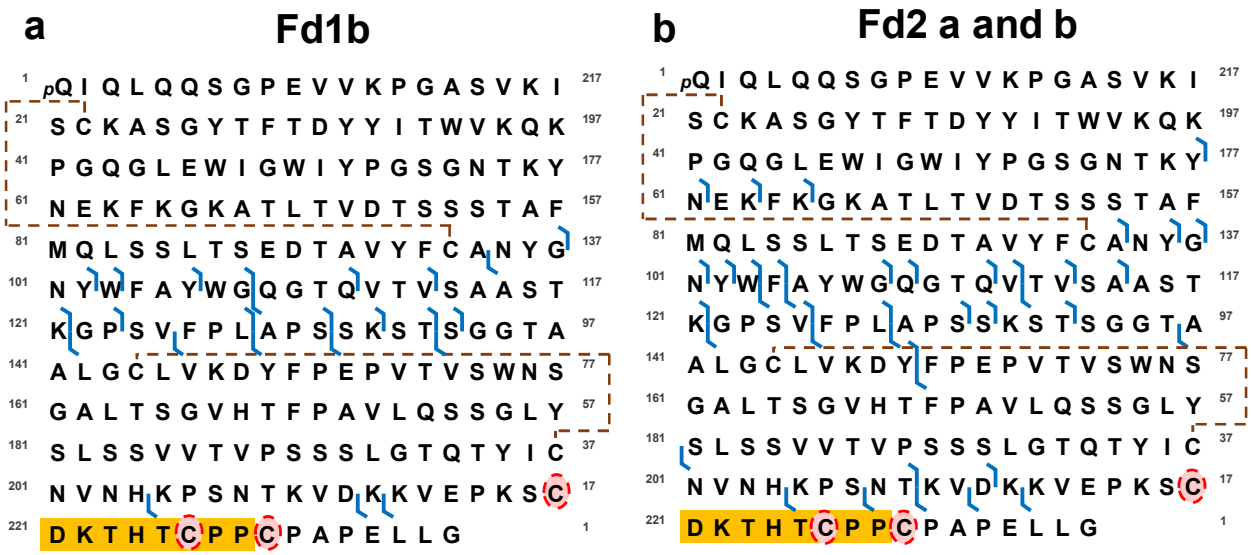
Supplemental Figure 6.6. ETD fragmentation maps of unconjugated subunits (Fc/2, Lc0, and Fd0) and drug conjugated subunits with confident conjugation site localization (Lc1 and Fd3) from fully reduced BV digested by IdeS. Green circles indicate confident localization of the conjugation sites.



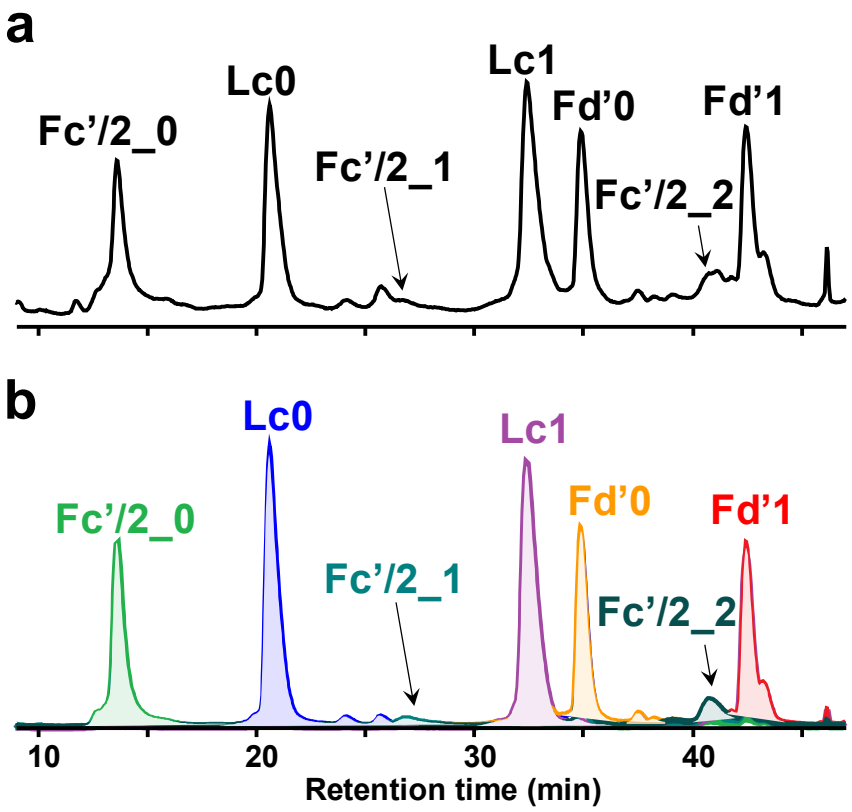
Supplemental Figure 6.7. ETD fragmentation maps of drug conjugated subunits with ambiguous conjugation site localization (Fd1a, Fd1b, Fd2a, and Fd2b) from fully reduced BV digested by IdeS. The red circles with a dash line indicate possible localization of the conjugation sites.



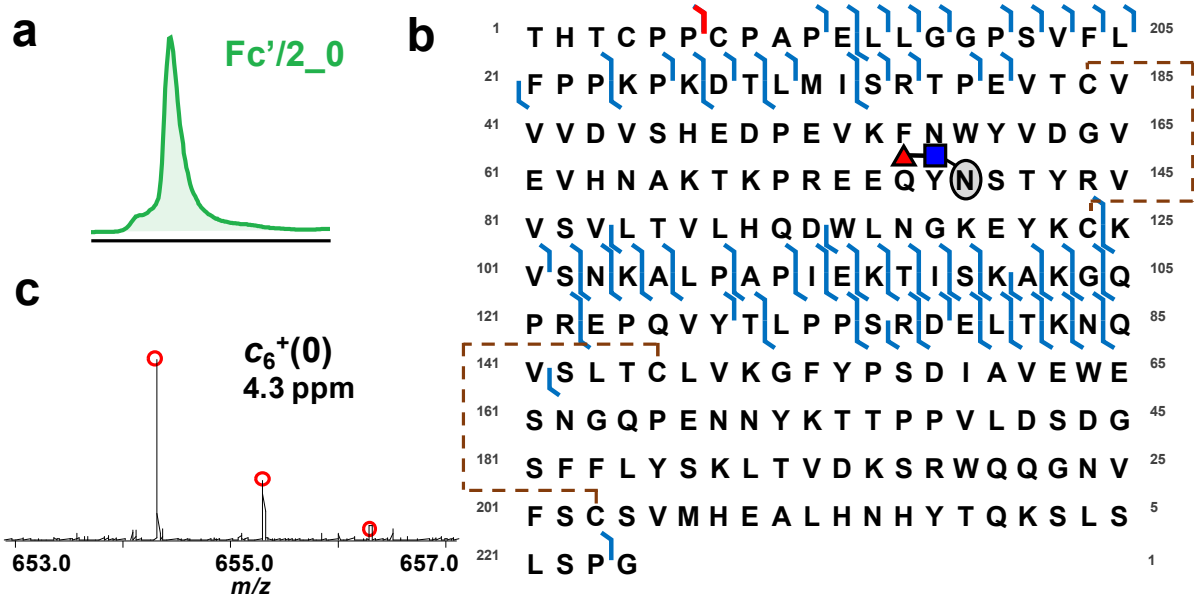
Supplemental Figure 6.8. ETD fragmentation maps of unconjugated subunits (Fc/2, Lc0, Fd0) and drug conjugated subunits with confident conjugation site localization (Lc1, Fd1a, and Fd3) from partially reduced BV digested by IdeS. Green circles indicate confident localization of the conjugation sites.



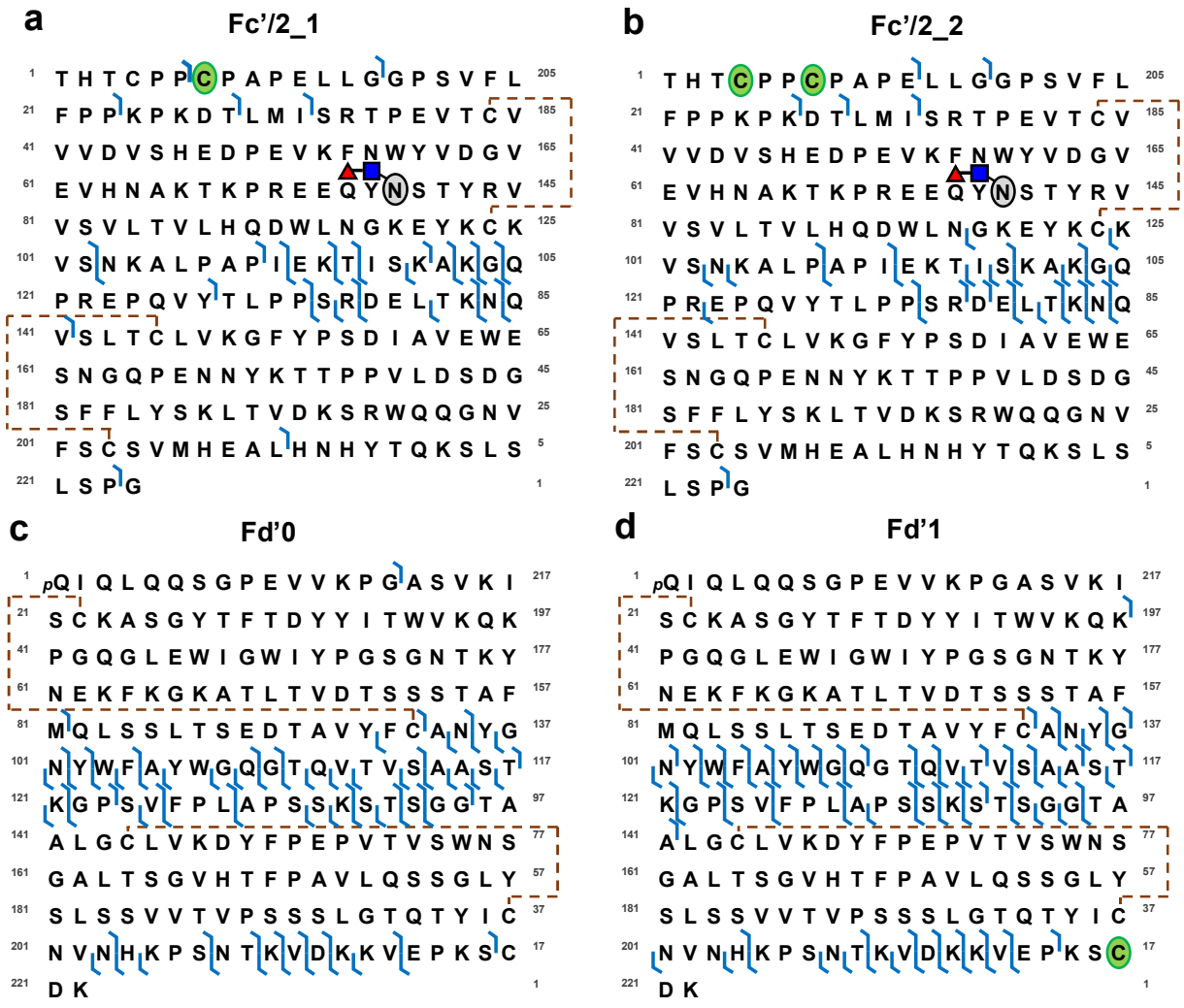
Supplemental Figure 6.9. ETD fragmentation maps of drug conjugated subunits with ambiguous conjugation site localization (Fd1b, Fd2a and b) from partially reduced BV digested by IdeS. The highlighted regions indicates the undetected cleavage sites required for diagnostic ions. The red circles with a dash line indicate possible localization of the conjugation sites.



