The Dissertation or Treatise Committee for Eleanor Watts Certifies that this is the approved version of the following Dissertation:

Characterization of Peptides and Proteins by Ultraviolet Photodissociation Mass Spectrometry for Biotechnology Applications

Committee:

Jennifer S. Brodbelt, Supervisor

Andrew D. Ellington

Thomas J. Brenna

George Georgiou

Characterization of Peptides and Proteins by Ultraviolet Photodissociation Mass Spectrometry for Biotechnology Applications

by

Eleanor Watts

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2023

Acknowledgements

The work demonstrated in this dissertation would not have been possible without the contributions of countless people who have helped and rooted for me along the way. My advisor, Jennifer Brodbelt enabled this work through her mentorship and by providing access to state of the art instrumentation. Professor Brodbelt helped and encouraged me in reaching my career goals by connecting me with industry experts allowing for the collaborations that resulted in this work. She also gave me substantial freedom to develop my own projects and propose and establish industry collaborations.

Current and former members of the Brodbelt group paved the way for this research, helped keep the lab and its many instruments running, and provided a sounding board to find problems and solutions. I have been extremely lucky to have worked with too many wonderful piers name.

I am also fortunate to have worked with many amazing collaborators during my time at UT Austin. I am extremely grateful to Melanie Patterson, John Williams, Greg Potts, Damien Ready, Alayna George Thompson, Janice Lee, Ross Thyer, and Christopher Crittenden for providing their time, expertise, and samples for the work presented here. I will be forever indebted to Melanie Patterson for her unwavering support and mentorship during the early days of the COVID-19 pandemic without which I may not have been able to complete this dissertation. I was also extremely fortunate to have the opportunity to explore a new field for me, and my research group, in Alternative Protein. Matt Crowe and Hank Kalinoski helped me to develop an additional collaboration and research goals meaningful to the food industry. I am also grateful to

Yiming Chen for his tireless excitement for science and for facilitating an internship to help me further pursue my careers goals.

Many people paved my way to a graduate education at the UT Austin, from amazing chemistry teaches going back to high school and middle school, to research mentors during my undergraduate education at Northeastern. I am especially grateful to Eric Shi and Steve Skinner who's shared passion for the field of Mass Spectrometry lead me here. I'm also grateful to the friends and family who have offered tireless encouragement over the years and lent me their ears on particularly long days.

Abstract

Characterization of Peptides and Proteins by Ultraviolet Photodissociation Mass Spectrometry for Biotechnology Applications

Eleanor Watts, Ph.D.

The University of Texas at Austin, 2023

Supervisor: Jennifer S. Brodbelt

Given the rapid and ubiquitous growth of biotechnology as a field, expanding the capabilities for the characterization of peptides and proteins remains paramount. Extending the breadth of applications for advanced tandem mass spectrometry methods has been critical to improving the understanding of biomolecules. The research presented in this dissertation focuses on the characterization of peptides and proteins with ultraviolet photodissociation-mass spectrometry (UVPD-MS). In addition to enhancing the body of knowledge on the applications of UVPD, this results in improved tools to characterize biotherapeutics and other biomolecules relevant to the biotechnology industry, including antibody drug conjugates (ADCs), synthetic selenoproteins, and human leukocyte antigen (HLA) immunopeptides. Through this work, improved localization of drug conjugation sites is presented for both cysteine- and lysine-linked ADCs, complete characterization of sequence and diselenide bond locations is reported for synthetic selenoproteins, and comprehensive sequence information, including leucine and isoleucine differentiation, is established for HLA immunopeptides.

Various sample preparation techniques are explored to suit each application, including bottom-up and middle-down proteolysis as well as top-down analysis. High resolution mass spectrometry instrumentation and modern tandem mass spectrometry platforms are integral to the research presented in this dissertation. In addition to UVPD, a variety of tandem mass spectrometry techniques are explored including collisional and electron transfer dissociation. External collaborations have contributed greatly to the impact of this work, ensuring the development of industry-relevant applications. Access to application-specific software suits has also facilitated and enhanced the analysis and presentation of high-quality data. Overall, the work represents a step forward in both the understanding of key biomolecules as well as the applicability of UVPD to key areas in biotechnology research.

Table of Contents

List of Tables

List of Figures

List of Schemes

Chapter 1: Mass Spectrometry Methods for Biotherapeutic Antibodies and Antibody-Drug Conjugates

1.1 Motivation and Scope of Research

Mass spectrometry has emerged as an indispensable tool in the field of biotechnology. From every stage of research and development through quality assurance, mass spectrometry is critical to our understanding of biological systems and the safety of drugs and consumer products.[1] Tandem mass spectrometry techniques have been extensively developed to enhance the characterization of peptides and proteins.[2] Despite the ubiquitous nature of mass spectrometry methods, challenges remain in their extension to complex systems. Antibodies have been and remain an important area of research with a plethora of therapeutic applications, and extensive work has been undertaken to characterize antibodies at an in-depth level.[3–6] As biotechnological fields continue to advance, however, so do the challenges faced by the field of mass spectrometry.[7] As a result, the necessity of expanding applications for increasingly complex systems becomes more pressing.

While it is often assumed that any advancements in the characterization of antibodies with mass spectrometry methods can be directly applied to antibody-drug conjugates (ADCs), little has been done to follow-up these claims. The development of increasingly complex biotherapeutics, including highly heterogeneous lysine-linked ADCs and custom selenoproteins, necessitates the careful validation of a robust toolbox of mass spectrometry methods.[7,8] Additional frontiers in biotechnology, including immunopeptidomics and foodomics, would also benefit from the broadening of the application of advanced mass spectrometry techniques.[9,10] The work described herein seeks to build on mass spectrometry applications to extend the level of characterization possible for biotherapeutic and other biotechnological applications.

1.2 Therapeutic Importance of Antibodies and ADCs

Antibody-based therapeutics are increasingly being developed to address new therapeutic challenges. Given the high specificity of mAbs, they often result in high efficacy and less adverse events when compared to traditional small-molecule therapeutics.[11] While cancer represents one of the most common therapeutic applications of antibodies, they are also used for treatment-resistant chronic conditions such as psoriasis, atopic dermatitis, sickle cell disease, and asthma.[12] Therapeutic antibodies have more recently been developed to treat viral infections including COVID-19.[12] As displayed in **Figure 1.1**, the number of antibody therapeutics approved in the US has increased substantially over the last 10 years. In addition to the 108 therapeutic antibodies currently approved for use in US, 20 additional therapeutic antibodies are currently in review,[12] representing a continued interest in the development of antibody-based therapeutics.

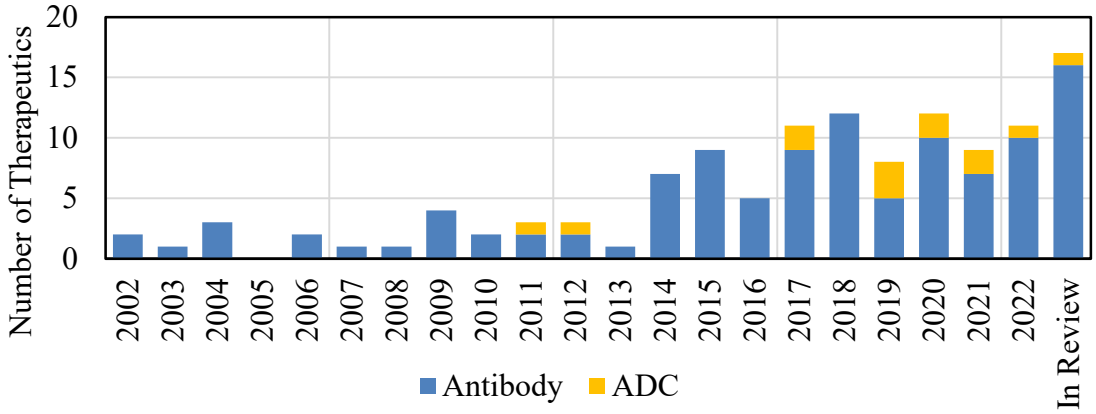


Figure 1.1: Table displaying the number of antibody and ADC therapeutics granted approval in the US each year from 2002-2022, including those under FDA review as of January 2023. Data was adapted from the Antibody Society.[12]

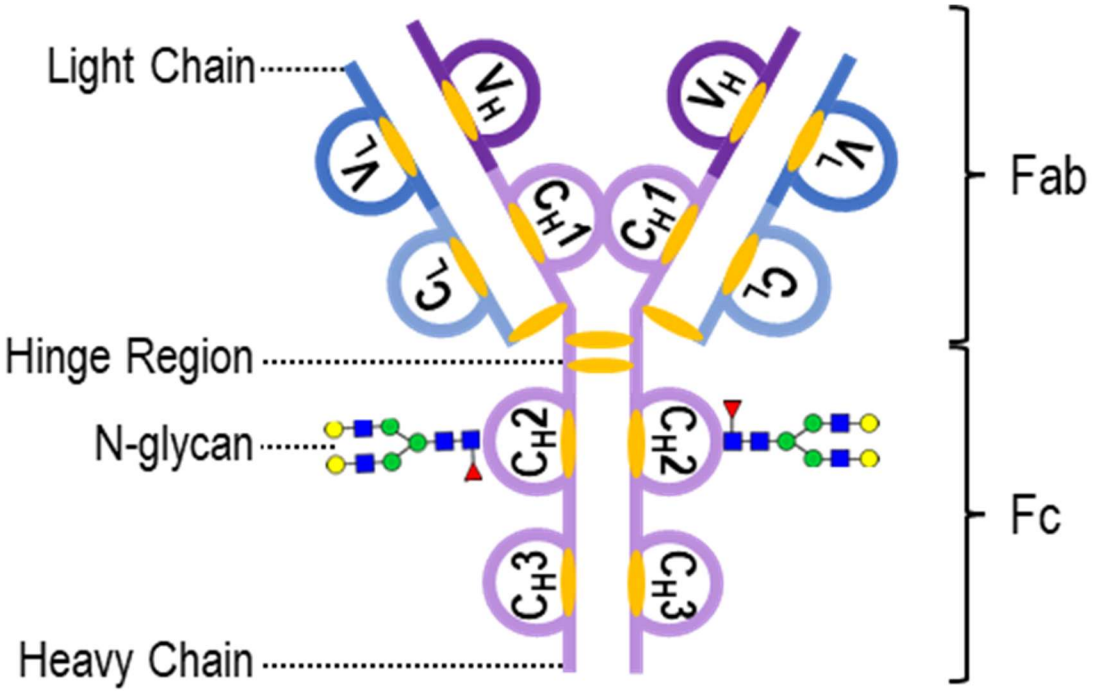
ADCs are a smaller group of therapeutics and a relatively newer frontier. They involve the linkage of cytotoxic drugs to monoclonal antibodies, in the hope of

generating a “magic bullet” that will specifically target diseased cells and unleash the cytotoxic drug.[13] Since the approval of the first ADC, gemtuzumab ozogamicin (Mylotarg), in 2001, a total of 12 ADCs have received FDA approval.[12] While this represents a relatively small subset of the total number of antibody therapeutics available, the interest in ADCs as a therapeutic tool has continued to expand and many more are under development.[13] The increasing prevalence of ADCs, as well as additional subcategories of antibody therapeutics, including bispecific antibodies and nanobodies, highlights the importance of expanding and validating the capabilities of mass spectrometry methodologies for increasingly complex systems beyond traditional monoclonal antibodies.

1.3 Challenges for Characterizing Antibodies and ADCs

1.3.1 Immunoglobulin G1

Most biotherapeutic antibodies and ADCs, as well as the main antibody components of human serum, retain the same basic structure of Immunoglobulin G1 (IgG1), as illustrated in **Scheme 1.1**. IgG1 is made up of two heavy chains, composed of a single variable region (V_H) and three constant regions (C_H1-3), as well as two light chains, which also contain a variable region (V_L) and a constant region (C_L). The variable portions of IgG1 contain the complementarity determining regions (CDRs) which define the epitope an antibody will bind to, while the constant regions are largely responsible for the antibody structure. In addition to the conserved portion of the primary peptide sequence, all IgG1s contain 16 disulfide bonds of which two link the heavy and light chains together and two link the heavy chains to each other. Two N-glycans also play important roles in the structure and stability of the antibody.



Scheme 1.1: Structure of IgG1. The light chains, composed of V_L and C_L, are shown in blue and the heavy chains, composed of V_H, C_H1, C_H2, and C_H3 are shown in purple. G2F glycans are included in the C_H2 region of each heavy chain to demonstrate the location of the N-glycans. Key regions of the antibody include the hinge region, which contains disulfide linkages between the two heavy chain subunits and between the heavy and light chain subunits, the Fab subunit, composed of the light chain and the V_H and C_H1 portions of the light chain, and the Fc subunit, which includes the C_H2 and C_H3 portions of the heavy chains as well as the N-glycans.

The confirmation of disulfide linkages within antibodies has been an important area of exploration. The formation and shuffling of disulfide bonds can cause problems for the stability and safety of therapeutic antibodies. Several methods, including ion mobility and tandem mass spectrometry, have been adopted to probe changes in the disulfide bridging patterns of antibodies.[14–17] The N-glycan also represents an important structural feature and a significant body of research has gone into characterizing the N-glycans of antibodies, typically by enzymatically releasing the N-glycans and

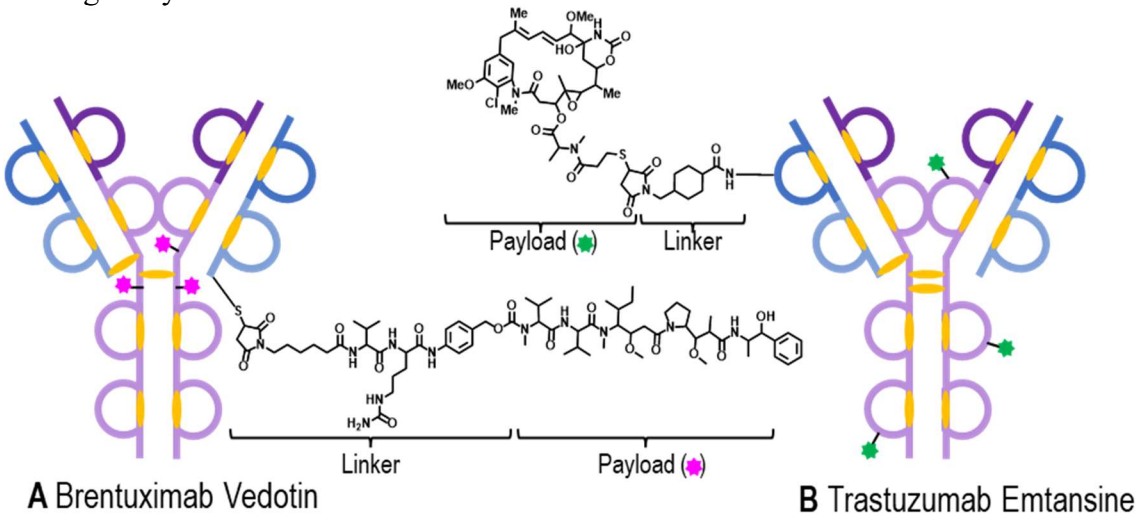
characterizing them separately from the protein.[18,19] Additional post-translational modifications (PTMs) including deamidation and oxidation, can occur during processing or as a result of degradation.[20] These modifications can be important to monitor, as they may impact bioactivity, especially if they occur on the binding portion of the antibody or at the termini.[21] In some cases, a PTM may even trigger an immunogenic response, leading to serious safety concerns.[1] A variety of mass spectrometry, and other analytical methods have been developed with the aim of to monitor antibody PTMs.[20]

1.3.2 Antibody-Drug Conjugates

For ADCs, the addition of payloads further complicates characterization. Common concerns when characterizing ADCs include assessing the average drug to antibody ratio (DAR) and confirming the location of payload linkages.[5] These quality attributes have received the most focus, but even more in-depth understanding is desirable to fully assess any impact the payload may have on the structure and function of the antibody.[4] Additional concerns may include the structure of the antibody after payload conjugation, as well as pharmacokinetic DAR changes in vitro.[3]

Scheme 1.2 includes two examples of the most common types of ADCs. Cysteine-linked ADCs, such as brentuximab vedotin, involve the partial reduction of the antibody disulfide bonds and the conjugation of cytotoxic payloads via maleimide attachment groups and protease-cleavable linkers.[22] Lysine-linked ADCs represent another common type of ADC. An exemplative lysine-linked ADC, trastuzumab emtansine, is illustrated in **Scheme 1.2B**. Trastuzumab emtansine utilizes a non-cleavable thioether linker that binds to random lysine residues via an amine reactive NHS ester group.[23] Given the large number of lysine residues (typically around 75-90 for IgG1), lysine-linked ADCs contain even greater heterogeneity then their cysteine-linked

counterparts. Identifying payload binding sites can become an even more significant concern, as conjugation within the complementarity determining region of the antibody could inhibit epitope binding. The location of the payload conjugation has also been shown to contribute significantly to the linker stability.[24,25] Additional ADCs are under development which utilize conjugation to the N-glycan or modified amino acid chemistries in order to limit and/or eliminate heterogeneity in payload binding sites.[22] While "next-generation" ADCs are a highly attractive alternative, allowing the selection of a conjugation site that will maximize linker stability and limit effect on antibody function, the majority of ADCs to date contain significant conjugation site heterogeneity.[22]



Scheme 1.2: Two exemplative ADC structures displaying common linker chemistries including (A) cysteine-linked ADC, brentuximab vedotin, and (B) lysine-linked ADC, trastuzumab emtansine. N-glycans are not shown.

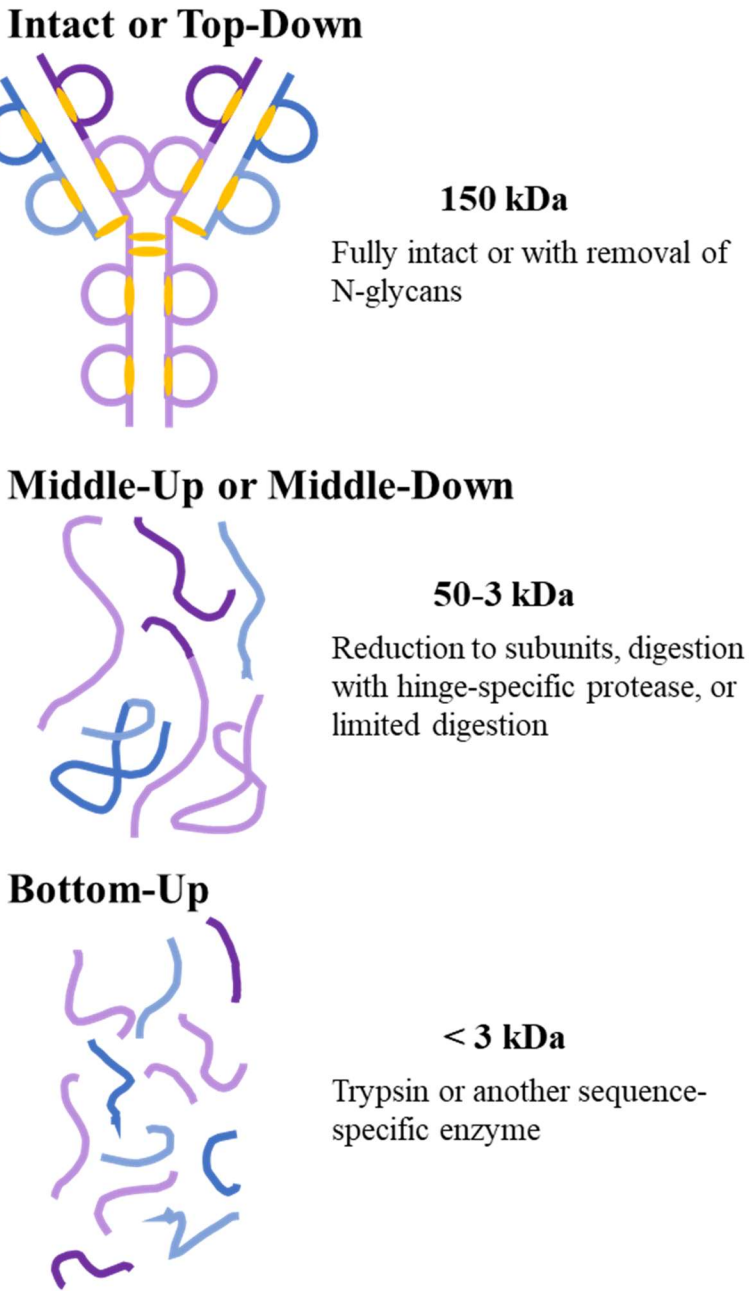
1.3.3 Other Immunoglobulins and Antibody Derivatives

The focus of most method development for the characterization of antibodies has centered around IgG1 (**Scheme 1.1**). While this is logical, considering that it is the most

common template for therapeutic antibodies, the rapid growth in the field has led to the development of a broader variety of therapeutic antibodies. IgG2 and IgG4 are two additional human IgG subclasses of interest for therapeutic antibodies. While IgG1 only has one possible disulfide bridging pattern, IgG2 and IgG4 can both undergo multiple disulfide bond rearrangements.[26] Disulfide rearrangements result in conformational changes at the Fab portion of IgG2.[26] Disulfide rearrangements in IgG4 can allow for half-molecule exchange and the generation of bivalent antibodies.[26] Beyond the 77 IgG1-based therapeutics that have been approved, there are currently 13 IgG2- and 18 IgG4-based therapeutic antibodies on the market.[12] In addition to the increasing variety of IgG types, additional antibody derivatives are gaining popularity. Examples include antibodies combining two epitopes (bivalent antibodies), and therapeutic antibodies retaining only the Fab region or single-chain variable fragment (nanobodies).[27,28] The broadening of the field highlights the importance of continuing to validate and improve upon mass spectrometry methodology. Chapter 5 focuses on the characterization of synthetic selenoproteins, an additional frontier for biotherapeutics.

1.4 Levels of Antibody Characterization

There are several different proteomic approaches developed for the characterization of antibodies. These approaches are selected based on the type of information needed and are often combined to gain a complete understanding of the antibody structure. The main sample processing methods to achieve different levels of characterization are overviewed in **Scheme 1.3** and described in additional detail in the following sections.



Scheme 1.3: MS strategies used of antibody characterization, including common sample preparation techniques for each method. Intact and middle-up techniques involve simple measurement of the intact mass of antibody, or subunits, while top-down, middle-down, and bottom-up techniques utilize tandem mass spectrometry.

1.4.1 Intact and Top-Down Methods

Intact or top-down methods require the least sample preparation and enable the direct analysis of an entire antibody or ADC. Intact methods involve the measurement of the protein molecular weight without any fragmentation. Intact methods are useful to determine highly accurate molecular weights, which can help to confirm the protein sequence, and allow the assessment of heterogeneity imposed by PTMs.[29] For ADCs, intact mass measurements are critical for the assessment of average DAR.[3] Native mass spectrometry strives to retain native (folded, conformationally active) structural features of proteins, enabling assessment of the higher order structures of proteins and other biomolecules.[30] An appealing route for antibodies, native mass spectrometry allows the characterization of antibody stoichiometry, topology, stability, binding properties, and structure.[29]

Mass spectrometry methods that analyze intact proteins are largely used in the characterization of DARs for ADCs. The development of chromatography methods coupled with mass spectrometry have proven integral to the characterization of DARs. This is particularly true for cysteine-linked ADCs, where the partial reduction of disulfide bonds necessitates native chromatography conditions.[31–34] Hydrophobic interaction chromatography (HIC) and size exclusion chromatography have both been implemented to that end.[31–34] HIC allows the separation of ADCs based on DAR, making it possible to perform an average DAR calculation with either mass spectrometry or UV-VIS results.[31–33] Size exclusion chromatography (SEC) does not separate species based DAR, but has been shown to be equally valid for assessing average DAR at a higher throughput.[34] The integration of ion mobility and high resolution mass spectrometry has also proven to aid in the characterization of DAR and PTM measurements without using chromatography.[35–37]

Top-down analysis also entails the ionization and analysis of intact proteins but is coupled with tandem mass spectrometry to achieve further characterization. Several recent strides have been made to generate informative sequence information through top-down characterization of antibodies. For a long time the goal was to develop advanced methods to increase sequence coverage, but results were limited to 30-60% sequence coverage.[38–42] More recently, the focus has shifted from total sequence coverage to how informative the fragment ions generated are. For example, it has been demonstrated that the primary fragment ions generated with electron capture dissociation of IgG complexes are N-terminal fragments from the variable region. Therefore, informative sequence information could be gained at a the top-down level without cleavage of disulfide bonds.[43,44] While top-down mass spectrometry of ADCs has been largely limited, a few studies have demonstrated the release of the light chain through tandem mass spectrometry, enabling assessment of subunit-level payload distribution without digestion or reduction, as required for middle-up analysis.[45,46]

1.4.2 Middle-Up and Middle-Down Methods

Middle-down characterization of proteins broadly involves the digestion into large 3-20 kDa or larger peptides. In the case of antibodies, middle-down analysis usually refers to specific digestion at the hinge region of the antibody with enzymes such as IdeS, IdeZ, and Kgp, and/or reduction of disulfide bonds, resulting in 25-50 kDa subunits. The term “middle-up analysis” refers to measurement of the intact mass of such subunits without tandem mass spectrometry. Such methodology can be helpful for the same reasons described for characterization of intact proteins, while the simplification into subunits facilitates measurements of proteins without the use of ultra high-performance mass spectrometers (ones with high resolution, high mass accuracy, and high mass

range). Middle-up analysis is often used to assess the drug-load on different subunits for ADCs.[47–52] Middle-down characterization represents a similar analog to the top-down method, meaning the emphasis is on using MS/MS techniques to obtain more detailed information rather than just measuring masses of the subunits. Middle-down analysis aims to gain more information than what can be achieved with bottom-up techniques, while generating peptides of more manageable sizes than an intact antibody. Middle-down characterization of antibody subunits has been of significant interest with sequence coverages of 50-90% reported.[53–57] Subunit characterization has also been extended to ADCs for which sequence coverages of 20-70% have been reported, along with modest characterization of payload location, depending on the level of heterogeneity.[58–60] In addition to the antibody specific subunit digestion, middle-down analysis can also refer to digestion with a less specific protease, aided by restricted digestion to generate larger peptides (~3-10 kDa) than achieved with traditional bottom-up (1-3 kDa peptides) methods.[61–63] The increased complexity of the subunits and peptides generated with middle-down techniques often results in the necessity for advanced tandem mass-spectrometry methods, which will be described in detail in section 1.5.

1.4.3 Bottom-up Methods

The simplest method in terms of mass spectrometry requirements, bottom-up analysis involves the complete reduction of disulfide bonds and digestion of protein sequences into small (<3 kDa) peptides. While top-down mass spectrometry is often preferred because the integrity of the entire intact mAb or ADC is retained, bottom-up methods offer improved abilities to identify low-level amino acid misincorporations, generate disulfide bridge assignments, and characterize N-glycans.[64] Bottom-up digestion without reduction has been shown to allow the characterization of disulfide bridges

without scrambling by maintaining a low pH.[65] Aspartic acid can isomerize, transferring a methylene group to the peptide backbone through a succinimide intermediate, resulting in iso-aspartic acid. The formation of iso-aspartic acid can result in structural changes and has been shown to lead to inactivity in therapeutic antibodies.[66] Given the difficulty associated with differentiating isomers, iso-aspartic acid is typically identified with bottom-up methods.[66,67] Bottom-up techniques are also those most often applied to determine the payload binding sites for ADCs. Bottom-up mass spectrometry methods have been used to identify 38-82 conjugated sites for lysine-linked ADCs.[68–71] Multi-dimensional mass spectrometry methods, often combining hinge digestion and bottom-up digestion, have also been shown to improve characterization of payload sites and additional PTMs for ADCs.[72–75]

1.5 Tandem Mass Spectrometry Methodology

The development of improved tandem mass spectrometry methods have been critical to broaden the capabilities of tandem mass spectrometry beyond the characterization of small peptides. This has been invaluable for antibody and ADC applications. The main tandem mass spectrometry techniques utilized in antibody characterization are described in detail below.

1.5.1 Collisional Dissociation

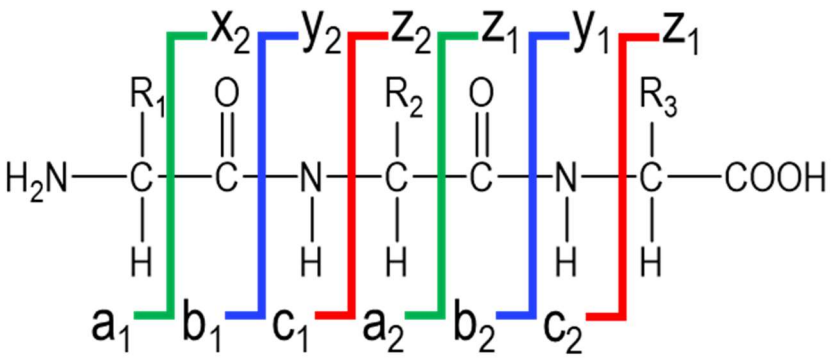
Collision-based dissociation methods represent the most common, and the "gold standard" of tandem mass spectrometry techniques.[2] The fragmentation of peptides and proteins through collisionally activated dissociation (CAD) occurs through a series of collisions, each converting kinetic energy of the ion into internal energy and ultimately allowing biomolecules to overcome the barrier for fragmentation at the peptide

backbone.[76] Given the stepwise nature of this process, CAD leads to the generation of products with a low energy barrier for activation, typically cleavage of the amide backbone bond or cleavage of labile PTMs.[76] The limitations of CAD become more significant as the size and complexity of the peptide or protein increases, leading to insufficient energy deposition.[77] Additionally, CAD has limitations in the characterization of peptides and proteins with highly labile PTMs as the modifications will tend to cleave before the peptide backbone, resulting in loss of PTM site informatoin.[77] This limitation can be problematic for ADCs, where the drug conjugation is a highly labile modification.

Despite these limitations, CAD has successfully been implemented with significant success in the characterization of antibodies and ADCs. Because the limitations in characterization only arise for larger peptides, CAD remains the most popular method for bottom-up characterization of antibodies.[78] CAD methods have also been implemented for the differentiation of aspartic and iso-aspartic acid, as discussed previously.[66,67] For ADCs, CAD has also been successfully employed for the identification of payload conjugation sites through bottom-up workflows.[68–71] Most middle-down studies do not utilize CAD, but a few have included CAD in comparison to or in conjunction with other fragmentation techniques with moderate success in characterizing antibodies and ADCs.[16,56,59] CAD for top-down workflows has been demonstrated to result in the release of the payload-conjugated light chain for cysteine linked ADCs, wherein the payload replaces a disulfide bond.[45,46] Beyond the success described here, CAD is frequently combined with electron transfer dissociation for powerful hybrid techniques, as discussed in detail in section 1.5.3.

1.5.2 Ultraviolet Photodissociation

Ultraviolet photodissociation (UVPD) is a newer tandem mass spectrometry methodology wherein fragmentation is enabled by the coupling of an ultraviolet laser to a mass spectrometer.[79] The absorption of one or more photons leads to excitation of the ion to an excited electronic state, enabling higher energy fragmentation as well as more diverse fragmentation pathways than can be achieved with CAD.[76] In the case of peptides and proteins, 157, 193, and 213 nm photons have all been used successfully to enable UVPD.[80] The fragment ions generated with UVPD include *a, b, c, x, y*, and *z* fragments through cleavage of different backbone positions, as illustrated in **Scheme 1.4**. This contrasts to collisional methods, which typically only yield *a, b*, and *y* fragments. Side-chain loss associated *d* and *w* ions are also generated with UVPD, which can lead to the differentiation of isomeric leucine and isoleucine residues.[81,82] UVPD has also been shown to cleave disulfide bonds.[17,83–86] Additionally, labile PTMs have been shown to be maintained with UVPD, including phosphorylation and glycosylation, thus allowing their localization based on monitoring the mass shifts of fragment ions that correspond to the presence of PTMs.[87–90] The potential to characterize isomeric residues, maintain labile modifications, and cleave disulfide bridges makes UVPD a promising technique for the characterization of antibodies and ADCs.



Scheme 1.4: Polypeptide fragmentation nomenclature, illustrated on a three amino acid long peptide. *a/x, b/y*, and *c/z* labels indicate the cleavage location along the peptide backbone, while the number subscript denotes the number of amino acids retained.

While UVPD is a powerful method in the characterization of peptides and proteins, UVPD spectra also tend to be highly complex due to the large number of possible product ions, which can lead to difficulty in assigning charge states and identifying fragment ions.[80] Several techniques have been developed to increase the signal to noise or reduce the spectral complexity of UVPD fragment ions, namely selective precursor ejection and proton transfer charge reduction (PTCR). Selective precursor ejection aims to increase fragment ion signal by eliminating the precursor and reducing excessive space charge related to the large population of non-dissociated precursor ions.[91] Proton transfer charge reduction reaction utilizes an ion-ion reaction to decrease the charge states of multi-charged ions and shift them to less dense (higher *m/z*) regions of the spectrum, facilitating deconvolution and identification of fragment ions.[92] The combination of PTCR and 193 nm UVPD has been demonstrated to enhance the characterization of ADCs.[92] When specialized techniques to boost the signal to noise of fragment ions are not used, determining accurate filters (based on S/N, mass accuracy or confirming isotopic fits) for fragment ion identification becomes particularly important

for UVPD data. As such, ensuring the quality of UVPD data will be a common theme in the remaining chapters.

UVPD has been implemented in the middle-down and top-down characterization of antibodies, however it is typically combined or compared to electron based dissociation. As such, specific results are detailed in the following section.

1.5.3 Electron Transfer and Electron Capture Dissociation

Additional advanced tandem mass spectrometry techniques include electron capture and electron transfer dissociation (ECD/ETD or ExD).[93] ECD results in a radical-driven fragmentation through the interaction of a positively-charged analyte with low energy electrons.[93] ETD is enabled by an ion-ion reaction in which a singly charged radical anion transfers an electron to the multiply charged cationic analyte.[94] ExD fragmentation occurs as a result of radical-driven cleavage at the amide bond, resulting in *c* and radical *z* ions.[95] While ExD has been responsible for dramatically improved characterization of peptides and proteins, some limitations do exist. For one, ExD is highly charge dependent, which significantly limits the characterization of peptides and proteins in low charge states.[96] Additionally, non-covalent interactions often prevent protein fragment ions from separating from each other and being detected. This phenomenon, nondissociative electron transfer dissociation (ETnoD), results in production of highly abundant and non-informative charge-reduced precursors in ETD spectra.[96] In order to overcome ETnoD and improve the characterization of low charge precursors, ETD is commonly combined in hybrid fragmentation methods by addition of supplemental activation to disrupt the non-covalent interactions and release fragment ions.[96] Common forms of supplemental activation include including CAD, as in electron-transfer/higher-energy collision dissociation (EThcD), and infrared photoactivation, as in

activated ion ETD (AI-ETD).[96] ETD and EThcD have been shown to retain labile PTMs that would be lost with collision based activation.[97] EThcD alone and ETD coupled with HCD event have both also been successfully applied to the differentiation of leucine and isoleucine through the generation of side-chain loss associated *w* ions.[98–100] ETD and EThcD have both been shown to lead to the cleavage of disulfide bonds.[101–103] While CAD is generally sufficient for the bottom-characterization of antibodies, ETD based methods have been among the most popular for middle-down and top-down.[6]

Middle-down characterization of antibodies with ETD alone has been shown to result in approximately 50% sequence coverage alone, and as high as 90% sequence coverage when combined with additional data from EThcD and 213 nm UVPD.[40,53] When applied to a site-specific ADC, middle-down ETD has still been able to achieve ~45% sequence coverage alone, with improved results arising from the addition of 213 nm UVPD and HCD.[59] ETD of heterogeneous cystine- and lysine-linked ADCs has shown further potential to localize the conjugation sites at the middle-down level.[58,104] Combining ETD with EThcD and 193 nm UVPD offers additional characterization of cystine-linked ADCs.[60] Electron-based dissociation methods have also been applied extensively to characterize antibodies at the top-down level. While overall sequence coverages of 30% and 40% have been reported with ETD alone and when combined with 213 nm UVPD, respectively, the highest overall antibody sequence coverage of over 60% was achieved using AI-ETD.[39,40,42] More recently, the focus in developing tandem mass spectrometry methods for top-down antibody characterization has shifted from generating superior sequence coverage to the producing the most informative characterization. High quality assessment of antibody chain pairing has been demonstrated with both of 157 nm UVPD and ECD.[17] Complete characterization of CDR regions from intact antibodies and Fab subunits has been implemented with ECD.[43,44] Overall, the exploration and

development of advanced tandem mass spectrometry techniques, including electron- and photon-based methods, have resulted in an extensive array of techniques to probe antibody drug conjugates.

1.6 Overview of Chapters

Antibody drug conjugates are one area of research that will benefit from the improved adaptation of advanced tandem mass spectrometry methods, including UVPD. As such, the development of UVPD-based methods for ADCs will be a major focus of this work. Several other active research areas and fields that could also benefit from broader applications of UVPD will also be explored.

In Chapter 3 a cystine-linked ADC is characterized with subunit-level middle-down characterization enabled by IdeZ digestion. Tandem mass spectrometry with 193 nm UVPD, ETD, and EThcD were all explored, and proved to provide complementary fragmentation. Because additional information was gained with each method, the combination of all three proved to allow for the most complete characterization of the antibody subunits with 0-2 payloads conjugated, and the differentiation of conjugation locations for the subunits with multiple positional isomers present.

The identification of payload binding sites for a more complex, lysine-linked ADC is examined in Chapter 4. To further simplify the characterization while retaining as much information as possible, a limited proteolysis-based middle-down approach is employed resulting in ~3-13 kDa peptides. This approach allowed for the generation of complex peptides, including those with multiple payloads bound. The characterization of complex peptides was further enabled by HCD-triggered UVPD and EThcD methods, which allowed for high resolution data acquisition without restriction of the duty cycle.

Chapter 5 explores the characterization of synthetic selenoproteins, an emerging technology for biotherapeutics. A combination of top-down, and bottom-up mass spectrometry was examined, and complete characterization was enabled by their combination, as well as complementary results obtained with EThcD and 213 nm UVPD. The method also led to the identification of previously unreported selenoprotein dimers, which shed insight on a potential cause of incorrect diselenide bridge formation.

As examined in Chapter 6, immunopeptides represent an important biological target that have been extensively explored with collisional and electron-based mass spectrometry methods but had yet to be explored with UVPD. The results in Chapter 6 indicated that the integration of 193 nm UVPD into an immunopeptidomic workflow resulted in improved characterization of the peptide sequences, including enhanced differentiation of isomeric leucine and isoleucine residues.

1.7 References

(1) Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R.; Wang, G.; Baykal, B.; Wang, S. Advances and Challenges in Analytical Characterization of Biotechnology Products: Mass Spectrometry-Based Approaches to Study Properties and Behavior of Protein Therapeutics. *Biotechnology Advances* **2012**, *30*, 210–222.

(2) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88*, 30–51.

(3) Huang, R. Y.-C.; Chen, G. Characterization of Antibody–Drug Conjugates by Mass Spectrometry: Advances and Future Trends. *Drug Discov. Today* **2016**, *21*, 850–855.

(4) Beck, A.; Terral, G.; Debaene, F.; Wagner-Rousset, E.; Marcoux, J.; Janin-Bussat, M.-C.; Colas, O.; Dorsselaer, A. V.; Cianférani, S. Cutting-Edge Mass Spectrometry Methods for the Multi-Level Structural Characterization of Antibody-Drug Conjugates. *Expert Rev. Proteomic* **2016**, *13*, 157–183.

(5) Fekete, S.; Guillarme, D.; Sandra, P.; Sandra, K. Chromatographic, Electrophoretic, and Mass Spectrometric Methods for the Analytical Characterization of Protein Biopharmaceuticals. *Anal. Chem.* **2016**, *88*, 480–507.

(6) Fornelli, L.; Ayoub, D.; Srzentic, K.; Nagornov, K.; Kozhinov, A.; Gasilova, N.; Menin, L.; Beck, A.; Tsybin, Y. Structural Analysis of Monoclonal Antibodies with Top-down and Middle-down Electron Transfer Dissociation Mass Spectrometry: The First Decade. *CHIMIA* **2022**, *76*, 114–114.

(7) Wagh, A.; Song, H.; Zeng, M.; Tao, L.; Das, T. K. Challenges and New Frontiers in Analytical Characterization of Antibody-Drug Conjugates. *MAbs* **2018**, *10*, 222–243.

(8) Thyer, R.; Shroff, R.; Klein, D. R.; d'Oelsnitz, S.; Cotham, V. C.; Byrom, M.; Brodbelt, J. S.; Ellington, A. D. Custom Selenoprotein Production Enabled by Laboratory Evolution of Recoded Bacterial Strains. *Nat Biotechnol* **2018**, *36*, 624–631.

(9) Chong, C.; Coukos, G.; Bassani-Sternberg, M. Identification of Tumor Antigens with Immunopeptidomics. *Nat Biotechnol* **2022**, *40*, 175–188.

(10) Picariello, G.; Mamone, G.; Addeo, F.; Ferranti, P. Novel Mass Spectrometry-Based Applications of the "Omic" Sciences in Food Technology and Biotechnology. *Food Technol. Biotechnol.* **2012**, *50*, 286–305.

(11) Lu, R.-M.; Hwang, Y.-C.; Liu, I.-J.; Lee, C.-C.; Tsai, H.-Z.; Li, H.-J.; Wu, H.-C. Development of Therapeutic Antibodies for the Treatment of Diseases. *Journal of Biomedical Science* **2020**, *27*, 1.

(12) Antibody therapeutics approved or in regulatory review in the EU or US https://www.antibodysociety.org/resources/approved-antibodies/ (accessed Jan 20, 2023).

(13) Fu, Z.; Li, S.; Han, S.; Shi, C.; Zhang, Y. Antibody Drug Conjugate: The "Biological Missile" for Targeted Cancer Therapy. *Sig Transduct Target Ther* **2022**, *7*, 1–25.

(14) Bagal, D.; Valliere-Douglass, J. F.; Balland, A.; Schnier, P. D. Resolving Disulfide Structural Isoforms of IgG2 Monoclonal Antibodies by Ion Mobility Mass Spectrometry. *Anal. Chem.* **2010**, *82*, 6751–6755.

(15) Terral, G.; Beck, A.; Cianférani, S. Insights from Native Mass Spectrometry and Ion Mobility-Mass Spectrometry for Antibody and Antibody-Based Product Characterization. *J. Chromatogr. B* **2016**, *1032*, 79–90.

(16) Srzentić, K.; Nagornov, K. O.; Fornelli, L.; Lobas, A. A.; Ayoub, D.; Kozhinov, A. N.; Gasilova, N.; Menin, L.; Beck, A.; Gorshkov, M. V.; Aizikov, K.; Tsybin, Y. O. Multiplexed Middle-Down Mass Spectrometry as a Method for Revealing Light and Heavy Chain Connectivity in a Monoclonal Antibody. *Anal. Chem.* **2018**, *90*, 12527–12535.

(17) Shaw, J. B.; Liu, W.; Vasil′ev, Y. V.; Bracken, C. C.; Malhan, N.; Guthals, A.; Beckman, J. S.; Voinov, V. G. Direct Determination of Antibody Chain Pairing

by Top-down and Middle-down Mass Spectrometry Using Electron Capture Dissociation and Ultraviolet Photodissociation. *Anal. Chem.* **2020**, *92*, 766–773.

(18) Zhang, X. Released N-Glycan Analysis for Biotherapeutic Development Using Liquid Chromatography and Mass Spectrometry. In *Proteomic Profiling: Methods and Protocols*; Posch, A., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2021; pp 35–53.

(19) Campuzano, I. D. G.; Sandoval, W. Denaturing and Native Mass Spectrometric Analytics for Biotherapeutic Drug Discovery Research: Historical, Current, and Future Personal Perspectives. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1861–1885.

(20) Gupta, S.; Jiskoot, W.; Schöneich, C.; Rathore, A. S. Oxidation and Deamidation of Monoclonal Antibody Products: Potential Impact on Stability, Biological Activity, and Efficacy. *Journal of Pharmaceutical Sciences* **2022**, *111*, 903–918.

(21) Alt, N.; Zhang, T. Y.; Motchnik, P.; Taticek, R.; Quarmby, V.; Schlothauer, T.; Beck, H.; Emrich, T.; Harris, R. J. Determination of Critical Quality Attributes for Monoclonal Antibodies Using Quality by Design Principles. *Biologicals* **2016**, *44*, 291–305.

(22) Beck, A.; Goetsch, L.; Dumontet, C.; Corvaïa, N. Strategies and Challenges for the next Generation of Antibody–Drug Conjugates. *Nat. Rev. Drug Discov.* **2017**, *16*, 315–337.

(23) Kim, M. T.; Chen, Y.; Marhoul, J.; Jacobson, F. Statistical Modeling of the Drug Load Distribution on Trastuzumab Emtansine (Kadcyla), a Lysine-Linked Antibody Drug Conjugate. *Bioconjug. Chem.* **2014**, *25*, 1223–1232.

(24) 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.; Chandrasekaran, R.; Samain, C.; Loo, C.; Rossi, A.; Rickert, M.; Krimm, S.; Wong, T.; Chin, S. M.; Yu, J.; Dilley, J.; Chaparro-Riggers, J.; Filzen, G. F.; O'Donnell, C. J.; Wang, F.; Myers, J. S.; Pons, J.; Shelton, D. L.; Rajpal, A. Location Matters: Site of Conjugation Modulates Stability and Pharmacokinetics of Antibody Drug Conjugates. *Chemistry & Biology* **2013**, *20*, 161–167.

(25) Shen, B.-Q.; Xu, K.; Liu, L.; Raab, H.; Bhakta, S.; Kenrick, M.; Parsons-Reponte, K. L.; Tien, J.; Yu, S.-F.; Mai, E.; Li, D.; Tibbitts, J.; Baudys, J.; Saad, O. M.; Scales, S. J.; McDonald, P. J.; Hass, P. E.; Eigenbrot, C.; Nguyen, T.; Solis, W. A.; Fuji, R. N.; Flagella, K. M.; Patel, D.; Spencer, S. D.; Khawli, L. A.; Ebens, A.; Wong, W. L.; Vandlen, R.; Kaur, S.; Sliwkowski, M. X.; Scheller, R. H.; Polakis, P.; Junutula, J. R. Conjugation Site Modulates the in Vivo Stability and Therapeutic Activity of Antibody-Drug Conjugates. *Nat Biotechnol* **2012**, *30*, 184–189.

(26) Thomson, C. A. IgG Structure and Function. In *Encyclopedia of Immunobiology*; Ratcliffe, M. J. H., Ed.; Academic Press: Oxford, 2016; pp 15–22.

(27) Lim, S. M.; Pyo, K.-H.; Soo, R. A.; Cho, B. C. The Promise of Bispecific Antibodies: Clinical Applications and Challenges. *Cancer Treatment Reviews* **2021**, *99*, 102240.

(28) Pirkalkhoran, S.; Grabowska, W. R.; Kashkoli, H. H.; Mirhassani, R.; Guiliano, D.; Dolphin, C.; Khalili, H. Bioengineering of Antibody Fragments: Challenges and Opportunities. *Bioengineering* **2023**, *10*, 122.

(29) J. Thompson, N.; Rosati, S.; J. Rose, R.; R. Heck, A. J. The Impact of Mass Spectrometry on the Study of Intact Antibodies: From Post-Translational Modifications to Structural Analysis. *Chemical Communications* **2013**, *49*, 538–548.

(30) Tamara, S.; den Boer, M. A.; Heck, A. J. R. High-Resolution Native Mass Spectrometry. *Chem. Rev.* **2022**, *122*, 7269–7326.

(31) Birdsall, R. E.; Shion, H.; Kotch, F. W.; Xu, A.; Porter, T. J.; Chen, W. A Rapid On-Line Method for Mass Spectrometric Confirmation of a Cysteine-Conjugated Antibody-Drug-Conjugate Structure Using Multidimensional Chromatography. *mAbs* **2015**, *7*, 1036–1044.

(32) Sarrut, M.; Corgier, A.; Fekete, S.; Guillarme, D.; Lascoux, D.; Janin-Bussat, M.-C.; Beck, A.; Heinisch, S. Analysis of Antibody-Drug Conjugates by Comprehensive on-Line Two-Dimensional Hydrophobic Interaction Chromatography x Reversed Phase Liquid Chromatography Hyphenated to High Resolution Mass Spectrometry. I − Optimization of Separation Conditions. *J. Chromatogr. B* **2016**, *1032*, 103–111.

(33) Jones, J.; Pack, L.; Hunter, J. H.; Valliere-Douglass, J. F. Native Size-Exclusion Chromatography-Mass Spectrometry: Suitability for Antibody–Drug Conjugate Drug-to-Antibody Ratio Quantitation across a Range of Chemotypes and Drug-Loading Levels. *mAbs* **2020**, *12*, 1682895.

(34) Hengel, S. M.; Sanderson, R.; Valliere-Douglass, J.; Nicholas, N.; Leiske, C.; Alley, S. C. Measurement of *in Vivo* Drug Load Distribution of Cysteine-Linked Antibody–Drug Conjugates Using Microscale Liquid Chromatography Mass Spectrometry. *Anal. Chem.* **2014**, *86*, 3420–3425.

(35) Debaene, F.; Bœuf, A.; Wagner-Rousset, E.; Colas, O.; Ayoub, D.; Corvaïa, N.; Van Dorsselaer, A.; Beck, A.; Cianférani, S. Innovative Native MS Methodologies for Antibody Drug Conjugate Characterization: High Resolution Native MS and IM-MS for Average DAR and DAR Distribution Assessment. *Anal. Chem.* **2014**, *86*, 10674–10683.

(36) He, J.; Su, D.; Ng, C.; Liu, L.; Yu, S.-F.; Pillow, T. H.; Del Rosario, G.; Darwish, M.; Lee, B.-C.; Ohri, R.; Zhou, H.; Wang, X.; Lu, J.; Kaur, S.; Xu, K. High-Resolution Accurate-Mass Mass Spectrometry Enabling In-Depth

Characterization of in Vivo Biotransformations for Intact Antibody-Drug Conjugates. *Anal. Chem.* **2017**, *89*, 5476–5483.

(37) Nagornov, K. O.; Gasilova, N.; Kozhinov, A. N.; Virta, P.; Holm, P.; Menin, L.; Nesatyy, V. J.; Tsybin, Y. O. Drug-to-Antibody Ratio Estimation via Proteoform Peak Integration in the Analysis of Antibody–Oligonucleotide Conjugates with Orbitrap Fourier Transform Mass Spectrometry. *Anal. Chem.* **2021**, *93*, 12930–12937.

(38) Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. Top-Down Structural Analysis of an Intact Monoclonal Antibody by Electron Capture Dissociation-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 4239–4246.

(39) Fornelli, L.; Ayoub, D.; Aizikov, K.; Liu, X.; Damoc, E.; Pevzner, P. A.; Makarov, A.; Beck, A.; Tsybin, Y. O. Top-down Analysis of Immunoglobulin G Isotypes 1 and 2 with Electron Transfer Dissociation on a High-Field Orbitrap Mass Spectrometer. *J. Proteom* **2017**, *159*, 67–76.

(40) Fornelli, L.; Srzentić, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **2018**, *90*, 8421–8429.

(41) van der Burgt, Y. E. M.; Kilgour, D. P. A.; Tsybin, Y. O.; Srzentić, K.; Fornelli, L.; Beck, A.; Wuhrer, M.; Nicolardi, S. Structural Analysis of Monoclonal Antibodies by Ultrahigh Resolution MALDI In-Source Decay FT-ICR Mass Spectrometry. *Anal. Chem.* **2019**, *91*, 2079–2085.

(42) Lodge, J. M.; Schauer, K. L.; Brademan, D. R.; Riley, N. M.; Shishkova, E.; Westphall, M. S.; Coon, J. J. Top-Down Characterization of an Intact Monoclonal Antibody Using Activated Ion Electron Transfer Dissociation. *Anal. Chem.* **2020**, *92*, 10246–10251.

(43) Boer, M. A. den; Greisch, J.-F.; Tamara, S.; Bondt, A.; R. Heck, A. J. Selectivity over Coverage in de Novo Sequencing of IgGs. *Chemical Science* **2020**, *11*, 11886–11896.

(44) Greisch, J.-F.; den Boer, M. A.; Beurskens, F.; Schuurman, J.; Tamara, S.; Bondt, A.; Heck, A. J. R. Generating Informative Sequence Tags from Antigen-Binding Regions of Heavily Glycosylated IgA1 Antibodies by Native Top-Down Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2021**.

(45) Dyachenko, A.; Wang, G.; Belov, M.; Makarov, A.; de Jong, R. N.; van den Bremer, E. T. J.; Parren, P. W. H. I.; Heck, A. J. R. Tandem Native Mass-Spectrometry on Antibody–Drug Conjugates and Submillion Da Antibody–

Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. *Anal. Chem.* **2015**, *87*, 6095–6102.

(46) Larson, E. J.; Roberts, D. S.; Melby, J. A.; Buck, K. M.; Zhu, Y.; Zhou, S.; Han, L.; Zhang, Q.; Ge, Y. High-Throughput Multi-Attribute Analysis of Antibody-Drug Conjugates Enabled by Trapped Ion Mobility Spectrometry and Top-Down Mass Spectrometry. *Anal. Chem.* **2021**.

(47) Martelet, A.; Garrigue, V.; Zhang, Z.; Genet, B.; Guttman, A. Multi-Attribute Method Based Characterization of Antibody Drug Conjugates (ADC) at the Intact and Subunit Levels. *J. Pharm. Biomed. Anal.* **2021**, *201*, 114094.

(48) Said, N.; Gahoual, R.; Kuhn, L.; Beck, A.; François, Y.-N.; Leize-Wagner, E. Structural Characterization of Antibody Drug Conjugate by a Combination of Intact, Middle-up and Bottom-up Techniques Using Sheathless Capillary Electrophoresis – Tandem Mass Spectrometry as NanoESI Infusion Platform and Separation Method. *Anal. Chim. Acta* **2016**, *918*, 50–59.

(49) Matsuda, Y.; Kliman, M.; Mendelsohn, B. A. Application of Native Ion Exchange Mass Spectrometry to Intact and Subunit Analysis of Site-Specific Antibody–Drug Conjugates Produced by AJICAP First Generation Technology. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 1706–1712.

(50) Kotapati, S.; Passmore, D.; Yamazoe, S.; Sanku, R. K. K.; Cong, Q.; Poudel, Y. B.; Chowdari, N. S.; Gangwar, S.; Rao, C.; Rangan, V. S.; Cardarelli, P. M.; Deshpande, S.; Strop, P.; Dollinger, G.; Rajpal, A. Universal Affinity Capture Liquid Chromatography-Mass Spectrometry Assay for Evaluation of Biotransformation of Site-Specific Antibody Drug Conjugates in Preclinical Studies. *Anal. Chem.* **2020**, *92*, 2065–2073.

(51) D'Atri, V.; Fekete, S.; Stoll, D.; Lauber, M.; Beck, A.; Guillarme, D. Characterization of an Antibody-Drug Conjugate by Hydrophilic Interaction Chromatography Coupled to Mass Spectrometry. *J. Chromatogr. B* **2018**, *1080*, 37–41.

(52) Firth, D.; Bell, L.; Squires, M.; Estdale, S.; McKee, C. A Rapid Approach for Characterization of Thiol-Conjugated Antibody–Drug Conjugates and Calculation of Drug–Antibody Ratio by Liquid Chromatography Mass Spectrometry. *Anal. Biochem.* **2015**, *485*, 34–42.

(53) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. *Anal. Chem.* **2014**, *86*, 3005–3012.

(54) Resemann, A.; Jabs, W.; Wiechmann, A.; Wagner, E.; Colas, O.; Evers, W.; Belau, E.; Vorwerg, L.; Evans, C.; Beck, A.; Suckau, D. Full Validation of Therapeutic Antibody Sequences by Middle-up Mass Measurements and Middle-down Protein Sequencing. *MAbs* **2016**, *8*, 318–330.

(55) He, L.; Anderson, L. C.; Barnidge, D. R.; Murray, D. L.; Hendrickson, C. L.; Marshall, A. G. Analysis of Monoclonal Antibodies in Human Serum as a Model for Clinical Monoclonal Gammopathy by Use of 21 Tesla FT-ICR Top-Down and Middle-Down MS/MS. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 827–838.

(56) Jin, Y.; Lin, Z.; Xu, Q.; Fu, C.; Zhang, Z.; Zhang, Q.; Pritts, W. A.; Ge, Y. Comprehensive Characterization of Monoclonal Antibody by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *mAbs* **2018**.

(57) Cotham, V. C.; Brodbelt, J. S. Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level Using Middle-Down 193 Nm Ultraviolet Photodissociation. *Anal. Chem.* **2016**, *88*, 4004–4013.

(58) Chen, B.; Lin, Z.; Zhu, Y.; Jin, Y.; Larson, E.; Xu, Q.; Fu, C.; Zhang, Z.; Zhang, Q.; Pritts, W. A.; Ge, Y. Middle-Down Multi-Attribute Analysis of Antibody-Drug Conjugates with Electron Transfer Dissociation. *Anal. Chem.* **2019**, *91*, 11661–11669.

(59) Hernandez-Alba, O.; Houel, S.; Hessmann, S.; Erb, S.; Rabuka, D.; Huguet, R.; Josephs, J.; Beck, A.; Drake, P. M.; Cianférani, S. A Case Study to Identify the Drug Conjugation Site of a Site-Specific Antibody-Drug-Conjugate Using Middle-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2419–2429.

(60) Watts, E.; Williams, J. D.; Miesbauer, L. J.; Bruncko, M.; Brodbelt, J. S. Comprehensive Middle-Down Mass Spectrometry Characterization of an Antibody-Drug Conjugate by Combined Ion Activation Methods. *Anal. Chem.* **2020**.

(61) Cotham, V. C.; Horton, A. P.; Lee, J.; Georgiou, G.; Brodbelt, J. S. Middle-Down 193-Nm Ultraviolet Photodissociation for Unambiguous Antibody Identification and Its Implications for Immunoproteomic Analysis. *Anal. Chem.* **2017**, *89*, 6498–6504.

(62) Pandeswari, P. B.; Sabareesh, V. Middle-down Approach: A Choice to Sequence and Characterize Proteins/Proteomes by Mass Spectrometry. *RSC Adv.* **2018**, *9*, 313–344.

(63) Sanders, J. D.; Greer, S. M.; Brodbelt, J. S. Integrating Carbamylation and Ultraviolet Photodissociation Mass Spectrometry for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89*, 11772–11778.

(64) Beck, A.; Sanglier-Cianférani, S.; Van Dorsselaer, A. Biosimilar, Biobetter, and Next Generation Antibody Characterization by Mass Spectrometry. *Anal. Chem.* **2012**, *84*, 4637–4646.

(65) Sung, W.-C.; Chang, C.-W.; Huang, S.-Y.; Wei, T.-Y.; Huang, Y.-L.; Lin, Y.-H.; Chen, H.-M.; Chen, S.-F. Evaluation of Disulfide Scrambling during the Enzymatic Digestion of Bevacizumab at Various PH Values Using Mass

Spectrometry. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2016**, *1864*, 1188–1194.

(66) Hui, J. O.; Flick, T.; Loo, J. A.; Campuzano, I. D. G. Unequivocal Identification of Aspartic Acid and IsoAspartic Acid by MALDI-TOF/TOF: From Peptide Standards to a Therapeutic Antibody. *J. Am. Soc. Mass Spectrom.* **2021**.

(67) Liu, M.; Cheetham, J.; Cauchon, N.; Ostovic, J.; Ni, W.; Ren, D.; Zhou, Z. S. Protein Isoaspartate Methyltransferase-Mediated 18O-Labeling of Isoaspartic Acid for Mass Spectrometry Analysis. *Anal. Chem.* **2012**, *84*, 1056–1062.

(68) Wu, G.; Gao, Y.; Liu, D.; Tan, X.; Hu, L.; Qiu, Z.; Liu, J.; He, H.; Liu, Y. Study on the Heterogeneity of T-DM1 and the Analysis of the Unconjugated Linker Structure under a Stable Conjugation Process. *ACS Omega* **2019**, *4*, 8834–8845.

(69) Arlotta, K. J.; Gandhi, A. V.; Chen, H.-N.; Nervig, C. S.; Carpenter, J. F.; Owen, S. C. In-Depth Comparison of Lysine-Based Antibody-Drug Conjugates Prepared on Solid Support Versus in Solution. *Antibodies* **2018**, *7*, 6.

(70) Sang, H.; Lu, G.; Liu, Y.; Hu, Q.; Xing, W.; Cui, D.; Zhou, F.; Zhang, J.; Hao, H.; Wang, G.; Ye, H. Conjugation Site Analysis of Antibody-Drug-Conjugates (ADCs) by Signature Ion Fingerprinting and Normalized Area Quantitation Approach Using Nano-Liquid Chromatography Coupled to High Resolution Mass Spectrometry. *Anal. Chim. Acta* **2017**, *955*, 67–78.

(71) Chen, L.; Wang, L.; Shion, H.; Yu, C.; Yu, Y. Q.; Zhu, L.; Li, M.; Chen, W.; Gao, K. In-Depth Structural Characterization of Kadcyla® (Ado-Trastuzumab Emtansine) and Its Biosimilar Candidate. *mAbs* **2016**, *8*, 1210–1223.

(72) Goyon, A.; Kim, M.; Dai, L.; Cornell, C.; Jacobson, F.; Guillarme, D.; Stella, C. Streamlined Characterization of an Antibody–Drug Conjugate by Two-Dimensional and Four-Dimensional Liquid Chromatography/Mass Spectrometry. *Anal. Chem.* **2019**, *91*, 14896–14903.

(73) Janin-Bussat, M.-C.; Dillenbourg, M.; Corvaia, N.; Beck, A.; Klinguer-Hamour, C. Characterization of Antibody Drug Conjugate Positional Isomers at Cysteine Residues by Peptide Mapping LC–MS Analysis. *J. Chromatogr. B* **2015**, *981–982*, 9–13.

(74) Sandra, K.; Vanhoenacker, G.; Vandenheede, I.; Steenbeke, M.; Joseph, M.; Sandra, P. Multiple Heart-Cutting and Comprehensive Two-Dimensional Liquid Chromatography Hyphenated to Mass Spectrometry for the Characterization of the Antibody-Drug Conjugate Ado-Trastuzumab Emtansine. *J. Chromatogr. B* **2016**, *1032*, 119–130.

(75) Li, K.; Lin, Z. J.; Shi, H.; Ma, Y. Characterization of Positional Isomers of Interchain Cysteine Linked Antibody−Drug Conjugates by High-Resolution Mass Spectrometry. *Anal. Chem.* **2019**, *91*, 8558–8563.

(76) Bayat, P.; Lesage, D.; Cole, R. B. Tutorial: Ion Activation in Tandem Mass Spectrometry Using Ultra-High Resolution Instrumentation. *Mass Spectrometry Reviews* **2020**, *39*, 680–702.

(77) Macias, L. A.; Santos, I. C.; Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2020**, *92*, 227–251.

(78) Révész, Á.; Rokob, T. A.; Jeanne Dit Fouque, D.; Hüse, D.; Háda, V.; Turiák, L.; Memboeuf, A.; Vékey, K.; Drahos, L. Optimal Collision Energies and Bioinformatics Tools for Efficient Bottom-up Sequence Validation of Monoclonal Antibodies. *Anal. Chem.* **2019**, *91*, 13128–13135.

(79) S. Brodbelt, J. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chemical Society Reviews* **2014**, *43*, 2757–2783.

(80) Brodbelt, J. S.; Morrison, L. J.; Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem. Rev.* **2020**, *120*, 3328–3380.

(81) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 Nm and Its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, *9*, 4205–4214.

(82) Zhang, L.; Reilly, J. P. De Novo Sequencing of Tryptic Peptides Derived from Deinococcus Radiodurans Ribosomal Proteins Using 157 Nm Photodissociation MALDI TOF/TOF Mass Spectrometry. *J. Proteome Res.* **2010**, *9*, 3025–3034.

(83) Fung, Y. M. E.; Kjeldsen, F.; Silivra, O. A.; Chan, T. W. D.; Zubarev, R. A. Facile Disulfide Bond Cleavage in Gaseous Peptide and Protein Cations by Ultraviolet Photodissociation at 157 Nm. *Angewandte Chemie International Edition* **2005**, *44*, 6399–6403.

(84) Agarwal, A.; Diedrich, J. K.; Julian, R. R. Direct Elucidation of Disulfide Bond Partners Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2011**, *83*, 6455–6458.

(85) Quick, M. M.; Crittenden, C. M.; Rosenberg, J. A.; Brodbelt, J. S. Characterization of Disulfide Linkages in Proteins by 193 Nm Ultraviolet Photodissociation (UVPD) Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 8523–8530.

(86) Macias, L. A.; Brodbelt, J. S. Investigation of Product Ions Generated by 193 Nm Ultraviolet Photodissociation of Peptides and Proteins Containing Disulfide Bonds. *J. Am. Soc. Mass Spectrom.* **2022**, *33*, 1315–1324.

(87) Robinson, M. R.; Taliaferro, J. M.; Dalby, K. N.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation Mass Spectrometry for Phosphopeptide

Characterization in the Positive and Negative Ion Modes. *J. Proteome Res.* **2016**, *15*, 2739–2748.

(88) Escobar, E. E.; Venkat Ramani, M. K.; Zhang, Y.; Brodbelt, J. S. Evaluating Spatiotemporal Dynamics of Phosphorylation of RNA Polymerase II Carboxy-Terminal Domain by Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2021**, *143*, 8488–8498.

(89) Ko, B. J.; Brodbelt, J. S. Comparison of Glycopeptide Fragmentation by Collision Induced Dissociation and Ultraviolet Photodissociation. *International Journal of Mass Spectrometry* **2015**, *377*, 385–392.

(90) Escobar, E. E.; Wang, S.; Goswami, R.; Lanzillotti, M. B.; Li, L.; McLellan, J. S.; Brodbelt, J. S. Analysis of Viral Spike Protein N-Glycosylation Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2022**, *94*, 5776–5784.

(91) Holden, D. D.; Brodbelt, J. S. Improving Performance Metrics of Ultraviolet Photodissociation Mass Spectrometry by Selective Precursor Ejection. *Anal. Chem.* **2017**, *89*, 837–846.

(92) Sanders, J. D.; Mullen, C.; Watts, E.; Holden, D. D.; Syka, J. E. P.; Schwartz, J. C.; Brodbelt, J. S. Enhanced Sequence Coverage of Large Proteins by Combining Ultraviolet Photodissociation with Proton Transfer Reactions. *Anal. Chem.* **2020**, *92*, 1041–1049.

(93) O. Zhurov, K.; Fornelli, L.; D. Wodrich, M.; A. Laskay, Ü.; O. Tsybin, Y. Principles of Electron Capture and Transfer Dissociation Mass Spectrometry Applied to Peptide and Protein Structure Analysis. *Chemical Society Reviews* **2013**, *42*, 5014–5030.

(94) Tureček, F.; Julian, R. R. Peptide Radicals and Cation Radicals in the Gas Phase. *Chem. Rev.* **2013**, *113*, 6691–6733.

(95) Coon, J. J. Collisions or Electrons? Protein Sequence Analysis in the 21st Century. *Anal. Chem.* **2009**, *81*, 3208–3215.

(96) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal. Chem.* **2018**, *90*, 40–64.

(97) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2006**, *1764*, 1811–1822.

(98) Kovalyov, S. V.; Zhokhov, S. S.; Onoprienko, L. V.; Vaskovsky, B. V.; Lebedev, A. T. Exploration of Doubtful Cases of Leucine and Isoleucine Discrimination in Mass Spectrometric Peptide Sequencing by Electron-Transfer and Higher-Energy Collision Dissociation-Based Method. *Eur J Mass Spectrom (Chichester)* **2017**, *23*, 376–384.

(99) Zhokhov, S. S.; Kovalyov, S. V.; Samgina, T. Yu.; Lebedev, A. T. An EThcD-Based Method for Discrimination of Leucine and Isoleucine Residues in Tryptic Peptides. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1600–1611.

(100) Lebedev, A. T.; Damoc, E.; Makarov, A. A.; Samgina, T. Yu. Discrimination of Leucine and Isoleucine in Peptides Sequencing with Orbitrap Fusion Mass Spectrometer. *Anal. Chem.* **2014**, *86*, 7017–7022.

(101) Cole, S. R.; Ma, X.; Zhang, X.; Xia, Y. Electron Transfer Dissociation (ETD) of Peptides Containing Intrachain Disulfide Bonds. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 310–320.

(102) Liu, F.; Breukelen, B. van; Heck, A. J. R. Facilitating Protein Disulfide Mapping by a Combination of Pepsin Digestion, Electron Transfer Higher Energy Dissociation (EThcD), and a Dedicated Search Algorithm SlinkS *. *Molecular & Cellular Proteomics* **2014**, *13*, 2776–2786.

(103) Gammelgaard, S. K.; Petersen, S. B.; Haselmann, K. F.; Nielsen, P. K. Characterization of Insulin Dimers by Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1910–1918.

(104) Larson, E. J.; Zhu, Y.; Wu, Z.; Chen, B.; Zhang, Z.; Zhou, S.; Han, L.; Zhang, Q.; Ge, Y. Rapid Analysis of Reduced Antibody Drug Conjugate by Online LC-MS/MS with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2020**, *92*, 15096–15103.

Chapter 2: Experimental Methods

2.1 Overview

The methods described here are intended to extend the applications of advanced tandem mass spectrometry (MS/MS) techniques, especially ultraviolet photodissociation (UVPD). To that end, antibody drug conjugates, synthetic selenoproteins, and immunopeptides are examined. In addition to UVPD, electron transfer dissociation (ETD), and collisional dissociation are explored, both to benchmark the capabilities of these methods, and to obtain complimentary information by combining information from multiple sources. General descriptions of the mass spectrometry, ion activation, chromatography, and data interpretation are described here with additional details included for each project in the remaining chapters.

