Title

Internal Fragments in Top-down and Middle-down Mass Spectrometry: From Fundamentals to Applications

Author

Wei, Benqian

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Internal Fragments in Top-down and Middle-down Mass Spectrometry: From Fundamentals to Applications

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy in Chemistry

by

Benqian Wei

2023

ABSTRACT OF THE DISSERTATION

Internal Fragments in Top-down and Middle-down

Mass Spectrometry: From Fundamentals to Applications

by

Benqian Wei

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2023

Professor Joseph A. Loo, Chair

The interrogation of protein structure, especially identifying and localizing post-translational modifications (PTMs) and sites of ligand/small molecule binding, is crucial for understanding protein function in biological systems. In particular, the rapid increase of the use and development of therapeutic monoclonal antibodies (mAbs) and antibody-drug conjugates (ADCs) for human health have necessitated the advancement of efficient and accurate analytical methods. Top-down and middle-down mass spectrometry (TD- and MD-MS) have become prominent analytical tools for protein characterization. However, obtaining complete protein sequence coverage by TD-/MD-MS can be limiting, particularly for proteins greater than 30 kDa, e.g., mAbs and ADCs. My dissertation research explores the utility of internal fragments, which

are largely ignored by the MS community as they are difficult to assign, from both fundamental and application perspectives, for more efficient and comprehensive protein sequence, structure, and PTM characterization. On the fundamental side, we demonstrated a relationship between internal fragments and conventional terminal fragments. A better understanding of the fundamental formation mechanism of internal fragments aids the development of sequencing algorithms to assign internal fragments more accurately and reliably. On the application side, we started by analyzing standard disulfide-intact proteins using TD-MS, in which assigning internal fragments helped achieve near complete sequence coverage, and obtain disulfide bond position and connectivity information. This encouraged us to apply TD-MS on proteins of therapeutic significance, i.e., mAbs and ADCs, which are extremely complex disulfide-intact proteins. Incorporating internal fragments analysis allowed us to achieve the highest sequence coverage to date on an intact mAb by TD-MS, and enabled the access of important PTM information and drug binding information on intact ADCs. We then expanded this application by applying MD-MS on reduced mAbs and ADCs. Analyzing internal fragments in MD-MS helped us achieve comprehensive characterization of mAbs and ADCs which is a significant improvement from TD-MS and is comparable to the results obtained from the routinely utilized bottom-up peptide mapping method. Lastly, we demonstrated that assigning internal fragments generated by collision-based fragmentation also helps deliver higher-order structure information of multi-subunit protein assemblies. In summary, this dissertation work contributes to advancing the technique and instrumentation surrounding TD- and MD-MS workflows to achieve better characterization of proteins, especially biotherapeutics, and identification of specific proteoforms.

The dissertation of Benqian Wei is approved.

Anne M. Andrews

James A. Wohlschlegel

Keriann M. Backus

Joseph A. Loo, Committee Chair

University of California, Los Angeles

2023

DEDICATION

In dedication to my loving family and friends

for all their support throughout these years

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

Chapter 6:

LIST OF TABLES

Chapter 2:

Chapter 3:

LIST OF SCHEMES

Chapter 2:

Chapter 6:

ACKNOWLEDGEMENTS

I would like to first express my deepest gratitude to my advisor, Professor Joseph Loo, for his continuous support and guidance throughout my PhD journey. Joe is a tremendous advisor who has supported and encouraged me in every aspect I can imagine. He allowed me to explore my own research interests and eventually found out what I would like to dedicate my whole PhD to. None of the work in this dissertation could have been completed without the mentorship, resources, and instrumentation that Joe provided. I am really thankful for so many great opportunities that Joe gave me to attend conferences, present my work, and network with brilliant scientists from many disciplines. Joe is also a fantastic and caring mentor. He is always there for me when I need to seek advice from him but at the same time allows me to explore and think independently. His support for me to do an internship enables me to find my career interest and prepare myself well for the next step of my career. I also want to sincerely thank Dr. Rachel Ogorzalek Loo for all her support, mentorship, and guidance. Her knowledge in proteomics, physical chemistry, and especially fragmentation mechanisms has shaped my understanding of many fundamental aspects of mass spectrometry and more importantly, facilitated me to always think critically.

I would also like to thank the entire Loo Lab members for these years of help, support, and company. The Loo Lab is truly a big family with whom I always love working and hanging out. I will never forget the good memories, both professionally and casually, with Dr. Muhammad Zenaidee, Dr. Carter Lantz, Dr. Wonhyeuk Jung, Dr. Janine Fu, Boyu Zhao, Andrew Goring, Jessie Le, Eileen Olivares, and Merin Rixen. I want to specifically thank Dr. Muhammad

Zenaidee who mentored me and guided me into the protein mass spectrometry world when I stepped into the office for the first time and had no idea what graduate school is about. Another special shout out goes to Dr. Carter Lantz who is such a great colleague, mentor, and friend. Chapter 6 is a work led by him and I was mostly assisting. I learned so much from him and we had so much fun working together these years that I'm always appreciative of. The Covid-19 pandemic is a tough time for everyone and I couldn't have navigated through those difficult times without these two fantastic buddies of mine.

I would like to thank my committee members, Professor Anne Andrews, Professor Keriann Backus, Professor James Wohlschlegel, and committee chair, Professor Joseph Loo. Thank you for reading and providing feedback on my proposals and this dissertation, and attending my oral exam, fourth year meeting, and finally the exit seminar. Special thanks to my invaluable collaborator, Dr. Iain Campuzano from Amgen, for not only providing samples but more importantly, giving me incredibly helpful feedback and suggestions to push forward the projects. Chapters 4 and 5 are both great collaboration products working with Dr. Campuzano. I would also like to thank the ACS Division of Analytical Chemistry and Agilent Technologies for their financial support.

Lastly, I would like to deeply thank my entire family for their support in exploring and navigating this journey with me. They have always been an emotional support and harbor for me whenever I need to relieve my stress. My mom has been so caring and supportive for not just the last four to five years, but throughout my entire life. Her intelligence, vision, and heart are a large reason that I can complete the PhD degree and write down these words here. I

deeply thank my mom and I truly love her.

Chapter Two of this dissertation is a version of a published manuscript: Wei, B.; Zenaidee, M. A.; Lantz, C.; Ogorzalek Loo, R. R.; Loo, J. A. Towards understanding the formation of internal fragments generated by collisionally activated dissociation for top-down mass spectrometry. *Anal. Chim. Acta* **2022**, *1194*, 339400. DOI: 10.1016/j.aca.2021.339400. Copyright © 2021 Elsevier B.V.

Chapter Three of this dissertation is a version of a published manuscript: Wei, B.; Zenaidee, M. A.; Lantz, C.; Williams, B. J.; Totten, S.; Ogorzalek Loo, R. R.; Loo, J. A., Top-down mass spectrometry and assigning internal fragments for determining disulfide bond positions in proteins. *The Analyst* **2023,** *148* (1), 26-37. DOI: https://doi.org/10.1039/D2AN01517J. Copyright © 2023 Royal Society of Chemistry.

Chapter Four of this dissertation is a version of a published manuscript: Wei, B.; Lantz, C.; Liu, W.; Viner, R.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A., Added Value of Internal Fragments for Top-Down Mass Spectrometry of Intact Monoclonal Antibodies and Antibody–Drug Conjugates. *Analytical Chemistry* **2023,** *95* (24), 9347–9356. DOI: https://doi.org/10.1021/acs.analchem.3c01426. Copyright © 2023 American Chemical Society.

Chapter Five is a submitted masnucript to *Analytical Chemistry*: Wei, B.; Lantz, C.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A. Internal Fragments Enhance Middle-down Mass Spectrometry Structural Characterization of Monoclonal Antibodies and Antibody-drug Conjugates.

Chapter Six of this dissertation is a version of a published manuscript: Lantz, C.; Wei, B.;

Zhao, B.; Jung, W.; Goring, A. K.; Le, J.; Miller, J.; Loo, R. R. O.; Loo, J. A., Native Top-Down Mass Spectrometry with Collisionally Activated Dissociation Yields Higher-Order Structure Information for Protein Complexes. *Journal of the American Chemical Society* **2022,** *144* (48), 21826-21830. DOI: https://doi.org/10.1021/jacs.2c06726.

VITA

EDUCATION

Master of Science, Chemistry 2019-2021
University of California, Los Angeles

Master of Science, Polymer Science 2017-2019
The University of Akron

Bachelor of Engineering, Polymer Materials and Engineering 2014-2018
Sichuan University

SELECTED PUBLICATIONS

Wei, B.; Lantz, C.; R.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A., Internal Fragments Enhance Middle-down Mass Spectrometry Structural Characterization of Monoclonal Antibodies and Antibody-drug Conjugates. *Analytical Chemistry* **2023**, submitted.

Wei, B.; Lantz, C.; Liu, W.; Viner, R.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A., Added Value of Internal Fragments for Top-Down Mass Spectrometry of Intact Monoclonal Antibodies and Antibody–Drug Conjugates. *Analytical Chemistry* **2023**, 95 (24), 9347–9356.

Wei, B.; Zenaidee, M. A.; Lantz, C.; Williams, B. J.; Totten, S.; Ogorzalek Loo, R. R.; Loo, J. A., Top-down mass spectrometry and assigning internal fragments for determining disulfide bond positions in proteins. *The Analyst* **2023,** *148* (1), 26-37. (**Featured on the journal front cover**)

Dunham, S. D.; **Wei, B.**; Lantz, C.; Loo, J. A.; Brodbelt, J. S., Impact of Internal Fragments on Top-Down Analysis of Intact Proteins by 193 nm UVPD. *Journal of Proteome Research* **2023,** *22* (1), 170-181.

Lantz, C.; **Wei, B.**; Zhao, B.; Jung, W.; Goring, A. K.; Le, J.; Miller, J.; Loo, R. R. O.; Loo, J. A., Native Top-Down Mass Spectrometry with Collisionally Activated Dissociation Yields Higher-Order Structure Information for Protein Complexes. *Journal of the American Chemical Society* **2022,** *144* (48), 21826-21830.

Wei, B.; Zenaidee, M. A.; Lantz, C.; Ogorzalek Loo, R. R.; Loo, J. A., Towards Understanding the Formation of Internal Fragments Generated by Collisionally Activated Dissociation for Top-Down Mass Spectrometry. *Analytica Chimica Acta* **2022,** *1194*, 339400.

Zenaidee, M. A.; **Wei, B.**; Lantz, C.; Wu, H. T.; Lambeth, T. R.; Diedrich, J. K.; Ogorzalek Loo, R. R.; Julian, R. R.; Loo, J. A., Internal Fragments Generated from Different Top-Down Mass Spectrometry Fragmentation Methods Extend Protein Sequence Coverage. *Journal of The American Society for Mass Spectrometry* **2021,** *32* (7), 1752-1758.

Lantz, C.; Zenaidee, M. A.; **Wei, B.**; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A., ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *Journal of Proteome Research* **2021,** *20* (4), 1928–1935.

SELECTED CONFERENCE PRESENTATIONS

Wei, B.; Lantz, C.; Loo, R. R.; Campuzano, I.; Loo, J. A., Internal Fragments Enhance Middle-down Mass Spectrometry Structural Characterization of Monoclonal Antibodies and Antibody-drug Conjugates. *2023 Advancing Mass Spectrometry for Biophysics and Structural Biology.* **2023**, poster presentation.

Wei, B.; Lantz, C.; Loo, R. R.; Campuzano, I.; Loo, J. A., Internal Fragments Enhance Middle-down Mass Spectrometry Structural Characterization of Monoclonal Antibodies and Antibody-drug Conjugates. *71st ASMS Conference on Mass Spectrometry & Allied Topics.* **2023**, oral presentation.

Wei, B.; Lantz, C.; Liu, W.; Viner, R.; Loo, R. R.; Campuzano, I.; Loo, J. A., Applying Internal Fragments for Native Top-down Mass Spectrometry of Biotherapeutic Proteins. *70th ASMS Conference on Mass Spectrometry & Allied Topics.* **2022**, oral presentation.

Wei, B.; Zenaidee, M.; Lantz, C.; Loo, R. R.; Loo, J. A., Internal Fragments in Top-Down Mass Spectrometry: From Fundamentals to Applications. *27th Lake Arrowhead Conference on Ion Chemistry and Mass Spectrometry.* **2022**, oral presentation.

Wei, B.; Zenaidee, M.; Lantz, C.; Loo, R. R.; Loo, J. A., Determining the Location of Disulfide Bonds by Analyzing Internal Fragments from Top-down Mass Spectrometry of Disulfide Intact Proteins. *69th ASMS Conference on Mass Spectrometry & Allied Topics*. **2021**, poster presentation.

Wei, B.; Zenaidee, M.; Lantz, C.; Loo, R. R.; Loo, J. A., Towards a General Understanding of the Formation of Internal Fragments Generated by Collisionally Activated Dissociation. *1st Southern California Mass Spectrometry Symposium*. **2020**, oral presentation.

SELECTED AWARDS

Roberts A. Smith Excellence in Research Award	**May 2023**
2023 Agilent/ACS Analytical Graduate Fellowship	**April 2023**
ASMS 2023 Graduate Student Travel Award	**March 2023**
Jim and Barbara Tsay Excellence in Second Year Research and Academics Award	**May 2022**
College of Letters and Science Stone Award	**April 2022**
2022 Pittcon/ACS Analytical Chemistry Summer Graduate Fellowship (declined)	**April 2022**

Chapter 1: Introduction

1. Mass spectrometry and tandem mass spectrometry

1.1. Mass spectrometry

Mass spectrometry (MS) is a powerful technique widely used to identify molecular compositions and analyze individual substances and complex mixtures including small molecules,[1] lipids,[2] carbohydrates,[3] oligonucleotides,[4] polymers,[5, 6] and proteins.[7-9] This is achieved by measuring the mass-to-charge ratio (*m/z*) of the analyte ions and separating them in the gas-phase.[10] Each MS instrument is comprised of five major components: inlet system, ionization source, mass analyzer, ion detector and data system (Figure 1).[10] In a single MS experiment, the sample is introduced into the mass spectrometer by direct infusion or chromatography (inlet system), followed by being converted into gas-phase ions within the ionization source. These analyte ions are measured and separated based on their *m/z* values in the mass analyzer, detected by the ion detector and then translated into electric signals. Finally, the data system converts these signals into digital information, presenting it as a mass spectrum.