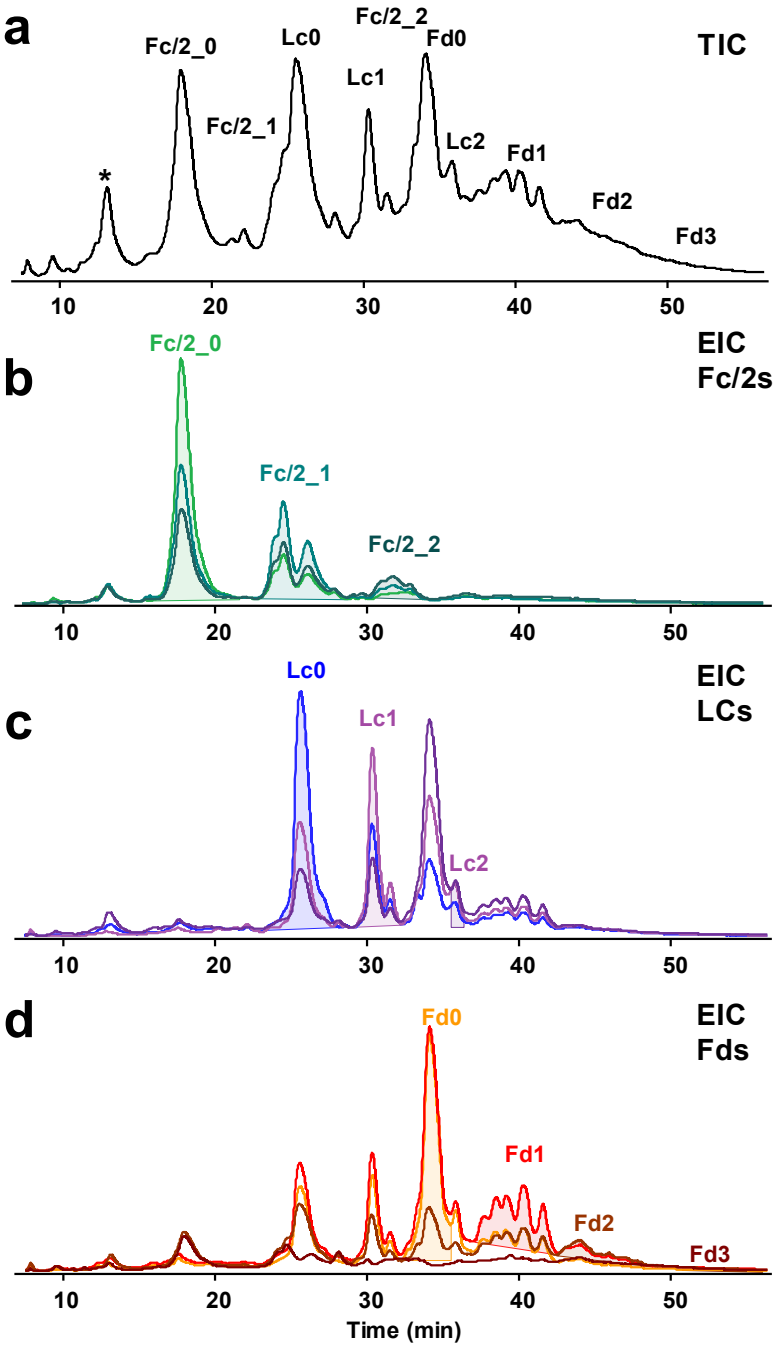
Supplemental Figure 6.10. Online RPLC-MS of partially reduced BV digested by KCP. (a) Total ion chromatogram (TIC) and (b) Extracted ion chromatograms of top-5 charge states of each subunits.



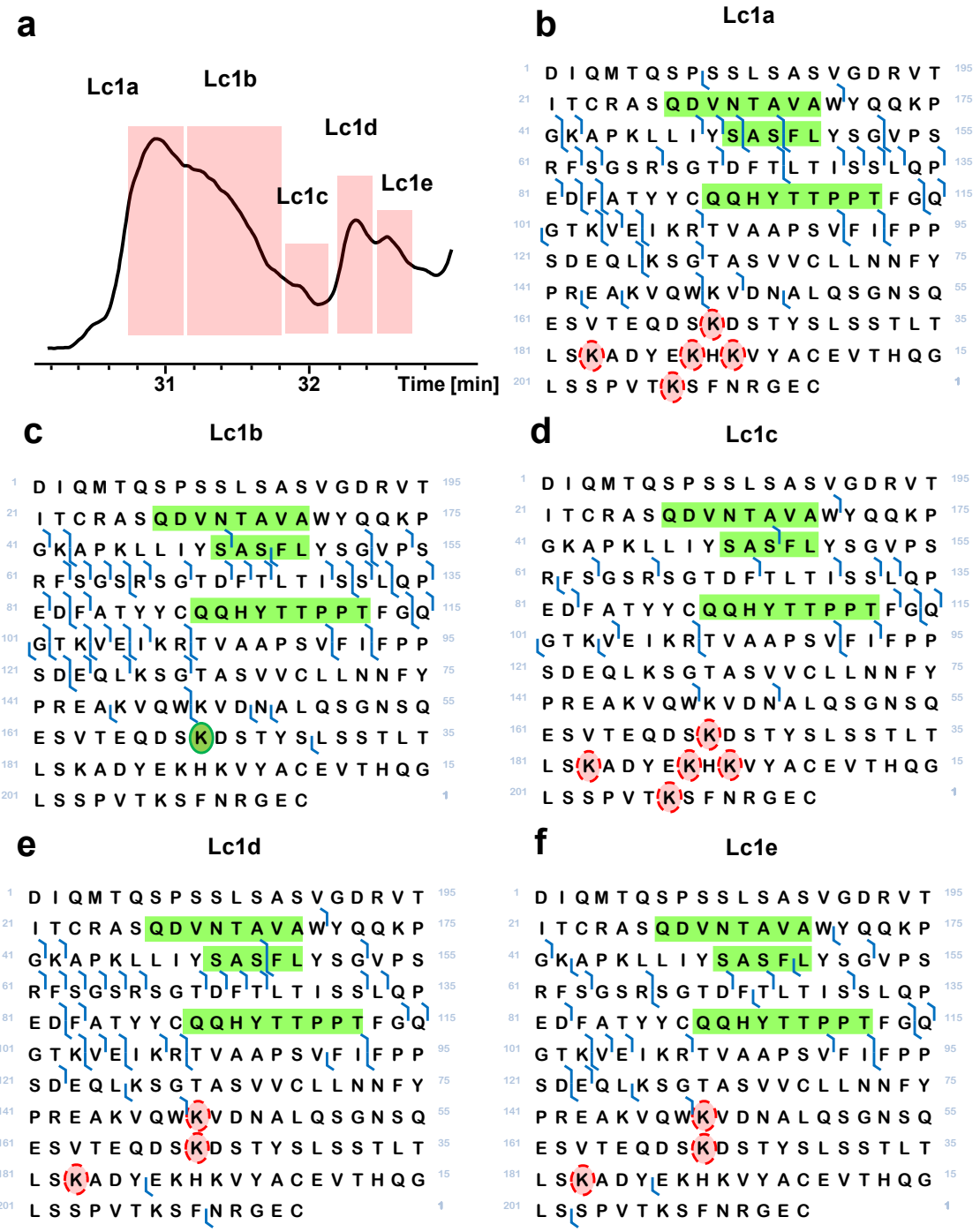
Supplemental Figure 6.11. ETD RPLC-MS/MS characterization of Fc'/2_0 subunit from BV produced by KGP digest. (a) Extracted ion chromatograms of top-5 charge states of Fc'/2_0. (b) ETD fragmentation map of Fc'/2_0 subunit. (c) Fragment ion c_6 demonstrating a possible cleavage between (CPPC) hinge region.



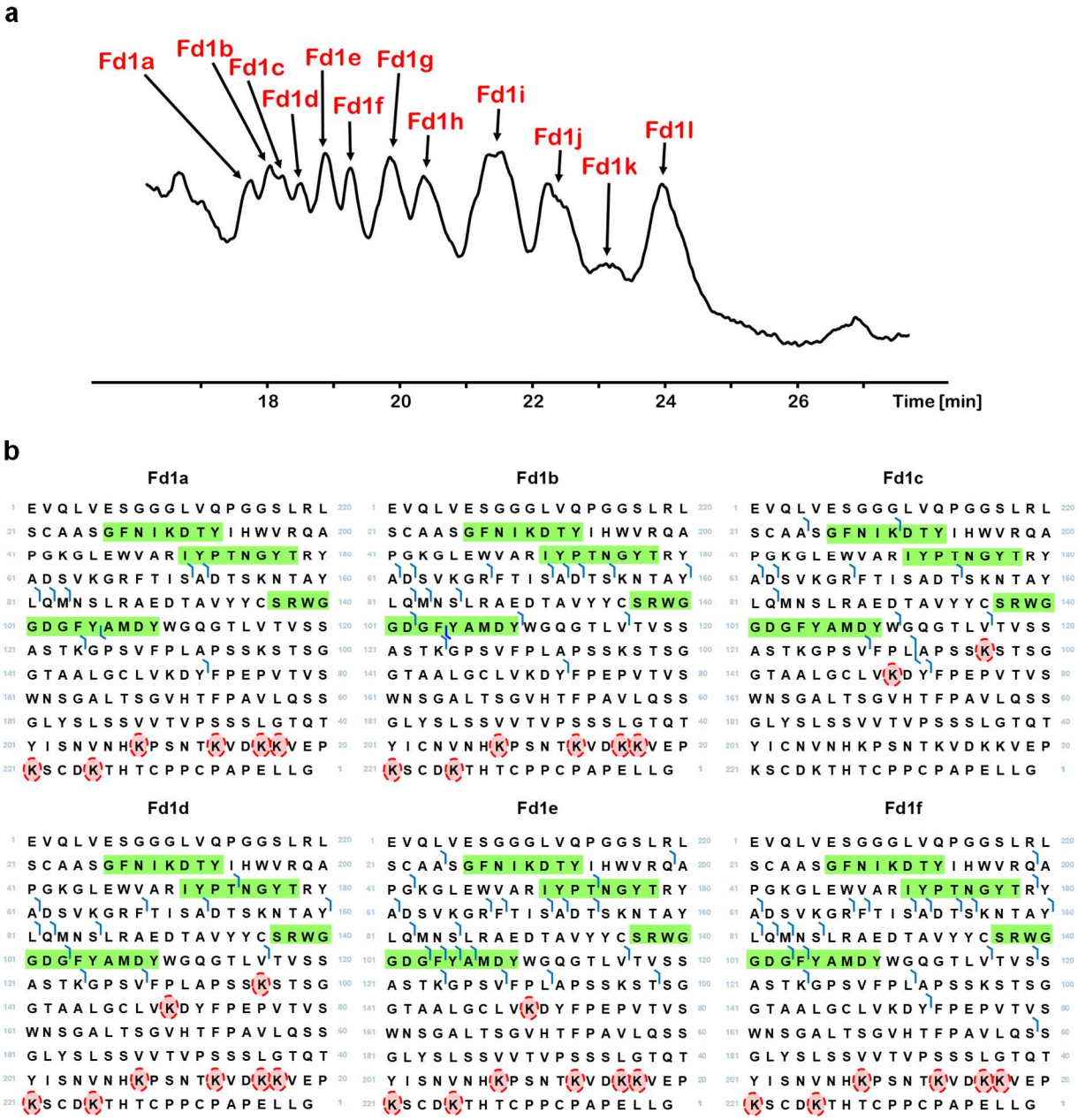
Supplemental Figure 6.12. Representative ETD fragmentation maps of subunits (Fc'2_1, Fc'2_2, Fd'0, and Fd'1) from partially reduced BV digested by KGP. Green circles indicate confident localization of the conjugation sites.