2.2 Mass Spectrometry Instrumentation

2.2.1 Orbitrap Mass Analyzer

The development of orbitrap mass analyzers revolutionized proteomics by increasing the accessibility of high-resolution Fourier transform mass spectrometry (FTMS).[1] The increased resolving power enables the characterization of intact proteins, where highly adjacent mass-to-charge rations (*m/z*) would otherwise be indistinguishable without the use of specialized Fourier transform ion cyclotron resonance (FT-ICR) analyzers.[2] The high mass accuracy offered by FTMS and orbitrap mass analyzers is also critical to establishing the identity of fragment ions in the dense spectra generated via UVPD.[3] The work described here was enabled by Thermo Scientific Orbitrap Fusion Lumos Mass Spectrometers.

2.2.2 Electrospray Ionization

The combination of the orbitrap mass analyzer with electrospray ionization (ESI) was also integral for the analysis of peptides and proteins. ESI is an electrostatic nebulization process that allows the desolvation and ionization of liquid samples, with no limit on the molecular weight of the analyte, greatly expanding the repertoire of biomolecules that can be characterized by mass spectrometry.[4] ESI is readily adaptable to liquid chromatography (LC), and gentle enough to ionize noncovalent protein complexes.[4] Nano-ESI has expanded the use of ESI, allowing for flow rates as low as 20 nL/min and generating increased sensitivity and desolvation.[5,6] The nano-ESI setup used for direct infusion here employs house-made pulled fused silica capillaries which have been coated with Au/Pd. Approximately 10 μL of sample are loaded into the pulled capillaries and approximately 0.8 to 1.2 kV is applied allowing for static infusion with a Thermo Scientific Nanospray Flex ion source. Nano-ESI can also be coupled with chromatography, as described in section 2.4.

2.2.3 Ion Activation

Enhancing the applications of UVPD is the focus of this work. UVPD entails the absorption of ultraviolet photons by the analyte ion, resulting in energization of the ion, excitation to an excited electronic state, and direct fragmentation from the excited state or internal conversion to the ground electronic state and intramolecular vibrational energy redistribution prior to dissociation. For 193 nm UVPD, a Thermo Scientific Orbitrap Fusion Lumos Mass Spectrometer was modified to preform UVPD in the high-pressure linear ion trap by coupling a 193 nm Coherent Existar excimer laser (500 Hz, 5 nsec pulse, 1-3 mJ/pulse) to the back end of the instrument, as previously described.[7] In some cases, a commercial version of UVPD, utilizing a 213 nm solid-state Nd:YAG laser

(2500 Hz, 1 nsec pulse, ~3 uJ/pulse), was also explored in an aim to increase the transferability of the methods.

Collisional dissociation-based methods, including higher-energy collision dissociation (HCD) represents the most popular methods for ion activation of peptides and proteins. As such, HCD was included in each application as a benchmark. For the process of collisional activated dissociation, ions are accelerated to higher kinetic energies prior to collisions with an inert target gas, resulting in conversion of kinetic energy of the ion into internal energy that results in fragmentation. ETD and electron-transfer/higher-energy collision dissociation (EThcD) are also promising advanced fragmentation techniques which are included to compare the information achievable with alternative methods, and to help develop combined strategies where complementary information can be gleaned. ETD is accomplished through an ion-ion reaction between a reagent anion and a positively charged analyte, resulting in exothermic transfer of an electron from the reagent anion to the analyte cation. EThcD is a hybrid activation method in which all ETD products are subjected to collisional activation prior to mass analysis.

2.3 LIQUID CHROMATOGRAPHY

Liquid chromatography experiments were enabled by a Dionex Ultimate nano LC system. Trap and elute configurations were used to allow for sample concentration and desalting on a trap column at a flow rate of 5 uL/min followed by a valve switch placing the trap column in-line with the analytical column. Typical loading conditions used 2% acetonitrile and 0.1% formic acid in water. Chromatographic separation occurred at a flow rate of 300 nL/min, with solvents consisting of water (A) and acetonitrile (B) with 0.1% formic acid. Gradient elution was optimized for each method. Reversed-phase

PLRP media was used to separate larger middle-down sized peptides and reversed-phase C18 media was used for bottom-up and immunopeptidomics workflows. Voltages of approximately 1.8 to 2.2 kV were applied for nanoESI.

2.4 Data Interpretation

For each study a combination of software and some form of manual validation was utilized. Key mass spectrometry data processing suites and their applications in relation to the work described in this dissertation are detailed here.

2.4.1 Byonic

Byonic (Protein Metrics Inc.) was utilized to process all bottom-up and immunopeptidomics data. Byonic matches peptides from a database to MS/MS through a combination of intact mass and sequence ions.[8] While several similar platforms are available, Byonic has been extensively developed for UVPD, making it ideal for the applications presented here.[9]

2.4.2 ProSight

ProSight Lite enables the identification and visualization of fragment ions generated in top-down mass spectrometry data by matching deconvoluted peak lists to a list of theoretical masses.[10] To generate deconvoluted peak lists, MS/MS data were deconvoluted and deisotoped with Xtract deconvolution in Thermo Scientific FreeStyle. Signal-to-noise ratios and deconvolution fit factors were established for each application to ensure only high-quality fragment ions were retained in deconvoluted peaks list. Predicted isotope patterns for exemplative fragment ions with both high and low signal to noise ratios and fit factors were typically visualized in TD Validator demonstrate the quality of the data.

ProSightPD for Proteome Discoverer enables the search of top-down proteomic data against a database of proteoforms. Each spectrum in a LC-MS/MS data file is deconvoluted with Xtract, and then matched to the proteoform database based on intact mass and fragmentation data. The use of deconvolution decongests and simplifies MS/MS data. As a result, ProSightPD is better suited to top-down and middle-down applications than Byonic because larger peptides and proteins yield more complex MS/MS spectra.

2.4.3 TDValidator

TDValidator also identifies top-down fragment ions, but directly uses raw MS/MS data and matches observed and predicted isotope patterns.[11] Because TDValidator directly interprets raw data based on a predicted sequence, it is often able to identify fragments missed by deconvolution algorithms.[11] This increased identification of fragment ions, means that TDValidator is best applied in cases where the protein sequence is already well established or as a follow-up to deconvolution-based identification.[11] In Chapter 3 TDValidator was used to facilitate the complete identification of fragment ions for ADC subunits, in which case each payload containing fragment ion was manually examined and excluded as necessary. TDValidator was also utilized in other chapters to compliment to ProSight data.

2.5 REFERENCES

(1) Zubarev, R. A.; Makarov, A. Orbitrap Mass Spectrometry. Anal. Chem. 2013, 85, 5288–5296.

(2) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. Annual Review of Analytical Chemistry 2015, 8, 61–80.

(3) Brodbelt, J. S.; Morrison, L. J.; Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. Chem. Rev. 2020, 120, 3328–3380.

(4) Wilm, M. Principles of Electrospray Ionization. Molecular & Cellular Proteomics 2011, 10.

(5) Wilm, M.; Mann, M. Analytical Properties of the Nanoelectrospray Ion Source. Anal. Chem. 1996, 68, 1–8.

(6) Shen, Y.; Zhao, R.; Berger, S. J.; Anderson, G. A.; Rodriguez, N.; Smith, R. D. High-Efficiency Nanoscale Liquid Chromatography Coupled On-Line with Mass Spectrometry Using Nanoelectrospray Ionization for Proteomics. Anal. Chem. 2002, 74, 4235–4249.

(7) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2016, 88, 1044–1051.

(8) Bern, M.; Kil, Y. J.; Becker, C. Byonic: Advanced Peptide and Protein Identification Software. Curr Protoc Bioinformatics 2012, CHAPTER, Unit13.20.

(9) Greer, S. M.; Bern, M.; Becker, C.; Brodbelt, J. S. Extending Proteome Coverage by Combining MS/MS Methods and a Modified Bioinformatics Platform Adapted for Database Searching of Positive and Negative Polarity 193 Nm Ultraviolet Photodissociation Mass Spectra. J. Proteome Res. 2018, 17, 1340–1347.

(10) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical Software to Analyze Top-down Mass Spectrometry Data. PROTEOMICS 2015, 15, 1235–1238.

(11) Fornelli, L.; Srzentić, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. Anal. Chem. 2018, 90, 8421–8429.

Chapter 3: Comprehensive Middle-Down Mass Spectrometry Characterization of an Antibody-Drug Conjugate by Combined Ion Activation Methods[1]

3.1 ABSTRACT

Antibody-drug conjugates (ADCs) are an increasingly prevalent drug class utilized as chemotherapeutic agents. The complexity of ADCs, including their large size, array of drug conjugation sites, and heterogeneous compositions containing from zero to several payloads, demand the use of advanced analytical characterization methods. Tandem mass spectrometry (MS/MS) strategies, including a variety of bottom-up, middle-down, and even top-down approaches, frequently applied for analysis of antibodies are increasingly being adapted for antibody-drug conjugates. Middle-down tandem mass spectrometry, often focusing on analysis of ~25 kDa protein subunits, offers the potential for complete sequence confirmation as well as the identification of multiple conjugation states. While middle-down studies have been extensively developed for monoclonal antibodies, middle-down characterization of ADCs has been limited by the high complexity of the drug molecules. This study seeks to bridge the gap by utilizing a combination of 193 nm ultraviolet photodissociation (UVPD), electron-transfer dissociation (ETD), and electron-transfer/higher-energy collision dissociation (EThcD). The compilation of these MS/MS methods leads to high sequence coverages of 60-80% for each subunit of the ADC. Moreover, the combined fragmentation patterns provide sufficient information to allow confirmation of both the sequence of the complementarity-determining regions as well as the payload conjugation sites.

[1]Watts, E.; Williams, J. D.; Miesbauer, L. J.; Bruncko, M.; Brodbelt, J. S. Comprehensive Middle-Down Mass Spectrometry Characterization of an Antibody–Drug Conjugate by Combined Ion Activation Methods. *Anal. Chem.* **2020**, *92*, 9790–9798.

3.2 INTRODUCTION

Antibody-drug conjugates (ADCs) have emerged as a highly promising drug class to address therapeutic challenges, particularly for the treatment of cancer.[1–4] ADCs promise to improve patient outcomes by combining a drug, often a cytotoxic payload used for chemotherapy, with an antibody.[1–3,5] Ideally, the antibody affords enhanced recognition of diseased cells and fewer off-target effects. Traditional ADCs are very heterogeneous with the payload bound at multiple sites, varying from molecule to molecule. This heterogeneity leads to special hurdles in characterization of ADCs and highlights the need to establish reliable analytical techniques that can resolve a variety of structural details.

ADCs comprise three parts: the antibody, the payload and the linker which bridges the payload to the antibody (typically via lysine sidechains or cysteine thiols). Design of the linker is critical to ensure that the payload remains attached to the antibody as it circulates the body but is released upon entering the diseased cell.[6,7] Over the years many linker-strategies have been developed.[4,8] One common strategy targets primary amines which results in the payload binding to a large array of locations (e.g., multiple lysines and the N-termini) of the antibody.[4,8] Alternatively, in cysteine-bound ADCs, one or more of the disulfide bonds that bridge the antibody subunits are reduced, and payloads are conjugated to one or more cysteines. This process leads to the incorporation of zero to eight payloads bound per antibody. While the selective conjugation to cysteine limits the maximum number and locations of payloads per antibody, cysteine conjugation remains one of the most efficient and well-understood processes.[9,10] More selective strategies have also been developed, often involving incorporation of modified amino acids such as engineered heavy chain cysteines that allow exact drug-to-antibody ratios of two or four.[4,8] These highly ordered ADCs are commonly referred to as “next-generation

ADCs."[8] As the performance metrics of ADCs become better characterized, they are being designed to facilitate the delivery and release of more acutely toxic drugs.[8] In these cases the stability of the ADC in and out of plasma becomes increasingly important, and the ability to monitor the payload localization and decomposition of ADCs is even more critical.[10–15] Seven ADCs, brentuximab vedotin (Seattle Genetics), polatuzumab vedotin (Roche), enfortumab vedotin (Seattle genetics), Trastuzumab deruxtecan (AstraZeneca), trastuzumab emtansine (Roche), gemtuzumab ozogamicin (Pfizer) and inotuzumab ozogamicin (Pfizer), have been developed and approved for use in patients for various cancer-related applications. While the latter three ADCs have lysine-linked payloads, the other four contain cysteine-bound payloads.[8,16]

Mass spectrometry is employed for ADC analysis at all stages throughout development. A key goal of ADC characterization is the determination of drug-to-antibody ratio (DAR). The DAR is often established by separating the antibodies with different numbers of payloads, normally via chromatography, quantifying each species, and averaging the results.[17,18] In the case of cysteine-linked antibodies, non-denaturing separations, such as hydrophobic interaction chromatography, or capillary electrophoresis, are frequently employed to prevent disassembly of ADCs that are destabilized owing to the cleavage of the original cysteine bridges linking the subunits and replacement with payloads.[19–21] In some cases DAR may be estimated by high resolution mass-spectrometry without chromatography.[22,23] Estimation of DARs may be alternatively accomplished at the subunit level either by reducing interchain disulfide bonds to produce separate heavy and light chain subunits or by using enzymes like Fabricator or IdeZ to cleave the heavy chain of the antibody in half, generating 25 kDa subunits.[24–29] For this strategy, the antibody is denatured and can be separated by conventional reversed phase chromatography. Examination of the molecular weights of

the individual subunits provides information about the distribution of drug on the antibody.

Additional key goals of ADC characterization are the confirmation of the sequence of the antibody and localization of the payloads.[13] Sequence confirmation is particularly important in the variable region of the antibody, especially the highly variable complementarity-determining region (CDR). Confirming the sequence of the CDR is necessary to ensure that the antibody targets the correct cells. Because of the inherent complexity of ADCs, sequencing and confirmation of the payload sites are traditionally accomplished via bottom-up MS/MS methods which rely on proteolysis to cleave the antibody into small peptides prior to analysis.[13,17,24] While the bottom-up strategy is robust and well-established, the ability to determine combinatorial payload locations is lost.

Alternatives to bottom-up methods that have gained attention are top-down and middle-down approaches which utilize no or selective proteolysis, respectively. In general, the sequence coverage obtained from analysis of intact molecules (i.e. top-down) is limited because of the large molecular size and significant disulfide bond composition of antibodies;[14,23,25,30,31] these analytical restraints have accelerated the exploration and adoption of middle-down approaches. Digestion of ADCs using IdeZ or Fabricator proteases, as mentioned above, generates large sub-units that retain nearly the same information as intact antibodies but present sizes that are better suited for analysis by current MS/MS methods. In the case of cysteine-bound ADCs, this type of selective proteolysis creates subunits containing 1-3 payloads, in theory permitting simultaneous examination of all possible payload sites on each heavy chain. Integrating top-down and middle-down analysis techniques offers the opportunity to achieve even more comprehensive characterization of ADCs.

Characterization of antibody subunits by tandem mass spectrometry (MS/MS), typically coupled with reversed phase liquid chromatography, has been extensively explored.[14,30,32–38] Multiple MS/MS methods, including combinations of 213 and 193 nm ultraviolet photodissociation (UVPD), electron transfer dissociation (ETD), collisionally activated dissociation (CAD), higher-energy collisional dissociation (HCD), electron capture dissociation (ECD) and hybrid ETD/HCD (EThcD), have been used to enhance characterization, with coverages of 50-80% typically reported.[30,33,35,39] Fraction collection of subunits followed by MALDI-insource decay has also been utilized to obtain sequence coverages as high as 90% for each subunit.[34,40] While many MS1-based top-down and middle-level characterization methods for ADCs have been developed,[27,41–44] MS/MS analysis of ADCs is still in the early stages of development. Two recent studies have demonstrated the initial application of middle-down MS/MS strategies for analysis of ADC subunits. [45-46] One expansive study reported the use of ETD to effectively identify payload locations on heterogeneous antibodies by focusing on specific diagnostic fragment ions, even attaining localization of conjugation sites of positional isomers as well as oxidation sites.[45] Another investigation compared ETD, 213 nm UVPD and HCD for analysis of a next-generation ADC.[46] For the antibody in this aforementioned study, the payload is fixed on C213 for the Fc/2 subunit and C163 for the Fd' subunit via modified amino acids.[46] While next-generation ADCs represent a growing class of ADCs, the analysis is simpler than that of more conventional cysteine-bound ADCs owing to site-specific payload conjugation. HCD proved to be the least informative owing to internal fragmentation of the payload (thus confounding interpretation), whereas 213 nm UVPD generated more payload-containing fragment ions than ETD.[46] These studies have laid important groundwork for ADC analysis by using tandem mass

spectrometry, motivating further optimization and assessment to promote wider adoption by the ADC community.

The heterogeneity of the number and locations of payloads complicates the analysis of ADCs. Here we describe a middle-down MS/MS strategy that combines nano-scale liquid chromatography, IdeZ digestion, and ETD, EThcD and UVPD as three MS/MS methods for characterization of ADCs. This multi-pronged approach is demonstrated for analysis of a surrogate cysteine-bound ADC in which the linker is bound to a propranolol payload. This surrogate ADC is used to demonstrate the wide-reaching method presented in this study which should be generally applicable to many other types of complex ADCs.

3.3 METHODS

The surrogate antibody-drug conjugate was provided by AbbVie. The sequence of the antibody is based on human IgG1. The Mc-Val-Cit-PABC-PNP linker for linker fragmentation experiment was obtained from MedChemExpress. HPLC solvents were obtained from EMD Millipore (Temecula, CA), and buffer components were obtained from Sigma-Aldrich (St. Louis, MO). Proteomics-grade trypsin was obtained from Promega (Madison, WI). IdeZ protease was obtained from New England Biolabs inc. (Ipswich, MA). All other reagents and solvents were obtained from Thermo Fisher Scientific (Fairlawn, NJ). The CDR of the ADC was calculated through the Paratome online tool.[47,48]

3.3.1 Trypsin Digestion Protocol

ADC stock solution diluted to 200 µg/mL was reduced and alkylated in 50 mM ammonium acetate buffer. Dithiothreitol (DTT) was added at 250 mM, and the solution

was incubated at 55°C for 30 min. The solution was allowed to cool for 10 min prior to addition of iodoacetamide at 500 mM. The mixture was allowed to react for 30 min in the dark. The reaction was quenched by adding a second aliquot of DTT for a final concentration of 500 mM. After alkylation trypsin was added in a 1:20 protease-protein ratio. The digestion was performed at 37°C for 15 hours. After digestion, the reaction was quenched with 1% formic acid and cleaned using a spin cartridge loaded with C18 resin (Pierce Biotechnology) prior to LC−MS analysis.

3.3.2 Bottom-up LC-MS/MS analysis

Reversed phase liquid chromatography (LC) was performed with a Dionex Ultimate nano LC system interfaced to a Thermo Scientific Orbitrap Elite mass spectrometer. Columns were packed in-house using 3.5 μm XBridge BEH C18 media. Mobile phases were water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B). Elution was achieved via a gradient starting at 2% B and going to 35% B over two hours. MS/MS events were performed in a data-dependent manner with HCD NCE 30-35. Data analysis was performed using Byonic with a score cut-off of 400. The data was searched against the UniProt human proteome coupled with the sequence of the ADC. The mass of the payload was included as a rare modification to cysteine with a maximum of two rare modifications per peptide.

3.3.3 MS/MS analysis of Linker

The linker was dissolved in DMSO at a concentration of 10 mM. This stock solution was diluted to 10 μM in 49.5/49.5/1 water/acetonitrile/formic acid. Samples were loaded into Au/Pd coated pulled tip emitters for nano-electrospray infusion and analyzed using a Thermo Scientific Orbitrap Elite mass spectrometer equipped with a 500

Hz 193 nm Coherent ExciStar XS excimer laser (Santa Clara, CA) for UVPD as described previously.[49] UVPD data was collected using one laser pulse at 1.0 mJ and HCD data was collected using a collision energy setting of 10 NCE. MS/MS spectra were collected with 60,000 resolution and 5 microscans. Twenty scans were averaged to produce the final MS/MS spectra.

3.3.4 IdeZ Digestion Protocol

Middle-down digestion was performed by directly diluting the ADC in water and adding an appropriate amount of IdeZ and concentrated Glycobuffer (New England Biolabs) per manufacturer instructions. The digest was incubated at 37°C for 1 hour. After the digestion, reduction was performed for 35 min on a shaker at room temperature with 3.4 M urea and 40 mM tris-(2-carboxyethyl)phosphine (TCEP). Following reduction, 1% formic acid was added to prevent disulfide bond reformation. The sample was cleaned up with Zeba spin desalting columns (Thermo Fisher Scientific) and buffer exchanged into 0.1% formic acid for LC-MS analysis.

3.3.5 Middle-down LC-MS/MS analysis

PLRP columns were packed in-house using 1000 Å, 50 µm bulk media from Agilent. Mobile phases were water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B) applied in a rapid initial gradient of 2 to 30% B over three minutes followed by a shallow gradient up to 40% B over 22 minutes at a flow rate of 300 nL/min. Mass spectrometry was performed on a Thermo Orbitrap Fusion Lumos mass spectrometer equipped with a 500 Hz 193 nm Coherent ExciStar XS excimer laser as previously described for UVPD.[50] The following activation parameters were optimized: activation time for ETD and EThcD, laser energy and number of laser pulses for UVPD.

In addition, the width of the precursor isolation window and the impact of the selected precursor charge state were evaluated. A detailed description of the results of these experiments is included in section 3.4.2. Unless otherwise specified, optimized parameters were used: 193 nm UVPD was performed using a pulse energy of 1 mJ and one laser pulse. The activation period for ETD was 15 ms. EThcD was performed using a 10 ms activation period with 10 NCE supplemental activation. An isolation window of 5 m/z was used across all methods centered on the 27+, 23+ and 25+ charge states for Fc/2, Lc, and Fd', respectively. Tandem mass spectra were collected in full profile mode.

Data analysis was completed using TDValidator 1.0 by Proteinaceous. In order to ensure validity of matching fragment assignments, strict data cutoffs were employed, including 5 ppm for the maximum error in fragment ions, 10 for the minimum signal-to-noise ratio, and 0.75, on a scale from 0 to 1 (with 1 representing a perfect match between predicted and observed isotope patterns), for the minimum score. Fragment ions were then further refined by manual validation, a process that entailed examining the isotopic profiles of each identified fragment ion to eliminate overlapping and uncertain assignments. Examples of cases in which fragment ions with poorly fitting isotopic profiles were eliminated include those where upon manual examination most of the isotope peaks were very low abundance and did follow an expected distribution or where it was clear that the identification was made based on a wrong charge state assignment, i.e. if a 14+ distribution was identified as a 7+ ion. For overlapping ions, the best matching ion was only included if there was an improvement in both the score (at least 0.5 better) or ppm error (at least 2 fold lower), otherwise both ions were eliminated from consideration. For UVPD there were many instances where both "a" and "a + 1", "x" and "x + 1", and "y" and "y - 1" ions were identified for the same ion distribution. While it is

likely that these distributions are comprised of a combination of multiple ion types, for simplicity the match with the highest score was considered in these cases.

An excel spreadsheet listing all fragment ions identified and manually confirmed using TDValidator for each subunit, including ppm errors, and signal to noise ratios, and scores for each fragment ion is available as a supporting information file at https://doi.org/10.1021/acs.analchem.0c01232.

3.4 Results And Discussion

3.4.1 Bottom-Up Analysis

To provide benchmark results using a traditional bottom-up workflow, the ADC was subjected to trypsin digestion and conventional LC-MS/MS using HCD. While sequence coverage of the protein was impressive (97% for both the heavy chain and light chain), few peptides containing the payload were identified. Only three MS/MS spectra (out of 388 matched spectra for the heavy chain) were matched to peptides containing the payload, representing two peptides from the heavy chain. One of these matches was identified as a peptide containing Cys233, a cysteine normally involved in an inter-chain disulfide bond. However, as displayed in **Figure 3.1**, none of the fragment ions contained the payload. The identification is believed to be a correct match as payload conjugation favors these cysteine residues;[24,41] however, it is difficult to confirm this based on the fragmentation pattern. The payload was also localized to Cys265 based on another peptide identified in the LC-MS/MS run, an unexpected finding owing to the prevalent involvement of this cysteine in an intra-chain disulfide bond which makes it an unlikely candidate for modification. The MS/MS spectrum for this peptide is displayed in **Figure 3.2**. There are, once again, no payload-containing fragments and additionally there are many unmatched fragment ions, making this a probable incorrect match. This second

payload binding site was not corroborated by the middle-down results (discussed later) and was thus classified as a false identification.

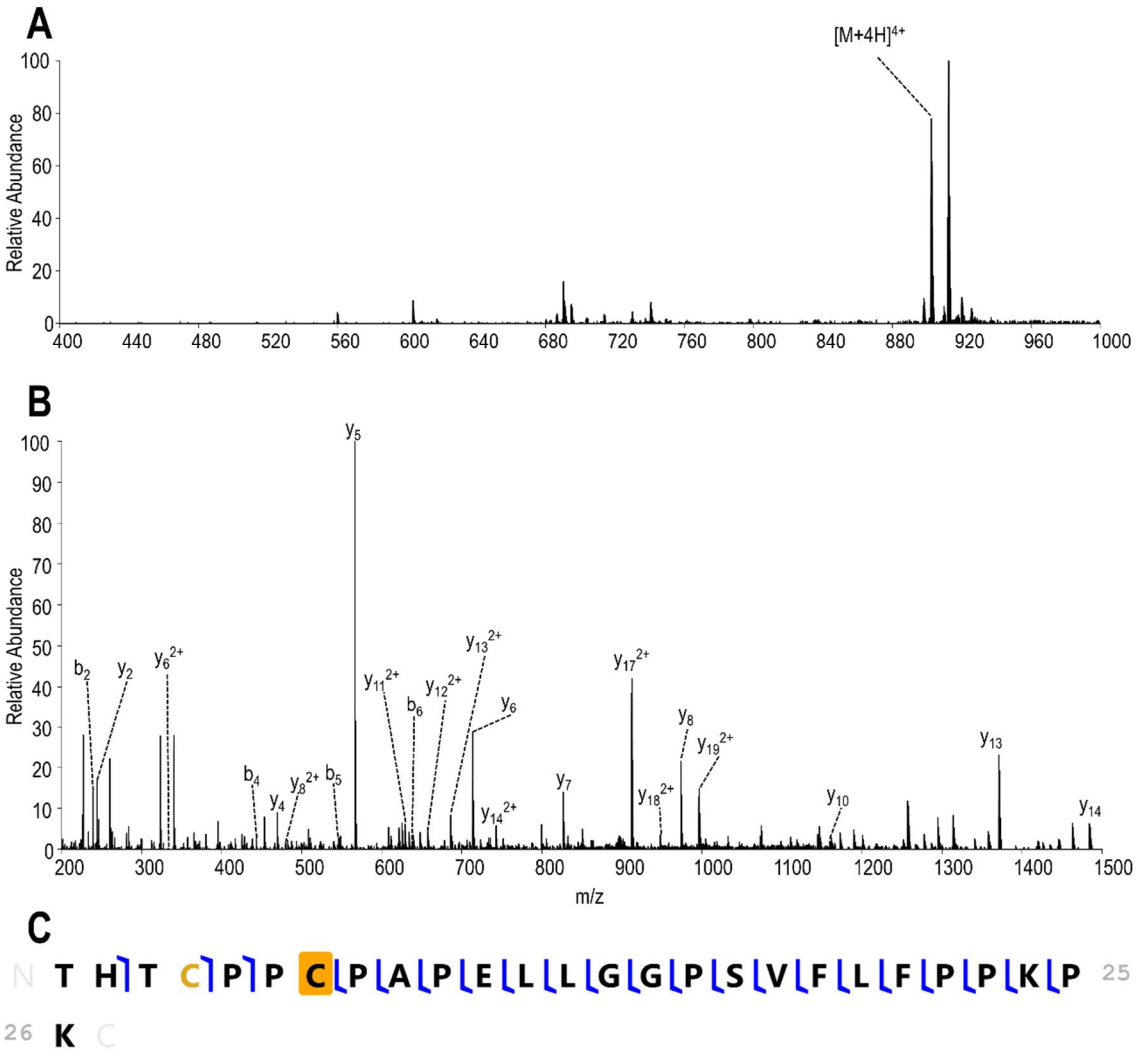


Figure 3.1: (A) MS1 and (B) MS/MS spectra (HCD NCE 30, 4+ charge state) of one payload-containing peptide identified from the Byonic database search of the bottom-up dataset. The spectra are shown for a peptide which is believed to be a correct match. (C) The resulting fragmentation map; no payload-containing product ions were identified. The cysteine highlighted in gold is presumed to contain the payload conjugation

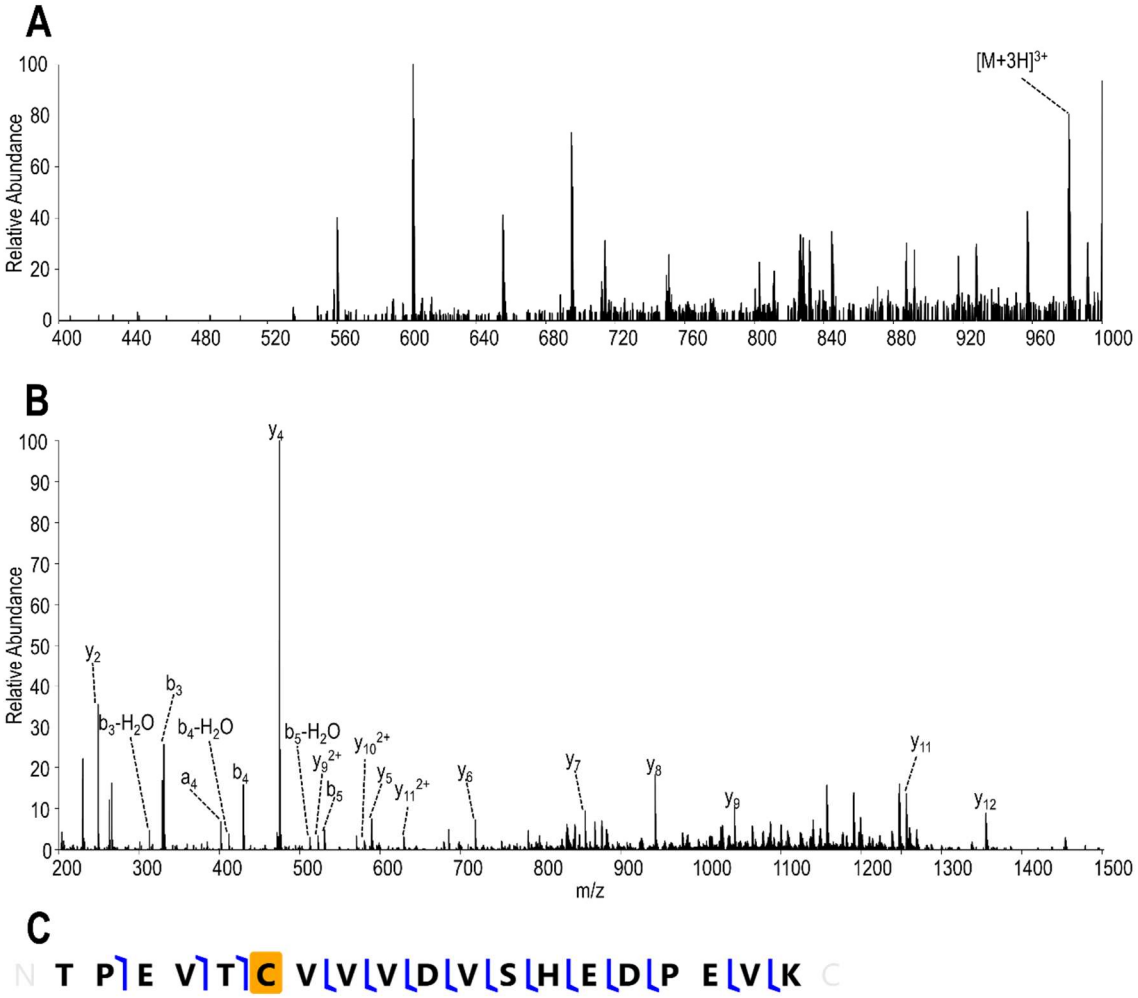


Figure 3.2: (A) MS1 and (B) MS/MS spectra (HCD NCE 30, 3+ charge state) of one payload-containing peptide identified from the Byonic database search of the bottom-up dataset. The spectra are shown for a peptide which is believed to be a false match. (C) The resulting fragmentation map; no payload-containing product ions were identified. The cysteine highlighted in gold is presumed to contain the payload conjugation

Two of the inter-chain cysteines (Cys230 and Cys233) of the heavy chain that are likely candidates for modification were characterized based on the bottom-up LC-MS/MS strategy, and one of them was identified as being conjugated to a payload. The other

inter-chain cysteine (Cys224) was not identified, likely owing to the presence of adjacent lysines leading to the production of peptides too small for successful MS/MS confirmation. Similarly, no peptides containing the inter-chain cysteine (Cys214) of the light chain were identified, thus preventing the successful characterization of the C-terminal section of the light chain and prohibiting confirmation of payload conjugation at that site. The tryptic peptide that should contain this cysteine (sequence GEC) would be 3-mer which is significantly smaller than the peptides that were identified. Further bottom-up studies, utilizing alternative enzymes or limited digestions, could improve the characterization of the ADC. However, given the current shortcoming of bottom-up methods in the context of ADCs, advancing the methods for subunit level middle-down characterization is desirable. Based on these results, the bottom-up approach proved ineffective for characterization of payload sites, reinforcing the need for alternative MS/MS strategies.

3.4.2 Linker Fragmentation

Another challenge associated with the MS/MS characterization of ADCs is the occurrence of additional fragmentation pathways originating from cleavages of the drug and/or the linker. These pathways are expected to contribute to the complexity of fragmentation patterns generated for peptides (bottom-up methods), subunits (middle-down) and intact ADCs (top-down). The fragmentation of the linker used in the present study was explored using HCD and UVPD. The resulting MS/MS spectra are displayed in **Figure 3.3**. Three major cleavage sites of the linker are mapped onto its structure in **Figure 3.3C**, and all are anticipated to be active during activation and MS/MS analysis of the ADC. The fragmentation of the linker as a result of ETD and EThcD could not be examined because the linker only produces singly charged ions upon ESI; however,

fragmentation of the linker is expected to occur upon ETD or EThcD of the multi-charged ADCs. Specific searches for fragment ions containing a partially cleaved linker yielded only a handful of additional fragment ions. While fragmentation of the linker certainly occurs for the ADC subunits, only fragments containing the entire payload were considered in the searches in the present study to minimize the impact on the false discovery rate arising from inclusion of multiple cleavage locations of the linker moiety.

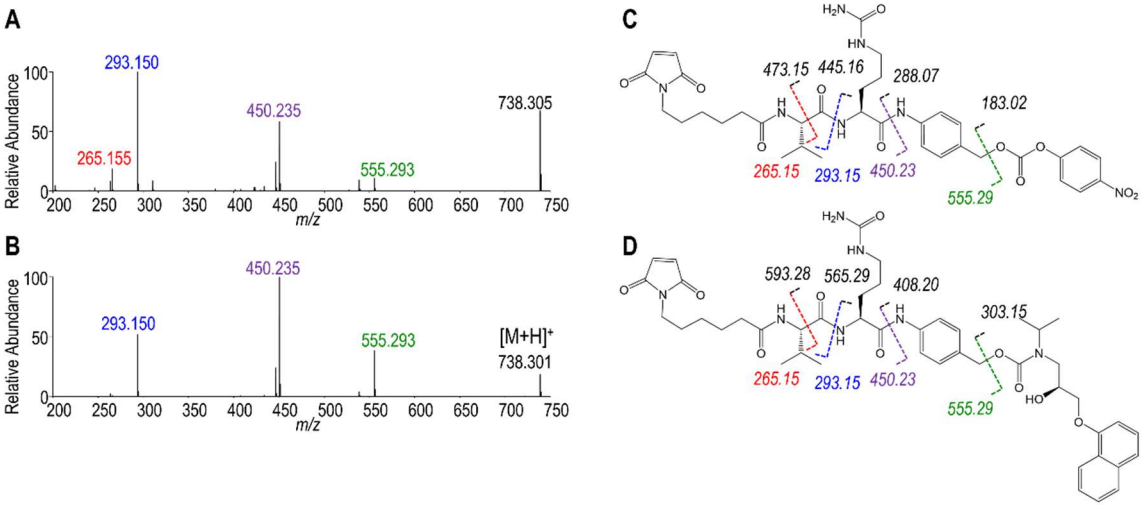


Figure 3.3: MS/MS spectra of the linker (1+ charge state). (A) UVPD (1 pulse, 1 mJ), (B) HCD (10 NCE). (C) The structure of the linker is illustrated with proposed color-coded cleavages that lead to the observed fragment ions along with the corresponding neutral losses indicated in black. (D) The same cleavages are mapped onto the linker-with-payload structure to indicate several predicted fragmentation pathways and expected neutral losses that might contribute to fragment ions in the MS/MS spectra of ADCs.

3.4.3 Optimization of Middle-Down Activation Parameters

Characterization of payload-containing subunits present additional challenges, thus requiring careful optimization of activation conditions to maximize coverage and bracket modified sites. Owing to gaps in sequence coverage by any single MS/MS method, it was anticipated that the results from multiple activation methods could be

combined to afford a higher level of characterization (a strategy also commonly employed for the analysis of antibodies). The experimental parameters of greatest impact include the laser energy and number of laser pulses for UVPD, the ETD activation time, the HCD collision energy, and the width of the mass isolation of the precursor ion. The influence of the charge state of the precursor ion was also evaluated. Each of the activation methods was performed throughout separate LC runs, during which the masses of the payload-conjugated subunits and unmodified subunits were targeted for isolation and activation at prescribed times based on the elution profile displayed in **Figures 3.4-3.6**. All figures display a representative standard deviation for one set of conditions in order to give an idea of the variability.

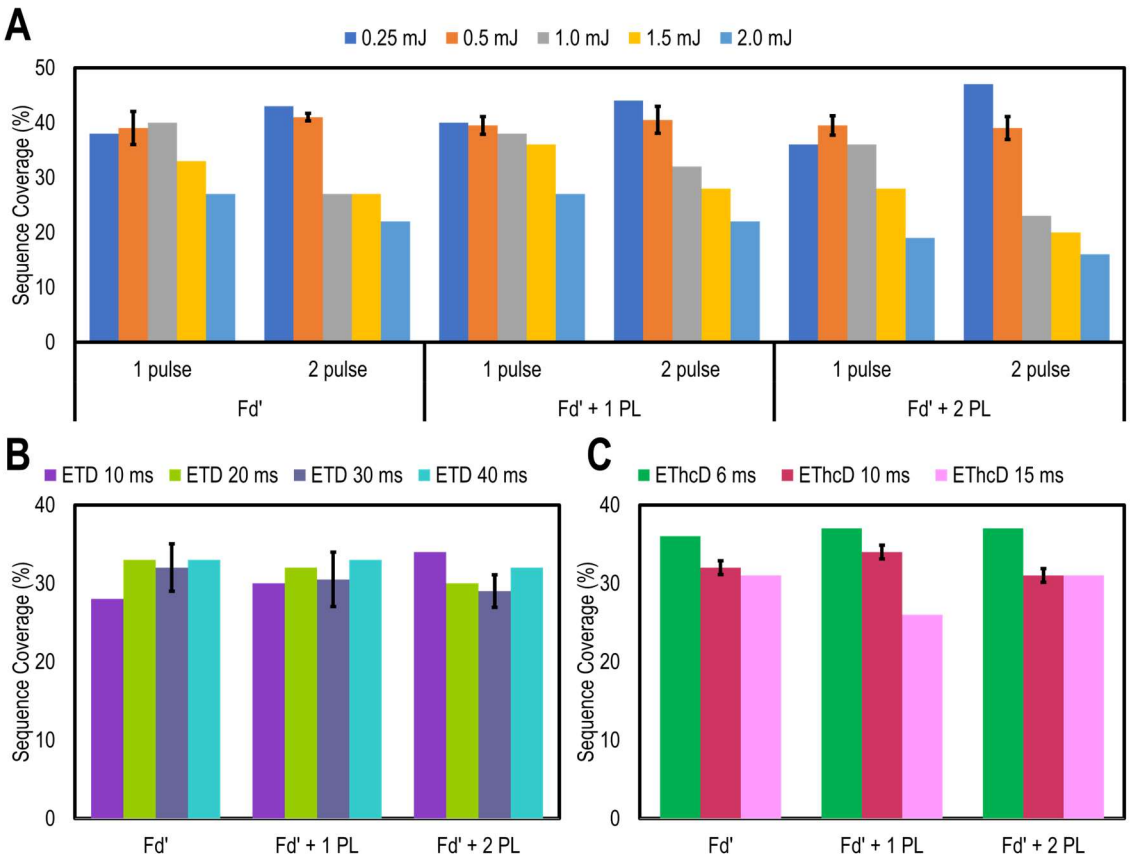


Figure 3.4: Optimization of activation conditions for (A) UVPD (1 pulse): variation of laser energy, (B) ETD: variation of activation time, and (C) EThcD (HCD energy 10 NCE): variation of activation time. For this set of comparisons, the 25+ charge state was selected with an isolation window of 10 *m/z*. Results are displayed with the payload (PL) located at C224 for Fd' + 1 PL and C230 and C233 for Fd' + 2 PL. The typical level of standard deviation is shown for one set of conditions for each activation mode. Error bars represent a 95% confidence interval based on four replicates.

While the data for the final study was collected in full profile mode and analyzed in TDValidator, optimization data was collected in reduced profile and a streamlined data analysis method was employed in order to allow efficient data analysis using Biopharma Finder. For Biopharma Finder data analysis, a signal-to-noise ratio of 3 and a maximum

ppm error of 5 were employed. These standards are typical for reduced profile data which results in lower levels of noise (and less spectral detail) than full profile data.

Figure 3.4A displays the impact of laser energy on the sequence coverage. There is not a dramatic change in the outcome of UVPD using one laser pulse versus two pulses in **Figure 3.4A.** Using a single pulse resulted in a slightly lower dependency on the specific laser energy, especially between 0.25-1.5 mJ. The rationale for this is that when multiple pulses are applied the effects of any variation in laser energy are compounded. In addition, while 0.25 mJ seemed to provide the highest sequence coverage using two laser pulses, slightly higher laser energy (0.50-1.0 mJ) offered better results when using a single laser pulse. Owing to these comparisons, for the remainder of the study a single pulse was used for UVPD because of the higher degree of reproducibility found in the optimization data and the general trend towards higher sequence coverages observed for each subunit. UVPD using 0.25-1.0 mJ yielded similar degrees of sequence coverage; however, increasing the energy per laser pulse beyond 1.0 mJ caused a loss of sequence coverage. This net decrease in sequence coverage is attributed to the prevalence of secondary dissociation and/or excessive energy deposition that contributes to formation of smaller and internal fragment ions, often ones that are redundant or unassignable. Based on the trends in **Figure 3.4A**, a laser energy of 1.0 mJ was selected. Following a similar strategy, the ETD activation time was optimized for ETD and EThcD. Variation of the ETD activation time resulted in relatively modest gains or losses of sequence coverage (+5% relative to the mean coverage) as shown in **Figure 3.4B,C**. In order to balance the potential to generate more fragments by using longer activation times with the need for collecting as many spectra as possible to allow averaging for the full profile data analysis, an ETD time of 15 ms and an EThcD time of 10 ms were selected.

Figures 3.5 and 3.6 illustrate the optimization of precursor charge state and isolation window. There are a few minor differences between the parameters used to initially optimize precursor charge state and isolation window compared to the finalized parameters. First, based on initial examination of **Figure 3.4** a laser energy of 0.5 mJ per pulse was selected in order to maximize sequence coverage for UVPD spectra acquired using either one or two pulses. After selection of one laser pulse, a final laser energy of 1.0 mJ was chosen owing to most consistent performance. Similarly, initially an ETD activation time of 30 ms was utilized for optimization of the size of the isolation window and precursor charge state in order to enhance the production of complementary fragment ions for ETD and EThcD (10 ms, 10 NCE). After extensive comparisons, it was determined that using an activation time that yielded the highest sequence coverage for ETD offered the best comparison to the complementary EThcD data. Because the results for charge state and isolation window were generally consistent across activation methods, further re-evaluation of these parameters was not undertaken.

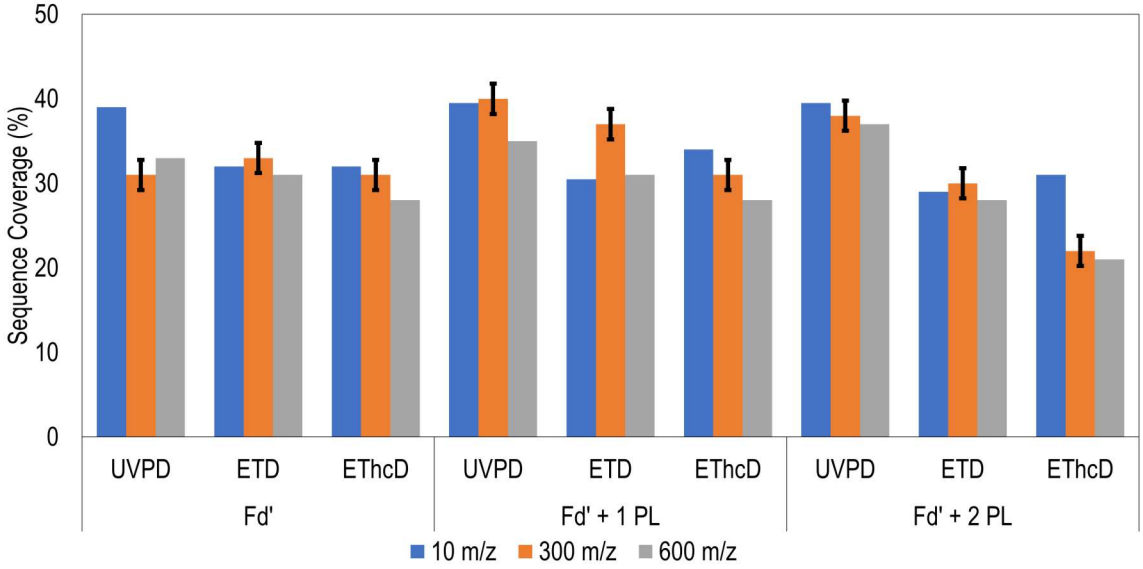


Figure 3.5: Optimization of precursor isolation window. The reaction time was set to 30 ms for ETD and 10 ms for EThcD (NCE 10). For UVPD one 0.5 mJ pulse was applied. The 25+ charge state was selected for this set of comparisons. Results are displayed with the payload (PL) located at C224 for Fd' + 1 PL and C230 and C233 for Fd' + 2 PL. The typical level of standard deviation is shown for one set of conditions for each activation mode. Error bars represent a 95% confidence interval based on four replicates.

The precursor isolation width was also optimized as displayed in **Figure 3.5** for UVPD (1 pulse, 0.5 mJ), ETD (30 ms) and EThcD (10 ms, NCE 10) for the three targeted Fd' species. While other studies have demonstrated improved sequence coverage for ETD and UVPD of antibody subunits using broad isolation windows,[30,32] decreased sequence coverage was observed when broad isolation windows (300 *m/z* or 600 *m/z*) were applied for the ADC subunits in the present study. The loss of sequence coverage is rationalized by considering the features of the MS/MS methods. UVPD produces a large variety of product ions. The fragmentation patterns become overly dense upon activation of multiple co-isolated charge states, resulting in overlapping isotopic envelopes of the ions and impeding ion assignments. For ETD and EThcD, co-isolation and reaction of multiple charge states results in rapid depletion of the ETD reagent ions and generation of

product ions in numerous charge states, resulting in lower S/N and loss of sequence coverage. Based on these results, individual charge states were isolated and activated for the remainder of the study.

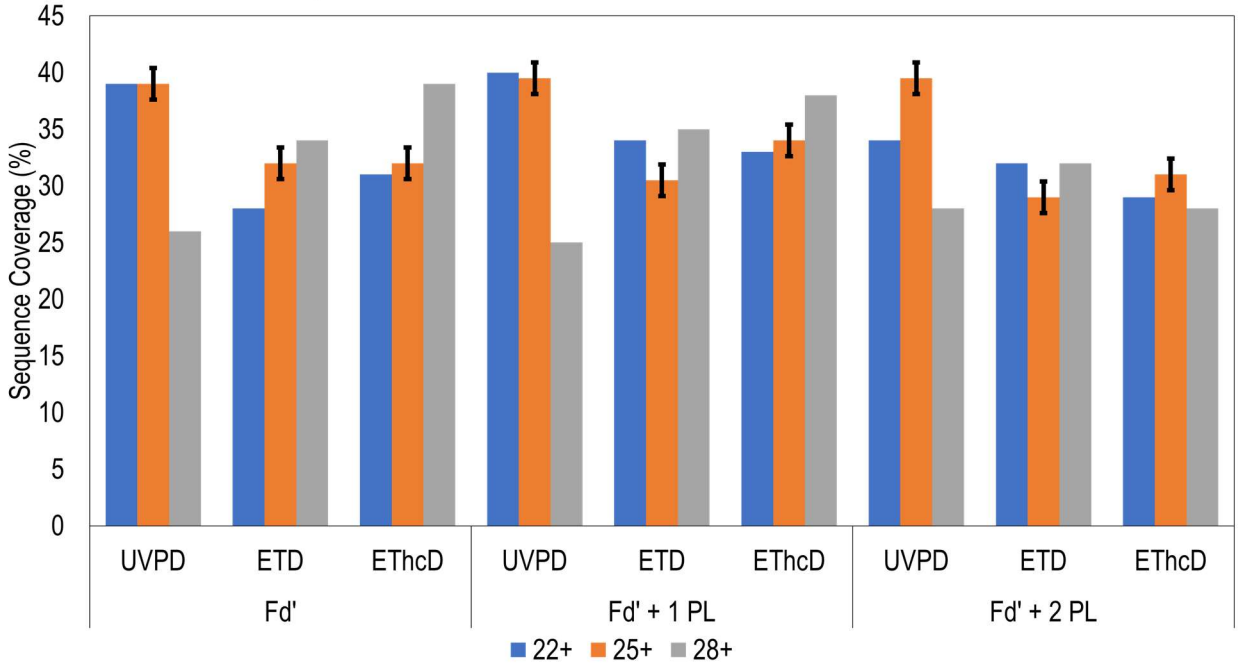


Figure 3.6: Optimization of charge state. The reaction time was set to 30 ms for ETD and 10 ms for EThcD (NCE 10). For UVPD one 0.5 mJ pulse was applied. For this comparison, the isolation window was set to 10 *m/z*. Results are displayed with the payload (PL) located at C224 for Fd' + 1 PL and C230 and C233 for Fd' + 2 PL. The typical level of standard deviation is shown for one set of conditions for each activation mode. Error bars represent a 95% confidence interval based on four replicates.

Previous reports have established that the performance of UVPD is not particularly charge state dependent,[49,52] and the same result is echoed in the present study as illustrated in **Figure 3.6**. For this assessment, the sequence coverage obtained for UVPD, ETD and EThcD was measured for three prominent charge states. For all three Fd' species in **Figure 3.6**, slightly higher sequence coverage was obtained for the 25+ charge state, a result attributed to the greater abundance of this charge state which

translates to somewhat better S/N of the MS/MS spectra. The efficacy of ETD is known to be charge state dependent, favoring precursors in higher charge states (higher charge density) owing to the greater exothermicity of the electron transfer reactions and reduction of stabilizing non-covalent interactions of the precursor.[53,54] However, as reported in **Figure 3.6**, a significant charge state dependence on sequence coverage was not observed for ETD nor EThcD. Instead, the sequence coverage was generally found to be slightly greater for ETD and EThcD of the 25+ (mid-range) charge state for all three targeted Fd' species. This outcome which seems to contradict conventional findings about the charge state dependence of ETD is rationalized by the impact of precursor ion abundance. For MS/MS analysis of ADC subunits, a factor that is more important than charge state in modulating sequence coverage is the abundance of the precursor ion. Based on these findings, the charge state of greatest abundance was selected for ETD and EThcD throughout the remainder of the study. Moreover, to obtain maximum coverage, the final method adopted in this study combined the results of three targeted LC runs, each utilizing a different optimized complementary MS/MS strategy (ETD (15 ms), EThcD (10 ms and NCE 10), UVPD (one pulse, 1 mJ)

3.4.4 Seven ADC Subunits

After IdeZ digestion and reduction of the ADC, the various unmodified and payload-conjugated subunits were successfully separated using nano-flow (300 nL/min) reversed phase chromatography. The use of the low flow system allowed a far smaller injection quantity (50 ng) per run compared to the more conventional microbore-scale LC systems operating at flow rates of 250 uL/min and requiring injection quantities of 10 ug or more for complex ADC samples. The low sample consumption afforded by using

nanoLC facilitated repeated injections, allowing extensive optimization of MS/MS parameters and better aligning with future analysis of biological samples. The chromatography was optimized for a shallow gradient to allow baseline resolution of seven products, including both unmodified and payload-conjugated subunits (Lc, Fc/2, Fd') as shown in **Figure 3.7.** The MS1 spectra obtained for each of the chromatographic peaks are displayed in **Figure 3.8**. Digestion and reduction led to the production of the Fc/2 portion of the heavy chain with no payload, the light chain (Lc) with either zero or one conjugated payloads, and the Fd' portion of the heavy chain carrying zero to three payloads. As displayed in the MS1 spectra in **Figure 3.8,** the charge state envelope of each subunit appears in a similar *m/z* range with the 27+, 23+ and 25+ charge states being the most abundant for the Fc/2, Lc and Fd' subunits, respectively. The conjugation of one or two payloads does not cause a significant change in the charge state distributions.

As observed in **Figure 3.7**, the relative abundances of the individual chromatographic peaks associated with various Fd' species (Fd' with zero to three payloads) were as much as 10 times lower than the Fc/2 subunit, even though the summed peak areas were comparable. The lower peak areas were also reflected in the lower sequence coverages obtained for the various Fd' species owing to the lower ion current (and concomitant reduction in S/N of the MS/MS spectra which limited fragment ion assignments). Extensive optimization of MS/MS parameters were needed to maximize sequence coverage and to ensure confident bracketing of payload sites. Combining results from multiple LC runs using different activation methods in each run was also found to be advantageous for enhancing the production of unique fragment ions essential for payload localization.

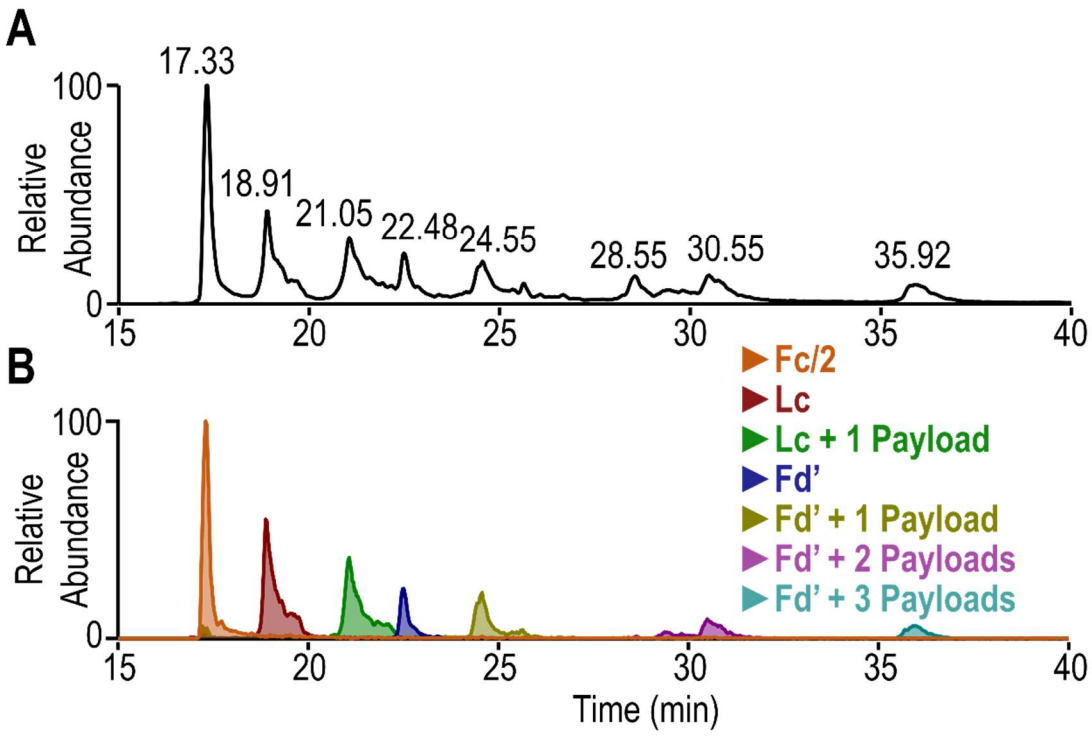


Figure 3.7: (A) Base peak chromatogram of the ADC-IdeZ digest, and (B) the extracted ion chromatograms representing each of the seven key subunit species based on the most abundant charge state observed in the MS1 spectra. The peak at 28.55 min in part A, which does not appear in part B, corresponds to IdeZ protease.

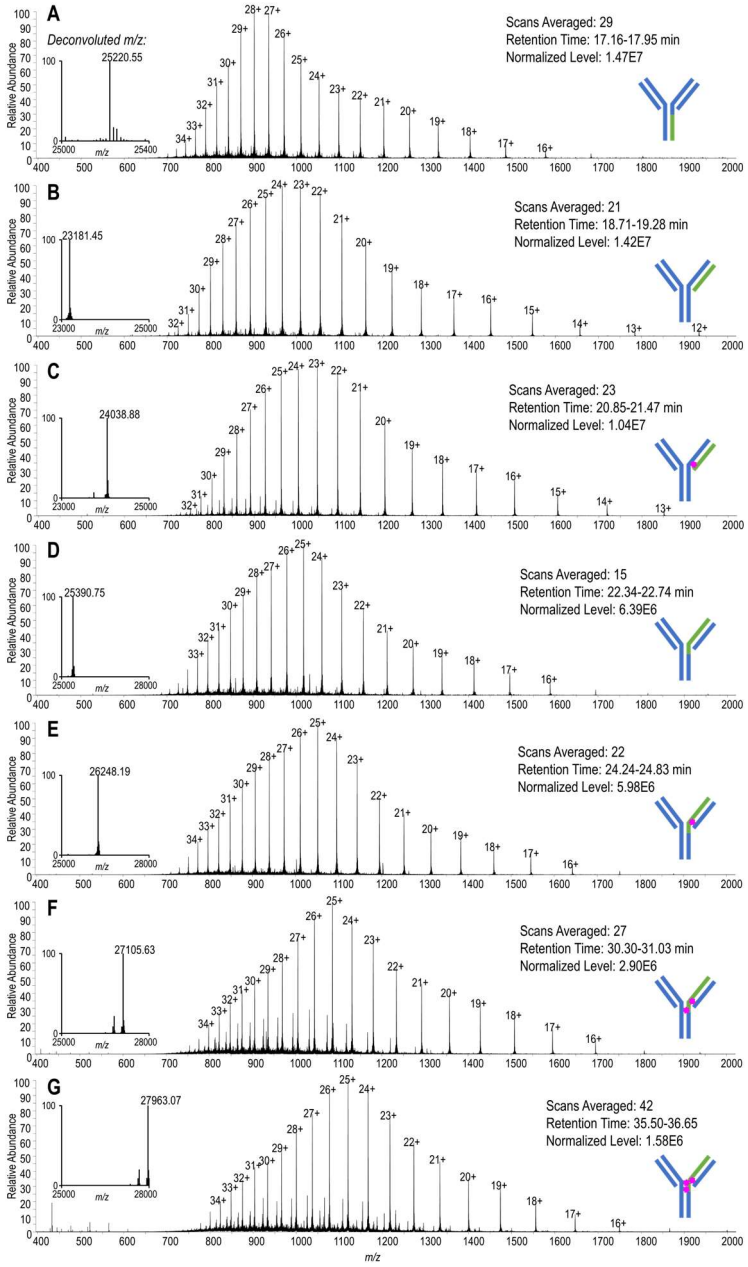


Figure 3.8: MS1 spectra for each of the seven corresponding LC peaks labeled in Figure 3.7. Subunits were identified from the deconvoluted mass, shown to the left of each spectrum. The subunits shown are: (A) Fc/2 (B) Lc, (C) Lc with one payload, (D) Fd', (E) Fd' with one payload, (F) Fd' with two payloads, and (G) Fd' with three payloads. For each MS1 spectrum, the ADC scheme to the right highlights the segment of the antibody in green and display the number of payloads as pink circles.

3.4.5 Characterization of Key Regions

Figure 3.9 displays product ion arrays obtained from the optimized ETD, EThcD and UVPD methods for the Lc subunit with zero or one payloads. The arrays provide a graphical depiction of the sequence coverage based on the identified N-terminal-containing fragment ions (*a,b,c*) and C-terminal-containing fragment ions (*x,y,z*). Although likely prevalent, internal ions (those which contain neither the N-terminus or C-terminus anchors) are difficult to assign with confidence and are thus not considered. While each MS/MS method provides extensive coverage towards the N-terminus (i.e. residues 10-100), all afford fragmentation that is sparser near the C-terminal region, thus reinforcing the beneficial gains of combining activation methods. If the analytical goal was solely focused on characterization of the CDR region, combining methods would not be essential. The payloads, however, are typically conjugated to cysteines initially involved in intermolecular disulfide bonds predominantly situated in the second (C-terminal) half of the Lc and Fd' sequences, amplifying the importance of comprehensive characterization of the C-terminal regions. Comparison of **Figure 3.9A** and **3.9B** reveals a dramatic reduction in C-terminal ions for the payload-conjugated Lc subunit, an outcome that is particularly notable given that in the light chain the payload is conjugated to the cysteine residue (Cys214) at the C-terminus. This means that all fragment ions containing the payload are C-terminal ions, and the presence of the payload clearly affects the fragmentation of the antibody. The reason for the loss or disappearance of C-terminal ions is attributed to fragmentation of the payload itself (as discussed earlier), leading to production of unassigned truncated payload-containing Lc fragments.

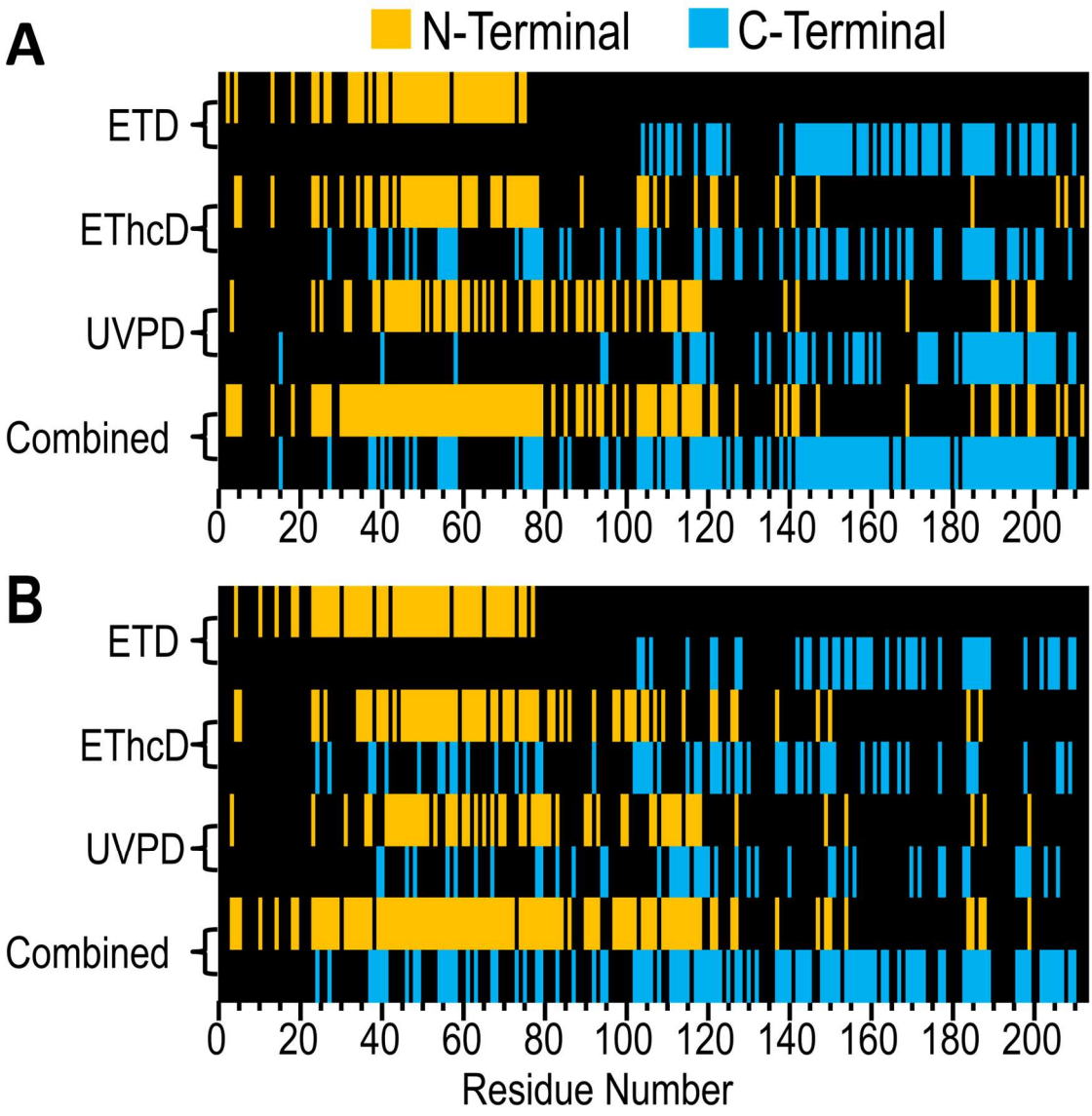


Figure 3.9: Product ion arrays showing sites of backbone cleavages for (A) Lc and (B) Lc with one payload at C214 based on the optimized MS/MS conditions including ETD (15 ms), EThcD (10 ms with 10 NCE), and UVPD (1 pulse at 1 mJ). The 23+ charge state was activated.

Despite the dramatic loss of C-terminal ions for the payload-conjugated Lc subunit for any single MS/MS method, combining the information from four MS/MS methods largely fills in the gaps in coverage. The last entries of **Figure 3.9A** and **3.9B**

show the combined product ion arrays, highlighting the net gain in coverage of the C-terminus regions. The N-terminal product ion arrays are similar for the Lc subunit in **Figure 3.9A** and **3.9B**, confirming the absence of payload in this region. **Figures 3.10** and **3.11** display additional product ion array maps for the Fd' (with and without payloads). As observed when comparing the Fd' subunit with zero or one conjugations in **Figure 3.10** to the Fd' subunit with two or three conjugations in **Figure 3.11**, incorporation of more than one payload leads to even greater depletion of C-terminal fragment ions. However, when the coverages from all methods are combined, as illustrated in the final rows of each set of product ion arrays, the collective sequence coverages are substantially improved. In each case, there are ample sequence ions near the C-terminus region that retain the payload, thus bracketing the payload location. By combining the three Lc runs, sufficient sequence detail is obtained to identify the area of the payload locations**.**

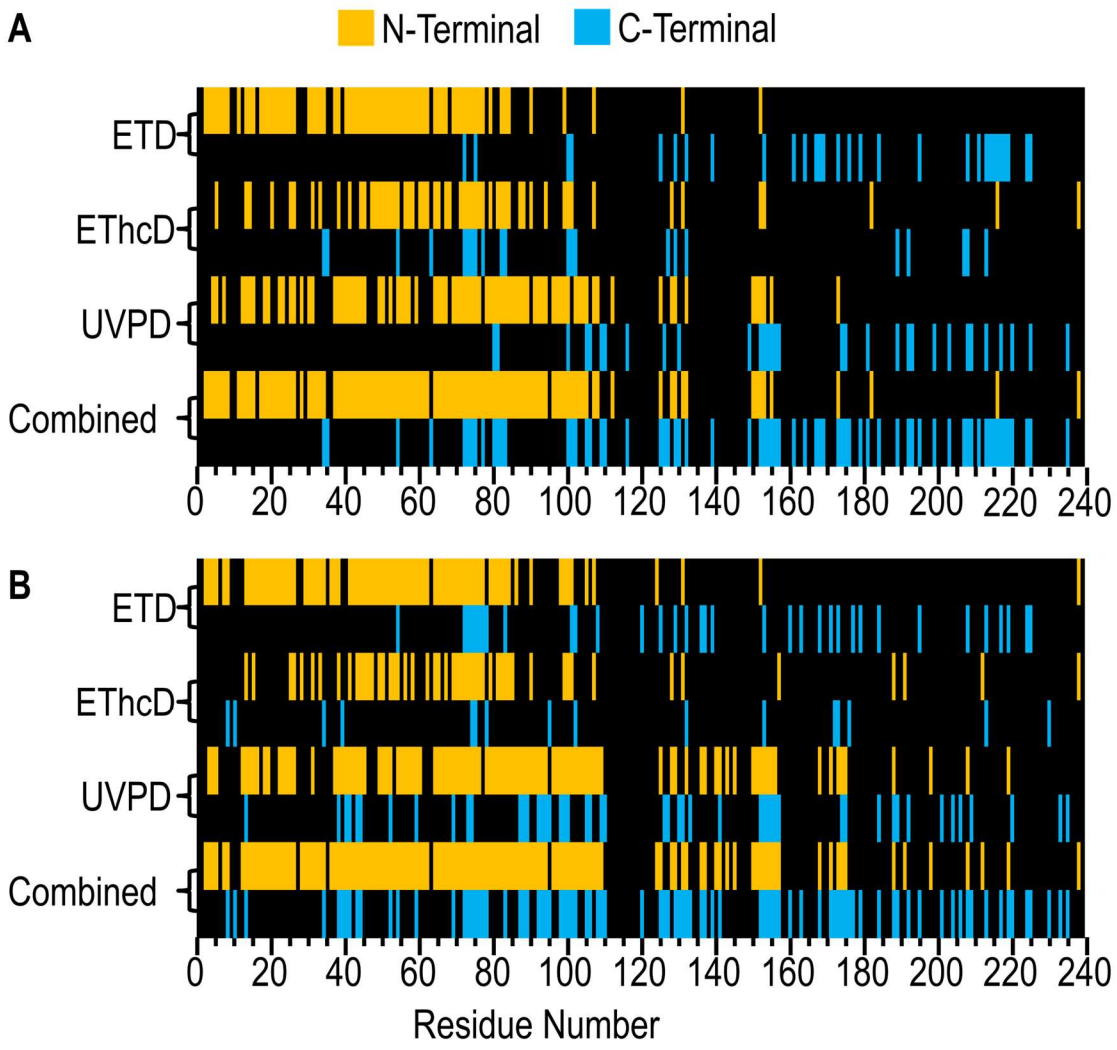


Figure 3.10: Product ion arrays showing sites of backbone cleavages for (A) Fd' and (B) Fd' with one payload on C224 based on the optimized activation conditions including ETD (15 ms), EThcD (10 ms with 10 NCE), and UVPD (1 pulse at 1 mJ). The 25+ charge state was activated.