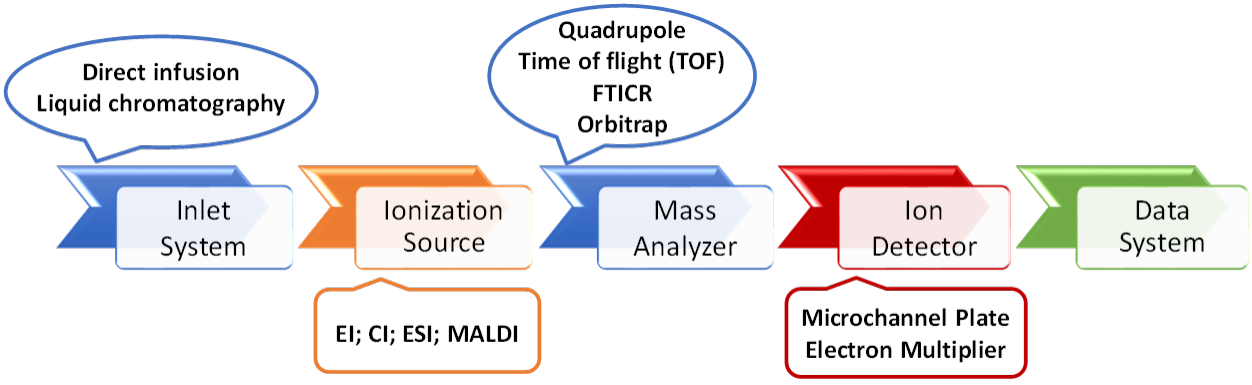


Figure 1. Five major components of a mass spectrometer.

After being introduced into the mass spectrometer by the inlet system, analytes are transferred into gas-phase ions which is facilitated by ionization. There are many different ionization techniques ranging from hard ionization such as electron ionization (EI),[11, 12] and chemical ionization[13, 14] to soft ionization such as electrospray ionization (ESI),[15, 16] matrix-assisted laser desorption ionization (MALDI),[17-19] and desorption electrospray ionization (DESI).[20] Among them, ESI and MALDI are the two most widely used ionization techniques as they allow the analysis of molecules in their intact forms. ESI has rapidly become one of the most popular ionization techniques since its introduction[16] due to its high sensitivity and the capability to couple with liquid chromatography (LC) for online LC-MS analysis.[21, 22] In ESI, the analytes are transferred into the gas-phase by applying a high voltage to a small tip. The strong electric field results in charge accumulation at the needle tip, creating a Taylor cone of solution which subsequently breaks apart and releases highly charged droplets. A combination of electrostatic repulsion and solvent evaporation desolvates these droplets as they eventually form gas-phase ions.[16] There are multiple mechanisms explaining the ESI process including charge residue model (CRM),[23, 24] ion evaporation model (IEM),[25, 26] and chain ejection model (CEM).[27] Ionization of analytes by MALDI involves the employment of laser light to bombard and ablate the sample mixed with excess amount of matrix on a solid plate. The matrix absorbs the laser energy and transfers it to the sample during the heating of the sample-matrix mixture, allowing its ionization without excessive fragmentation.[28] Both ESI and MALDI were awarded the Nobel prize in 2002 for their contribution to the mass spectrometry characterization of biological molecules.[29] Thanks to their ballooned and continuous development since the late

1980s, ESI and MALDI have been widely utilized to ionize peptides,[18, 30] proteins,[31, 32] and large protein complexes,[33, 34] enabling both ionization techniques to become an essential part of the modern proteomics field.[21, 22, 35-38]

Once the analyte ions are formed in the gas-phase, they are transmitted to a mass analyzer through ion funnels[39, 40] and quadrupoles.[41] This process transfers ions from low vacuum to high vacuum which causes ions to scatter in the gas-phase, hence loss of signal eventually. Ion funnels which consist of a series of electrodes with decreasing diameter are designed to focus ions during transmission to overcome this issue.[42] This greatly reduces the impact of ion scattering and increases signal to noise ratio, thus eventually increases the sensitivity of mass spectrometers.

Mass analyzer is another very important component of a mass spectrometer as it detects and separates ions based on their *m/z*. There are many different types of mass analyzers today including quadrupole,[43-45] Time of flight (TOF),[46-49] Orbitrap,[50-52] and Fourier transform ion cyclotron resonance (FT-ICR).[53-55] Quadrupoles are one of the most widely used mass analyzers.[45] They are comprised of four cylindrical rod electrodes aligned in parallel to each other, forming a channel through which ions can pass through and be separated based on their *m/z* when an electric field is applied to the rods.[44] Quadrupoles are also usually used as mass selectors or mass filters. When they function like this, only ions with a narrow range of *m/z* are allowed to pass through, enabling quadrupoles to be easily coupled with other mass analyzers for tandem MS experiments (discussed later).[56, 57] This is done by applying direct current (DC) and radio frequency (RF) potentials to the rods simultaneously to direct and filter

ions.[58] The basic principle of a TOF analyzer involves passing an ion packet through a field free region to hit an electron multiplier tube.[47] The ion trajectory in this process is entirely dependent on the kinetic energy gained by the initial ion acceleration and not impacted by the electric field. Therefore, the time it takes for each ion to reach the detector is correlated to the *m/z* of the ion and is measured to eventually determine the mass of the ion.[46, 49] TOF analyzers has multiple facets of benefits including fast speed and high sensitivity, making them one of the most used mass analyzers in modern mass spectrometry technology.[57, 59-63] Orbitrap and FTICR are both a branch of Fourier Transform MS (FTMS).[64] In an Orbitrap analyzer, ions circulate in a spiral manner around the central electrode and at the same time, oscillate along the axial direction. This ion trajectory is caused by a combination of radial and axial electric fields applied between central and outer electrodes. The axial oscillations of ions are measured and detected by outer electrodes as image current.[50] In FTICR, ions are excited into an orbit and the cyclotron frequency with which they oscillate are measured based on the ions' *m/z* and recorded as image current.[55] In both FTMS technologies, a Fourier transform[65] is performed to convert the resulting digitized image current in the time domain to frequency domain and eventually a mass spectrum.[64] FTMS has the ultimate advantage of achieving high resolution and accuracy,[66] thus they have been widely used in the modern mass spectrometry analysis.[67, 68]

1.2. Tandem mass spectrometry and fragmentation techniques

MS can offer molecular weight information of gaseous ions based on their *m/z*. However, MS alone does not yield comprehensive structural and conformational information of the analyte. Tandem MS or MS/MS capabilities that involve isolation and subsequent fragmentation of

precursor ions are then introduced to unveil these important molecular characteristics. Protein characterization by tandem mass spectrometry can be done in three different ways, i.e., bottom-up, top-down, and middle-down mass spectrometry (Figure 2). Bottom-up MS involves the reduction and digestion of all proteins in a sample such as cell lysates, followed by online liquid chromatography tandem MS (LC-MS/MS).[69] Alternatively, top-down MS is performed by directly measuring intact gas-phase protein ions without reduction or digestion, which are subsequently dissociated by different fragmentation methods to generate sequence-informative fragment ions.[70] The intermediate approach is middle-down MS in which proteins of interest undergo limited reduction or partial proteolysis resulting in large peptides or subunits, which are analyzed by LC-MS/MS afterwards.[71]

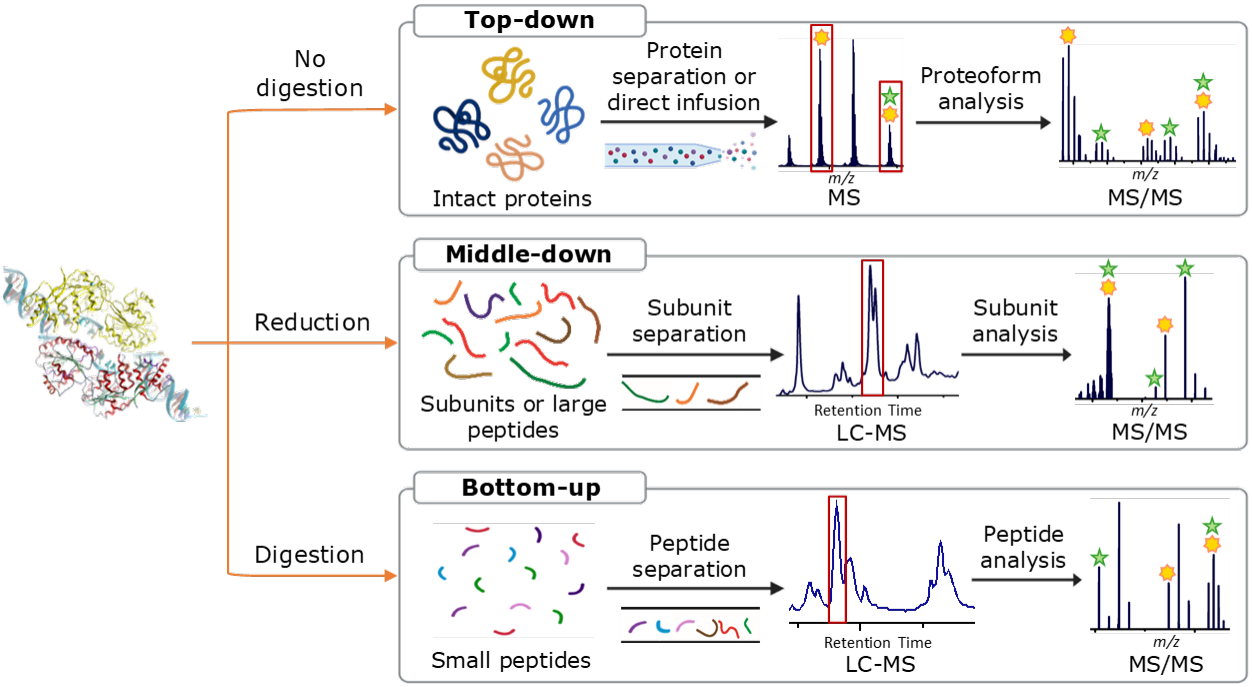


Figure 2. Concepts and workflows of top-down, middle-down, and bottom-up mass spectrometry for protein characterization.

Tandem mass spectrometry or MS/MS has been successfully used to characterize the primary sequence and higher-order structure of peptides,[38, 72] proteins,[73-75] and even protein complexes.[76-79] In MS/MS, either isolated precursor ions or a broadband of ions with a large *m/z* window are dissociated into fragment ions, which can be mapped back to the protein sequence.[80] These fragments can either be *a*, *b*, or *c* fragment ions which contain the N-terminus, or they can be *x*, *y*, or *z* fragment ions which contain the C-terminus (Figure 3).[81] In addition, internal fragments that contain neither terminus can also be generated (structures shown in Figure 3 and discussed later). Throughout the years, many ion activation methods have been developed including collision-, electron-, and photon-based fragmentation techniques to unveil protein sequence and structure information, with each having discrete advantages (Figure 4).[82, 83]

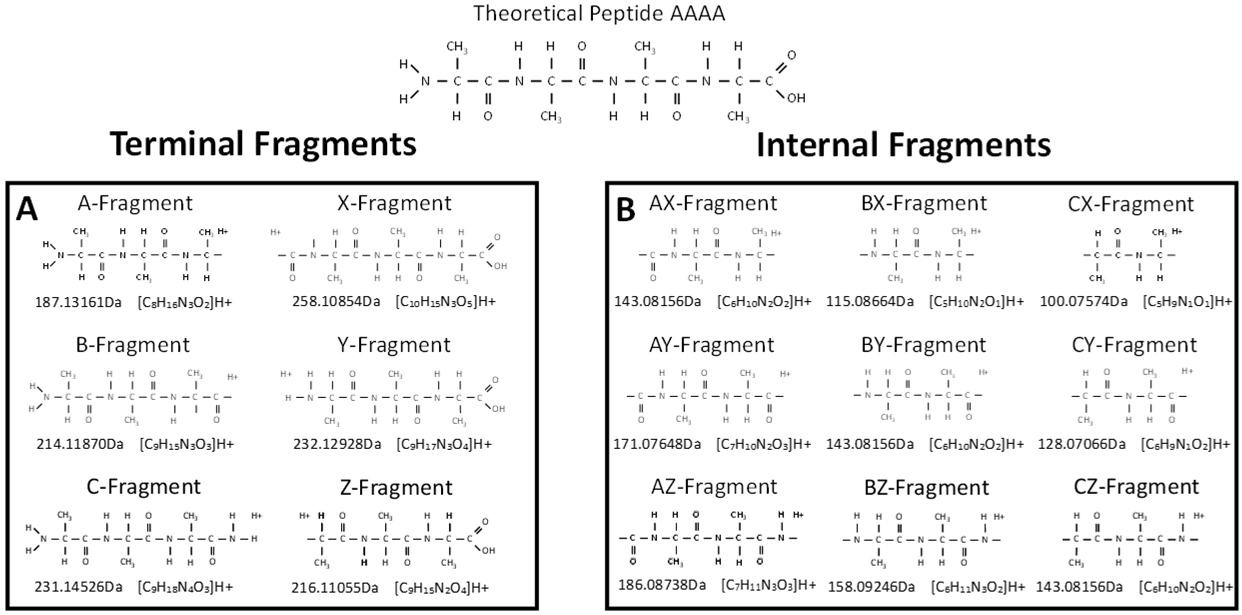


Figure 3. Theoretical masses and structures of all types of (A) terminal fragments and (B) internal fragments.

Reprinted with permission from Ref.[84]. Copyright 2021 American Chemical Society.

1.2.1. Collision-based fragmentation

Collision activated dissociation (CAD) is the first introduced collision-based fragmentation technique and arguably the most widely used.[85, 86] Later, High energy C-trap dissociation (HCD), another collision-based technique that occurs with higher energy on a faster time scale, was introduced with the introduction of linear ion trap–orbitrap mass spectrometers (LTQ Orbitraps).[87] Over the years, collision-based fragmentation techniques especially CAD has been the "gold standard" thanks to its high efficiency of energy accumulation and fragmentation.[83, 88] In CAD and HCD, energetic collisions between the analyte ions and inert gas molecules such as nitrogen, helium or argon transfer part of the kinetic energy of the ion into its internal energy, of which the accumulation ultimately leads to fragmentation of the analyte ions.[85] Energy deposits in a step-wise manner in CAD and HCD, resulting in a ceiling of the total energy that can be accumulated in this multi-collision process. Therefore, the weakest bond on the protein backbone, i.e., the peptide bond is usually cleaved in CAD and HCD, generating *b*- and *y*-type terminal fragment ions. Although overwhelmingly popular, collision-based fragmentation still has some weaknesses. For example, CAD suffers from low mass cutoff (LMCO) in ion trap mass spectrometers, meaning that the truncation of the low *m/z* region may occur.[89] In addition, CAD and HCD are considered as "harsh" fragmentation techniques that usually result in the loss of important structural information such as post-translational modifications (PTMs), non-covalent protein-protein or protein-ligand interactions. To combat these problems, alternative fragmentation techniques including electron- and photon-based fragmentation have been developed.