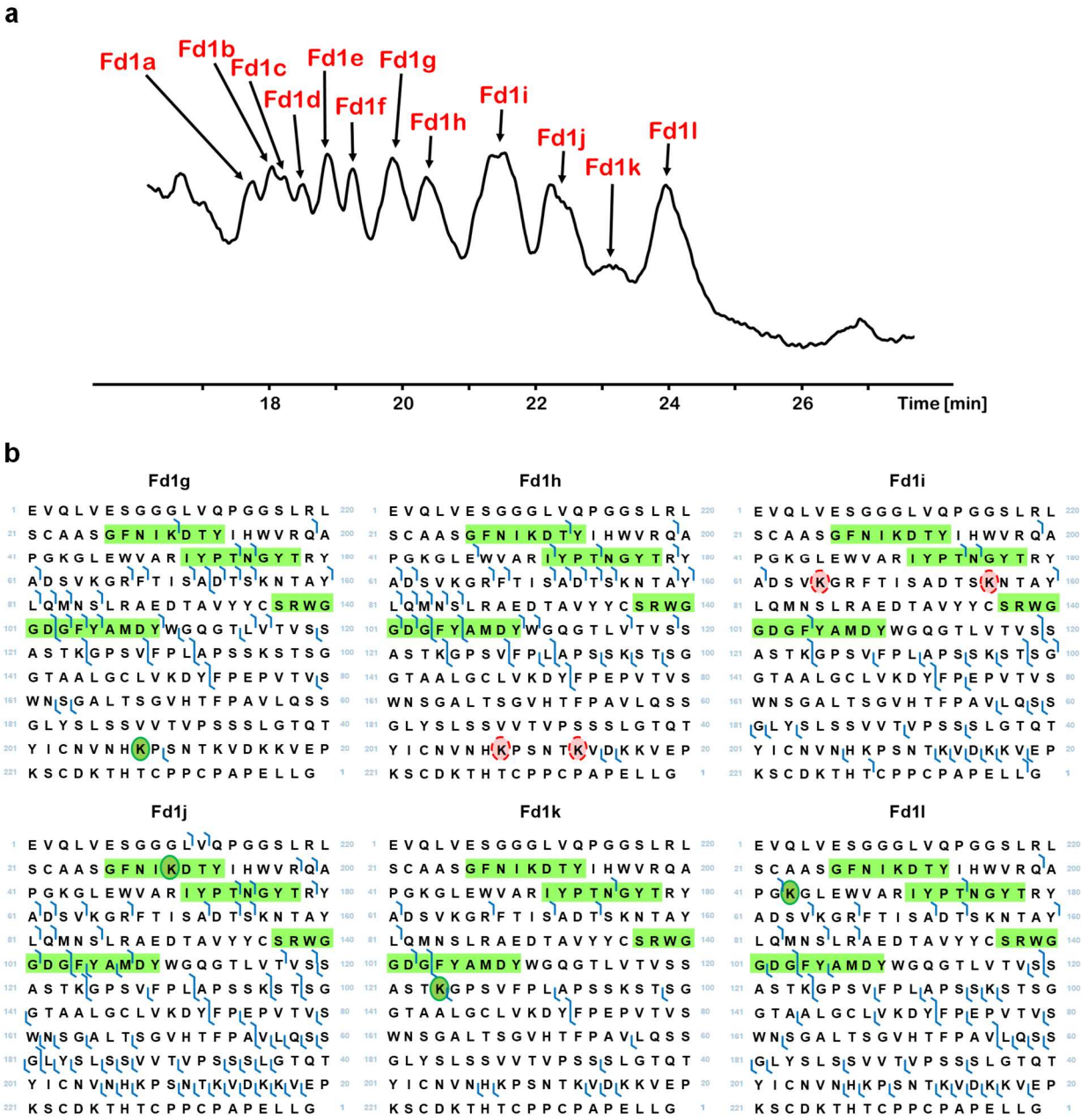
Supplemental Figure 6.13. LC-MS/(MS) characterization of subunits of T-DM1 digested by IdeS. (a) Total ion chromatogram of the subunits. (b) Overlaid extracted ion chromatogram (EIC) of Fc/2 with 0, 1, and 2 drugs (Fc/2_0, Fc/2_1, and Fc/2_2) (c) Overlaid EIC of Lc with 0, 1, and 2 drugs (Lc0, Lc1, and Lc2) (d) Overlaid EIC of Fd with 0, 1, 2, and 3 drugs (Fd0, Fd1, Fd2, and Fd3). * corresponds to a possible truncation with a mass of 20404.23 Da.



Supplemental Figure 6.14. ETD fragmentation maps of light chain with 1 drug (Lc1) isomers from fully reduced IdeS digested T-DM1. Green circles with a solid line indicate confident localization of the conjugation site; and red circles with a dash line indicate ambiguous localization of the single conjugation site. No conjugation sites were found from the N-terminal side of Lc1.



Supplemental Figure 6.15. LC-MS/(MS) with ETD characterizing Fd with 1 drug (Fd1) from fully reduced T-DM1 digested by IdeS. (a) Total ion chromatogram of an ETD LC-MS/MS run showing the Fd1 isomers. (b) ETD fragmentations maps of fd1 with site(s) localized far away from complementary determining regions. Red circles with a dash line indicate ambiguous localization of the single conjugation site.



Supplemental Figure 6.16. LC-MS/(MS) with ETD characterizing Fd with 1 drug (Fd1) from fully reduced T-DM1 digested by IdeS. (a) Total ion chromatogram of an ETD LC-MS/MS run showing the Fd1 isomers. (b) ETD fragmentations maps of fd1 with site(s) with close proximity to complementary determining regions. Green circles with a solid line indicate confident localization of the conjugation site; and red circles with a dash line indicate ambiguous localization of the single conjugation site.

6.7 Acknowledgements

This work was supported by AbbVie Inc. We acknowledge NIH grants, R01HL109810, R01HL096971 GM117058, GM125085 and high-end instrument grant S10OD018475 (to Y.G.). We would also like to thank Dr. Linjie Han and Dr. Shiyue Zhou for helpful discussion and critical reading of the manuscript.

Conclusions and future outlook

In summary, the projects described above provide new methods and tools to analyze intact proteins by top-down mass spectrometry. Using functionalized nanoparticles to enrich low abundance proteins (e.g. phosphoproteins in this work) proves to be a viable way to facilitate comprehensive characterization of proteins of interest. With ligands that have higher specificity and affinity to particular targets and biomarkers, the top-down approach could provide useful inside in many clinical studies given its ability to profile all proteoforms. The development of online HIC-MS opens up doors to exploit native separation based on hydrophobicity, which was previously unexplored with MS compatible mobile phases. This work demonstrates that it is possible to preserve noncovalent interactions with the addition of organic in the presence of volatile salts with moderate concentration. Finally, the top-down (at subunit level) approach shows great potential to obtain useful quality attributes without a prolonged sample preparation time. I envision the top-down approach will be further adopted across all areas and disciplines.

In the past few years, top-down proteomics has gained tremendous momentum especially after the establishment of the Consortium for Top-down Proteomics (http://www.topdownproteomics.org/). Instrument developments and technical advancement in the areas discussed above (sample preparation, separation, tandem MS, data acquisition and analysis, native MS, and quantitation), have equipped the top-down approach with wider dynamic range, higher sensitivity, and better PTM characterization to elucidate proteoforms in a deeper fashion than ever before. Thousands of proteoforms can now be identified, quantified, and characterized in a top-down study.[10,11,93] Nevertheless, efforts are still needed to achieve a complete coverage of all proteoforms.

Despite the significant advancement in top-down hardware and software, some on-going challenges remain. For example, solubilization of hydrophobic proteins, fractionation or separation of proteins with good recovery, detection and characterization of high MW and low-abundance proteins, localization of (labile) PTMs in an accurate and high-throughput manner during LC-MS/MS among others, remain to be further addressed by novel methodologies. In parallel with the technical advancement, we anticipate accelerated and much-needed development of data analysis and bioinformatics tools to meet the growing demands of top-down practitioners, and to fully capitalize on the state-of-the-art instruments and top-down MS workflows.

Of the recent trends, many technological developments have focused on pushing the limit of large-scale proteoform profiling. Indeed, quantitative global profiling can result in specific proteoform targets and pathways that can potentially be linked to disease phenotypes in translational research. However, targeted analysis remains powerful and necessary, as it often leads to untangling of underlying molecular mechanisms of disease-associated cellular events. We expect a renaissance and resurgence of the targeted approach to focus on molecular specificity. Recent attempts to use multiple reaction monitoring to quantify intact proteins or proteoforms exemplify the need for further development of top-down targeted analysis.[179,180]

Is top-down proteomics ready for prime time? The recent advances in the field in all the areas discussed above demonstrate that significant progress has been made toward this goal. Clearly, top-down proteomics has gained remarkable space in the proteomics landscape over the past few years. With the development of high-resolution instrumentation and the maturation of the techniques, the age when top-down proteomics was specialized in a few laboratories, has long passed. Centered around proteoforms, top-down proteomics attracts numerous mass spectrometrists and biomedical researchers alike with its power to decipher the complex proteome

and provide a holistic view on the molecular level, which might better reflect disease phenotypes. Nevertheless, prime time comes with both opportunities and challenges. As the Consortium for Top-Down Proteomics provides the fostering ground for education, collaboration, and development of comprehensive analysis of proteoforms, the field continues to progress and grow. We optimistically envision that more and more laboratories will adopt, practice, and advance top-down proteomics to answer important biological questions and decipher underlying disease mechanisms.

References

(1) Yates, J. R.; Ruse, C. I.; Nakorchevsky, A. *Annu. Rev. Biomed. Eng.* **2009**, *11*, 49-79.
(2) Pandey, A.; Mann, M. *Nature* **2000**, *405*, 837-846.
(3) Altelaar, A. F. M.; Munoz, J.; Heck, A. J. R. *Nat. Rev. Genet.* **2013**, *14*, 35-48.
(4) Gregorich, Z. R.; Ge, Y. *Proteomics* **2014**, *14*, 1195-1210.
(5) Aebersold, R.; Mann, M. *Nature* **2016**, *537*, 347-355.
(6) Smith, L. M.; Kelleher, N. L.; Consortium for Top Down, P. *Nat. Methods* **2013**, *10*, 186-187.
(7) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255-261.
(8) Cai, W. X.; Tucholski, T. M.; Gregorich, Z. R.; Ge, Y. *Expert Rev. Proteomics* **2016**, *13*, 717-730.
(9) Toby, T. K.; Fornelli, L.; Kelleher, N. L. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499-519.
(10) Fornelli, L.; Durbin, K. R.; Fellers, R. T.; Early, B. P.; Greer, J. B.; LeDuc, R. D.; Compton, P. D.; Kelleher, N. L. *J. Proteome Res.* **2017**, *16*, 609-618.
(11) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; Moore, R. J.; Liu, T.; Petyuk, V. A.; Tolić, N.; Paša-Tolić, L.; Smith, R. D.; Payne, S. H.; Kim, S. *Nat. Methods* **2017**, *14*, 909-914.
(12) Brodbelt, J. S. *Anal. Chem.* **2016**, *88*, 30-51.
(13) Nikolaev, E. N.; Boldin, I. A.; Jertz, R.; Baykut, G. *J. Am. Soc. Mass. Spectrom.* **2011**, *22*, 1125-1133.
(14) Shaw, J. B.; Lin, T. Y.; Leach, F. E.; Tolmachev, A. V.; Tolić, N.; Robinson, E. W.; Koppenaal, D. W.; Paša-Tolić, L. *J. Am. Soc. Mass. Spectrom.* **2016**, *27*, 1929-1936.
(15) Weisbrod, C. R.; Kaiser, N. K.; Syka, J. E. P.; Early, L.; Mullen, C.; Dunyach, J. J.; English, A. M.; Anderson, L. C.; Blakney, G. T.; Shabanowitz, J.; Hendrickson, C. L.; Marshall, A. G.; Hunt, D. F. *J. Am. Soc. Mass. Spectrom.* **2017**, *9*, 1787-1795.
(16) Ge, Y.; Rybakova, I. N.; Xu, Q. G.; Moss, R. L. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12658-12663.
(17) Chamot-Rooke, J.; Mikaty, G.; Malosse, C.; Soyer, M.; Dumont, A.; Gault, J.; Imhaus, A. F.; Martin, P.; Trellet, M.; Clary, G.; Chafey, P.; Camoin, L.; Nilges, M.; Nassif, X.; Dumenil, G. *Science* **2011**, *331*, 778-782.
(18) Zhou, H.; Ning, Z.; Starr, A. E.; Abu-Farha, M.; Figeys, D. *Anal. Chem.* **2012**, *84*, 720-734.
(19) Savaryn, J. P.; Catherman, A. D.; Thomas, P. M.; Abecassis, M. M.; Kelleher, N. L. *Genome Med.* **2013**, *5*, 53.
(20) Kelleher, N. L.; Thomas, P. M.; Ntai, I.; Compton, P. D.; LeDuc, R. D. *Expert Rev. Proteomics* **2014**, *11*, 649-651.
(21) Pan, P.; McLuckey, S. A. *Anal. Chem.* **2003**, *75*, 5468-5474.
(22) Metwally, H.; McAllister, R. G.; Konermann, L. *Anal. Chem.* **2015**, *87*, 2434-2442.
(23) Laganowsky, A.; Reading, E.; Hopper, J. T. S.; Robinson, C. V. *Nat. Protoc.* **2013**, *8*, 639-651.
(24) Pohl, T.; Kamp, R. M. *Anal. Biochem.* **1987**, *160*, 388-391.
(25) Zhang, H.; Ge, Y. *Circ. Cardiovasc. Genet.* **2011**, *4*, 711.
(26) Susa, A. C.; Xia, Z.; Williams, E. R. *Anal. Chem.* **2017**, *89*, 3116-3122.
(27) Speers, A. E.; Wu, C. C. *Chem. Rev.* **2007**, *107*, 3687-3714.
(28) Catherman, A. D.; Li, M.; Tran, J. C.; Durbin, K. R.; Compton, P. D.; Early, B. P.; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2013**, *85*, 1880-1888.