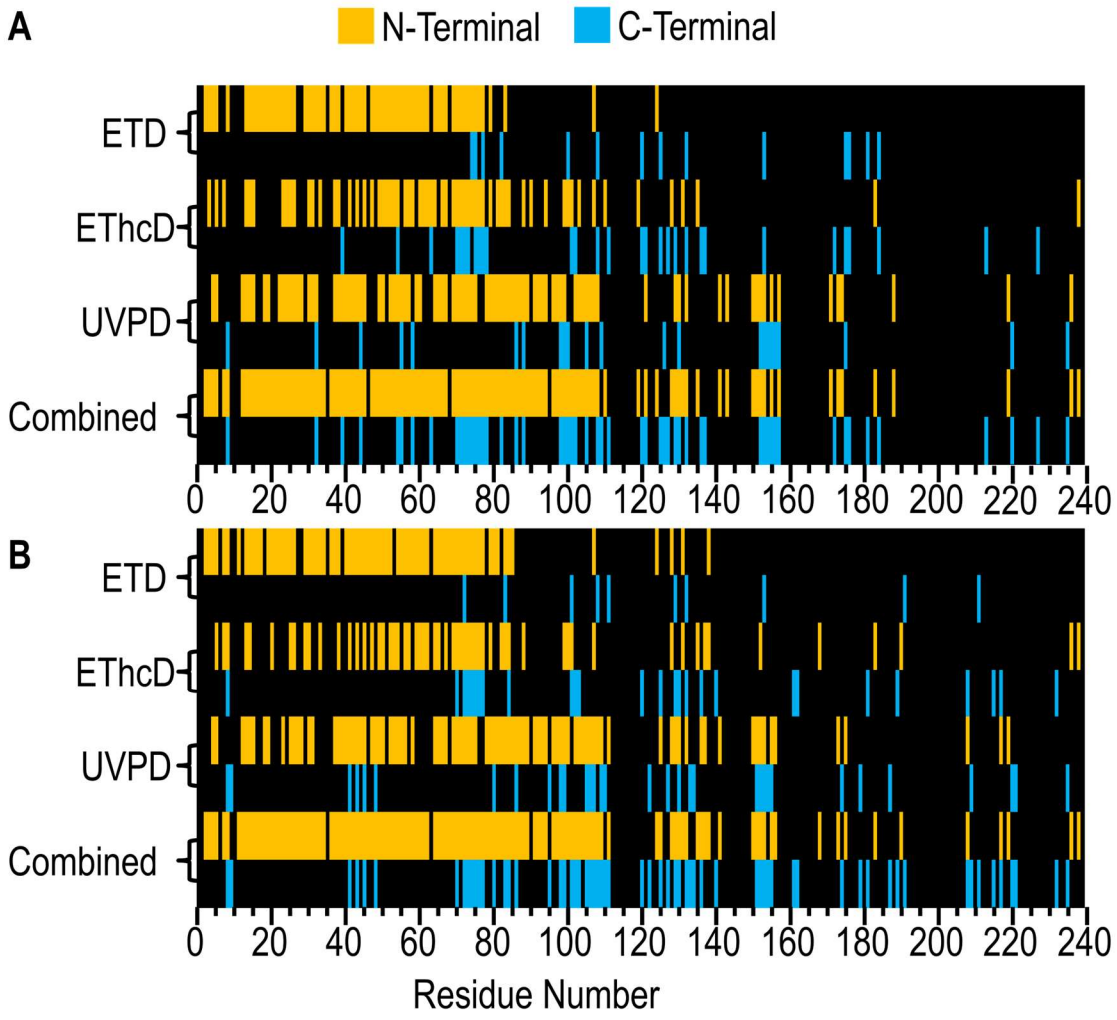


Figure 3.11: Product ion arrays showing sites of backbone cleavages for (A) Fd' with two payloads on C230 and C233 and (B) Fd' with three payload on C224, C230 and C233 based on the optimized activation conditions including ETD (15 ms), EThcD (10 ms with 10 NCE), and UVPD (1 pulses at 1 mJ). The 25+ charge state was activated.

3.4.6 Payload Conjugation Sites on the Light Chain (Lc)

Combining the sequence maps from the four optimized fragmentation methods makes payload localization possible. **Figure 3.12** illustrates the payload characterization results for the light chain Lc conjugated to a single payload. The Venn diagrams in

Figure 3.12A display the fragment ions unique to each of the three activation methods (ETD, EThcD, UVPD) categorized as those which contain a payload and those that do not contain a payload. In total across all activation methods, 99 fragment ions that do not contain a payload and 105 fragment ions that retain a payload were identified. The distribution of unique fragment ions identified by each activation method is similar between the two Venn diagrams. While EThcD identifies the greatest number of fragment ions overall, UVPD contributes the most unique fragment ions, making it a valuable complementary method. This outcome will become even more important for the more complex Fd' subunits, as described in the next section. The sequence coverage map displayed in **Figure 3.12B** further illustrates the power of combining activation methods. The nearly complete fragmentation map of the light chain, including numerous ions that originate from backbone cleavages between C194 and C214, offers both unambiguous localization of the payload and extensive characterization of the CDRs.

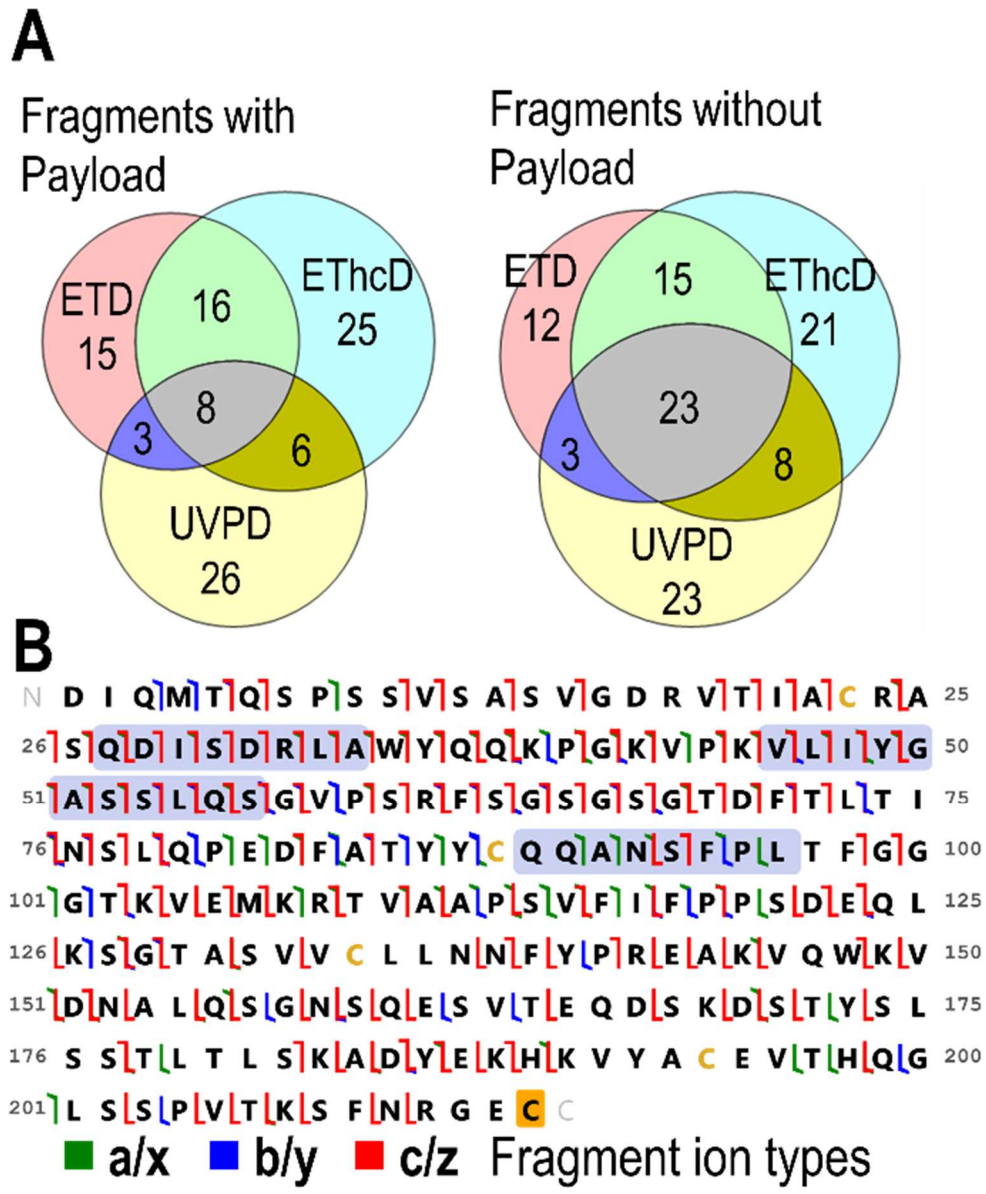


Figure 3.12: (A) Venn diagrams displaying the number of fragment ions originating from unique cleavage locations for MS/MS of the Lc subunit bound to one payload for UVPD (1 pulse, 1 mJ), ETD (15 ms) and EThcD (10 ms, 10 NCE) of the 23+ charge state, categorized into the fragment ions containing the payload (left) and those that do not (right). (B) Sequence coverage map for the combined sequence overage from UVPD, ETD and EThcD. The payload is positioned on C214, highlighted in gold.

3.4.7 Identification of Heavy Chain (Fd') Conjugation Sites

There are three inter-chain disulfide bonds where conjugation of the payload is probable on the heavy chain (C224, C230, C233). This collection of conjugation sites affords the possibility of heterogeneity in the resulting Fd' products for those that contain one or two payloads. For example, the pairs of modification sites of Fd' containing two payloads are C224/C230, C224/C233, and C230/C233. Chromatographic separation did not completely resolve the conjugation states of the Fd' subunit containing one or two payloads (**Figure 3.7**). For the Fd' containing either one or two payloads, in each case there appeared to be two chromatographic peaks with one being a shoulder peak of low abundance. Previously, such shoulder peaks have been identified as different conjugation states for similar cysteine-bound ADCs.[41] In the present study, the shoulder peaks also corresponded to payload-modified Fd' ions. However, these low abundance Fd' ions generated subpar MS/MS spectra, preventing confirmation of specific conjugation sites.

The detailed results for characterization of the Fd' subunit with 1-3 payloads obtained by each activation method are displayed in **Figures 3.13-3.19**. The analysis of the Fd' subunit with 1 or 2 payloads focuses on the species corresponding to the most abundant peak in the chromatographic trace (**Figure 3.7**), not the smaller shoulder peaks because no unique information was obtained from the latter. The payload was placed at C224 for the Fd' subunit with a single payload location for all analysis because diagnostic fragment ions were identified using ETD to support that location, and no fragment ions were identified to support an alternative interchain cysteine conjugation site. **Figure 3.13** displays the results for UVPD of the Fd' subunit containing one payload (25+). While no fragment ions were identified to help determine which interchain disulfide bond was responsible for payload conjugation, a number of fragment ions were identified that helped confirm payload location on the interchain disulfide bonds (rather

than the intrachain disulfide bonds). Individual mass regions around key fragment ions of interest are expanded in order to highlight the quality of identification. In each case the predicted isotopic profiles of the fragment ions match the predicted assignments. Two of the ions showcased in **Figure 3.13B,C** are y_{31}^{5+} and y_{20}^{3+}, both of which are short payload-containing fragments that unambiguously localize the payload to the last three cysteines. Similarly, several large payload-free fragment ions, including b_{189}^{20+} (**Figure 3.13D)**, help to eliminate cysteines C22-C204 from consideration as payload-conjugation sites.

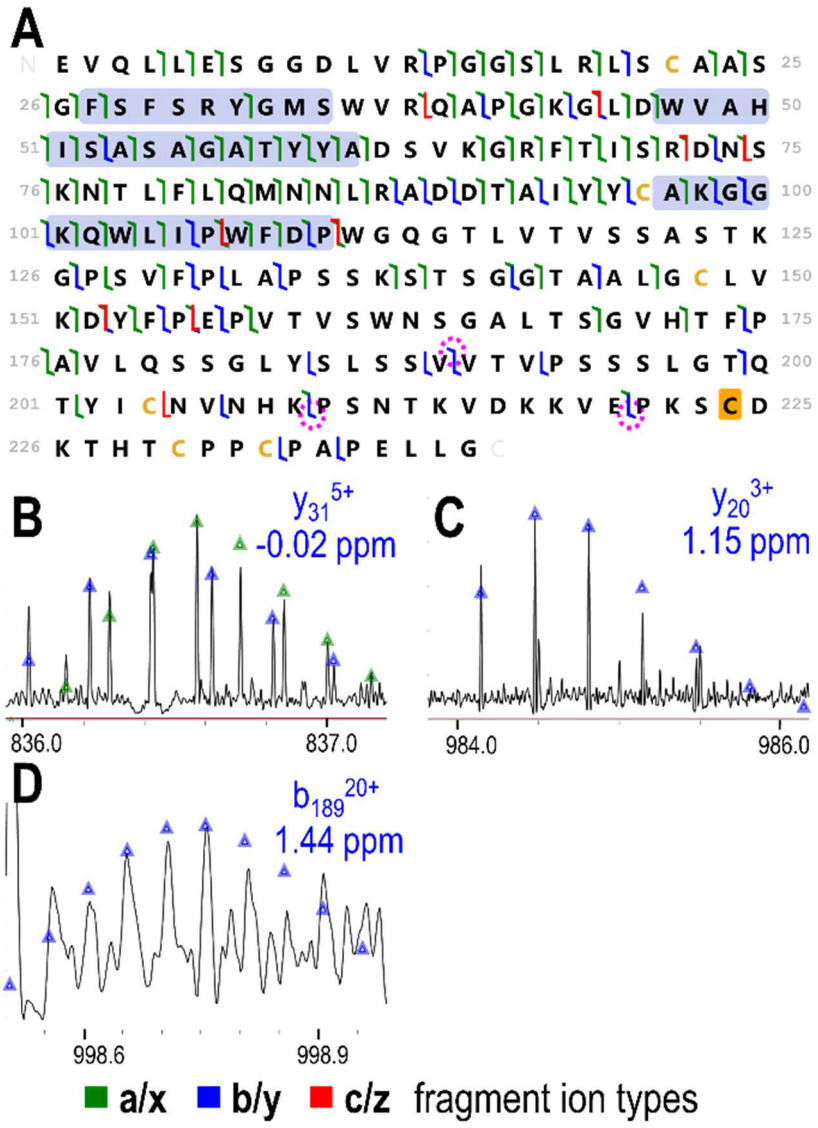


Figure 3.13: Fragmentation data for UVPD (1 pulse 1 mJ) of the 25+ charge state of subunit Fd' with one payload. (A) Sequence coverage map with the payload positioned at C224. (B) Expansion of payload-containing y_{31}^{5+} (with one hydrogen loss). The a_{57}^{7+} fragment ion is also included in the expansion, marked in green. (C) Expansion of payload-containing y_{20}^{3+} (with one hydrogen loss). (D) Expansion of payload-free b_{189}^{20+}. The backbone cleavage sites that lead to the fragment ions that were selected for expansion are circled in purple on the sequence coverage map. CDRs are shaded in blue on the sequence coverage maps, and the payload conjugation site is highlighted in gold.

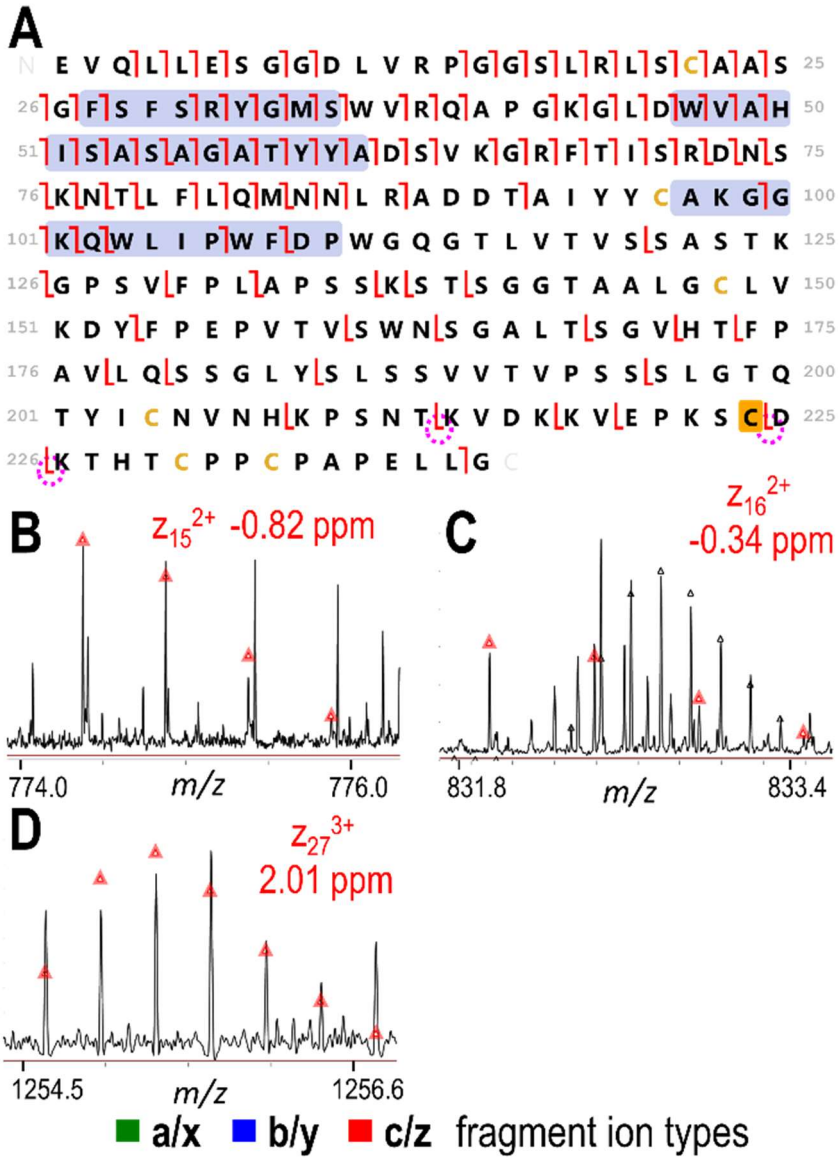


Figure 3.14: Fragmentation data for ETD (15 ms) of the 25+ charge state of subunit Fd' with one payload. (A) Sequence coverage map with the payload positioned at C224. (B) Expansion of payload-free z_{15}^{2+} which localizes the payload to C224. (C) Expansion of payload-free z_{16}^{2+} which localizes the payload to C224. (D) Expansion of payload-containing z_{27}^{3+}. The backbone cleavage sites that lead to the fragment ions that were selected for expansion are circled in purple on the sequence coverage map. CDRs are shaded in blue on the sequence coverage maps, and the payload conjugation site is highlighted in gold.

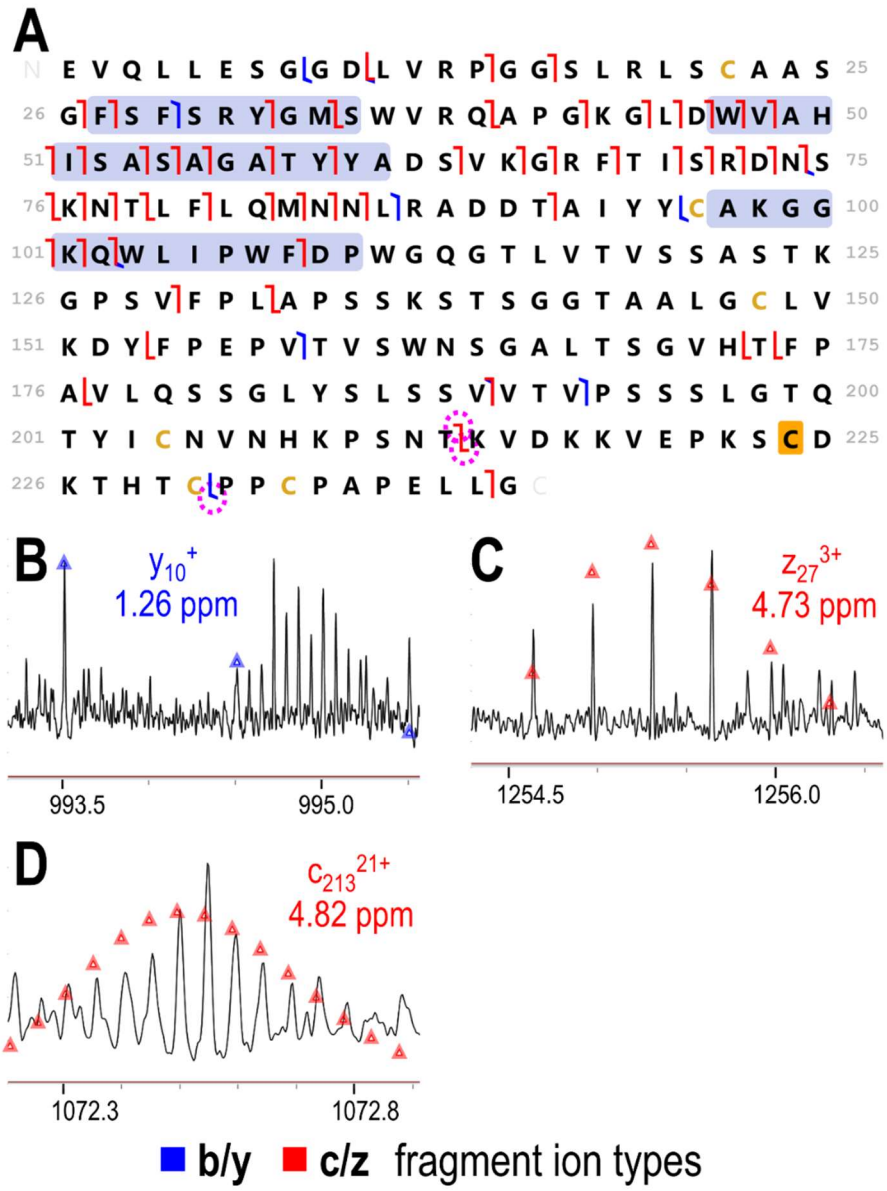


Figure 3.15: MS/MS data for EThcD (10 ms 10 NCE) of the 25+ charge state of Fd' with one payload. (A) Sequence coverage map with the payload positioned at C224. (B) Expansion of payload-free y_{10}^{+} which localizes the payload to C224 or C230. (C) Expansion of bis-payload-containing ion z_{27}^{3+}. (D) Expansion of payload-free ion c_{213}^{21+}. The backbone cleavage sites that lead to the fragment ions that were selected for expansion are circled in purple on the sequence coverage map. CDRs are shaded in blue on the sequence coverage maps, and the payload conjugation site is highlighted in gold.

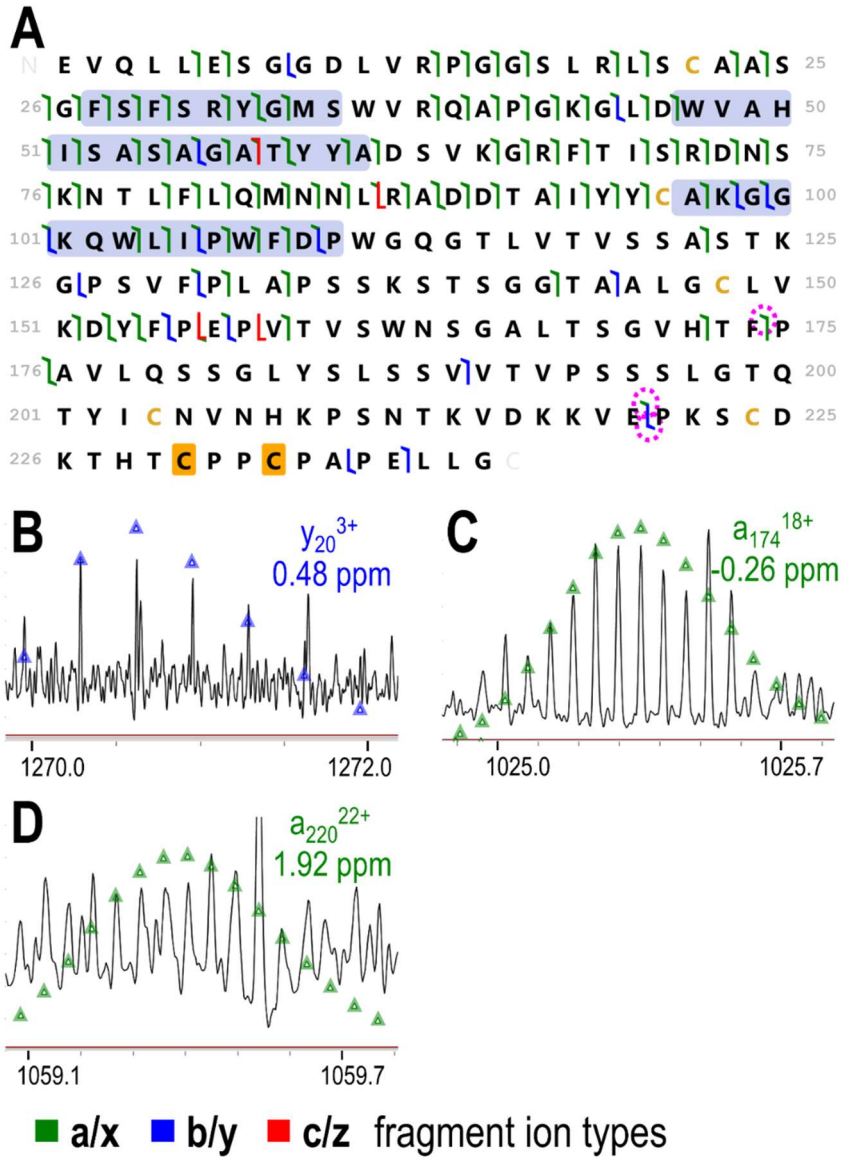


Figure 3.16: MS/MS data for UVPD (1 pulse 1 mJ) of the 25+ charge state of Fd' with two payloads. (A) Sequence coverage map with the payloads positioned at C230 and C233. (B) Expansion of bis-payload-containing y_{20}^{3+} (with one hydrogen loss). (C) Expansion of payload-free a_{174}^{18+} (with one additional hydrogen). (D) Expansion of payload-free a_{220}^{22+} (with one additional hydrogen). The backbone cleavage sites that lead to the fragment ions that were selected for expansion are circled in purple on the sequence coverage map. CDRs are shaded in blue on the sequence coverage maps, and the payload conjugation site is highlighted in gold.

Figure 3.17: Sequence coverage map for ETD (15 ms) of the 25+ charge state of Fd' with two payloads positioned at C230 and C233. CDRs are shaded in blue on the sequence coverage maps, and the payload conjugation site is highlighted in gold.

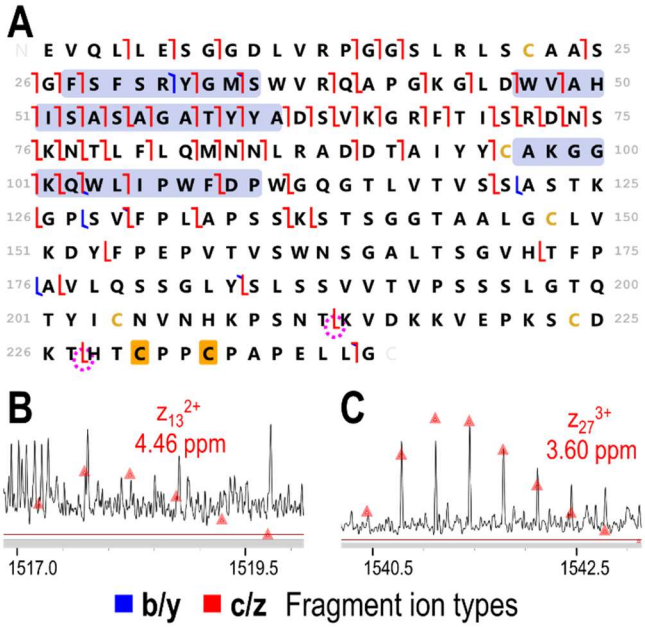


Figure 3.18: MS/MS data for EThcD (10 ms 10 NCE) of the 25+ charge state of Fd' with two payloads. (A) Sequence coverage map with the payloads positioned at C230 and C233. (B) Expansion of bis-payload-containing z_{13}^{2+} which localizes the payloads to C230 and C233. (C) Expansion of bis-payload-containing z_{27}^{3+}. The backbone cleavage sites that lead to the fragment ions that were selected for expansion are circled in purple on the sequence coverage map. CDRs are shaded in blue on the sequence coverage maps, and the payload conjugation site is highlighted in gold.

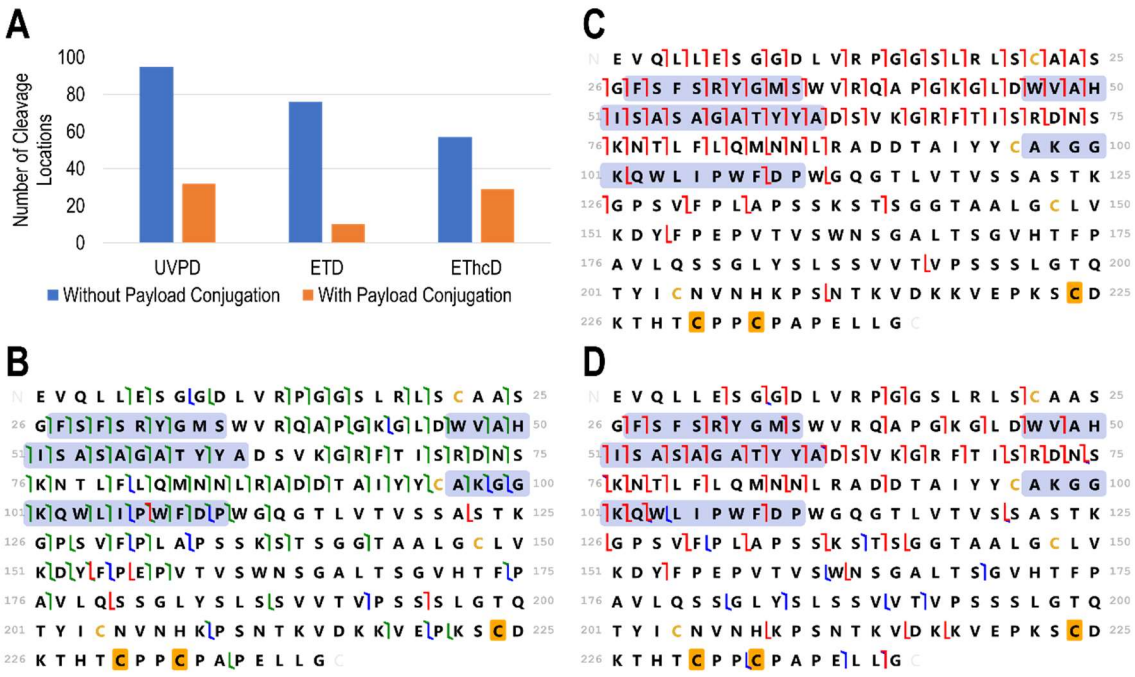


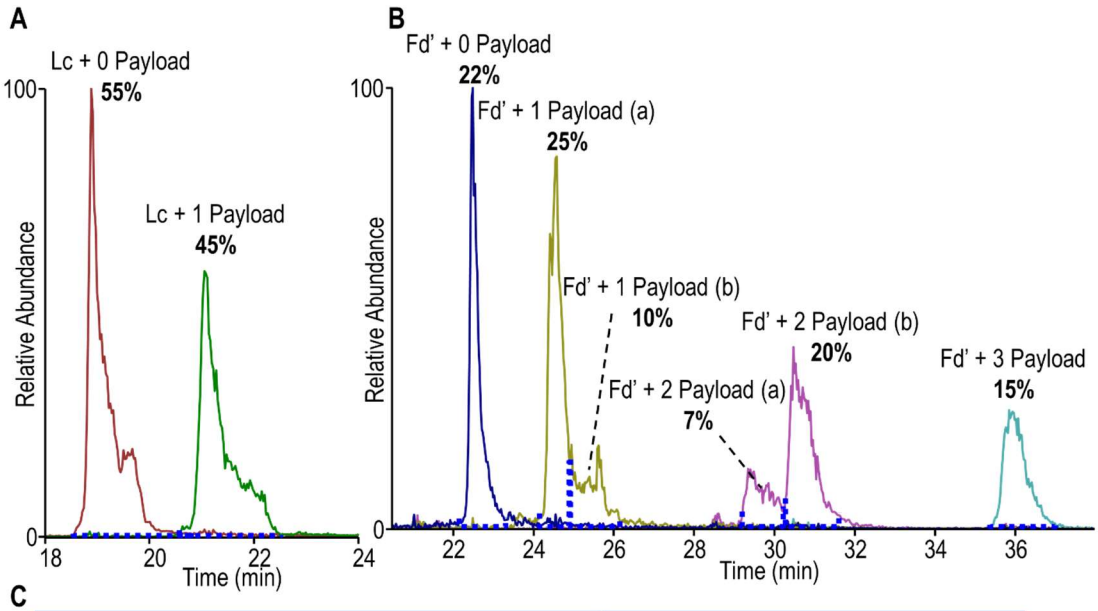
Figure 3.19: (A) Histogram displaying the number of backbone cleavage locations for Fd' with three payloads from UVPD (1 pulse, 1 mJ), ETD (15ms), and EThcD (10 ms, 10 NCE) of the 25+ charge state, segregated into fragment ions that contain the payload (orange bars) and those that do not (blue bars). Sequence coverage maps are displayed for (B) UVPD, (C) ETD and (D) EThcD. CDRs are shaded in blue on the sequence coverage maps, and the payload conjugation sites (C224,C230, C233) are highlighted in gold.

Fragment ion assignments identified from ETD of the Fd' subunit containing one payload are summarized in **Figure 3.14**. Of note are two short payload-free C-terminal ions, z_{15}^{2+} and z_{16}^{2+}, which indicate that the payload is located at C224. Similarly, several payload-containing C-terminal ions, such as z_{27}^{3+} highlighted in **Figure 3.14D**, bracket the payload and isolate it to C224. The results here are consistent with a previous ETD-focused study which examined a similar cysteine-bound antibody and also localized the payload of one eluting species to the interchain disulfide bond associated with the bridge between the heavy and light chains.[45] The results for EThcD of this Fd' subunit

containing one payload are displayed in **Figure 3.15**. A payload-free fragment ion, y_{10}^{+}, is highlighted in **Figure 3.15B** because it originates from a backbone cleavage that occurs between C230 and C231, indicating that the payload is not situated among the final ten residues of the Fd' subunit. This reinforces the assignment of C224 as the payload conjugation site by eliminating C233 as a possibility.

In a similar manner, **Figures 3.16, 3.17** and **3.18** display the results for UVPD, ETD and EThcD, respectively, obtained for the species corresponding to the most abundant chromatographic peak for the Fd' subunit with two payloads. UVPD and EThcD produce several fragment ions that help to confirm that both payloads are located on the hinge cysteines, as also found for the Fd' with one payload. UVPD and ETD do not produce ions that localize the payload to a specific hinge cysteine. EThcD, however, generated a z_{13}^{2+} fragment ion that helped to map the payload to the two interchain cysteines that connect the heavy chains, C230 and C233. The mass profile of the z_{13}^{2+} ion is shown in **Figure 3.18B**. While the low signal-to-noise of this fragment ion somewhat obscures the isotope pattern in this congested region of the mass spectrum, the high resolution and mass accuracy of the Orbitrap mass spectrometer allows preliminary identification of this low abundance fragment ion. The low abundance fragment ion can reasonably be considered a possible match because its isotope pattern aligns with the isotope pattern predicted for an ion of the same molecular composition, as illustrated in **Figure 3.18B.** The conjugation sites identified for the more prominent peak of Fd' with one payload are consistent with the results from a previous study of brentuximab vedotin.[41] Even so, this fragment ion could still arguably be removed in the process of manual validation, so it by itself is insufficient to identify the location of the payload. The combination of UVPD, ETD, and EThcD do not allow confident characterization of the specific locations of the two payloads on the Fd' subunit.

Figure 3.19 shows the results obtained for MS/MS analysis of the Fd' subunit containing three payloads. This figure conveys a dramatic difference between the number of payload-containing and payload-lacking fragment ions produced. The histogram in **Figure 3.19A** shows that there are less than half as many payload-containing fragment ions. This difference is more pronounced for the Fd' with three payloads than it is for any other subunit. The rationale for this outcome is that the effects of linker fragmentation are compounded given the enhanced number of payloads on the more complex Fd' subunit. Even given this limitation, by combining the three fragmentation methods sufficient information is obtained to isolate the three payloads to the three hinge cysteines.



C

Peaks added together	Percent of C224 Containing a Payload
Fd' + 1 Payload (a), Fd' + 2 Payload (b), Fd' + 3 Payload	60%
Fd' + 1 Payload (a), Fd' + 2 Payload (a), Fd' + 3 Payload	47%
Fd' + 1 Payload (b), Fd' + 2 Payload (b), Fd' + 3 Payload	45%
Fd' + 1 Payload (b), Fd' + 2 Payload (a), Fd' + 3 Payload	32%

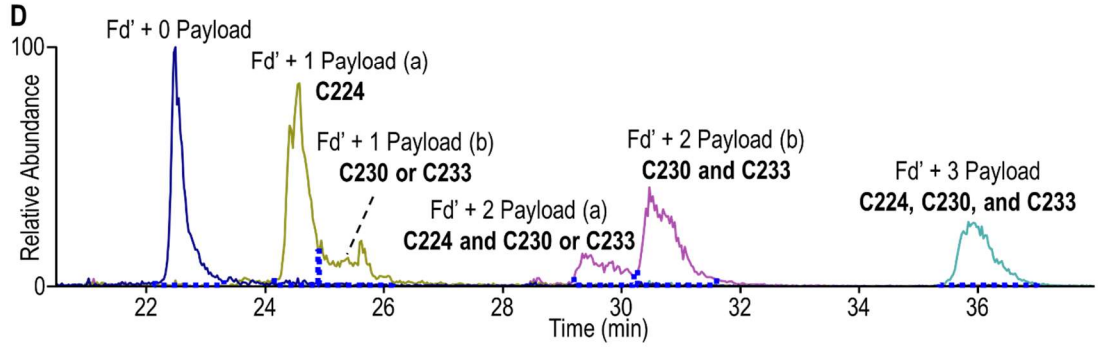


Figure 3.20: Percent contributions for different peaks of (A) Lc and (B) Fd' based on areas of the extracted ion chromatograms. For (A) Lc (26+, 25+, 24+, 23+, and 22+ charge states are all included) and for (B) Fd' (28+, 27+, 26+, 25+, 24+, and 23+ charge states are all included). The Fd' + 1 payload and Fd' + 2 payload species are each split into two peaks (labelled (a) and (b)). The dotted blue lines indicate the boundaries of the areas utilized for each contribution. (C) Summary of the possible ways to add together the peaks in part B to estimate the percentage of C224 containing a payload. (D) The final extracted ion chromatogram displays the expected conjugation sites for each peak associated with the Fd' subunit.

As a supplement to the tandem mass spectrometry data, some details can be gleaned from the MS1 spectra to help validate the assignment of payload locations for Fd' with two payloads. **Figure 3.20** displays the percent distribution of the different chromatographic peaks of Lc and Fd' based on the comparative areas of the peaks in the extracted ion chromatograms. The percentages are only relative but can help decipher the distribution of conjugation sites. **Figure 3.20A** indicates that 45% of the Lc subunit has one conjugated payload. Based on the even number of conjugated payloads consistently observed for cysteine-conjugated ADCs, this information can be extrapolated to indicate that 45% of the heavy chain cysteines involved in heavy chain to light chain inter-chain bonds will contain a payload conjugation. In other words, 45% of C224 on Fd' will be conjugated. The tandem mass spectrometry data definitively reveals that the larger peak for Fd' with one payload indicates conjugation at C224, as does Fd' with three payloads. If the percentage of Fd + one payload (25%) and the percentage Fd + three payloads (15%) in **Figure 3.20A** are summed and subtracted from the 45% of C224 conjugation of Fd', the remainder is 5% for Fd' with 2 payloads.

Owing to the very close positions of C230 and C233, this difference will not result in dramatically different retention times, and therefore the two peaks observed for Fd' with one and two payloads correspond to conjugation or lack of conjugation of C224. The smaller peak, representing 7% of Fd', is much closer to the 5% remaining than the larger peak. This also supports the, albeit less confident, results from tandem mass spectrometry discussed earlier. The final conjugation state assignments for each LC peak based on all of this information are displayed in **Figure 3.20D**.

The most prominent conjugation site for Fd' with one payload is not the most prominent for Fd' with two payloads. A possible explanation for this discrepancy could

be that while the light chain to heavy chain inter-chain bond is most readily conjugated on its own, once one of the heavy chain to heavy chain interchain bonds is conjugated a structural change occurs that facilitates conjugation of the second. The difference also helps to explain why the larger peak is the first to elute for Fd' with one payload, and the second to elute for Fd' with two payloads. In both cases the conjugation to C224 resulted in an earlier elution time.

A summary of the results for the payload-containing Fd' subunits is displayed in **Figure 3.21**, including integrated sequence coverage maps for the species corresponding to each of the three mass-shifted Fd' chromatographic peaks in **Figure 3.7** with the payloads conjugated to representative inter-chain cysteine residues. The Venn diagrams associated with each sequence coverage map in **Figure 3.21** display the sequence coverage and number of backbone cleavages yielding payload-containing fragment ions afforded by each of the MS/MS methods. As a benchmark comparison, the overall sequence coverage obtained for the Fd' subunit with no payload was 65%. This value was comparable to the overall sequence coverages obtained for the Fd' subunit with one, two or three payloads, which were 68%, 60% and 65%, respectively. While the total sequence coverage reveals the overall characterization of the protein, the payload-containing fragment ions are the most critical ones for payload localization. In order to summarize the contributions of each activation method, the cleavage locations resulting in payload-containing fragment ions are displayed in the Venn diagrams. UVPD provided the greatest number of unique fragment ions, but ETD and EThcD offered numerous complementary fragments, thus making it possible to localize the payloads even for heterogeneous compositions of the subunits (e.g., ones where the specific payload locations vary). The Venn diagrams in **Figure 3.21** highlight the benefit of combining the results from three activation methods. As a culmination of the research presented here,

the payloads on the most abundant chromatographic peaks corresponding to Fd' with one payload and Fd' with two payloads were localized to cysteines consistent with previous studies of cysteine-conjugated ADCs.[41,45]

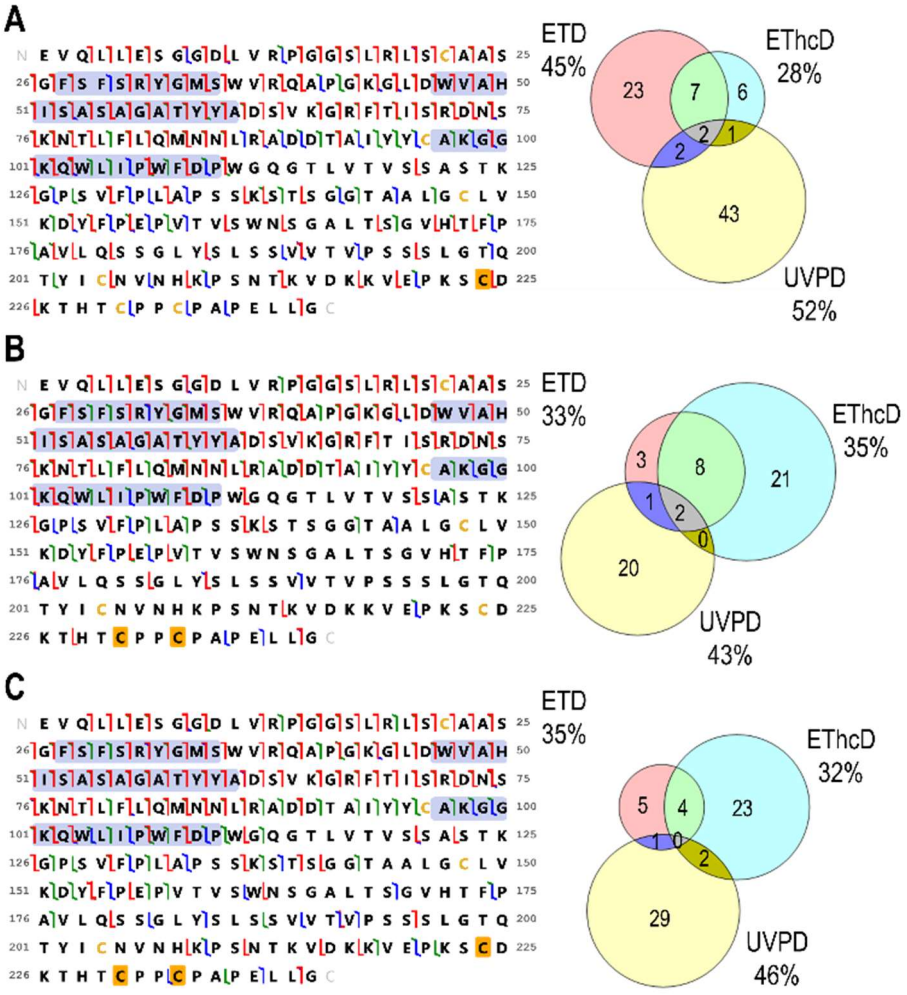


Figure 3.21: Sequence coverage maps for the heavy chain fragment Fd': (A) Fd' + 1 payload, (B) Fd' + 2 payloads and (C) Fd' + 3 payloads. Corresponding Venn diagrams display the number of unique cleavage locations leading to payload-containing fragment ions and the total sequence coverage for each fragmentation method. Data represents a combination of three LC runs using each of the optimized activation methods previously discussed. One representative conjugation state is displayed for each map. DRs are shaded in blue on the sequence coverage maps, and the payload conjugation sites are highlighted in gold. The 25+ charge state was activated for each subunit.

3.5 CONCLUSIONS

A middle-down method that integrates three activation methods (UVPD, ETD, and EThcD) provides extensive characterization of antibody-drug conjugates. Optimization of the MS/MS parameters was essential for maximizing sequence coverage. Given the location of the payloads, C-terminal ions containing the payload proved to be key in confirming the payload location, but the abundances of these ions were depleted as payloads were added to the subunits of interest, presumably owing to truncation of the payload moiety during MS/MS that resulted in unassignable product ions. However, by combining multiple complementary activation methods, sufficiently diverse arrays of C-terminal ions were generated to allow payload sites to be pinpointed in addition to confirmation of the important CDR regions. High sequence coverages were observed for each of the subunits: 80% and 75% for Lc with zero or one payloads, respectively, and 65%, 68%, 60%, and 65% for Fd' with zero, one, two or three payloads, respectively. Consideration of relative abundance contributions of Lc and Fd' related peaks helped to strengthen payload localization claims. This integrated MS/MS strategy should be readily adapted for characterization of therapeutic antibodies; other antibody-drug conjugates, including ones with different conjugation sites or more elaborate payloads; and even for biological samples in drug metabolism studies. Coupling the methods presented here with innovative approaches to increase the number of diagnostic ions confidently assigned in the complicated MS/MS spectra of ADCs, such as using advanced spectral averaging and multiplexing[36] or proton transfer reactions,[51] is expected to further improve the comprehensive characterization of ADCs. Additionally, pairing this middle-down approach with bottom-up or top-down methods, as recently presented for antibody analysis,[30] offers the potential for an even more powerful analytical strategy.

3.6 REFERENCES

(1) Chari, R. V. J.; Miller, M. L.; Widdison, W. C. Antibody–Drug Conjugates: An Emerging Concept in Cancer Therapy. *Angew. Chem. Int. Ed.* **2014**, *53*, 3796–3827.

(2) Peters, C.; Brown, S. Antibody-Drug Conjugates as Novel Anti-Cancer Chemotherapeutics. *Biosci. Rep.* **2015**, *35*, e00225.

(3) Bouchard, H.; Viskov, C.; Garcia-Echeverria, C. Antibody–Drug Conjugates—A New Wave of Cancer Drugs. *Bioorg. Med. Chem. Lett* **2014**, *24*, 5357–5363.

(4) Abdollahpour-Alitappeh, M.; Lotfinia, M.; Gharibi, T.; Mardaneh, J.; Farhadihosseinabadi, B.; Larki, P.; Faghfourian, B.; Sepehr, K. S.; Abbaszadeh-Goudarzi, K.; Abbaszadeh-Goudarzi, G.; Johari, B.; Zali, M. R.; Bagheri, N. Antibody–Drug Conjugates (ADCs) for Cancer Therapy: Strategies, Challenges, and Successes. *J. Cell. Physiol* **2019**, *234*, 5628–5642.

(5) Zolot, R. S.; Basu, S.; Million, R. P. Antibody–Drug Conjugates. *Nat. Rev. Drug Discov.* **2013**, *12*, 259–260.

(6) Chalouni, C.; Doll, S. Fate of Antibody-Drug Conjugates in Cancer Cells. *J. Exp. Clin. Cancer Res* **2018**, *37*, 20.

(7) Joubert, N.; Denevault-Sabourin, C.; Bryden, F.; Viaud-Massuard, M.-C. Towards Antibody-Drug Conjugates and Prodrug Strategies with Extracellular Stimuli-Responsive Drug Delivery in the Tumor Microenvironment for Cancer Therapy. *Eur. J. Med. Chem* **2017**, *142*, 393–415.

(8) Beck, A.; Goetsch, L.; Dumontet, C.; Corvaïa, N. Strategies and Challenges for the next Generation of Antibody–Drug Conjugates. *Nat. Rev. Drug Discov.* **2017**, *16*, 315–337.

(9) Behrens, C. R.; Ha, E. H.; Chinn, L. L.; Bowers, S.; Probst, G.; Fitch-Bruhns, M.; Monteon, J.; Valdiosera, A.; Bermudez, A.; Liao-Chan, S.; Wong, T.; Melnick, J.; Theunissen, J.-W.; Flory, M. R.; Houser, D.; Venstrom, K.; Levashova, Z.; Sauer, P.; Migone, T.-S.; van der Horst, E. H.; Halcomb, R. L.; Jackson, D. Y. Antibody–Drug Conjugates (ADCs) Derived from Interchain Cysteine Cross-Linking Demonstrate Improved Homogeneity and Other Pharmacological Properties over Conventional Heterogeneous ADCs. *Mol. Pharm* **2015**, *12*, 3986–3998.

(10) Valliere-Douglass, J. F.; Hengel, S. M.; Pan, L. Y. Approaches to Interchain Cysteine-Linked ADC Characterization by Mass Spectrometry. *Mol. Pharm* **2015**, *12*, 1774–1783.

(11) Tumey, L. N.; Leverett, C. A.; Vetelino, B.; Li, F.; Rago, B.; Han, X.; Loganzo, F.; Musto, S.; Bai, G.; Sukuru, S. C. K.; Graziani, E. I.; Puthenveetil, S.; Casavant, J.; Ratnayake, A.; Marquette, K.; Hudson, S.; Doppalapudi, V. R.;

Stock, J.; Tchistiakova, L.; Bessire, A. J.; Clark, T.; Lucas, J.; Hosselet, C.; O'Donnell, C. J.; Subramanyam, C. Optimization of Tubulysin Antibody–Drug Conjugates: A Case Study in Addressing ADC Metabolism. *ACS Med. Chem. Lett.* **2016**, *7*, 977–982.

(12) Dorywalska, M.; Strop, P.; Melton-Witt, J. A.; Hasa-Moreno, A.; Farias, S. E.; Casas, M. G.; Delaria, K.; Lui, V.; Poulsen, K.; Sutton, J.; Bolton, G.; Zhou, D.; Moine, L.; Dushin, R.; Tran, T.-T.; Liu, S.-H.; Rickert, M.; Foletti, D.; Shelton, D. L.; Pons, J.; Rajpal, A. Site-Dependent Degradation of a Non-Cleavable Auristatin-Based Linker-Payload in Rodent Plasma and Its Effect on ADC Efficacy. *PLOS ONE* **2015**, *10*, e0132282.

(13) Huang, R. Y.-C.; Chen, G. Characterization of Antibody–Drug Conjugates by Mass Spectrometry: Advances and Future Trends. *Drug Discov. Today* **2016**, *21*, 850–855.

(14) He, L.; Anderson, L. C.; Barnidge, D. R.; Murray, D. L.; Hendrickson, C. L.; Marshall, A. G. Analysis of Monoclonal Antibodies in Human Serum as a Model for Clinical Monoclonal Gammopathy by Use of 21 Tesla FT-ICR Top-Down and Middle-Down MS/MS. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 827–838.

(15) Wagh, A.; Song, H.; Zeng, M.; Tao, L.; Das, T. K. Challenges and New Frontiers in Analytical Characterization of Antibody-Drug Conjugates. *MAbs* **2018**, *10*, 222–243.

(16) Nakada, T.; Sugihara, K.; Jikoh, T.; Abe, Y.; Agatsuma, T. The Latest Research and Development into the Antibody–Drug Conjugate, [Fam-] Trastuzumab Deruxtecan (DS-8201a), for HER2 Cancer Therapy. *Chem. Pharm. Bull.* **2019**, *67*, 173–185.

(17) Said, N.; Gahoual, R.; Kuhn, L.; Beck, A.; François, Y.-N.; Leize-Wagner, E. Structural Characterization of Antibody Drug Conjugate by a Combination of Intact, Middle-up and Bottom-up Techniques Using Sheathless Capillary Electrophoresis – Tandem Mass Spectrometry as NanoESI Infusion Platform and Separation Method. *Anal. Chim. Acta* **2016**, *918*, 50–59.

(18) He, J.; Su, D.; Ng, C.; Liu, L.; Yu, S.-F.; Pillow, T. H.; Del Rosario, G.; Darwish, M.; Lee, B.-C.; Ohri, R.; Zhou, H.; Wang, X.; Lu, J.; Kaur, S.; Xu, K. High-Resolution Accurate-Mass Mass Spectrometry Enabling In-Depth Characterization of in Vivo Biotransformations for Intact Antibody-Drug Conjugates. *Anal. Chem.* **2017**, *89*, 5476–5483.

(19) Terral, G.; Beck, A.; Cianférani, S. Insights from Native Mass Spectrometry and Ion Mobility-Mass Spectrometry for Antibody and Antibody-Based Product Characterization. *J. Chromatogr. B* **2016**, *1032*, 79–90.

(20) D'Atri, V.; Fekete, S.; Stoll, D.; Lauber, M.; Beck, A.; Guillarme, D. Characterization of an Antibody-Drug Conjugate by Hydrophilic Interaction

Chromatography Coupled to Mass Spectrometry. *J. Chromatogr. B* **2018**, *1080*, 37–41.

(21) Chen, T.-H.; Yang, Y.; Zhang, Z.; Fu, C.; Zhang, Q.; Williams, J. D.; Wirth, M. J. Native Reversed-Phase Liquid Chromatography: A Technique for LCMS of Intact Antibody–Drug Conjugates. *Anal. Chem.* **2019**, *91*, 2805–2812.

(22) Debaene, F.; Bœuf, A.; Wagner-Rousset, E.; Colas, O.; Ayoub, D.; Corvaïa, N.; Van Dorsselaer, A.; Beck, A.; Cianférani, S. Innovative Native MS Methodologies for Antibody Drug Conjugate Characterization: High Resolution Native MS and IM-MS for Average DAR and DAR Distribution Assessment. *Anal. Chem.* **2014**, *86*, 10674–10683.

(23) Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. Top-Down Structural Analysis of an Intact Monoclonal Antibody by Electron Capture Dissociation-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 4239–4246.

(24) Beck, A.; Terral, G.; Debaene, F.; Wagner-Rousset, E.; Marcoux, J.; Janin-Bussat, M.-C.; Colas, O.; Dorsselaer, A. V.; Cianférani, S. Cutting-Edge Mass Spectrometry Methods for the Multi-Level Structural Characterization of Antibody-Drug Conjugates. *Expert Rev. Proteomic* **2016**, *13*, 157–183.

(25) Fekete, S.; Guillarme, D.; Sandra, P.; Sandra, K. Chromatographic, Electrophoretic, and Mass Spectrometric Methods for the Analytical Characterization of Protein Biopharmaceuticals. *Anal. Chem.* **2016**, *88*, 480–507.

(26) Wagner-Rousset, E.; Janin-Bussat, M.-C.; Colas, O.; Excoffier, M.; Ayoub, D.; Haeuw, J.-F.; Rilatt, I.; Perez, M.; Corvaïa, N.; Beck, A. Antibody-Drug Conjugate Model Fast Characterization by LC-MS Following IdeS Proteolytic Digestion. *mAbs* **2014**, *6*, 173–184.

(27) Botzanowski, T.; Erb, S.; Hernandez-Alba, O.; Ehkirch, A.; Colas, O.; Wagner-Rousset, E.; Rabuka, D.; Beck, A.; Drake, P. M.; Cianférani, S. Insights from Native Mass Spectrometry Approaches for Top- and Middle- Level Characterization of Site-Specific Antibody-Drug Conjugates. *mAbs* **2017**, *9*, 801–811.

(28) Dai, J.; Zhang, Y. A Middle-Up Approach with Online Capillary Isoelectric Focusing/Mass Spectrometry for In-Depth Characterization of Cetuximab Charge Heterogeneity. *Anal. Chem.* **2018**, *90*, 14527–14534.

(29) Haselberg, R.; De Vijlder, T.; Heukers, R.; Smit, M. J.; Romijn, E. P.; Somsen, G. W.; Domínguez-Vega, E. Heterogeneity Assessment of Antibody-Derived Therapeutics at the Intact and Middle-up Level by Low-Flow Sheathless Capillary Electrophoresis-Mass Spectrometry. *Anal. Chim. Acta* **2018**, *1044*, 181–190.

(30) Fornelli, L.; Srzentić, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence

Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **2018**, *90*, 8421–8429.

(31) Fornelli, L.; Ayoub, D.; Aizikov, K.; Liu, X.; Damoc, E.; Pevzner, P. A.; Makarov, A.; Beck, A.; Tsybin, Y. O. Top-down Analysis of Immunoglobulin G Isotypes 1 and 2 with Electron Transfer Dissociation on a High-Field Orbitrap Mass Spectrometer. *Journal of Proteomics* **2017**, *159*, 67–76.

(32) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. *Anal. Chem.* **2014**, *86*, 3005–3012.

(33) Cotham, V. C.; Brodbelt, J. S. Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level Using Middle-Down 193 Nm Ultraviolet Photodissociation. *Anal. Chem.* **2016**, *88*, 4004–4013.

(34) Resemann, A.; Jabs, W.; Wiechmann, A.; Wagner, E.; Colas, O.; Evers, W.; Belau, E.; Vorwerg, L.; Evans, C.; Beck, A.; Suckau, D. Full Validation of Therapeutic Antibody Sequences by Middle-up Mass Measurements and Middle-down Protein Sequencing. *MAbs* **2016**, *8*, 318–330.

(35) Jin, Y.; Lin, Z.; Xu, Q.; Fu, C.; Zhang, Z.; Zhang, Q.; Pritts, W. A.; Ge, Y. Comprehensive Characterization of Monoclonal Antibody by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *mAbs* **2019**, *11*, 106–115.

(36) Srzentić, K.; Nagornov, K. O.; Fornelli, L.; Lobas, A. A.; Ayoub, D.; Kozhinov, A. N.; Gasilova, N.; Menin, L.; Beck, A.; Gorshkov, M. V.; Aizikov, K.; Tsybin, Y. O. Multiplexed Middle-Down Mass Spectrometry as a Method for Revealing Light and Heavy Chain Connectivity in a Monoclonal Antibody. *Anal. Chem.* **2018**, *90*, 12527–12535.

(37) An, Y.; Zhang, Y.; Mueller, H.-M.; Shameem, M.; Chen, X. A New Tool for Monoclonal Antibody Analysis. *mAbs* **2014**, *6*, 879–893.

(38) Biacchi, M.; Said, N.; Beck, A.; Leize-Wagner, E.; François, Y.-N. Top-down and Middle-down Approach by Fraction Collection Enrichment Using off-Line Capillary Electrophoresis – Mass Spectrometry Coupling: Application to Monoclonal Antibody Fc/2 Charge Variants. *J. Chromatogr. A* **2017**, *1498*, 120–127.

(39) Shaw, J. B.; Liu, W.; Vasil′ev, Y. V.; Bracken, C. C.; Malhan, N.; Guthals, A.; Beckman, J. S.; Voinov, V. G. Direct Determination of Antibody Chain Pairing by Top-down and Middle-down Mass Spectrometry Using Electron Capture Dissociation and Ultraviolet Photodissociation. *Anal. Chem.* **2020**, *92*, 766–773.

(40) van der Burgt, Y. E. M.; Kilgour, D. P. A.; Tsybin, Y. O.; Srzentić, K.; Fornelli, L.; Beck, A.; Wuhrer, M.; Nicolardi, S. Structural Analysis of Monoclonal

Antibodies by Ultrahigh Resolution MALDI In-Source Decay FT-ICR Mass Spectrometry. *Anal. Chem.* **2019**, *91*, 2079–2085.

(41) Janin-Bussat, M.-C.; Dillenbourg, M.; Corvaia, N.; Beck, A.; Klinguer-Hamour, C. Characterization of Antibody Drug Conjugate Positional Isomers at Cysteine Residues by Peptide Mapping LC–MS Analysis. *J. Chromatogr. B* **2015**, *981–982*, 9–13.

(42) Beck, A.; D'Atri, V.; Ehkirch, A.; Fekete, S.; Hernandez-Alba, O.; Gahoual, R.; Leize-Wagner, E.; François, Y.; Guillarme, D.; Cianférani, S. Cutting-Edge Multi-Level Analytical and Structural Characterization of Antibody-Drug Conjugates: Present and Future. *Expert Rev. Proteomic* **2019**, *16*, 337–362.

(43) Li, K.; Lin, Z. J.; Shi, H.; Ma, Y. Characterization of Positional Isomers of Interchain Cysteine Linked Antibody−Drug Conjugates by High-Resolution Mass Spectrometry. *Anal. Chem.* **2019**, *91*, 8558–8563.

(44) Tian, Y.; Lippens, J. L.; Netirojjanakul, C.; Campuzano, I. D. G.; Ruotolo, B. T. Quantitative Collision-Induced Unfolding Differentiates Model Antibody–Drug Conjugates. *Protein Science* **2019**, *28*, 598–608.

(45) Chen, B.; Lin, Z.; Zhu, Y.; Jin, Y.; Larson, E.; Xu, Q.; Fu, C.; Zhang, Z.; Zhang, Q.; Pritts, W. A.; Ge, Y. Middle-Down Multi-Attribute Analysis of Antibody-Drug Conjugates with Electron Transfer Dissociation | Analytical Chemistry. *Anal. Chem.* **2019**, *91*, 11661–11669.

(46) Hernandez-Alba, O.; Houel, S.; Hessmann, S.; Erb, S.; Rabuka, D.; Huguet, R.; Josephs, J.; Beck, A.; Drake, P. M.; Cianférani, S. A Case Study to Identify the Drug Conjugation Site of a Site-Specific Antibody-Drug-Conjugate Using Middle-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2419–2429.

(47) Kunik, V.; Ashkenazi, S.; Ofran, Y. Paratome: An Online Tool for Systematic Identification of Antigen-Binding Regions in Antibodies Based on Sequence or Structure. *Nucleic Acids Res* **2012**, *40*, W521–W524.

(48) Kunik, V.; Peters, B.; Ofran, Y. Structural Consensus among Antibodies Defines the Antigen Binding Site. *PLoS Comput Biol* **2012**, *8*, e1002388.

(49) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135*, 12646–12651.

(50) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 1044–1051.

(51) Sanders, J. D.; Mullen, C.; Watts, E.; Holden, D. D.; Syka, J. E. P.; Schwartz, J. C.; Brodbelt, J. S. Enhanced Sequence Coverage of Large Proteins by Combining Ultraviolet Photodissociation with Proton Transfer Reactions. *Anal. Chem.* **2020**, *92*, 1041–1049.

(52) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. Integration of Ultraviolet Photodissociation with Proton Transfer Reactions and Ion Parking for Analysis of Intact Proteins. *Anal. Chem.* **2016**, *88*, 1008–1016.

(53) Liu, J.; McLuckey, S. A. Electron Transfer Dissociation: Effects of Cation Charge State on Product Partitioning in Ion/Ion Electron Transfer to Multiply Protonated Polypeptides. *Int. J. Mass Spectrom.* **2012**, *330–332*, 174–181.

(54) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell. Proteom.* **2007**, *6*, 1942–1951.

Chapter 4: Enhanced Characterization of Lysine-Linked Antibody Drug Conjugates Enabled by Middle-Down Mass Spectrometry and HCD-Triggered EThcD and UVPD[2]

4.1 Abstract

As the development of new biotherapeutics advances, increasingly sophisticated tandem mass spectrometry methods are needed to characterize the most complex molecules, including antibody drug conjugates (ADCs). Lysine-linked ADCs, such as trastuzumab-emtansine (T-DM1), are among the most heterogeneous biotherapeutics. Here we implement a workflow that combines limited proteolysis with HCD-triggered EThcD and UVPD mass spectrometry for characterization of the resulting large middle-down sized peptides of T-DM1. A total of 48 payload-containing peptides were identified, ranging in mass from 1.8 to 14.7 kDa, and leading to the unambiguous identification of 44 out of 92 possible conjugation sites. In addition, two peptides were identified containing multiple payloads. The characterization of these types of heterogeneous peptides represents an important step in unravelling the combinatorial nature of lysine-conjugated ADCs.

4.2 Introduction

Antibody-drug conjugates (ADCs) represent a promising class of therapeutic treatments. The combination of a selective antibody and cytotoxic payload in the form of an ADC has been hailed as a "magic bullet" given their promise to dramatically improve treatments.[1,2] While a variety of different modalities, including site-specific ADCs, are under development and in clinical trials, all ADCs that have received FDA approval fall

[2]Watts, E.; Crittenden, C.M.; Brodbelt, J. S. Enhanced Characterization of Lysine-Linked Antibody Drug Conjugates Enabled by Middle-Down Mass Spectrometry and HCD-Triggered EThcD and UVPD . *Anal. Chem.* **2023**, In Review.

into the categories of either cysteine or lysine linked ADCs.[3] The use of either reduced cysteines or solvent accessible lysine residues adds heterogeneity to the antibodies, increasing the challenges of characterization. Recent advances in mass spectrometry techniques have resulted in enhanced structural characterization of ADCs as well as improved differentiation of positional isomers with middle-down techniques.[4,5] However, the translation of these methods to lysine-linked ADCs has been limited.

The complexity of an ADC and the difficulty associated with its characterization arises largely from the modality of the payload-to-antibody linkage. The least complex case is site-specific, or next-generation ADCs, which utilize modified amino acid residues, typically on the fragment crystallizable (Fc) domain of the heavy chain (HC), to link the payload to two or four specific amino acids.[6] Cysteine-linked ADCs typically contain an average of four payloads attached to interchain disulfide bonds, resulting in eight possible conjugation sites per ADC.[6] Finally, lysine-linked ADCs are derived from linker conjugation to random lysine residues along the entire amino-acid sequence of the antibody.[6] The large number of lysine residues on antibodies, typically around 90 in total, increases the challenge of characterizing lysine-linked ADCs.

Despite the challenges in characterization, significant strides have been made to improve the characterization of ADCs. Until recently, drug-to-antibody ratios (DAR) were typically measured through liquid chromatography coupled to UV-visible spectroscopy, and bottom-up proteomic methods were the only feasible means to identify the locations of payload binding sites.[7,8] Enhanced native mass spectrometry, chromatographic methods, and the increased availability of high resolution mass spectrometry instrumentation have facilitated more advanced characterization of intact ADCs through intact mass spectrometry, making it the new gold standard for DAR assessment.[9–19] Moreover, recent developments in ion-mobility and hydrogen-deuterium

exchange mass spectrometry have further elevated the capabilities of mass spectrometry for structural characterization of ADCs.[20–27] Finally, the growing prevalence of subunit-based middle-down strategies have eliminated the notion that drug conjugation site identification can only be achieved with bottom-up proteomics.[4,5,28,29] Despite the prolific achievements described thus far, very few studies have translated the successes of intact and structural characterization or middle-down mass spectrometry to lysine-linked ADCs.

Given the complexity associated with lysine-linked ADCs, bottom-up mass spectrometry remains the primary method to identify payload locations. Varying numbers of payloads have been identified for lysine linked ADCs.[30] Most studies report close to 40 out of 92 possible sites identified, while one reported 82 conjugation site.[31–34] These studies typically rely on digestion with trypsin, resulting in peptides containing only one possible lysine conjugation site per peptide which greatly simplifies the localization of the payloads but eliminates all chances of identifying combinatorial modifications. Most bottom-up ADC studies use collisionally activated dissociation (CAD) for characterization of the peptides which, in addition to generating sequence ions, may result in fragmentation of the labile payload or its cleavage from the ADC.[31–34] The generation of highly abundant payload-related fragment ions has been reported for CAD of ADCs containing emtansine (DM1), the payload commonly used in lysine-linked ADCs.[31–34] The presence of these fragment ions has been exploited to unambiguously detect the presence of payload-containing peptides.[31–34] This feature becomes a significant attribute in the development of CAD-based methods that aim to screen digests for the presence of payload-containing peptides, as utilized in the present study.