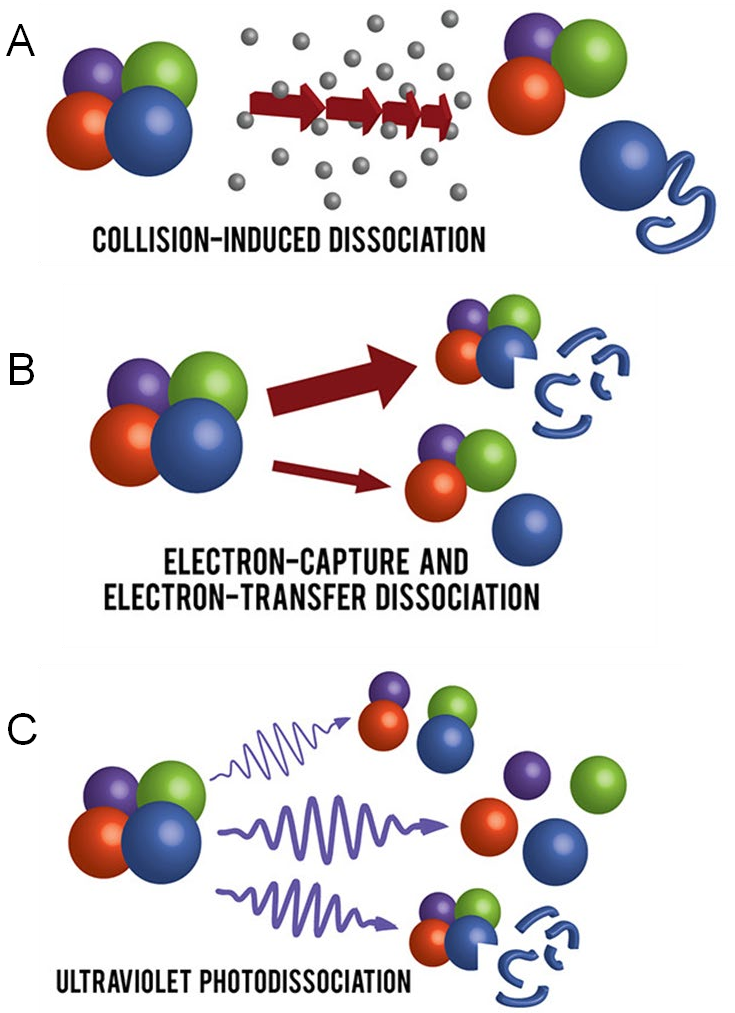


Figure 4. Cartoon illustration representing the major products of various common dissociation methods in the study of protein complexes.
Small blue fragments correspond to covalent cleavage of an individual protein chain. Adapted with permission from Ref[90]. Copyright 2019 American Chemical Society.

1.2.2. Electron-based fragmentation

Electron-based fragmentation including electron capture dissociation (ECD)[91, 92] and electron transfer dissociation (ETD)[93] are generally perceived as "softer" dissociation techniques.[94] In ECD and ETD, reactions between multicharged positive analyte ions and either low-energy electrons (< 2eV, ECD) or negatively charged radical ions (ETD) lead to the absorption of

electrons by the analyte ions, resulting in charge reduction and fragmentation.[95, 96] During this electron capture or electron transfer process, odd-electron radical species are formed with minimal redistribution of vibrational energy. The energy absorbed by the charge-reduced protein ions through this exoergic process leads to cleavages of the N-C_α bonds rather than the weaker peptide bonds, producing *c*- and *z*- type terminal fragments.[97] Compared to collision-based fragmentation, ECD and ETD hold the advantage of being able to preserve relatively labile non-covalent interactions during fragmentation, enabling the acquisition of protein structure information.[98, 99] In addition, ECD and ETD cleave different protein backbone bonds than CAD and HCD, thus providing complementary primary sequence information.[100] However, the fragmentation efficiency of ECD and ETD is relatively low which can limit the coverage of protein sequence. In recent years, a new electron-based fragmentation technique, electron induced dissociation (EID) that employs the use of high energy electrons (> 20 eV),[101-103] has been developed to overcome these limitations. In EID, charge oxidation instead of charge reduction occurs as a result of the interaction between high energy electrons and positive analyte ions. Following this, oxidized species undergo rearrangement and/or capture a second electron, initiating protein backbone fragmentation.[101] EID has been demonstrated to improve the fragmentation efficiency of peptides and proteins to 100%.[102] In addition, our lab have successfully applied EID to cause extensive fragmentation of protein complexes and generate internal fragments ions.[104, 105]

1.2.3. Photon-based fragmentation

Photon-based fragmentation techniques have a rich history and can be categorized into infrared

multiple-photon dissociation (IRMPD) and ultraviolet photodissociation (UVPD), depending on the wavelength of the photon sources.[106-108] In IRMPD, protein ions absorb one or more low frequency photons (e.g., 10.6 μm) resulting in the excitation of vibrational modes, leading to a slow-heating fragmentation process similar to CAD and HCD.[109] Therefore, IRMPD generates the same *b*- and *y*-type terminal fragments as collision-based fragmentation. UVPD employs photons in the ultraviolet range including 157nm,[110] 193nm,[111, 112] 213nm,[113] and 266nm[114] to cleave the protein backbone, allowing the multicharged positive ions to access excited electronic states. This fast-heating process enables UVPD to cleave the C-C_α bond in addition to peptide bonds and the N-C_α bond, yielding all six types of terminal fragments (*a*, *b*, *c* N-terminal fragments and *x*, *y*, *z* C-terminal fragments). Photon-based dissociation has been widely applied to characterize peptides and proteins,[115-119] as well as revealing protein structural information by localizing covalent and non-covalent interactions on proteins and protein complexes.[78, 79, 120, 121]

2. Internal fragments of proteins

Protein fragment ions can either be i) a terminal fragment resulting from a single backbone bond cleavage event of the precursor ion to generate amino-terminal-containing fragments (*a*, *b,* or *c* fragments) or carboxy-terminal-containing fragments (*x*, *y*, or *z* fragment), or ii) an internal fragment resulting from multiple cleavage events of the precursor ion to generated *ax*, *ay*, *az*, *bx*, *by*, *bz*, *cx*, *cy*, and *cz* fragment ions that contain neither terminus (with the first letter designating cleavage on the N-terminal side and the second letter designating cleavage on the C-terminal side, Figure 3 and 5).[72, 105] Although the concept of internal fragments has been around for

decades,[122-126] there is still some debate about the nomenclature of internal fragments. The nomenclature used here is not a consensus and our lab is actively engaging in discussions with other academic and industrial labs to establish a nomenclature system of internal fragments that will be widely accepted by the mass spectrometry community.

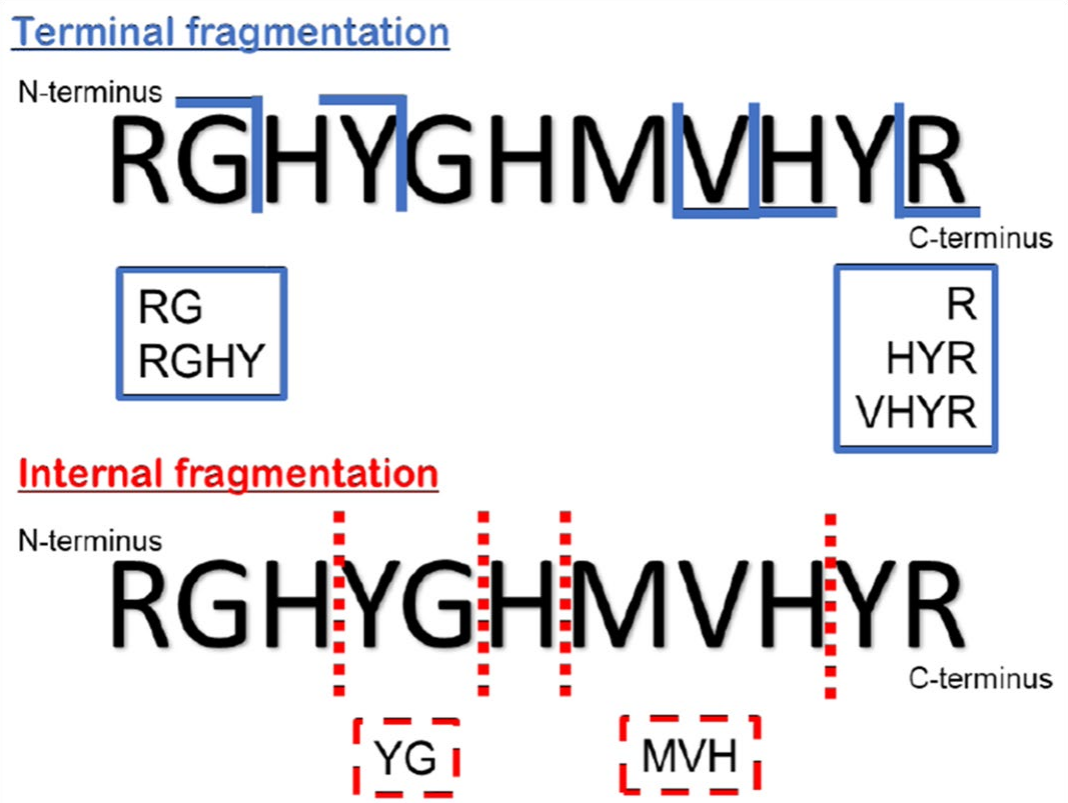


Figure 5. The fragmentation pathways of an example peptide showing that terminal fragments contain either N- or C-terminus while internal fragments contain neither termini.

2.1. Research on internal fragments in the early years

To the best of my knowledge, internal fragments were first reported in 1973 from a study investigating fragments generated from EI-MS of Schiff bases of peptide esters.[127] Since then, the great Klaus Biemann has mentioned and discussed the concept of internal fragments throughout his research career of studying peptide and protein sequencing. It started in the

early 1980s, successive Edman degradations[128] of internal fragments were used to help determine the amino acid sequence of rabbit muscle creatine phosphokinase.[129] Internal fragments were then mentioned in a study exploring novel side chain fragmentation of peptides in CAD[130], and were subsequently shown to be present in fast atom bombardment (FAB) MS in the late 1980s.[131] The notation system of internal fragments was first discussed by Dr. Biemann in 1988[122] and later again in 1992.[123] He and coworkers prefer to label internal fragments as a string of single amino acid letters to represent the sequence of the fragments (e.g., PT, PTL, VGE).[122] We now choose to adopt a more specific notation system because the mass of a protein fragment sequence can differ significantly when different fragmentation methods are used. For example, the mass of the same PTL fragment generated by CAD and ECD can have differing masses. Internal fragments started to have more applications subsequently, with internal acyl and immonium ions included in the score of a computer algorithm called SEQPEP to aid the interpretation of high-energy collision tandem mass spectra of peptides,[132] and used to help distinguish similar candidate peptide sequences.[133] Coming to 1990s, the assignment of internal fragments has been demonstrated to help differentiate hydroxyproline isomers in CAD of synthetic peptides,[134] determine the sequence of *Sarcophaga bullata* protease inhibitor,[135] and assign the correct sequence of horse and sperm whale myoglobins.[136] One study specifically dedicated to the production of internal fragments in a hybrid sector/quadrupole mass spectrometer investigated multiple possible formation pathways of internal fragments by CAD in sequential MS experiments.[137]

2.2. Why research on internal fragments is rare?

Research on internal fragments has never ballooned despite the fact that they are shown to be helpful on many occasions. One major reason that internal fragments have been largely ignored by the mass spectrometry community is that their formation mechanism is poorly understood, resulting in a lack of available software to accurately and reliably assign them.[84, 138] In a single MS/MS experiment, the number of theoretical internal fragments is significantly greater than the number of possible terminal fragments that can be generated. In addition, the disparity of this number scales exponentially with the increase of protein size, resulting in a substantial computational demand and an ambiguity of assigning internal fragments.[105, 139, 140] Therefore, efforts have been made by multiple groups to address this issue.[141]

Agar and coworkers classified the ambiguity of assigning internal fragments into three subcategories: arrangement ambiguity, frameshift ambiguity, and mass accuracy ambiguity.[142] A study by our lab which is also presented in Chapter 2 of this dissertation, used a statistical approach to investigate the formation mechanism of internal fragments by CAD to resolve the arrangement ambiguity and frameshift ambiguity, thus increasing the confidence of assigning internal fragments.[72] This study has provided insights into the development by our group of ClipsMS, a python-based program for internal fragment analysis.[84] Mann and coworkers conducted a systematic investigation on the nature of tryptic HCD spectra and found that proline is often the first amino acid of an internal fragment and there is a slight preference for aspartic acid, glutamic acid, glutamine, tryptophan and histidine on the C-terminal side of an internal fragment.[143] Nevertheless, studies on fundamentals of internal fragments, especially by fragmentation methods other than CAD or HCD, are still scarce. For electron- and photon-based

techniques which have more complex fragmentation mechanisms thus more diverse and accessible cleavage pathways and fragment types, more research is needed to better understand internal fragments formed in these methods to aid their assignments and reduce the false-discovery rates.