(29) Skinner, O. S.; Catherman, A. D.; Early, B. P.; Thomas, P. M.; Compton, P. D.; Kelleher, N. L. *Anal. Chem.* **2014**, *86*, 4627-4634.
(30) Loo, R. R.; Dales, N.; Andrews, P. C. *Protein Sci.* **1994**, *3*, 1975-1983.
(31) Wessel, D.; Flugge, U. I. *Anal. Biochem.* **1984**, *138*, 141-143.
(32) Doucette, A. A.; Vieira, D. B.; Orton, D. J.; Wall, M. J. *J. Proteome Res.* **2014**, *13*, 6001-6012.
(33) Moore, S. M.; Hess, S. M.; Jorgenson, J. W. *J. Proteome Res.* **2016**, *15*, 1243-1252.
(34) Whitelegge, J. P.; Gundersen, C. B.; Faull, K. F. *Protein Sci.* **1998**, *7*, 1423-1430.
(35) Hengel, S. M.; Floyd, E.; Baker, E. S.; Zhao, R.; Wu, S.; Paša-Tolić, L. *Proteomics* **2012**, *12*, 3138-3142.
(36) Crowell, A. M.; MacLellan, D. L.; Doucette, A. A. *J. Proteomics* **2015**, *118*, 140-150.
(37) Kachuk, C.; Faulkner, M.; Liu, F.; Doucette, A. A. *J. Proteome Res.* **2016**, *15*, 2634-2642.
(38) Kim, K. H.; Compton, P. D.; Tran, J. C.; Kelleher, N. L. *J. Proteome Res.* **2015**, *14*, 2199-2206.
(39) Chang, Y. H.; Gregorich, Z. R.; Chen, A. J.; Hwang, L.; Guner, H.; Yu, D.; Zhang, J.; Ge, Y. *J. Proteome Res.* **2015**, *14*, 1587-1599.
(40) Meng, F. Y.; Cargile, B. J.; Patrie, S. M.; Johnson, J. R.; McLoughlin, S. M.; Kelleher, N. L. *Anal. Chem.* **2002**, *74*, 2923-2929.
(41) Barrera, N. P.; Robinson, C. V. *Annu. Rev. Biochem* **2011**, *80*, 247-271.
(42) Yen, H. Y.; Hopper, J. T. S.; Liko, I.; Allison, T. M.; Zhu, Y.; Wang, D. J.; Stegmann, M.; Mohammed, S.; Wu, B. L.; Robinson, C. V. *Science Advances* **2017**, *3*, e1701016.
(43) Heck, A. J. R. *Nat. Methods* **2008**, *5*, 927-933.
(44) Carroll, J.; Fearnley, I. M.; Walker, J. E. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16170-16175.
(45) Waas, M.; Bhattacharya, S.; Chuppa, S.; Wu, X.; Jensen, D. R.; Omasits, U.; Wollscheid, B.; Volkman, B. F.; Noon, K. R.; Gundry, R. L. *Anal. Chem.* **2014**, *86*, 1551-1559.
(46) Cox, B.; Emili, A. *Nat. Protoc.* **2006**, *1*, 1872-1878.
(47) Catherman, A. D.; Durbin, K. R.; Ahlf, D. R.; Early, B. P.; Fellers, R. T.; Tran, J. C.; Thomas, P. M.; Kelleher, N. L. *Mol. Cell. Proteomics* **2013**, *12*, 3465-3473.
(48) Zhou, M. W.; Wu, S.; Stenoien, D. L.; Zhang, Z. R.; Connolly, L.; Freitag, M.; Paša-Tolić, L. In *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Wajapeyee, N.; Gupta, R., Eds., 2017, pp 153-168.
(49) Fei, R.; Zhang, T.; Huang, Y.; Hu, Y. *Anal. Chim. Acta* **2017**, *986*, 161-170.
(50) He, Y. T.; Liu, W.; Chen, L.; Lin, G.; Xiao, Q.; Gao, C. L.; Wu, J. L.; Lin, Z. *J. Sep. Sci.* **2017**, *40*, 1516-1523.
(51) Hwang, L.; Ayaz-Guner, S.; Gregorich, Z. R.; Cai, W.; Valeja, S. G.; Jin, S.; Ge, Y. *J. Am. Chem. Soc.* **2015**, *137*, 2432-2435.
(52) Chen, B.; Hwang, L.; Ochowicz, W.; Lin, Z.; Guardado-Alvarez, T. M.; Cai, W.; Xiu, L.; Dani, K.; Colah, C.; Jin, S.; Ge, Y. *Chem. Sci.* **2017**, *8*, 4306-4311.
(53) Zhang, J.; Guy, M. J.; Norman, H. S.; Chen, Y. C.; Xu, Q. G.; Dong, X. T.; Guner, H.; Wang, S. J.; Kohmoto, T.; Young, K. H.; Moss, R. L.; Ge, Y. *J. Proteome Res.* **2011**, *10*, 4054-4065.
(54) Xu, F. M.; Xu, Q. G.; Dong, X. T.; Guy, M.; Guner, H.; Hacker, T. A.; Ge, Y. *Int. J. Mass spectrom.* **2011**, *305*, 95-102.
(55) Dong, X. T.; Sumandea, C. A.; Chen, Y. C.; Garcia-Cazarin, M. L.; Zhang, J.; Balke, C. W.; Sumandea, M. P.; Ge, Y. *J. Biol. Chem.* **2012**, *287*, 848-857.
(56) Peng, Y.; Chen, X.; Sato, T.; Rankin, S. A.; Tsuji, R. F.; Ge, Y. *Anal. Chem.* **2012**, *84*, 3339-3346.

(57) Burnaevskiy, N.; Fox, T. G.; Plymire, D. A.; Ertelt, J. M.; Weigele, B. A.; Selyunin, A. S.; Way, S. S.; Patrie, S. M.; Alto, N. M. *Nature* **2013**, *496*, 106-109.
(58) Savaryn, J. P.; Skinner, O. S.; Fornelli, L.; Fellers, R. T.; Compton, P. D.; Terhune, S. S.; Abecassis, M. M.; Kelleher, N. L. *J. Proteomics* **2016**, *134*, 76-84.
(59) Doucette, A. A.; Tran, J. C.; Wall, M. J.; Fitzsimmons, S. *Expert Rev. Proteomics* **2011**, *8*, 787-800.
(60) Capriotti, A. L.; Cavaliere, C.; Foglia, P.; Samperi, R.; Lagana, A. *J. Chromatogr. A* **2011**, *1218*, 8760-8776.
(61) Tran, J. C.; Doucette, A. A. *Anal. Chem.* **2008**, *80*, 1568-1573.
(62) Vellaichamy, A.; Tran, J. C.; Catherman, A. D.; Lee, J. E.; Kellie, J. F.; Sweet, S. M.; Zamdborg, L.; Thomas, P. M.; Ahlf, D. R.; Durbin, K. R.; Valaskovic, G. A.; Kelleher, N. L. *Anal. Chem.* **2010**, *82*, 1234-1244.
(63) Sun, L.; Knierman, M. D.; Zhu, G.; Dovichi, N. J. *Anal. Chem.* **2013**, *85*, 5989-5995.
(64) Xiu, L. C.; Valeja, S. G.; Alpert, A. J.; Jin, S.; Ge, Y. *Anal. Chem.* **2014**, *86*, 7899-7906.
(65) Rogers, B. A.; Wu, Z.; Wei, B.; Zhang, X.; Cao, X.; Alabi, O.; Wirth, M. J. *Anal. Chem.* **2015**, *87*, 2520-2526.
(66) Valeja, S. G.; Xiu, L. C.; Gregorich, Z. R.; Guner, H.; Jin, S.; Ge, Y. *Anal. Chem.* **2015**, *87*, 5363-5371.
(67) Chen, B. F.; Peng, Y.; Valeja, S. G.; Xiu, L. C.; Alpert, A. J.; Ge, Y. *Anal. Chem.* **2016**, *88*, 1885-1891.
(68) Shen, Y.; Tolić, N.; Piehowski, P. D.; Shukla, A. K.; Kim, S.; Zhao, R.; Qu, Y.; Robinson, E.; Smith, R. D.; Paša-Tolić, L. *J. Chromatogr. A* **2017**, *1498*, 99-110.
(69) Cai, W.; Tucholski, T.; Chen, B.; Alpert, A. J.; McIlwain, S.; Kohmoto, T.; Jin, S.; Ge, Y. *Anal. Chem.* **2017**, *89*, 5467-5475.
(70) Zhen, W.; Bingchuan, W.; Ximo, Z.; Wirth, M. J. *Anal. Chem.* **2014**, *86*, 1592–1598.
(71) Eeltink, S.; Wouters, B.; Desmet, G.; Ursem, M.; Blinco, D.; Kemp, G. D.; Treumann, A. *J. Chromatogr. A* **2011**, *1218*, 5504-5511.
(72) Guiochon, G. *J. Chromatogr. A* **2007**, *1168*, 101-168.
(73) Simone, P.; Pierri, G.; Foglia, P.; Gasparrini, F.; Mazzoccanti, G.; Capriotti, A. L.; Ursini, O.; Ciogli, A.; Lagana, A. *J. Sep. Sci.* **2016**, *39*, 264-271.
(74) Muneeruddin, K.; Thomas, J. J.; Salinas, P. A.; Kaltashov, I. A. *Anal. Chem.* **2014**, *86*, 10692-10699.
(75) Muneeruddin, K.; Nazzaro, M.; Kaltashov, I. A. *Anal. Chem.* **2015**, *87*, 10138-10145.
(76) Muneeruddin, K.; Bobst, C. E.; Frenkel, R.; Houde, D.; Turyan, I.; Sosic, Z.; Kaltashov, I. A. *Analyst* **2017**, *142*, 336-344.
(77) Valaskovic, G. A.; Kelleher, N. L.; McLafferty, F. W. *Science* **1996**, *273*, 1199-1202.
(78) Zhou, F.; Johnston, M. V. *Anal. Chem.* **2004**, *76*, 2734-2740.
(79) Haselberg, R.; de Jong, G. J.; Somsen, G. W. *Anal. Chem.* **2013**, *85*, 2289-2296.
(80) Zhao, Y.; Sun, L.; Champion, M. M.; Knierman, M. D.; Dovichi, N. J. *Anal. Chem.* **2014**, *86*, 4873-4878.
(81) Han, X. M.; Wang, Y. J.; Aslanian, A.; Fonslow, B.; Graczyk, B.; Davis, T. N.; Yates, J. R. *J. Proteome Res.* **2014**, *13*, 6078-6086.
(82) Zhao, Y.; Riley, N. M.; Sun, L.; Hebert, A. S.; Yan, X.; Westphall, M. S.; Rush, M. J.; Zhu, G.; Champion, M. M.; Mba Medie, F.; Champion, P. A.; Coon, J. J.; Dovichi, N. J. *Anal. Chem.* **2015**, *87*, 5422-5429.