While bottom-up proteomics methods have proven successful for identifying payload locations, they are frequently unable to capture the full heterogeneity of ADCs nor unravel the context of multiple co-existing payload locations. Employing a middle-

down approach in which larger peptides are generated and analyzed is a promising option for improved characterization of complex biomolecules.[36–42] Several studies have now reported characterization of site-specific, cysteine-linked and even lysine-linked ADCs using middle-down strategies enabled by IdeS or IdeZ proteases which cleave ADCs into large subunits.[29,4,5] Despite the achievements of these studies, the characterization of the lysine-linked ADCs at the subunit level was hampered by the increased complexity compared to the cysteine-linked counterparts.[5] The limitations of subunit-level characterization can be subverted by employing alternative proteases or conditions to modulate the peptide sizes, generating ones larger than tryptic peptides but smaller than intact subunits. Limited proteolysis has resulted in improved characterization of proteins, including monoclonal antibodies.[36,42,43] This approach could augment the characterization of lysine-linked ADCs by allowing the generation of longer peptides representative of proteoforms containing multiple drug linkages and is adopted in the present study.

Collision-based dissociation methods are the most well established for the identification of peptide sequences. Recently, alternative MS/MS methods, including electron-transfer/higher-energy collision dissociation (EThcD) and ultraviolet photodissociation (UVPD) have gained popularity. EThcD is a hybrid method combining electron transfer dissociation (ETD) and CAD to enhance the conversion of charge-reduced peptides into diagnostic *b/y* and *c/z* fragment ions along with retention of labile modifications.[44] UVPD is a higher energy activation method that causes extensive fragmentation of peptides and proteins and also allows retention of labile PTMs.[45] The application of these MS/MS methods to lysine-linked ADCs proved vital for the present investigation.

In this study, we focus on advancing the characterization of lysine-linked ADCs, as exemplified by trastuzumab-emtansine (T-DM1, brand name Kadcyla), via a middle-

down strategy that utilizes limited proteolysis with Lysyl-endopeptidase (Lys-C). To increase the throughput of the liquid chromatography-MS/MS workflow and the confidence in identification of payload-containing peptides, a higher energy collisional dissociation (HCD) triggered method is employed, using payload-related reporter ions which have already been reported for DM1 to trigger a second high resolution MS^2 event utilizing either UVPD or EThcD. The performance of the two auxiliary MS/MS methods is compared in the context of localizing multiple payloads.

4.3 EXPERIMENTAL

4.3.1 Sample Preparation

Lyophilized lysine-linked antibody-drug conjugate (T-DM1) was provided by Genentech. The T-DM1 sample was reconstituted in water to a concentration of approximately 2.5 mg/mL then desalted with Micro Bio-Spin 6 (Bio-Rad) prior to digestion. After cleanup, samples were diluted to 1 mg/mL in 50 mM sodium phosphate pH 7.5 (New England Biolabs). Limited proteolysis was achieved by adding Lys-C (Santa Cruz Biotechnology) in a 1:75 enzyme to protein ratio and digesting for 4 hours at 37°C. Reduction was completed after digestion with 50 mM tris(2-carboxyethyl)phosphine and 2 M urea. Samples were diluted in water prior to analysis to allow 0.5 µg injections. All reagents were purchased from Thermo-Fisher Scientific unless otherwise noted.

4.3.2 Liquid Chromatography-Mass Spectrometry

Liquid chromatography-mass spectrometry was performed using a Dionex Ultimate nano liquid chromatography system coupled to a Thermo Scientific Instruments Orbitrap Fusion Lumos mass spectrometer equipped with 193 nm UVPD as previously

described.[46] Chromatographic separation was achieved with trap and elute using columns house-packed with Agilent polymeric reversed-phase (PLRP-S) bulk media. Trap (100 μm ID/ 360 μm OD) and analytical columns (75 μm ID/ 360 μm OD) were packed with 5 μm bulk PLRP media (1000 Å pore size) from Agilent to a length of 3 and 20 cm, respectively. Samples were injected under starting conditions of 2% acetonitrile and 0.1% formic acid in water at 5 μL/min. After 5 min of loading, a valve switch placed the trap column in-line with the analytical column. Analytical mobile phases comprising water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B) were applied at a rapid initial gradient of 2 to 20% B over 2 min followed by a slower gradient up to 40% B over 33 min at a flow rate of 300 nL/min.

Following chromatographic separation, the eluate was introduced to the mass spectrometer by electrospray ionization with an applied voltage of 2 kV. MS1 spectra were collected with a resolution of 60,000 at *m/z* 200, an AGC target of 4e5, and a 50 ms maximum ion injection time. Data dependent properties were set to allow 10 scans with exclusion after two repeats within 7 s and a 30 s exclusion duration. All runs included a preliminary HCD MS^2 event with 30% normalized collisional energy, a resolution of 30,000 at *m/z* 200, an AGC target of 5e5, and a 54 ms maximum ion injection time. Targeted inclusion triggered a second MS^2 event on the same precursor if a fragment ion of *m/z* 547.22 was detected in the HCD scan. The secondary MS^2 event utilized either 193 nm UVPD with two pulses (applied during a 5 ms activation period) with 2 mJ per pulse, or EThcD with calibrated charge-dependent activation period ranging from 3 to 150 ms and 15% normalized collisional energy supplemental activation. Secondary MS^2 scans utilized a resolution of 240,000 at *m/z* 200, an AGC target of 5e5, and a 502 ms maximum ion injection time. Five technical replicates were collected for both UVPD and EThcD.

4.3.3 Data Analysis

Data analysis was performed with ProSight PD 4.2 within Proteome Discover 3.0. To implement middle-down data processing, a custom middle-down database was generated using Protein Digestion Simulator (Pacific Northwest National Labs) to create a list of peptides produced from any number of missed cleavages at lysine. This list was input into Proteome Discoverer as a FASTA file. The addition of a 956.364 Da payload was included as a custom variable modification for each lysine residue and for the N-terminus of the heavy and light chains. Spectra were processed with ProSight PD High/High cRAWler and matched with ProSight PD 4.2 Annotated Proteoform search. In the cRAWler module a fit factor of 0.80, a remainder of 0.25, and a S/N threshold of 3 was used for deconvolution of fragmentation spectra with the Xtract algorithm. For the annotated proteoforms search precursor mass tolerance was set to 2.2 Da and fragment mass tolerance was set to 10 ppm. The "UVPD 9" setting was used to allow inclusion of *a, a+1, b, c, x, x+1, y, y-1*, and *z* ions in UVPD search. Matches were filtered to only include those identified with "Medium Confidence" or higher. Payload-containing hits were only considered if they were matched in secondary MS^2 scans (*i.e.* EThcD or UVPD) in at least three out of the five replicates and matches were individually validated. Details of the manual validation process are explained in the discussion. In some cases, particularly for the doubly conjugated species, manual analysis involved examination of spectra in runs where the peptide was not matched using ProSight PD. In this case Xtract deconvolution was achieved in FreeStyle (Thermo) and fragment ions were matched with ProSight Lite using the same parameters listed above for Proteome Discoverer.

All raw data files and an excel table displaying mass errors of identified fragment ions of the peptides listed in **Table 4.1** and **Table 4.2** are available in massive database

(massive.ucsd.edu) with accession number MSV000090758. *For reviewer access use username: MSV000090758_reviewer and password: ECWNov22.*

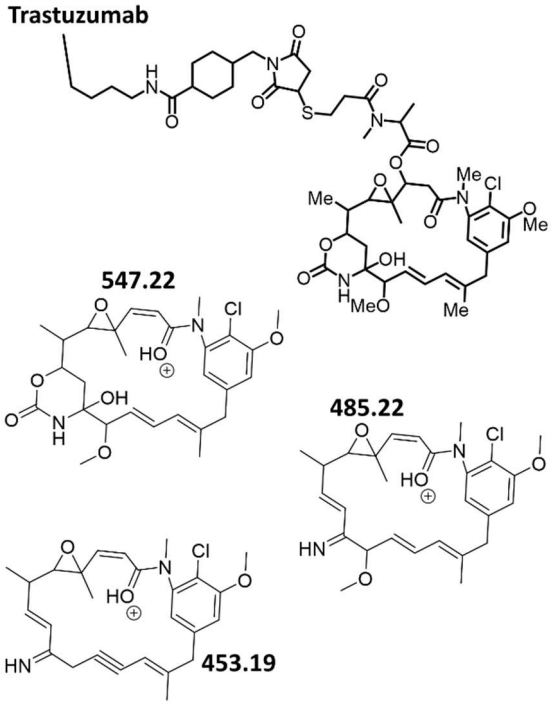
4.4 Results and Discussion

4.4.1 Development of Middle-Down Method for Payload Containing Peptides

While specific enzymatic digestion at the hinge region of an antibody is a popular approach for middle-down characterization, it has significant limitations in localizing payloads to specific lysine residues due to the large sizes (25-100 kDa) of the resulting subunits (i.e., Fc, Fc/2, $F(ab')_2$, Fd') and the substantial number of lysines per subunit (13 to 52). The generation of middle-down peptides in the 3-10 kDa range has shown promise for antibodies, histones, and other proteins containing complex post-translational modifications.[36,41,43] In these prior studies, limited proteolysis was achieved using Lys-C with a high antibody-to-protease ratio and shorter digestion time than would be used for conventional bottom-up proteolysis. Because this method was successful for generation of large peptides for the characterization of antibodies in the past,[36] this strategy seemed promising for analysis of T-DM1.

While CAD (including beam-type HCD implemented on Thermo mass spectrometers) has been highly successful for characterization of small peptides, such as those generated by tryptic digestion, alternative higher energy ion activation methods like EThcD and UVPD are often better suited for characterization of larger peptides. For EThcD and UVPD, the longer signal averaging required to adequately resolve the denser fragmentation patterns generated relative to CAD reduces throughput. Methods that capitalize on the high throughput of CAD and the enhanced peptide characterization of ETD or UVPD have been developed and are generally known as "triggered" methods, as demonstrated for analysis of phosphopeptides and glycopeptides.[47,48] In essence, the

slower ETD or UVPD scans are only acquired if an initial CAD scan generates a reporter ion characteristic of a particular type or class of peptide. The second stage of MS/MS is undertaken on the same precursor ion to enable more detailed structural characterization. By only collecting a higher resolution EThcD or UVPD scan if payload-specific fragment ions are detected by HCD, the number of high-resolution scans is reduced, improving the duty cycle and maximizing the time spent analyzing the peptides of interest. Activation of peptides containing the DM1 payload of T-DM1 with CAD has previously been reported to generate unique fragment ions which have even been exploited to aid in data processing.[31–34] The presumed structures of these fragment ions, with *m/z* values of 453.19, 485.22, and 547.22 are displayed in **Scheme 4.1**. These same ions were adopted as potential reporter ions in the present study.



Scheme 4.1: Structure of DM1 payload conjugated to trastuzumab and possible fragment ion structures that correspond to *m/z* 547.22, 485.22, and 453.19 Da payload reporter ions observed in the HCD spectra.

The chromatogram in **Figure 4.1** illustrates the separation of the peptides generated by limited proteolysis of T-DM1, along with the extracted ion chromatograms corresponding to the contribution of the payload-specific fragment ions. The total ion chromatogram in **Figure 4.1A** includes all peptides, regardless of whether they are key payload-containing peptides or not. **Figure 4.1B** displays only those peptides which produced DM1-payload specific fragments upon HCD of the eluting peptides. Closer examination of the individual HCD spectra revealed that the fragment ion of *m/z* 547.22 was consistently the highest abundance and most prevalent of the three reporter ions, as exemplified in **Figures 4.2A** and **4.3A**. Thus, for the HCD-triggered UVPD and EThcD methods in our strategy, the *m/z* 547.22 reporter ion was utilized.

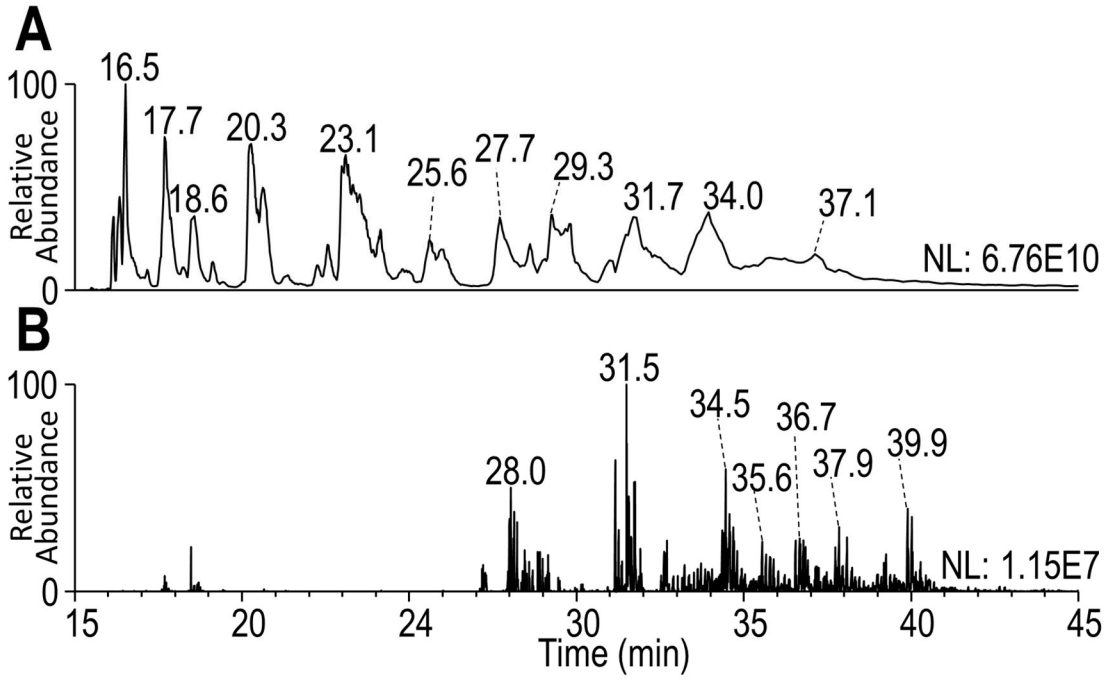


Figure 4.1: (A) Total ion MS^1 chromatogram and (B) HCD MS^2 extracted ion chromatogram for reporter ions of *m/z* 547.22, 485.22, and 453.19 from T-DM1 digest.

As expected, HCD resulted in adequate characterization for smaller peptides without multiple payload sites, as shown in the example in **Figure 4.2A**. Interestingly, even for this relatively small 4.5 kDa peptide containing 30 residues, the level of characterization by HCD is limited in that no payload-containing fragment ions are generated, indicating that the DM1 payload is labile and readily cleaved by HCD. In contrast, EThcD and UVPD (**Figure 4.2B-C**) generated many payload-containing products in addition to providing more extensive sequence coverage.

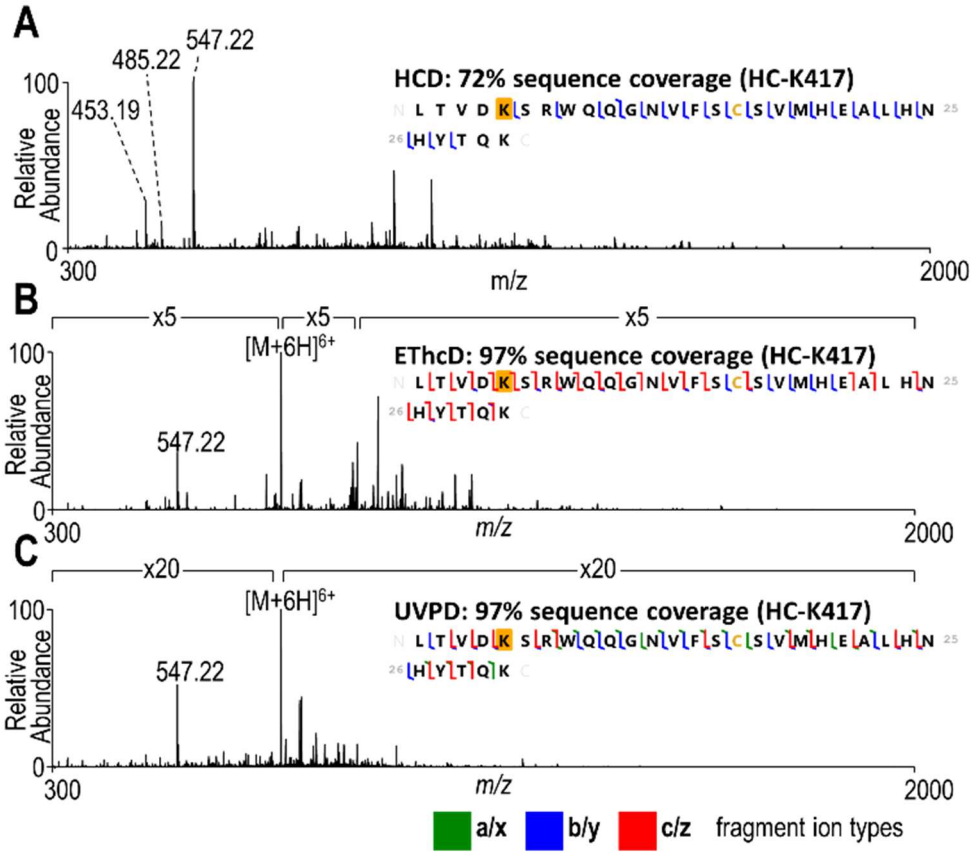


Figure 4.2: MS/MS spectra of a 4.5 kDa peptide (6+) using: (A) HCD (30% NCE), (B) EThcD (charge calibrated activation and 15% NCE supplemental activation), and (C) 193 nm (UVPD 2 pulses, 2 mJ per pulse). Sequence coverage maps along with sequence coverages are included for each spectrum. K5 of the peptide sequence shown here corresponds to K417 of the antibody HC.

For the larger 8.9 kDa peptide shown in **Figure 4.3**, the enhanced fragmentation offered by EThcD and UVPD is even more beneficial. In this example the HCD spectrum (**Figure 4.3A**) displays the abundant payload reporter ions in the low *m/z* range but only one fragment ion (b_{32}) occurs between any of the three lysine residues. The EThcD spectrum in **Figure 4.3B** offers higher sequence coverage (72%), including 25 payload-containing fragments that localize the conjugation site to HC-K417. UVPD yields an even higher sequence coverage (78%) and 46 payload-containing fragment ions (**Figure 4.3C**). While EThcD resulted in more backbone cleavages between HC-K412 and HC-K417, UVPD yielded greatly increased coverage between HC-K392 and HC-K412, highlighting the complementarity of the two activation methods that amplifies confidence in site localization.

In the present strategy, acquisition of the HCD spectrum is the primary step used for the HCD-triggered methods, and thus this screening MS/MS spectrum was collected at a lower resolution and with a lower maximum ion injection time than the subsequent triggered EThcD and UVPD spectra. Characterization of the peptide by HCD is improved by increasing the resolution of the HCD scan, as shown in **Figure 4.4** for the same 8.9 kDa peptide analyzed in **Figure 4.3**. The gain in sequence coverage is minimal (going from 31% to 38%) and still fails to localize the payload site. The limitations of HCD in characterizing very large peptides have already been well established,[36,42] and thus the optimal strategy utilizes the high speed and sensitivity of HCD (at lower spectral resolution) to generate reporter ions to trigger the subsequent slower but more informative EThcD or UVPD spectra (at higher spectral resolution). One additional benefit of the HCD screening step is the confidence gained by identifying the highly specific payload reporter ions which uniquely differentiate conjugated peptides from non-conjugated ones.

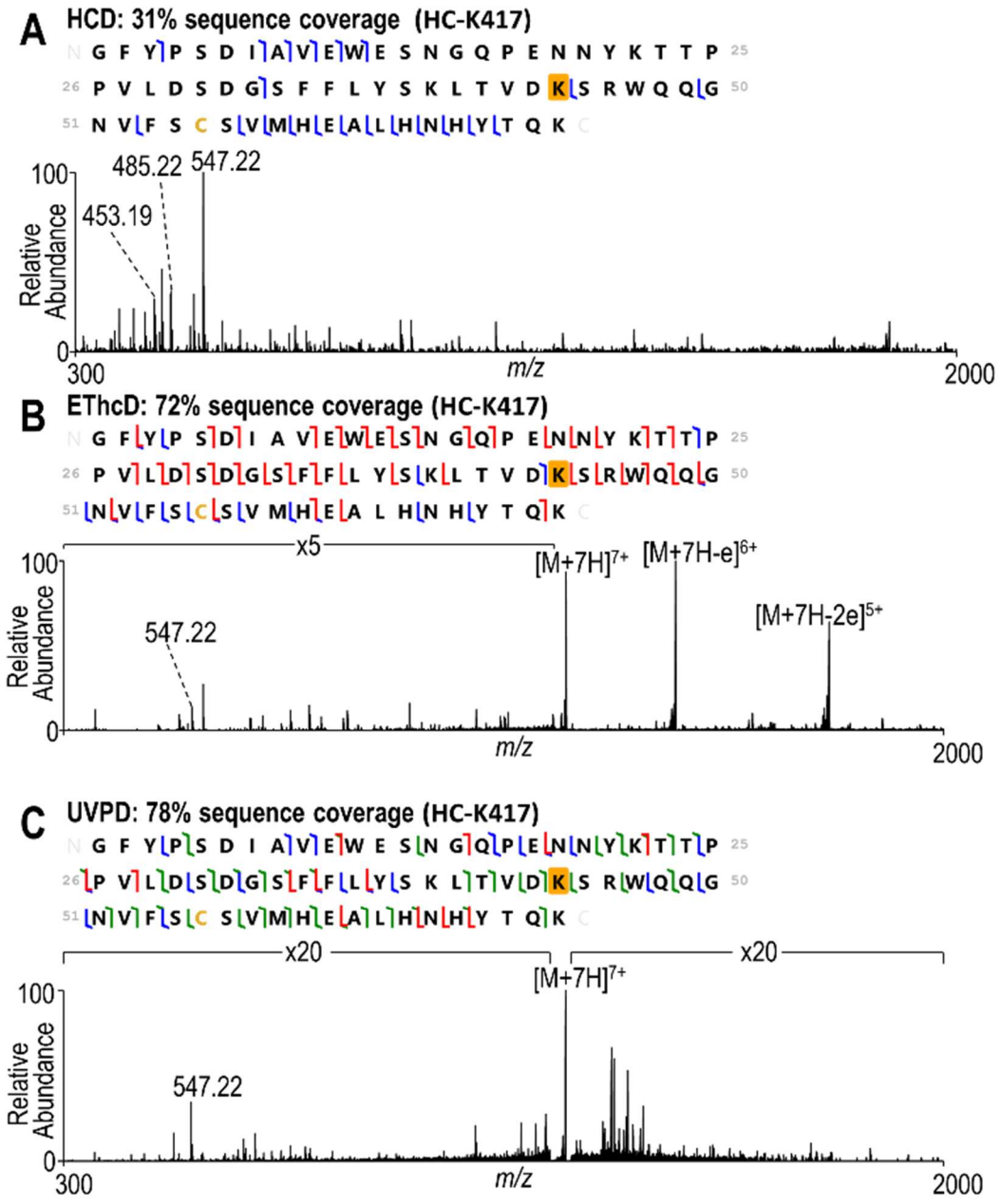


Figure 4.3: MS/MS spectra of an 8.9 kDa peptide by: (A) HCD (30% NCE), (B) EThcD (charge calibrated activation and 15% NCE supplemental activation), and (C) 193 nm (UVPD 2 pulses, 2 mJ per pulse). The 6+ charge state is displayed for HCD and the 7+ for UVPD and EThcD in order to achieve the best characterization for each method. Sequence coverage maps along with sequence coverages are included for each spectrum. K44 of the peptide sequence shown here corresponds to K417 of the antibody HC.

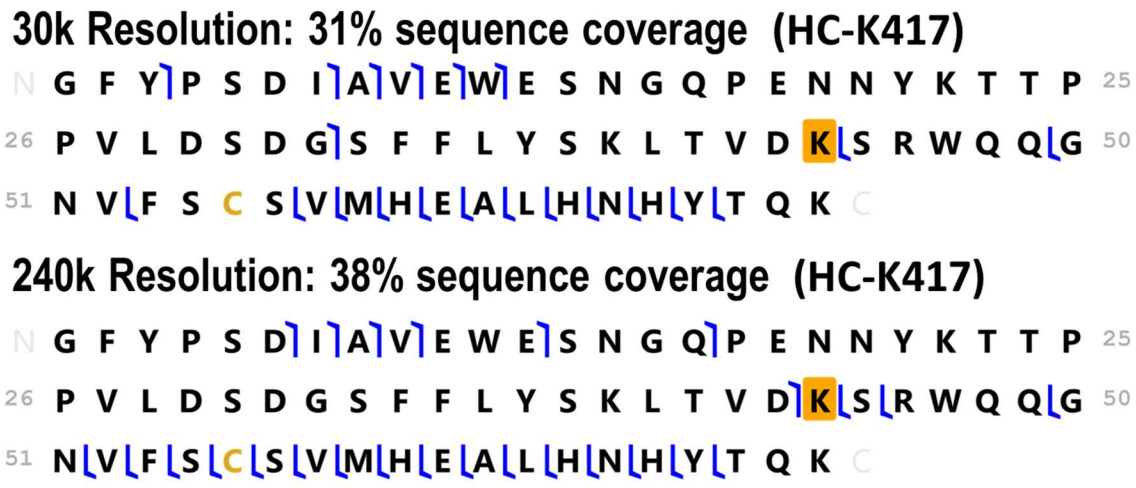


Figure 4.4: Sequence Coverage plots of an 8.9 kDa peptide (6+) for HCD with (A) 30,000 and (B) 240,000 resolution at *m/z* 200. K44 of the peptide sequence shown here corresponds to K417 of the antibody HC.

Complexity of the spectra increases with peptide size, necessitating the careful inspection of peptide spectral matches, particularly in cases where multiple lysine residues are present. For example, in the spectra shown in **Figure 4.3B,C**, a number of site-localizing fragment ions (ones which originate from backbone cleavages in regions of the sequence between lysine residues) were scrutinized to ensure that their isotope profiles matched the theoretical composition of the assigned ions. Examples of isotopic fits for fragment ions bracketing potential modified lysine sites of the peptide are shown in **Figure 4.5**. The isotopic fits shown in **Figures 4.5B** and **4.5C** include examples of signal-to-noise and fit factors near the cutoff applied to all identified fragment ions. The validity of these threshold fragment ions demonstrates the quality of the data. Given the large number of peptides examined in this study, it is not feasible to manually validate all isotopic fits for all fragment ions for every peptide, as is often done in top-down studies that focus on localization of PTMs or in subunit-level middle-down studies of ADCs in which far fewer spectra are collected and curated.[4,5] As an additional criterion, a payload

site was not considered unambiguous unless multiple (at least two) fragment ions bracketed and thus differentiated the lysine residues in multiple replicates. In some cases, multiple positional isomers were present, in which case both positional isomers were considered unambiguously identified if at least four fragment ions specific to each position were present in at least three out of five replicates.

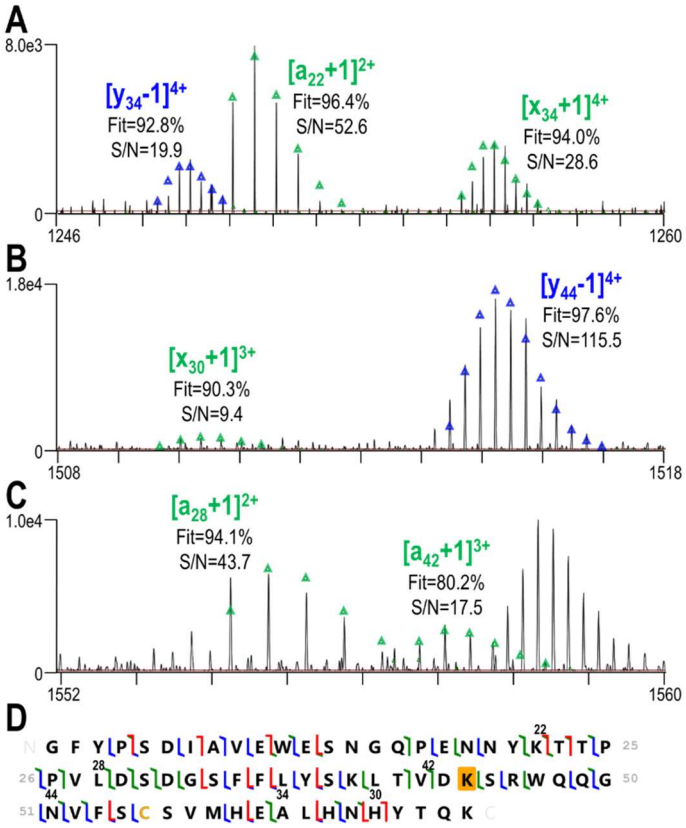


Figure 4.5: Expanded regions of the UVPD mass spectrum shown in **Figure 4.3C** illustrating examples of fragment ion isotope patterns with (A) high fit factors and signal to noise, (B) low signal to noise, and (C) low fit factors. Also included is (D) a sequence map which highlights the fragment ions identified. A variety of fragment ions were identified with and without the payload, including C-terminal payload containing fragments (A) $[y_{34}\text{-}1]^{4+}$, (A) $[x_{34}\text{+}1]^{4+}$, (B) $[x_{30}\text{+}1]^{3+}$, and (B) $[y_{44}\text{-}1]^{4+}$, and N-terminal fragments without the payload (A) $[a_{22}\text{+}1]^{2+}$, (C) $[a_{28}\text{-}\text{+}]^{2+}$, and (C) $[a_{42}\text{+}1]^{3+}$. Fit factors are calculated in Thermo Xtract and signal to noise values and the theoretical isotopic distributions are generated with TDValidator.

For peptides containing multiple lysine residues, peptides with different conjugation sites sometimes had different elution times (see **Figure 4.6**) which facilitated differentiation of conjugation sites. Other peptides were only observed as single chromatographic peaks which may or may not be composed of multiple conjugated isomers (**Figures 4.7-4.8**). Examination of the EThcD and UVPD fragmentation patterns in tandem was used to confirm the conjugation states. For each conjugated peptide, the sequence maps obtained by re-positioning the payload at each lysine were assessed for diagnostic lysine-bracketing fragment ions. For example, in **Figure 4.7** if the payload is positioned at HC-K124 then the sequence coverage is slightly lower than if the payload is positioned at HC-K136; however, there are many additional fragment ions generated by both EThcD and UVPD that support localization of the payload to HC-K124. Therefore, it can be concluded that both conjugation states are likely present, and the peptides are not chromatographically resolved. A similar conclusion was reached for the peptides in **Figure 4.8**: both K65 and K76 conjugation sites likely contribute to this HC peptide. The sequence maps for the example in **Figure 4.6** have sufficient payload-localizing fragment ions to confidently pinpoint the conjugated lysine positions (HC-K30 and HC-K43).

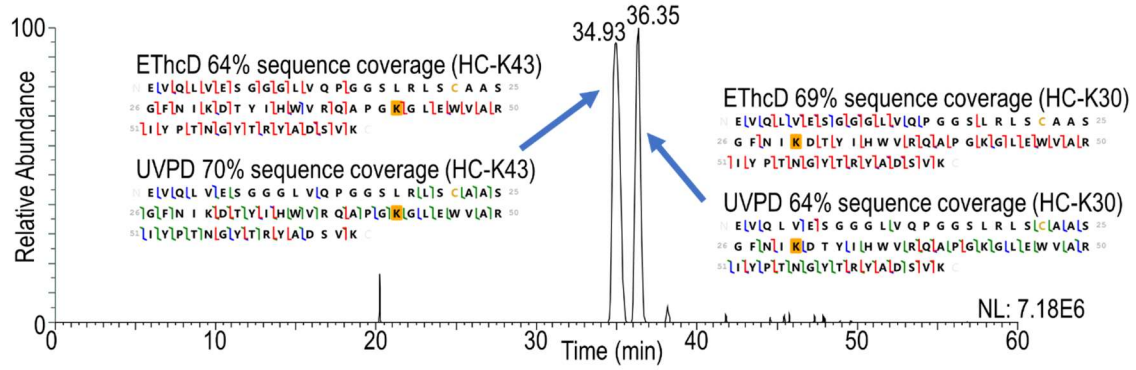


Figure 4.6: Extracted ion chromatogram revealing two HC peptides (8 kDa, 6+), each containing 65 residues (E1 through K65) and a single payload. Sequence coverage maps obtained by EThcD and UVPD localize the payload to K43 (peptide at 34.93 min) or K30 (peptide at 36.35 min).

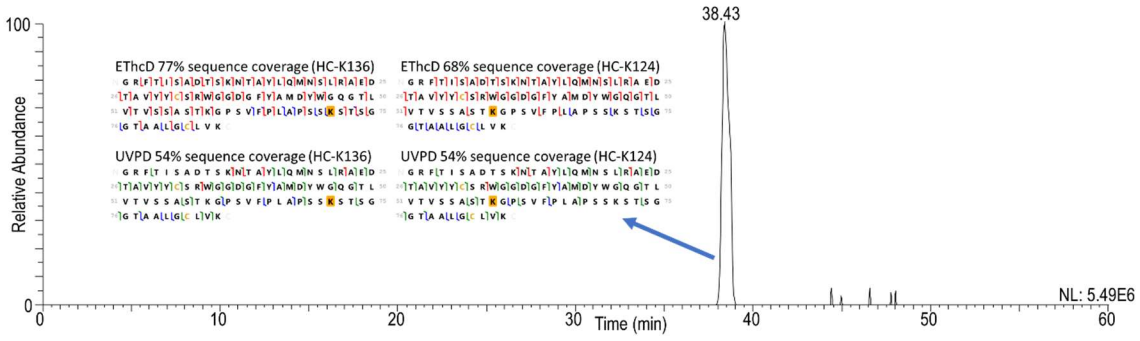


Figure 4.7: Extracted ion chromatogram of a HC peptide (9.9 kDa, 7+) containing G66 through K150 and a single payload. Sequence coverage maps are included for EThcD and UVPD which localize the payload to both K124 and K136, although they are not chromatically resolved and are therefore co-isolated. K59 and K71 of the sequence correspond to K124 and K136 of the HC.

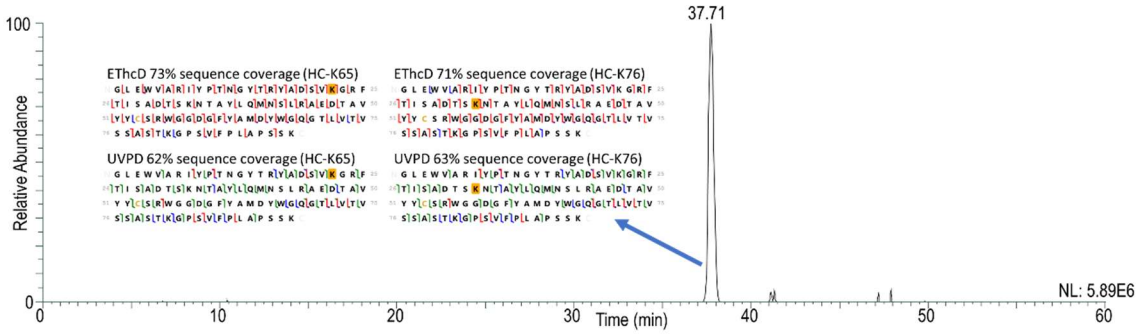


Figure 4.8: Extracted ion chromatogram of a HC peptide (11 kDa, 8+) containing G44 through K136 and a single payload. Sequence coverage maps are included for EThcD and UVPD which localize the payload to both K65 and K76, although they are not chromatically resolved and are therefore co-isolated. K22 and K33 of the sequence correspond to K65 and K76 of the HC.

4.4.2 Characterization of Heterogeneous Species with Multiple Conjugation Sites

While the peptides identified using ProSight PD, as listed in **Table 4.1** and **S2**, only included peptides with single payload conjugations, the production of peptides containing multiple modified sites was also considered. Based on the typical average DAR of around 3.5 for T-DM1,[32] and the fact that the payloads may be located across four subunits (two HCs and two LCs) the probability that two payloads will be conjugated in sufficiently close proximity to result in doubly-modified peptides is very low when the ADCs are subjected to conventional bottom-up proteolytic methods. However, the large sizes of the peptides generated by limited proteolysis have the potential to allow detection and characterization of multiply conjugated species, likely in low abundance. For example, trastuzumab has three lysines at positions HC-213, HC-216 and HC-217 which span only five residues. To increase the ability to identify the highly interesting but low abundance multiply-conjugated peptides, peptides with two payload conjugations were considered even if they were only identified in a single replicate. Four peptides containing two payloads were identified by ProSight PD in at least one replicate. Two could be confirmed by manual validation, as illustrated in **Figure 4.9** and **Figure 4.10**, and the more elaborate systematic strategy is described in the following section.

Table 4.1: List of payload-containing peptides identified with ProSight PD for UVPD replicates. For each peptide, the residue to which the payload was localized, the theoretical mass, as well as the sequence coverage and retention time for each replicate are listed. Some replicate entries are blank in the case that a peptide was not identified in all five technical replicates. Manually identified peptides from Figure 4.9 and Figure 4.10 are not included.

Annotated Sequence	Payload Localization (residue number on peptide sequence)	Payload Localization (residue number on protein sequence)	Theo. Mass [Da]	Replicate 1		Replicate 2		Replicate 3		Replicate 4		Replicate 5	
				SC	RT	SC	RT	SC	RT	SC	RT	SC	RT
ADYEKHK	K5	Lc-K188	1845.79	83%	27.41	83%	28.2	50%	28.37	83%	27.54	100%	27.66
ADYEKHKVYACEVTHQGLSSPVTK	K5	Lc-K188	3645.68	87%	26.61	96%	26.57	78%	26.78	91%	26.82	100%	26.92
ADYEKHKVYACEVTHQGLSSPVTKSFNRGEC	Ambiguous between K5 and K7	Ambiguous between Lc-K188 and Lc-K190	4439	43%	26.54	20%	26.61	40%	26.83				
AKGQPREPQVYTLPPSREEMTK	K2	HC-K343	3497.66	81%	28.36	86%	27.41	86%	27.54	86%	27.49	81%	27.64
AKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEW ESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFS CSVMHEALHNHYTQKSLSLSPG	Ambiguous between K71 and K76	Ambiguous between HC-K412 and K417	13129.29	50%	35.09	51%	35.4	41%	35.22	44%	35.28	51%	35.12
DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKA PK	K42	Lc-K42	5775.77	73%	30.64	66%	30.79	73%	30.98	80%	30.85	77%	30.97
DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKA PKLLIYSASFLYSGVPSRFSGSRSGTDFTLTISSLQPEDFATYYCQ QHYTTPPTFGQGTKVEIKRTVAAPSVFIFPPSDEQLK	K45	Lc-K45	14728.24			38%	36.48	44%	36.26	28%	36.49	43%	36.32
DSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTK	K19	Lc-K188	5129.42	57%	30.21	60%	30.31	57%	30.54	51%	30.39	46%	30.6
DTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAK	K26	HC-K277	5512.56	26%	37.22	41%	37.53	49%	37.52	41%	37.64	41%	37.67
DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSS SLGTQTYICNVNHKPSNTKVDK	Ambiguous between K58 and K63	Ambiguous between HC-K208 and HC-K213	7953.84	63%	36.69	68%	36.65	63%	37	63%	37.24	71%	37.15
DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSS SLGTQTYICNVNHKPSNTKVDKKVEPK	Ambiguous between K58, K63, and K66	Ambiguous between HC-K208, HC-K213, and HC-K216	8535.19	46%	35.17	46%	35.34	49%	35.28	54%	35.46	39%	35.72
EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGK	K30	HC-K30	5480.7	57%	35.85	62%	36.2	43%	35.71	52%	36.36	57%	36.43
EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGK GLEWVARIYPTNGYTRYADSVK	K30	HC-K30	8020.98	42%	36.27	59%	36.53	64%	36.51	64%	36.63	59%	36.61
EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGK GLEWVARIYPTNGYTRYADSVK	K43	HC-K43	8020.98	61%	34.89	63%	34.87	63%	34.95	70%	35.1	64%	35.02
GFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSK	K22	HC-K395	5354.39	55%	39.99	74%	40.12	63%	40.18	58%	40.34	74%	40.36

Table 4.1 Continued

GFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDK	K22	HC-K395	5910.71	72%	39.02	72%	39.24	70%	39.41	74%	39.35	70%	39.45
GFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDK SRWQQGNVFSCSVMHEALHNHYTQK	K44	HC-K417	8879.08	74%	36.14	74%	36.33	78%	36.44	53%	36.65	72%	36.57
GLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYLQMNSL RAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPS VFPLAPSSK	K22	HC-K65	11161.28	49%	37.76	73%	37.93	62%	38.01	63%	38.02	54%	38.17
GLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYLQMNSL RAEDTAVYYCSRWGGDGFY AMDYWGQGTLVTVSSASTKGPSVFPLAPSSK	K33	HC-K76	11161.28	63%	37.71	57%	38	63%	38.01	66%	38.02	64%	38.08
GQPREPQVYTLPPSREEMTKNQVSLTCLVK	K20	HC-K363	4384.12	69%	32.9	72%	32.98	55%	33.13	86%	32.28	66%	32.53
GQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESN GQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSV MHEALHN HYTQKSLSLSPG	K52	HC-K395	12930.16	52%	34.6	64%	34.71	45%	34.84	70%	34.83	60%	34.96
GRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAM DYWGQGTLVTVSSASTKGPSVFPLAPSSK	K11	HC-K76	8621	89%	38.84	80%	39.09	83%	39.11	80%	39.17	79%	38.77
GRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAM DYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVK	K71	HC-K136	9866.64	50%	38.41	54%	38.53	57%	38.69	46%	38.99	51%	38.75
GRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAM DYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVK	K59	HC-K124	9866.64	44%	38.41	54%	38.53	40%	38.58	45%	38.99	49%	38.72
HKVYACEVTHQGLSSPVTK	K2	Lc-K190	3039.42	89%	27.82	83%	27.8	89%	28.02	94%	27.96	94%	28.61
HKVYACEVTHQGLSSPVTKSFNRGEC	K19	Lc-K207	3832.73	76%	28.96	76%	28.98	64%	29.16	72%	29.18	80%	29.35
LLIYSASFLYSGVPSRFSGSRSGTDFTLTISSLQPEDFATYYCQQH YTTPPTFGQGTKVEIKRTVAAPSVFIFPPSDEQLK	K62	Lc-K107	9926.85	66%	37.95	65%	37.96	69%	38.14	64%	38.21	64%	38.37
LTVDKSRWQQGNVFSCSVMHEALHNHYTQK	K5	HC-K417	4499.06	97%	30.93	83%	30.97	97%	31.32	97%	31.15	93%	31.31
NTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLV TVSSASTKGPSVFPLAPSSK	K48	HC-K124	7457.41	39%	40.46	34%	40.62	46%	41.06	27%	40.82	32%	40.87
NTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLV TVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVK	K60	HC-K136	8703.05	45%	39.67	43%	39.86	60%	39.94				
SCDKTHTCPPCPAPELLGGPSVFLFPPK	K4	HC-K225	3893.79	74%	38.45	63%	38.88	74%	38.87	56%	38.82	74%	38.95
SCDKTHTCPPCPAPELLGGPSVFLFPPKPK	K4	HC-K225	4118.93	55%	36.3	59%	36.79	41%	36.79	59%	36.76	55%	37.18
SGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDS KDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTK	K62	Lc-K188	9837.7	59%	30.89	40%	31.02	60%	31.16	55%	31.11	56%	31.19
SGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDS KDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNR	K62	Lc-K188	10631.02			43%	30.83	43%	30.98	52%	30.92	46%	30.95

Table 4.1 Continued

GEC													
THTCPPCPAPELLGGPSVFLFPPKPK	K24	HC-K249	3685.77	76%	39.52	76%	39.75	76%	39.96	68%	39.97	80%	39.82
THTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDV SHEDPEVKFNWYVDGVEVHNAK	K52	HC-K277	8223.96	48%	37.15	46%	37.4	23%	37.59	46%	37.54	35%	37.55
VQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYE KHK	K43	Lc-K188	5987.79	80%	28.6	77%	28.64	73%	29.45	71%	28.74	84%	28.81
VQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYE KHKVYACEVTHQGLSSPVTK	K43	Lc-K188	7787.67	79%	29.05	69%	29.27	74%	29.39	71%	29.26	75%	29.38
VSNKALPAPIEK	K4	HC-K329	2222.1	91%	32.79	82%	32.16	100%	32.38	91%	32.41	82%	33.13
VYACEVTHQGLSSPVTKSFNRGEC	K17	Lc-K207	3567.58					78%	32.28	48%	32.23	61%	32.41

Table 4.2: List of payload-containing peptides identified with ProSight PD for EThcD replicates. For each peptide, the residue to which the payload was localized, the theoretical mass, as well as the sequence coverage and retention time for each replicate are listed. Some replicate entries are blank in the case that a peptide was not identified in all five technical replicates. Manually identified peptides from Figure 4.9 and Figure 4.10 are not included.

Annotated Sequence	Payload Localization (residue number on peptide sequence)	Payload Localization (residue number on protein sequence)	Theo. Mass [Da]	Replicate 1		Replicate 2		Replicate 3		Replicate 4		Replicate 5	
				SC	RT	SC	RT	SC	RT	SC	RT	SC	RT
ADYEKHK	K5	Lc-K188	1845.79	100%	28.03	100%	27.99	100%	28.17	83%	28.07	100%	28.35
ADYEKHKVYACEVTHQGLSSPVTK	K5	Lc-K188	3645.68	78%	27.18	83%	27.19	83%	27.26	83%	27.26	83%	27.37
AKGQPREPQVYTLPPSREEMTK	K2	HC-K343	3497.66	81%	28.99	76%	27.93	81%	28.09	81%	28.97	76%	28.05
AKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEW ESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFS CSVMHEALHNHYTQKSLSLSPG	Ambiguous between K71 and K76	Ambiguous between HC-K412 and K417	13129.29	34%	35.37	33%	35.44	32%	35.4	35%	35.46	29%	35.5
AKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEW ESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFS CSVMHEALHNHYTQKSLSLSPG	K54	HC-K395	13129.29	28%	34.78			42%	34.28	26%	34.65		
ALPAPIEKTISK	K8	HC-K337	2223.12	82%	34.94	73%	35.04	73%	34.33	73%	35.06	82%	34.53
DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKA PK	K42	Lc-K42	5775.77	77%	31.26	68%	31.37	61%	31.64	71%	31.02	73%	31.42
DSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTK	K19	Lc-K188	5129.42			54%	30.75	76%	30.9			57%	30.89
DTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAK	K26	HC-K277	5512.56	41%	37.83	49%	37.74	64%	37.68	54%	37.82	51%	37.84
DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSS SLGTQTYICNVNHKPSNTKVDK	Ambiguous between K58 and K63	Ambiguous between HC-K208 and HC-K213	7953.84	31%	37.43	52%	37.27	62%	37.34	43%	37.14	46%	37.26
DYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSS SLGTQTYICNVNHKPSNTKVDKKVEPK	Ambiguous between K58 and K63	Ambiguous between HC-K208 and HC-K213	8535.19	69%	35.67	51%	35.46	61%	35.61	66%	35.5	67%	35.56
EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGK	K30	HC-K30	5480.7	60%	36.55	83%	36.42	86%	36.43	74%	36.39	74%	36.6
EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGK GLEWVARIYPTNGYTRYADSVK	K30	HC-K30	8020.98	59%	36.78	66%	36.68	64%	36.6	56%	36.74	69%	36.8
EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGK GLEWVARIYPTNGYTRYADSVK	K43	HC-K43	8020.98	61%	35.43	58%	35.28	67%	35.24	58%	35.44	64%	35.38
GFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSK	K22	HC-K395	5354.39	76%	40.42	42%	40.5			76%	40.48	68%	40.53
GFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDK	K22	HC-K395	5910.71	47%	39.57	42%	39.6	47%	39.48	44%	39.93	47%	39.64
GFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDK SRWQQGNVFSCSVMHEALHNHYTQK	K44	HC-K417	8879.08	60%	36.55	72%	36.57	68%	36.45	69%	36.54	69%	36.67

Table 4.2 Continued

GLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSK	K22	HC-K65	11161.28	51%	38.27	52%	38.27	73%	38.08	53%	38.34	61%	38.31
GLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSK	K33	HC-K76	11161.28	59%	38.33	72%	38.25	71%	38.17	61%	38.32	62%	38.41
GQPREPQVYTLPPSREEMTKNQVSLTCLVK	K20	HC-K363	4384.12	66%	33.03	59%	32.93	62%	32.75	66%	32.78	59%	32.86
GQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQK	K52	HC-K395	12288.82	42%	34.76	47%	34.83	47%	34.76	53%	34.94	47%	34.91
GQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG	K52	HC-K395	12930.16	40%	35.07	49%	35.01	36%	35.13	39%	35.1	54%	35.05
GRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSK	K11	HC-K76	8621	76%	39.2	84%	39.3	86%	38.97	86%	39.03	84%	39.27
GRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVK	K59	HC-K124	9866.64	32%	39.19	55%	38.93	58%	38.8	44%	38.89	68%	38.92
GRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVK	K71	HC-K136	9866.64	70%	38.87	56%	38.93	70%	38.8	68%	38.82	77%	38.89
HKVYACEVTHQGLSSPVTK	K2	Lc-K190	3039.42	100%	28.58	89%	28.49	100%	28.4	89%	28.43	94%	28.5
HKVYACEVTHQGLSSPVTKSFNRGEC	K19	Lc-K207	3832.73	52%	29.96			64%	29.53	72%	29.46	32%	29.58
LLIYSASFLYSGVPSRFSGSRSGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAPSVFIFPPSDEQLK	K62	Lc-K107	9926.85	59%	38.24	64%	38.19	76%	38.05	68%	38.32	60%	38.65
LTVDKSRWQQGNVFSCSVMHEALHNHYTQK	K5	HC-K417	4499.06	90%	31.51	90%	31.49	93%	31.53	97%	31.69	83%	31.51
NTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSK	K48	HC-K124	7457.41	46%	40.93	71%	40.84	63%	40.77	66%	40.88	59%	41.07
NTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVK	K60	HC-K136	8703.05	75%	40.18	78%	40.11	75%	40.14	80%	40.15	22%	40.19
PKDTLMISRTPEVTCVVVDVSHEDPEVK	K2	HC-K251	4078.93	37%	33.63	48%	33.6	56%	33.59	30%	33.75		
SCDKTHTCPPCPAPELLGGPSVFLFPPK	K4	HC-K225	3893.79	59%	39.26	59%	38.94	74%	38.83	74%	39.23	56%	39.05
SCDKTHTCPPCPAPELLGGPSVFLFPPKPK	K4	HC-K225	4118.93	76%	36.78	76%	36.66	72%	36.66	72%	36.65	83%	36.72
THTCPPCPAPELLGGPSVFLFPPKPK	K24	HC-K249	3685.77	76%	39.9	76%	39.96	88%	39.87	76%	39.97	80%	39.91
THTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVK	Ambiguous between K24 and K26	Ambiguous between HC-K249 and HC-K251	6565.18	45%	37.62	37%	37.63	28%	37.76	49%	37.73	41%	37.72
THTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVS	K52	HC-K277	8223.96	42%	37.6	49%	37.49	45%	37.41	51%	37.46		

Table 4.2 Continued

HEDPEVKFNWYVDGVEVHNAK													
THTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVS HEDPEVKFNWYVDGVEVHNAK	Ambiguous between K24 and K26	Ambiguous between HC-K249 and HC-K251	8223.96	48%	37.52	45%	37.5			42%	37.59		
VQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEK HK	K43	Lc-K188	5987.79	91%	29.18	89%	29.24	93%	29.14	84%	29.13	86%	29.24
VSNKALPAPIEK	K4	HC-K329	2222.1	82%	32.71	82%	32.73	82%	32.65	91%	32.71	91%	32.73
VYACEVTHQGLSSPVTKSFNRGEC	K17	Lc-K207	3567.58	26%	32.62	30%	32.63	30%	32.56				

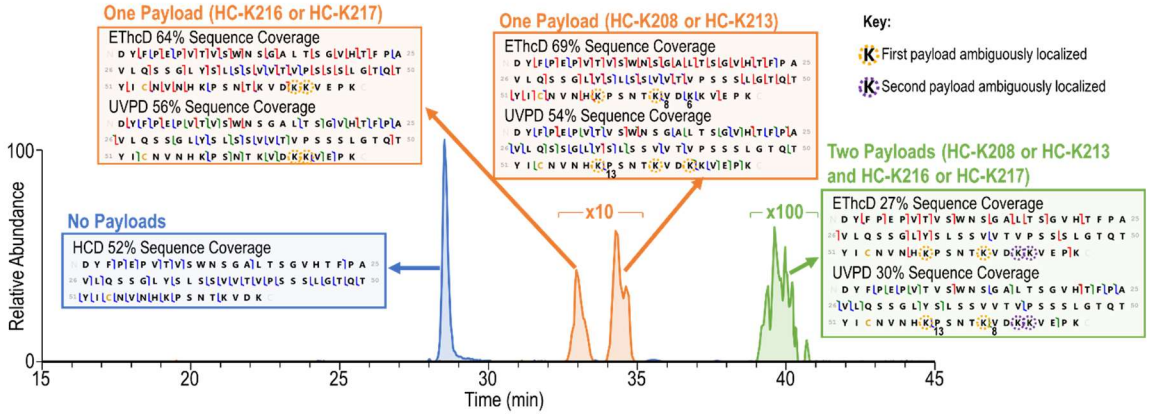


Figure 4.9: Extracted ion chromatograms (EICs) for a HC peptide (71 amino acids, 6+) containing D151 through K221 with zero (*m/z* 1264.82), one (*m/z* 1424.21) or two (*m/z* 1583.78) payloads. Sequence coverage plots are included for each chromatographic peak observed in the EIC. K58, K63, K66, and K67 of these sequences correspond to K208, K213, K216, and K217 of the HC. Fragment ions discussed in the text are indicated with number labels.

To identify and characterize peptides with multiple conjugated payloads, both the chromatographic data and the MS/MS spectra are examined. **Figure 4.9** displays the extracted ion chromatogram for a peptide containing 71 amino acids with no payloads (blue trace), with one payload (orange traces), and with two payloads (green trace). The elution order of the three species offers one important feature. The payload is expected to increase the hydrophobicity of the peptide, resulting in a greater retention time as the number of payloads increases, as reflected in **Figure 4.9**. An HCD sequence coverage map is included for the peptide with no payloads. Given the high abundance and simplicity of this peptide, the HCD spectrum adequately characterizes it. Neither EThcD nor UVPD was triggered because the *m/z* 547.22 reporter ion was not generated by HCD. For the species with a single payload, two chromatographically resolved peaks were observed and attributed to two different conjugation states. Given the proximity of the

four different lysine residues in this peptide (HC residues K208, K213, K216, and K217), it is challenging to distinguish the exact conjugation sites. The earlier eluting chromatographic peak (t_r 33 min) is well characterized by both UVPD and EThcD, and in both cases there are fragment ions originating from backbone cleavages between HC-K213 and HC-K216/HC-K217 indicating that the payload must be conjugated to HC-K216 or HC-K217. The site localization is particularly compelling based on the UVPD sequence coverage map for which there are a several backbone cleavages between HC-K213 and HC-K216/HC-K217. This earlier eluting peak was not consistently identified with Prosight PD, but manual examination of chromatographic data revealed it was present in every replicate. For the later eluting chromatographic peak (t_r 34.5 min), there are only a handful of fragment ions originating from backbone cleavages spanning the four lysine residues. While a single fragment ion without a payload (y_{13}) identified by UVPD seems to indicate that HC-K213 does not contain the payload, a few additional fragment ions without payloads (y_6 *and* y_8) generated by EThcD seem to localize the payload to either HC-K208 or HC-K213. While these few fragment ions, which only appear in some replicates, do not allow confident localization, the significant retention time shift in combination with the differences in sequence coverage maps between the two isomeric peptides (t_r 33 min versus 34.5 min) implies that the payload is conjugated to a different site for the second peptide, which supports payload localization to HC-K208 or HC-K213.

Finally, the peptide with multiple payloads was examined. Based on the green chromatographic profile in **Figure 4.9,** the abundance of the peptide is low, diminishing the quality of the MS/MS spectra (and reducing the probability that the peptide would be targeted in any conventional data-dependent method). While the doubly conjugated peptide was only identified in two out of five replicates for both UVPD and EThcD based

on automated ProSight PD analysis, the same precursor was targeted for UVPD and EThcD in every replicate with the HCD-triggered workflow, and the resulting MS/MS spectra could be manually deconvoluted in FreeStyle and characterized with ProSight Lite. In the sequence maps of the double-conjugated peptide in **Figure 4.9,** the sequence coverage is sufficient to identify the peptide but is inadequate to localize the payloads. Even in the case of UVPD, which yielded a higher sequence coverage of 30% compared to 27% obtained with EThcD, the overall degree of characterization is low. UVPD did however, yield two informative fragment ions, y_{13} and x_8, which contained single payload conjugations. By combining the MS/MS data from the peptide with two payloads and the peptide with one payload, it can be reasoned that one of the conjugation sites for the doubly modified species resides at HC-K208 or HC-K213 and the second is HC-K216 or HC-K217. **Figure 4.9** also demonstrates the power of the HCD-triggered method to optimize data acquisition time for the most critical peptides. If only HCD spectra had been acquired to investigate the peptides, characterization of the sequence and payload sites would have been impeded. If only high resolution EThcD or UVPD data had been collected, this peptide might have been missed because of its low abundance. By utilizing the HCD-triggered method, even low abundance payload-containing peptides are targeted for characterization.

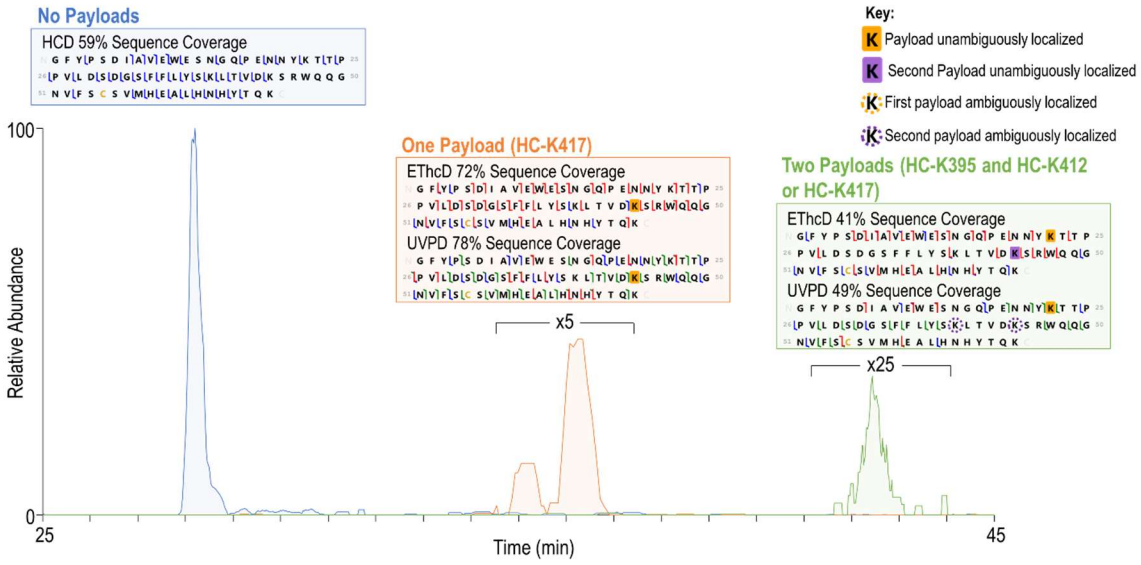


Figure 4.10: EICs for the 7+ charge state of a 69 amino acid long HC peptide containing G374 through K442 with zero (*m/z* 1132.59), one (*m/z* 1270.31) or two (*m/z* 1406.93) payload conjugations. Sequence coverage plots are included for each chromatographic peak observed in the EIC. The payload localization sites are highlighted by a square when unambiguous and circled when ambiguous. K22, K39, and K44 on these maps correspond to K395, K412, and K417 on the HC.

Figure 4.10 displays the results for the other peptide for which two conjugation states (one payload or two payloads) were identified. In this case, only one conjugation site for the peptide containing a single payload was identified. As displayed by the EIC and sequence coverage plots highlighted in orange in **Figure 4.10**, the payload was unambiguously localized to HC-K417 by UVPD. A secondary peak eluting earlier than the identified peptide is observed in the chromatogram; however the abundance was too low to trigger any MS^2 scans. Additional peptides containing HC-K395 were also identified, as listed in **Table 4.1** and **Table 4.2**, which resulted from LysC cleavage at HC-K412 and HC-K417 and lead to unambiguous identification of the payload at HC-K395. It is likely that the presence of the payload on HC-K417 hindered the LysC digestion and aided in the generation of the larger peptide identified in **Figure 4.10**. The

fact that other peptides were identified with a payload conjugated to HC-K395 aids in the confidence that a multiply conjugated species could exist for this peptide. The chromatographic peak corresponding to the doubly-conjugated peptide is shaded in green in **Figure 4.10** along with the sequence maps produced by EThcD (41% coverage) and UVPD (49% coverage). The first payload was unambiguously localized to HC-K395 with both UVPD and EThcD based on payload bracketing fragment ions. While there were no backbone cleavages between HC-K412 and HC-K417 for UVPD, one fragment ion (z_{27}) containing a single payload originating from backbone cleavage between HC-K412 and HC-K417, was identified by EThcD. Through the combination of information about the peptides containing a single payload and the supplemental insight from the MS/MS spectra of the peptide containing two payloads, we can identify two conjugation sites of the bis-modified peptide in **Figure 4.10** as HC-K395 and HC-K417.

4.4.3 Complete Characterization of Payload Binding Sites

In summary, high-quality identification of peptides across the entire antibody sequence was achieved via a middle-down HCD-triggered MS/MS strategy. **Figure 4.11** displays the comprehensive map of all the confirmed payload sites, including 6 unambiguous LC sites, 16 unambiguous HC sites, and 5 ambiguous HC sites in which the payload could only be localized to a span of adjacent lysine residues. **Figures 4.12** and **4.13** display maps of the peptides identified, helping to visualize the redundancies present in the payload-containing peptides identified. While the total number of conjugation sites identified was lower than the number mapped in a recent bottom-up study,[30] the results here represent a step towards unravelling the heterogeneity of lysine-conjugated ADCs. By examining the payload conjugation sites in the context of larger peptides, an improved understanding of the most prevalent conjugation sites and interplay among those sites is

obtained. The focus of this study has been the HC, given the inherent increased complexity imparted by its size, and the larger number of payload containing peptides identified relative to the LC. As displayed in **Figure 4.11**, six payload conjugation sites were identified along the LC. The peptides leading to these six site localizations on the LC are listed in **Tables 4.1** and **4.2**. While no multiply conjugated species were identified on the LC, the contribution of the LC towards the overall ADC heterogeneity should be considered. **Figure 4.11** also highlights the regions in which the middle-down HCD-triggered MS/MS method revealed that multiple payloads were conjugated to one HC peptide. By identifying and characterizing peptides containing multiple lysine conjugation sites, a better understanding of regions containing high levels of ADC conjugation that might cause changes to antibody structure and function is gained.

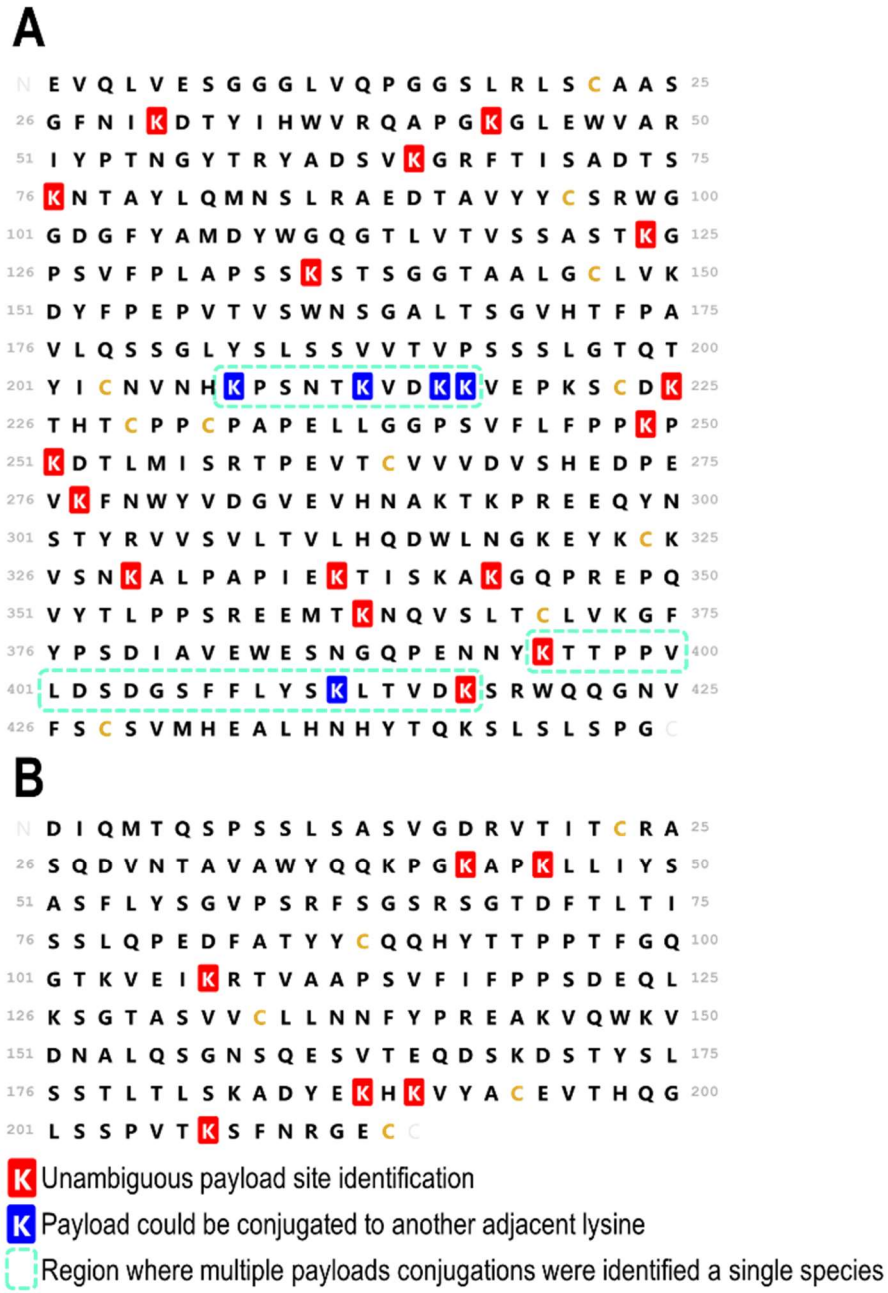


Figure 4.11: Sequence maps of the (A) HC and (B) LC of T-DM1 displaying the locations of the payloads, including those that were unambiguously localized (shaded in red) and those that remain ambiguous with adjacent or nearby lysine residues (shaded in blue).

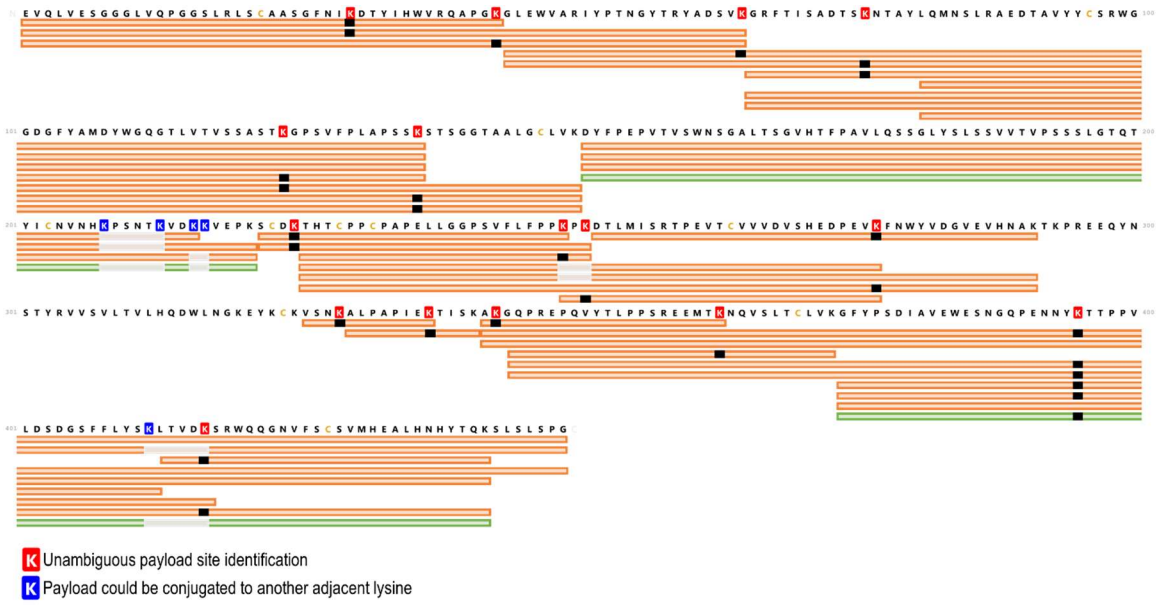


Figure 4.12: Peptide map displaying the global sequence coverage for the heavy chain. The same legend used in **Figure 4.11** was retained for the heavy chain sequence. Peptides displayed in orange contain a single payload conjugation and peptides displayed in green contain two payload conjugations. Unambiguously localized payload conjugations sites are shown in black on for each peptide, and payloads that could be localized two multiple sites are shown in gray.