2.3. Applications of internal fragments in the new millennium

Although not extensive, research efforts have been made to explore the utility of internal fragments in various applications in the new millennium. ECD followed by sustained off-resonance irradiation collision activated dissociation (SORI-CAD) in an FTICR instrument generated internal fragments diagnostic of disulfide bridges of naturally occurring small peptide systems.[144] Internal fragments in HCD have also been used to differentiate between similar candidate peptides, thus improving de novo peptide sequencing.[145] Several studies have highlighted a significant increase of the total percentage of matched fragments and protein sequence coverage by analyzing internal fragments of intact proteins.[105, 139, 146-148] The fragmentation techniques used in these studies span from CAD[146, 147] and HCD[139] to ECD[105, 148] and EID[105, 148] on various types of instruments including QTOF,[147] FTICR,[105, 146, 148] and Orbitrap,[139] highlighting that the value of analyzing internal fragments of intact proteins is not affected by fragmentation techniques and instrumentation platforms used. However, UVPD has been shown to produce a large population of fragments that result in an increased risk of false-positive identifications of internal fragments.[149, 150] This highlights the need of developing strategies to increase the confidence of assigning internal fragments by UVPD. The identification of various types of PTMs can be helped by the analysis of internal fragments. For

example, for disulfide bonds which is one of the most important PTMs that ensures proper protein folding,[151, 152] Chin and coworkers applied internal fragments to decipher disulfide bonds in disulfide-rich peptides.[153] A recent study by Schmitt and coworkers demonstrated the utility of internal fragments in identifying sequence motifs within a disulfide-constrained loop of the SOD1 protein that could not be achieved solely by terminal fragments.[142] Our lab have shown that internal fragments can be used to determine the disulfide connectivity of disulfide-intact proteins, which is also presented in Chapter 3 of this dissertation.[154] Furthermore, modifications and variants on therapeutic biologics including natural glycosylations, liabilities like deamidation, and single sequence variants can be identified by assigning internal fragments in top-down and middle-down mass spectrometry.[155, 156] In addition to being applied on small peptides and intact proteins, internal fragments have been demonstrated useful in the characterization of protein complexes to improve sequence coverage[78, 104] and localize a methylation site.[78] Proteome-level utilization of internal fragments have also been evaluated and proved to be directly or indirectly beneficial in improving the identification of proteoforms.[138, 157] Recently, internal fragments have been integrated into and shown to contribute to more comprehensive characterization of top-down sequencing of oligonucleotides.[158]

In summary, the assignment of internal fragments has been demonstrated valuable in various types of applications. However, more research efforts, both on the fundamental and application side, are needed to push forward the incorporation of internal fragment analysis into the mainstream mass spectrometry techniques. An important aspect of my dissertation research focuses on exploring the fundamentals and applications of internal fragments in top-down and

middle-down mass spectrometry. I hope to make a contribution to the advancement of internal fragment research through my dissertation research.

3. Mass spectrometry in biotherapeutics analysis

The first monoclonal antibody (mAb) therapeutic was approved for the treatment of T-cell-mediated rejection of renal allografts in 1985.[159] Almost 40 years later, mAb-based therapeutics have gained growing significance and there are nearly 200 mAb-based drugs[160] on the market right now to treat a host of human diseases such as cancer, metabolic disorders, and viral infections.[161-164] Mabs possess distinct pharmacological properties including cognate target specificity and long circulating half-life, resulting in various potential mechanisms of action.[165-168] In recent years, various new mAb formats have been developed including nanobodies, fusion proteins, multispecifics, and antibody-drug conjugates (ADCs).[169-172] Among them, ADCs have emerged as one of the most promising classes of protein therapeutics.[173, 174] ADCs are designed to integrate the high target specificity of mAbs and the potent cytotoxicity of small molecule payloads to improve their antitumor efficacy.[175] Currently, there are 13 ADC products approved by US, EU, or China for use and over 80 ADC molecules are undergoing clinical trials at various stages.[165, 176] Both mAbs and ADCs are highly heterogeneous molecules with large size (~150 kDa), a series of intra- and inter-chain disulfide bonds and numerous PTMs (Figure 6).[177, 178] In the case of ADCs, even more molecular complexity is introduced due to the conjugation of various numbers of payloads onto either a set of fixed sites or a large array of locations on the antibody, depending on the linker chemistry utilized.[179-182] Throughout the drug development process, full characterization by robust analytical techniques of the critical quality attributes

(CQAs)[183] including the primary sequence, PTMs and potential liabilities, higher order structure, drug conjugation profiles, and any related impurities of mAbs and ADCs is required to produce high-quality drug products.[184]

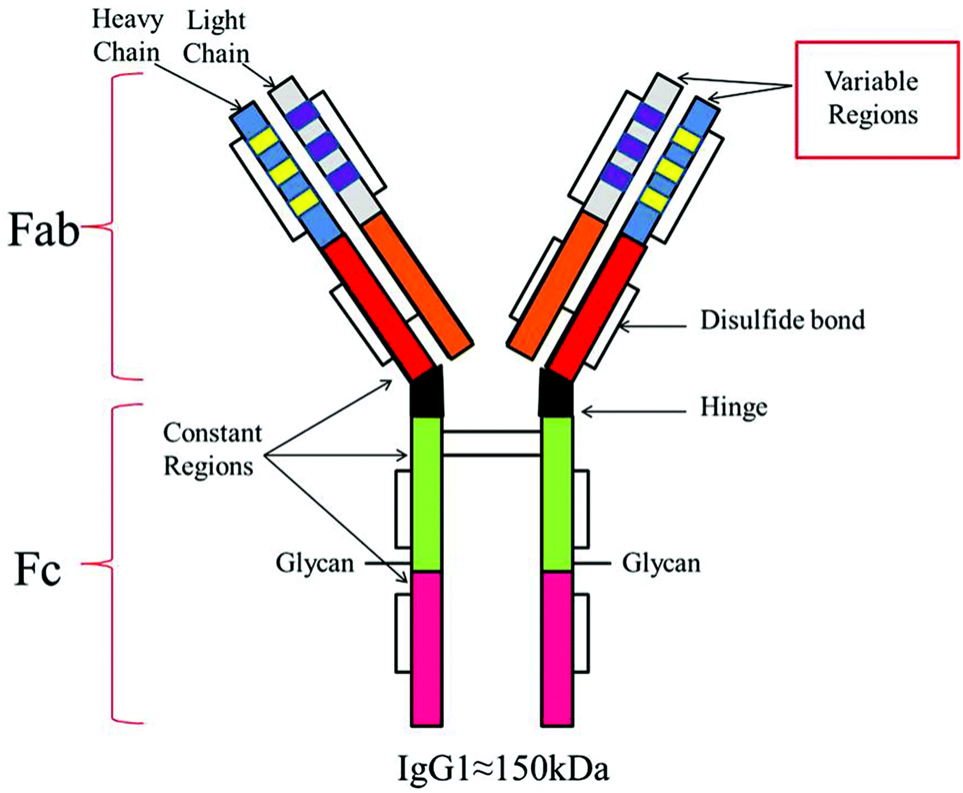


Figure 6. General structure of an Immunoglobulin G1. Reprinted with permission from Ref[185]. Copyright 2015 Royal Society of Chemistry.

Extensive research efforts have been conducted to develop various analytical tools to characterize mAb-based therapeutic molecules including capillary electrophoresis,[186] differential scanning calorimetry,[187] differential scanning fluorimetry,[188] imaged capillary isoelectric focusing,[189] Fourier transform infrared and Raman spectroscopy,[190] and various chromatography methods such as size exclusion chromatography (SEC),[191] hydrophobic interaction chromatography (HIC),[192] and ion exchange chromatography (IEX).[193] Mass spectrometry based

techniques, especially reversed phase liquid chromatography (rpLC) coupled with MS (rpLC-MS), have played a critical role for the characterization of mAb based therapeutics throughout all phases of drug development.[194-197] Denaturing rpLC-MS provides accurate molecular weight information of the biotherapeutic molecule to make sure the correct molecule has been cloned and expressed. In addition, bottom-up rpLC-MS peptide mapping methods offer high sequence coverage with amino acid resolution and can identify drug conjugation sites of ADCs with high confidence.[198, 199] Native MS methods have also been used to obtain molecular weight and stoichiometric information of noncovalently assembled therapeutic molecules, as native MS can retain the “native-like” structure of proteins in the gas-phase.[200, 201] Native MS can also be hyphenated with various native chromatography methods such as SEC, HIC, and IEX to enable size and charge variant analysis of mAb based therapeutics.[202-205]

Top-down (TD) and middle-down (MD) MS have gained in popularity in recent years for the characterization of mAb based therapeutics.[206-212] Both TD- and MD-MS have their unique advantages compared to bottom-up peptide mapping methods including minimal sample preparation and being able to avoid the introduction of artifactual modifications such as deamidation and oxidation.[213, 214] However, they suffer from limited fragmentation hence not ideal sequence, PTM, and drug conjugation coverage due to the high molecular complexity of the biotherapeutic targets. Furthermore, unlike the well-established bottom-up MS method, there is currently no universal workflow for the TD- and MD-MS data analysis process. Therefore, the application of TD- and MD-MS for the characterization of mAb based therapeutics have been limited to individual academia labs and they are not routinely used in

biopharma industry. Chapters 4 and 5 of this dissertation seek to explore the utility of internal fragments to improve the fragmentation efficiency of TD- and MD-MS of mAb based therapeutic molecules, in the hope of advancing these MS techniques to become a routine participant in biopharma industry in the future.

References

1. Kind, T.; Fiehn, O., Advances in structure elucidation of small molecules using mass spectrometry. *Bioanalytical Reviews* **2010,** *2* (1), 23-60.

2. Cajka, T.; Fiehn, O., Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *TrAC Trends in Analytical Chemistry* **2014,** *61*, 192-206.

3. Zaia, J., Mass spectrometry of oligosaccharides. *Mass Spectrometry Reviews* **2004,** *23* (3), 161-227.

4. Beck, J. L.; Colgrave, M. L.; Ralph, S. F.; Sheil, M. M., Electrospray ionization mass spectrometry of oligonucleotide complexes with drugs, metals, and proteins. *Mass Spectrometry Reviews* **2001,** *20* (2), 61-87.

5. Jackson, C. A.; Simonsick, W. J., Application of mass spectrometry to the characterization of polymers. *Current Opinion in Solid State and Materials Science* **1997,** *2* (6), 661-667.

6. Hanton, S. D., Mass Spectrometry of Polymers and Polymer Surfaces. *Chemical Reviews* **2001,** *101* (2), 527-570.

7. Domon, B.; Aebersold, R., Mass Spectrometry and Protein Analysis. *Science* **2006,** *312* (5771), 212-217.

8. Heck, A. J. R., Native mass spectrometry: a bridge between interactomics and structural biology. *Nature Methods* **2008,** *5* (11), 927-933.

9. Zhou, M.; Lantz, C.; Brown, K. A.; Ge, Y.; Paša-Tolić, L.; Loo, J. A.; Lermyte, F., Higher-order structural characterisation of native proteins and complexes by top-down mass spectrometry. *Chemical Science* **2020,** *11* (48), 12918-12936.

10. De Hoffmann, E.; Stroobant, V., *Mass spectrometry: principles and applications*. John Wiley & Sons: 2007.

11. Bleakney, W., A New Method of Positive Ray Analysis and Its Application to the Measurement of Ionization Potentials in Mercury Vapor. *Physical Review* **1929,** *34* (1), 157-160.

12. Nier, A. O., A Mass Spectrometer for Isotope and Gas Analysis. *Review of Scientific Instruments* **2004,** *18* (6), 398-411.

13. Munson, M. S.; Field, F.-H., Chemical ionization mass spectrometry. I. General introduction. *Journal of the American Chemical Society* **1966,** *88* (12), 2621-2630.

14. Harrison, A. G., *Chemical ionization mass spectrometry*. Routledge: 2018.

15. Yamashita, M.; Fenn, J. B., Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry* **1984,** *88* (20), 4451-4459.

16. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989,** *246* (4926), 64-71.

17. Karas, M.; Bachmann, D.; Hillenkamp, F., Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical chemistry* **1985,** *57* (14), 2935-2939.

18. Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F., Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes* **1987,** *78*, 53-68.

19. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry* **1988,** *60* (20), 2299-2301.

20. Takáts, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G., Mass Spectrometry Sampling Under Ambient Conditions with Desorption Electrospray Ionization. *Science* **2004,** *306* (5695), 471-473.

21. Aebersold, R.; Goodlett, D. R., Mass Spectrometry in Proteomics. *Chemical Reviews* **2001,** *101* (2), 269-296.

22. Han, X.; Aslanian, A.; Yates, J. R., Mass spectrometry for proteomics. *Current Opinion in Chemical Biology* **2008,** *12* (5), 483-490.

23. Winger, B. E.; Light-Wahl, K. J.; Ogorzalek Loo, R. R.; Udseth, H. R.; Smith, R. D., Observation and implications of high mass-to-charge ratio ions from electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* **1993,** *4* (7), 536-545.

24. Fernandez de la Mora, J., Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta* **2000,** *406* (1), 93-104.

25. Thomson, B.; Iribarne, J., Field induced ion evaporation from liquid surfaces at atmospheric pressure. *The Journal of Chemical Physics* **1979,** *71* (11), 4451-4463.

26. Hogan, C. J., Jr.; Carroll, J. A.; Rohrs, H. W.; Biswas, P.; Gross, M. L., Combined Charged Residue-Field Emission Model of Macromolecular Electrospray Ionization. *Analytical Chemistry* **2009,** *81* (1), 369-377.

27. Konermann, L.; Rodriguez, A. D.; Liu, J., On the Formation of Highly Charged Gaseous Ions from Unfolded Proteins by Electrospray Ionization. *Analytical Chemistry* **2012,** *84* (15), 6798-6804.

28. Knochenmuss, R., A Quantitative Model of Ultraviolet Matrix-Assisted Laser Desorption/Ionization Including Analyte Ion Generation. *Analytical Chemistry* **2003,** *75* (10), 2199-2207.

29. Grayson, M. A., John Bennett Fenn: A Curious Road to the Prize. *Journal of the American Society for Mass Spectrometry* **2011,** *22* (8), 1301-1308.

30. Emmett, M. R.; Caprioli, R. M., Micro-electrospray mass spectrometry: Ultra-high-sensitivity analysis of peptides and proteins. *Journal of the American Society for Mass Spectrometry* **1994,** *5* (7), 605-613.

31. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T., Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **1988,** *2* (8), 151-153.

32. Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R., New developments in biochemical mass spectrometry: electrospray ionization. *Analytical chemistry* **1990,** *62* (9), 882-899.

33. Cohen, L. R. H.; Strupat, K.; Hillenkamp, F., Analysis of quaternary protein ensembles by matrix assisted laser desorption/ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* **1997,** *8* (10), 1046-1052.