(83) Bush, D. R.; Zang, L.; Belov, A. M.; Ivanov, A. R.; Karger, B. L. *Anal. Chem.* **2016**, *88*, 1138-1146.
(84) Stepanova, S.; Kasicka, V. *Anal. Chim. Acta* **2016**, *933*, 23-42.
(85) Zhao, Y.; Sun, L.; Zhu, G.; Dovichi, N. J. *J. Proteome Res.* **2016**, *15*, 3679-3685.
(86) Wang, Z.; Ma, H.; Smith, K.; Wu, S. *Int. J. Mass spectrom.* **2017**, *427*, 43-51.
(87) Compton, P. D.; Zamdborg, L.; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2011**, *83*, 6868-6874.
(88) 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. X.; Wu, C.; Sweet, S. M. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. *Nature* **2011**, *480*, 254-258.
(89) Catherman, A. D.; Durbin, K. R.; Ahlf, D. R.; Early, B. P.; Fellers, R. T.; Tran, J. C.; Thomas, P. M.; Kelleher, N. L. *Mol. Cell. Proteomics* **2013**, *12*, 3465-3473.
(90) Morrison, L. J.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2016**, *138*, 10849-10859.
(91) Tian, Z. X.; Tolić, N.; Zhao, R.; Moore, R. J.; Hengel, S. M.; Robinson, E. W.; Stenoien, D. L.; Wu, S.; Smith, R. D.; Paša-Tolić, L. *Genome Biol.* **2012**, *13*, R86.
(92) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nat. Methods* **2007**, *4*, 709-712.
(93) Durbin, K. R.; Fornelli, L.; Fellers, R. T.; Doubleday, P. F.; Narita, M.; Kelleher, N. L. *J. Proteome Res.* **2016**, *15*, 976-982.
(94) Siuti, N.; Kelleher, N. L. *Nat. Methods* **2007**, *4*, 817-821.
(95) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 563-573.
(96) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9528-9533.
(97) Peng, Y.; Gregorich, Z. R.; Valeja, S. G.; Zhang, H.; Cai, W. X.; Chen, Y. C.; Guner, H.; Chen, A. J.; Schwahn, D. J.; Hacker, T. A.; Liu, X. W.; Ge, Y. *Mol. Cell. Proteomics* **2014**, *13*, 2752-2764.
(98) Gregorich, Z. R.; Peng, Y.; Lane, N. M.; Wolff, J. J.; Wang, S. J.; Guo, W.; Guner, H.; Doop, J.; Hacker, T. A.; Ge, Y. *J. Mol. Cell. Cardiol.* **2015**, *87*, 102-112.
(99) Jin, Y. T.; Peng, Y.; Lin, Z. Q.; Chen, Y. C.; Wei, L. M.; Hacker, T. A.; Larsson, L.; Ge, Y. *J. Muscle Res. Cell Motil.* **2016**, *37*, 41-52.
(100) Yu, D. Y.; Peng, Y.; Ayaz-Guner, S.; Gregorich, Z. R.; Ge, Y. *J. Am. Soc. Mass. Spectrom.* **2016**, *27*, 220-232.
(101) Pan, J. X.; Zhang, S. P.; Borchers, C. H. *J. Proteomics* **2016**, *134*, 138-143.
(102) Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O. *Mol. Cell. Proteomics* **2012**, *11*, 1758-1767.
(103) Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. *Anal. Chem.* **2013**, *85*, 4239-4246.
(104) Tran, B. Q.; Barton, C.; Feng, J. H.; Sandjong, A.; Yoon, S. H.; Awasthi, S.; Liang, T.; Khan, M. M.; Kilgour, D. P. A.; Goodlett, D. R.; Goo, Y. A. *J. Proteomics* **2016**, *134*, 93-101.
(105) He, L. D.; Anderson, L. C.; Barnidge, D. R.; Murray, D. L.; Hendrickson, C. L.; Marshall, A. G. *J. Am. Soc. Mass. Spectrom.* **2017**, *28*, 827-838.
(106) Fornelli, L.; Ayoub, D.; Aizikov, K.; Liu, X. W.; Damoc, E.; Pevzner, P. A.; Makarov, A.; Beck, A.; Tsybin, Y. O. *J. Proteomics* **2017**, *159*, 67-76.
(107) Brunner, A. M.; Lossl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. *Anal. Chem.* **2015**, *87*, 4152-4158.

(108) Riley, N. M.; Westphall, M. S.; Coon, J. J. *Anal. Chem.* **2015**, *87*, 7109-7116.
(109) Riley, N. M.; Westphall, M. S.; Coon, J. J. *J. Proteome Res.* **2017**, *16*, 2653-2659.
(110) Horn, D. M.; Ge, Y.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 4778-4784.
(111) Shaw, J. B.; Li, W. Z.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2013**, *135*, 12646-12651.
(112) Thyer, R.; Robotham, S. A.; Brodbelt, J. S.; Ellington, A. D. *J. Am. Chem. Soc.* **2015**, *137*, 46-49.
(113) Cammarata, M. B.; Thyer, R.; Rosenberg, J.; Ellington, A.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2015**, *137*, 9128-9135.
(114) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. *Anal. Chem.* **2016**, *88*, 1008-1016.
(115) Mayfield, J. E.; Robinson, M. R.; Cotham, V. C.; Irani, S.; Matthews, W. L.; Ram, A.; Gilmour, D. S.; Cannon, J. R.; Zhang, Y. J.; Brodbelt, J. S. *ACS Chem. Biol.* **2017**, *12*, 153-162.
(116) Cleland, T. P.; DeHart, C. J.; Fellers, R. T.; VanNispen, A. J.; Greer, J. B.; Leduc, R. D.; Parker, W. R.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Proteome Res.* **2017**, *16*, 2072-2079.
(117) McLuckey, S. A.; Reid, G. E.; Wells, J. M. *Anal. Chem.* **2002**, *74*, 336-346.
(118) Earley, L.; Anderson, L. C.; Bai, D. L.; Mullen, C.; Syka, J. E.; English, A. M.; Dunyach, J. J.; Stafford, G. C., Jr.; Shabanowitz, J.; Hunt, D. F.; Compton, P. D. *Anal. Chem.* **2013**, *85*, 8385-8390.
(119) Anderson, L. C.; Karch, K. R.; Ugrin, S. A.; Coradin, M.; English, A. M.; Sidoli, S.; Shabanowitz, J.; Garcia, B. A.; Hunt, D. F. *Mol. Cell. Proteomics* **2016**, *15*, 975-988.
(120) Riley, N. M.; Mullen, C.; Weisbrod, C. R.; Sharma, S.; Senko, M. W.; Zabrouskov, V.; Westphall, M. S.; Syka, J. E. P.; Coon, J. J. *J. Am. Soc. Mass. Spectrom.* **2016**, *27*, 520-531.
(121) Holden, D. D.; Brodbelt, J. S. *Anal. Chem.* **2017**, *89*, 837-846.
(122) Durbin, K. R.; Fellers, R. T.; Ntai, I.; Kelleher, N. L.; Compton, P. D. *Anal. Chem.* **2014**, *86*, 1485-1492.
(123) Hofstadler, S. A.; Bruce, J. E.; Rockwood, A. L.; Anderson, G. A.; Winger, B. E.; Smith, R. D. *Int. J. Mass Spectrom. Ion Processes* **1994**, *132*, 109-127.
(124) Anderson, L. C.; DeHart, C. J.; Kaiser, N. K.; Fellers, R. T.; Smith, D. F.; Greer, J. B.; LeDuc, R. D.; Blakney, G. T.; Thomas, P. M.; Kelleher, N. L.; Hendrickson, C. L. *J. Proteome Res.* **2017**, *16*, 1087-1096.
(125) Kaiser, N. K.; Savory, J. J.; Hendrickson, C. L. *J. Am. Soc. Mass. Spectrom.* **2014**, *25*, 943-949.
(126) Liu, X.; Inbar, Y.; Dorrestein, P. C.; Wynne, C.; Edwards, N.; Souda, P.; Whitelegge, J. P.; Bafna, V.; Pevzner, P. A. *Mol. Cell. Proteomics* **2010**, *9*, 2772-2782.
(127) Kou, Q.; Wu, S.; Liu, X. *BMC Genomics* **2014**, *15*, 1140.
(128) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. *Proteomics* **2015**, *15*, 1235-1238.
(129) Cai, W. X.; Guner, H.; Gregorich, Z. R.; Chen, A. J.; Ayaz-Guner, S.; Peng, Y.; Valeja, S. G.; Liu, X. W.; Ge, Y. *Mol. Cell. Proteomics* **2016**, *15*, 703-714.
(130) Xiao, K. J.; Yu, F.; Tian, Z. X. *J. Proteomics* **2017**, *152*, 41-47.
(131) Sun, R. X.; Luo, L.; Wu, L.; Wang, R. M.; Zeng, W. F.; Chi, H.; Liu, C.; He, S. M. *Anal. Chem.* **2016**, *88*, 3082-3090.
(132) Kou, Q.; Xun, L. K.; Liu, X. W. *Bioinformatics* **2016**, *32*, 3495-3497.
(133) Petrotchenko, E. V.; Borchers, C. H. *J. Am. Soc. Mass. Spectrom.* **2015**, *26*, 1895-1898.

(134) Vyatkina, K.; Wu, S.; Dekker, L. J. M.; VanDuijn, M. M.; Liu, X. W.; Tolić, N.; Luider, T. M.; Paša-Tolić, L.; Pevzner, P. A. *Bioinformatics* **2016**, *32*, 2753-2759.
(135) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass. Spectrom.* **2000**, *11*, 320-332.
(136) Jaitly, N.; Mayampurath, A.; Littlefield, K.; Adkins, J. N.; Anderson, G. A.; Smith, R. D. *BMC Bioinformatics* **2009**, *10*, 87.
(137) Mayampurath, A. M.; Jaitly, N.; Purvine, S. O.; Monroe, M. E.; Auberry, K. J.; Adkins, J. N.; Smith, R. D. *Bioinformatics* **2008**, *24*, 1021-1023.
(138) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. *Nucleic Acids Res.* **2007**, *35*, W701-W706.
(139) Guner, H.; Close, P. L.; Cai, W. X.; Zhang, H.; Peng, Y.; Gregorich, Z. R.; Ge, Y. *J. Am. Soc. Mass. Spectrom.* **2014**, *25*, 464-470.
(140) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K.; Benesch, J. L.; Robinson, C. V. *Anal. Chem.* **2015**, *87*, 4370-4376.
(141) Li, L.; Tian, Z. X. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 1267-1277.
(142) Liu, X. W.; Sirotkin, Y.; Shen, Y. F.; Anderson, G.; Tsai, Y. S.; Ting, Y. S.; Goodlett, D. R.; Smith, R. D.; Bafna, V.; Pevzner, P. A. *Mol. Cell. Proteomics* **2012**, *11*.
(143) Tsai, Y. S.; Scherl, A.; Shaw, J. L.; MacKay, C. L.; Shaffer, S. A.; Langridge-Smith, P. R. R.; Goodlett, D. R. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 2154-2166.
(144) Liu, X. W.; Hengel, S.; Wu, S.; Tolić, N.; Paša-Tolić, L.; Pevzner, P. A. *J. Proteome Res.* **2013**, *12*, 5830-5838.
(145) Kou, Q.; Zhu, B. H.; Wu, S.; Ansong, C.; Tolić, N.; Paša-Tolić, L.; Liu, X. W. *J. Proteome Res.* **2016**, *15*, 2422-2432.
(146) Kou, Q.; Wu, S.; Tolić, N.; Paša-Tolić, L.; Liu, Y. L.; Liu, X. W. *Bioinformatics* **2017**, *33*, 1309-1316.
(147) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. *J. Proteome Res.* **2014**, *13*, 3231-3240.
(148) Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L. *J. Am. Soc. Mass. Spectrom.* **2015**, *26*, 782-787.
(149) Belov, A. M.; Viner, R.; Santos, M. R.; Horn, D. M.; Bern, M.; Karger, B. L.; Ivanov, A. R. *J. Am. Soc. Mass. Spectrom.* **2017**.
(150) Belov, M. E.; Damoc, E.; Denisov, E.; Compton, P. D.; Horning, S.; Makarov, A. A.; Kelleher, N. L. *Anal. Chem.* **2013**, *85*, 11163-11173.
(151) Skinner, O. S.; Havugimana, P. C.; Haverland, N. A.; Fornelli, L.; Early, B. P.; Greer, J. B.; Fellers, R. T.; Durbin, K. R.; Do Vale, L. H. F.; Melani, R. D.; Seckler, H. S.; Nelp, M. T.; Belov, M. E.; Horning, S. R.; Makarov, A. A.; LeDuc, R. D.; Bandarian, V.; Compton, P. D.; Kelleher, N. L. *Nat. Methods* **2016**, *13*, 237-240.
(152) Haverland, N. A.; Skinner, O. S.; Fellers, R. T.; Tariq, A. A.; Early, B. P.; Leduc, R. D.; Fornelli, L.; Compton, P. D.; Kelleher, N. L. *J. Am. Soc. Mass. Spectrom.* **2017**, *28*, 1203-1215.
(153) Li, H. L.; Wolff, J. J.; Van Orden, S. L.; Loo, J. A. *Anal. Chem.* **2014**, *86*, 317-320.
(154) Schneeberger, E. M.; Breuker, K. *Angew. Chem. Int. Ed.* **2017**, *56*, 1254-1258.
(155) Yan, J.; Zhou, M.; Gilbert, J. D.; Wolff, J. J.; Somogyi, A.; Pedder, R. E.; Quintyn, R. S.; Morrison, L. J.; Easterling, M. L.; Paša-Tolić, L.; Wysocki, V. H. *Anal. Chem.* **2017**, *89*, 895-901.
(156) Skinner, O. S.; McAnally, M. O.; Van Duyne, R. P.; Schatz, G. C.; Breuker, K.; Compton, P. D.; Kelleher, N. L. *Anal. Chem.* **2017**, *89*, 10711-10716.
(157) Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. *Anal. Chem.* **2006**, *78*, 4271-4280.