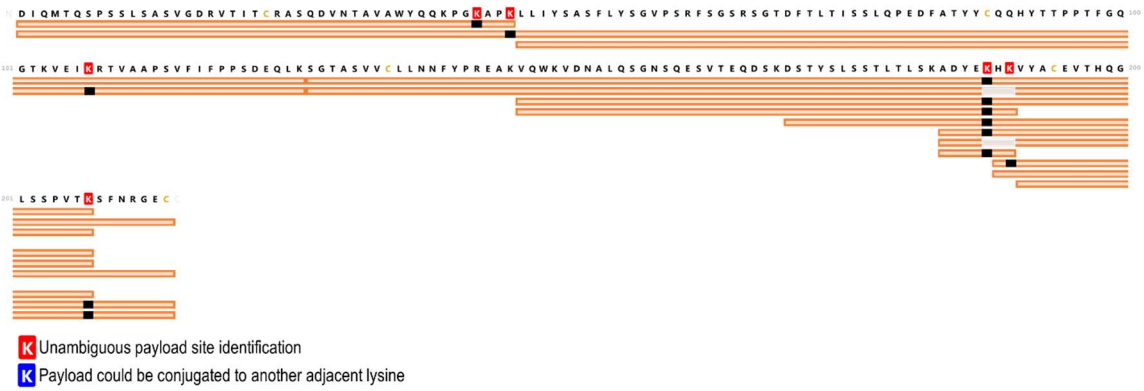


Figure 4.13: Peptide map displaying the global sequence coverage for the light chain. The same legend used in **Figure 4.11** was retained for the heavy chain sequence. Peptides displayed in orange contain a single payload conjugation. Unambiguously localized payload conjugations sites are shown in black on for each peptide, and payloads that could be localized two multiple sites are shown in gray.

Typical bottom-up methods report the identification of 38-54 conjugation sites on lysine-linked ADCs out of 92 possible sites, with one study reporting the identification of 82 sites.[31–34]For trastuzumab, there are 92 possible conjugation sites, including 31 and 13 lysine residues on each HC and LC, respectively, plus an additional conjugation site at each N-terminal amine group (a total of 4). Out of those 92 conjugation sites, 44 were unambiguously identified based on the present middle-down approach, making these results comparable to most bottom-up studies. The ability to pinpoint locations where multiple payloads are conjugated in close proximity represents a level of information that has yet to be achieved by bottom-up or middle-down studies of ADCs. The information could be complemented by middle-up MS1 data generated from IdeS-digested ADCs,[35,49] which would reveal the number of payloads bound to each subunit but not the locations of those payloads. The methods presented here have the potential to be used for relative quantification between different samples. For example, different batches of ADCs could be evaluated to determine if the relative abundance of individual conjugation sites differs,

as recently demonstrated using bottom-up methods.[33] While the limited digestion strategy demonstrated here represents one promising application of the HCD-triggered method for characterization of T-DM1, it could also be utilized to improve other middle-down, as well as bottom-up workflows for T-DM1 and other ADCs that also generate signature fragment ions upon collisional activation and could benefit from a more targeted approach.

4.5 Conclusions

The HCD-triggered MS/MS methods presented here represent a promising improvement in the characterization of highly heterogeneous lysine-linked ADCs. By applying the HCD-triggered methods to large peptides generated from a middle-down proteolysis method, high-quality data were generated for large payload-containing peptides. The combination of results from EThcD and UVPD was particularly helpful in the case of large peptides containing multiple potential payload sites; in some cases one MS/MS method outperformed the other and in other cases only the combination of EThcD and UVPD yielded sufficient confidence to localize payloads. A key finding was that higher sequence coverage for large peptides did not always translate to complete localization of payloads, as localization often hinged on key regions between adjacent lysine residues. The generation of payload-containing fragment ions, enabled by EThcD and UVPD, often proved critical to comprehensive characterization. Overall, 44 sites out of 92 were unambiguously identified through the characterization of 48 payload-containing peptides of varying molecular weights. Additionally, two peptides containing multiple conjugations were identified and their characterization was described in detail. The ability to identify multiply conjugated species, as enabled by these methods, proved

crucial for deciphering heterogeneous lysine-linked ADCs and offers a compelling approach for more detailed characterization of biotherapeutics.

4.6 REFERENCES

(1) Joubert, N.; Beck, A.; Dumontet, C.; Denevault-Sabourin, C. Antibody–Drug Conjugates: The Last Decade. *Pharmaceuticals* **2020**, *13*, 245.

(2) Fu, Z.; Li, S.; Han, S.; Shi, C.; Zhang, Y. Antibody Drug Conjugate: The "Biological Missile" for Targeted Cancer Therapy. *Sig Transduct Target Ther* **2022**, *7*, 1–25.

(3) Tong, J. T. W.; Harris, P. W. R.; Brimble, M. A.; Kavianinia, I. An Insight into FDA Approved Antibody-Drug Conjugates for Cancer Therapy. *Molecules* **2021**, *26*, 5847.

(4) Watts, E.; Williams, J. D.; Miesbauer, L. J.; Bruncko, M.; Brodbelt, J. S. Comprehensive Middle-Down Mass Spectrometry Characterization of an Antibody–Drug Conjugate by Combined Ion Activation Methods. *Anal. Chem.* **2020**, *92*, 9790–9798.

(5) Chen, B.; Lin, Z.; Zhu, Y.; Jin, Y.; Larson, E.; Xu, Q.; Fu, C.; Zhang, Z.; Zhang, Q.; Pritts, W. A.; Ge, Y. Middle-Down Multi-Attribute Analysis of Antibody-Drug Conjugates with Electron Transfer Dissociation. *Anal. Chem.* **2019**, *91*, 11661–11669.

(6) Beck, A.; Goetsch, L.; Dumontet, C.; Corvaïa, N. Strategies and Challenges for the next Generation of Antibody–Drug Conjugates. *Nat. Rev. Drug Discov.* **2017**, *16*, 315–337.

(7) Friese, O. V.; Smith, J. N.; Brown, P. W.; Rouse, J. C. Practical Approaches for Overcoming Challenges in Heightened Characterization of Antibody-Drug Conjugates with New Methodologies and Ultrahigh-Resolution Mass Spectrometry. *mAbs* **2018**, *10*, 335–345.

(8) Huang, R. Y.-C.; Chen, G. Characterization of Antibody–Drug Conjugates by Mass Spectrometry: Advances and Future Trends. *Drug Discov. Today* **2016**, *21*, 850–855.

(9) Nagornov, K. O.; Gasilova, N.; Kozhinov, A. N.; Virta, P.; Holm, P.; Menin, L.; Nesatyy, V. J.; Tsybin, Y. O. Drug-to-Antibody Ratio Estimation via Proteoform Peak Integration in the Analysis of Antibody–Oligonucleotide Conjugates with Orbitrap Fourier Transform Mass Spectrometry. *Anal. Chem.* **2021**, *93*, 12930–12937.

(10) Huang, Y.; Mou, S.; Wang, Y.; Mu, R.; Liang, M.; Rosenbaum, A. I. Characterization of Antibody–Drug Conjugate Pharmacokinetics and in Vivo

Biotransformation Using Quantitative Intact LC-HRMS and Surrogate Analyte LC-MRM. *Anal. Chem.* **2021**, *93*, 6135–6144.

(11) Jones, J.; Pack, L.; Hunter, J. H.; Valliere-Douglass, J. F. Native Size-Exclusion Chromatography-Mass Spectrometry: Suitability for Antibody–Drug Conjugate Drug-to-Antibody Ratio Quantitation across a Range of Chemotypes and Drug-Loading Levels. *mAbs* **2020**, *12*, 1682895.

(12) He, J.; Su, D.; Ng, C.; Liu, L.; Yu, S.-F.; Pillow, T. H.; Del Rosario, G.; Darwish, M.; Lee, B.-C.; Ohri, R.; Zhou, H.; Wang, X.; Lu, J.; Kaur, S.; Xu, K. High-Resolution Accurate-Mass Mass Spectrometry Enabling In-Depth Characterization of in Vivo Biotransformations for Intact Antibody-Drug Conjugates. *Anal. Chem.* **2017**, *89*, 5476–5483.

(13) Sarrut, M.; Corgier, A.; Fekete, S.; Guillarme, D.; Lascoux, D.; Janin-Bussat, M.-C.; Beck, A.; Heinisch, S. Analysis of Antibody-Drug Conjugates by Comprehensive on-Line Two-Dimensional Hydrophobic Interaction Chromatography x Reversed Phase Liquid Chromatography Hyphenated to High Resolution Mass Spectrometry. I − Optimization of Separation Conditions. *J. Chromatogr. B* **2016**, *1032*, 103–111.

(14) Pacholarz, K. J.; Barran, P. E. Use of a Charge Reducing Agent to Enable Intact Mass Analysis of Cysteine-Linked Antibody-Drug-Conjugates by Native Mass Spectrometry. *EuPA Open Proteom.* **2016**, *11*, 23–27.

(15) Redman, E. A.; Mellors, J. S.; Starkey, J. A.; Ramsey, J. M. Characterization of Intact Antibody Drug Conjugate Variants Using Microfluidic Capillary Electrophoresis–Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 2220–2226.

(16) Birdsall, R. E.; Shion, H.; Kotch, F. W.; Xu, A.; Porter, T. J.; Chen, W. A Rapid On-Line Method for Mass Spectrometric Confirmation of a Cysteine-Conjugated Antibody-Drug-Conjugate Structure Using Multidimensional Chromatography. *mAbs* **2015**, *7*, 1036–1044.

(17) Dyachenko, A.; Wang, G.; Belov, M.; Makarov, A.; de Jong, R. N.; van den Bremer, E. T. J.; Parren, P. W. H. I.; Heck, A. J. R. Tandem Native Mass-Spectrometry on Antibody–Drug Conjugates and Submillion Da Antibody–Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. *Anal. Chem.* **2015**, *87*, 6095–6102.

(18) Chen, T.-H.; Yang, Y.; Zhang, Z.; Fu, C.; Zhang, Q.; Williams, J. D.; Wirth, M. J. Native Reversed-Phase Liquid Chromatography: A Technique for LCMS of Intact Antibody–Drug Conjugates. *Anal. Chem.* **2019**, *91*, 2805–2812.

(19) Dyachenko, A.; Wang, G.; Belov, M.; Makarov, A.; de Jong, R. N.; van den Bremer, E. T. J.; Parren, P. W. H. I.; Heck, A. J. R. Tandem Native Mass-Spectrometry on Antibody–Drug Conjugates and Submillion Da Antibody–

Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. *Anal. Chem.* **2015**, *87*, 6095–6102.

(20) Larson, E. J.; Roberts, D. S.; Melby, J. A.; Buck, K. M.; Zhu, Y.; Zhou, S.; Han, L.; Zhang, Q.; Ge, Y. High-Throughput Multi-Attribute Analysis of Antibody-Drug Conjugates Enabled by Trapped Ion Mobility Spectrometry and Top-Down Mass Spectrometry. *Anal. Chem.* **2021**, *93*, 10013–10021.

(21) Deslignière, E.; Ehkirch, A.; Duivelshof, B. L.; Toftevall, H.; Sjögren, J.; Guillarme, D.; D'Atri, V.; Beck, A.; Hernandez-Alba, O.; Cianférani, S. State-of-the-Art Native Mass Spectrometry and Ion Mobility Methods to Monitor Homogeneous Site-Specific Antibody-Drug Conjugates Synthesis. *Pharmaceuticals* **2021**, *14*, 498.

(22) Botzanowski, T.; Erb, S.; Hernandez-Alba, O.; Ehkirch, A.; Colas, O.; Wagner-Rousset, E.; Rabuka, D.; Beck, A.; Drake, P. M.; Cianférani, S. Insights from Native Mass Spectrometry Approaches for Top- and Middle- Level Characterization of Site-Specific Antibody-Drug Conjugates. *mAbs* **2017**, *9*, 801–811.

(23) Huang, R. Y.-C.; Deyanova, E. G.; Passmore, D.; Rangan, V.; Deshpande, S.; Tymiak, A. A.; Chen, G. Utility of Ion Mobility Mass Spectrometry for Drug-to-Antibody Ratio Measurements in Antibody-Drug Conjugates. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1791–1794.

(24) Cho, E.; Mayhugh, B. M.; Srinivasan, J. M.; Sacha, G. A.; Nail, S. L.; Topp, E. M. Stability of Antibody Drug Conjugate Formulations Evaluated Using Solid-State Hydrogen-Deuterium Exchange Mass Spectrometry. *J. Pharm. Sci.* **2021**, *110*, 2379–2385.

(25) Pan, L. Y.; Salas-Solano, O.; Valliere-Douglass, J. F. Antibody Structural Integrity of Site-Specific Antibody-Drug Conjugates Investigated by Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* **2015**, *87*, 5669–5676.

(26) Pan, L. Y.; Salas-Solano, O.; Valliere-Douglass, J. F. Conformation and Dynamics of Interchain Cysteine-Linked Antibody-Drug Conjugates as Revealed by Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* **2014**, *86*, 2657–2664.

(27) Debaene, F.; Bœuf, A.; Wagner-Rousset, E.; Colas, O.; Ayoub, D.; Corvaïa, N.; Van Dorsselaer, A.; Beck, A.; Cianférani, S. Innovative Native MS Methodologies for Antibody Drug Conjugate Characterization: High Resolution Native MS and IM-MS for Average DAR and DAR Distribution Assessment. *Anal. Chem.* **2014**, *86*, 10674–10683.

(28) Larson, E. J.; Zhu, Y.; Wu, Z.; Chen, B.; Zhang, Z.; Zhou, S.; Han, L.; Zhang, Q.; Ge, Y. Rapid Analysis of Reduced Antibody Drug Conjugate by Online LC-

MS/MS with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2020**, *92*, 15096–15103.

(29) Hernandez-Alba, O.; Houel, S.; Hessmann, S.; Erb, S.; Rabuka, D.; Huguet, R.; Josephs, J.; Beck, A.; Drake, P. M.; Cianférani, S. A Case Study to Identify the Drug Conjugation Site of a Site-Specific Antibody-Drug-Conjugate Using Middle-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2419–2429.

(30) Zhu, X.; Huo, S.; Xue, C.; An, B.; Qu, J. Current LC-MS-Based Strategies for Characterization and Quantification of Antibody-Drug Conjugates. *J Pharm Anal.* **2020**, *10*, 209–220.

(31) Arlotta, K. J.; Gandhi, A. V.; Chen, H.-N.; Nervig, C. S.; Carpenter, J. F.; Owen, S. C. In-Depth Comparison of Lysine-Based Antibody-Drug Conjugates Prepared on Solid Support Versus in Solution. *Antibodies* **2018**, *7*, 6.

(32) Chen, L.; Wang, L.; Shion, H.; Yu, C.; Yu, Y. Q.; Zhu, L.; Li, M.; Chen, W.; Gao, K. In-Depth Structural Characterization of Kadcyla® (Ado-Trastuzumab Emtansine) and Its Biosimilar Candidate. *mAbs* **2016**, *8*, 1210–1223.

(33) Sang, H.; Lu, G.; Liu, Y.; Hu, Q.; Xing, W.; Cui, D.; Zhou, F.; Zhang, J.; Hao, H.; Wang, G.; Ye, H. Conjugation Site Analysis of Antibody-Drug-Conjugates (ADCs) by Signature Ion Fingerprinting and Normalized Area Quantitation Approach Using Nano-Liquid Chromatography Coupled to High Resolution Mass Spectrometry. *Anal. Chim. Acta* **2017**, *955*, 67–78.

(34) Wu, G.; Gao, Y.; Liu, D.; Tan, X.; Hu, L.; Qiu, Z.; Liu, J.; He, H.; Liu, Y. Study on the Heterogeneity of T-DM1 and the Analysis of the Unconjugated Linker Structure under a Stable Conjugation Process. *ACS Omega* **2019**, *4*, 8834–8845.

(35) Cotham, V. C.; Horton, A. P.; Lee, J.; Georgiou, G.; Brodbelt, J. S. Middle-Down 193-Nm Ultraviolet Photodissociation for Unambiguous Antibody Identification and Its Implications for Immunoproteomic Analysis. *Anal. Chem.* **2017**, *89*, 6498–6504.

(36) Cotham, V. C.; Brodbelt, J. S. Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level Using Middle-Down 193 Nm Ultraviolet Photodissociation. *Anal. Chem.* **2016**, *88*, 4004–4013.

(37) Fornelli, L.; Ayoub, D.; Srzentic, K.; Nagornov, K.; Kozhinov, A.; Gasilova, N.; Menin, L.; Beck, A.; Tsybin, Y. Structural Analysis of Monoclonal Antibodies with Top-down and Middle-down Electron Transfer Dissociation Mass Spectrometry: The First Decade. *CHIMIA* **2022**, *76*, 114–114.

(38) Fornelli, L.; Srzentić, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap

Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **2018**, *90*, 8421–8429.

(39) Srzentić, K.; Nagornov, K. O.; Fornelli, L.; Lobas, A. A.; Ayoub, D.; Kozhinov, A. N.; Gasilova, N.; Menin, L.; Beck, A.; Gorshkov, M. V.; Aizikov, K.; Tsybin, Y. O. Multiplexed Middle-Down Mass Spectrometry as a Method for Revealing Light and Heavy Chain Connectivity in a Monoclonal Antibody. *Anal. Chem.* **2018**, *90*, 12527–12535.

(40) Greer, S. M.; Sidoli, S.; Coradin, M.; Schack Jespersen, M.; Schwämmle, V.; Jensen, O. N.; Garcia, B. A.; Brodbelt, J. S. Extensive Characterization of Heavily Modified Histone Tails by 193 Nm Ultraviolet Photodissociation Mass Spectrometry via a Middle–Down Strategy. *Anal. Chem.* **2018**, *90*, 10425–10433.

(41) Sanders, J. D.; Greer, S. M.; Brodbelt, J. S. Integrating Carbamylation and Ultraviolet Photodissociation Mass Spectrometry for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89*, 11772–11778.

(42) Pandeswari, P. B.; Sabareesh, V. Middle-down Approach: A Choice to Sequence and Characterize Proteins/Proteomes by Mass Spectrometry. *RSC Adv.* **2018**, *9*, 313–344.

(43) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal. Chem.* **2018**, *90*, 40–64.

(44) Brodbelt, J. S.; Morrison, L. J.; Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem. Rev.* **2020**, *120*, 3328–3380.

(45) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 1044–1051.

(46) Penkert, M.; Yates, L. M.; Schümann, M.; Perlman, D.; Fiedler, D.; Krause, E. Unambiguous Identification of Serine and Threonine Pyrophosphorylation Using Neutral-Loss-Triggered Electron-Transfer/Higher-Energy Collision Dissociation. *Anal. Chem.* **2017**, *89*, 3672–3680.

(47) Escobar, E. E.; King, D. T.; Serrano-Negrón, J. E.; Alteen, M. G.; Vocadlo, D. J.; Brodbelt, J. S. Precision Mapping of O-Linked N-Acetylglucosamine Sites in Proteins Using Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2020**, *142*, 11569–11577.

(48) Sandra, K.; Vanhoenacker, G.; Vandenheede, I.; Steenbeke, M.; Joseph, M.; Sandra, P. Multiple Heart-Cutting and Comprehensive Two-Dimensional Liquid Chromatography Hyphenated to Mass Spectrometry for the Characterization of the Antibody-Drug Conjugate Ado-Trastuzumab Emtansine. *J. Chromatogr. B* **2016**, *1032*, 119–130.

(49) Martelet, A.; Garrigue, V.; Zhang, Z.; Genet, B.; Guttman, A. Multi-Attribute Method Based Characterization of Antibody Drug Conjugates (ADC) at the Intact and Subunit Levels. *J. Pharm. Biomed. Anal.* **2021**, *201*, 114094.

Chapter 5: Integrated Top-down and Bottom-up Mass Spectrometry for Characterization of Diselenide Bridging Patterns of Synthetic Selenoproteins[3]

5.1 Abstract

With the rapid acceleration in the design and development of new biotherapeutics, ensuring consistent quality and understanding degradation pathways remains paramount, requiring an array of analytical methods including mass spectrometry. The incorporation of non-canonical amino acids, such as for synthetic selenoproteins, creates additional challenges. A comprehensive strategy to characterize selenoproteins should serve dual purposes of providing sequence confirmation and allowing mapping of selenocysteine bridge locations and the identification of unanticipated side-products. In the present study, a combined approach exploiting the benefits of both top-down and bottom-up mass spectrometry was developed. Both electron-transfer/higher-energy collision dissociation (EThcD) and 213 nm ultraviolet photodissociation (UVPD) were utilized to provide highly complementary information, allowing high quality characterization, localization of diselenide bridges for complex proteins, and the identification of previously unreported selenoprotein dimers.

5.2 Introduction

Disulfide bridges, formed by the oxidation and linkage of two cysteine residues, contribute significantly to the stabilities and structures of proteins.[1,2] The escalating interest in the assembly, degradation, and characterization of disulfide bridges is unsurprising given their ubiquitous presence in antibody-based biotherapeutics.[3,4]

[3]Watts, E.; Thyer, R.; Ellington, A. D.; Brodbelt, J. S. Integrated Top-Down and Bottom-Up Mass Spectrometry for Characterization of Diselenide Bridging Patterns of Synthetic Selenoproteins. Anal. Chem. 2022, 94, 11175–11184.

Incorporation of selenocysteine instead of cysteine in proteins has recently been explored as an innovative strategy to advance the development of biotherapeutics.[5,6] Selenocysteine has a significantly lower pKa than cysteine, resulting in the facile deprotonation of selenocysteine at physiological pH.[7] Moreover, selenocysteine has a much lower redox potential than cysteine, leading to favorable generation of diselenide bridges with increased stability relative to conventional disulfide bonds.[8] As a result of this unique chemistry, the incorporation of selenocysteine in the place of cysteine has the potential to improve the production and stability of biotherapeutic proteins. However, the unusual biosynthetic pathway involving the enzymatic conversion of serine to selenocysteine on its tRNA, and constraints within the coding sequence of the target protein imposed by the native protein and RNA factors made this previously unachievable in recombinant protein systems.[9,10] Breakthroughs in synthetic biology have led to the site-specific incorporation of selenocysteine and formation of diselenide bridges in recombinant proteins generated in *E. coli*.[5,6]

Given the great potential for enhancing design and development of novel biotherapeutics via incorporation of selenocysteines and/or other non-canonical amino acids, it is critical to develop versatile analytical methods for their characterization.[2,3] The characterization of proteins containing disulfide bridges and their diselenide analogs presents a challenge, and an array of approaches ranging from bottom-up to top-down tandem mass spectrometry (MS/MS) methods have been employed.[2,11,12] For bottom-up methods, proteins are typically reduced prior to enzymatic digestion, thus facilitating characterization of the primary protein sequences without the hinderance of disulfide bridges. Structural analysis of disulfide bridges has also been achieved through strategic digestion without reduction.[2] Identification of peptides resulting from enzymatic digestion is traditionally accomplished by utilizing collision induced dissociation (CID),

however, in the case of peptides containing disulfide bridges, little success is achieved with CID alone.[2] As a means to reduce disulfide bonds and streamline the mapping of disulfide bridges, innovative ionization methods, including corona discharge[13] and solvent assisted photoionization,[14] as well as on-line electrochemical reduction,[15] have been implemented. Improved characterization has also been reported for peptides containing intramolecular disulfide bonds by examining internal ions as well as their disulfide bridge-containing complements, or "external ions" produced by CID.[16] Other MS/MS methods using alternative ion activation techniques have been used to augment the characterization of disulfide-bridged peptides.[2] Both MALDI in-source decay[17–20] and electron transfer dissociation (ETD)[21–28] have been shown to yield fragment ions resulting from the cleavage of the disulfide bond. Hybrid ETD methods, including electron-transfer/CID and electron-transfer/higher energy collisional dissociation (EThcD), boost the generation of sequence-related fragment ions.[23,25] The capabilities of ultraviolet photodissociation (UVPD) for analysis of disulfide-bridged peptides have been explored.[29–31] 266 nm UVPD results in site-specific cleavage of disulfide bonds, offering higher confidence in peptide pairing and better characterization when combined with electron capture dissociation (ECD).[30,31] 193 nm UVPD has also achieved complete characterization of peptides containing complex disulfide bridges by integrating informative fragment ions resulting from cleavage of disulfide bridges with fragment ions retaining disulfide linkages.[29] The ability of 193 nm UVPD to achieve high quality characterization without requiring MS^3 approaches represented a promising benchmark for analysis of disulfide- and diselenide-bridged peptides.

While bottom-up mass spectrometry has been ideal for identifying individual disulfide bridges, top-down mass spectrometry offers the potential for greater confidence in protein characterization and allows the opportunity for comprehensive differentiation

of complex proteoforms.[32,33] As also witnessed in bottom-up workflows, disulfide bridges and their analogs present a major challenge for top-down methods because the disulfide bonds often remain intact upon protein fragmentation, limiting the sequence coverage in sections spanned by the bridged regions. Electron transfer dissociation (ETD) and electron capture dissociation (ECD) of intact proteins results in limited cleavage of disulfide bridges, with most identified fragments resulting from backbone cleavage outside of constrained regions.[11,34–37] While the lack of fragment ions over the disulfide-bridged sections limits sequence coverage, it can help confirm the presence and location of disulfide bonds, and the reduced complexity of the resulting spectra can increase confidence in the fragment ion identifications.[11,38] Hybrid electron-capture/higher energy collisional dissociation and 157 nm UVPD promote cleavage of disulfide-bridged subunits, as demonstrated for intact antibodies; however characterization remains limited owing to survival of some intrapeptide disulfide bonds.[12] Another hybrid method, activated ion ETD (AI-ETD), has exhibited the greatest success in characterizing the constrained regions of disulfide bridged proteins, with enhanced cleavage of intrapeptide disulfide bonds resulting in sequence coverage comparable to that of reduced proteins.[39,40] A recent study of insulin dimers demonstrated the ability of 213 nm UVPD to successfully analyze large peptides resulting from complex linkages.[41] Although the insulin dimers examined in this seminal study were smaller than the proteins considered in most top-down studies, the results represented a promising adaptation of 213 nm UVPD for disulfide characterization.[41]

While the characterization of therapeutic proteins containing disulfide bridges has been explored extensively with tandem mass spectrometry, there has been little emphasis on proteins containing diselenide linkages.[42] A few bottom-up studies have successfully identified cysteine-selenocysteine linkages in naturally occurring selenoproteins by using

CID methods.[43,44] Recent advances in development of custom recombinant selenoproteins have reinforced the need for improved analytical methods for characterization of diselenide bonds.[5,6] Initial studies of synthetic selenoproteins utilized top-down mass spectrometry, facilitated by 193 nm UVPD.[5,6] While the methodology was successful in confirming the mass of the proteins and localizing the diselenide linkages, the sequence coverage was constrained by the diselenide bridges, hindering complete characterization of the proteins.[5,6] As described in the present study, 213 nm UVPD, EThcD, ETD, and higher energy collisional dissociation (HCD) are integrated in a combined top-down/bottom-up workflow to characterize synthetic selenoproteins with the aim of characterizing the entire sequence of the proteins as well as interrogating diselenide linkages.

5.3 EXPERIMENTAL

5.3.1 Samples and Reagents

Recombinant selenoproteins samples were prepared as previously reported.[5] All Millipore OmniSolv LC-MS grade solvents were obtained from VWR. Lysyl-endopeptidase (Lys-C) was purchased from Santa Cruz Biotechnology. All other reagents were purchased from Thermo-Fisher Scientific.

5.3.2 Top-Down Mass Spectrometry

Each protein sample was buffer exchanged into water with Micro Bio-Spin 6 columns (BioRad) and then diluted to approximately 10 μM in 50:50 water/acetonitrile with 1% formic acid. Samples were then loaded into house-made borosilicate emitters coated in Au/Pd for nano-electrospray infusion with an applied voltage of 0.9-1.3 kV. All experiments were performed with an Orbitrap™ Fusion Lumos mass spectrometer

(Thermo Scientific) equipped with 213 nm UVPD. For all MS1 spectra 100 microscans were collected with a resolving power of 120,000 at m/z 200. MS2 experiments combined 500 microscans with a resolving power of 240,000 at m/z 200. Except when otherwise noted top-down activation parameters were as follows, 15% normalized collisional energy (NCE) for HCD, 5 ms activation time for ETD and EThcD with 15% NCE supplemental activation for EThcD, and 50 ms activation period for 213 nm UVPD equivalent to 125 pulses (approximately 3 μJ/pulse) at 2500 Hz. Top-down mass spectra were deconvoluted with the Xtract algorithm in Thermo Scientific Freestyle using a signal to noise (S/N) threshold of 3, fit factor of 80% and remainder threshold of 25%. Fragment ions were identified using ProSight Lite with a 10-ppm tolerance for fragment ions. TDValidator was used to generate the fragment ion distributions included in **Figure 5.1** to demonstrate the quality of the isotope patterns generated for representative fragment ions with high S/N and fit factors (**Figure 5.1A**) as well as lower S/N (**Figure 5.1B**) and fit factors (**Figure 5.1C**) close to the cutoff values. The unusual isotope pattern of selenium resulted in deconvoluted masses shifted by approximately -1.0 Da. As a compensatory measure, 2.01 Da was subtracted from each selenocysteine resides to account for the isotope pattern and the loss of one hydrogen to form a diselenide bridge.

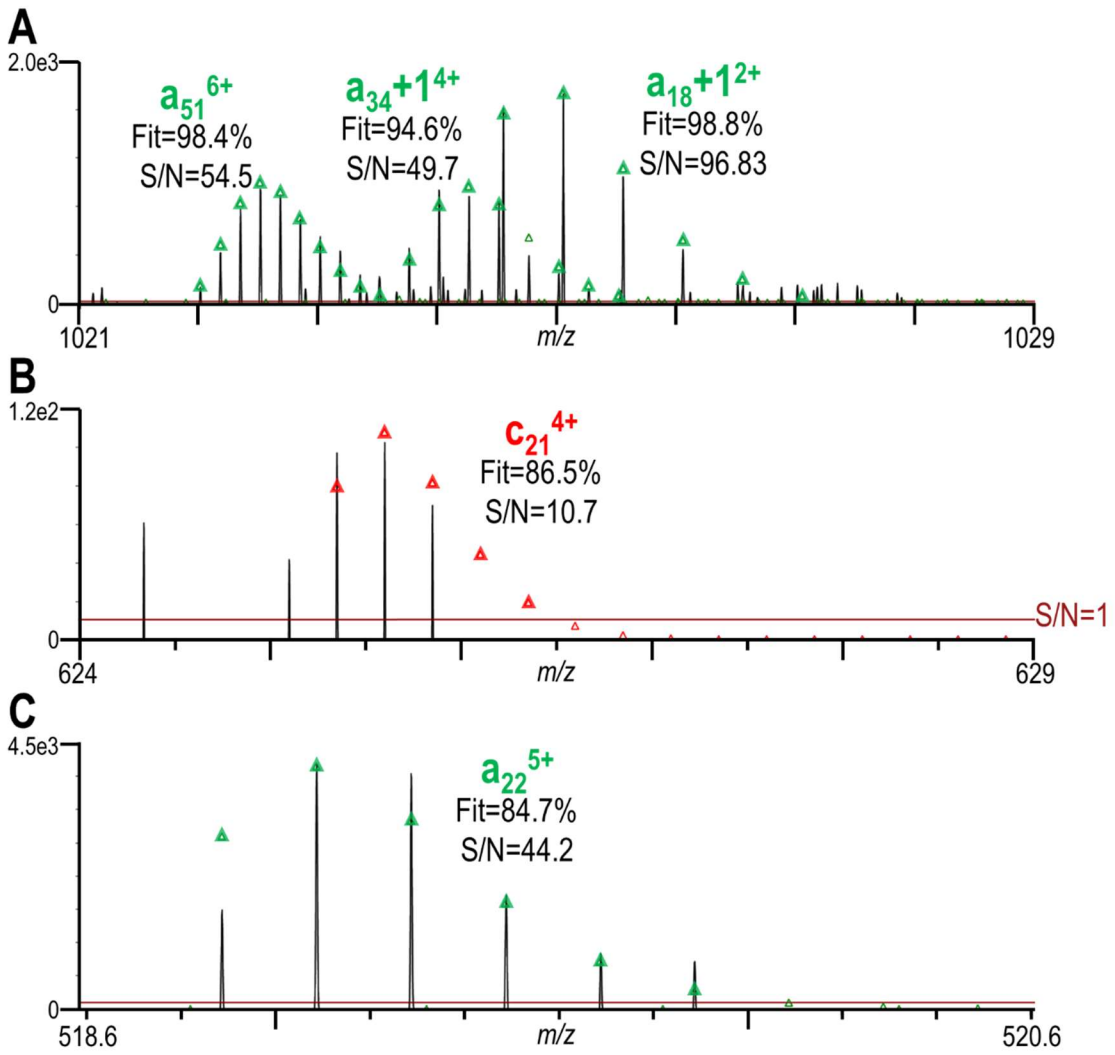


Figure 5.1: Expanded m/z regions from top-down spectrum of HGH 16+ fragmented with 213 nm UVPD (50 ms activation period) demonstrate the isotopic distributions for fragment ions with (A) high S/N values and fit factors, (B) S/N close to the cutoff value, and (C) fit factor close to the cutoff value. The triangles overlayed on each spectra demonstrate the predicted isotope pattern generated with TDValidator.

An excel spreadsheet listing all fragment ions identified for each protein, is available at: https://doi.org/10.1021/acs.analchem.2c01433. Each protein and fragmentation type is listed on a separate sheet. There is one line for every charge state of

each fragment ion. For each fragment ion matched, the ion name, theoretical mass, observed mass, the mass difference in both Daltons and ppm, the intensity, and the number of charge states matched is included. For each charge state the calculated m/z, most abundant m/z, the charge normalized intensity, and the fit factor (%) reported by FreeStyle are reported.

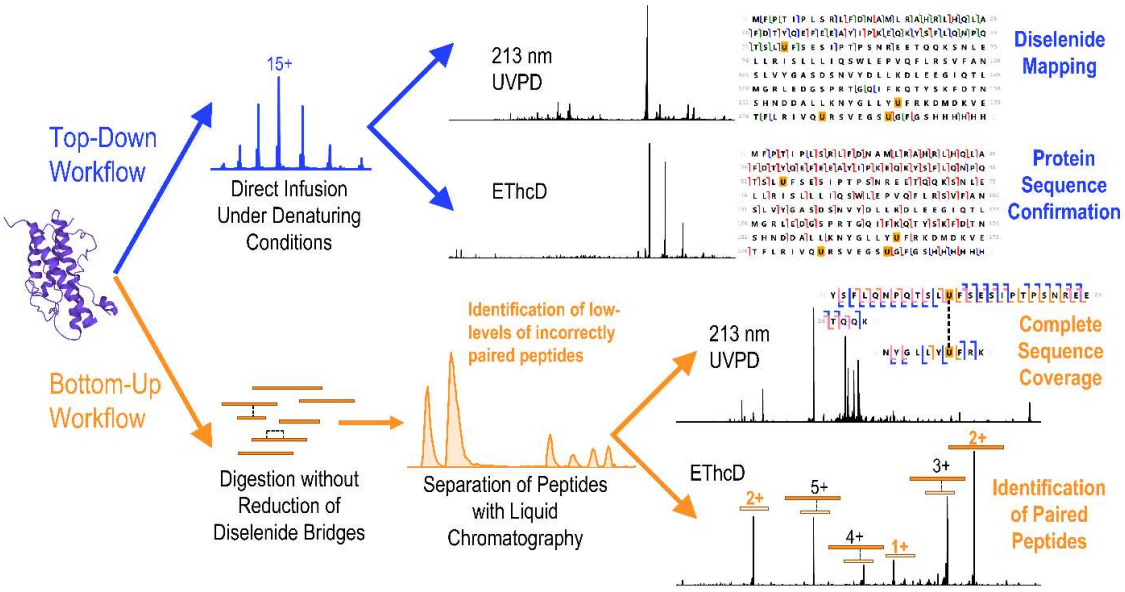
5.3.3 Bottom-Up Liquid Chromatography-Mass Spectrometry

For digestion, 50 µg of sample was prepared in Tris-HCL at pH 6.0. Lys-C was then added at a 1:25 enzyme to protein ratio and samples were incubated at 37°C for 18 hours. Digests were buffer exchanged into water with Amicon Ultra 3 kDa molecular weight cut off centrifugal filters. Samples were stored at -80°C until analysis. Samples were diluted to 0.1-0.3 mg/mL for a 1 µl injection.

Reversed phase liquid chromatography (LC) was performed using a Dionex Ultimate nano LC system configured to trap and elute. Trap and analytical columns were packed in-house to a length of 3 and 20 cm, respectively, using 1000 Å, 50 µm bulk PLRP media from Agilent. Samples were loaded onto the trap column with starting conditions of 2% acetonitrile and 0.1% formic acid in water at 5 µL/min. After 5 minutes of loading, a valve switch placed the trap column in-line with the analytical column. Analytical mobile phases comprised water (A) and acetonitrile (B) with 0.1% formic acid (A and B) were applied at an initial gradient of 2% to 6% B over 2 minutes followed by a gradient up to 35% B over 33 minutes at a flow rate of 300 nL/ min. Introduction into the Orbitrap™ Fusion Lumos mass spectrometer occurred with an applied voltage of 2-2.4 kV. MS1 scans with a resolving power of 60,000 at *m/z* 200 and an AGC target of $4x10^5$, using an *m/z* range of 400-2000. Top ten data dependent MS2 scans were selected with an intensity filter of $1x10^5$. All MS2 scan were collected with an AGC target of $5x10^5$,

within an *m/z* range of 300-2000. Except for where otherwise noted, 3 microscans were averaged for each MS^2 spectrum with a resolving power of 120,000 at *m/z* 200. HCD scans were collected with 30% NCE. ETD and EThcD scans were acquired with calibrated charge-dependent ETD parameters and 15% NCE supplemental activation in the case of EThcD. 213 nm UVPD activation occurred over a 100 ms activation period, equivalent to 250 pulses (approximately 3 μJ/pulses) at 2500 Hz, unless otherwise indicated.

Bottom-up data was analyzed with Byonic (Protein Metrics version 4.2.10) and searched against a single protein database with the appropriate protein sequence. Cleavage site was set to lysine, fully specific, with two missed cleavages allowed, and both precursor and fragment mass tolerances were set to 20 ppm. The conversion of selenocysteine to serine was set as a custom modification with a mass loss of 63.92 Da at selenocysteine. Diselenide, disulfide, and Se-S bonds were searched as a custom crosslink with the loss of two hydrogens. When calculating total protein sequence coverage, only peptides that were identified in two out of three replicates were included. All peptides identified containing selenocysteine, including those with serine incorporation in the place of selenocysteine, were manually verified by examining and deconvoluting the associated 213 nm UVPD mass spectra and identifying matching fragments with ProSight Lite. EThcD mass spectra were also manually verified for specific high abundance fragments related to cleavage across the diselenide bond to confirm the presence or absence of interpeptide diselenide bridges.



Scheme 5.1: Workflow diagram displaying key information obtained using a combination of top-down and bottom-up mass spectrometry and a combination of 213 nm UVPD and EThcD as illustrated for HGH (crystal structure PDB ID: 1HGU).

5.4 DISCUSSION

As illustrated in the workflow overview in **Scheme 5.1,** the combination of top-down and bottom-up strategies provides complementary information that allows the most comprehensive characterization of selenoproteins. While top-down mass spectrometry affords high-level characterization of sequences and locations of diselenide bridges, bottom-up methods offer confident identification of incorrectly paired peptides that contribute to the protein heterogeneity. Additionally, while ETD and HCD were also explored to evaluate their attributes for mapping diselenide bridges, 213 nm UVPD and EThcD provided the most detailed information and were highly complementary in both top-down and bottom-up approaches. The benefits and limitations of each method, as well as the advantages in combining the workflows, are described in more detail in the following sections.

5.4.1 Top-Down Analysis

High-resolution (120,000 at *m/*z 200) MS1 spectra acquired for each intact selenoprotein, as exemplified in **Figure 5.2**, allows the first level of assessment of the three selenoprotein constructs. The first construct examined is green fluorescent protein (GFP) which only contains one selenocysteine pair. The only notable masses in the deconvoluted spectrum of GFP (**Figure 5.2A**) correspond to the intact protein (28011 Da) and the intact protein after methionine loss (27880 Da), a common truncation caused by the enzyme methionine aminopeptidase which cleaves methionine co-translationally.[45] The other two proteins, trastuzumab single chain variable fragment (scFv) and human growth hormone (HGH), both contained two pairs of selenocysteine residues. While methionine loss was not observed for either protein, another truncated product corresponding to a mass shift of 124.9 Da was noted for each protein in the deconvoluted mass spectra (**Figure 5.2B-C**). The net mass shift of 124.9 Da is explained by the incorporation of two serine residues in place of two selenocysteine residues. The incorporation of serine instead of selenocysteine is a known side reaction that occurs because serylated $tRNA^{Sec}$ is an intermediate in the generation of selenylated $tRNA^{Sec}$.[5] Interestingly, the incorporation of serine in place of selenocysteine was only observed in pairs, a phenomenon that will be examined in greater detail in the last section of the discussion.

After initial mass analysis of the intact proteins, MS/MS spectra were collected utilizing HCD, ETD, EThcD, and UVPD. The optimization of activation parameters (e.g, collision energy, activation time) for each activation method is demonstrated for two representative charge states (27+, 30+) of GFP in **Figure 5.3,** and examples of optimized MS/MS spectra for GFP (27+) are displayed in **Figure 5.4**. Little difference in sequence coverage was observed based on charge state, with the lower and more abundant charge

state (27+) typically resulting in slightly higher sequence coverage than the higher and less abundant charge state (30+). Considering these results, high abundance charge states were selected for each protein for MS/MS analysis. For trastuzumab the 20+ and 19+ charge states had similar abundances, and the 20+ charge state was selected to maximize fragmentation for charge-dependent methods, whereas for GFP the 16+ charge state was significantly more abundant than adjacent charge states and was therefore utilized to maximize fragment ion abundances. Based on the outcomes in **Figure 5.3** for GFP and similar assessments for the other two proteins (data not shown), the optimal activation parameters were 15% NCE for HCD, 5 ms activation time for ETD and EThcD with 15% NCE supplemental activation for EThcD, and 50 ms activation period for 213 nm UVPD. A series of representative sequence maps and MS/MS spectra are shown in **Figure 5.5** and **Figure 5.6, respectively,** for trastuzumab scFv (20+). While the highest sequence coverage was obtained using EThcD (**Figure 5.5C**), a significant number of fragment ions resulted from cleavage of the diselenide bridges linking U24 to U89 and U129 to U233, creating ambiguity in bridge locations. Bracketing of the diselenide bonds was more cleanly accomplished by HCD and ETD (**Figure 5.5A-B**), as indicated by the large stretches of the sequence that reveal no or few backbone cleavages. However, HCD and ETD resulted in low sequence coverage, particularly in the N-terminal regions. 213 nm UVPD (**Figure 5.5D**) resulted in higher sequence coverage than HCD and ETD and better bracketing of diselenide bonds than EThcD. Similar results are illustrated for GFP in **Figure 5.4**. EThcD and 213 nm UVPD proved to be highly complementary methods and the utilization of both led to confident identification of diselenide bridge locations as well as characterization of the protein sequence.

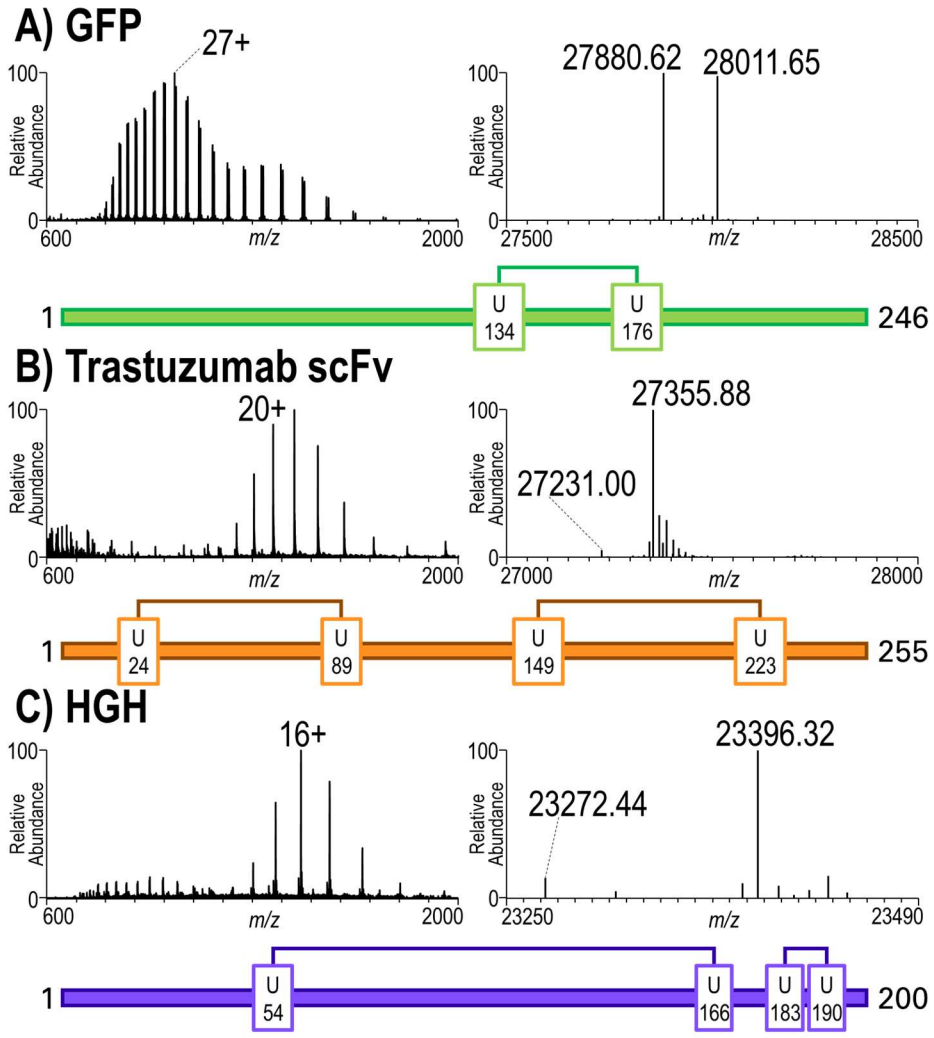


Figure 5.2: Electrospray ionization (left) and deconvoluted (right) mass spectra for three modified selenocysteine-containing proteins including (A) GFP, (B) trastuzumab scFv, and (C) HGH. All spectra were collected with a resolution of 120,000 at *m/z* 200. Schematic depictions of the pairings of diselenides for each protein are shown below each spectrum.

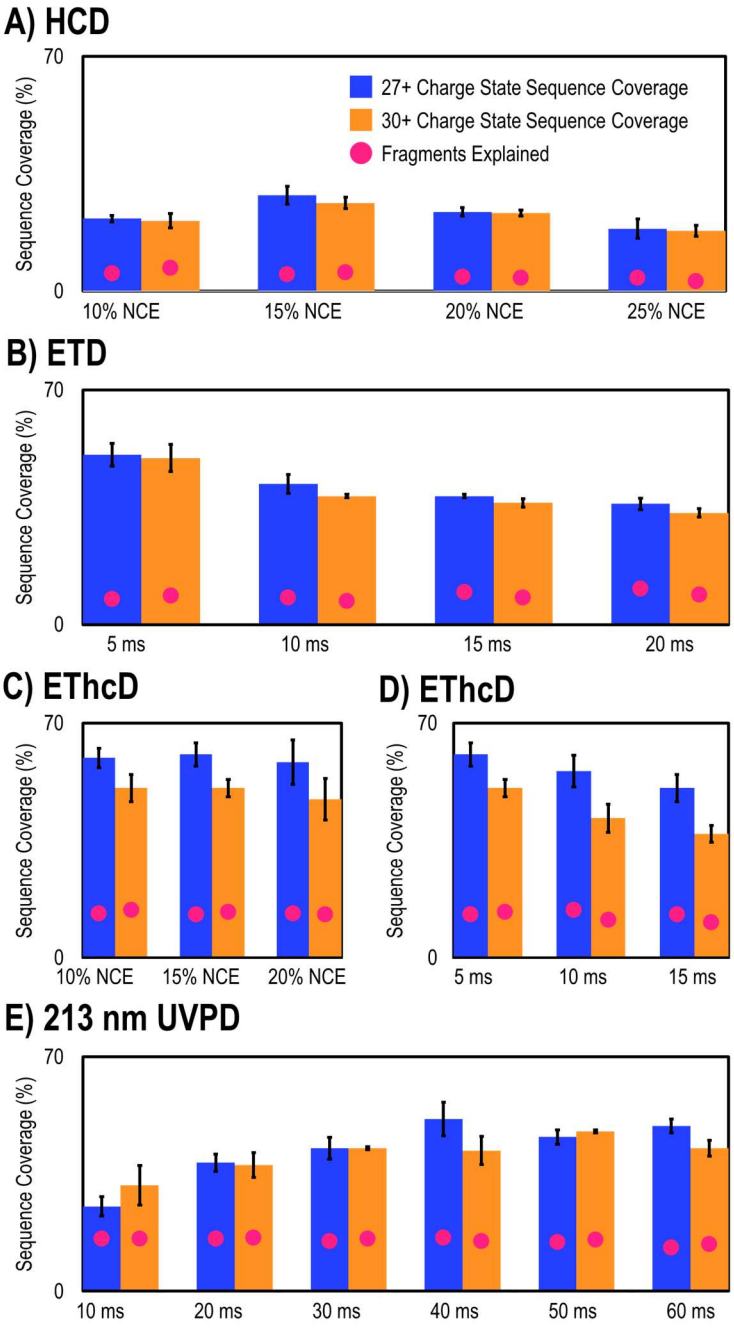


Figure 5.3: Sequence coverage (bar charts) and percent of fragment ions explained (dots) as a function of (A) HCD collisional energy, (B) ETD activation time, (C) EThcD supplemental collisional energy, (D) EThcD activation time, or (E) 213 nm UVPD activation period for GFP (23kDa) in both the 27+ and 30+ charge states.

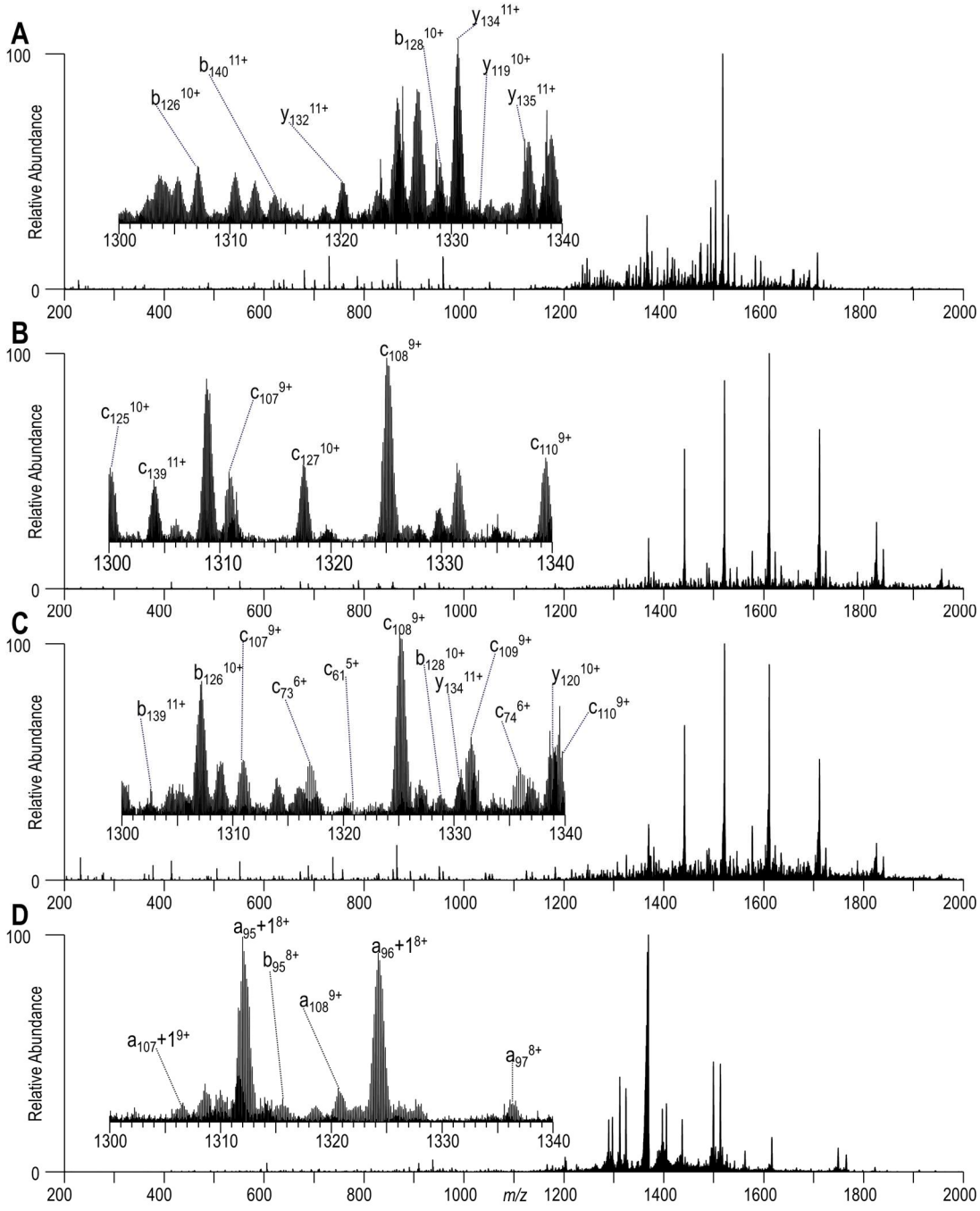


Figure 5.4: MS/MS spectra with labeled insets showing expansions of selected fragment ions and sequence coverage maps for trastuzumab scFv (20+) obtained using (A) HCD (15% NCE), (B) ETD (5 ms activation time), (C) EThcD (5 ms activation time and 15% NCE supplemental activation), and (D) 213 nm UVPD (50 ms activation period).

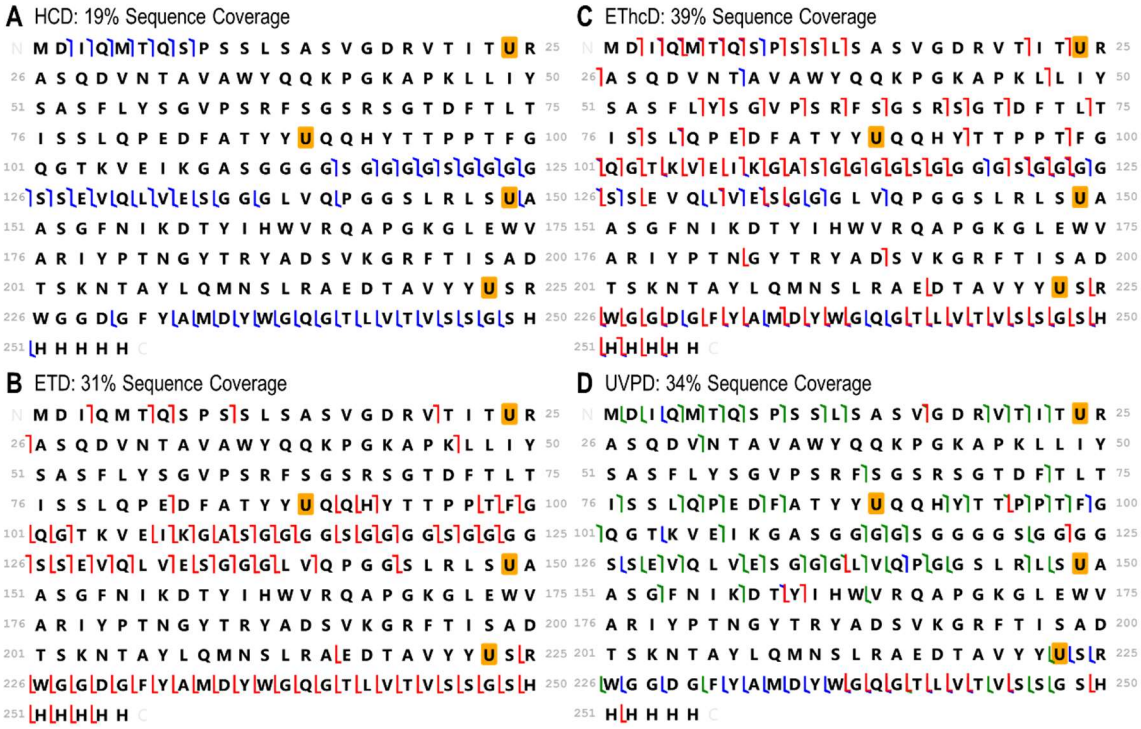


Figure 5.5: Sequence coverage maps for trastuzumab scFv (20+) obtained using (A) HCD (15% NCE), (B) ETD (5 ms activation time), (C) EThcD (5 ms activation time and 15% NCE supplemental activation), and (D) 213 nm UVPD (50 ms activation period).

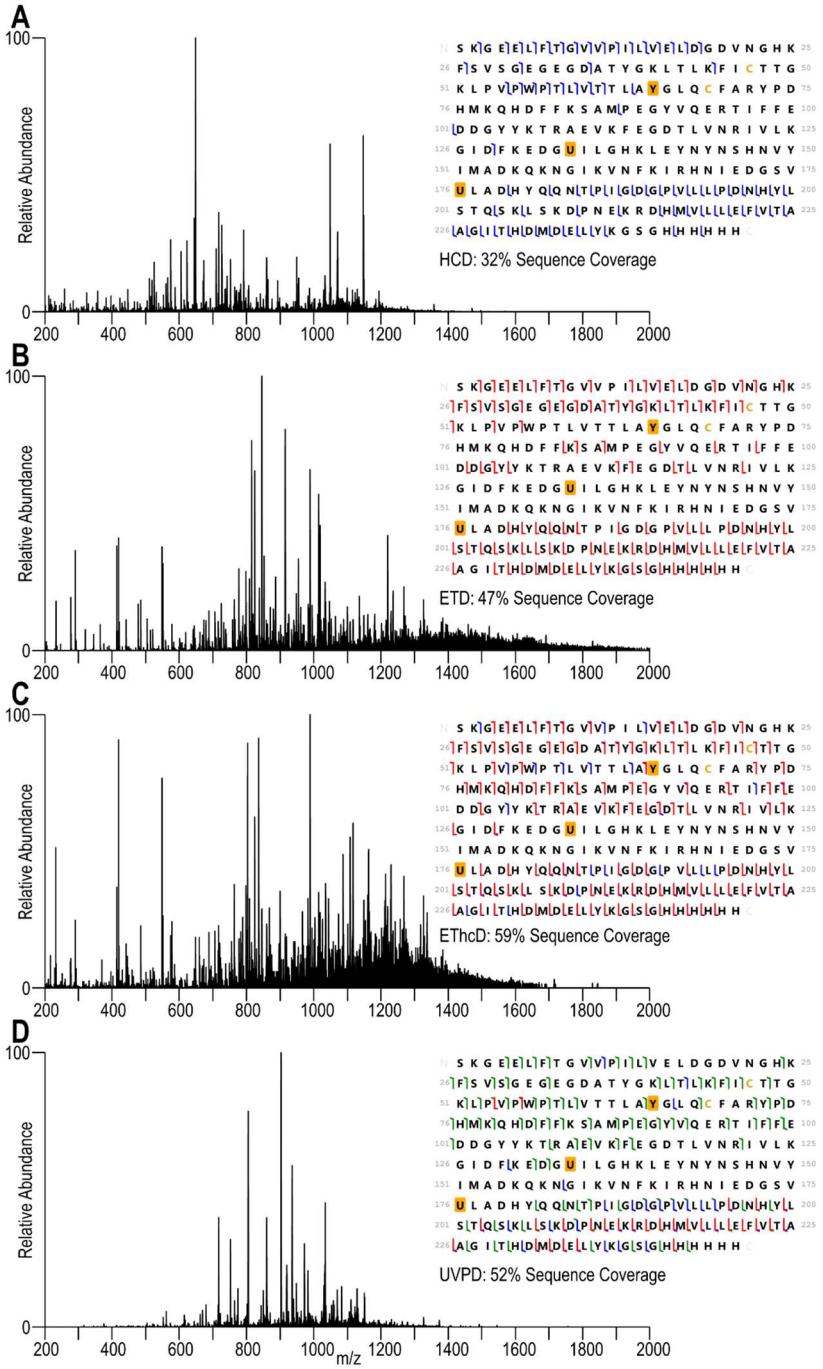


Figure 5.6: MS/MS spectra with sequence coverage maps for GFP (27+) produced by (A) HCD (15% NCE), (B) ETD (5 ms activation time), (C) EThcD (5 ms activation time and 15% NCE supplemental activation), and (D) 213 nm UVPD (50 ms activation period).

The final protein examined with top-down mass spectrometry was HGH, and examples of sequence maps and MS/MS spectra are displayed in **Figure 5.7** and **Figure 5.8**. The two diselenide bridges in HGH constrain most of the protein sequence, and three (U166, U183, and U190) of the selenocysteines are nearly adjacent, curbing sequence coverage and complicating the identification of diselenide bridges. This limitation of the top-down approach is most evident for HCD and ETD (**Figure 5.7A-B**) for which very low sequence coverage is achieved and the locations of the diselenide bridges are ambiguous. For EThcD (**Figure 5.7C**) the sequence coverage remains high due to significant cleavage of diselenide bridges, but the diselenide bridge locations cannot be identified. 213 nm UVPD (**Figure 5.7D**) offers the greatest success in bracketing the diselenide bridges of HGH, but given the proximity of U166, U183, and U190, as well as backbone cleavages between U54 and U166, the confidence in pinpointing the location of the diselenide bridge is low. An additional impediment observed in the series of MS/MS spectra for HGH is that although EThcD offered consistently high sequence coverage, patchy fragmentation between bridged selenocysteine residues restricted coverage. To adequately characterize complex patterns of diselenide bridges, such as those observed for HGH, and decipher entire protein sequences constrained by diselenide bridges, a bottom-up method is required to fill in the gaps.

Figure 5.7: Sequence coverage maps for HGH (16+) obtained using (A) HCD (15% NCE), (B) ETD (5 ms activation time), (C) EThcD (5 ms activation time and 15% NCE supplemental activation), and (D) 213 nm UVPD (50 ms activation period).

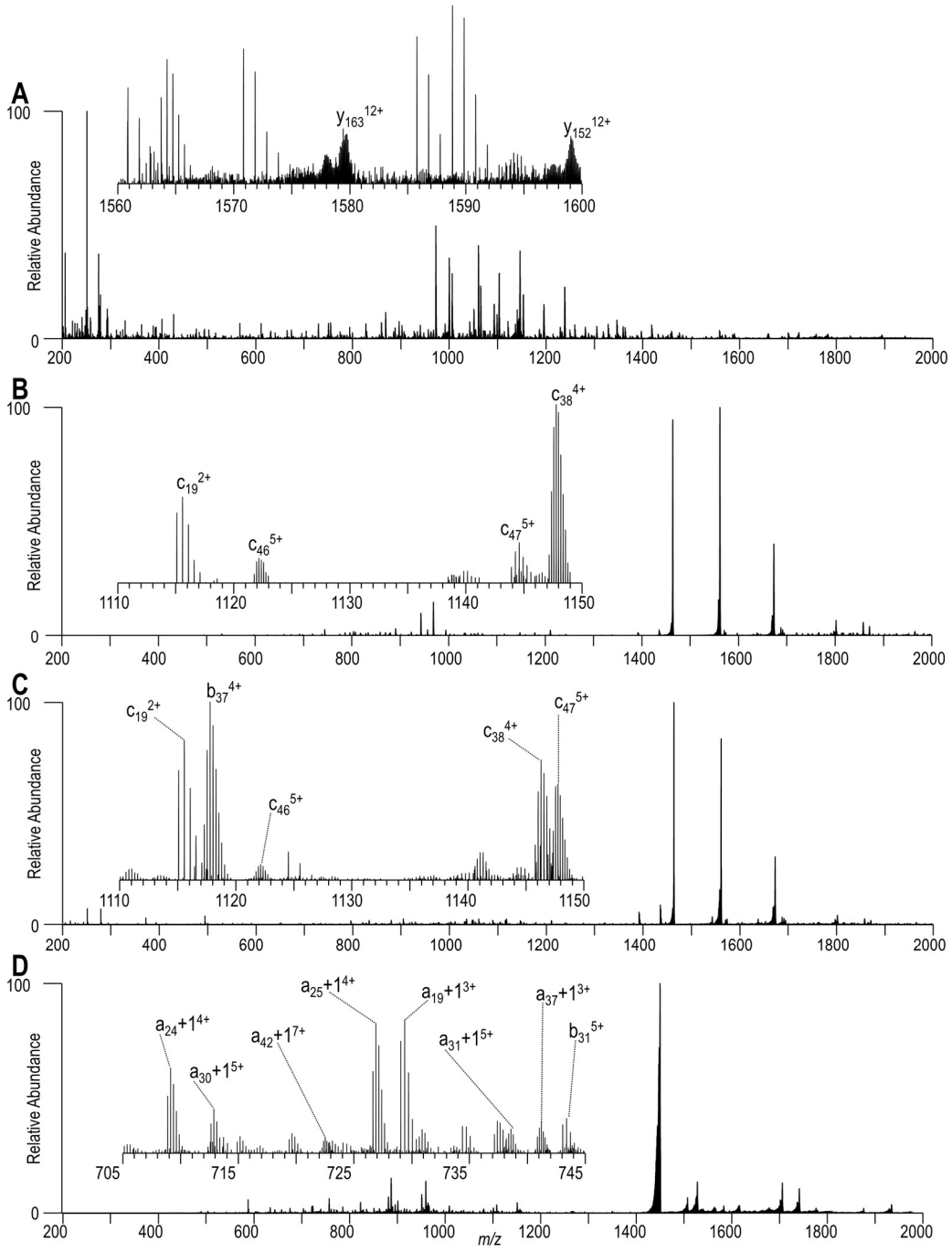


Figure 5.8: MS/MS spectra with labeled insets showing expansions of selected fragment ions and sequence coverage maps for HGH (16+) obtained using (A) HCD (15% NCE), (B) ETD (5 ms activation time), (C) EThcD (5 ms activation time and 15% NCE supplemental activation), and (D) 213 nm UVPD (50 ms activation period).

5.4.2 Bottom-up Characterization of Predicted Peptides

Analysis of HGH using a bottom-up (LysC digestion) LC-MS/MS approach eliminated the shortcomings imposed by diselenide-constrained regions in the top-down strategies. High sequence coverages were observed for each MS/MS method, including 98% for HCD, 76% for ETD, 98% for EThcD and 96% for 213 nm UVPD based on identification of 8, 6, 8, and 7 unique peptides, respectively, excluding replicate peptides that appeared with multiple modifications as listed in **Table 5.1**. A representative chromatogram is included in **Figure 5.9.** The peptides of primary interest for HGH are the two diselenide-bridged species, one containing two peptides with 29 and 10 residues each and the U54-U166 linkage and the other containing 28 residues and the U183-U190 linkage, highlighted in **Scheme 5.2**. Both peptides were identified in the LC-MS data and were selected to evaluate MS/MS parameters. Standard parameters of 30% and 15% NCE were used for HCD and EThcD, respectively, and calibrated charge-dependent activation periods were employed for ETD and EThcD. For 213 nm UVPD, ideal parameters were selected based on mass resolution, number of microscans, and activation period (e.g., number of laser pulses) (**Figure 5.10**).

Table 5.1: All peptides identified in the bottom-up LysC digest of HGH.