34. Loo, J. A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrometry Reviews* **1997,** *16* (1), 1-23.

35. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry* **2007,** *389* (4), 1017-1031.

36. Gygi, S. P.; Aebersold, R., Mass spectrometry and proteomics. *Current Opinion in Chemical Biology* **2000,** *4* (5), 489-494.

37. Cravatt, B. F.; Simon, G. M.; Yates Iii, J. R., The biological impact of mass-spectrometry-based proteomics. *Nature* **2007,** *450* (7172), 991-1000.

38. Ong, S.-E.; Mann, M., Mass spectrometry–based proteomics turns quantitative. *Nature Chemical Biology* **2005,** *1* (5), 252-262.

39. Shaffer, S. A.; Prior, D. C.; Anderson, G. A.; Udseth, H. R.; Smith, R. D., An Ion Funnel Interface for Improved Ion Focusing and Sensitivity Using Electrospray Ionization Mass Spectrometry. *Analytical Chemistry* **1998,** *70* (19), 4111-4119.

40. Kelly, R. T.; Tolmachev, A. V.; Page, J. S.; Tang, K.; Smith, R. D., The ion funnel: Theory, implementations, and applications. *Mass Spectrometry Reviews* **2010,** *29* (2), 294-312.

41. March, R. E., An Introduction to Quadrupole Ion Trap Mass Spectrometry. *Journal of Mass Spectrometry* **1997,** *32* (4), 351-369.

42. Julian, R. R.; Mabbett, S. R.; Jarrold, M. F., Ion funnels for the masses: Experiments and simulations with a simplified ion funnel. *Journal of the American Society for Mass Spectrometry* **2005,** *16* (10), 1708-1712.

43. Paul, W.; Steinwedel, H., Ein neues massenspektrometer ohne magnetfeld. *Zeitschrift f□ ⊾ r Naturforschung A* **1953,** *8* (7), 448-450.

44. Wolfgang, P.; Helmut, S., Apparatus for separating charged particles of different specific charges. Google Patents: 1960.

45. Dawson, P. H., *Quadrupole mass spectrometry and its applications*. Elsevier: 2013.

46. Guilhaus, M., Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. *Journal of Mass Spectrometry* **1995,** *30* (11), 1519-1532.

47. Wolff, M. M.; Stephens, W. E., A Pulsed Mass Spectrometer with Time Dispersion. *Review of Scientific Instruments* **2004,** *24* (8), 616-617.

48. Mamyrin, B. A., Time-of-flight mass spectrometry (concepts, achievements, and prospects). *International Journal of Mass Spectrometry* **2001,** *206* (3), 251-266.

49. Boesl, U., Time-of-flight mass spectrometry: Introduction to the basics. *Mass Spectrometry Reviews* **2017,** *36* (1), 86-109.

50. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R., The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry* **2005,** *40* (4), 430-443.

51. Zubarev, R. A.; Makarov, A., Orbitrap Mass Spectrometry. *Analytical Chemistry* **2013,** *85* (11), 5288-5296.

52. Scigelova, M.; Makarov, A., Orbitrap Mass Analyzer – Overview and Applications in Proteomics. *PROTEOMICS* **2006,** *6* (S2), 16-21.

53. Comisarow, M. B.; Marshall, A. G., Fourier transform ion cyclotron resonance spectroscopy. *Chemical Physics Letters* **1974,** *25* (2), 282-283.

54. Comisarow, M. B.; Marshall, A. G., Resolution-enhanced Fourier transform ion cyclotron resonance spectroscopy. *The Journal of Chemical Physics* **1975,** *62* (1), 293-295.

55. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews* **1998,** *17* (1), 1-35.

56. Yost, R. A.; Boyd, R. K., [7] Tandem mass spectrometry: Quadrupole and hybrid instruments. In *Methods in Enzymology*, Academic Press: 1990; Vol. 193, pp 154-200.

57. Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A., An introduction to quadrupole–time-of-flight mass spectrometry. *Journal of Mass Spectrometry* **2001,** *36* (8), 849-865.

58. Miller, P. E.; Denton, M. B., The quadrupole mass filter: basic operating concepts. *Journal of chemical education* **1986,** *63* (7), 617.

59. Wang, L.; Amphlett, G.; Lambert, J. M.; Bl¡¤ttler, W.; Zhang, W., Structural Characterization of a Recombinant Monoclonal Antibody by Electrospray Time-of-Flight Mass Spectrometry. *Pharmaceutical Research* **2005,** *22* (8), 1338-1349.

60. Cotter, R. J., Time-of-flight mass spectrometry for the structural analysis of biological molecules. *Analytical chemistry* **1992,** *64* (21), 1027A-1039A.

61. Nielen, M. W. F., Maldi time-of-flight mass spectrometry of synthetic polymers. *Mass Spectrometry Reviews* **1999,** *18* (5), 309-344.

62. Weickhardt, C.; Moritz, F.; Grotemeyer, J. r., Time-of-flight mass spectrometry: State-of the-art in chemical analysis and molecular science. *Mass Spectrometry Reviews* **1996,** *15* (3), 139-162.

63. Williams, J. P.; Morrison, L. J.; Brown, J. M.; Beckman, J. S.; Voinov, V. G.; Lermyte, F., Top-Down Characterization of Denatured Proteins and Native Protein Complexes Using Electron Capture Dissociation Implemented within a Modified Ion Mobility-Mass Spectrometer. *Anal Chem* **2020,** *92* (5), 3674-3681.

64. Scigelova, M.; Hornshaw, M.; Giannakopulos, A.; Makarov, A., Fourier Transform Mass Spectrometry. *Molecular & Cellular Proteomics* **2011,** *10* (7).

65. Cochran, W. T.; Cooley, J. W.; Favin, D. L.; Helms, H. D.; Kaenel, R. A.; Lang, W. W.; Maling, G. C.; Nelson, D. E.; Rader, C. M.; Welch, P. D., What is the fast Fourier transform? *Proceedings of the IEEE* **1967,** *55* (10), 1664-1674.

66. Marshall, A. G.; Hendrickson, C. L., High-Resolution Mass Spectrometers. *Annual Review of Analytical Chemistry* **2008,** *1* (1), 579-599.

67. Tucholski, T.; Ge, Y., Fourier-transform ion cyclotron resonance mass spectrometry for characterizing proteoforms. *Mass Spectrometry Reviews* **2022,** *41* (2), 158-177.

68. Lange, O.; Damoc, E.; Wieghaus, A.; Makarov, A., Enhanced Fourier transform for Orbitrap mass spectrometry. *International Journal of Mass Spectrometry* **2014,** *369*, 16-22.

69. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422* (6928), 198-207.

70. Cui, W.; Rohrs, H. W.; Gross, M. L., Top-down mass spectrometry: recent developments, applications and perspectives. *Analyst* **2011,** *136* (19), 3854-3864.

71. Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C., High-Throughput Middle-Down Analysis Using an Orbitrap. *Journal of Proteome Research* **2010,** *9* (8), 3886-3890.

72. Wei, B.; Zenaidee, M. A.; Lantz, C.; Ogorzalek Loo, R. R.; Loo, J. A., Towards understanding the formation of internal fragments generated by collisionally activated dissociation for top-down mass spectrometry. *Anal. Chim. Acta* **2022,** *1194*, 339400.

73. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M., Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal. Chem.* **1996,** *68* (5), 850-858.

74. Yin, S.; Loo, J. A., Elucidating the site of protein-ATP binding by top-down mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2010,** *21*, 899-907.

75. Zhang, J.; Ogorzalek Loo, R. R.; Loo, J. A., Increasing fragmentation of disulfide-bonded proteins for top–down mass spectrometry by supercharging. *Int. J. Mass Spectrom.* **2015,** *377*, 546-556.

76. Chen, F.; G[illegible]lbakan, B.; Weidmann, S.; Fagerer, S. R.; Ib[illegible]ez, A. J.; Zenobi, R., Applying mass spectrometry to study non-covalent biomolecule complexes. *Mass Spectrometry Reviews* **2016,** *35* (1), 48-70.

77. Li, H.; Wolff, J. J.; Van Orden, S. L.; Loo, J. A., Native top-down electrospray ionization-mass spectrometry of 158 kDa protein complex by high-resolution fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2014,** *86*, 317-320.

78. Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A., An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* **2018,** *10* (2), 139-148.

79. Brien, J. P.; Li, W.; Zhang, Y.; Brodbelt, J. S., Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry. *Journal of the American Chemical Society* **2014,** *136* (37), 12920-12928.

80. Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G., Mass spectrometry of peptides and proteins. *Methods* **2005,** *35* (3), 211-22.

81. ROEPSTORFE, P., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **1984,** *11*, 601-605.

82. Sleno, L.; Volmer, D. A., Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **2004,** *39* (10), 1091-1112.

83. Macias, L. A.; Santos, I. C.; Brodbelt, J. S., Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2020,** *92* (1), 227–251.

84. Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A., ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* **2021,** *20* (4), 1928–1935.

85. Mcluckey, S. A., Principles of collisional activation in analytical mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1992,** *3* (6), 599-614.

86. Mitchell Wells, J.; McLuckey, S. A., Collision‐Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*, Academic Press: 2005; Vol. 402, pp 148-185.

87. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **2007,** *4* (9), 709-712.

88. Brodbelt, J. S., Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016,** *88* (1), 30-51.

89. Yang, Y.-H.; Lee, K.; Jang, K.-S.; Kim, Y.-G.; Park, S.-H.; Lee, C.-S.; Kim, B.-G., Low mass cutoff evasion with qz value optimization in ion trap. *Analytical Biochemistry* **2009,** *387* (1), 133-135.

90. Stiving, A. Q.; Vanaernum, Z. L.; Busch, F.; Harvey, S. R.; Sarni, S. H.; Wysocki, V. H., Surface-Induced Dissociation: An Effective Method for Characterization of Protein Quaternary Structure. *Analytical Chemistry* **2019,** *91*, 190-209.

91. Zubarev, R. A.; Kelleher, N. L.; Mclafferty, F. W., Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998,** *120* (13), 3265-3266.

92. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; Mclafferty, F. W., Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Anal. Chem.* **2000,** *72* (3), 563-573.

93. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (26), 9528-9533.

94. McLuckey, S. A.; Goeringer, D. E., SPECIAL FEATURE: TUTORIAL Slow Heating Methods in Tandem Mass Spectrometry. *Journal of Mass Spectrometry* **1997,** *32* (5), 461-474.

95. Qi, Y.; Volmer, D. A., Electron-based fragmentation methods in mass spectrometry: An overview. *Mass Spectrom Rev* **2017,** *36* (1), 4-15.

96. Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F., Radical solutions: Principles and application of electron-based dissociation in mass spectrometry-based analysis of protein structure. *Mass Spectrometry Reviews* **2018**.

97. Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, Ü. A.; Tsybin, Y. O., Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chemical Society Reviews* **2013,** *42*, 5014-5030.

98. Xie, Y.; Zhang, J.; Yin, S.; Loo, J. A., Top-down ESI-ECD-FT-ICR mass spectrometry localizes noncovalent protein-ligand binding sites. *Journal of the American Chemical Society* **2006,** *128*, 14432-14433.

99. Riley, N. M.; Coon, J. J., The Role of Electron Transfer Dissociation in Modern Proteomics. *Analytical Chemistry* **2018,** *90*, 40-64.

100. Coon, J. J., Collisions or Electrons? Protein Sequence Analysis in the 21st Century. *Analytical Chemistry* **2009,** *81* (9), 3208-3215.

101. Fung, Y. M. E.; Adams, C. M.; Zubarev, R. A., Electron ionization dissociation of singly and multiply charged peptides. *J. Am. Chem. Soc.* **2009,** *131*, 9977-9985.

102. Zubarev, R. A.; Yang, H., Multiple soft ionization of gas-phase proteins and swift backbone dissociation in collisions with $\leqslant$ 99 eV electrons. *Angewandte Chemie - International Edition* **2010,** *49*, 1439-1441.

103. Kaczorowska, M. A.; Cooper, H. J., Electron induced dissociation (EID) tandem mass spectrometry of octaethylporphyrin and its iron(iii) complex. *Chemical Communications* **2011,** *47*, 418-420.

104. Li, H.; Sheng, Y.; McGee, W.; Cammarata, M.; Holden, D.; Loo, J. A., Structural Characterization of Native Proteins and Protein Complexes by Electron Ionization Dissociation-Mass Spectrometry. *Anal. Chem.* **2017,** *89*, 2731-2738.

105. Zenaidee, M. A.; Lantz, C.; Perkins, T.; Jung, W.; Loo, R. R. O.; Loo, J. A., Internal Fragments Generated by Electron Ionization Dissociation Enhance Protein Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020,** *31* (9), 1896–1902.

106. Brodbelt, J. S., Photodissociation mass spectrometry: New tools for characterization of biological molecules. *Chemical Society Reviews* **2014,** *43*, 2757-2783.

107. Brodbelt, J. S.; Morrison, L. J.; Santos, I. s., Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chemical Reviews* **2020,** *120* (7), 3328-3380.

108. Ly, T.; Julian, R. à., Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angewandte Chemie International Edition* **2009,** *48* (39), 7130-7137.

109. Little, D. P.; Speir, J. P.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W., Infrared multiphoton dissociation of large multiply charged ions for biomolecule sequencing. *Analytical Chemistry* **1994,** *66* (18), 2809-2815.

110. 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. *Angew. Chem. Int. Ed.* **2005,** *44* (39), 6399-6403.

111. Williams, E. R.; Furlong, J. J. P.; McLafferty, F. W., Efficiency of collisionally-activated dissociation and 193-nm photodissociation of peptide ions in fourier transform mass spectrometry. *Journal of the American Society for Mass Spectrometry* **1990,** *1* (4), 288-294.

112. Guan, Z.; Kelleher, N. L.; O'Connor, P. B.; Aaserud, D. J.; Little, D. P.; McLafferty, F. W., 193 nm photodissociation of larger multiply-charged biomolecules. *International Journal of Mass Spectrometry and Ion Processes* **1996,** *157-158*, 357-364.

113. Theisen, A.; Black, R.; Corinti, D.; Brown, J. M.; Bellina, B.; Barran, P. E., Initial Protein Unfolding Events in Ubiquitin, Cytochrome c and Myoglobin Are Revealed with the Use of 213 nm UVPD Coupled to IM-MS. *Journal of the American Society for Mass Spectrometry* **2019,** *30* (1), 24-33.