(158) Ansong, C.; Wu, S.; Meng, D.; Liu, X. W.; Brewer, H. M.; Kaiser, B. L. D.; Nakayasu, E. S.; Cort, J. R.; Pevzner, P.; Smith, R. D.; Heffron, F.; Adkins, J. N.; Paša-Tolić, L. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 10153-10158.
(159) Dang, X. B.; Singh, A.; Spetman, B. D.; Nolan, K. D.; Isaacs, J. S.; Dennis, J. H.; Dalton, S.; Marshall, A. G.; Young, N. L. *J. Proteome Res.* **2016**, *15*, 3196-3203.
(160) Rea, M.; Jiang, T. T.; Eleazer, R.; Eckstein, M.; Marshall, A. G.; Fondufe-Mittendorf, Y. N. *Mol. Cell. Proteomics* **2016**, *15*, 2411-2422.
(161) Zheng, Y. P.; Fornelli, L.; Compton, P. D.; Sharma, S.; Canterbury, J.; Mullen, C.; Zabrouskov, V.; Fellers, R. T.; Thomas, P. M.; Licht, J. D.; Senko, M. W.; Kelleher, N. L. *Mol. Cell. Proteomics* **2016**, *15*, 776-790.
(162) Stefanowicz, P.; Kijewska, M.; Szewczuk, Z. *Anal. Chem.* **2014**, *86*, 7247-7251.
(163) Chen, B. F.; Guo, X.; Tucholski, T.; Lin, Z. Q.; McIlwain, S.; Ge, Y. *J. Am. Soc. Mass. Spectrom.* **2017**, *28*, 1805-1814.
(164) Collier, T. S.; Sarkar, P.; Rao, B.; Muddiman, D. C. *J. Am. Soc. Mass. Spectrom.* **2010**, *21*, 879-889.
(165) Wiese, S.; Reidegeld, K. A.; Meyer, H. E.; Warscheid, B. *Proteomics* **2007**, *7*, 340-350.
(166) Hung, C. W.; Tholey, A. *Anal. Chem.* **2012**, *84*, 161-170.
(167) Fang, H. Q.; Xiao, K. J.; Li, Y. H.; Yu, F.; Liu, Y.; Xue, B. B.; Tian, Z. X. *Anal. Chem.* **2016**, *88*, 7198-7205.
(168) Rhoads, T. W.; Rose, C. M.; Bailey, D. J.; Riley, N. M.; Molden, R. C.; Nestler, A. J.; Merrill, A. E.; Smith, L. M.; Hebert, A. S.; Westphall, M. S.; Pagliarini, D. J.; Garcia, B. A.; Coon, J. J. *Anal. Chem.* **2014**, *86*, 2314-2319.
(169) Shortreed, M. R.; Frey, B. L.; Scalf, M.; Knoener, R. A.; Cesnik, A. J.; Smith, L. M. *J. Proteome Res.* **2016**, *15*, 1213-1221.
(170) Dai, Y.; Shortreed, M. R.; Scalf, M.; Frey, B. L.; Cesnik, A. J.; Solntsev, S.; Schaffer, L. V.; Smith, L. M. *J. Proteome Res.* **2017**, *16*, 4156-4165.
(171) Waanders, L. F.; Hanke, S.; Mann, M. *J. Am. Soc. Mass. Spectrom.* **2007**, *18*, 2058-2064.
(172) Geis-Asteggiante, L.; Ostrand-Rosenberg, S.; Fenselau, C.; Edwards, N. J. *Anal. Chem.* **2016**, *88*, 10900-10907.
(173) Quijada, J. V.; Schmitt, N. D.; Salisbury, J. P.; Auclair, J. R.; Agar, J. N. *Anal. Chem.* **2016**, *88*, 11139-11146.
(174) Mazur, M. T.; Cardasis, H. L.; Spellman, D. S.; Liaw, A.; Yates, N. A.; Hendrickson, R. C. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 7728-7733.
(175) Wu, S.; Brown, J. N.; Tolić, N.; Meng, D.; Liu, X.; Zhang, H.; Zhao, R.; Moore, R. J.; Pevzner, P.; Smith, R. D.; Paša-Tolić, L. *Proteomics* **2014**, *14*, 1211-1222.
(176) Ntai, I.; Kim, K.; Fellers, R. T.; Skinner, O. S.; Smith, A. D. t.; Early, B. P.; Savaryn, J. P.; LeDuc, R. D.; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2014**, *86*, 4961-4968.
(177) Ntai, I.; LeDuc, R. D.; Fellers, R. T.; Erdmann-Gilmore, P.; Davies, S. R.; Rumsey, J.; Early, B. P.; Thomas, P. M.; Li, S.; Compton, P. D.; Ellis, M. J.; Ruggles, K. V.; Fenyo, D.; Boja, E. S.; Rodriguez, H.; Townsend, R. R.; Kelleher, N. L. *Mol. Cell. Proteomics* **2016**, *15*, 45-56.
(178) Schmit, P. O.; Vialaret, J.; Wessels, H. J. C. T.; van Gool, A. J.; Lehmann, S.; Gabelle, A.; Wood, J.; Bern, M.; Paape, R.; Suckau, D.; Kruppa, G.; Hirtz, C. *J. Proteomics* **2017**, doi: 10.1016/j.jprot.2017.1008.1003.
(179) Wang, E. H.; Combe, P. C.; Schug, K. A. *J. Am. Soc. Mass. Spectrom.* **2016**, *27*, 886-896.
(180) Wang, E. H.; Appulage, D. K.; McAllister, E. A.; Schug, K. A. *J. Am. Soc. Mass. Spectrom.* **2017**.

(181) Kilpatrick, L. E.; Kilpatrick, E. L. *J. Proteome Res.* **2017**, *16*, 3255-3265.
(182) Petras, D.; Heiss, P.; Sussmuth, R. D.; Calvete, J. J. *J. Proteome Res.* **2015**, *14*, 2539-2556.
(183) Petras, D.; Heiss, P.; Harrison, R. A.; Sussmuth, R. D.; Calvete, J. J. *J. Proteomics* **2016**, *146*, 148-164.
(184) Mao, P.; Wang, D. J. *J. Proteome Res.* **2014**, *13*, 1560-1569.
(185) Cheon, D. H.; Nam, E. J.; Park, K. H.; Woo, S. J.; Lee, H. J.; Kim, H. C.; Yang, E. G.; Lee, C.; Lee, J. E. *J. Proteome Res.* **2016**, *15*, 229-244.
(186) Savaryn, J. P.; Toby, T. K.; Catherman, A. D.; Fellers, R. T.; LeDuc, R. D.; Thomas, P. M.; Friedewald, J. J.; Salomon, D. R.; Abecassis, M. M.; Kelleher, N. L. *Proteomics* **2016**, *16*, 2048-2058.
(187) Toby, T. K.; Abecassis, M.; Kim, K.; Thomas, P. M.; Fellers, R. T.; LeDuc, R. D.; Kelleher, N. L.; Demetris, J.; Levitsky, J. *Am. J. Transplantation* **2017**, *17*, 2458-2467.
(188) Desiderio, C.; D'Angelo, L.; Rossetti, D. V.; Iavarone, F.; Giardina, B.; Castagnola, M.; Massimi, L.; Tamburrini, G.; Di Rocco, C. *Proteomics* **2012**, *12*, 2158-2166.
(189) Fania, C.; Arosio, B.; Capitanio, D.; Torretta, E.; Gussago, C.; Ferri, E.; Mari, D.; Gelfi, C. *PLoS ONE* **2017**, *12*, e0179280.
(190) Cabras, T.; Pisano, E.; Montaldo, C.; Giuca, M. R.; Iavarone, F.; Zampino, G.; Castagnola, M.; Messana, I. *Molecular & Cellular Proteomics* **2013**, *12*, 1844-1852.
(191) Iauarone, F.; Melis, M.; Platania, G.; Cabras, T.; Manconi, B.; Petruzzelli, R.; Cordaro, M.; Siracusano, A.; Faa, G.; Messana, I.; Zanasi, M.; Castagnola, M. *J. Proteomics* **2014**, *103*, 15-22.
(192) Labas, V.; Spina, L.; Belleannee, C.; Teixeira-Gomes, A. P.; Gargaros, A.; Dacheux, F.; Dacheux, J. L. *J. Proteomics* **2015**, *113*, 226-243.
(193) Bystrom, C.; Sheng, S. J.; Zhang, K.; Caulfield, M.; Clarke, N. J.; Reitz, R. *PLoS ONE* **2012**, *7*, e43457.
(194) Kellie, J. F.; Higgs, R. E.; Ryder, J. W.; Major, A.; Beach, T. G.; Adler, C. H.; Merchant, K.; Knierman, M. D. *Sci. Rep.* **2014**, *4*.
(195) Steffen, P.; Kwiatkowski, M.; Robertson, W. D.; Zarrine-Afsar, M.; Deterra, D.; Richter, V.; Schluter, H. *J. Proteomics* **2016**, *134*, 5-18.
(196) Gafvels, M.; Bengtson, P. *Clin. Chim. Acta* **2015**, *442*, 87-95.
(197) Carel, C.; Marcoux, J.; Reat, V.; Parra, J.; Latge, G.; Laval, F.; Demange, P.; Burlet-Schiltz, O.; Milon, A.; Daffe, M.; Tropis, M. G.; Renault, M. A. M. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4231-4236.
(198) Gregorich, Z. R.; Peng, Y.; Cai, W. X.; Jin, Y. T.; Wei, L. M.; Chen, A. J.; McKiernan, S. H.; Aiken, J. M.; Moss, R. L.; Diffee, G. M.; Ge, Y. *J. Proteome Res.* **2016**, *15*, 2706-2716.
(199) Wei, L.; Gregorich, Z. R.; Lin, Z.; Cai, W.; Jin, Y.; McKiernan, S. H.; McIlwain, S.; Aiken, J. M.; Moss, R. L.; Diffee, G. M.; Ge, Y. *Mol. Cell. Proteomics* **2017**, doi: 10.1074/mcp.RA1117.000124.
(200) Hunter, T. *Cell* **2000**, *100*, 113-127.
(201) Humphrey, S. J.; Azimifar, S. B.; Mann, M. *Nat. Biotechnol.* **2015**, *33*, 990-995.
(202) Julien, S. G.; Dubé, N.; Hardy, S.; Tremblay, M. L. *Nat. Rev. Cancer* **2011**, *11*, 35-49.
(203) Banks, R. E.; Dunn, M. J.; Hochstrasser, D. F.; Sanchez, J. C.; Blackstock, W.; Pappin, D. J.; Selby, P. J. *Lancet* **2000**, *356*, 1749-1756.
(204) Gregorich, Z. R.; Ge, Y. *Proteomics* **2014**, *14*, 1195-1210.
(205) Zhang, J.; Guy, M. J.; Norman, H. S.; Chen, Y. C.; Xu, Q.; Dong, X.; Guner, H.; Wang, S.; Kohmoto, T.; Young, K. H.; Moss, R. L.; Ge, Y. *J. Proteome Res.* **2011**, *10*, 4054-4065.