Sequence	Modifications	Paired Peptide (If interpeptide diselenide is present)	Calculated Mass (Da)	RT (min)
NYGLLYUFRK	U166 Ser		1259.67	29.60
NYGLLYUFRK	U166-U166 Interpeptide Diselenide	NYGLLYUFRK	2645.16	32.98
NYGLLYUFRKDMDKVETFLRIVQURSVEGSUGFGSHHHHHH	U166 Ser; U183-U190 Intrapeptide diselenide		4920.17	37.15
NYGLLYUFRKDMDKVETFLRIVQURSVEGSUGFGSHHHHHH	U166-U54 Interpeptide Diselenide; U183-U190 Intrapeptide diselenide	YSFLQNPQTSLUFSESIPTPSNREETQQK	8388.59	43.41
QTYSKFDTNSHNDDALLKNYGLLYUFRK	U166-U166 Interpeptide Diselenide	NYGLLYUFRK	4723.13	36.82
VETFLRIVQURSVEGSUGFGSHHHHHH	U183 Ser; U190 Ser		3063.50	26.82
VETFLRIVQURSVEGSUGFGSHHHHHH	U183-U190 Intrapeptide diselenide		3189.32	27.73
VETFLRIVQURSVEGSUGFGSHHHHHH	U183 Ser; U190-U166 Interpeptide Diselenide	NYGLLYUFRK	4448.99	31.62
VETFLRIVQURSVEGSUGFGSHHHHHH	U183-U166 Interpeptide Diselenide; U190 Ser	NYGLLYUFRK	4448.99	32.51
VETFLRIVQURSVEGSUGFGSHHHHHH	U183-U54 Interpeptide Diselenide; U190 Ser	YSFLQNPQTSLUFSESIPTPSNREETQQK	6531.92	33.29
VETFLRIVQURSVEGSUGFGSHHHHHH	U183 Ser; U190-U54 Interpeptide Diselenide	YSFLQNPQTSLUFSESIPTPSNREETQQK	6531.92	33.92
YSFLQNPQTSLUFSESIPTPSNREETQQK	U54 Ser		3342.60	34.17
YSFLQNPQTSLUFSESIPTPSNREETQQK	U54-U166 Interpeptide Diselenide	NYGLLYUFRK	4728.09	36.82
YSFLQNPQTSLUFSESIPTPSNREETQQK	U54-U54 Interpeptide Diselenide	YSFLQNPQTSLUFSESIPTPSNREETQQK	6811.03	38.12
MFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPK			4708.37	36.99
DLEEGIQTLMGRLEDGSPRTGQIFK			2789.4	29.47
FDTNSHNDDALLK			1488.68	15.87
SNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLK			5122.77	47.54
SNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPRTGQIFK			7894.17	47.59

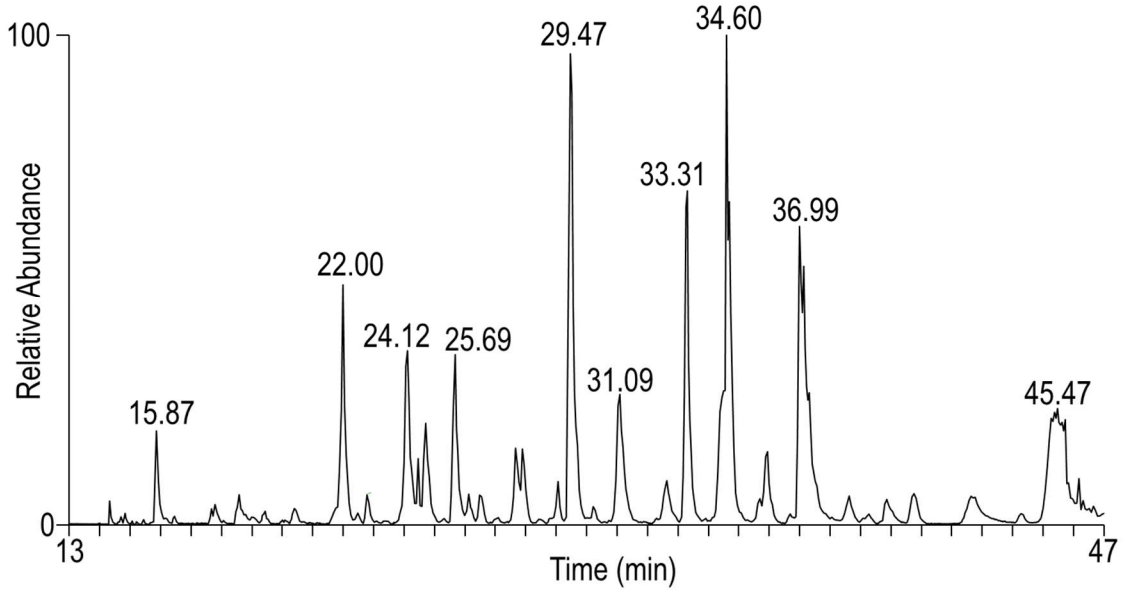
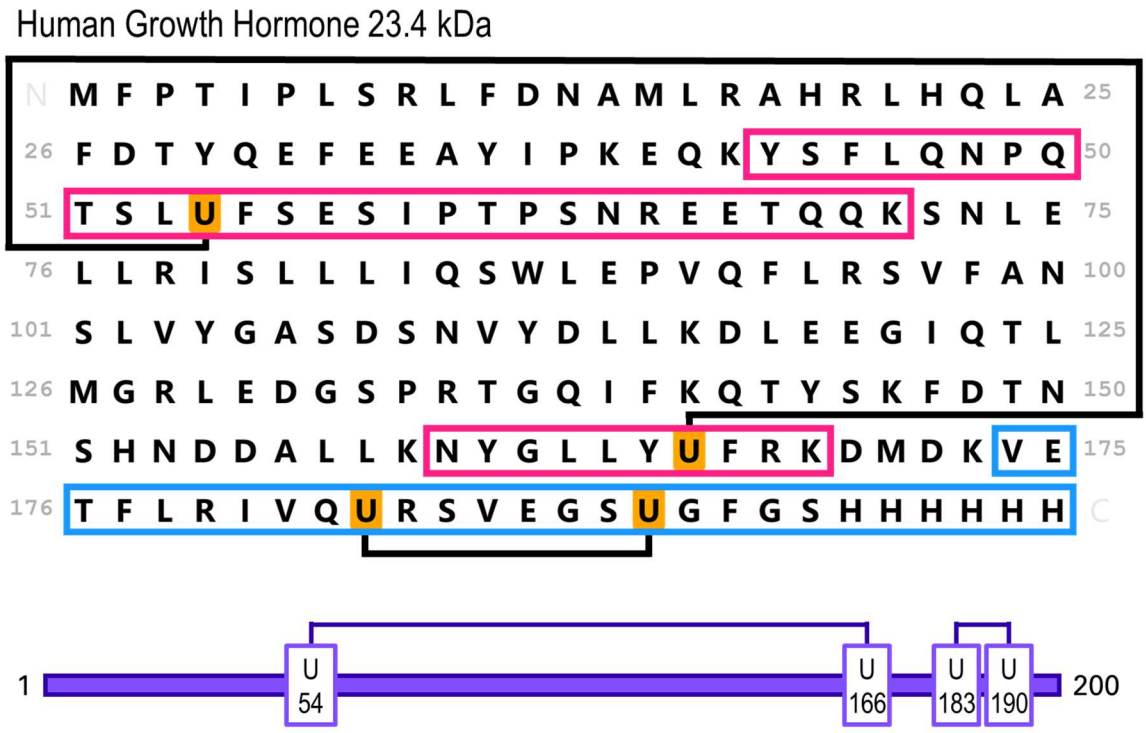


Figure 5.9: Base peak chromatogram of HGH digest



Scheme 5.2: Illustration of predicted diselenide-bridged tryptic peptides for HGH. Each predicted diselenide bridge is shown by a black line. The first pair of diselenide-bridged peptides is outlined in pink (4730.11 Da), and the second diselenide-bridged peptide is outlined in blue (3190.33 Da). Both selenocysteines for the latter peptide are contained within one peptide. Schematic depiction of the pairings of diselenides is also shown.

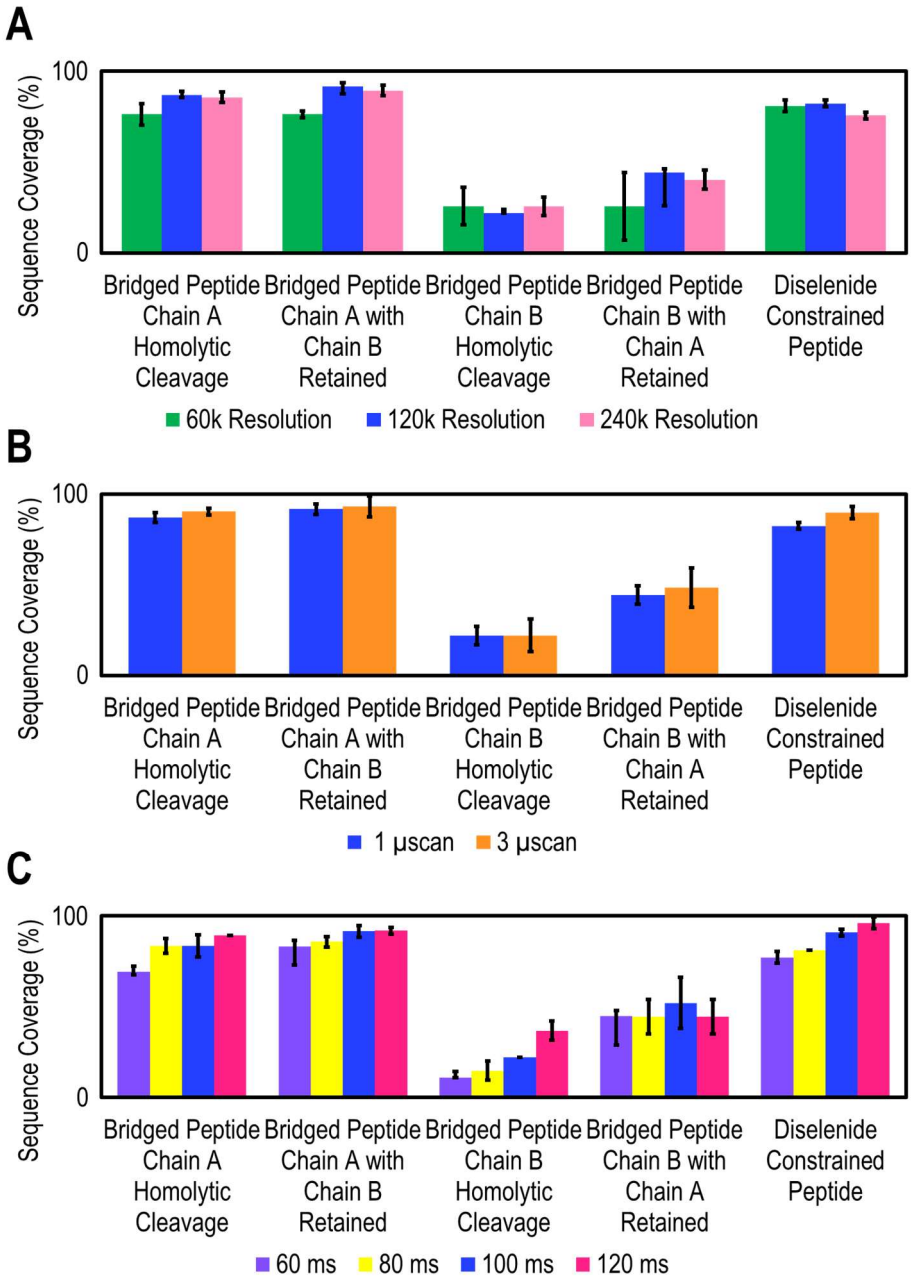


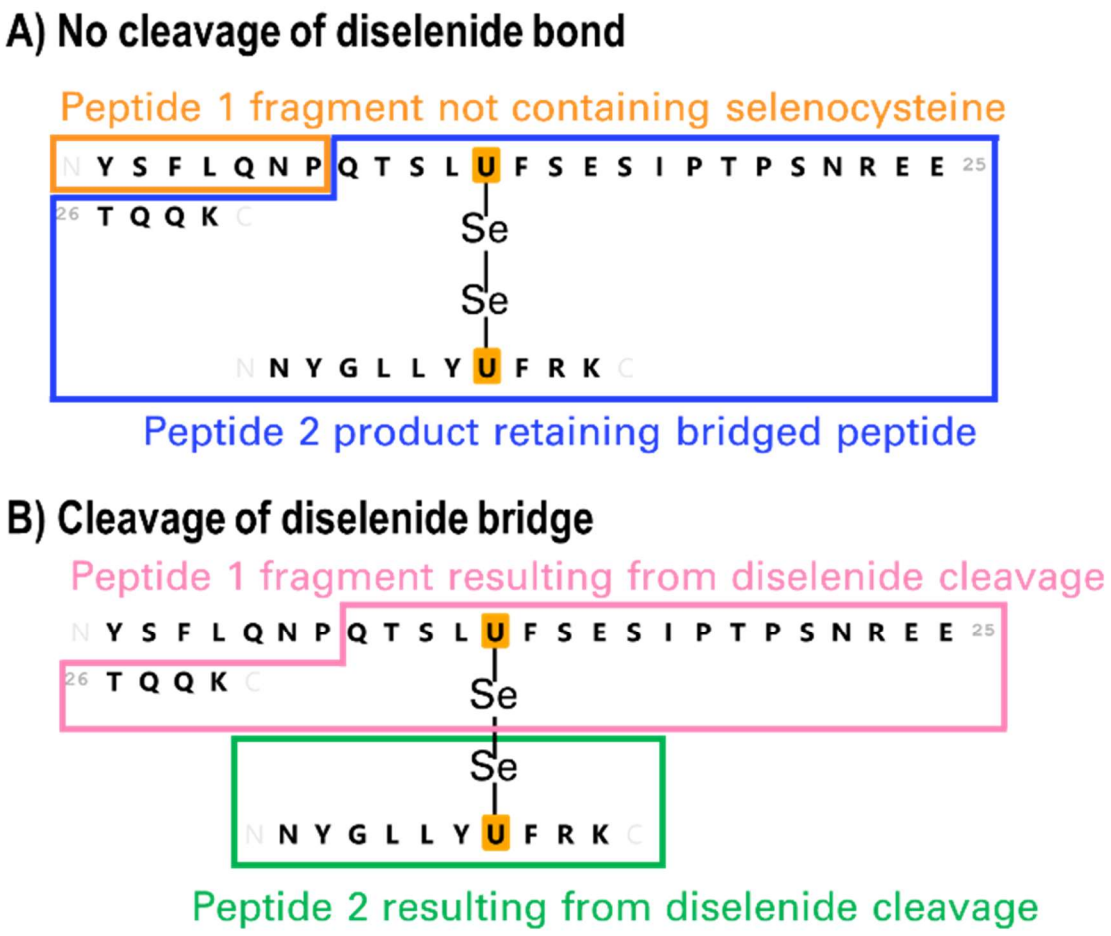
Figure 5.10: Optimization of (A) resolution, (B) number of µscans, and (C) activation period based on 213 nm UVPD of the correctly bridged peptide pair and the diselenide-constrained peptide from the bottom-up digest of HGH. All results in A were acquired using 100 ms activation period and 3 µscans, all data in B was acquired using a 120k resolution and 100 ms activation period and all data in C was acquired using 3 µscans and 120k resolution.

For MS/MS analysis of diselenide-linked peptides produced after digestion of the proteins, there are a variety of possible fragment ions that can be used for sequencing. Each possibility is illustrated in **Scheme 5.3**. In theory, peptide fragment ions can be produced without cleavage of the diselenide bond (**Scheme 5.3A**) or with cleavage of the diselenide bridge (**Scheme 5.3B**). In the case of diselenide cleavage, the resulting products can either undergo additional backbone cleavage to result in a sequence ion (pink fragment in **Scheme 5.3B**) or release a partner peptide without backbone fragmentation (green fragment in **Scheme 5.3B**). When cleavage of the diselenide bridge occurs, it may entail cleavage of the Se-Se bond (homolytic) or may occur at either of the C-Se bonds in an asymmetric manner that results in products containing two selenium atoms or none. Very few peptide fragment ions were identified that resulted from asymmetric cleavage, and it was therefore not considered further in the identification of sequence ions.

As illustrated in the MS/MS spectra shown in **Figure 5.11** and **Figure 5.12**, HCD, EThcD and 213 nm UVPD all resulted in high quality characterization of the U54-U166 diselenide-bridged peptide pair from the HGH protein digest. While all activation methods have merits, significant differences were observed in the resulting MS/MS spectra. For HCD (**Figure 5.11A**) high quality characterization of the sequence was achieved; however, limited production of complementary bi-directional *b/y* ion pairs and the lack of cleavage across the diselenide bond resulted in lower confidence compared to the other methods. The performance of ETD for the same diselenide-bridged peptide was less impressive (**Figure 5.11B**). ETD has been reported to result in peptide partners after cleavage across disulfide bonds, confirming the presence of the disulfide bridges between peptides.[21,22,26,27] This unique cleavage was observed in low abundance upon ETD of the

selenocysteine-bridged peptide in the 5+ charge state (**Figure 5.11B**), and the production of other fragment ions was sparse. Cleavage of the diselenide bond was observed in much greater abundance for the same peptide in the higher 6+ charge state (**Figure 5.13B**), and a larger array of sequence ions was produced. While the charge dependence of ETD could be mitigated by analyzing multiple charge states of each peptide in a run, the same limitation was not observed for EThcD (**Figures 5.11C** and **5.13C**) or 213 nm UVPD (**Figures 5.12** and **5.13D**). EThcD promoted homolytic cleavage of the diselenide bond, resulting in the two complementary peptide partner products as the highest abundance fragment ions, regardless of charge state of the precursor ion. EThcD also resulted in improved sequence coverage compared to ETD, although HCD still surpassed EThcD in total sequence coverage. Only 213 nm UVPD resulted in both the generation of the peptide partner products from homolytic diselenide bond cleavage and complete sequence coverage for both peptide partners (**Figure 5.12**). In addition to homolytic Se-Se cleavage, some peptide products originated from alternative C-Se cleavage upon 213 nm UVPD, resulting in fragment ions that contained two selenium atoms (or none). Although the observation of these C-Se cleavage products was interesting, they were only identified in the peptide partners, not the sequence ions, and therefore did not contribute to the peptide characterization. The high sequence coverage obtained by 213 nm UVPD was enabled by the generation of a significant number of high abundance fragment ions arising from cleavage of the diselenide bond (**Figure 5.12**). While the overall best characterization and sequence coverage was afforded by 213 nm UVPD, EThcD yielded the highest abundance fragment ions. The peptide partners from homolytic diselenide cleavage are critical indicators for unambiguous confirmation of the presence of diselenide bridges in tryptic peptides and are uniquely observed in high abundance with

EThcD. The combination of EThcD and 213 nm UVPD is thus ideal for the identification and characterization of these bridged peptides.



Scheme 5.3: Illustration of possible fragments resulting from MS/MS of a pair of peptides (1 and 2) bridged by a diselenide bond. (A) One possibility is cleavage of a peptide bond without cleavage of the diselenide bonds, resulting in a terminal fragment ion from peptide 1 and not containing selenocysteine, outlined in orange, and a fragment retaining all of peptide 2 and part of peptide 1, outlined in blue. (B) Another possibility is cleavage across the diselenide bond, which can either occur in addition to cleavage of the peptide backbone, exemplified by the product outlined in pink, or without cleavage of the peptide backbone, resulting in the product outlined in green.

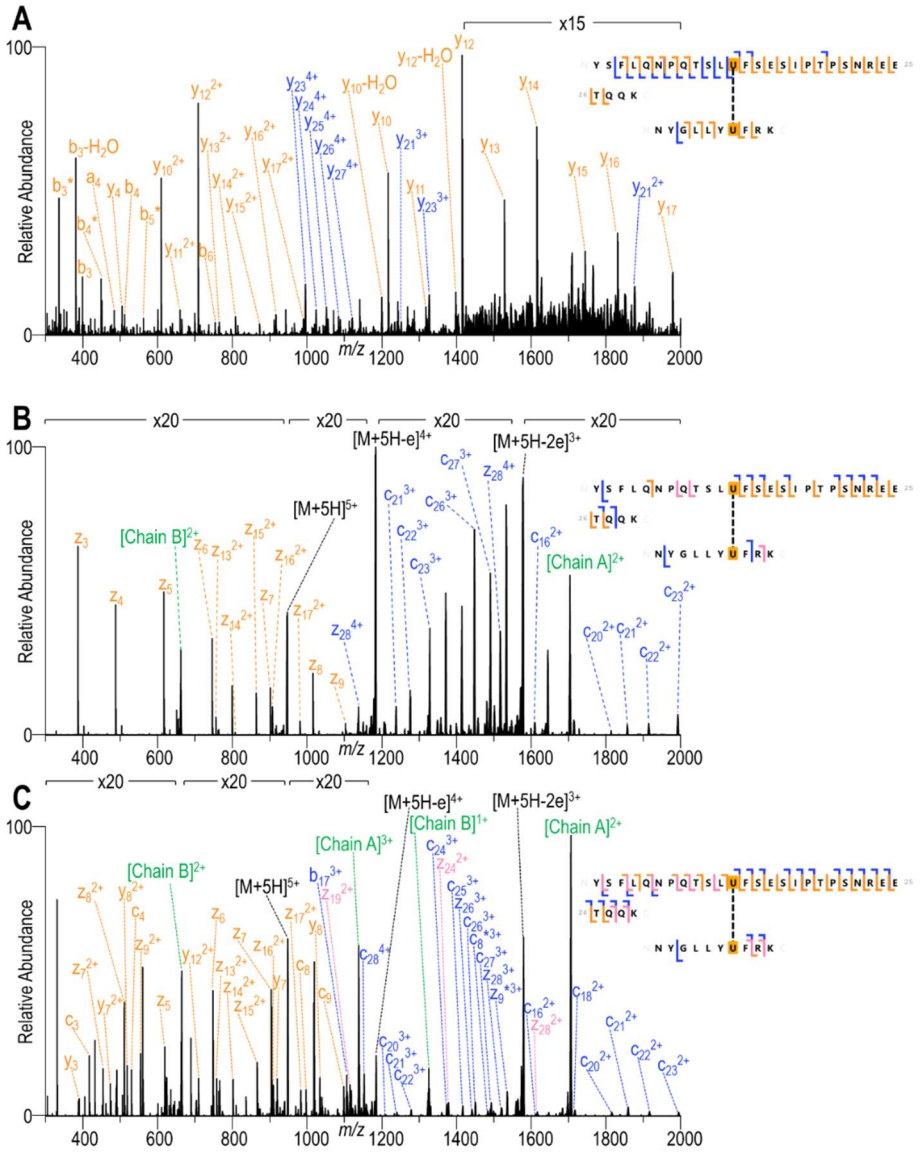


Figure 5.11: MS/MS spectra and sequence coverage maps for one selenocysteine-bridged peptide pair from HGH (5+) obtained using (A) HCD (30% NCE), (B) ETD (charge calibrated activation), and (C) EThcD (charge calibrated activation and 15% NCE supplemental activation), All spectra were collected with 120k resolution and 3 μscans, and 7-11 spectra were averaged. Only the most abundant fragment ions in each region are labeled. Fragment ions relating to free peptides after cleavage of diselenide bonds are marked as Chain A (larger peptide) and Chain B (smaller peptide). Any fragment ion labelled with a star originated from Chain B. The fragment ion coloring is defined in Scheme 5.3.

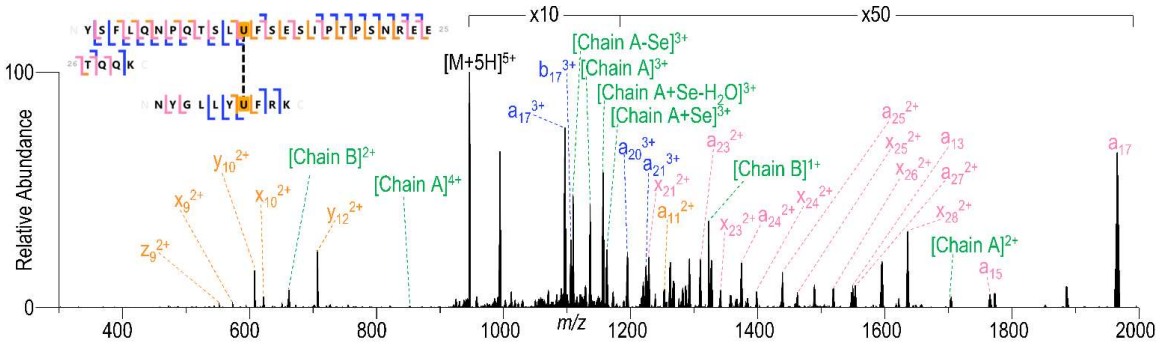


Figure 5.12: MS/MS spectrum and sequence coverage map for one selenocysteine-bridged peptide pair from HGH (5+) obtained using 213 nm UVPD (100 ms activation period). All spectra were collected with 120k resolution and 3 µscans, and 8 spectra were averaged. Only the most abundant fragment ions in each region are labeled. Fragment ions relating to free peptides after cleavage of diselenide bonds are marked as Chain A (larger peptide) and Chain B (smaller peptide). Any fragment ion labelled with a star originated from Chain B. The fragment ion coloring is defined in Scheme 5.3.

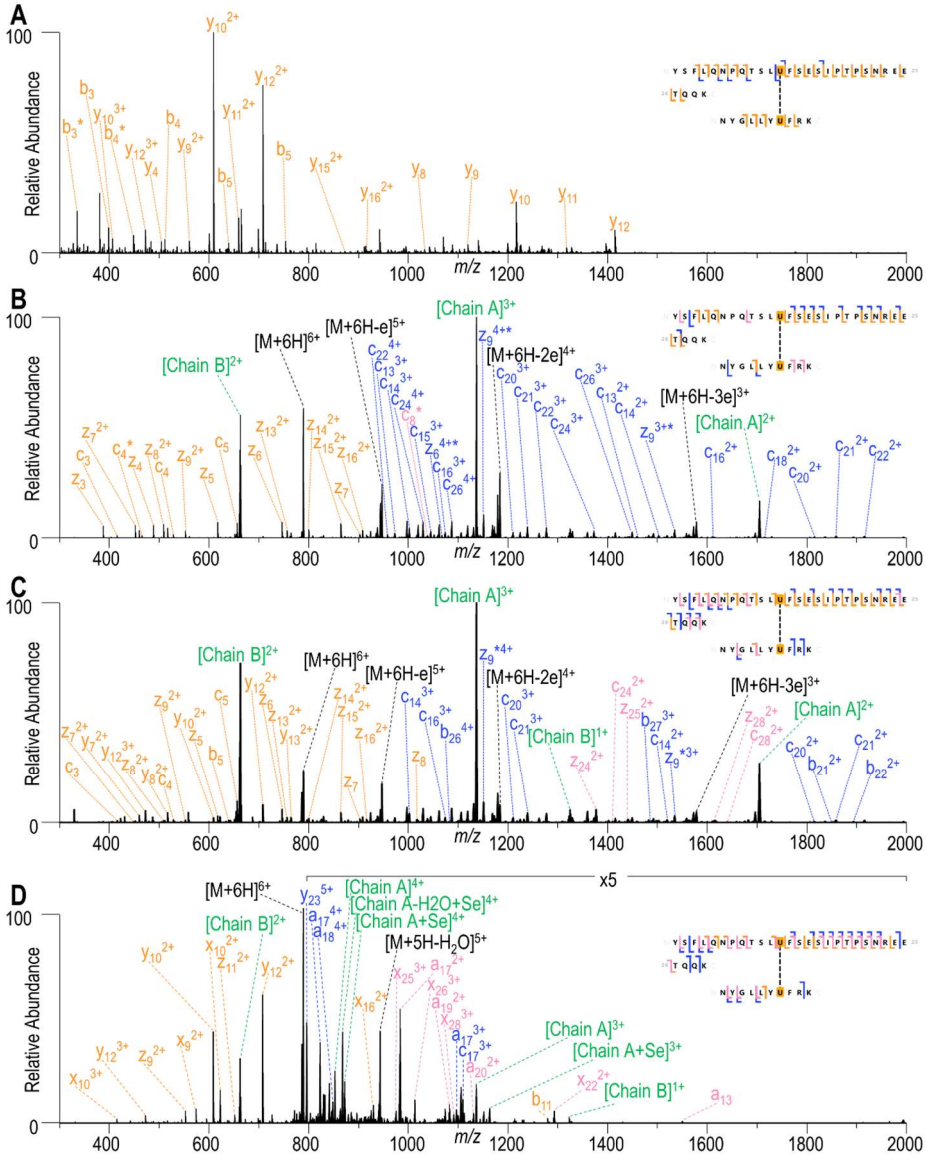


Figure 5.13: MS/MS spectra and sequence coverage maps for selenocysteine-bridged peptide pair from HGH (6+) obtained using (A) HCD (30% NCE), (B) ETD (charge calibrated activation), (C) EThcD (charge calibrated activation and 15% NCE supplemental activation), and (D) 213 nm UVPD (100 ms activation period). All spectra were collected with 120k resolution and 3 uscans, and 7-10 spectra were averaged. Only the most abundant fragment ions in each region are labeled. Fragment ions relating to free peptides after cleavage of diselenide bonds are marked as Chain A (larger peptide) and Chain B (smaller peptide). Any fragment ion labelled with a star originated from Chain B. The fragment ion coloring is defined in Scheme 5.3.

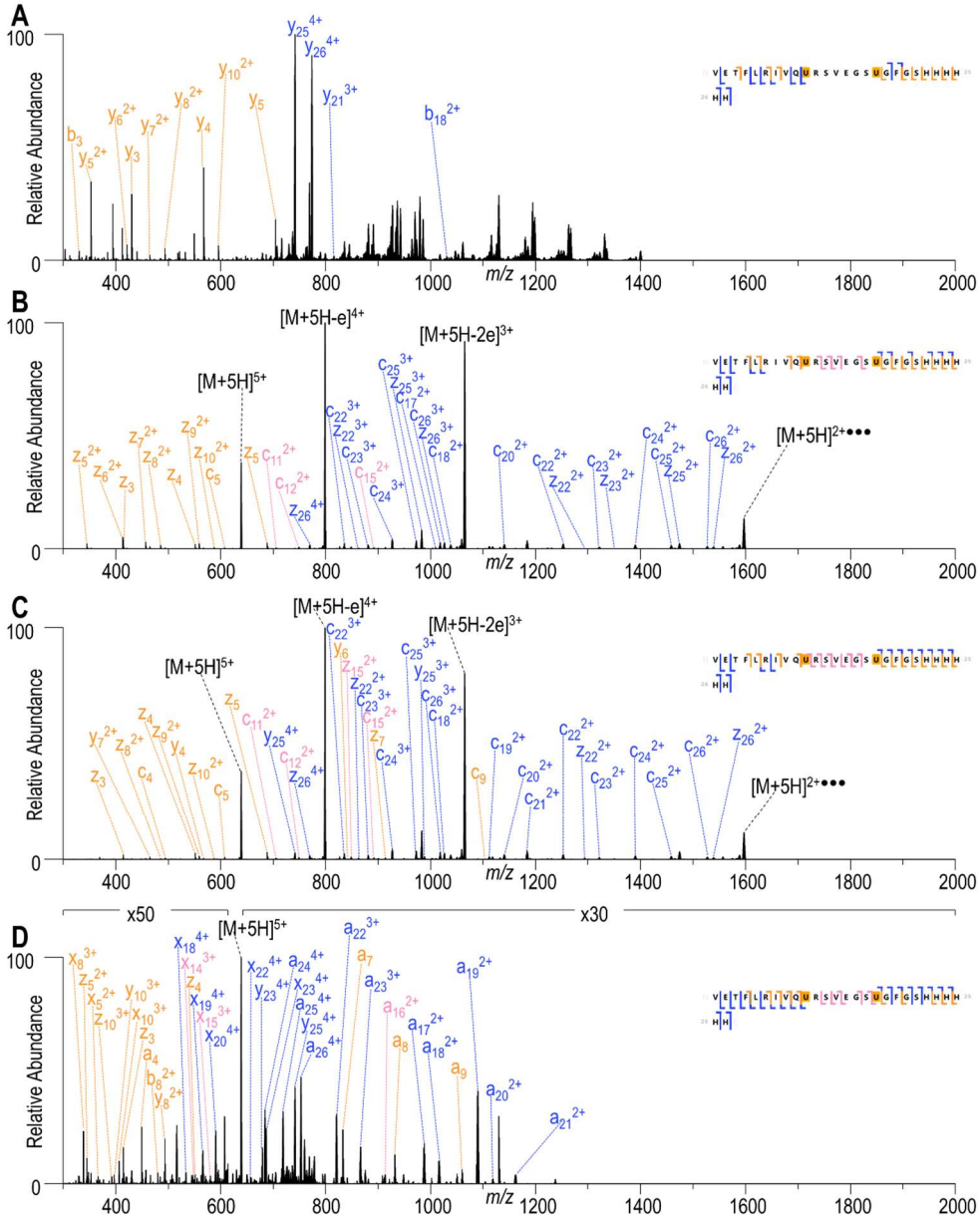


Figure 5.14: MS/MS spectra and sequence coverage maps for peptide with intra-peptide selenocysteine bridge from HGH (5+) obtained using (A) HCD (30% NCE), (B) ETD (charge calibrated activation), (C) EThcD (charge calibrated activation and 15% NCE supplemental activation), and (D) 213 nm UVPD (100 ms activation period). All spectra were collected with 120k resolution and 3 uscans, and 7-9 spectra were averaged. Only the most abundant fragment ions in each region are labeled. The fragment ion coloring is defined in Scheme 5.3.

The second predicted diselenide-bridged peptide from HGH, with an intrapeptide U183-U190 diselenide bridge, was also analyzed in the 5+ charge state (**Figure 5.14**). For this type of species, fragment ions originating from backbone cleavages between the two selenocysteine residues are only observed if the constraining diselenide bond is also cleaved, allowing separation of the two peptide partners. These homolytic cleavage fragment ions are crucial to obtaining complete characterization of the constrained peptide. Sequence ions resulting from concurrent homolytic cleavage and backbone cleavage were observed for ETD, EThcD, and 213 nm UVPD (but not HCD), and this second diselenide-bridged peptide is well-characterized as illustrated in the maps in **Figure 5.14**.

5.4.3 Characterization of Diselenide Bridge Heterogeneity

Assessing the best way to decipher diselenide-bridged peptides utilizing the predicted digestion products also allows more facile analysis of peptides containing unanticipated diselenide bridges. During the proteolytic digestion process used for bottom-up analysis, the possibility of diselenide scrambling must be considered because disulfide scrambling has been reported to occur during digestion.[2,46] The replacement of sulfur with diselenide for cysteine bridges, however, has repeatedly been show to lead to improved stability.[5,47,48] To mitigate diselenide scrambling, the digestion was performed at pH 6, which has previously been shown to limit disulfide scrambling.[2,46] After establishing the conditions to limit scrambling, the identification of unanticipated diselenide bonds is possible.

In addition to the successful characterization of the correctly linked peptides described in the previous section, a significant number of peptides containing serine in the place of selenocysteine and/or containing an incorrect diselenide bond were identified

in the HGH digest based on the database search. To confirm the diselenide linkages, EThcD and 213 nm UVPD mass spectra were examined for each match. For a match to be validated for a pair of peptides with interpeptide diselenide bonds, the EThcD spectra must display high abundance peptide partner ions resulting from homolytic cleavage of diselenide bonds. For a peptide containing an intrapeptide diselenide bond, the highest abundance ions in the EThcD spectra were expected to be consistent with charge reduction products (no change in mass). Because 213 nm UVPD tended to result in the greatest number of fragment ions and the highest sequence coverage, 213 nm UVPD mass spectra were also examined for each peptide match to confirm sequence identities. All confirmed peptides from selenocysteine-containing HGH are listed in **Table 5.1**. Most of the incorrectly paired diselenide bridges are clearly the result of the mis-incorporation of serine in the place of selenocysteine, as showcased in the extracted ion chromatograms (EIC) in **Figure 5.15** generated for all HGH peptides containing U183 and U190, which are expected to participate in an intrapeptide diselenide bridge. The highest abundance peptide, eluting at 27.73 min (highlighted in blue), corresponds to the predicted sequence and correct diselenide bridge. The next highest abundance peptide eluting at 26.82 min (highlighted in purple) originates from the incorporation of two serine resides in the place of each selenocysteine and thus has no diselenide bond. The presence of a peptide containing two serine residues is not surprising given the identification of a protein containing two serine replacements in the top-down MS1 spectrum for HGH (**Figure 5.2C**). The remaining four peaks (highlighted in pink and orange) in the EIC trace in **Figure 5.15** result from the replacement of either U183 or U190 with serine, allowing the other selenocysteine to pair with a different selenocysteine residue within the protein. No peptides were identified that indicated participation of U183 and U190 in an incorrect pairing unless either U183 or U190 was

replaced with serine. These results convey that incorrect selenocysteine pairing was due to the unintended incorporation of serine.

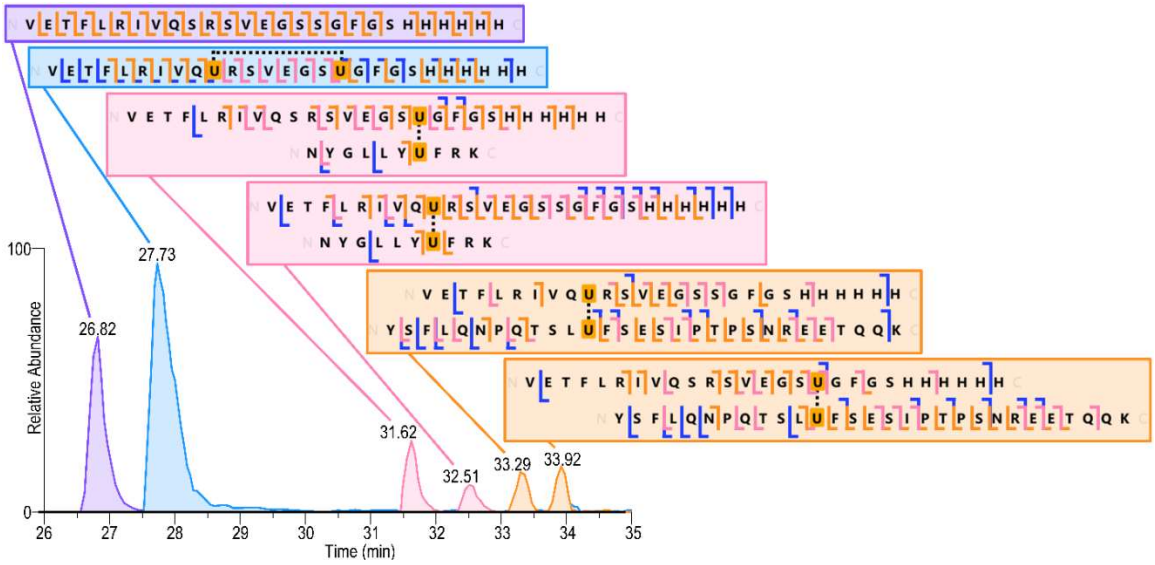


Figure 5.15: Extracted ion chromatograms for all identified peptides containing U183 and U190 with no missed cleavages from the HGH digest. The most abundant charge state from the MS1 spectrum acquired for each chromatographic peak is extracted for each peptide, and a sequence coverage map for each peptide is included based on the results from 213 nm UVPD. The peptides highlighted in pink and orange are pairs of isomers that elute separately and share extracted ion chromatograms.

In addition to the diselenide bridges illustrated in **Figure 5.15**, which could easily be explained by the misincorporation of serine within the observed peptides, there were a few other peptides identified that were consistent with more surprising diselenide bridges. Namely, the peptides containing U166-U166 and U54-U54 diselenide bridges listed in **Table 5.1** could only have resulted from the pairing of two separate HGH proteins. This was unexpected because an intact HGH dimer was not observed in the high-resolution top-down MS1 spectrum (**Figure 5.2C**). While one explanation is that these linkages resulted from diselenide scrambling during digestion, the low pH utilized during

digestion should have limited that possibility. An alternative explanation is that the dimer is present in the sample but was simply not detected in the high-resolution mass spectrum. In an Orbitrap analyzer, a higher mass ion, such as a dimer, may experience accelerated signal decay relative to a higher abundance monomer, a phenomenon resulting from more frequent gas collisions for the species with the larger size (i.e., greater collision cross-section).[49,50] This accelerated signal decay of the higher mass dimer may result in an artificially low signal abundance. This signal decay discrepancy is substantiated by collection of MS1 spectra using lower resolution by curtailing the time that the transient is acquired and truncating the signal decay. The lower resolution (R 15,000 at *m/z* 200) spectrum displayed in **Figure 5.16A** reveals the presence of an HGH dimer not observed in the higher resolution (R 120,000 at *m/z* 200) spectrum in **Figure 5.16B**. While the low-resolution data could not be deconvoluted with traditional software, deconvolution with Unidec was applied to reveal additional information about the dimer (**Figure 5.17**). The monomer population in the low-resolution mass spectrum was deconvoluted as a single species corresponding to the complete incorporation of four selenocysteine residues (4xU). The dimer population, however, was comprised of several species consistent with the incorporation of either two serine residues (6xU 2xS, nominally 46,700 Da) or four serine residues (4xU 4xS, nominally 46,570 Da) in place of selenocysteine. The fact that no HGH dimers were identified containing complete incorporation of selenocysteine (8xU 0xS) supports the idea that dimer formation is uniquely caused by the incorporation of at least one serine residue per protein. HGH dimers can occur naturally as a consequence of a genetic mutation resulting in the incorporation of serine in the place of C54, leading to bridging of unpaired C166 residues of two HGH molecules.[51] Given the greater electronegativity of selenium relative to sulfur,[8] it is unsurprising that a similar phenomenon would be observed in the

selenocysteine system. It can be reasoned that the HGH dimer containing two serine residues resulted from a partnership of a pair of HGH monomers, each with one serine replacement, leaving one free selenocysteine of each protein to form a dimer. Similarly, the dimer with four serine misincorporations likely formed from interaction of a protein containing one serine incorporation with a protein containing three serine incorporations, resulting in one free selenocysteine on each HGH molecule. The production of dimers owing to free (unpaired) selenocysteines also explains a phenomenon first observed in the high-resolution mass spectra in **Figure 5.2B,C.** The monomers were always observed either with complete selenocysteine incorporation, as expected, or with the misincorporation of two serine residues (low abundance species), but never with an odd number of serine residues. The spontaneous generation of dimers whenever a free selenocysteine residue is present would eliminate any monomeric proteins containing uneven numbers of selenocysteine residues.

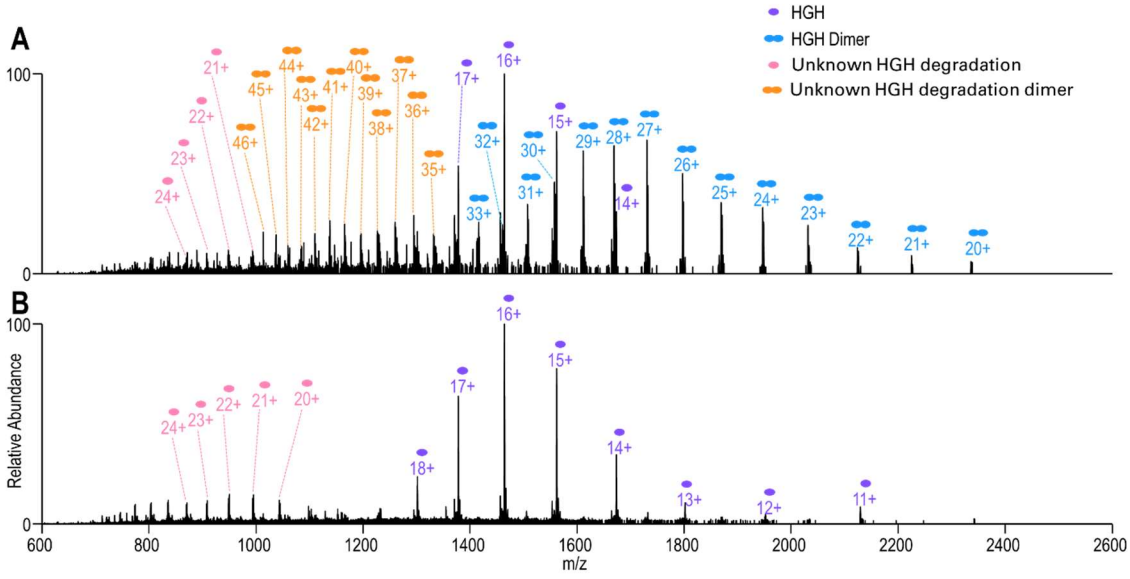
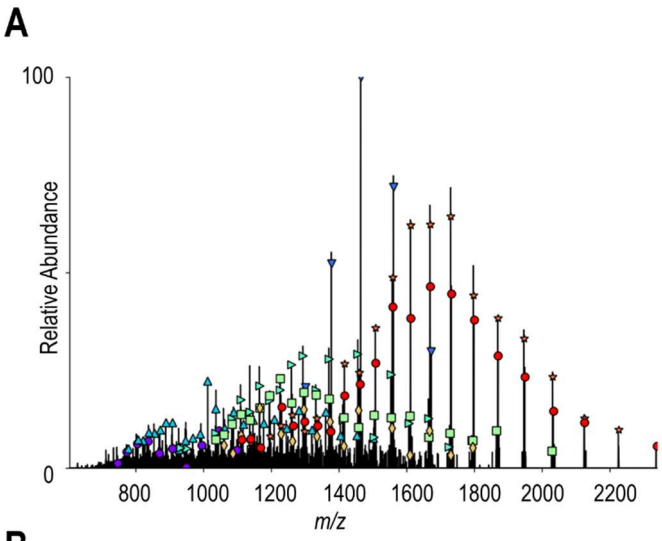
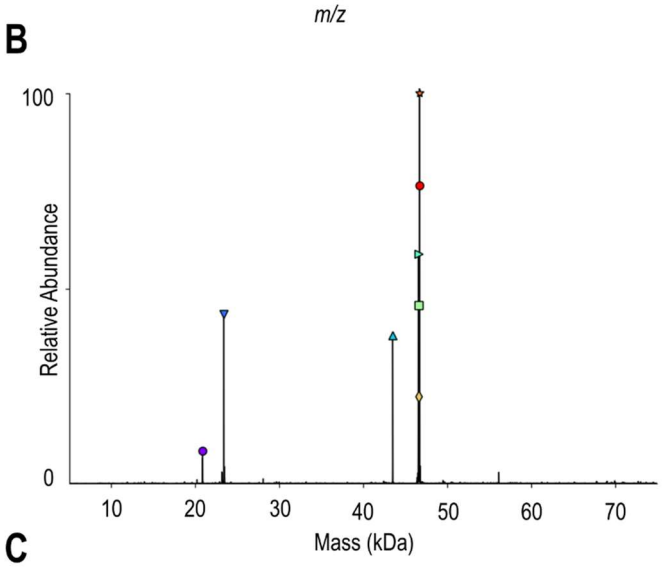


Figure 5.16: Mass spectra for HGH with a resolution of (A) 15,000 or (B) 120,000 at m/z 200.





C

	Observed Mass	Theoretical Mass
Unknown HGH degradation	20880	
HGH 4xU	23410	23410
Unknown HGH degradation dimer	43530	
HGH Dimer 4xU 4xS	46570	46570
HGH Dimer 4xU 4xS + K	46610	
HGH Dimer 4xU 4xS + 2K	46650	
HGH Dimer 6xU 2xS	46700	46700
HGH Dimer 6xU 2xS + K	46740	

Figure 5.17: Mass spectrum for HGH with a resolution of 15,000 at m/z 200 (A) processed and (B) deconvoluted with Unidec and (C) table listing the proteins identified.

Given evidence of dimer formation for HGH, the MS1 spectra of trastuzumab scFv and GFP proteins were re-acquired at lower resolution. Unlike HGH, dimers were not observed in the lower resolution data for GFP or trastuzumab scFv **Figure 5.18**). To further explore the possible incorporation of serine residues, the LysC digest of GFP was examined in greater detail (**Table 5.2** and **Figure 5.19**), revealing the presence of additional peptides consistent with the mis-incorporation of Ser (**Figure 5.20).** In addition to the correctly bridged pair of peptides, peptides resulting from the misincorporation of serine at both U134 and U176 were observed (retention times 23.88, 17.22 and 24.41 min). Additionally, peptides resulting from the pairing of U134-U134 and U176-U176, which could only be explained by the presence of a dimer, were identified (retention times 16.35 and 27.60 min). The fact that peptides originating from the GFP dimers were identified in the tryptic digest, but intact dimers could not be observed in the top-down MS1 spectra, suggests that detection of these low abundance intact mixed S/U proteins was suppressed in top-down analysis and supports the value of the bottom-up approach for identification of low abundance side-products.

Table 5.2: All peptides identified in bottom-up digest of GFP.

Sequence	Modifications	Paired Peptide (If interpeptide diselenide is present)	Calculated Mass (Da)	RT (min)
EDGUILGHK	U134 Ser		954.48	17.22
EDGUILGHK	U134-U176 Interpeptide Diselenide	IRHNIEDGSVULADHYQQNTPIGDGPVLLPDNHYLSTQSK	5508.48	23.88
EDGUILGHK	U134-U134 Interpeptide Diselenide	EDGUILGHK	2034.78	16.35
GIDFKEDGUILGHK	U134 Ser		1514.77	18.62
GIDFKEDGUILGHK	U134-U176 Interpeptide Diselenide	IRHNIEDGSVULADHYQQNTPIGDGPVLLPDNHYLSTQSK	6068.77	25.50
GIDFKEDGUILGHK	U134-U134 Interpeptide Diselenide	EDGUILGHK	2595.07	19.33
GIDFKEDGUILGHKLEYNYNSHNVYIMADK	U134 Ser		3469.66	26.25
IRHNIEDGSVULADHYQQNTPIGDGPVLLPDNHYLSTQSK	U176 Ser		4428.17	24.41
IRHNIEDGSVULADHYQQNTPIGDGPVLLPDNHYLSTQSK	U176-U176 Interpeptide Diselenide	IRHNIEDGSVULADHYQQNTPIGDGPVLLPDNHYLSTQSK	8982.18	27.64
IRHNIEDGSVULADHYQQNTPIGDGPVLLPDNHYLSTQSK	U176-C47 Interpeptide Se-S bond	FICTTGK	5258.46	24.41
IRHNIEDGSVULADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEK	U176 Ser		5339.64	23.65
NGIKVNFKIRHNIEDGSVULADHYQQNTPIGDGPVLLPDNHYLSTQSK	U176 Ser		5328.69	26.09
FEGDTLVNRIVLK			1502.85	21.96
FICTTGK			768.38	16.17
FSVSGEGEGDATYGK			1502.65	17.02
GEELFTGVVPILVELDGDVNGHK			2436.25	35.86
LEYNYNSHNVYIMADK			1972.9	20.75
LSKDPNEKRDHMVLLEFVTAAGITHDMDELYK			3714.84	33.26
QHDFFK			820.39	16.4

Table 5.2 Continued

QHDFFKSAMPEGYVQERTIFFEDDGYYK			3446.56	29.79
QKNGIKVNFK			1174.68	15.83
RDHMVLLEFVTAAGITHDMDELYK			2803.37	34.11
SAMPEGYVQERTIFFEDDGYYK			2644.18	27.86
SAMPEGYVQERTIFFEDDGYYKTRAEVKFEGDTLVNRIVLK			4813.41	30.74
TRAEVKFEGDTLVNRIVLK			2187.24	21.79
SKGEELFTGVVPILVELDGDVNGHK			2651.38	34.06

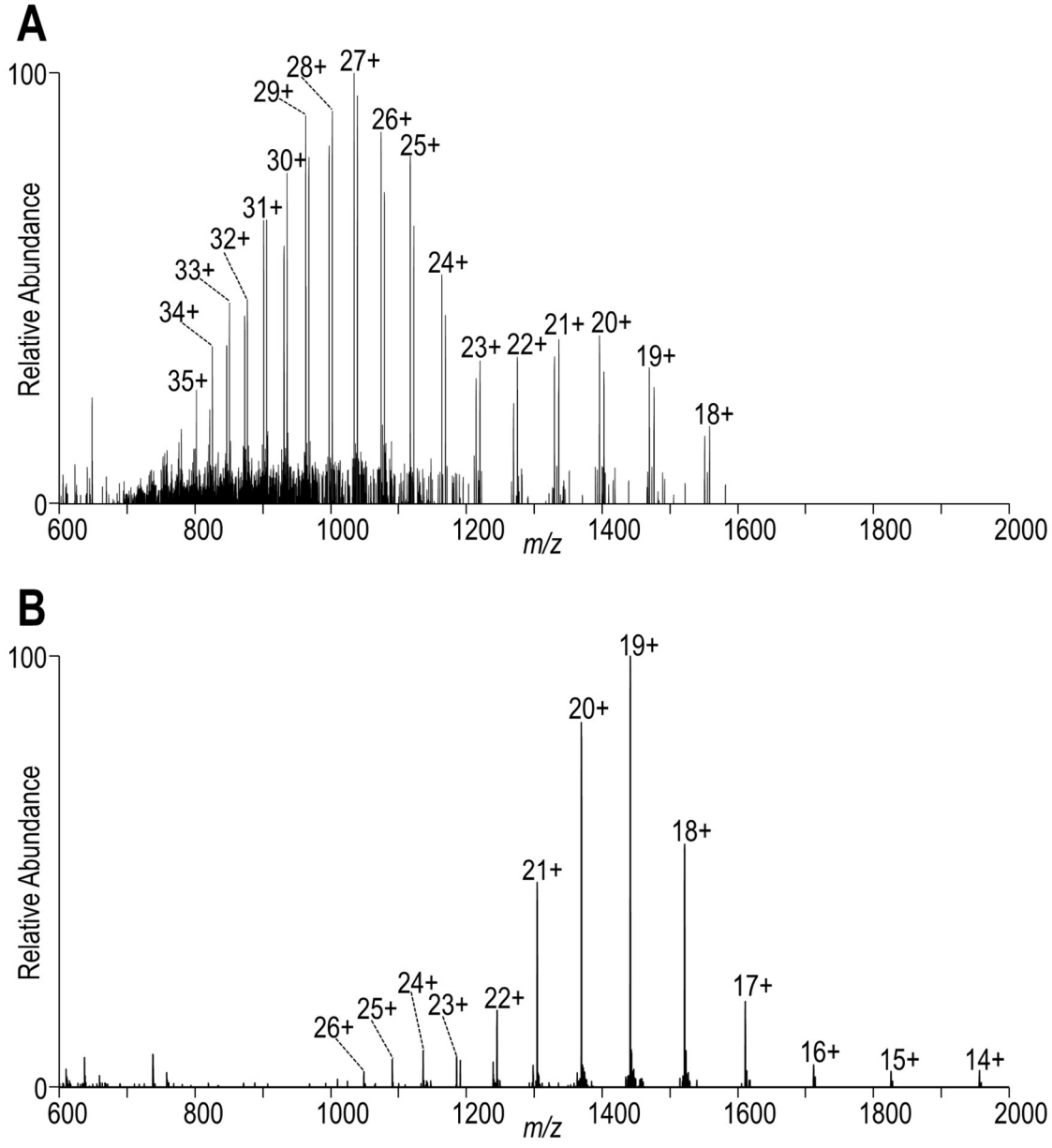


Figure 5.18: Low resolution (15,000 at m/z 200) mass spectra of (A) GFP and (B) trastuzumab ScV

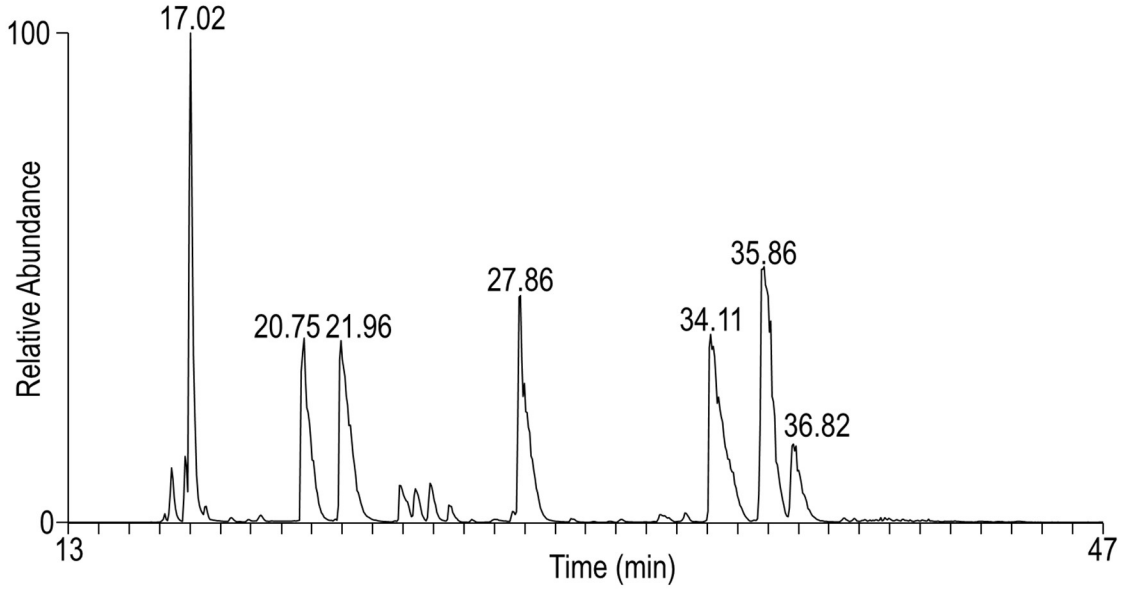


Figure 5.19: Base peak chromatogram of GFP digest

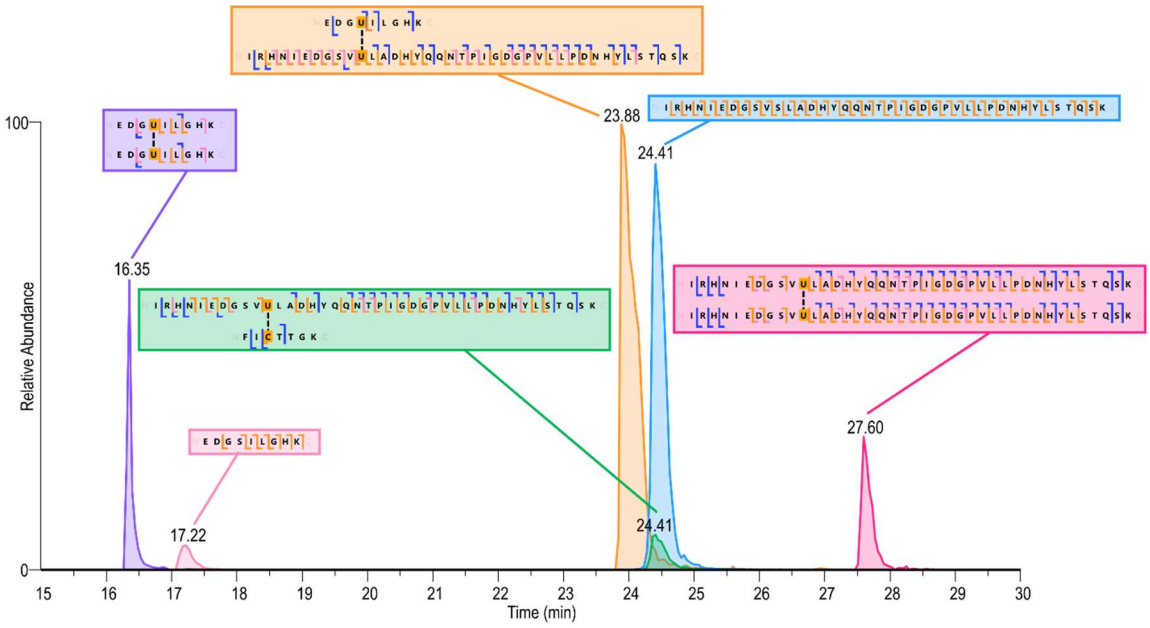


Figure 5.20: Extracted ion chromatograms for all identified peptides containing U134 and U176 with no missed cleavages from GFP digest. The most abundant charge state from the MS1 spectrum is extracted for each peptide. The peptide outlined in orange contains the correct diselenide bridge between U134 and U176. Fragment ions indicated on sequence coverage maps are the result of ProSight matches for 213 nm UVPD spectra deconvoluted with Xtract, except for the 954 Da peptide eluting at 17.22 min, which displays the fragment ions matched in Byonic from the 213 nm UVPD mass spectra. The fragment ion coloring is defined in Scheme 5.3.

5.5 Conclusion

Given the substantial potential for selenoproteins to aid in the development of biotherapeutics, it is critical to establish adequate characterization techniques. The combination of top-down and bottom-up mass spectrometry, presented herein, led to the complete characterization of customized selenoprotein constructs. While the bottom-up approach is critical for the identification of peptide pairings not identified with top-down analysis, the context of combinatorial modifications is lost. Additionally, when correctly paired peptides are identified, it is not possible to confirm they did not originate from a

dimer owing to the contextual information lost in the digestion process. With the top-down approach, it is possible to individually isolate the proteins with complete selenocysteine incorporation, allowing the corroboration of diselenide bridges in the proteins without serine misincorporation. The link between incorrect diselenide bridge formation and serine misincorporation suggests that improving the enzymatic incorporation of selenocysteine would eliminate the generation of incorrectly bridged proteins. In addition, it was gratifying to not observe any apparent mis-folded proteins with a full complement of selenocysteine residues. These results indicate that many complex selenoproteins can potentially be made by simple replacement of cysteine with selenocysteine. Furthermore, the presence of the dimeric species upon incorporation of a single serine residue presents the opportunity to purify selenoproteins with size-based separation techniques.

The combination of EThcD and 213 nm UVPD proved beneficial for comprehensive top-down characterization. EThcD cleaved diselenide bonds and yielded the highest sequence coverage for diselenide-bridged proteins. In a complementary manner, 213 nm UVPD resulted in many sequence ions without cleavage of diselenide bonds, allowing localization of diselenide bridges. In the bottom-up strategy, a combination of 213 nm UVPD and EThcD also offered the highest quality characterization of diselenide-bridged peptides given the unique ions generated by EThcD and the high sequence coverage achieved with 213 nm UVPD. Using these methods, the identification of both monomers with incomplete diselenide incorporation and dimers resulting from unpaired selenocysteine residues also sheds light on factors mediating the production of synthetic selenoproteins that could have key implications on

protein function. The integrated approach presented here offers a compelling strategy to advance the development of selenoproteins for biotherapeutic applications.

5.6 REFERENCES

(1) Feige, M. J.; Braakman, I.; Hendershot, L. M. CHAPTER 1.1:Disulfide Bonds in Protein Folding and Stability. In *Oxidative Folding of Proteins*; 2018; pp 1–33.

(2) Lakbub, J. C.; Shipman, J. T.; Desaire, H. Recent Mass Spectrometry-Based Techniques and Considerations for Disulfide Bond Characterization in Proteins. *Anal Bioanal Chem* **2018**, *410*, 2467–2484.

(3) Campuzano, I. D. G.; Sandoval, W. Denaturing and Native Mass Spectrometric Analytics for Biotherapeutic Drug Discovery Research: Historical, Current, and Future Personal Perspectives. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1861–1885.

(4) Weinfurtner, D. CHAPTER 1.4:Analysis of Disulfide Bond Formation in Therapeutic Proteins. In *Oxidative Folding of Proteins*; 2018; pp 81–98.

(5) Thyer, R.; Shroff, R.; Klein, D. R.; d'Oelsnitz, S.; Cotham, V. C.; Byrom, M.; Brodbelt, J. S.; Ellington, A. D. Custom Selenoprotein Production Enabled by Laboratory Evolution of Recoded Bacterial Strains. *Nat Biotechnol* **2018**, *36*, 624–631.

(6) Thyer, R.; Robotham, S. A.; Brodbelt, J. S.; Ellington, A. D. Evolving TRNASec for Efficient Canonical Incorporation of Selenocysteine. *J. Am. Chem. Soc.* **2015**, *137*, 46–49.

(7) Mousa, R.; Notis Dardashti, R.; Metanis, N. Selenium and Selenocysteine in Protein Chemistry. *Angewandte Chemie International Edition* **2017**, *56*, 15818–15827.

(8) Arnér, E. S. J. Selenoproteins—What Unique Properties Can Arise with Selenocysteine in Place of Cysteine? *Experimental Cell Research* **2010**, *316*, 1296–1303.

(9) Cardoso, B. R.; Ganio, K.; Roberts, B. R. Expanding beyond ICP-MS to Better Understand Selenium Biochemistry. *Metallomics* **2019**, *11*, 1974–1983.

(10) Thyer, R.; Filipovska, A.; Rackham, O. Engineered RRNA Enhances the Efficiency of Selenocysteine Incorporation during Translation. *J. Am. Chem. Soc.* **2013**, *135*, 2–5.

(11) Srzentić, K.; Nagornov, K. O.; Fornelli, L.; Lobas, A. A.; Ayoub, D.; Kozhinov, A. N.; Gasilova, N.; Menin, L.; Beck, A.; Gorshkov, M. V.; Aizikov, K.; Tsybin, Y. O. Multiplexed Middle-Down Mass Spectrometry as a Method for Revealing

Light and Heavy Chain Connectivity in a Monoclonal Antibody. *Anal. Chem.* **2018**, *90*, 12527–12535.

(12) Shaw, J. B.; Liu, W.; Vasil′ev, Y. V.; Bracken, C. C.; Malhan, N.; Guthals, A.; Beckman, J. S.; Voinov, V. G. Direct Determination of Antibody Chain Pairing by Top-down and Middle-down Mass Spectrometry Using Electron Capture Dissociation and Ultraviolet Photodissociation. *Anal. Chem.* **2020**, *92*, 766–773.

(13) Li, G.; Pei, J.; Yin, Y.; Huang, G. Direct Sequencing of a Disulfide-Linked Peptide with Electrospray Ionization Tandem Mass Spectrometry. *Analyst* **2015**, *140*, 2623–2627.

(14) Adhikari, S.; Yang, X.; Xia, Y. Acetone/Isopropanol Photoinitiating System Enables Tunable Disulfide Reduction and Disulfide Mapping via Tandem Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 13036–13043.

(15) Cramer, C. N.; Haselmann, K. F.; Olsen, J. V.; Nielsen, P. K. Disulfide Linkage Characterization of Disulfide Bond-Containing Proteins and Peptides by Reducing Electrochemistry and Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 1585–1592.

(16) Chin, S.; Chen, T.; Hannoush, R. N.; Crittenden, C. M. Tracking Internal and External Ions for Constrained Peptides Leads to Enhanced Sequence Coverage and Disulfide Bond Deciphering. *Journal of Pharmaceutical and Biomedical Analysis* **2021**, *195*, 113893.

(17) Schnaible, V.; Wefing, S.; Resemann, A.; Suckau, D.; Bücker, A.; Wolf-Kümmeth, S.; Hoffmann, D. Screening for Disulfide Bonds in Proteins by MALDI In-Source Decay and LIFT-TOF/TOF-MS. *Anal. Chem.* **2002**, *74*, 4980–4988.

(18) Massonnet, P.; Haler, J. R. N.; Upert, G.; Smargiasso, N.; Mourier, G.; Gilles, N.; Quinton, L.; De Pauw, E. Disulfide Connectivity Analysis of Peptides Bearing Two Intramolecular Disulfide Bonds Using MALDI In-Source Decay. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1995–2002.

(19) Huwiler, K. G.; Mosher, D. F.; Vestling, M. M. Optimizing the MALDI-TOF-MS Observation of Peptides Containing Disulfide Bonds. *J Biomol Tech* **2003**, *14*, 289–297.

(20) van der Burgt, Y. E. M.; Kilgour, D. P. A.; Tsybin, Y. O.; Srzentić, K.; Fornelli, L.; Beck, A.; Wuhrer, M.; Nicolardi, S. Structural Analysis of Monoclonal Antibodies by Ultrahigh Resolution MALDI In-Source Decay FT-ICR Mass Spectrometry. *Anal. Chem.* **2019**, *91*, 2079–2085.

(21) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal. Chem.* **2018**, *90*, 40–64.

(22) Rombouts, I.; Lagrain, B.; Scherf, K. A.; Lambrecht, M. A.; Koehler, P.; Delcour, J. A. Formation and Reshuffling of Disulfide Bonds in Bovine Serum Albumin Demonstrated Using Tandem Mass Spectrometry with Collision-Induced and Electron-Transfer Dissociation. *Sci Rep* **2015**, *5*, 12210.

(23) Bishop, A.; Brodbelt, J. S. Selective Cleavage upon ETD of Peptides Containing Disulfide or Nitrogen–Nitrogen Bonds. *International Journal of Mass Spectrometry* **2015**, *378*, 127–133.

(24) Liu, F.; Breukelen, B. van; Heck, A. J. R. Facilitating Protein Disulfide Mapping by a Combination of Pepsin Digestion, Electron Transfer Higher Energy Dissociation (EThcD), and a Dedicated Search Algorithm SlinkS *. *Molecular & Cellular Proteomics* **2014**, *13*, 2776–2786.

(25) Cole, S. R.; Ma, X.; Zhang, X.; Xia, Y. Electron Transfer Dissociation (ETD) of Peptides Containing Intrachain Disulfide Bonds. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 310–320.

(26) Clark, D. F.; Go, E. P.; Desaire, H. Simple Approach to Assign Disulfide Connectivity Using Extracted Ion Chromatograms of Electron Transfer Dissociation Spectra. *Anal. Chem.* **2013**, *85*, 1192–1199.

(27) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2006**, *1764*, 1811–1822.

(28) Zhang, L.; English, A. M.; Bai, D. L.; Ugrin, S. A.; Shabanowitz, J.; Ross, M. M.; Hunt, D. F.; Wang, W.-H. Analysis of Monoclonal Antibody Sequence and Post-Translational Modifications by Time-Controlled Proteolysis and Tandem Mass Spectrometry. *Mol. Cell. Proteom.* **2016**, *15*, 1479–1488.

(29) Quick, M. M.; Crittenden, C. M.; Rosenberg, J. A.; Brodbelt, J. S. Characterization of Disulfide Linkages in Proteins by 193 Nm Ultraviolet Photodissociation (UVPD) Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 8523–8530.

(30) Agarwal, A.; Diedrich, J. K.; Julian, R. R. Direct Elucidation of Disulfide Bond Partners Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2011**, *83*, 6455–6458.

(31) Wongkongkathep, P.; Li, H.; Zhang, X.; Ogorzalek Loo, R. R.; Julian, R. R.; Loo, J. A. Enhancing Protein Disulfide Bond Cleavage by UV Excitation and Electron Capture Dissociation for Top-down Mass Spectrometry. *International Journal of Mass Spectrometry* **2015**, *390*, 137–145.

(32) Siuti, N.; Kelleher, N. L. Decoding Protein Modifications Using Top-down Mass Spectrometry. *Nat Methods* **2007**, *4*, 817–821.

(33) Chen, B.; Brown, K. A.; Lin, Z.; Ge, Y. Top-Down Proteomics: Ready for Prime Time? *Anal. Chem.* **2018**, *90*, 110–127.

(34) Ganisl, B.; Breuker, K. Does Electron Capture Dissociation Cleave Protein Disulfide Bonds? *ChemistryOpen* **2012**, *1*, 260–268.

(35) Mao, Y.; Valeja, S. G.; Rouse, J. C.; Hendrickson, C. L.; Marshall, A. G. Top-Down Structural Analysis of an Intact Monoclonal Antibody by Electron Capture Dissociation-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 4239–4246.

(36) Fornelli, L.; Parra, J.; Hartmer, R.; Stoermer, C.; Lubeck, M.; Tsybin, Y. O. Top-down Analysis of 30–80 KDa Proteins by Electron Transfer Dissociation Time-of-Flight Mass Spectrometry. *Anal Bioanal Chem* **2013**, *405*, 8505–8514.

(37) Zhang, J.; Loo, R. R. O.; Loo, J. A. Increasing Fragmentation of Disulfide-Bonded Proteins for Top–down Mass Spectrometry by Supercharging. *International Journal of Mass Spectrometry* **2015**, *377*, 546–556.

(38) Greisch, J.-F.; den Boer, M. A.; Beurskens, F.; Schuurman, J.; Tamara, S.; Bondt, A.; Heck, A. J. R. Generating Informative Sequence Tags from Antigen-Binding Regions of Heavily Glycosylated IgA1 Antibodies by Native Top-Down Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1326–1335.