114. Ly, T.; Julian, R. R., Elucidating the Tertiary Structure of Protein Ions in Vacuo with Site Specific Photoinitiated Radical Reactions. *Journal of the American Chemical Society* **2010,** *132* (25), 8602-8609.

115. Maitre, P.; Scuderi, D.; Corinti, D.; Chiavarino, B.; Crestoni, M. E.; Fornarini, S., Applications of Infrared Multiple Photon Dissociation (IRMPD) to the Detection of Posttranslational Modifications. *Chemical Reviews* **2020,** *120* (7), 3261-3295.

116. Polfer, N. C., Infrared multiple photon dissociation spectroscopy of trapped ions. *Chemical Society Reviews* **2011,** *40* (5), 2211-2221.

117. Crowe, M. C.; Brodbelt, J. S., Infrared multiphoton dissociation (IRMPD) and collisionally activated dissociationof peptides in a quadrupole ion trapwith selective IRMPD of phosphopeptides. *Journal of the American Society for Mass Spectrometry* **2004,** *15* (11), 1581-1592.

118. 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. *Journal of the American Chemical Society* **2013,** *135* (34), 12646-12651.

119. Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S., Ultrafast Ultraviolet Photodissociation at 193 nm and its Applicability to Proteomic Workflows. *Journal of Proteome Research* **2010,** *9* (8), 4205-4214.

120. Mayfield, J. E.; Robinson, M. R.; Cotham, V. C.; Irani, S.; Matthews, W. L.; Ram, A.; Gilmour, D. S.; Cannon, J. R.; Zhang, Y. J.; Brodbelt, J. S., Mapping the Phosphorylation Pattern of Drosophila melanogaster RNA Polymerase II Carboxyl-Terminal Domain Using Ultraviolet Photodissociation Mass Spectrometry. *ACS Chemical Biology* **2017,** *12* (1), 153-162.

121. Crittenden, C. M.; Novelli, E. T.; Mehaffey, M. R.; Xu, G. N.; Giles, D. H.; Fies, W. A.; Dalby, K. N.; Webb, L. J.; Brodbelt, J. S., Structural Evaluation of Protein/Metal Complexes via Native Electrospray Ultraviolet Photodissociation Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (5), 1140-1150.

122. Biemann, K., Contributions of mass spectrometry to peptide and protein structure. *Biomedical & Environmental Mass Spectrometry* **1988,** *16* (1-12), 99-111.

123. Biemann, K., Mass spectrometry of peptides and proteins. *Annual review of biochemistry* **1992,** *61* (1), 977-1010.

124. Roth, K. D. W.; Huang, Z. ô.; Sadagopan, N.; Watson, J. T., Charge derivatization of peptides for analysis by mass spectrometry. *Mass Spectrometry Reviews* **1998,** *17* (4), 255-274.

125. Medzihradszky, K. F.; Chalkley, R. J., Lessons in de novo peptide sequencing by tandem mass spectrometry. *Mass Spectrometry Reviews* **2013,** *34* (1), 43-63.

126. Reilly, J. P., Ultraviolet photofragmentation of biomolecular ions. *Mass Spectrom. Rev.* **2009,** *28* (3), 425-447.

127. Patil, G. V.; Hamilton, R. E.; Day, R. A., Schiff base derivatives of peptide esters: Relative abundance of N-terminal, C-terminal and ق | ¬ internalق | └ fragments as a function of the blocking group. *Organic Mass Spectrometry* **1973,** *7* (7), 817-838.

128. Laursen, R. A., Solid-Phase Edman Degradation. *European Journal of Biochemistry* **1971,** *20* (1), 89-102.

129. Putney, S.; Herlihy, W.; Royal, N.; Pang, H.; Aposhian, H. V.; Pickering, L.; Belagaje, R.; Biemann, K.; Page, D.; Kuby, S., Rabbit muscle creatine phosphokinase. CDNA cloning, primary structure and detection of human homologues. *Journal of Biological Chemistry* **1984,** *259* (23), 14317-14320.

130. Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T., Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Analytical chemistry* **1987,** *59* (21), 2621-2625.

131. Martin, S. A.; Biemann, K., A comparison of keV atom bombardment mass spectra of peptides obtained with a two-sector mass spectrometer with those from a four-sector tandem mass spectrometer. *International Journal of Mass Spectrometry and Ion Processes* **1987,** *78*, 213-228.

132. Johnson, R. S.; Biemann, K., Computer program (SEQPEP) to aid in the interpretation of high-energy collision tandem mass spectra of peptides. *Biomedical & Environmental Mass Spectrometry* **1989,** *18* (11), 945-957.

133. Falick, A. M.; Hines, W. M.; Medzihradszky, K. F.; Baldwin, M. A.; Gibson, B. W., Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. *Journal of the American Society for Mass Spectrometry* **1993,** *4* (11), 882-893.

134. Kassel, D. B.; Biemann, K., Differentiation of hydroxyproline isomers and isobars in peptides by tandem mass spectrometry. *Analytical chemistry* **1990,** *62* (15), 1691-1695.

135. Papayannopoulos, I. A.; Biemann, K., Amino acid sequence of a protease inhibitor isolated from Sarcophaga bullata determined by mass spectrometry. *Protein Science* **1992,** *1* (2), 278-288.

136. Zaia, J.; Annan, R. S.; Biemann, K., The correct molecular weight of myoglobin, a common calibrant for mass spectrometry. *Rapid Communications in Mass Spectrometry* **1992,** *6* (1), 32-36.

137. Ballard, K. D.; Gaskell, S. J., Sequential mass spectrometry applied to the study of the formation of "internal" fragment ions of protonated peptides. *International Journal of Mass Spectrometry and Ion Processes* **1991,** *111*, 173-189.

138. Xiao, K.; Yu, F.; Fang, H.; Xue, B.; Liu, Y.; Li, Y.; Tian, Z., Are neutral loss and internal product ions useful for top-down protein identification? *J Proteomics* **2017,** *160*, 21-27.

139. Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L., Analyzing internal fragmentation of electrosprayed ubiquitin ions during beam-type collisional dissociation. *J. Am. Soc. Mass Spectrom.* **2015,** *26* (5), 782-787.

140. Lyon, Y. A.; Riggs, D.; Fornelli, L.; Compton, P. D.; Julian, R. R., The Ups and Downs of Repeated Cleavage and Internal Fragment Production in Top-Down Proteomics. *J. Am. Soc. Mass Spectrom.* **2018,** *29* (1), 150-157.

141. Palzs, B.; Suhal, S., Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews* **2005,** *24*, 508-548.

142. Schmitt, N. D.; Berger, J. M.; Conway, J. B.; Agar, J. N., Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem.* **2021,** *93* (16), 6355-6362.

143. Michalski, A.; Neuhauser, N.; Cox, J.; Mann, M., A Systematic Investigation into the Nature of Tryptic HCD Spectra. *J. Proteome Res.* **2012,** *11* (11), 5479-5491.

144. Barran, P. E.; Polfer, N. C.; Campopiano, D. J.; Clarke, D. J.; Langridge-Smith, P. R. R.; Langley, R. J.; Govan, J. R. W.; Maxwell, A.; Dorin, J. R.; Millar, R. P.; Bowers, M. T., Is it

biologically relevant to measure the structures of small peptides in the gas-phase? *Int. J. Mass Spectrom.* **2005,** *240* (3), 273-284.

145. Chi, H.; Sun, R.-X.; Yang, B.; Song, C.-Q.; Wang, L.-H.; Liu, C.; Fu, Y.; Yuan, Z.-F.; Wang, H.-P.; He, S.-M.; Dong, M.-Q., pNovo: De novo Peptide Sequencing and Identification Using HCD Spectra. *Journal of Proteome Research* **2010,** *9* (5), 2713-2724.

146. Cobb, J. S.; Easterling, M. L.; Agar, J. N., Structural characterization of intact proteins is enhanced by prevalent fragmentation pathways rarely observed for peptides. *J. Am. Soc. Mass Spectrom.* **2010,** *21* (6), 949-959.

147. Chen, J.; Shiyanov, P.; Green, K. B., Top-down mass spectrometry of intact phosphorylated β-casein: Correlation between the precursor charge state and internal fragments. *J. Mass Spectrom.* **2019,** *54* (6), 527-539.

148. Zenaidee, M. A.; Wei, B.; Lantz, C.; Wu, H. T.; Lambeth, T. R.; Diedrich, J. K.; Ogorzalek Loo, R. R.; Julian, R. R.; Loo, J. A., Internal Fragments Generated from Different Top-Down Mass Spectrometry Fragmentation Methods Extend Protein Sequence Coverage. *J. Am. Soc. Mass Spectrom.* **2021,** *32* (7), 1752-1758.

149. Holden, D. D.; Brodbelt, J. S., Improving Performance Metrics of Ultraviolet Photodissociation Mass Spectrometry by Selective Precursor Ejection. *Analytical Chemistry* **2017,** *89* (1), 837-846.

150. Dunham, S. D.; Wei, B.; Lantz, C.; Loo, J. A.; Brodbelt, J. S., Impact of Internal Fragments on Top-Down Analysis of Intact Proteins by 193 nm UVPD. *Journal of Proteome Research* **2023,** *22* (1), 170-181.

151. Matsumura, M.; Signor, G.; Matthews, B. W., Substantial increase of protein stability by multiple disulphide bonds. *Nature* **1989,** *342* (6247), 291-293.

152. Wedemeyer, W. J.; Welker, E.; Narayan, M.; Scheraga, H. A., Disulfide Bonds and Protein Folding. *Biochem.* **2000,** *39* (15), 4207-4216.

153. 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. *J. Pharm. Biomed. Anal.* **2021,** *195*, 113893.

154. Wei, B.; Zenaidee, M. A.; Lantz, C.; Williams, B. J.; Totten, S.; Ogorzalek Loo, R. R.; Loo, J. A., Top-down mass spectrometry and assigning internal fragments for determining disulfide bond positions in proteins. *The Analyst* **2023,** *148* (1), 26-37.

155. Griaud, F.; Denefeld, B.; Kao-Scharf, C.-Y.; Dayer, J.; Lang, M.; Chen, J.-Y.; Berg, M., All Ion Differential Analysis Refines the Detection of Terminal and Internal Diagnostic Fragment Ions for the Characterization of Biologics Product-Related Variants and Impurities by Middle-down Mass Spectrometry. *Anal. Chem.* **2019,** *91* (14), 8845-8852.

156. Wei, B.; Lantz, C.; Liu, W.; Viner, R.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A., Added Value of Internal Fragments for Top-Down Mass Spectrometry of Intact Monoclonal Antibodies and Antibody–Drug Conjugates. *Analytical Chemistry* **2023,** *95* (24), 9347–9356.

157. Rolfs, Z.; Smith, L. M., Internal Fragment Ions Disambiguate and Increase Identifications in Top-Down Proteomics. *J. Proteome Res.* **2021,** *20* (12), 5412-5418.

158. Kenderdine, T.; McIntyre, W.; Yassaghi, G.; Rollo, D.; Bunkowski, A.; Goerlach, L.; Suckau, D.; Tremintin, G.; Greig, M.; Bell, C.; Fabris, D., Integrating Internal Fragments in the Interpretation of Top-Down Sequencing Data of Larger Oligonucleotides. *Journal of the American Society for Mass Spectrometry* **2023**.

159. Goldstein, G. J. S.; Tsai, H. Z.; Cosimi, B. A.; Russell, P. S.; Norman, D.; Barry, J.; Shield, C. F.; Cho, S. I.; Levey, A. S.; Burdick, J. F.; Williams, G. M.; Stuart, F. P.; Alexander, J. W.; First, R.; Helderman, J. H.; Wathen, R. L.; Lordon, R. E.; Sampson, D.; Levin, B. S., A Randomized Clinical Trial of OKT3 Monoclonal Antibody for Acute Rejection of Cadaveric Renal Transplants. *New England Journal of Medicine* **1985,** *313* (6), 337-342.

160. https://www.antibodysociety.org/antibody-therapeutics-product-data/.

161. Schrama, D.; Reisfeld, R. A.; Becker, J. C., Antibody targeted drugs as cancer therapeutics. *Nature Reviews Drug Discovery* **2006,** *5* (2), 147-159.

162. Bebbington, C.; Yarranton, G., Antibodies for the treatment of bacterial infections: current experience and future prospects. *Current Opinion in Biotechnology* **2008,** *19* (6), 613-619.

163. Buss, N. A. P. S.; Henderson, S. J.; McFarlane, M.; Shenton, J. M.; de Haan, L., Monoclonal antibody therapeutics: history and future. *Current Opinion in Pharmacology* **2012,** *12* (5), 615-622.

164. Ross, J. S.; Slodkowska, E. A.; Symmans, W. F.; Pusztai, L.; Ravdin, P. M.; Hortobagyi, G. N., The HER-2 Receptor and Breast Cancer: Ten Years of Targeted Anti – HER-2 Therapy and Personalized Medicine. *The Oncologist* **2009,** *14* (4), 320-368.

165. 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* (8), 1861–1885.

166. Walsh, G.; Walsh, E., Biopharmaceutical benchmarks 2022. *Nature Biotechnology* **2022,** *40* (12), 1722-1760.

167. Saunders, K. O., Conceptual Approaches to Modulating Antibody Effector Functions and Circulation Half-Life. *Frontiers in Immunology* **2019,** *10*.

168. Redman, J. M.; Hill, E. M.; AlDeghaither, D.; Weiner, L. M., Mechanisms of action of therapeutic antibodies for cancer. *Molecular Immunology* **2015,** *67* (2, Part A), 28-45.

169. Grilo, A. L.; Mantalaris, A., The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends in Biotechnology* **2019,** *37* (1), 9-16.

170. Deshaies, R. J., Multispecific drugs herald a new era of biopharmaceutical innovation. *Nature* **2020,** *580* (7803), 329-338.

171. Labrijn, A. F.; Janmaat, M. L.; Reichert, J. M.; Parren, P. W. H. I., Bispecific antibodies: a mechanistic review of the pipeline. *Nature Reviews Drug Discovery* **2019,** *18* (8), 585-608.

172. Beck, A.; Goetsch, L.; Dumontet, C.; Corvaأ»a, N., Strategies and challenges for the next generation of antibodyق | ôdrug conjugates. *Nature Reviews Drug Discovery* **2017,** *16* (5), 315-337.