(206) Peng, Y.; Gregorich, Z. R.; Valeja, S. G.; Zhang, H.; Cai, W.; Chen, Y. C.; Guner, H.; Chen, A. J.; Schwahn, D. J.; Hacker, T. A.; Liu, X.; Ge, Y. *Mol. Cell. Proteomics* **2014**, *13*, 2752-2764.
(207) Smith, L. M.; Kelleher, N. L.; Consortium Top Down, P. *Nat. Methods* **2013**, *10*, 186-187.
(208) Sharma, K.; D'Souza, R. C.; Tyanova, S.; Schaab, C.; Wiśniewski, J. R.; Cox, J.; Mann, M. *Cell Rep.* **2014**, *8*, 1583-1594.
(209) Huang, P. H.; White, F. M. *Mol. Cell* **2008**, *31*, 777-781.
(210) Chait, B. T. *Science* **2006**, *314*, 65-66.
(211) 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.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. *Nature* **2011**, *480*, 254-258.
(212) Zabrouskov, V.; Ge, Y.; Schwartz, J.; Walker, J. W. *Mol. Cell. Proteomics* **2008**, *7*, 1838-1849.
(213) Cai, W.; Tucholski, T. M.; Gregorich, Z. R.; Ge, Y. *Expert Rev. Proteomics* **2016**, *13*, 717-730.
(214) Riley, N. M.; Coon, J. J. *Anal. Chem.* **2016**, *88*, 74-94.
(215) Xue, L.; Wang, W. H.; Iliuk, A.; Hu, L.; Galan, J. A.; Yu, S.; Hans, M.; Geahlen, R. L.; Tao, W. A. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 5615-5620.
(216) Bodenmiller, B.; Mueller, L. N.; Mueller, M.; Domon, B.; Aebersold, R. *Nat. Methods* **2007**, *4*, 231-237.
(217) Oda, Y.; Nagasu, T.; Chait, B. T. *Nat. Biotechnol.* **2001**, *19*, 379-382.
(218) Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. *Nature* **1975**, *258*, 598-599.
(219) Schmidt, S. R.; Schweikart, F.; Andersson, M. E. *J. Chromatogr. B* **2007**, *849*, 154-162.
(220) Kaur-Atwal, G.; Weston, D. J.; Bonner, P. L.; Crosland, S.; Green, P. S.; Creaser, C. S. *Curr. Anal. Chem.* **2008**, *4*, 127-135.
(221) Wu, S.; Yang, F.; Zhao, R.; Tolic, N.; Robinson, E. W.; Camp, D. G.; Smith, R. D.; Pasa-Tolic, L. *Anal. Chem.* **2009**, *81*, 4210-4219.
(222) Sytnyk, M.; Kirchschlager, R.; Bodnarchuk, M. I.; Primetzhofer, D.; Kriegner, D.; Enser, H.; Stangl, J.; Bauer, P.; Voith, M.; Hassel, A. W. *Nano Lett.* **2013**, *13*, 586-593.
(223) Gu, Z.; Xiang, X.; Fan, G.; Li, F. *J. Phys. Chem. C* **2008**, *112*, 18459-18466.
(224) Sun, S.; Zeng, H.; Robinson, D. B.; Raoux, S.; Rice, P. M.; Wang, S. X.; Li, G. *J. Am. Chem. Soc.* **2003**, *126*, 273-279.
(225) Bateer, B.; Tian, C.; Qu, Y.; Du, S.; Yang, Y.; Ren, Z.; Pan, K.; Fu, H. *Dalton Trans.* **2014**, *43*, 9885-9891.
(226) Kinoshita, E.; Kinoshita-Kikuta, E.; Takiyama, K.; Koike, T. *Mol. Cell. Proteomics* **2006**, *5*, 749-757.
(227) De Palma, R.; Peeters, S.; Van Bael, M. J.; Van den Rul, H.; Bonroy, K.; Laureyn, W.; Mullens, J.; Borghs, G.; Maes, G. *Chem. Mater.* **2007**, *19*, 1821-1831.
(228) Aubin-Tam, M.-E.; Hamad-Schifferli, K. *Biomed. Mater.* **2008**, *3*, 034001.
(229) Pan, Y.; Du, X.; Zhao, F.; Xu, B. *Chem. Soc. Rev.* **2012**, *41*, 2912-2942.
(230) Pan, Y.; Long, M. J.; Lin, H.-C.; Hedstrom, L.; Xu, B. *Chem. Sci.* **2012**, *3*, 3495-3499.
(231) Zhang, J. A.; Dong, X. T.; Hacker, T. A.; Ge, Y. *J. Am. Soc. Mass. Spectrom.* **2010**, *21*, 940-948.
(232) Liu, X. W.; Sirotkin, Y.; Shen, Y. F.; Anderson, G.; Tsai, Y. S.; Ting, Y. S.; Goodlett, D. R.; Smith, R. D.; Bafna, V.; Pevzner, P. A. *Mol. Cell. Proteomics* **2012**, *11*, M111.008524.
(233) Valeja, S. G.; Xiu, L.; Gregorich, Z. R.; Guner, H.; Jin, S.; Ge, Y. *Anal. Chem.* **2015**, *87*, 5363-5371.

(234) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. *Anal. Chem.* **2016**, *88*, 2303-2310.
(235) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.
(236) Cooper, H. J.; Hakansson, K.; Marshall, A. G. *Mass Spectrom. Rev.* **2005**, *24*, 201-222.
(237) Zhang, H.; Ge, Y. *Circulation-Cardiovascular Genetics* **2011**, *4*, 711-+.
(238) Schennach, M.; Breuker, K. *J. Am. Soc. Mass. Spectrom.* **2015**, *26*, 1059-1067.
(239) Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 2857-2862.
(240) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, U. A.; Tsybin, Y. O. *Chem. Soc. Rev.* **2013**, *42*, 5014-5030.
(241) Williams, E. R.; Furlong, J. J.; McLafferty, F. W. *J. Am. Soc. Mass. Spectrom.* **1990**, *1*, 288-294.
(242) Senko, M. W.; Speir, J. P.; Mclafferty, F. W. *Anal. Chem.* **1994**, *66*, 2801-2808.
(243) Little, D. P.; Paul Speir, J.; Senko, M. W.; O'Connor, P. B. *Anal. Chem.* **1994**, *66*, 2809-2815.
(244) Breuker, K.; Oh, H. B.; Lin, C.; Carpenter, B. K.; McLafferty, F. W. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14011-14016.
(245) McLafferty, F. W.; Horn, D. M.; Breuker, K.; Ge, Y.; Lewis, M. A.; Cerda, B.; Zubarev, R. A.; Carpenter, B. K. *J. Am. Soc. Mass. Spectrom.* **2001**, *12*, 245-249.
(246) Zubarev, R. A. *Mass Spectrom. Rev.* **2003**, *22*, 57-77.
(247) Shi, S. D. H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W. *Anal. Chem.* **2001**, *73*, 19-22.
(248) Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A. *Anal. Chem.* **1999**, *71*, 4431-4436.
(249) Mann, M.; Ong, S. E.; Gronborg, M.; Steen, H.; Jensen, O. N.; Pandey, A. *Trends Biotechnol.* **2002**, *20*, 261-268.
(250) Steen, H.; Jebanathirajah, J. A.; Rush, J.; Morrice, N.; Kirschner, M. W. *Mol. Cell. Proteomics* **2006**, *5*, 172-181.
(251) Sweet, S. M. M.; Bailey, C. M.; Cunningham, D. L.; Heath, J. K.; Cooper, H. J. *Mol. Cell. Proteomics* **2009**, *8*, 904-912.
(252) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. *J. Mass Spectrom.* **2009**, *44*, 861-878.
(253) Ayaz-Guner, S.; Zhang, J.; Li, L.; Walker, J. W.; Ge, Y. *Biochemistry* **2009**, *48*, 8161-8170.
(254) Peng, Y.; Chen, X.; Zhang, H.; Xu, Q. G.; Hacker, T. A.; Ge, Y. *J. Proteome Res.* **2013**, *12*, 187-198.
(255) Zabrouskov, V.; Ge, Y.; Schwartz, J.; Walker, J. W. *Molecular & Cellular Proteomics* **2008**, *7*, 1838-1849.
(256) Creese, A. J.; Cooper, H. J. *J. Am. Soc. Mass. Spectrom.* **2008**, *19*, 1263-1274.
(257) Lopez-Clavijo, A. F.; Duque-Daza, C. A.; Creese, A. J.; Cooper, H. J. *Int. J. Mass spectrom.* **2015**, *390*, 63-70.
(258) Zhang, J.; Dong, X.; Hacker, T. A.; Ge, Y. *J Am Soc Mass Spectrom* **2010**, *21*, 940-948.
(259) Iavarone, A. T.; Paech, K.; Williams, E. R. *Anal. Chem.* **2004**, *76*, 2231-2238.
(260) Krusemark, C. J.; Frey, B. L.; Belshaw, P. J.; Smith, L. M. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 1617-1625.
(261) Liu, J.; Konermann, L. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 819-828.
(262) Tsybin, Y. O.; Witt, M.; Baykut, G.; Hakansson, P. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 1607-1613.