(39) Lodge, J. M.; Schauer, K. L.; Brademan, D. R.; Riley, N. M.; Shishkova, E.; Westphall, M. S.; Coon, J. J. Top-Down Characterization of an Intact Monoclonal Antibody Using Activated Ion Electron Transfer Dissociation. *Anal. Chem.* **2020**, *92*, 10246–10251.

(40) Rush, M. J. P.; Riley, N. M.; Westphall, M. S.; Coon, J. J. Top-Down Characterization of Proteins with Intact Disulfide Bonds Using Activated-Ion Electron Transfer Dissociation. *Anal. Chem.* **2018**, *90*, 8946–8953.

(41) Gammelgaard, S. K.; Petersen, S. B.; Haselmann, K. F.; Nielsen, P. K. Characterization of Insulin Dimers by Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1910–1918.

(42) Lamarche, J.; Ronga, L.; Szpunar, J.; Lobinski, R. Characterization and Quantification of Selenoprotein P: Challenges to Mass Spectrometry. *International Journal of Molecular Sciences* **2021**, *22*, 6283.

(43) Ma, S.; Hill, K. E.; Burk, R. F.; Caprioli, R. M. Mass Spectrometric Identification of N- and O-Glycosylation Sites of Full-Length Rat Selenoprotein P and Determination of Selenide−Sulfide and Disulfide Linkages in the Shortest Isoform. *Biochemistry* **2003**, *42*, 9703–9711.

(44) Ma, S.; Hill, K. E.; Burk, R. F.; Caprioli, R. M. Mass Spectrometric Determination of Selenenylsulfide Linkages in Rat Selenoprotein P. *Journal of Mass Spectrometry* **2005**, *40*, 400–404.

(45) Wingfield, P. N-Terminal Methionine Processing. *Curr Protoc Protein Sci* **2017**, *88*, 6.14.1-6.14.3.

(46) Sung, W.-C.; Chang, C.-W.; Huang, S.-Y.; Wei, T.-Y.; Huang, Y.-L.; Lin, Y.-H.; Chen, H.-M.; Chen, S.-F. Evaluation of Disulfide Scrambling during the Enzymatic Digestion of Bevacizumab at Various PH Values Using Mass Spectrometry. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2016**, *1864*, 1188–1194.

(47) Mousa, R.; Hidmi, T.; Pomyalov, S.; Lansky, S.; Khouri, L.; Shalev, D. E.; Shoham, G.; Metanis, N. Diselenide Crosslinks for Enhanced and Simplified Oxidative Protein Folding. *Commun Chem* **2021**, *4*, 1–9.

(48) Shimodaira, S.; Asano, Y.; Arai, K.; Iwaoka, M. Selenoglutathione Diselenide: Unique Redox Reactions in the GPx-Like Catalytic Cycle and Repairing of Disulfide Bonds in Scrambled Protein. *Biochemistry* **2017**, *56*, 5644–5653.

(49) Makarov, A.; Denisov, E. Dynamics of Ions of Intact Proteins in the Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1486–1495.

(50) Sanders, J. D.; Grinfeld, D.; Aizikov, K.; Makarov, A.; Holden, D. D.; Brodbelt, J. S. Determination of Collision Cross-Sections of Protein Ions in an Orbitrap Mass Analyzer. *Anal. Chem.* **2018**, *90*, 5896–5902.

(51) Sander, M.; Wu, Z.; Strasburger, C. J. Short Stature Explained by Dimerization of Human Growth Hormone Induced by a p.C53S Point Mutation. *Journal of Biological Chemistry* **2020**, *295*, 4893–4901.

Chapter 6: Characterization of HLA-A*02:01 MHC Immunopeptide Antigens Enhanced by Ultraviolet Photodissociation Mass Spectrometry[4]

6.1 Abstract

Identifying major histocompatibility complex (MHC) class I immunopeptide antigens represents a key step in the development of immune-based targeted therapeutics and vaccines. However, the complete characterization of these antigens by tandem mass spectrometry remains challenging due to their short sequence length, high degree of hydrophobicity, and/or lack of sufficiently basic amino acids. This study seeks to address the potential for 193 nm ultraviolet photodissociation (UVPD) to improve the analysis of MHC class I immunopeptides by offering enhanced characterization of these sequences in lower charge states and differentiation of prominent isomeric leucine and isoleucine residues in the HLA-A*02:01 motif. While electron transfer dissociation-higher energy collisional dissociation (EThcD) offered some success in the differentiation of leucine and isoleucine, 193 nm UVPD was able to confirm the identity of nearly 60% of leucine and isoleucine residues in a synthetic peptide mixture. Furthermore, 193 nm UVPD led to significantly more peptide identifications and higher scoring metrics than EThcD for peptides obtained from immunoprecipitation of MHC class I immunopeptides from *in vitro* cell culture. Additionally, 193 nm UVPD represented a promising complementary technique to HCD, in which 424 of the 2593 peptides identified by 193 nm UVPD were not identified by HCD in a HLA-A*02:01-specific immunoprecipitation and 804 of the 3300 peptides identified by 193 nm UVPD were not identified by HCD for a broad HLA-

[4]Watts, E.; Potts, G. K.; Ready, D. B.; George Thompson, A. M.; Lee, J.; Escobar, E. E.; Patterson, M. J.; Brodbelt, J. S. Characterization of HLA-A*02:01 MHC Immunopeptide Antigens Enhanced by Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2021, 93, 13134–13142.

A, -B, and -C immunoprecipitation. These results highlight that 193 nm UVPD offers an option for the characterization of immunopeptides, including robust differentiation of leucine and isoleucine residues.

6.2 Introduction

Therapies that augment the immune system offer the potential for more specialized approaches for the treatment of cancer than chemotherapy and radiation.[1,2] Peptide vaccines or T-cell receptor (TCR) based therapies are developed to account for the complex processes constituting the immune response. The major histocompatibility complex (MHC), also called the human leukocyte antigen (HLA),[2] presents immunopeptides derived from the degradation of cytosolic proteins onto the cell surface to communicate the status of the cell to the immune system. These peptides originate from the proteasomal degradation of intracellular proteins.[3] Some of the resulting peptides, determined by the binding groove motif requirements within the HLA protein, are bound as a non-covalent complex and presented on the surface of the cell.[2,4] The presentation of a foreign immunopeptide antigen can lead to recognition by $CD8^+$ T cells, resulting in cell death.[3] In cancerous cells, expression of tumor-specific antigens or overexpression of self-antigens can occur,[1,5,6] offering opportunities to identify immunopeptides for targeted therapy.[1,7] An immunopeptide-based approach has been previously demonstrated to facilitate the generation of vaccines by focusing on the epitope regions of an antigen.[8]

While the development of such therapeutic approaches has been advanced by mass spectrometric identification of epitopes and MHC-binding peptides, characterization of immunopeptides remains challenging.[9] MHC class I peptides are

typically short, ranging from 8 to 10 amino acids in length, owing to selective cleavage by endoplasmic reticulum aminopeptidase 1 after proteasome degradation.[10,11] The short length of immunopeptides necessitates nearly complete sequence coverage for confident identification. The MHC class I serotype HLA-A*02:01 is commonly studied because it is one of the more prevalent alleles in the human population.[12,13] This motif tends to contain many hydrophobic residues, particularly prevalent at positions 2 and 9.[14] Another key feature of HLA-A*02:01 peptides is the incorporation of multiple leucine and isoleucine residues, typically between two and five per peptide.[14] These isomeric residues are not commonly characterized in routine MS/MS analysis, leading to an additional hurdle in the comprehensive analysis of immunopeptides.

Owing to the broad importance of immunotherapies, mass spectrometry-based immunopeptidomics has been a growing field for the past 30 years.[15] Peptides are commonly characterized using collisionally-activated dissociation (CAD) or higher-energy collisional dissociation (HCD).[2,16–20] Electron transfer dissociation (ETD) as well as the hybrid method electron-transfer/higher energy collisional dissociation (EThcD) have also been explored for analysis of immunopeptides.[21–25] EThcD methods have been shown to increase the number of unique peptide matches when combined with CAD.[21,22] Moreover, EThcD has been shown to generate *w* ions which allow differentiation between leucine and isoleucine.[26,27] ETD has also been applied to the characterization of immunopeptides containing post-translational modifications (PTMs) not successfully identified by HCD.[24,25] While these methods have been promising, ETD and EThcD remain limited in their ability to characterize the broad spectrum of immunopeptides owing to the charge state dependence of these methods. ETD and EThcD exhibit their

best performance metrics for the identification of longer and more highly charged immunopeptides.[22]

CAD has also been implemented to differentiate leucine and isoleucine, typically via an MS^3 strategy in which the 86 Da immonium ion, specific to either leucine or isoleucine, is analyzed.[28–31] This basic MS^3 method is not successful in characterizing peptides for which multiple leucine or isoleucine residues are present, as is the case for many immunopeptides. Some methods have addressed this limitation; however, they require either more complex MS^n staging or derivatization of the peptides.[28,31] Additionally, while *d* and *w* ions were first observed with high energy (keV) CAD,[32] the production of *d* and/or *w* ions via UVPD or EThcD is more readily accomplished using modern high performance mass spectrometers.

Ultraviolet photodissociation (UVPD) has garnered growing attention as an alternative fragmentation method for peptides and proteins.[33–41] Two hallmarks of UVPD are the generation of a wide variety of fragment ions, including but not limited to *a, b, c, d, x, y, z,* and *w* ions, and the retention of post-translational modifications.[41–44] The variety of fragment ions typically leads to an increased number of identified fragment ions, resulting in higher sequence coverage. Another feature of UVPD is the lack of significant charge state dependence compared to other fragmentation techniques.[34,45] UVPD has been shown to generate both *d* and *w* ions from side-chain losses, each affording diagnostic neutral losses, which facilitate differentiation of leucine and isoleucine.[34,46] Moreover, the production of both *d* and *w* ions by UVPD offers the opportunity for bi-directional cleavages that can double the number of potential fragment ions, augmenting successful leucine and isoleucine differentiation. By capitalizing on

these features, UVPD has the potential to improve the number and quality of identifications from immunopeptide analysis.

The present study examines the capabilities of 193 nm UVPD to increase the number of identifications of the HLA-A*02:01 specific peptides motif as well as the assessment of UVPD for analysis of the full HLA-A, -B, and -C allelic repertoire. The evaluation of synthetic peptide standards is conducted for method development as well as to benchmark comparisons between several fragmentation techniques. Finally, MHC peptides isolated from cells are examined to compare the number and quality of matches obtained by HCD, EThcD, and 193 nm UVPD.

6.3 EXPERIMENTAL

6.3.1 Synthetic Peptide Standards

The synthetic peptide mixture matching the HLA-A*02:01 motif was designed and provided by AbbVie. Synthetic peptides of at least 85% purity were acquired from Genscript and reconstituted in dimethyl sulfoxide (DMSO) and pooled for the 157 peptide mixture. Prior to LC-MS analysis, all peptide solutions were diluted to 6.4 nM in 10% acetic acid (Fisher Scientific)/90% water (VWR). A total of 6.4 fmol of each synthetic peptide or peptide mixture was injected for LC-MS analysis.

6.3.2 Immunoprecipitated HLA Samples

Immunoprecipitated samples were prepared by Alayna George Thompson and Janice Lee (AbbVie). BV173 (HLA-A*02:01, -A*30:01, -B*18:01, -B*15:01, -C*12:03*,-C*03:04) cells (DSMZ) were utilized. Immunoprecipitation was completed following procedures previously reported.[47,48] Clarified cell lysates went through two rounds of immunoprecipitation, first using BB7.2 antibodies to select immunopeptides

matching the motif of HLA-A*02:01.[49] Flow through from BB7.2 went through a second round of immunoprecipitation with W6/32 antibodies to isolate pan-HLA (HLA-A, -B, and -C) peptides.[49] HLA peptide samples from each immunoprecipitation set were pooled, dried, and reconstituted in 10% acetic acid for LC-MS analysis.

6.3.3 Liquid Chromatography

Liquid chromatography (LC) was performed using a Dionex Ultimate nanoLC system. Trap and analytical columns were packed in-house to a length of 3.5 cm and 20 cm, respectively, with UChrom C18 packing material. Five-micron packing material was used for the trap column. Three micron and 1.8 micron packing materials were used for the analytical columns employed for the synthetic peptide mixture and immunoprecipitation samples, respectively. All UChrom bulk C18 packing material (nanoLCMS solutions) had a 120 Å pore size. Forward trapping was employed for all samples. A five microliter/minute gradient was employed for the trapping column along with a 0.300 microliter/minute for the analytical. For all analyses, analytical mobile phases A and B were composed of water with 0.1% formic acid and acetonitrile (VWR) with 0.1% formic acid, respectively. The mobile phase applied to the trapping column was held at 98% A and 2% B

For all LC methods, the solvent applied to the analytical column was initially held at 2% mobile phase B for a 5-minute equilibration period while the sample was loaded to the trapping column. The gradients began after a valve switch. Similarly, at the end of each method, the column was flushed for five minutes with 90% acetonitrile, followed by a ten-minute equilibration using the initial solvent conditions. Sawtooth washes were run between each injection of immunoprecipitated samples. A 2% to 40% mobile phase B

gradient was employed for both the 157-peptide mixture and the immunoprecipitated samples. The gradient was applied for two and three hours for the synthetic peptides and the immunoprecipitated samples, respectively.

6.3.4 Mass Spectrometry

All LC-MS experiments were performed using a Thermo Orbitrap™ Fusion Lumos Tribrid mass spectrometer equipped with a 193 nm excimer laser as previously described.[50] 193 nm UVPD was performed in the high-pressure linear ion trap. For MS1 scans, a resolution of 60,000 at 200 *m/z* and automated gain control (AGC) target of 5e5 were employed. Quadrupole isolation was utilized with a 1.2 *m/z* isolation window. Charge states from 1+ to 3+ were included for HCD and UVPD methods, and 2+ and 3+ were included for EThcD. *M/z* filters were applied for all charge states such that only ions between *m/z* 700-2000, *m/z* 350-1000, and *m/z* 200-500 Da were included for the 1+, 2+ and 3+ charge states, respectively. Dynamic exclusion was applied with a top speed three second cycle, with two repeats within seven seconds and a 30 second exclusion duration. Across all MS/MS scans the resolution was set to 60,000 at 200 *m/*z, the AGC target to 5e5, and one microscan was collected. For HCD, 21% normalized collision energy (NCE) was applied. EThcD was applied for 40 ms with 15% NCE supplemental activation. 193 nm UVPD was performed using four pulses (in an 8 ms period) with 3 mJ pulse energy.

6.3.5 Data Analysis

Data was analyzed using Byonic (Protein Metrics version 3.11.3) and searched against the Uniprot Swiss-Prot human proteome (2021_02 release) with nonspecific cleavage. Oxidation of methionine and cysteinylation of cystine were set as variable

modifications. Precursor and fragment tolerance were both set to 10 ppm. The Byonic protein level FDR was set to "no cuts", and a 1% false discovery rate (FDR) cutoff post-processing was applied at a peptide-spectral match level after sorting the hits by PEP 1D, as recommended by the vendor. All datasets were further refined by collecting samples in triplicate and only including peptides identified in all three replicates. Byonic was also used to facilitate *d* and *w* ion identifications. Each matching fragment ion was limited to those within 10 ppm error of the mass of the theoretical fragment ion and with a signal to noise greater than 3. To further verify these results, *d* and *w* ions were only reported if they were identified in two of the three replicate runs. Hydrophobicity indexes of peptides identified in W6/32 were calculated using the Sequence Specific Retention Calculator (SSRCalc) version Q from the Manitoba Centre for Proteomics and Systems Biology.[51] All spectra are archived and available at: https://repository.jpostdb.org/, and accession numbers are PXD027830 for ProteomeXchange and JPST001269 for jPOST.[52]

6.4 DISCUSSION

6.4.1 Optimization of MS/MS Activation Parameters

A set of 157 HLA-A*02:01 immunopeptides selected from the immune epitope database[53] (IEDB) was synthesized to use as standards for evaluation and optimization of the three MS/MS methods, HCD, EThcD, and 193 nm UVPD. A list of the peptides selected for the mixture are included in the appendix. Key activation parameters were examined for 193 nm UVPD, HCD, and EThcD, as illustrated in **Figures 6.1-6.3**.Each ion activation method was optimized based on the Byonic scores for each peptide matched, the sequence coverage, and the fragment ion types generated, especially the number of *d* and *w* ions for UVPD and EThcD. The sequence coverages are calculated by

Byonic as a part peptide spectral matches (PSM) scoring. This information was retained by running Byonic with debug enabled and then compiled from multiple files into a single table using an R program written in-house. Byonic includes *a, b*, and *y* ions when calculating sequence coverage for UVPD and HCD and *a, b, c, y*, and *z* ions for EThcD. After examining the Byonic scores and sequence coverages for all conditions considered, a handful of top conditions were selected, and the ion types identified in the Byonic output and examined. Only 18 peptides, ones consistently well-matched across all techniques and conditions, were included owing to the time-consuming nature of compiling the ion types (done manually). In the ion type tables, all *d* and *w* ions only represent those associated with diagnostic neutral losses leading to the specific identification of a leucine or isoleucine residue. The highest scoring PSM was used for the ion type figures.

193 nm UVPD conditions were optimized for laser pulse energy and the number of pulses, HCD for the normalized collision energy (%NCE), and EThcD for the activation time. **Figure 6.1** summarizes the results for the optimization of the laser pulse energy and the number of pulses for 193 nm UVPD. While there were no significant improvements in the sequence coverages above one pulse with a 2 mJ laser energy, the Byonic scores continued to increase as the number of pulses and laser energy increased. Ordinarily, the lack of increase in sequence coverage would be taken as an indicator that no additional informative fragment ions are being identified, and the continued increase in score is explained by an increase in the fragment ion abundances. In this case however, the *d* and *w* ions were of particular interest due to the goal of localizing leucine and isoleucine residues. By examining **Figure 6.1C**, it is observed that while the number of *a, b*, and *y* ion types that contribute to the Byonic calculation of sequence coverage did not

increase significantly as the laser energy and the number of pulses increased, the number of *c, x*, and *z* ions, as well as the diagnostic *d* and *w* ions, continued to increase. While these particular fragment ions do not contribute to the sequence coverage or to the Byonic score, they are produced from higher energy fragmentation pathways, and it follows that higher laser energies and pulse counts might be necessary to access them. For this reason, four pulses 3 mJ were selected for UVPD. While the identification of these additional *d/w* fragment ions was critical for the differentiation of leucine and isoleucine, it may not be sufficient reason to add these additional ion types to the Byonic scoring algorithm. Searches based on the *a, b*, and *y* types yielded the largest number of matches and therefore seem an appropriate choice when trying to generate an overall score for peptide correctness that balances identifying as many PSMs as possible without increasing the FDR by considering too many ion types of low frequency.

For HCD, examination of the Byonic scores and sequence coverage led to the adoption of 21% NCE. As illustrated in **Figure 6.2**, a dramatic increase in sequence coverage was observed when NCE was increased from 18% to 21%, and both the Byonic scores and sequence coverages decreased with higher NCE. While 21% NCE is low compared to the NCE often used for immunopeptide analysis, the specific NCE value can vary significantly depending on the instrument used. Previous work done using a Thermo Orbitrap™ Fusion Lumos Tribrid mass spectrometer similar to the one utilized here reported that an NCE as low as 17% can be optimal when analyzing HLA-A*02:01 immunopeptides.[1] **Figure 6.3** includes the optimization of ETD reaction time for EThcD, segregated based on peptides for which the highest-scoring PSM was either from the 2+ or 3+ charge state. This charge segregation was done because the high charge state dependence of EThcD is well established. Overall, for a range of activation times,

relatively modest changes were observed in both the Byonic scores and sequence coverages. There were no dramatic differences between the findings for the 2+ and 3+ charge states, and a small decrease in score was observed for both charge states when increasing the reaction time from 40 ms to 50 ms. When examining the fragment ions identified in **Figure 6.3C**, slightly more diagnostic *w* ions were identified using the 40 ms activation time. For this reason, the 40 ms activation time was selected for EThcD.

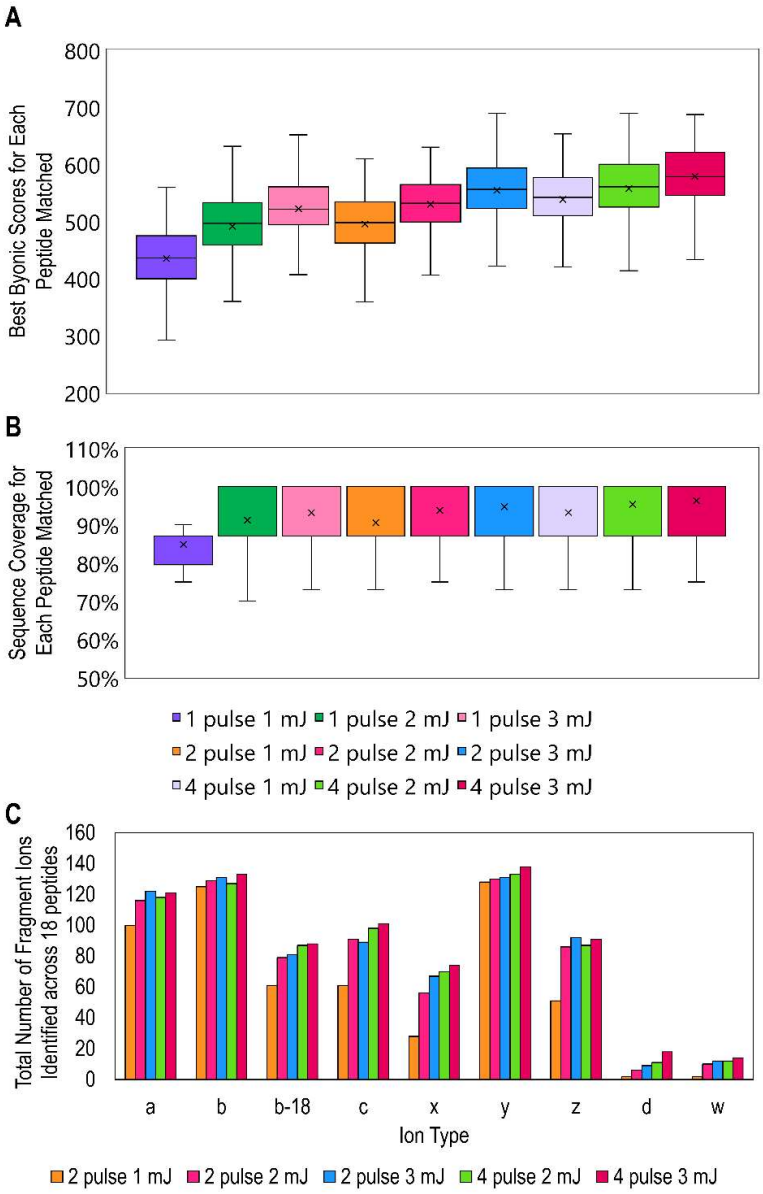


Figure 6.1: Optimization of laser energy and the number of pulses for 193 nm UVPD, including box plots of (A) the Byonic score and (B) the sequence coverage of the highest-scoring peptide-spectra match for each peptide in the 157-synthetic peptide mixture. Median lines do not appear in part B because medians were the same as either lower or upper quartiles. (C) Also considered in the optimization was the total number of each ion type observed. Ion types were summed for 18 of the peptides from the 157-peptide mixture.

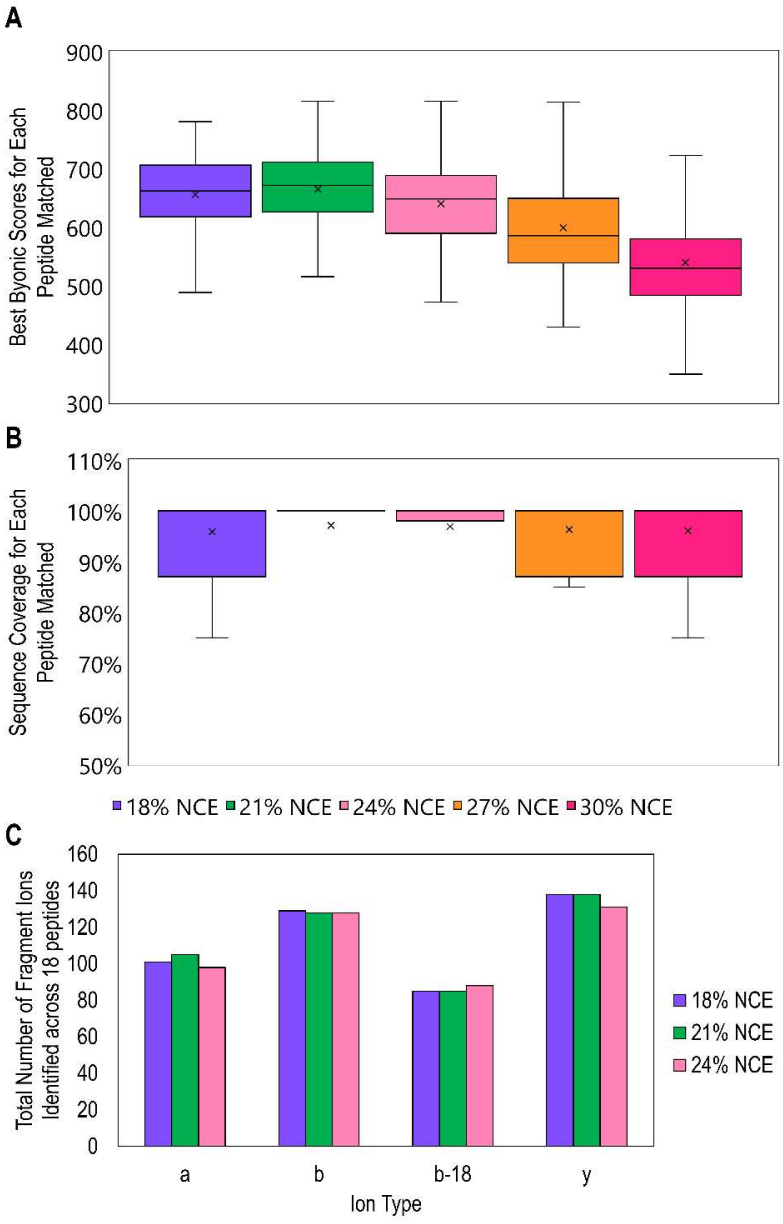


Figure 6.2: Optimization of normalized collisional energy (% NCE) for HCD, including box plots of (A) the Byonic score and (B) the sequence coverage of the highest-scoring peptide-spectra match for each peptide in the 157-synthetic peptide mixture. Median lines do not appear in part B because medians were the same as either lower or upper quartiles. (C) Also considered in the optimization was the total number of each ion type observed. Ion types were summed for 18 of the peptides from the 157-peptide mixture.

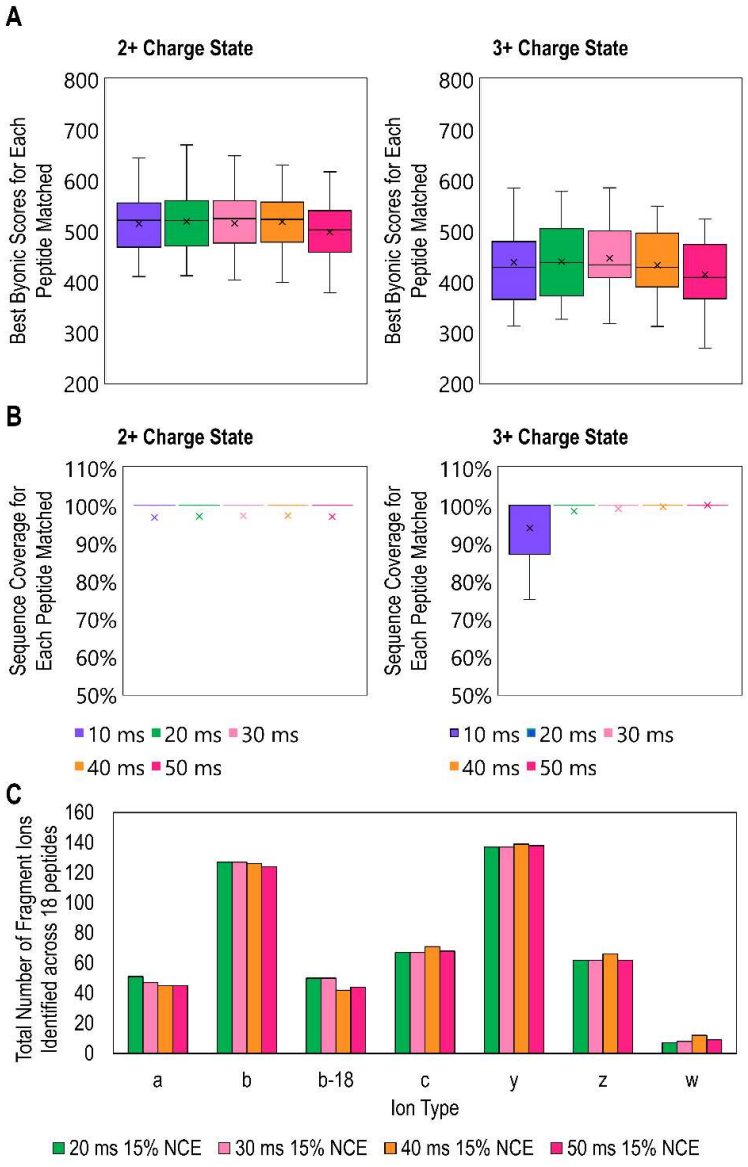


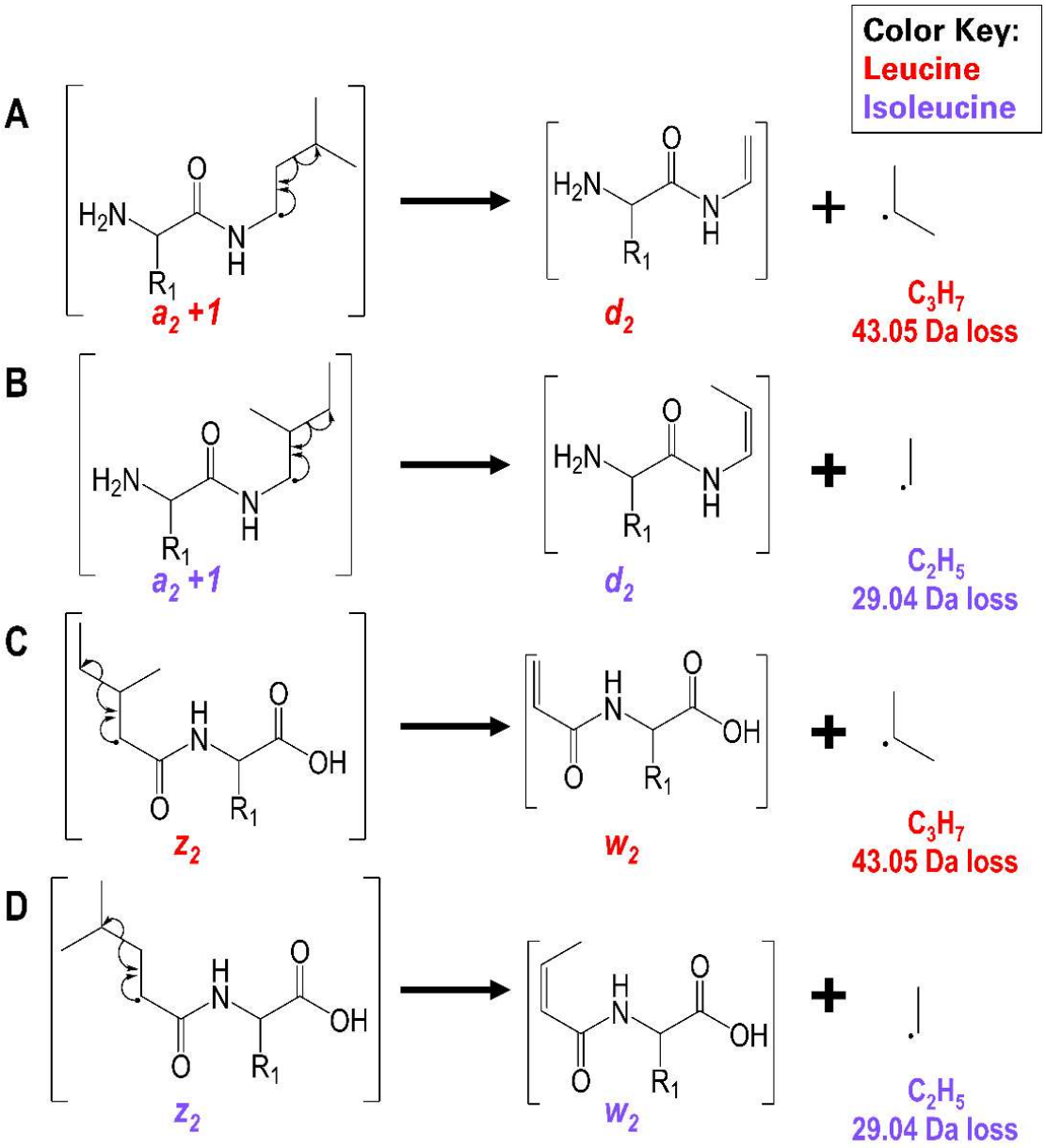
Figure 6.3: Optimization of activation time for EThcD, broken down by charge state, including box plots of (A) the Byonic score and (B) the sequence coverage of the highest-scoring peptide-spectra match for each peptide in the 157-synthetic peptide mixture. Supplemental activation of 15% NCE was applied with all activation times. Median lines do not appear in part B because medians were the same as either lower or upper quartiles. (C) Also considered in the optimization was the total number of each ion type observed. Ion types were summed for 18 of the peptides from the 157-peptide mixture.

6.4.2 Analysis of a Mixture Containing 157 Synthetic Peptides

Differentiation of leucine and isoleucine is an important facet for the characterization of immunopeptides, and thus the generation of diagnostic *d* and *w* side-chain loss ions was examined. **Scheme 6.1** illustrates the pathways for the formation of N-terminal *d* ions from radical *a+1* ions and C-terminal *w* ions from radical *z* ions.[54] While EThcD is capable of yielding *w* ions and has been reported to differentiate leucine and isoleucine in tryptic peptides,[26,54] 157 nm and 193 nm UVPD have been shown to generate both *d* and *w* ions for tryptic peptides.[34,46,55] 193 UVPD could offer an opportunity to improve immunopeptide characterization. Producing diagnostic ions containing either the N- or C-terminus upon UVPD is beneficial for analyzing immunopeptides which often lack a basic residue at the C-terminus (a feature of tryptic peptides that facilitates the formation of C-terminal *w* ions with EThcD).

Initial characterization of synthetic peptides by 193 nm UVPD was promising as several *d* and *w* ions were identified. Representative MS/MS spectra produced by 193 nm UVPD, EThcD, and HCD are displayed in **Figure 6.4** for YLFERIKEL, a peptide containing three I/L residues. As highlighted in **Figure 6.4A,** w_8 ions are identified in multiple charge states leading to the confirmation of L2, and both z_4 and d_6 ions are identified which confirm the identity of I6. The terminal L9 residue could not be identified owing to the low *m/z* of the potential z_1 fragment falling below the examined range (100-1250 *m/z*), and the *m/z* of the potential a_9 fragment appearing close to the precursor where it was obscured by other less informative neutral loss products. **Figure 6.4B** displays the EThcD spectrum for the 3+ charge state, and the same two I/L diagnostic fragment ions are identified as with 193 nm UVPD. The ratios of the related *z*/*w* ion pairs are comparable for both MS/MS methods. Owing to EThcD charge-reduction, the 3+ charge

state was selected for this MS/MS method. This need for a higher charge state poses a constraint on the analysis of other immunopeptides containing fewer basic residues. The severity of this constraint can be quantified based on how frequently each charge state is matched. In the mixture of synthetic peptides, out of the 154 peptides identified with HCD, only 11 were observed in the 3+ charge state, whereas 142 peptides were identified in the 2+ charges state, and nine peptides were identified in the 1+ charge state. In general, a greater array of diagnostic side-chain loss ions was generated by 193 nm UVPD than by EThcD, offering an advantage for analysis of complicated mixtures containing uncharacterized peptides where I/L identities are unknown. While *d* and *w* ions are not typically included in routine database searches owing to the inability of conventional MS/MS methods to generate these ions, the *d/w* ions are instrumental for confirmation of I/L assignments by 193 nm UVPD. The HCD spectrum is also included in **Figure 6.4C,** and while high sequence coverage was achieved, there was no generation of diagnostic *d* and *w* ions to differentiate leucine and isoleucine.



Scheme 6.1: Illustration of diagnostic neutral losses resulting in *d* and *w* ions that allow differentiation of leucine and isoleucine. The schema includes (A) a *d* ion resulting from the loss of C_3H_7 from a leucine-containing *a+1* ion, (B) a *d* ion resulting from the loss of C_2H_5 from an isoleucine-containing *a+1* ion, (C) a *w* ion resulting from the loss of C_3H_7 from a leucine-containing *z* ion, and (D) a *w* ion resulting from the loss of C_2H_5 from a leucine-containing *z* ion. The pathways are adapted from Xiao *et al*.[56]

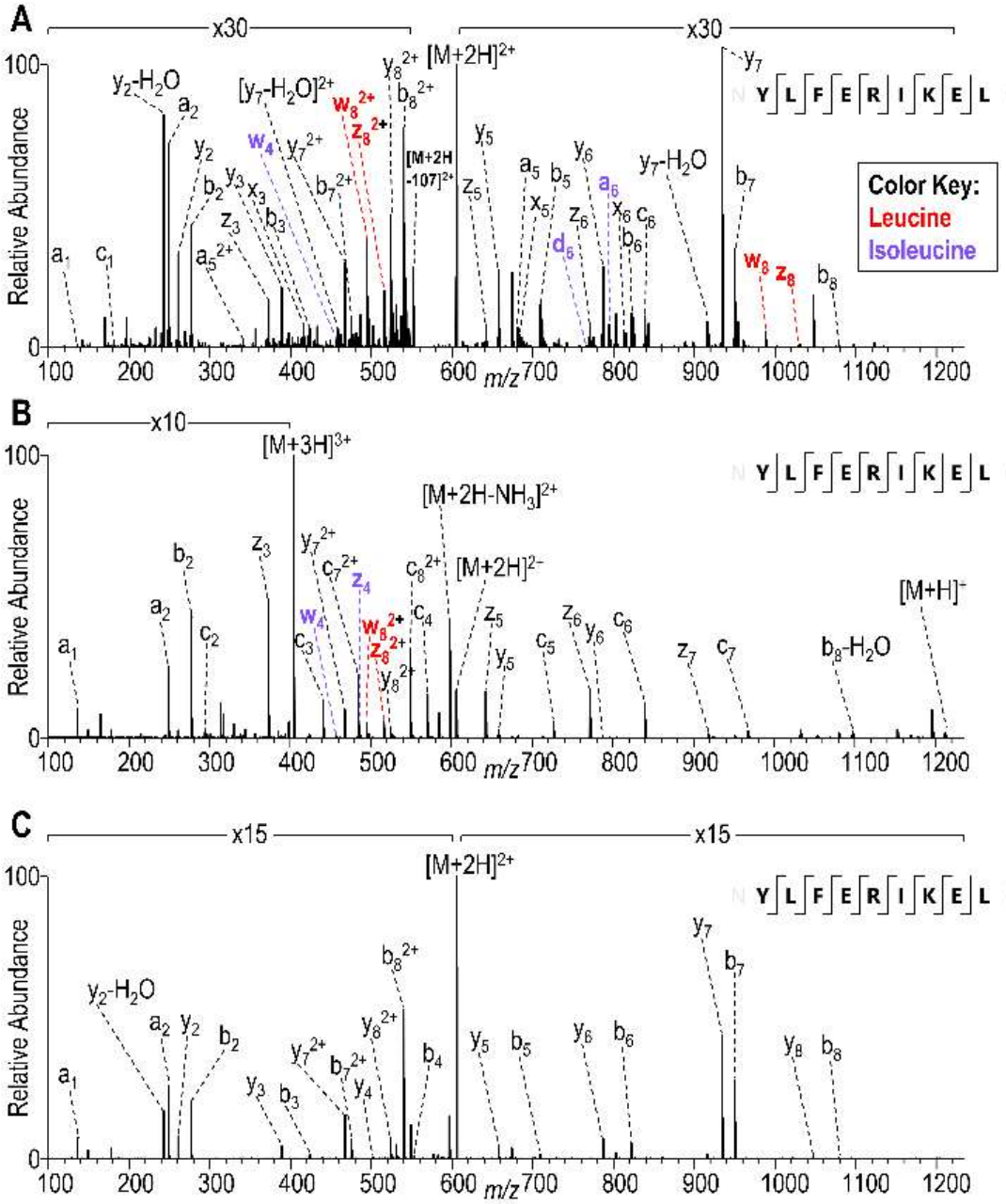


Figure 6.4: MS/MS spectra of the synthetic peptide, YLFERIKEL, subjected to (A) 193 nm UVPD (4 pulses, 3 mJ per pulse), (B) EThcD (40 ms and 15% NCE), or (C) HCD (21% NCE). The 2+ charge state was selected for 193 nm UVPD and HCD, and the 3+ charge state was selected for EThcD. Key fragments related to several diagnostic *d* and *w* ions are highlighted for 193 nm UVPD and EThcD.

Assessment of the mixture containing 157 synthetic peptides resulted in 153 peptides successfully identified by 193 nm UVPD, 154 peptides by HCD, and 150 peptides by EThcD. The set included 144 leucine-containing peptides and 74 isoleucine-containing peptides. Leucine and isoleucine identifications were examined systematically for the 157 synthetic peptides, as summarized in **Table 6.1**. Each leucine or isoleucine loss as annotated by Byonic was manually recorded for the peptides in the 157-peptide mixture, and only *d* and *w* ions that appeared in two of the three replicates were reported. Of the 242 leucine residues in the peptide mixture, 143 were confirmed by either a *d* or *w* ion, while 49 of the 92 isoleucine residues were confirmed with 193 nm UVPD. EThcD, limited to the generation of *w* ions, resulted in 75 leucine identifications and only ten isoleucine identifications. The significantly fewer identifications observed for EThcD can be explained by the tendency of EThcD to generate longer fragments, impeding the identification of sequences in which the leucine/isoleucine residues were located near the C-terminus of the peptides. This limitation is overcome by 193 nm UVPD, in which small fragment ions were effectively generated, and the ability to generate both *d* and *w* ions by 193 nm UVPD led to a greater chance of observing a productive ion. While not every leucine and isoleucine residue was confirmed with a *d* or *w* ion by 193 nm UVPD, the data displayed here is the result of consistent and reliable identifications, and it is expected that this strategy could be expanded to confirm the identity of individual peptides of interest via targeted analysis. The approach developed here represents a promising step towards improving immunopeptide identification and could be especially valuable if applied to spliced immunopeptide sequencing or de-novo sequencing, where the identifications of leucine and isoleucine residues are either ambiguous or rely on a post-analysis library search after the peptide sequences are characterized.[57–59]

		Number of isomeric amino acids	Number of isomeric amino acids confirmed	Percent confirmed	Number of *w* ions identified	Number of *d* ions identified
193 nm UVPD	Leucine	242	143	59%	111	90
	Isoleucine	92	49	53%	40	18
EThcD	Leucine	241	75	31%	75	--
	Isoleucine	89	10	11%	10	--

Table 6.1: Confirmation of I/L identities from a mixture of 157 synthetic peptides achieved using either 193 nm UVPD or EThcD. All *d* and *w* ions were identified by Byonic and were only included if they appeared in duplicate LC runs. The activation parameters used were: 193 nm UVPD (4 pulses 3 mJ) and EThcD (40 ms and 15% NCE).

Another prominent pathway observed upon 193 nm UVPD in **Figure 6.4A**, and highlighted in **Figure 6.5B**, is the loss of 107.05 Da from the $[M+2H]^{2+}$ ion peak, which corresponds to the loss of a tyrosine side-chain. Tyrosine and other aromatic neutral losses have previously been reported for 193 nm, 213 nm, 220 nm, and 266 nm UVPD.[41,60–62] The high abundance of this product makes it an attractive candidate to integrate into a search algorithm to further refine the accuracy of peptide matching. In addition to the highly abundant tyrosine neutral loss, we also observed similar neutral losses from tryptophan- and phenylalanine-containing peptides, as depicted in **Figure 6.5**. The loss of 91.05 Da, characteristic of phenylalanine, is shown in the 193 nm UVPD mass spectrum of AMSSKFFLV (2+) in **Figure 6.5C**, and the loss of 130.06 Da, associated with tryptophan, is shown in the 193 nm UVPD mass spectrum of QLIDKVWQL (2+) in **Figure 6.5D**. The prevalence of aromatic side-chain neutral losses observed across all peptides in the synthetic peptide mixture is compiled in **Figure 6.5A**. The presence of tyrosine afforded the neutral loss of highest abundance, but side-chain neutral losses were also observed for all phenylalanine- and tryptophan-containing peptides. Owing to the frequency and significant abundances of the aromatic neutral

losses, these fragment ions should be incorporated in future versions of UVPD search algorithms, allowing enhanced scoring and unambiguous identification of peptides containing aromatic residues.

A

	Number of peptides containing aromatic residues	Number of sidechain losses observed	Percent Observed
All aromatics	59	59	100%
Tyrosine (107.05 Da)	29	29	100%
Phenylalanine (91.05 Da)	25	25	100%
Tryptophan (130.06 Da)	8	8	100%

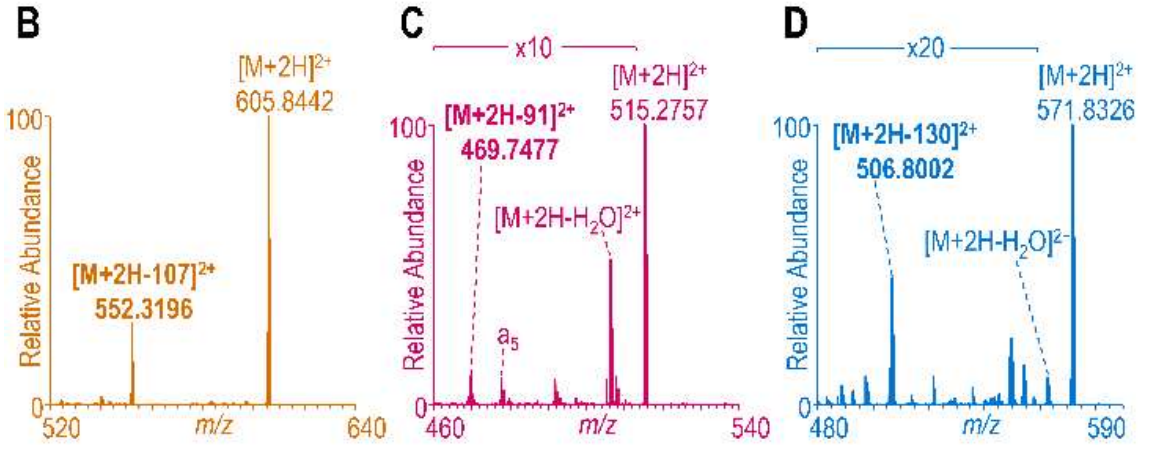


Figure 6.5: (A) Table displaying the frequency of side-chain losses from aromatic residues observed for the HLA-A*02:01 synthetic peptide mixture using 193 nm UVPD (4 pulses, 3 mJ per pulse). For each peptide containing aromatic amino acids, the spectrum was manually examined to determine if the neutral loss occurred (Tyr 107.05 Da, Phe 91.05 Da, Trp 130.06 Da). Expansions of selected regions of the UVPD mass spectra are also included to demonstrate each type of neutral loss. The representative peptides are (B) tyrosine-containing peptide YLFERIKEL (2+), (C) phenylalanine-containing peptide AMSSKFFLV (2+), and (D) tryptophan-containing peptide QLIDKVWQL (2+).

6.4.3 Immunoprecipitation samples

To expand the examination of 193 UVPD for characterization of immunopeptides, samples were generated by immunoprecipitation of peptides from BV173 cells. Immunoprecipitation of cell lysates was first performed with the BB7.2 antibody, which specifically isolates the peptides associated with the HLA-A*02:01 allele. The flow-through was precipitated with the W6/32 antibody to isolate the complete MHC class I repertoire (HLA -A, -B, and -C). While the focus of this study is the HLA-A*02:01 allele due to its relatively high prevalence[12,13] and the specific challenges involved in its characterization, the HLA-A, -B, and -C peptides were included to evaluate the applicability of UVPD to a broader range of motifs. 193 nm UVPD, EThcD, and HCD methods were utilized for benchmark comparisons.

Examples of the 193 nm UVPD, EThcD, and HCD mass spectra obtained for one exemplative peptide in the BB7.2 immunoprecipitated sample are displayed in **Figure 6.6**. This peptide (SLAQYLINV) was showcased because it was consistently identified in each run by every MS/MS method. As highlighted in **Figure 6.6A**, the identities of the isoleucine residue and one of the leucine residues in this peptide were confirmed by *w* ions in the 193 nm UVPD spectrum of the 2+ charge state. As a complement, the w_8 ion resulting in the confirmation of the remaining leucine residue was identified in the 193 nm UVPD spectrum of the 1+ charge state (**Figure 6.6B**) as well as the EThcD spectrum of the 2+ charge state (**Figure 6.6C**). While the abundances of the *w* ions in the UVPD spectra are low, these ions had high mass accuracies and could be differentiated from the other low abundance sequence and internal ions in the spectra, as highlighted in subsections of the spectra expanded in **Figures 6.7-6.9**. It was promising to see that in this case, complete confirmation of the leucine and isoleucine residues could be obtained

by combining information from the 2+ and 1+ charge states. This trend of the generation of longer *d* and *w* ions for the 1+ charge state was often observed with 193 nm UVPD, assisting in the confident identification of leucine and isoleucine residues near the termini. A similar pattern was observed for EThcD of 2+ charge state, whereas shorter fragment ions were more frequently generated upon EThcD of the 3+ charge state (as in **Figure 6.4B**). While ideally these patterns would result in complete sequence information from both UVPD and EThcD, most peptides only appeared in the 2+ charge state (see **Figure 6.10**). Therefore, it was reasoned that the greatest number of leucine and isoleucine confirmations could be achieved by combining the *d* and *w* ions identified from UVPD with the *w* ions identified from EThcD.

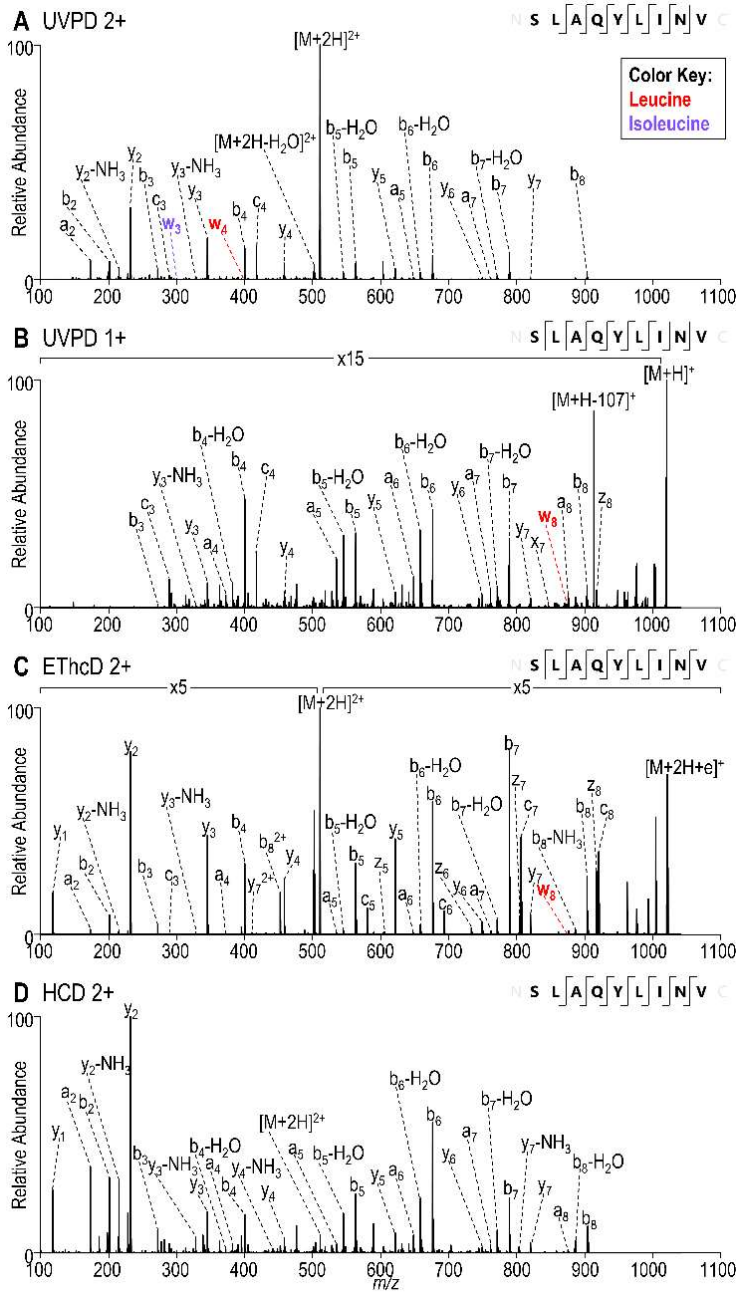
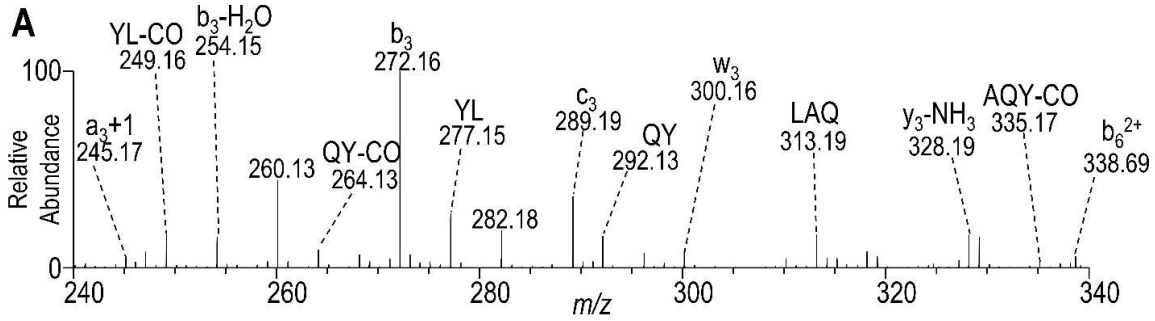


Figure 6.6: Mass spectra of the peptide identified as SLAQYLINV from BB7.2 immunoprecipitation obtained using 193 nm UVPD (4 pulses, 3 mJ) of the (A) 1+ and (B) 2+ charge states, (C) EThcD (2+ charge state, 40 ms/15% NCE) and (D) HCD (2+ charge state, 21% NCE). The *w* ions leading to the confirmation of leucine and isoleucine residues are highlighted in the UVPD and EThcD mass spectra.

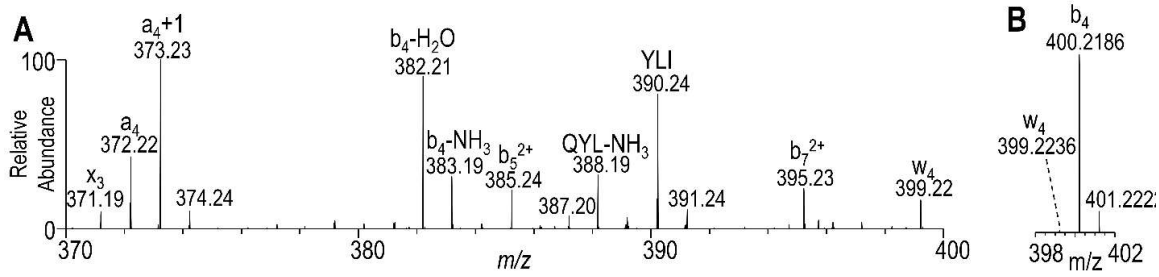
A



B

	Theoretical m/z	Observed m/z	Error (ppm)
a_3+1	245.1734	245.1732	0.82
YL-CO	249.1598	249.1594	1.61
b_3-H_2O	254.1499	254.1497	0.79
QY-CO	264.1343	264.1340	1.14
b_3	272.1605	272.1602	1.10
YL	277.1547	277.1544	1.08
c_3	289.1870	289.1867	1.04
QY	292.1292	292.1290	0.68
w_3	300.1554	300.1551	1.00
LAQ	313.1870	313.1867	0.96
y_3-NH_3	328.1867	328.1862	1.52
AQY-CO	335.1714	335.1709	1.49
b_6^{2+}	338.6872	338.6866	1.62

Figure 6.7: (A) Expansion of region around *w3* fragment ion in the UVPD (4 pulses, 3 mJ) mass spectrum acquired for the peptide SLAQYLINV (2+ charge state) from the BB7.2 immunoprecipitation. (B) Table listing the theoretical *m/z*, observed *m/z*, and ppm error for each identified ion in the spectrum. All theoretical *m/z* values were calculated using Protein Prospector. The mass of the *w3* ion that confirms the identification of isoleucine was calculated by subtracting 29.0391 Da from the mass of the *z3* ion, and the mass of the *a3+1* ion was calculated by adding 1.0078 Da to the mass of the *a3* ion.



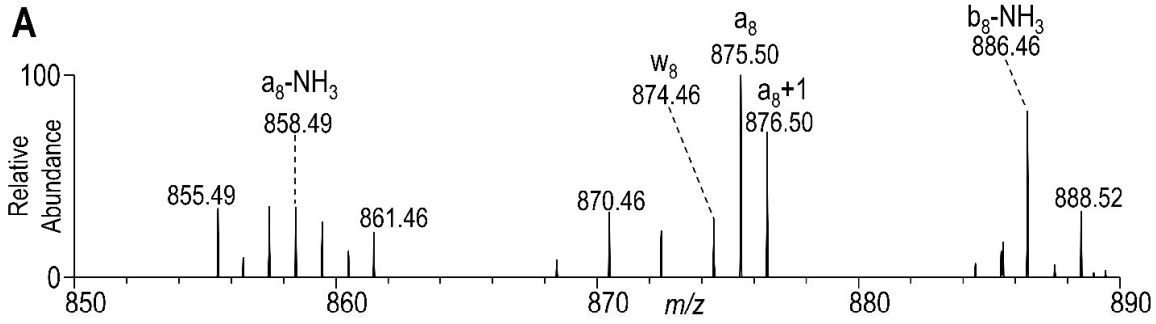
C

	Theoretical *m/z*	Observed *m/z*	Error (ppm)
AQY	363.1663	363.1659	1.10
x_3	371.1925	371.1924	0.27
a_4	372.2241	372.2237	1.07
a_4+1	373.2319	373.2316	0.87
b_4-H_2O	382.2085	382.2081	1.05
b_4-NH_3	383.1925	383.1921	1.04
b_5^{2+}	395.2292	395.2286	1.39
QYL-NH_3	388.1867	388.1863	1.03
YLI	390.2387	390.2382	1.28
b_7^{2+}	395.2292	395.2286	1.39
w_4	399.2238	399.2235	0.75

D

	Theoretical m/z	Observed m/z	Error (ppm)
w_4	399.2238	399.2235	0.75
b_4	400.2191	400.2188	0.75

Figure 6.8: Expansion of region around *w4* ion in the UVPD (4 pulses, 3 mJ) mass spectrum acquired for the peptide SLAQYLINV (2+ charge state) from the BB7.2 immunoprecipitation. Both (A) the lower *m/z* region as well as (B) a smaller excerpt around the *b4* ion are included because the high abundance of the *b4* ion makes it challenging to observe the other ions when presented using the same abundance scale. (C/D) Tables of the identified ions in A and B display the theoretical *m/z*, observed *m/z*, and ppm error. The mass of the *w4* ion that confirms the identification of leucine was calculated by subtracting 43.0548 Da from the mass of the *z4* ion.



B

	Theoretical m/z	Observed m/z	Error (ppm)
a_8-NH_3	858.4720	858.4686	3.96
w_8	874.4669	874.4644	2.86
a_8	875.4985	875.4964	2.40
a_8+1	876.5063	876.5034	3.31
b_8-NH_3	886.4669	886.4641	3.16

Figure 6.9: (A) Expansion of region around *w8* ion in the UVPD (4 pulses, 3 mJ) mass spectrum acquired for the peptide SLAQYLINV (1+ charge state) from the BB7.2 immunoprecipitation. (B) Table of the identified ions displays the theoretical *m/z*, observed *m/z*, and ppm error for each identified ion in the spectrum. All theoretical *m/z* values were calculated using Protein Prospector. The mass of the *w8* ion that confirms the identification of isoleucine was calculated by subtracting 43.0548 Da from the mass of the z_8 ion, and the mass of the a_8+1 ion was calculated by adding 1.0078 Da to the mass of a_3 ion.

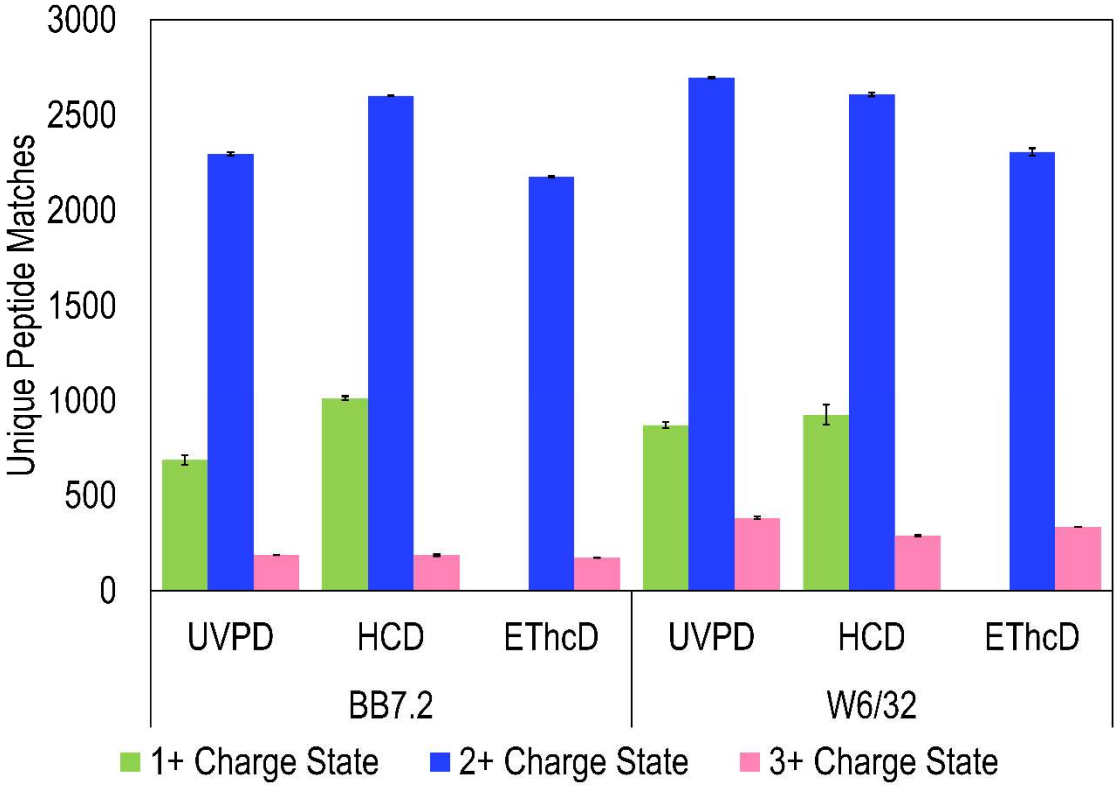


Figure 6.10: The number of peptide-spectrum matches (PSMs) for the BB7.2 and W6/32 immunoprecipitated samples categorized by charge state (1+, 2+, 3+) for UVPD, EThcD, and HCD mass spectra from Byonic output.

After manually examining a few individual peptides for production of *d* and *w* ions, a larger set of peptides identified with both 193 nm UVPD and EThcD from the BB7.2 immunoprecipitation was systematically examined. The results for I/L identifications are displayed in **Table 6.2**. Given the preliminary status of this methodology, I/L identifications were only included in the table if they confirmed the identity of the residue from the UniProt database and if the same neutral loss fragment ion appeared in two of the three replicates. The percent of leucine identifications with UVPD alone dropped from 59% in the synthetic mixture to 34% for the BB7.2 immunoprecipitated sample and from 53% to 28% for isoleucine. The number of identifications dropped comparably for EThcD, from 31% to 14% for leucine and from 11% to 7% for isoleucine. These decreases are explained by the lower peptide abundances in the immunoprecipitation sample compared to the synthetic peptide mixture, as well as the greater complexity which led to fewer MS/MS repeats for each peptide. Due to complementary information from 193 nm UVPD and EThcD, a handful of additional leucine and isoleucine identifications was achieved by combining the diagnostic ions from the two methods. By combining UVPD and EThcD results, the percent of leucine and isoleucine identifications increased to 40% and 30%, respectively. Depending on the sample availability, this slight increase may not justify including EThcD runs in the analysis. Notably, the increase was more significant for leucine than for isoleucine identifications. This outcome is not a result of a bias for EThcD to generate more *w* ions near leucine, but rather because more leucine residues were present near the C-terminus, allowing the confirmation of leucine residues through the generation of long *w* ions. As observed in the motif included in **Figure 6.11**, a large proportion of leucine residues was present in position two near the C-terminus, while a similar bias towards

positions two, four, and 9 was observed for isoleucine. The less significant bias towards leucine identifications observed with 193 nm UVPD than EThcD is a result of 193 UVPD achieving confirmation of leucine and isoleucine residues at a larger variety of amino acid positions.

		Number of isomeric amino acids	Number of isomeric amino acids confirmed	Percent confirmed	Number of *w* ions identified	Number of *d* ions identified
193 nm UVPD	leucine	453	156	34%	102	82
	Isoleucine	184	51	28%	45	10
EThcD	leucine	453	65	14%	65	-
	Isoleucine	184	12	7%	12	-
Combined	leucine	453	180	40%	126	82
	Isoleucine	184	56	30%	50	10

Table 6.2: Confirmation of I/L identities from BB7.2 immunoprecipitation samples achieved using 193 nm UVPD and or EThcD. All *d* and *w* ions were identified by Byonic and were only included if they appeared in at least twice in triplicate LC runs. Out of the peptides that were identified with both 193 nm UVPD and EThcD, only the 250 peptides with the highest average Byonic UVPD scores are enumerated in this table.

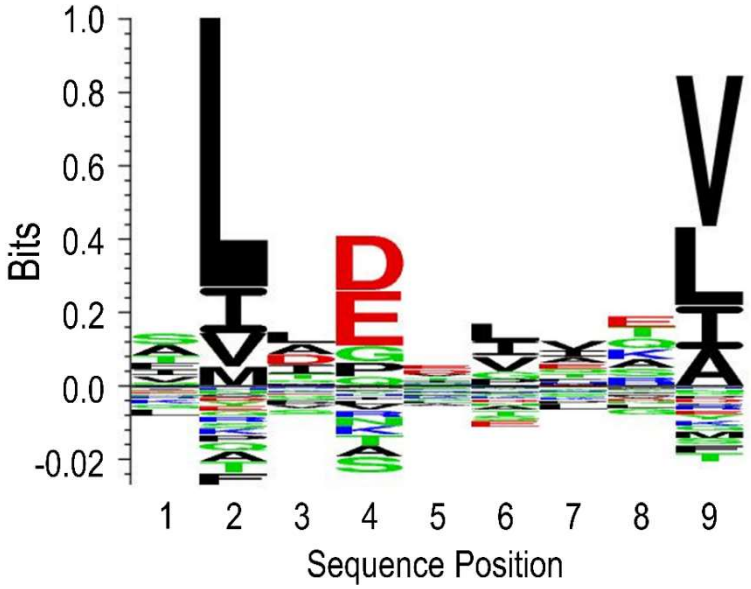


Figure 6.11: Sequence logo generated with Gibbs Cluster 2.0 from the 250 peptides used to produce Table 6.2.

While less consistent identification of leucine and isoleucine residues was achieved for the complex immunoprecipitation samples compared to the mixture of synthetic peptides, the total number of identifications was significant. The ability to differentiate leucine and isoleucine residues for some of the HLA-A*02:01 immunopeptides represents a promising step for the MS/MS workflow. Additional identifications are expected if manual analysis is employed, allowing averaging multiple spectra, as was achieved in **Figure 6.6**.

The results for both BB7.2 and W6/32 immunoprecipitation samples in terms of the Byonic scores and the sequence coverages are highlighted in **Figure 6.12**. While the average sequence coverage was high for all three MS/MS methods, the Byonic scores were all around 100 points lower than what was observed under the same optimized conditions with the synthetic peptide mixtures. This difference is explained by the higher

complexity and lower abundances of peptides of the immunoprecipitation sample compared to the synthetic mixture and highlights the importance of only comparing the Byonic scores within a given sample and fragmentation method. While the Byonic scores are significantly higher for HCD than the other MS/MS methods, the sequence coverages were consistent across all three methods, indicating a comparable identification quality. The number of unique peptides matched for each charge state and MS/MS method are also highlighted in **Figure 6.10**. Most peptides for both immunoprecipitation samples appeared in the 2+ charge state with small numbers of identifications of peptides in the 1+ and 3+ charge states. The limited prevalence of peptides in higher charge states for these two samples indicates that HCD and UVPD are preferable choices for characterization due to their lower charge state dependence compared to EThcD.

In order to further validate the findings in this study, **Figure 6.13** displays the plots of the hydrophobicity index of each peptide, versus its retention time.[51] SSRCalc has previously been employed to help assess false discovery rates of immunopeptide identifications.[63,64] In the current study, a tight correlation between hydrophobicity and retention time was observed. The linearity provides confidence that the peptides identified in each LC-MS/MS run have reasonable assignments that align with expected physio-chemical properties.