173. 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. *Journal of Cellular Physiology* **2019,** *234* (5), 5628-5642.

174. Maecker, H.; Jonnalagadda, V.; Bhakta, S.; Jammalamadaka, V.; Junutula, J. R., Exploration of the antibody–drug conjugate clinical landscape. *mAbs* **2023,** *15* (1), 2229101.

175. Ornes, S., Antibody–drug conjugates. *Proceedings of the National Academy of Sciences* **2013,** *110* (34), 13695-13695.

176. Joubert, N.; Beck, A.; Dumontet, C.; Denevault-Sabourin, C. Antibody–Drug Conjugates: The Last Decade *Pharmaceuticals* [Online], 2020.

177. Liu, H.; May, K., Disulfide bond structures of IgG molecules. *mAbs* **2012,** *4* (1), 17-23.

178. Nimmerjahn, F.; Vidarsson, G.; Cragg, M. S., Effect of posttranslational modifications and subclass on IgG activity: from immunity to immunotherapy. *Nature Immunology* **2023**.

179. Tsuchikama, K.; An, Z., Antibody-drug conjugates: recent advances in conjugation and linker chemistries. *Protein & Cell* **2018,** *9* (1), 33-46.

180. Haque, M.; Forte, N.; Baker, J. R., Site-selective lysine conjugation methods and applications towards antibody–drug conjugates. *Chemical Communications* **2021,** *57* (82), 10689-10702.

181. Lambert, J. M.; Chari, R. V. J., Ado-trastuzumab Emtansine (T-DM1): An Antibody–Drug Conjugate (ADC) for HER2-Positive Breast Cancer. *Journal of Medicinal Chemistry* **2014,** *57* (16), 6949-6964.

182. Yoder, N. C.; Bai, C.; Tavares, D.; Widdison, W. C.; Whiteman, K. R.; Wilhelm, A.; Wilhelm, S. D.; Mcshea, M. A.; Maloney, E. K.; Ab, O.; Wang, L.; Jin, S.; Erickson, H. K.;

Keating, T. A.; Lambert, J. M., A Case Study Comparing Heterogeneous Lysine- and Site-Specific Cysteine-Conjugated Maytansinoid Antibody-Drug Conjugates (ADCs) Illustrates the Benefits of Lysine Conjugation. *Molecular Pharmaceutics* **2019,** *16* (9), 3926-3937.

183. Rathore, A. S.; Winkle, H., Quality by design for biopharmaceuticals. *Nature Biotechnology* **2009,** *27* (1), 26-34.

184. Meibohm, B., Pharmacokinetics and Pharmacodynamics of Therapeutic Peptides and Proteins. In *Pharmaceutical Biotechnology: Fundamentals and Applications*, Crommelin, D. J. A.; Sindelar, R. D.; Meibohm, B., Eds. Springer International Publishing: Cham, 2019; pp 105-137.

185. Villacorta, P. J.; Salmer╡ ╠n-Garc╡╤a, A.; Pelta, D. A.; Cabeza, J.; Lario, A.; Navas, N., Cluster-based comparison of the peptide mass fingerprint obtained by MALDI-TOF mass spectrometry. A case study: long-term stability of rituximab. *Analyst* **2015,** *140* (5), 1717-1730.

186. Suntornsuk, L., Recent advances of capillary electrophoresis in pharmaceutical analysis. *Analytical and Bioanalytical Chemistry* **2010,** *398* (1), 29-52.

187. Temel, D. B.; Landsman, P.; Brader, M. L., Chapter Fourteen - Orthogonal Methods for Characterizing the Unfolding of Therapeutic Monoclonal Antibodies: Differential Scanning Calorimetry, Isothermal Chemical Denaturation, and Intrinsic Fluorescence with Concomitant Static Light Scattering. In *Methods in Enzymology*, Feig, A. L., Ed. Academic Press: 2016; Vol. 567, pp 359-389.

188. Garidel, P.; Hegyi, M.; Bassarab, S.; Weichel, M., A rapid, sensitive and economical assessment of monoclonal antibody conformational stability by intrinsic tryptophan fluorescence spectroscopy. *Biotechnology Journal* **2008,** *3* (9-10), 1201-1211.

189. Li, N.; Kessler, K.; Bass, L.; Zeng, D., Evaluation of the iCE280 Analyzer as a potential high-throughput tool for formulation development. *Journal of Pharmaceutical and Biomedical Analysis* **2007,** *43* (3), 963-972.

190. Luypaert, J.; Massart, D. L.; Vander Heyden, Y., Near-infrared spectroscopy applications in pharmaceutical analysis. *Talanta* **2007,** *72* (3), 865-883.

191. Hong, P.; Koza, S.; Bouvier, E. S. P., A REVIEW SIZE-EXCLUSION CHROMATOGRAPHY FOR THE ANALYSIS OF PROTEIN BIOTHERAPEUTICS AND THEIR AGGREGATES. *Journal of Liquid Chromatography & Related Technologies* **2012,** *35* (20), 2923-2950.

192. Haverick, M.; Mengisen, S.; Shameem, M.; Ambrogelly, A., Separation of mAbs molecular variants by analytical hydrophobic interaction chromatography HPLC: overview and applications. *mAbs* **2014,** *6* (4), 852-858.

193. Harris, R. J.; Kabakoff, B.; Macchi, F. D.; Shen, F. J.; Kwong, M.; Andya, J. D.; Shire, S. J.; Bjork, N.; Totpal, K.; Chen, A. B., Identification of multiple sources of charge heterogeneity in a recombinant antibody. *Journal of Chromatography B: Biomedical Sciences and Applications* **2001,** *752* (2), 233-245.

194. Zhu, X.; Huo, S.; Xue, C.; An, B.; Qu, J., Current LC-MS-based strategies for characterization and quantification of antibody-drug conjugates. *Journal of Pharmaceutical Analysis* **2020,** *10* (3), 209-220.

195. Graf, T.; Heinrich, K.; Grunert, I.; Wegele, H.; Haberger, M.; Bulau, P.; Leiss, M., Recent advances in LC–MS based characterization of protein-based bio-therapeutics –

mastering analytical challenges posed by the increasing format complexity. *Journal of Pharmaceutical and Biomedical Analysis* **2020,** *186*, 113251.

196. Yang, F.; Zhang, J.; Buettner, A.; Vosika, E.; Sadek, M.; Hao, Z.; Reusch, D.; Koenig, M.; Chan, W.; Bathke, A.; Pallat, H.; Lundin, V.; Kepert, J. F.; Bulau, P.; Deperalta, G.; Yu, C.; Beardsley, R.; Camilli, T.; Harris, R.; Stults, J., Mass spectrometry-based multi-attribute method in protein therapeutics product quality monitoring and quality control. *mAbs* **2023,** *15* (1).

197. Bobály, B.; Fleury-Souverain, S.; Beck, A.; Veuthey, J.-L.; Guillarme, D.; Fekete, S., Current possibilities of liquid chromatography for the characterization of antibody-drug conjugates. *Journal of Pharmaceutical and Biomedical Analysis* **2018,** *147*, 493-505.

198. Bongers, J.; Cummings, J. J.; Ebert, M. B.; Federici, M. M.; Gledhill, L.; Gulati, D.; Hilliard, G. M.; Jones, B. H.; Lee, K. R.; Mozdzanowski, J.; Naimoli, M.; Burman, S., Validation of a peptide mapping method for a therapeutic monoclonal antibody: what could we possibly learn about a method we have run 100 times? *Journal of Pharmaceutical and Biomedical Analysis* **2000,** *21* (6), 1099-1128.

199. Luo, Q.; Chung, H. H.; Borths, C.; Janson, M.; Wen, J.; Joubert, M. K.; Wypych, J., Structural Characterization of a Monoclonal Antibody–Maytansinoid Immunoconjugate. *Analytical Chemistry* **2016,** *88* (1), 695-702.

200. Campuzano, I. D. G.; Netirojjanakul, C.; Nshanian, M.; Lippens, J. L.; Kilgour, D. P. A.; Van Orden, S. L.; Loo, J., Native-MS Analysis of Monoclonal Antibody Conjugates by

Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **2017**, acs.analchem.7b03021.

201. Campuzano, I. D. G.; Nshanian, M.; Spahr, C.; Lantz, C.; Netirojjanakul, C.; Li, H.; Wongkongkathep, P.; Wolff, J. J.; Loo, J. A., High Mass Analysis with a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: From Inorganic Salt Clusters to Antibody Conjugates and Beyond. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (5), 1155-1162.

202. Yan, Y.; Xing, T.; Wang, S.; Daly, T. J.; Li, N., Online coupling of analytical hydrophobic interaction chromatography with native mass spectrometry for the characterization of monoclonal antibodies and related products. *Journal of Pharmaceutical and Biomedical Analysis* **2020,** *186*, 113313.

203. 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* (1), 1682895.

204. 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. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (8), 1706-1712.

205. Liu, A. P.; Yan, Y.; Wang, S.; Li, N., Coupling Anion Exchange Chromatography with Native Mass Spectrometry for Charge Heterogeneity Characterization of Monoclonal Antibodies. *Analytical Chemistry* **2022**.

206. 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. *Analytical Chemistry* **2018,** *90* (14), 8421-8429.

207. Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O., Analysis of Intact Monoclonal Antibody IgG1 by Electron Transfer Dissociation Orbitrap FTMS. *Molecular & Cellular Proteomics* **2012,** *11* (12), 1758-1767.

208. 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. *Analytical Chemistry* **2013,** *85* (9), 4239-4246.

209. 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. *Analytical Chemistry* **2014,** *86* (6), 3005-3012.

210. 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* (1), 766-773.

211. 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* **2019,** *91* (18), 11661-11669.

212. 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. *Analytical Chemistry* **2020,** *92* (14), 9790-9798.

213. Ren, D.; Pipes, G. D.; Liu, D.; Shih, L.-Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V., An improved trypsin digestion method minimizes digestion-induced modifications on proteins. *Analytical Biochemistry* **2009,** *392* (1), 12-21.

214. Hains, P. G.; Robinson, P. J., The Impact of Commonly Used Alkylating Agents on Artifactual Peptide Modification. *Journal of Proteome Research* **2017,** *16* (9), 3443-3447.

215. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker, J. M., Ed. Springer: Humana Press Inc., Totowa, NJ, 2005; pp 571-607.

Chapter 2: Towards understanding the formation of internal fragments generated by collisionally activated dissociation for top-down mass spectrometry

Reprinted with permission from

Wei, B.; Zenaidee, M. A.; Lantz, C.; Ogorzalek Loo, R. R.; Loo, J. A. Towards understanding the formation of internal fragments generated by collisionally activated dissociation for top-down mass spectrometry. *Anal. Chim. Acta* **2022**, *1194*, 339400. DOI: 10.1016/j.aca.2021.339400.

Benqian Wei[a], Muhammad A. Zenaidee[a, c], Carter Lantz[a], Rachel R. Ogorzalek Loo[a], Joseph A. Loo[a, b, *]

[a] Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA; [b] Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA; [c] Australian Proteome Analysis Facility, Macquarie University, NSW, Australia.

* Corresponding Author: Joseph A. Loo, Department of Chemistry and Biochemistry, Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA, E-mail: jloo@chem.ucla.edu.

Abstract

Top-down mass spectrometry (TD-MS) generates fragment ions that returns information on the polypeptide amino acid sequence. In addition to terminal fragments, internal fragments that result from multiple cleavage events can also be formed. Traditionally, internal fragments are largely ignored due to a lack of available software to reliably assign them, mainly caused by a poor understanding of their formation mechanism. To accurately assign internal fragments, their formation process needs to be better understood. Here, we applied a statistical method to compare fragmentation patterns of internal and terminal fragments of peptides and proteins generated by collisionally activated dissociation (CAD). Internal fragments share similar fragmentation propensities with terminal fragments (e.g., enhanced cleavages N-terminal to proline and C-terminal to acidic residues), suggesting that their formation follows conventional CAD pathways. Internal fragments should be generated by subsequent cleavages of terminal fragments and their formation can be explained by the well-known mobile proton model. In addition, internal fragments can be coupled with terminal fragments to form complementary product ions that span the entire protein sequence. These enhance our understanding of internal fragment formation and can help improve sequencing algorithms to accurately assign internal fragments, which will ultimately lead to more efficient and comprehensive TD-MS analysis of proteins and proteoforms.

Keywords: Top-down mass spectrometry (TD-MS); Internal fragment; Collisionally activated dissociation (CAD); Fragmentation propensity; Mobile proton model

Highlights:

- Internal fragments can enhance peptide and protein sequence coverage up to 100%.
- The mechanism of internal fragment formation by CAD is reported for the first time.
- CAD-internal fragments are formed by cleavages of terminal fragments and can be explained by the mobile proton model.
- Combining internal and terminal fragments can generate complementary ion pairs that cover the entire protein sequence.

1. Introduction

Traditional mass spectrometry (MS) sequence analysis of proteins is typically performed by the “bottom-up” strategy in which intact proteins are digested into small peptides and then analyzed by MS [1,2]. Methods such as top-down mass spectrometry (TDMS) have gained popularity to characterize the structure of proteins and proteoforms. In TD-MS, the digestion and separation steps required for “bottom-up” are bypassed, allowing for the preservation of labile posttranslational modifications (PTMs) [3-6]. TD-MS measurements start by generating intact gas-phase protein ions using electrospray ionization (ESI), which are subsequently fragmented by different activation/dissociation techniques to return information on the protein primary structure, i.e., sequence. Many ion activation methods have been developed throughout the years with each having discrete advantages [7,8]. The most widely used fragmentation method is collisionally activated dissociation (CAD) [9], but other fragmentation techniques including electron capture dissociation (ECD) [10,11], electron transfer dissociation (ETD) [12], electron

ionization dissociation (EID) [13,14] and ultraviolet photodissociation (UVPD) [15,16] have gained in popularity. Nevertheless, CAD remains popular and has been the “gold standard” for all ion activation methods to be compared with, owing to its high efficiency of energy accumulation and fragmentation, as well as its compatibility with nearly all MS instrumentation platforms [17].