(263) Vorobyev, A.; Hamidane, H. B.; Tsybin, Y. O. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 2273-2283.
(264) Krueger, K. E.; Srivastava, S. *Mol. Cell. Proteomics* **2006**, *5*, 1799-1810.
(265) Karve, T. M.; Cheema, A. K. *J. Amino Acids* **2011**, *2011*, 207691.
(266) Ge, Y.; Rybakova, I. N.; Xu, Q.; Moss, R. L. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12658-12663.
(267) Garcia, B. A. *J. Am. Soc. Mass. Spectrom.* **2010**, *21*, 193-202.
(268) Armirotti, A.; Damonte, G. *Proteomics* **2010**, *10*, 3566-3576.
(269) Cui, W.; Rohrs, H. W.; Gross, M. L. *Analyst* **2011**, *136*, 3854-3864.
(270) Zhang, Z. R.; Wu, S.; Stenoien, D. L.; Pasa-Tolic, L. *Annu. Rev. Anal. Chem.* **2014**, *7*, 427-454.
(271) Roth, M. J.; Plymire, D. A.; Chang, A. N.; Kim, J.; Maresh, E. M.; Larson, S. E.; Patrie, S. M. *Anal. Chem.* **2011**, *83*, 9586-9592.
(272) Stobaugh, J. T.; Fague, K. M.; Jorgenson, J. W. *J. Proteome Res.* **2013**, *12*, 626-636.
(273) Vellaichamy, A.; Tran, J. C.; Catherman, A. D.; Lee, J. E.; Kellie, J. F.; Sweet, S. M. M.; Zamdborg, L.; Thomas, P. M.; Ahlf, D. R.; Durbin, K. R.; Valaskovic, G. A.; Kelleher, N. L. *Anal. Chem.* **2010**, *82*, 1234-1244.
(274) Strader, M. B.; VerBerkmoes, N. C.; Tabb, D. L.; Connelly, H. M.; Barton, J. W.; Bruce, B. D.; Pelletier, D. A.; Davison, B. H.; Hettich, R. L.; Larimer, F. W.; Hurst, G. B. *J. Proteome Res.* **2004**, *3*, 965-978.
(275) Eschelbach, J. W.; Jorgenson, J. W. *Anal. Chem.* **2006**, *78*, 1697-1706.
(276) Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A. *Mol. Cell. Proteomics* **2009**, *8*, 2266-2284.
(277) Tetaz, T.; Detzner, S.; Friedlein, A.; Molitor, B.; Mary, J. L. *J. Chromatogr. A* **2011**, *1218*, 5892-5896.
(278) Chen, X.; Ge, Y. *Proteomics* **2013**, *13*, 2563-2566.
(279) Queiroz, J. A.; Tomaz, C. T.; Cabral, J. M. S. *J. Biotechnol.* **2001**, *87*, 143-159.
(280) Alpert, A. J. *J. Chromatogr.* **1986**, *359*, 85-97.
(281) Xiu, L.; Valeja, S. G.; Alpert, A. J.; Song, J.; Ying, G. *Anal. Chem.* **2014**, *86*, 8p.
(282) Flick, T. G.; Cassou, C. A.; Chang, T. M.; Williams, E. R. *Anal. Chem.* **2012**, *84*, 7511-7517.
(283) Gooding, D. L.; Schmuck, M. N.; Nowlan, M. P.; Gooding, K. M. *J. Chromatogr.* **1986**, *359*, 331-337.
(284) McNay, J. L.; O'Connell, J. P.; Fernandez, E. J. *Biotechnol. Bioeng.* **2001**, *76*, 233-240.
(285) Jungbauer, A.; Machold, C.; Hahn, R. *J. Chromatogr. A* **2005**, *1079*, 221-228.
(286) A Loo, J.; A Benchaar, S.; Zhang, J. *Mass spectrometry (Tokyo, Japan)* **2013**, *2*, S0013.
(287) Haverick, M.; Mengisen, S.; Shameem, M.; Ambrogelly, A. *Mabs* **2014**, *6*, 852-858.
(288) Di Donato, A.; Ciardiello, M. A.; de Nigris, M.; Piccoli, R.; Mazzarella, L.; D'Alessio, G. *J. Biol. Chem.* **1993**, *268*, 4745-4751.
(289) Haimer, E.; Tscheliessnig, A.; Hahn, R.; Jungbauer, A. *J. Chromatogr. A* **2007**, *1139*, 84-94.
(290) Lin, F. Y.; Chen, W. Y.; Ruaan, R. C.; Huang, H. M. *J. Chromatogr. A* **2000**, *872*, 37-47.
(291) Lin, F. Y.; Chen, W. Y.; Hearn, M. T. W. *Anal. Chem.* **2001**, *73*, 3875-3883.
(292) Leavy, O. *Nat. Rev. Immunol.* **2010**, *10*, 297-297.
(293) Buss, N. A. P. S.; Henderson, S. J.; McFarlane, M.; Shenton, J. M.; de Haan, L. *Curr. Opin. Pharmacol.* **2012**, *12*, 615-622.

(294) Weiner, G. J. *Nat. Rev. Cancer* **2015**, *15*, 361-370.
(295) Ecker, D. M.; Jones, S. D.; Levine, H. L. *Mabs* **2015**, *7*, 9-14.
(296) Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N. *Nat. Rev. Immunol.* **2010**, *10*, 345-352.
(297) Beck, A.; Wagner-Rousset, E.; Ayoub, D.; Van Dorsselaer, A.; Sanglier-Cianferani, S. *Anal. Chem.* **2013**, *85*, 715-736.
(298) Fekete, S.; Guillarme, D.; Sandra, P.; Sandra, K. *Anal. Chem.* **2016**, *88*, 480-507.
(299) Zhang, H.; Cui, W. D.; Gross, M. L. *FEBS Lett.* **2014**, *588*, 308-317.
(300) Cusumano, A.; Guillarme, D.; Beck, A.; Fekete, S. *J. Pharm. Biomed. Anal.* **2016**, *121*, 161-173.
(301) Chen, B.; Brown, K. A.; Lin, Z.; Ge, Y. *Anal. Chem.* **2018**, *90*, 110-127.
(302) Rosati, S.; Yang, Y.; Barendregt, A.; Heck, A. J. *Nat. Protoc.* **2014**, *9*, 967-976.
(303) Campuzano, I. D. G.; Netirojjanakul, C.; Nshanian, M.; Lippens, J. L.; Kilgour, D. P. A.; Van Orden, S.; Loo, J. A. *Anal. Chem.* **2018**, *90*, 745-751.
(304) Formolo, T.; Ly, M.; Levy, M.; Kilpatrick, L.; Lute, S.; Phinney, K.; Marzilli, L.; Brorson, K.; Boyne, M.; Davis, D.; Schiel, J. *State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization, Vol 2: Biopharmaceutical Characterization: The Nistmab Case Study* **2015**, *1201*, 1-62.
(305) Hilliard, M.; Alley, W. R.; McManus, C. A.; Yu, Y. Q.; Hallinan, S.; Gebler, J.; Rudd, P. M. *mAbs* **2017**, *9*, 1349-1359.
(306) Alain, B.; François, D.; Hélène, D.; Elsa, W. R.; Olivier, C.; Van, D. A.; Sarah, C. *J. Mass Spectrom.* **2015**, *50*, 285-297.
(307) Ganisl, B.; Breuker, K. *ChemistryOpen* **2012**, *1*, 260-268.
(308) Zhang, J.; Ogorzalek Loo, R. R.; Loo, J. A. *Int. J. Mass spectrom.* **2015**, *377*, 546-556.
(309) Zhang, H.; Cui, W. D.; Wen, J. Z.; Blankenship, R. E.; Gross, M. L. *Anal. Chem.* **2011**, *83*, 5598-5606.
(310) Li, H. L.; Nguyen, H. H.; Loo, R. R. O.; Campuzano, I. D. G.; Loo, J. A. *Nat. Chem.* **2018**, *10*, 139-148.
(311) Karageorgos, I.; Gallagher, E. S.; Galvin, C.; Gallagher, D. T.; Hudgens, J. W. *Biologicals* **2017**, *50*, 27-34.
(312) Beck, A.; Goetsch, L.; Dumontet, C.; Corvaia, N. *Nat. Rev. Drug Discov.* **2017**, *16*, 315-337.
(313) Ducry, L.; Stump, B. *Bioconjugate Chem.* **2010**, *21*, 5-13.
(314) Chari, R. V. J.; Miller, M. L.; Widdison, W. C. *Angewandte Chemie-International Edition* **2014**, *53*, 3796-3827.
(315) Lambert, J. M.; Berkenblit, A. In *Annual Review of Medicine, Vol 69*, Caskey, C. T., Ed., 2018, pp 191-207.
(316) Sievers, E. L.; Senter, P. D. In *Annual Review of Medicine, Vol 64*, Caskey, C. T., Ed., 2013, pp 15-29.
(317) Panowski, S.; Bhakta, S.; Raab, H.; Polakis, P.; Junutula, J. R. *Mabs* **2014**, *6*, 34-45.
(318) Strop, P.; Liu, S. H.; Dorywalska, M.; Delaria, K.; Dushin, R. G.; Tran, T. T.; Ho, W. H.; Farias, S.; Casas, M. G.; Abdiche, Y.; Zhou, D. H.; Chandrasekaran, R.; Samain, C.; Loo, C.; Rossi, A.; Rickert, M.; Krimm, S.; Wong, T.; Chin, S. M.; Yu, J., et al. *Chem. Biol.* **2013**, *20*, 161-167.
(319) Wagh, A.; Song, H. T.; Zeng, M.; Tao, L.; Das, T. K. *Mabs* **2018**, *10*, 222-243.
(320) Mouchahoir, T.; Schiel, J. E. *Anal. Bioanal. Chem.* **2018**, *410*, 2111-2126.

(321) Ren, D.; Pipes, G. D.; Liu, D. J.; Shih, L. Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V. *Anal. Biochem.* **2009**, *392*, 12-21.
(322) Krokhin, O. V.; Antonovici, M.; Ens, W.; Wilkins, J. A.; Standing, K. G. *Anal. Chem.* **2006**, *78*, 6645-6650.
(323) Sang, H.; Lu, G. Y.; Liu, Y. Z.; Hu, Q. Y.; Xing, W. F.; Cui, D. B.; Zhou, F.; Zhang, J. W.; Hao, H. P.; Wang, G. J.; Ye, H. *Anal. Chim. Acta* **2017**, *955*, 67-78.
(324) Chen, B.; Lin, Z.; Alpert, A. J.; Fu, C.; Zhang, Q.; Pritts, W. A.; Ge, Y. *Anal. Chem.* **2018**, *90*, 7135-7138.
(325) Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W. P.; Mohammed, S.; Heck, A. J. R. *Anal. Chem.* **2017**, *89*, 3318-3325.
(326) Jin, Y. T.; Wei, L. M.; Cai, W. X.; Lin, Z. Q.; Wu, Z. J.; Peng, Y.; Kohmoto, T.; Moss, R. L.; Ge, Y. *Anal. Chem.* **2017**, *89*, 4922-4930.
(327) Dyachenko, A.; Wang, G. B.; Belov, M.; Makarov, A.; de Jong, R. N.; van den Bremer, E. T. J.; Parren, P. W. H. I.; Heck, A. J. R. *Anal. Chem.* **2015**, *87*, 6095-6102.
(328) Janin-Bussat, M. C.; Dillenbourg, M.; Corvaia, N.; Beck, A.; Klinguer-Hamour, C. *J. Chromatogr. B* **2015**, *981*, 9-13.
(329) Tsybin, Y. O.; Fornelli, L.; Stoermer, C.; Luebeck, M.; Parra, J.; Nallet, S.; Wurm, F. M.; Hartmer, R. *Anal. Chem.* **2011**, *83*, 8919-8927.
(330) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. *Anal. Chem.* **2014**, *86*, 3005-3012.
(331) Cotham, V. C.; Brodbelt, J. S. *Anal. Chem.* **2016**, *88*, 4004-4013.
(332) Fornelli, L.; Srzentic, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskoy, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. *Anal. Chem.* **2018**, *90*, 8421-8429.
(333) Wagner-Rousset, E.; Janin-Bussat, M. C.; Colas, O.; Excoffier, M.; Ayoub, D.; Haeuw, J. F.; Rilatt, I.; Perez, M.; Corvaia, N.; Beck, A. *Mabs* **2014**, *6*, 173-184.
(334) Chevreux, G.; Tilly, N.; Bihoreau, N. *Anal. Biochem.* **2011**, *415*, 212-214.
(335) Moelleken, J.; Endesfelder, M.; Gassner, C.; Lingke, S.; Tomaschek, S.; Tyshchuk, O.; Lorenz, S.; Reiff, U.; Molhoj, M. *Mabs* **2017**, *9*, 1076-1087.
(336) Jin, Y. T.; Lin, Z. Q.; Xu, Q. G.; Fu, C. X.; Zhang, Z. R.; Zhang, Q. Y.; Pritts, W. A.; Ge, Y. *Mabs* **2019**, *11*, 106-115.
(337) Wang, W. R.; Vlasak, J.; Li, Y. S.; Pristatsky, P.; Fang, Y. L.; Pittman, T.; Roman, J.; Wang, Y.; Prueksaritanont, T.; Ionescu, R. *Mol. Immunol.* **2011**, *48*, 860-866.
(338) Bertolotti-Ciarlet, A.; Wang, W. R.; Lownes, R.; Pristatsky, P.; Fang, Y. L.; McKelvey, T.; Li, Y. Z.; Li, Y. S.; Drummond, J.; Prueksaritanont, T.; Vlasak, J. *Mol. Immunol.* **2009**, *46*, 1878-1882.
(339) Rush, M. J. P.; Riley, N. M.; Westphall, M. S.; Coon, J. J. *Anal. Chem.* **2018**, *90*, 8946-8953.
(340) Kim, M. T.; Chen, Y.; Marhoul, J.; Jacobson, F. *Bioconjugate Chem.* **2014**, *25*, 1223-1232.
(341) Chen, L. X.; Wang, L.; Shion, H.; Yu, C. F.; Yu, Y. Q.; Zhu, L.; Li, M.; Chen, W. B.; Gao, K. *Mabs* **2016**, *8*, 1210-1223.