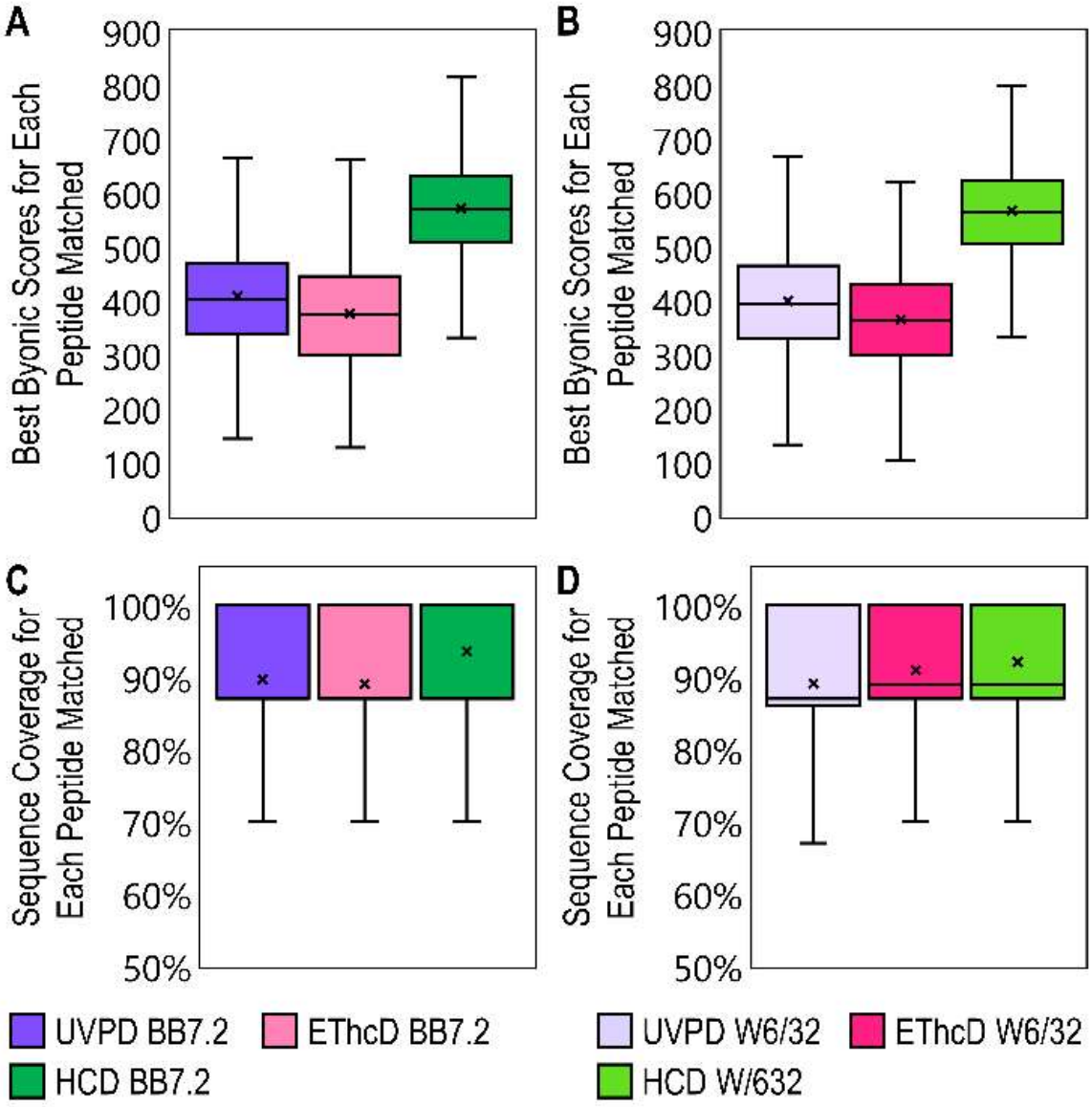


Figure 6.12: Summary of metrics for the immunoprecipitation samples. Box and whisker plots are included showing the Byonic scores for all peptides matched for both (A) BB7.2 and (B) W6/32 immunoprecipitations, as well as the sequence coverages for both (C) BB7.2 and (D) W6/32 immunoprecipitations for all three MS/MS methods. The activation parameters used were: 193 nm UVPD (4 pulses 3 mJ), EThcD (40 ms and 15% NCE), and HCD (21% NCE). Median lines do not appear in part C because medians were the same as either lower or upper quartiles.

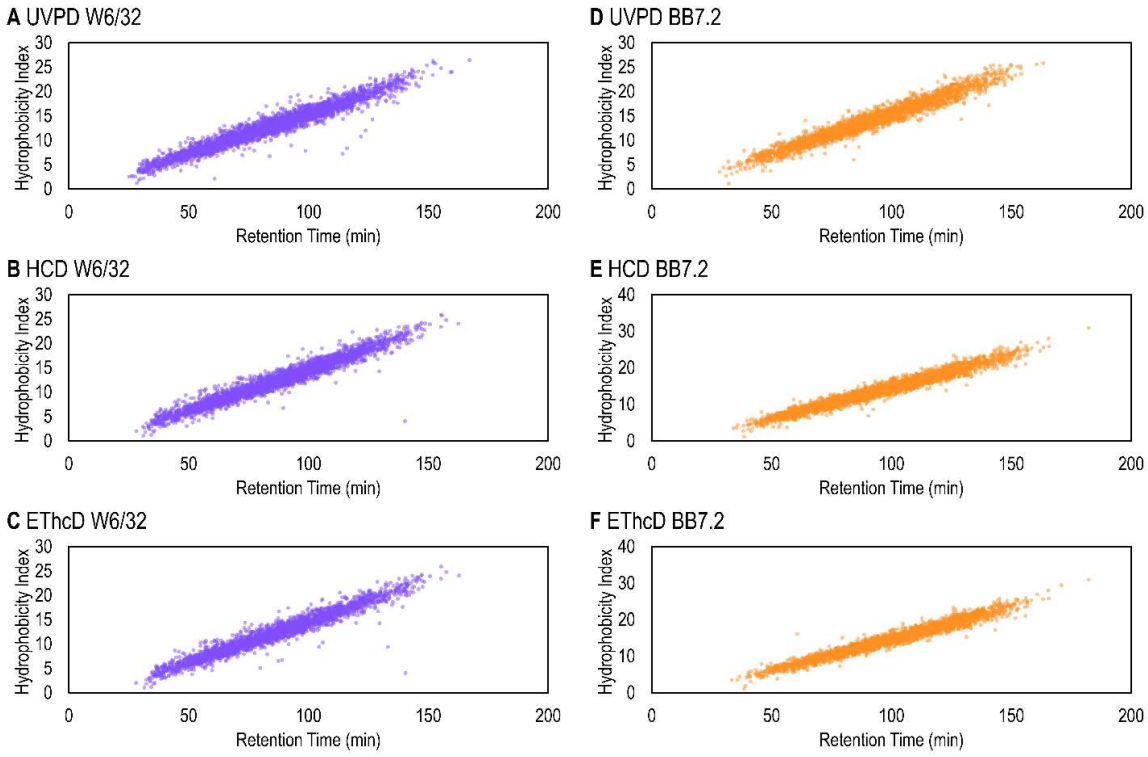


Figure 6.13: Plots of the retention time vs. the hydrophobicity index for the peptides in the W6/32 Immunoprecipitation, calculated using SSR calc for (A) UVPD, (B) HCD, and (C) EThcD spectra and in the BB7.2 immunoprecipitation for (D) UVPD, (E) HCD, and (F) EThcD spectra.

To reduce the likelihood of false positive matches, only those peptides identified in triplicate runs by each MS/MS method were retained. Based on this criterion, peptides identified by each of the MS/MS methods are summarized in the Venn diagrams in **Figure 6.14** for the two sets of immunopeptide enrichments (BB7.2 and W6/32). While there is considerable overlap between all three fragmentation methods, EThcD consistently matched fewer peptides than the other two methods, an outcome rationalized based on its charge state dependence. For the BB7.2 sample, HCD matched the most peptides, while UVPD matched the most peptides in the W6/32 sample. Most of the peptides identified in the W6/32 sample belonged to either the motif HLA-B*18:01,

which tends to contain a negatively charged glutamic acid residue, or HLA-B*15:10, which tends to contain a positively charged histidine residue. The motifs are a good complement to HLA-A*02:01 peptides from BB7.2, which tend towards neutral hydrophobic peptides. The peptides identified for each fragmentation technique were separated by motif for the W6/32 samples with Gibbs cluster 2.0.[65] Venn Diagrams summarizing the number of peptides identified for the distinctive motifs based on MS/MS method are included in **Figure 6.15**. UVPD resulted in a handful more peptide identification than HCD for both motifs, with a slightly larger difference in the histidine-containing HLA-B*15:10 motif. The differences between the subsets in **Figure 6.15** are too small to make conclusions about how the charge of the peptide motifs affects fragmentation by UVPD, HCD, and EThcD, respectively. However, it is clear that UVPD provides enhanced complementary identifications when paired with HCD rather than EThcD for the characterization of immunopeptides.

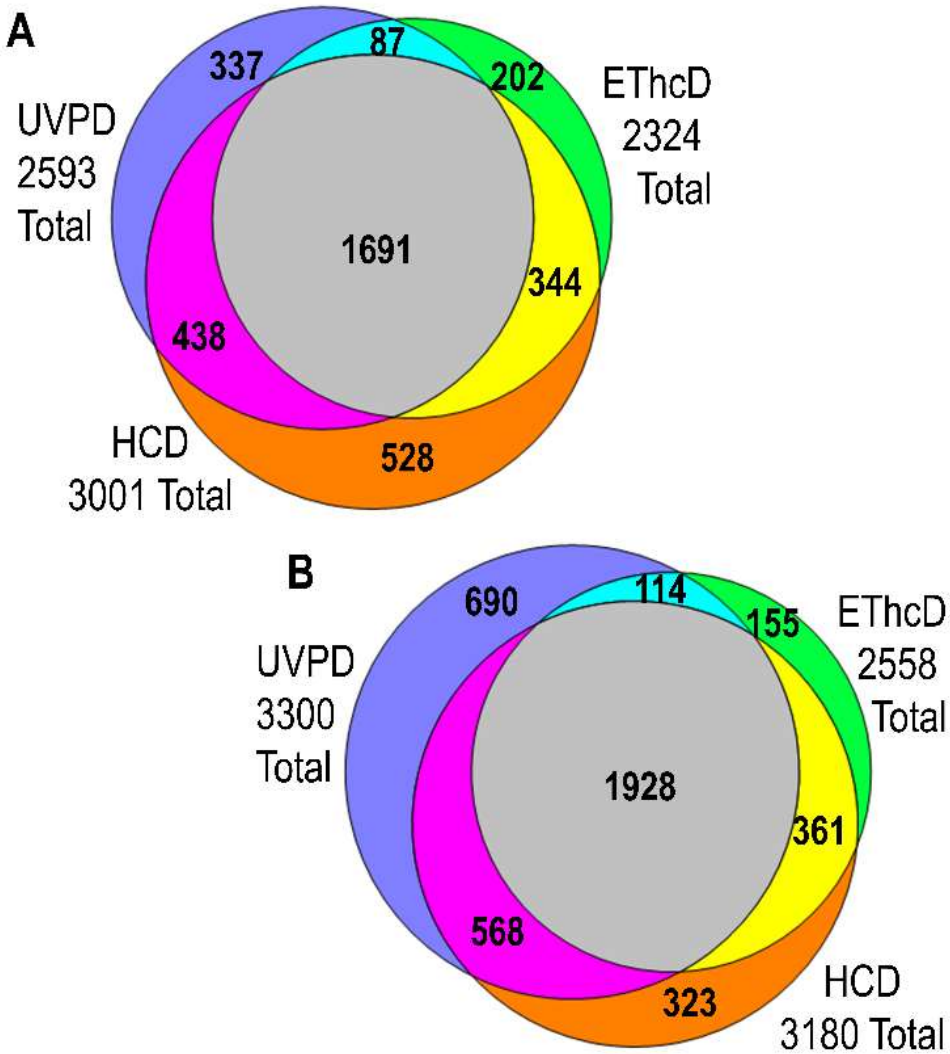


Figure 6.14: Venn diagrams displaying the total number of peptides identified utilizing UVPD, EThcD, and HCD for (A) BB7.2 and (B) W6/32 immunoprecipitated samples. The activation parameters used were: 193 nm UVPD (4 pulses 3 mJ), HCD (21% NCE), and EThcD (40 ms and 15% NCE).

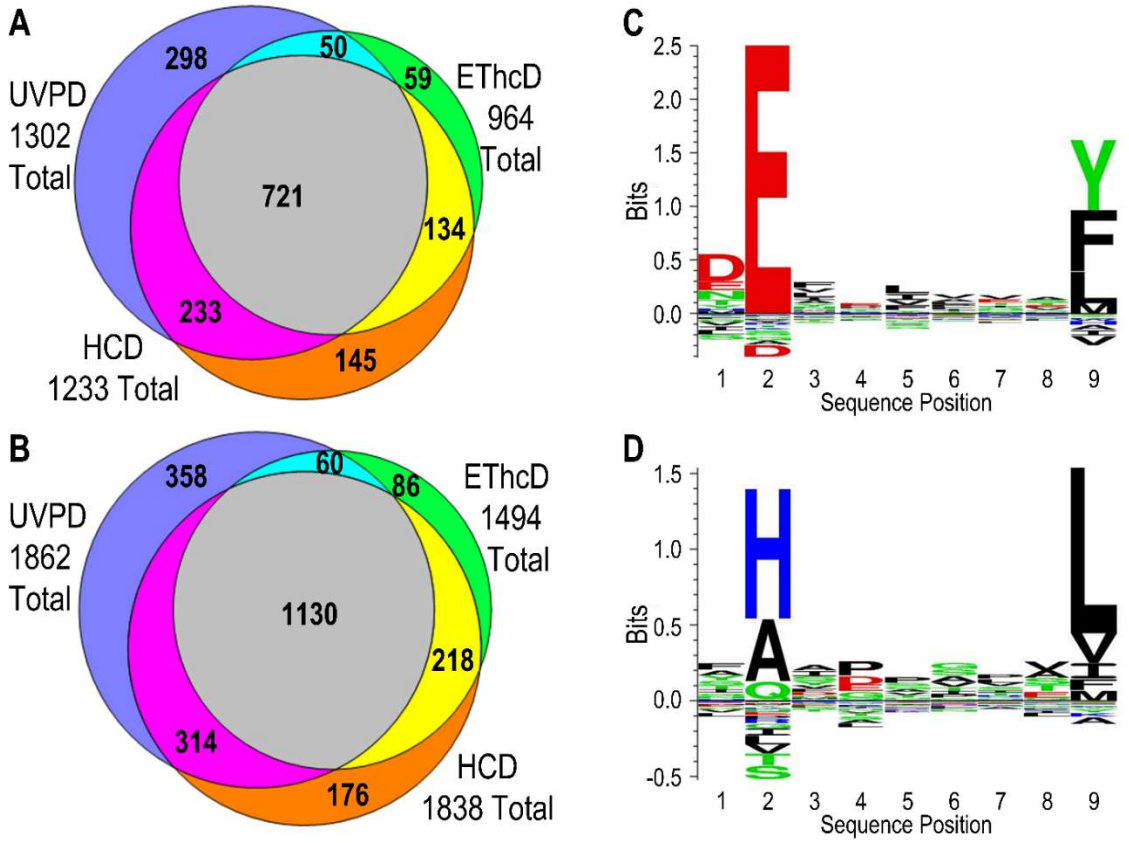


Figure 6.15: Venn diagrams displaying the total number of peptides identified utilizing UVPD, EThcD, and HCD for W6/32 immunoprecipitated sample. The peptides have been separated into either (A) HLA-B*15:10 or (B) HLA*18:01 based on cluster alignment in Gibbs Cluster 2.0. The sequence logos generated with Gibbs Cluster 2.0 are also included for both motifs (C) HLA-B*15:10 and (D) HLA*18:01 based on the combination of unique peptide matches across UVPD, EThcD, and HCD.

6.5 Conclusion

The capabilities of 193 nm UVPD for characterization of immunopeptides were compared to two other popular MS/MS methods, EThcD and HCD, based on evaluation of a mixture of 157 synthetic peptides as well as two different immunoprecipitation samples. Byonic allowed semi-automated confirmation of isomeric leucine and isoleucine

residues from EThcD and UVPD data. Neutral losses associated with aromatic amino acids, particularly for tyrosine, consistently emerged in the UVPD mass spectra, offering a promising new diagnostic marker to further enhance peptide identifications. For the BB7.2 immunoprecipitated sample, 3001, 2593, and 2324 total unique peptides were identified with HCD, 193 nm UVPD, and EThcD, respectively, and for the W6/32 immunoprecipitated sample, 3180, 3300, and 2558 peptides were identified using HCD, 193 nm UVPD, and EThcD. The successful performance of 193 nm UVPD is attributed in part to its lack of significant dependence on peptide charge state, a factor that likely hampered the performance of EThcD for HLA-A*02:01 and HLA-B*15:10 immunopeptides that lack basic sites and preferentially produce low charge states.

6.6 REFERENCES

(1) Creech, A. L.; Ting, Y. S.; Goulding, S. P.; Sauld, J. F. K.; Barthelme, D.; Rooney, M. S.; Addona, T. A.; Abelin, J. G. The Role of Mass Spectrometry and Proteogenomics in the Advancement of HLA Epitope Prediction. *Proteomics* **2018**, *18*, 1700259.

(2) Schumacher, F.-R.; Delamarre, L.; Jhunjhunwala, S.; Modrusan, Z.; Phung, Q. T.; Elias, J. E.; Lill, J. R. Building Proteomic Tool Boxes to Monitor MHC Class I and Class II Peptides. *Proteomics* **2017**, *17*, 1600061.

(3) Kronenberg-Versteeg, D.; Eichmann, M.; Russell, M. A.; Ru, A. de; Hehn, B.; Yusuf, N.; Veelen, P. A. van; Richardson, S. J.; Morgan, N. G.; Lemberg, M. K.; Peakman, M. Molecular Pathways for Immune Recognition of Preproinsulin Signal Peptide in Type 1 Diabetes. *Diabetes* **2018**, db170021.

(4) Blees, A.; Januliene, D.; Hofmann, T.; Koller, N.; Schmidt, C.; Trowitzsch, S.; Moeller, A.; Tampé, R. Structure of the Human MHC-I Peptide-Loading Complex. *Nature* **2017**, *551*, 525–528.

(5) Lemmel, C.; Weik, S.; Eberle, U.; Dengjel, J.; Kratt, T.; Becker, H.-D.; Rammensee, H.-G.; Stevanović, S. Differential Quantitative Analysis of MHC Ligands by Mass Spectrometry Using Stable Isotope Labeling. *Nat. Biotechnol* **2004**, *22*, 450–454.

(6) Yadav, M.; Jhunjhunwala, S.; Phung, Q. T.; Lupardus, P.; Tanguay, J.; Bumbaca, S.; Franci, C.; Cheung, T. K.; Fritsche, J.; Weinschenk, T.; Modrusan, Z.;

Mellman, I.; Lill, J. R.; Delamarre, L. Predicting Immunogenic Tumour Mutations by Combining Mass Spectrometry and Exome Sequencing. *Nature* **2014**, *515*, 572–576.

(7) Carreno, B. M.; Magrini, V.; Becker-Hapak, M.; Kaabinejadian, S.; Hundal, J.; Petti, A. A.; Ly, A.; Lie, W.-R.; Hildebrand, W. H.; Mardis, E. R.; Linette, G. P. A Dendritic Cell Vaccine Increases the Breadth and Diversity of Melanoma Neoantigen-Specific T Cells. *Science* **2015**, *348*, 803–808.

(8) Ahmed, S. F.; Quadeer, A. A.; McKay, M. R. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* **2020**, *12*, 254.

(9) Tran, N. H.; Qiao, R.; Xin, L.; Chen, X.; Liu, C.; Zhang, X.; Shan, B.; Ghodsi, A.; Li, M. Deep Learning Enables de Novo Peptide Sequencing from Data-Independent-Acquisition Mass Spectrometry. *Nat. Methods* **2019**, *16*, 63–66.

(10) York, I. A.; Chang, S.-C.; Saric, T.; Keys, J. A.; Favreau, J. M.; Goldberg, A. L.; Rock, K. L. The ER Aminopeptidase ERAP1 Enhances or Limits Antigen Presentation by Trimming Epitopes to 8–9 Residues. *Nat. Immunol* **2002**, *3*, 1177–1184.

(11) Abelin, J. G.; Keskin, D. B.; Sarkizova, S.; Hartigan, C. R.; Zhang, W.; Sidney, J.; Stevens, J.; Lane, W.; Zhang, G. L.; Eisenhaure, T. M.; Clauser, K. R.; Hacohen, N.; Rooney, M. S.; Carr, S. A.; Wu, C. J. Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-Allelic Cells Enables More Accurate Epitope Prediction. *Immunity* **2017**, *46*, 315–326.

(12) Cano, P.; Klitz, W.; Mack, S. J.; Maiers, M.; Marsh, S. G. E.; Noreen, H.; Reed, E. F.; Senitzer, D.; Setterholm, M.; Smith, A.; Fernández-Viña, M. Common and Well-Documented HLA Alleles: Report of the Ad-Hoc Committee of the American Society for Histocompatiblity and Immunogenetics. *Hum. Immunol* **2007**, *68*, 392–417.

(13) Solberg, O. D.; Mack, S. J.; Lancaster, A. K.; Single, R. M.; Tsai, Y.; Sanchez-Mazas, A.; Thomson, G. Balancing Selection and Heterogeneity across the Classical Human Leukocyte Antigen Loci: A Meta-Analytic Review of 497 Population *Studies. Hum. Immunol* **2008**, *69*, 443–464.

(14) Barouch, D.; Friede, T.; Stevanovic, S.; Tussey, L.; Smith, K.; Rowland-Jones, S.; Braud, V.; McMichael, A.; Rammensee, H.-G. HLA-A2 Subtypes Are Functionally Distinct in Peptide Binding and Presentation. *J. Exp. Med.* **1995**, *182*, 1847–1856.

(15) Hunt, D. F.; Henderson, R. A.; Shabanowitz, J.; Sakaguchi, K.; Michel, H.; Sevilir, N.; Cox, A. L.; Appella, E.; Engelhard, V. H. Characterization of Peptides

Bound to the Class I MHC Molecule HLA-A2.1 by Mass Spectrometry. *Science* **1992**, *255*, 1261–1263.

(16) Blatnik, R.; Mohan, N.; Bonsack, M.; Falkenby, L. G.; Hoppe, S.; Josef, K.; Steinbach, A.; Becker, S.; Nadler, W. M.; Rucevic, M.; Larsen, M. R.; Salek, M.; Riemer, A. B. A Targeted LC-MS Strategy for Low-Abundant HLA Class-I-Presented Peptide Detection Identifies Novel Human Papillomavirus T-Cell Epitopes. *Proteomics* **2018**, *18*, 1700390.

(17) Ternette, N.; Nordkamp, M. J. M. O.; Müller, J.; Anderson, A. P.; Nicastri, A.; Hill, A. V. S.; Kessler, B. M.; Li, D. Immunopeptidomic Profiling of HLA-A2-Positive Triple Negative Breast Cancer Identifies Potential Immunotherapy Target Antigens. *Proteomics* **2018**, *18*, 1700465.

(18) Caron, E.; Kowalewski, D. J.; Koh, C. C.; Sturm, T.; Schuster, H.; Aebersold, R. Analysis of Major Histocompatibility Complex (MHC) Immunopeptidomes Using Mass Spectrometry. *Mol. Cell. Proteom.* **2015**, *14*, 3105–3117.

(19) Bassani-Sternberg, M.; Pletscher-Frankild, S.; Jensen, L. J.; Mann, M. Mass Spectrometry of Human Leukocyte Antigen Class I Peptidomes Reveals Strong Effects of Protein Abundance and Turnover on Antigen Presentation. *Mol. Cell. Proteom.* **2015**, *14*, 658–673.

(20) Jensen, S. M.; Potts, G. K.; Ready, D. B.; Patterson, M. J. Specific MHC-I Peptides Are Induced Using PROTACs. *Front. Immunol.* **2018**, *9*.

(21) Mommen, G. P. M.; Frese, C. K.; Meiring, H. D.; van Gaans-van den Brink, J.; de Jong, A. P. J. M.; van Els, C. A. C. M.; Heck, A. J. R. Expanding the Detectable HLA Peptide Repertoire Using Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). *Proc. Natl. Acad. Sci. U.S.A* **2014**, *111*, 4507–4512.

(22) Mommen, G. P. M.; Marino, F.; Meiring, H. D.; Poelen, M. C. M.; Brink, J. A. M. van G. den; Mohammed, S.; Heck, A. J. R.; Els, C. A. C. M. van. Sampling From the Proteome to the Human Leukocyte Antigen-DR (HLA-DR) Ligandome Proceeds Via High Specificity. *Mol. Cell. Proteom.* **2016**, *15*, 1412–1423.

(23) Cobbold, M.; Peña, H. D. L.; Norris, A.; Polefrone, J. M.; Qian, J.; English, A. M.; Cummings, K. L.; Penny, S.; Turner, J. E.; Cottine, J.; Abelin, J. G.; Malaker, S. A.; Zarling, A. L.; Huang, H.-W.; Goodyear, O.; Freeman, S. D.; Shabanowitz, J.; Pratt, G.; Craddock, C.; Williams, M. E.; Hunt, D. F.; Engelhard, V. H. MHC Class I–Associated Phosphopeptides Are the Targets of Memory-like Immunity in Leukemia. *Sci. Transl. Med.* **2013**, *5*, 203ra125-203ra125.

(24) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal. Chem.* **2018**, *90*, 40–64.

(25) Abelin, J. G.; Trantham, P. D.; Penny, S. A.; Patterson, A. M.; Ward, S. T.; Hildebrand, W. H.; Cobbold, M.; Bai, D. L.; Shabanowitz, J.; Hunt, D. F.

Complementary IMAC Enrichment Methods for HLA-Associated Phosphopeptide Identification by Mass Spectrometry. *Nat. Protoc.* **2015**, *10*, 1308–1318.

(26) Zhokhov, S. S.; Kovalyov, S. V.; Samgina, T. Yu.; Lebedev, A. T. An EThcD-Based Method for Discrimination of Leucine and Isoleucine Residues in Tryptic Peptides. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1600–1611.

(27) Kovalyov, S. V.; Zhokhov, S. S.; Onoprienko, L. V.; Vaskovsky, B. V.; Lebedev, A. T. Exploration of Doubtful Cases of Leucine and Isoleucine Discrimination in Mass Spectrometric Peptide Sequencing by Electron-Transfer and Higher-Energy Collision Dissociation-Based Method. *Eur. J. Mass Spectrom.* **2017**, *23*, 376–384.

(28) Armirotti, A.; Millo, E.; Damonte, G. How to Discriminate between Leucine and Isoleucine by Low Energy ESI-TRAP MSn. *J Am Soc Mass Spectrom* **2007**, *18*, 57–63.

(29) Lebedev, A. T.; Damoc, E.; Makarov, A. A.; Samgina, T. Yu. Discrimination of Leucine and Isoleucine in Peptides Sequencing with Orbitrap Fusion Mass Spectrometer. *Anal. Chem.* **2014**, *86*, 7017–7022.

(30) Nakada, T.; Sugihara, K.; Jikoh, T.; Abe, Y.; Agatsuma, T. The Latest Research and Development into the Antibody–Drug Conjugate, [Fam-] Trastuzumab Deruxtecan (DS-8201a), for HER2 Cancer Therapy. *Chem. Pharm. Bull.* **2019**, *67*, 173–185.

(31) Maibom-Thomsen, S.; Heissel, S.; Mørtz, E.; Højrup, P.; Bunkenborg, J. Discrimination of Isoleucine and Leucine by Dimethylation-Assisted MS3. *Anal. Chem.* **2018**, *90*, 9055–9059.

(32) Johnson, R. S.; Martin, S. A.; Biemann, Klaus.; Stults, J. T.; Watson, J. Throck. Novel Fragmentation Process of Peptides by Collision-Induced Decomposition in a Tandem Mass Spectrometer: Differentiation of Leucine and Isoleucine. *Anal. Chem.* **1987**, *59*, 2621–2625.

(33) 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. High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *J. Proteome Res.* **2017**, *16*, 2072–2079.

(34) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 Nm and Its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, *9*, 4205–4214.

(35) Greer, S. M.; Bern, M.; Becker, C.; Brodbelt, J. S. Extending Proteome Coverage by Combining MS/MS Methods and a Modified Bioinformatics Platform Adapted for Database Searching of Positive and Negative Polarity 193 Nm Ultraviolet Photodissociation Mass Spectra. *J. Proteome Res.* **2018**, *17*, 1340–1347.

(36) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. *Anal. Chem.* **2016**, *88*, 2303–2310.

(37) Reilly, J. P. Ultraviolet Photofragmentation of Biomolecular Ions. *Nucleic Acids Res* **2009**, *28*, 425–447.

(38) Moon, J. H.; Yoon, S. H.; Kim, M. S. Photodissociation of Singly Protonated Peptides at 193 Nm Investigated with Tandem Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3248–3252.

(39) Julian, R. The Mechanism behind Top-Down UVPD Experiments: Making Sense of Apparent Contradictions. *J Am Soc Mass Spectrom* **2017**, *28*, 1823–1826.

(40) Fornelli, L.; Srzentić, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **2018**, *90*, 8421–8429.

(41) Robinson, M. R.; Taliaferro, J. M.; Dalby, K. N.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation Mass Spectrometry for Phosphopeptide Characterization in the Positive and Negative Ion Modes. *J. Proteome Res.* **2016**, *15*, 2739–2748.

(42) Burkholder, N. T.; Sipe, S. N.; Escobar, E. E.; Venkatramani, M.; Irani, S.; Yang, W.; Wu, H.; Matthews, W. M.; Brodbelt, J. S.; Zhang, Y. Mapping RNAPII CTD Phosphorylation Reveals That the Identity and Modification of Seventh Heptad Residues Direct Tyr1 Phosphorylation. *ACS Chem. Biol.* **2019**, *14*, 2264–2275.

(43) Ramani, M. K. V.; Escobar, E. E.; Irani, S.; Mayfield, J. E.; Moreno, R. Y.; Butalewicz, J. P.; Cotham, V. C.; Wu, H.; Tadros, M.; Brodbelt, J. S.; Zhang, Y. J. Structural Motifs for CTD Kinase Specificity on RNA Polymerase II during Eukaryotic Transcription. *ACS Chem. Biol.* **2020**, *15*, 2259–2272.

(44) Escobar, E. E.; King, D. T.; Serrano-Negrón, J. E.; Alteen, M. G.; Vocadlo, D. J.; Brodbelt, J. S. Precision Mapping of O-Linked *N* -Acetylglucosamine Sites in Proteins Using Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2020**, *142*, 11569–11577.

(45) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135*, 12646–12651.

(46) Zhang, L.; Reilly, J. P. De Novo Sequencing of Tryptic Peptides Derived from Deinococcus Radiodurans Ribosomal Proteins Using 157 Nm Photodissociation MALDI TOF/TOF Mass Spectrometry. *J. Proteome Res.* **2010**, *9*, 3025–3034.

(40) Chong, C.; Marino, F.; Pak, H.; Racle, J.; Daniel, R. T.; Müller, M.; Gfeller, D.; Coukos, G.; Bassani-Sternberg, M. High-Throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferonγ-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. *Mol. Cell Proteom.* **2018**, *17*, 533–548.

(48) Marino, F.; Chong, C.; Michaux, J.; Bassani-Sternberg, M. High-Throughput, Fast, and Sensitive Immunopeptidomics Sample Processing for Mass Spectrometry. *Immune Checkpoint Blockade: Methods and Protocols* **2019**, 67–79.

(49) Kowalewski, D. J.; Stevanović, S. Biochemical Large-Scale Identification of MHC Class I Ligands. *Methods Mol. Biol* **2013**, No. 960, 145–157.

(50) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 1044–1051.

(51) Krokhin, O. V.; Craig, R.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. An Improved Model for Prediction of Retention Times of Tryptic Peptides in Ion Pair Reversed-Phase HPLC: Its Application to Protein Peptide Mapping by Off-Line HPLC-MALDI MS. *Mol. Cell. Proteom.* **2004**, *3*, 908–919.

(52) Okuda, S.; Watanabe, Y.; Moriya, Y.; Kawano, S.; Yamamoto, T.; Matsumoto, M.; Takami, T.; Kobayashi, D.; Araki, N.; Yoshizawa, A. C.; Tabata, T.; Sugiyama, N.; Goto, S.; Ishihama, Y. JPOSTrepo: An International Standard Data Repository for Proteomes. *Nucleic Acids Research* **2017**, *45*, D1107–D1111.

(53) Vita, R.; Overton, J. A.; Greenbaum, J. A.; Ponomarenko, J.; Clark, J. D.; Cantrell, J. R.; Wheeler, D. K.; Gabbard, J. L.; Hix, D.; Sette, A.; Peters, B. The Immune Epitope Database (IEDB) 3.0. *Nucleic Acids Res* **2015**, *43*, D405–D412.

(54) Xiao, Y.; Vecchi, M. M.; Wen, D. Distinguishing between Leucine and Isoleucine by Integrated LC–MS Analysis Using an Orbitrap Fusion Mass Spectrometer. *Anal. Chem.* **2016**, *88*, 10757–10766.

(55) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-Nm Photodissociation of Singly and Multiply Charged Peptide Anions for Acidic Proteome Characterization. *Proteomics* **2011**, *11*, 1329–1334.

(56) Xiao, Y.; Vecchi, M. M.; Wen, D. Distinguishing between Leucine and Isoleucine by Integrated LC–MS Analysis Using an Orbitrap Fusion Mass Spectrometer. *Anal. Chem.* **2016**, *88*, 10757–10766.

(57) Mylonas, R.; Beer, I.; Iseli, C.; Chong, C.; Pak, H.-S.; Gfeller, D.; Coukos, G.; Xenarios, I.; Müller, M.; Bassani-Sternberg, M. Estimating the Contribution of Proteasomal Spliced Peptides to the HLA-I Ligandome*. *Mol. Cell Proteomics* **2018**, *17*, 2347–2357.

(58) Bettencourt, P.; Müller, J.; Nicastri, A.; Cantillon, D.; Madhavan, M.; Charles, P. D.; Fotso, C. B.; Wittenberg, R.; Bull, N.; Pinpathomrat, N.; Waddell, S. J.; Stylianou, E.; Hill, A. V. S.; Ternette, N.; McShane, H. Identification of Antigens Presented by MHC for Vaccines against Tuberculosis. *NPJ Vaccines* **2020**, *5*, 1–14.

(59) Liepe, J.; Sidney, J.; Lorenz, F. K. M.; Sette, A.; Mishto, M. Mapping the MHC Class I–Spliced Immunopeptidome of Cancer Cells. *Cancer Immunol Res* **2019**, *7*, 62–76.

(60) Halim, M. A.; Girod, M.; MacAleese, L.; Lemoine, J.; Antoine, R.; Dugourd, P. 213 Nm Ultraviolet Photodissociation on Peptide Anions: Radical-Directed Fragmentation Patterns. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 474–486.

(61) Lemoine, J.; Tabarin, T.; Antoine, R.; Broyer, M.; Dugourd, P. UV Photodissociation of Phospho-Seryl-Containing Peptides: Laser Stabilization of the Phospho-Seryl Bond with Multistage Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 507–511.

(62) Zhang, X.; Julian, R. R. Radical Additions to Aromatic Residues in Peptides Facilitate Unexpected Side Chain and Backbone Losses. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 626–635.

(63) Rolfs, Z.; Solntsev, S. K.; Shortreed, M. R.; Frey, B. L.; Smith, L. M. Global Identification of Post-Translationally Spliced Peptides with Neo-Fusion. *J. Proteome Res.* **2019**, *18*, 349–358.

(64) Rolfs, Z.; Müller, M.; Shortreed, M. R.; Smith, L. M.; Bassani-Sternberg, M. Comment on "A Subset of HLA-I Peptides Are Not Genomically Templated: Evidence for Cis- and Trans-Spliced Peptide Ligands." *Sci. Immunol.* **2019**, *4*.

(65) Andreatta, M.; Lund, O.; Nielsen, M. Simultaneous Alignment and Clustering of Peptide Data Using a Gibbs Sampling Approach. *Bioinformatics* **2013**, *29*, 8–14.

Chapter 7: Conclusion

7.1 Conclusions

Together, the work presented in this dissertation offers an expanded understanding of the mass spectrometry toolkit available to aid in the characterization and development of biotherapeutics. Several biotherapeutic proteins are explored, including multiple modalities of antibody-drug conjugates (ADCs) and synthetic selenoproteins. Human leucocyte antigen (HLA) immunopeptides are an additional therapeutic avenue that was explored in detail. For each project, multiple MS/MS methods were investigated including 193 and 213 nm ultraviolet photodissociation (UVPD), higher-energy collision dissociation (HCD), electron transfer dissociation (ETD), and electron-transfer/higher-energy collision dissociation (EThcD).

In Chapter 3 a subunit-based middle-down approach integrating 193 nm UVPD, ETD, and EThcD was presented, resulting in extensive characterization of cysteine-linked ADCs. To maximize the quality of the data generated, a data-independent MS/MS acquisition was implemented. Payload-containing C-terminal ions proved essential for confirming conjugation location due to the position of the payloads on the peptide sequence. While the abundances of these key ions were low, combining ion activation methods produced sufficient complementary fragment ions to identify payload locations. Claims about payload conjugation sites for positional isomers from each chromatographic peak were supported by relative quantitation of the subunits. The integrated approach allowed for confident localization of the payloads.

While the subunit-based approach was successfully in characterizing the cysteine-linked ADC, a more complex lysine-linked ADC required a different approach. In Chapter 4, a peptide-based middle-down method, powered by limited proteolysis and

HCD-triggered MS/MS, enabled the characterization of a highly heterogeneous lysine-linked ADC. The HCD-triggered method enabled the generation of high-quality data for large payload-containing peptides, without sacrificing duty cycle. EThcD and 193 nm UVPD data were both examined independently, but the combination proved the most powerful for large peptides with multiple potential payload sites. Payload-containing fragment ions generated through EThcD and 193 nm UVPD proved integral to localizing conjugation sites, differentiating positional isomers, and enabling complete characterization. A total of 44 conjugation sites out of 92 modifiable positions were unambiguously identified. Deeper understanding of the combinatorial nature of conjugation was enabled through the detailed characterization of two peptides containing multiple conjugations.

ADCs represent one growing and prominent modality for biotherapeutics. A newer technique for the modification of therapeutic proteins is the incorporation of selenocysteine to facilitate the generation of bridged proteins, as examined in Chapter 5. Chapter 5 demonstrates the combination of intact, top-down, and bottom-up methods to aid in the characterization of synthetic selenoproteins. For both top-down and bottom-up characterization, a combination of 213 nm UVPD and EThcD proved ideal. For the top-down approach, EThcD resulted in cleavage of diselenide bonds, enabling the most complete sequence coverage. 213 nm UVPD also generated high coverage, but because diselenide bonds remained intact, it was possible to confirm the locations of the diselenide bridges. For the bottom-up method, 213 nm UVPD yielded the most complete sequence coverage, while EThcD generated diagnostic fragments that aided in the confirmation of the presence of diselenide-bridged peptides. The bottom-up strategy also lead to the discovery of heterogeneity in the diselenide bridges, some of which could only

be explained by the presence of dimeric species. The existence of the dimeric species was confirmed through lower resolution intact mass measurements. The methods developed helped to improve the understanding of synthetic selenoproteins and offered a strategy to aid in the development of selenocysteine-based biotherapeutics.

While ADCs and other large proteins are one of the more largely discussed area for biotherapeutic development, personalized peptide therapies are an exciting avenue to offer selective treatment powered by immunopeptidomics.[1] In Chapter 6, the potential for 193 nm UVPD to improve the characterization of HLA immunopeptides was explored and compared to HCD and EThD, which are more often utilized in immunopeptidomics workflows. 193 nm UVPD proved to generate the most complete characterization of immunopeptides, including the generation of more diagnostic neutral loss fragments to aid in the differentiation of leucine and isoleucine than EThcD. 193 nm UVPD resulted in more peptide identification than EThcD, but less than HCD, indicating that 193 nm UVPD is a preferred complementary method to enhance the characterization afforded by HCD.

7.2 Future Directions

The work presented in this dissertation represents and exciting step towards expanding the applications of UVPD. The expansion in applications should continue to be explored. Top-down mass spectrometry was one area described in Chapter 1 that has been investigated for antibodies, but the expansion to ADCs has been limited. Preliminary data demonstrating the potential for top-down to be expanded to lysine-linked ADCs is detailed in section 7.2.1. To further improve the tools available to

characterize antibodies, additional studies should also focus on the characterization of therapeutic antibody modalities beyond IgG1, as discussed in Chapter 1.

HLA immunopeptides represent another key area where UVPD has shown promise to improve characterization. The results presented in this dissertation demonstrated the potential for UVPD to drastically improve the understanding of individual therapeutically relevant peptides. While the identification and characterization of thousands of peptides was an exciting outcome, further studies can increase the biological relevance by specifically examining tumor associated antigens (TAAs). TAAs include overexpressed self-antigens and tumor specific neoantigens.[1] A proposed approach would involve a preliminary HCD based screening to identify TAAs followed by a targeted UVPD based method to completely characterize them. This approach would ensure high quality UVPD data is generated, improving characterization including leucine and isoleucine differentiation.

While biotherapeutics and immunopeptides have already been explored with advanced MS/MS methodologies, the translation of these methods to food applications has been relatively limited.[2] The increasing interest in developing plant-based and other alternative protein solutions as food products through biotechnology research offers several new avenues for investigation. The devastating environmental impacts of animal agriculture bring a further urgency to this work.[3] Section 7.2.2 outlines a preliminary exploration of glycoproteomics-based characterization of pea protein samples. In addition to building on the work outlined here, additional studies may be developed to characterize other plant-based proteins, and further build on the transfer of mass spectrometry-based methods for applications in the emerging fields of precision fermentation and cultured meat.[4,5]

7.2.1 Top-Down of Lysine-Linked ADC

Improving the characterization of individual ADC conjugation sites as demonstrated in Chapters 3 and 4 is one important approach to increasing the understanding of ADCs. Generating information about the relative contribution of key regions of the ADCs to the DAR is another fruitful avenue. Middle-up approaches can be used to assess the number of drugs conjugated to each antibody subunit.[6,7] While the examination of individual subunits can be informative, digestion and/or reduction prior to mass spectrometry analysis negates the ability to relate the level of conjugation on each subunit to the conjugation state of the intact ADC. One study of a cysteine-linked ADC utilized reduction of disulfide bonds followed by native mass spectrometry.[8] Individual intact drug to antibody ratio (DAR) species were then isolated and fragmented with HCD, yielding released light chain and enabling assessment of the fraction of drug conjugated light chain present for multiple intact DARs.[8] It was more recently demonstrated that electron capture dissociation and 157 nm UVPD both lead to release of antibody light chains through the cleavage of disulfide bonds.[9] This opens up an exciting avenue to characterize the relative contribution of the light chain to the antibody DAR with minimal sample processing. In the remainder of this section, it will be demonstrated that 193 nm UVPD can lead to the release of the light chain and the drug conjugated light chain for a lysine-linked ADC. Furthermore, the top-down method allows the possibility of generating a relationship between the number of payloads bound to the intact antibody and the distribution between the heavy and light chains.

Figure 7.1 displays the intact spectrum of the lysine linked ADC. While the intact mass spectrum in **Figure 7.1A** is highly heterogeneous, deconvolution, as demonstrated in **Figure 7.1B,** allows the differentiation of different antibody DAR species as well as a

variety of glycoforms. Multiple glycoforms were co-isolated in MS/MS experiments to maintain high signal while isolating individual intact antibody DAR species. **Figure 7.2A** demonstrates the fragmentation of the DAR4 species in the 25+ charge state, resulting in the release of the light chain. In addition to the intact light chain, the light chain was observed with one and two payload conjugations. The average number of payloads conjugated to the light chain was then calculated based on the relative abundance of each light chain species. This was repeated for intact antibody DARs 1 through 5, resulting in the chart included in **Figure 7.2B**. The relatively linear relation observed between intact antibody DAR and the average number of payloads conjugated to the light chain, indicates that increasing intact antibody DAR does not change the distribution of payloads between the heavy and light chain. Further studies to complete the work demonstrated here should assess the capabilities of electron capture dissociation and HCD to release ADC light chains. The assessment of multiple intact antibody charge states, as well as multiple samples using complementary MS/MS methods will further increase confidence in the method. Overall, the top-down fragmentation of lysine-linked ADCs represents a promising method to gain further insight into ADC structure.

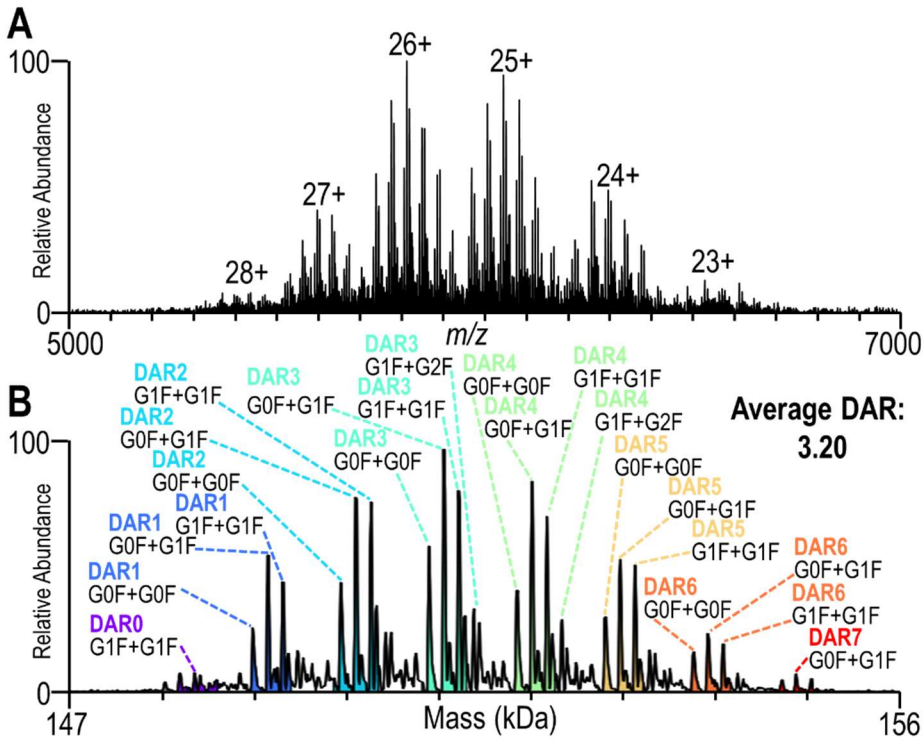


Figure 7.1: Mass spectrum of intact Trastuzumab emtansine (A) with charge states labeled and (B) after deconvolution with UniDec. A resolution of 50,000 at *m/z* 400 was used and 100 spectra were averaged.

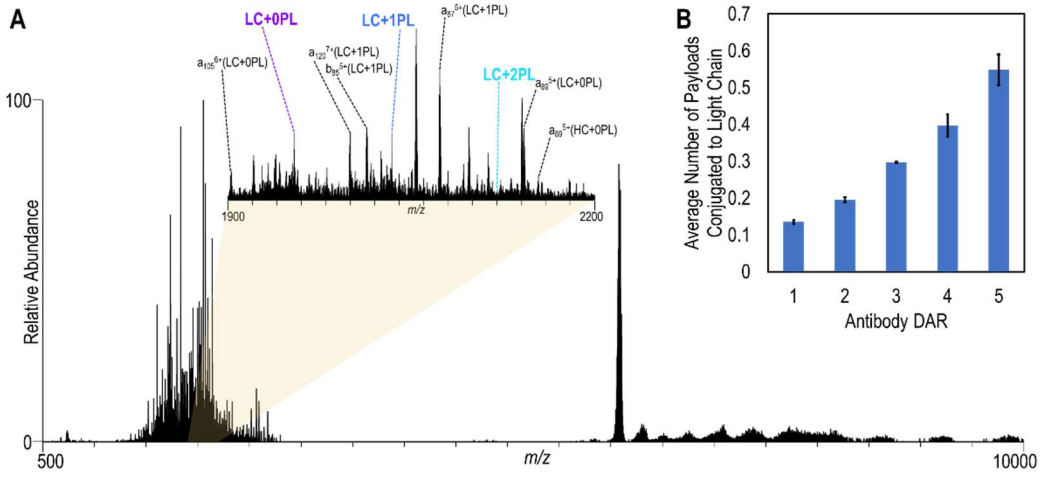


Figure 7.2: (A) Top-down UVPD spectrum of the DAR4 species in the 25+ charge state. Expanded region shown in the inset is displays fragment ions identified in the dense spectral region, including the released light chain with 0, 1 or 2 conjugated payloads, all observed in the 12+ charge state. For 193 nm UVPD one 0.25 mJ laser pulse was applied. (B) Bar chart displaying the relationship between the intact antibody DAR and the average number of payloads conjugated to the light chain.

7.2.2 Characterization of Pea Protein

While the application of proteomics to mammalian proteins has been extensively studied, plant proteomics is a relatively new field. Plant proteomics presents unique challenges for the generation of reference proteomes and the identification of different post-translational modifications (PTMs).[10] Given the growing interest in plant-based meat and dairy alternatives, the adaptation of advanced proteomics techniques to plants is long overdue.[11,12] Both primary protein sequence and PTMs can affect functional properties of proteins resulting in changes to flavor, solubility, texture, and other properties critical to food development.[13] Protein glycosylation is of particular interest. Chemical glycosylation of plant-based proteins, typically through a Maillard reaction, has been

shown to improve functional properties.[13] While the reported results were promising, the industry is moving away from chemical modification to favor clean label products, making the exploration of naturally occurring glycosylation an interesting alternative.[13] For an initial proof of concept, a bottom-up strategy utilizing 193 nm ultraviolet photodissociation (UVPD) was developed to compare the type and degree of glycosylation in two pea protein isolate samples.

Initial sample preparation utilized ProtiFi S-Trap spin columns to enable complete digestion of the pea protein isolate despite its poor solubility. Following digestion, electrostatic repulsion-hydrophilic interaction chromatography-based enrichment allowed the specific enhancement of the N-glycopeptides. After enrichment the signal of each glycopeptide was increased, affording both an increased number of glycopeptide identifications and improved characterization. Glycopeptides were identified and quantified using Byonic and Byologic modules of Byos 4.4 by Protein Metrics. While N-glycans were identified on several species of peptides, the most notable was a peptide originating from vicilin (47 kDa). Vicilin is a seed storage protein and a major component of legumes, making glycosylation on this protein a noteworthy observation. For that reason, the peptide related to vicilin was examined in more detail, starting with tandem MS fragmentation of an exemplative glycopeptide. While HCD fragmentation of the glycopeptide did not lead to complete characterization of the peptide sequence or the glycan structure, it produced highly abundant oxonium ions specific to glycosylation (**Figure 7.3A**), enabling the implementation of an HCD-triggered UVPD method.[14] For this method, the identification of oxonium ions in the HCD spectra was used to prompt isolation and acquisition of UVPD in the subsequent scans. The inclusion of UVPD resulted in enhanced characterization of both the peptide sequence and the glycan

structure, as demonstrated in **Figure 7.3B.** While HCD predominantly generated glycan fragments released from the peptide (B ions), UVPD resulted in many glycan fragments containing the intact peptide (Y ions). Glycan fragmentation is illustrated in greater detail in **Figure 7.3C**. UVPD also resulted in the generation of peptide sequence ions retaining the entire N-glycan, an outcome not achieved with HCD. This strategic combination of UVPD and HCD, in the form of an HCD-triggered UVPD method, led to the best characterization of N-glycosylated peptides.

To better understand the correlation between N-glycan structures and food functional properties, the proteomic signatures of two different pea-protein isolates were examined. Pea protein isolates 1 and 2 were comprised of 80% and 84% protein content, respectively. **Figure 7.4** displays the extracted ion chromatograms (EICs) for the unmodified vicilin peptide and three different glycosylation states for each sample examined. The integrated area for the area under the curve (AUC) calculation is also highlighted. To compare the amount of glycosylated vicilin between each sample, the AUC is normalized to the unmodified peptide. **Figure 7.5** displays the result of this normalization, thus conveying the relative degrees of glycosylation. Although the glycosylation values were calculated based on the enriched samples and do not represent absolute percentages of vicilin glycoproteins, these ratios can be used to compare the degree of glycosylation between the two samples because both were processed in the same way. Overall, the results in **Figure 7.5** indicate greater glycosylation of vicilin in pea protein isolate 2, the 84% protein sample. Although the overall differences are minor, this initial assessment illustrates the relative quantitation of N-glycosylation between the two pea protein isolates.

The initial results demonstrate the capabilities of 193 nm UVPD to characterize glycopeptides found in pea protein isolate and the potential to apply relative quantitation of glycopeptides to different pea protein samples to generate more informative data. Additional samples should be examined to evaluate the degree of glycosylation based on a greater variation in processing techniques. Doing so will aid in determining if a relationship exists between the N-glycosylation present in pea protein samples and their functional properties.

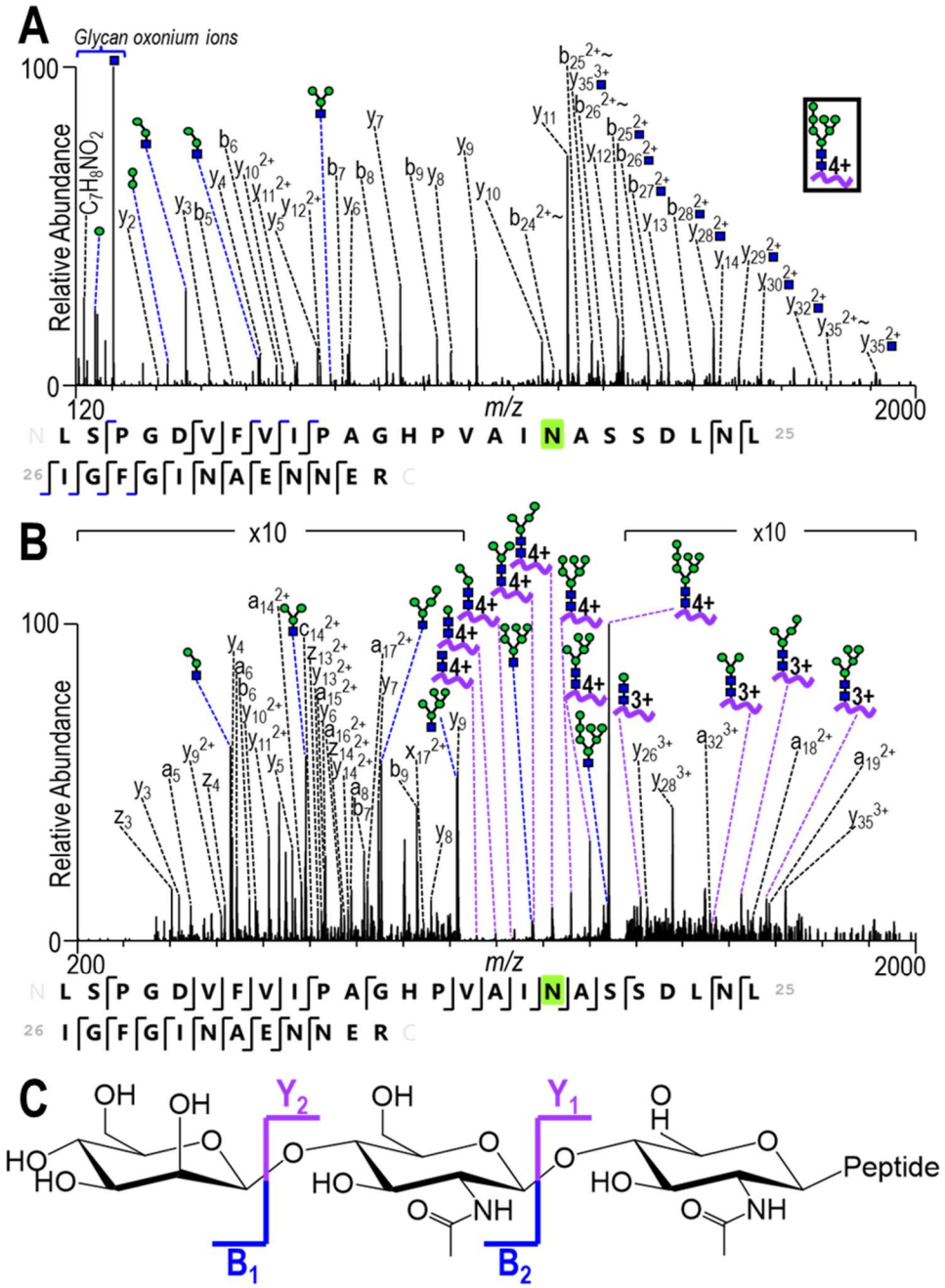


Figure 7.3: (A) HCD and (B) UVPD spectra for the vicilin glycopeptide with HexNAc(2)Hex(7) N-glycan (4+ charge state). (C) Scheme demonstrating cleavage locations of B and Y fragment ions mapped onto an N-glycopeptide containing HexNAc(2)Hex(1).

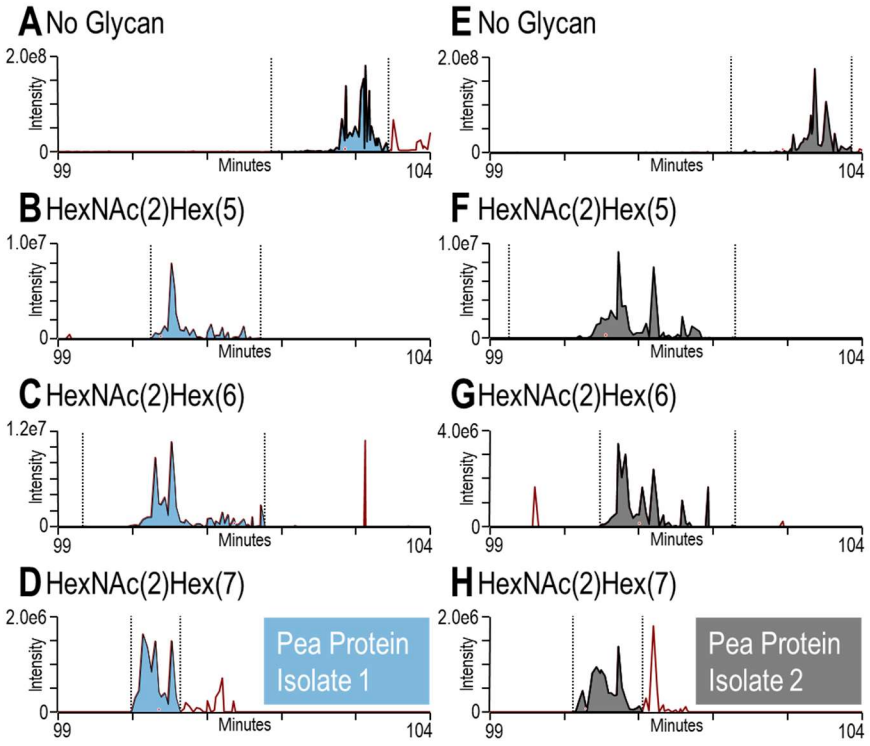


Figure 7.4: Extracted ion chromatograms used to quantify the glycopeptides in the two samples. All EICs are based on the 3+ charge state of the peptide LSPGDVFVIPAGHPVAIN*ASSDLNLIGFGINAENNER in different glycosylation states.

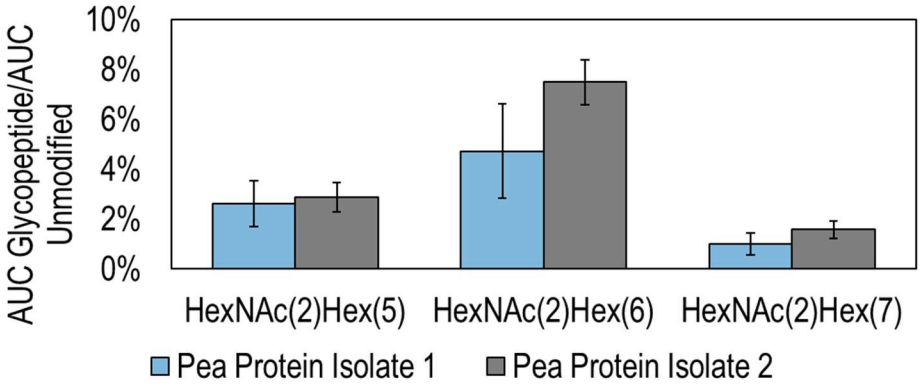


Figure 7.5: The ratios of the area under the curve (AUC) of the glycopeptide divided by the AUC of the unmodified peptide are shown for three glycopeptides of vicilin. All AUCs were generated using EICs of the 3+ charge state.

7.3 REFERENCES

(1) Creech, A. L.; Ting, Y. S.; Goulding, S. P.; Sauld, J. F. K.; Barthelme, D.; Rooney, M. S.; Addona, T. A.; Abelin, J. G. The Role of Mass Spectrometry and Proteogenomics in the Advancement of HLA Epitope Prediction. *Proteomics* **2018**, *18*, 1700259.

(2) Valdés, A.; Álvarez-Rivera, G.; Socas-Rodríguez, B.; Herrero, M.; Ibáñez, E.; Cifuentes, A. Foodomics: Analytical Opportunities and Challenges. *Anal. Chem.* **2022**, *94*, 366–381.

(3) Eisen, M. B.; Brown, P. O. Rapid Global Phaseout of Animal Agriculture Has the Potential to Stabilize Greenhouse Gas Levels for 30 Years and Offset 68 Percent of CO2 Emissions This Century. *PLOS Climate* **2022**, *1*, e0000010.

(4) Teng, T. S.; Chin, Y. L.; Chai, K. F.; Chen, W. N. Fermentation for Future Food Systems. *EMBO reports* **2021**, *22*, e52680.

(5) Rubio, N. R.; Xiang, N.; Kaplan, D. L. Plant-Based and Cell-Based Approaches to Meat Production. *Nat Commun* **2020**, *11*, 6276.

(6) Martelet, A.; Garrigue, V.; Zhang, Z.; Genet, B.; Guttman, A. Multi-Attribute Method Based Characterization of Antibody Drug Conjugates (ADC) at the Intact and Subunit Levels. *J. Pharm. Biomed. Anal.* **2021**, *201*, 114094.

(7) Sandra, K.; Vanhoenacker, G.; Vandenheede, I.; Steenbeke, M.; Joseph, M.; Sandra, P. Multiple Heart-Cutting and Comprehensive Two-Dimensional Liquid Chromatography Hyphenated to Mass Spectrometry for the Characterization of the Antibody-Drug Conjugate Ado-Trastuzumab Emtansine. *J. Chromatogr. B* **2016**, *1032*, 119–130.

(8) Dyachenko, A.; Wang, G.; Belov, M.; Makarov, A.; de Jong, R. N.; van den Bremer, E. T. J.; Parren, P. W. H. I.; Heck, A. J. R. Tandem Native Mass-Spectrometry on Antibody–Drug Conjugates and Submillion Da Antibody–Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. *Anal. Chem.* **2015**, *87*, 6095–6102.

(9) Shaw, J. B.; Liu, W.; Vasil′ev, Y. V.; Bracken, C. C.; Malhan, N.; Guthals, A.; Beckman, J. S.; Voinov, V. G. Direct Determination of Antibody Chain Pairing by Top-down and Middle-down Mass Spectrometry Using Electron Capture Dissociation and Ultraviolet Photodissociation. *Anal. Chem.* **2020**, *92*, 766–773.

(10) Mergner, J.; Kuster, B. Plant Proteome Dynamics. *Annual Review of Plant Biology* **2022**, *73*, null.

(11) The Not Company Is Looking to Start a Food Revolution from Chile. *TechCrunch*.

(12) Hampton Creek's Data Scientists Team up with Chefs to Find the Holy Grail of Plant Proteins. *VentureBeat*, 2014.

(13) Akharume, F. U.; Aluko, R. E.; Adedeji, A. A. Modification of Plant Proteins for Improved Functionality: A Review. *Comprehensive Reviews in Food Science and Food Safety* **2021**, *20*, 198–224.

(14) Escobar, E. E.; Wang, S.; Goswami, R.; Lanzillotti, M. B.; Li, L.; McLellan, J. S.; Brodbelt, J. S. Analysis of Viral Spike Protein N-Glycosylation Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2022**, *94*, 5776–5784.

Appendix

List of peptides included in the 157 synthetic peptide mixture discussed in Chapter 6.

Number	Peptide sequence
1	AIVDKVPSV
2	ALADGVQKV
3	ALADIVWRA
4	ALAEEVEQV
5	ALAGDQPSV
6	ALAPAPPQV
7	ALASLIRSV
8	ALATLIHQV
9	ALDKALTSV
10	ALDSQVPKV
11	ALISNSHQL
12	ALITRIFGV
13	ALLDKLYAL
14	ALLDRIVSV
15	ALLDSAHLL
16	ALLETVNRL
17	ALLGDLTKA
18	ALLNIKVKL
19	ALLQSDVRL
20	ALLSRLEQI
21	ALLTYLEQV
22	ALMDEVVKA
23	ALMPVLNQV
24	ALNEKLVNL
25	ALNELLQHV
26	ALQEMVHQV
27	ALSDHHIYL
28	ALSEKIVSV
29	ALSPHNILL

Appendix continued

30	ALTNAVAHV
31	ALVVQVAEA
32	ALYDEVRTV
33	AMSSKFFLV
34	AVANIVNSV
35	FIASKGVKLV
36	FLADPSAFVAA
37	FLDGNELTL
38	FLDGNEMTL
39	FLDHIIASV
40	FLFEPVVKA
41	FLGEKIASV
42	FLLDKKIGV
43	FLQEHNTTL
44	FLQEKSPAV
45	FLVDGVCTV
46	GLAKAVASV
47	GLALKVQEV
48	GLAPKPVQV
49	GLATDVQTV
50	GLIDGVVEA
51	GLIDHQTYL
52	GLIDKVNEL
53	GLIDRQVTV
54	GLMDTVKKV
55	GLNEEIARV
56	GLVDEKALAQA
57	GLWEEAYRV
58	GLWGPVHEL
59	GLYSGVTTV
60	GVIDGHIYAV
61	HIIENIVAV
62	HVLEEVQQV
63	ILAAVETRL
64	ILDKKVEKV

Appendix continued

65	ILDQKINEV
66	ILHDDEVTV
67	ILLKDILSV
68	ILMEHIHKL
69	ILQAHLHSL
70	ILQQHIATV
71	ILSEVQQAV
72	ILTDITKGV
73	ILWETVPSM
74	ILYGEVEKL
75	ILYGKIIHL
76	IMLEALERV
77	IVAESLQQV
78	KILPTLEAV
79	KIYEGQVEV
80	KLDAFVEGV
81	KLDPTKTTL
82	KLDQDLNEV
83	KLGSVPVTV
84	KLHGVNINV
85	KLIDDVHRL
86	KLNPQQFEV
87	KLQELNYNL
88	KLSELEAAL
89	KVAPAPAVV
90	KVLSKEFHL
91	LLANKVPAA
92	LLDEEISRV
93	LLDRFLATV
94	LLDVPTAAV
95	LLIENVASL
96	LLPPQPALA
97	MLFGHPLLV
98	MVAPAVASV
99	NLAEDIMRL

Appendix continued

100	NLAEKLIGV
101	NLAENISRV
102	NLASFIEQV
103	NLLPKLHIV
158	NLVPMVATV
104	PMFIVNTNVPRASVPD
105	QLAQFVHEV
106	QLDDLKVEL
107	QLIDKVWQL
108	QLLGSAHEV
109	RLASYLDKV
110	RLASYLDRV
111	RLAVYIDRV
112	RLWGEPVNL
113	SIYPSPTGV
114	SLADVHIEV
115	SLAEGLRTV
116	SLAGGIIGV
117	SLAQYLINV
118	SLAQYNPKL
119	SLDAKEIYL
120	SLDLTTTCV
121	SLFGGSVKL
122	SLFPGKLEV
123	SLIRNLEQL
124	SLKLQASNVTNKND
125	SLLDKIIGA
126	SLLENLEKI
127	SLQDEIQRV
128	SLQEEIAFL
129	SLSEKTVLL
130	SLVSKGTLVQTK
131	SLWGQPAEA
132	SLYDYNPNL
133	TLADLVHHV

Appendix continued

134	TLAEHIIEA
135	TLDDKKVYL
136	TLSEVTNQL
137	TLVYHVVGV
138	TLWVDPYEV
139	TLYEAVREV
140	TMADQIVTV
141	TMAKESSIIGV
142	VIDGHIYAV
143	VLAEKLAAI
144	VLAETLTQV
145	VLDDKLVFV
146	VLFTGVKEV
147	VLIDYQRNV
148	VLQGAVPTV
149	VLSPADKTNV
150	VLSPADKTNVK
151	VLSPADKTNVKAA
152	VLYHVETEV
153	VMDSKIVQV
154	VMLDVPIRL
155	VSDGVIKV
156	YLFERIKEL
157	YLLPAIVHI

Vita

Ellie grew up in a Harvard Massachusetts, an idyllic New England town approximately 30 miles northwest of Boston. At the beginning of her senior year of high school Ellie applied to college as an "undeclared" major but having completed every chemistry class offered by her high school, she realized that something was missing. Ellie started a Bachelor of Science in chemistry at Northeastern University in the Fall of 2012. During her five years at Northeastern, Ellie participated in undergraduate research under the supervision of Professor Ke Zhang, studied abroad at the University of Edinburgh, and completed a total of four six-month co-ops, including one at the Harvard School of Public, one at Moderna, and two at GlaxoSmithKline. The opportunity to rapidly move between research experiences strengthened Ellie's excitement about science and inspired her to pursue a PhD with a focus on mass spectrometry. In 2017, Ellie joined the research group of Jennifer Brodbelt at the University of Texas at Austin. During her time there she had the opportunity to participate in and develop external collaborations with industry professionals and complete a summer internship at Impossible Foods. Vegetarian since 2013 and vegan since 2018, Ellie is passionate about animal rights, climate change, and other environmental issues. In her free time, Ellie enjoys hiking with her rescue dog, a beagle mix named Daryl.

Permanent email address: eleanorwatts@utexas.edu