The thermal collision process between inert gas molecules and polypeptide ions during CAD transfers part of the precursor ion's kinetic energy into internal energy, of which the accumulation ultimately leads to dissociation, generating *b*- and *y*-type fragment ions [9]. In CAD-type experiments, many factors related to the polypeptide ion including amino acid composition, gas-phase basicity, basic residue content, secondary structure and charge state can significantly affect the fragmentation pattern. The mobile proton model has been developed to describe the gas-phase fragmentation propensities of peptides to produce these *b* and *y* fragment ions [18-20]. If the number of available charges of a peptide ion is greater than the number of its strongly basic residues (*e.g.*, arginine), there will be protons mobilizing along the peptide backbone to induce cleavages at various amide bonds [18,21]. In contrast, the presence of fewer charges than strongly basic residues will lead to protons sequestered at these sites, resulting in higher energy requirements to induce backbone fragmentation [18,22,23]. The mobile proton model has been applied to elucidate different selective cleavage observations of gas-phase peptide ions including enhanced dissociation N-terminal to proline [24-29], C-terminal to histidine [30,31] and C-terminal to acidic residues (aspartic acid, glutamic acid) [23,32-34]. Large-scale statistical analyses have also been reported to support the mobile proton model, which show other selective and non-selective fragmentation propensities [35-

38].

Protein cleavage products formed by TD-MS can either be i) a terminal fragment ion that contains the amino-terminus (*a*, *b* or *c* fragment) or carboxy-terminus (*x*, *y*, or *z* fragment) of the precursor ion from a single bond cleavage event, or ii) an internal fragment ion generated by multiple cleavage events forming *ax*, *ay*, *az*, *bx*, *by*, *bz*, *cx*, *cy*, and *cz* fragment ions (with the first letter designating cleavage on the N-terminal side and the second letter designating cleavage on the C-terminal side), depending on the ion activation method utilized and the cleavage sites [39-41]. Internal fragment ions have been largely ignored by the TD-MS community due to a poor understanding of their formation process, resulting in a lack of software to accurately and reliably assign them. In a single TD-MS spectrum, the number of theoretical internal fragments that can be produced is significantly greater than the number of theoretical terminal fragments, and this gap increases exponentially as the size of the protein increases [42,43], resulting in substantial computational demand. As a result, a large proportion of the mass spectral signals can go unassigned by ignoring the possibility of internal fragments. Potentially, the inclusion of internal fragments in TD-MS analysis can offer much richer protein sequence information if accurately assigned.

Previous studies that included internal fragment analysis was initially limited to peptides [44,45], and has been expanded to TDMS of intact proteins [39,43,46-49] and protein complexes [14,50] in recent years. Among these studies, various ion activation methods have been utilized including CAD [39,46-49], ECD [48,50] and EID [14,43,48]. Regardless of the technique used for fragmentation, all of these studies showcase that the inclusion of internal

fragments can result in greater protein sequence coverage, significantly benefitting TD-MS experiments.

A major obstacle to the inclusion of internal fragments in the TD-MS workflow is the ambiguity of assigning internal fragments that likely scales as the size of the protein increases. Agar and coworkers classified this ambiguity into three subcategories: arrangement ambiguity, frameshift ambiguity, and mass accuracy ambiguity [49]. A better understanding of the formation of internal fragments could be useful for resolving the arrangement ambiguity and frameshift ambiguity, thus increasing the confidence of assigning internal fragments.

Here, we applied a statistical approach to compare the CAD fragmentation patterns of internal fragments and terminal fragments of 42 polypeptides ranging in size from 1.5 to 8.8 kDa. The experiments generated 1412 terminal fragments and 1861 internal fragments, constituting our dataset to perform the statistical analysis. From this data, we demonstrate a relationship between internal and terminal fragments generated by CAD. This is crucial to enhance our understanding of the formation of internal fragments at the molecular level and to improve MS sequencing algorithms that can help incorporate internal fragment analysis into TD-MS workflow [51]. This can ultimately lead to more efficient and comprehensive TD-MS characterization of intact proteins, protein complexes and identification of specific proteoforms, which have been a major challenge in the analytical chemistry field. Furthermore, our results can also be applied to bottom-up and middle-down MS experiments, benefitting the application of internal fragments to the entire protein MS community.

2. Experimental

2.1. Materials and sample preparation

Human [Glu1]-fibrinopeptide B (Glu-fib), insulin chain A (ammonium salt from bovine pancreas), fibronectin type III connecting segment fragment 1-25, melittin, 3X FLAG peptide, C-peptide fragment 3-33 (human), glucagon, oxidized insulin chain B (bovine pancreas), and ubiquitin (bovine erythrocytes) were purchased from Sigma-Aldrich (St. Louis, MO, USA). UOM-6 peptide (1574.84 Da), Tummino peptide (2068.39 Da), synthetic I peptide (3271.88 Da) and synthetic II peptide (3032.53 Da) were synthesized by the University of Michigan Protein Facility. LARL peptide (2014.30 Da) and β-amyloid (1-42; human) were acquired from AnaSpec, Inc. (Fremont, CA, USA). ACTH (1-17) and ACTH (18-39) (human), gastrin releasing peptide (human), xenin, tau peptide (45-73) (exon 2/insert 1 domain), peripheral myelin protein P2 (53-78) (bovine), calcitonin gene related peptide (CGRP 8-37; human), galanin like peptide (GALP; N-terminal fragment, porcine), tau peptide (275-305) (repeat 2 domain), VIP (human, porcine or rat), proinsulin C-peptide (31-63; porcine), OVA (241-270), apelin-36 (human), neuropeptide Y (free acid; human or rat), and *anti*-BetaGamma (MPS - Phosducin - like protein C terminus) were obtained from InnoPep Inc. (San Diego, CA, USA). All peptides were used without further purification. Proteins apomyoglobin (equine skeletal muscle), α-casein and β-casein (bovine milk) and carbonic anhydrase II (bovine) were purchased from Sigma-Aldrich (St. Louis, MO, USA). These proteins were dissolved in water and digested at 37 °C with Glu-C protease for 10 h at a 1:100 protease/protein ratio in 100 mM ammonium acetate solution to obtain polypeptides of less than 10 kDa. For electrospray ionization, all peptides were prepared in 49.5:49.5:1

methanol/water/formic acid solution to a final concentration of 20 μM. The peptides resulting from protein digestion were resuspended with 1% formic acid before mass spectrometry analysis.

For TD-MS of apomyoglobin and carbonic anhydrase II, protein solutions were prepared in 49.5:49.5:1 methanol/water/formic acid solution to a final concentration of 20 μM before mass spectrometry analysis.

2.2. Mass spectrometry

All experiments were conducted on a 15-T SolariX Fourier transform ion cyclotron resonance (FTICR)-MS instrument equipped with an infinity ICR cell (Bruker Daltonics, Billerica, MA, USA). All analyte solutions were loaded into in-house pulled capillaries coated with gold, and electrosprayed by applying a voltage between 0.7 and 1.5 kV on the ESI capillary. Peptide ions were isolated in the quadrupole, with an isolation window between 5 and 15 *m/z* to ensure the minimum precursor ion abundance to be above the 107 level before CAD fragmentation. For CAD MS/MS experiments, the most abundant charge state for each peptide was isolated as the precursor ion to undergo fragmentation. A series of collision energies were applied, ranging from a low energy to reduce the precursor ion signal by ca. 10% to a high energy to reduce the precursor ion signal to ca. 95% of the original level. For some peptides, data from other charge states were acquired if these lower abundance charge states were able to be isolated efficiently and reach the minimum 10^7 signal level threshold. A similar series of collision energies were applied to these lower abundance charge states. For each charge state,1, 2, 3, 5, 6 or 10 collision energies were applied, spanning the collision energy range discussed above.

CAD-MS/MS of apomyoglobin (apoMb) and carbonic anhydrase II (CAII) was done by isolating [apoMb + 17H]$^{17+}$ and [CAII + 32H]$^{32+}$ with an isolation window of 10 *m/z*. The CAD energy was set at 24 V for apoMb and 11 V for CAII to reduce the precursor ion signal to ca. 95% of the original level.

2.3. Data processing and fragment assignment

For CAD MS/MS of polypeptides, raw MS/MS spectra were deconvoluted using either Bruker Data Analysis software (SNAP algorithm) or mMass software version 5.5.0 [52]. Every deconvoluted mass list was internally calibrated against a theoretical fragment list of that specific peptide and uploaded into the ClipsMS program [51] to obtain a matched fragments list. The error for fragment matching was set at 1 ppm and the smallest internal fragment size was set at 2 amino acids. Up to 2 water and ammonia losses were included as unlocalized modifications to avoid masses overlapping between internal fragments and neutral losses of terminal fragments. Only *by* internal fragments were searched for and assigned, and all terminal fragments were assigned before considering internal fragments. All overlapping internal fragments due to the arrangement and frameshift ambiguity [49] were retained in order to include all fragmentation propensity possibilities. After matching, all assigned internal fragments were manually validated against the raw MS/MS spectra to ensure: i) these internal fragments were real peaks rather than noise or isotopes and ii) the masses of matched internal fragments were not overlapping with terminal fragments or neutral losses.

For the TD-MS measurements of apoMb and CAII, similar data analysis parameters were used, with the following exceptions noted. The error for fragment matching was set at 2 ppm

and the smallest internal fragment size was set at 5 amino acids. No localized or unlocalized modifications were imported. The searched fragment types include *a*, *x*, *b*, and *y* for terminal fragments, and *ay*, *bx*, *by* for internal fragments.

2.4. Statistical analysis

2.4.1. Peptide sequence coverage

Peptide sequence coverage is calculated by the number of observed inter-residue cleavage sites divided by the total number of possible inter-residue cleavage sites on the peptide backbone.

2.4.2. Abundance normalization

Normalized abundances were calculated separately for terminal and internal fragments. For each peptide or protein, the absolute abundance of every terminal fragment is divided by the absolute abundance of the most abundant terminal fragment to obtain the normalized abundance of that terminal fragment. Likewise, the normalized abundance of an internal fragment is calculated by dividing the absolute abundance of that internal fragment by the absolute abundance of the most abundant internal fragment. To plot the distribution of normalized abundance adjacent to each amino acid residue, after adding normalized abundances of all peptides and proteins adjacent to an amino acid residue (terminal and internal fragments separately), all 20 normalized abundance values were divided by the largest value to obtain the normalized abundance adjacent to that specific residue. For example, for terminal fragments, N-terminal fragmentation adjacent to proline has the largest normalized abundance after all peptides and proteins are added; therefore, it has a value of 1.00. To plot

the heatmap depicting the normalized abundance deconstructed by residue pair, after adding normalized abundances of all peptides and proteins between a specific residue pair (terminal and internal fragments separately), all 400 normalized abundance values were first cube rooted to avoid extremely light-colored cells for better visualization. These cube-rooted normalized abundance values were divided by the largest value to obtain the normalized abundance of that specific residue pair. For example, for terminal fragments, fragmentation occurring at L|P inter-residue site has the largest normalized abundance after all peptides and proteins added; therefore, it has a value of 1.00.

2.4.3. Delta normalized abundance

To plot the bar graph depicting the difference of normalized abundance between internal and terminal fragments (internal e terminal) for each amino acid residue, all 20 normalized abundances of terminal fragments adjacent to an amino acid residue were subtracted from all 20 normalized abundances of internal fragments, respectively, to obtain 20 delta abundances adjacent to a specific residue. Every delta abundance was then divided by the largest absolute value of all 20 delta abundances to obtain 20 delta normalized abundances. For example, internal fragments generated by C-terminal cleavages adjacent to proline have the largest advantage to terminal fragments adjacent to proline; therefore, the delta normalized abundance adjacent to proline has a value of1.00. Similarly, to plot the heatmap depicting the difference of normalized abundance between internal and terminal fragments (internal e terminal) deconstructed by residue pair, all 400 normalized abundances of terminal fragments deconstructed by residue pair were first subtracted from all 400 normalized abundances of

internal fragments, respectively, to obtain 400 delta abundances between a residue pair. Every delta abundance was then divided by the largest absolute value of all 400 delta abundances to obtain 400 delta normalized abundances. For example, terminal fragments generated by cleaving E|G sites have the largest advantage to internal fragments generated by cleaving E|G sites; therefore, the delta normalized abundance of the E|G inter-residue site has a value of -1.00.

2.4.4. "N-bias" calculation

"N-bias" is calculated by the following equation:

$$\text{N-bias} = (\text{Abun}_{\text{N-term}} - \text{Abun}_{\text{C-term}}) / (\text{Abun}_{\text{N-term}} + \text{Abun}_{\text{C-term}})$$

Where $\text{Abun}_{\text{N-term}}$ is the normalized abundance of N-terminal fragments of an amino acid residue while $\text{Abun}_{\text{C-term}}$ is the normalized abundance of C-terminal fragments of an amino acid residue.

3. Results and discussion

3.1. Internal fragments can enhance peptide sequence information

To demonstrate the use of internal fragments for enhancing peptide sequence information, consider the CAD mass spectrum of the peptide, glucagon (29 amino acids, 3.4 kDa; Figure. 1). CAD of $[\text{glucagon} + 3\text{H}]^{3+}$ causes amide bond cleavages that resulted in not only terminal fragments but also internal fragments (Figure. 1a). Many peaks in this spectrum that cannot be assigned as terminal fragments can be assigned as internal fragments. For example, isotopically resolved (singly charged) peaks at *m/z* 674.3663 (674.3661, theory), 805.4071 (805.4065, theory), and 1483.7302 (1483.7288, theory) were assigned as $b_{26}y_8$, $b_{27}y_8$, and $b_{20}y_{21}$,

respectively. Internal fragments can span much of the amino acid sequence, as shown in the fragment location map (Figure. 1b) for glucagon from our ClipsMS program.

Internal fragments span more of the interior regions of the glucagon sequence, but more importantly provide complementary information to terminal fragments (Figure. 1b and Figure. S1b). Furthermore, terminal fragments generated by CAD of [glucagon + 3H]$^{3+}$ have 5 missed cleavages out of 28 inter-residue sites while internal fragments cover every single inter-residue site to result in 100% sequence coverage (Figure. S1a). These data demonstrate that internal fragments can provide rich sequence information in a single mass spectrum.

For glucagon, both the number and relative abundances of internal fragments increase with collision energy (Figure. 1c). Throughout the energies applied, internal fragments only account for a small portion of the total mass spectral signals (~20-30%); however, they represent a large fraction of the assigned fragments (>50%), enhancing the sequence coverage of glucagon to 100%.

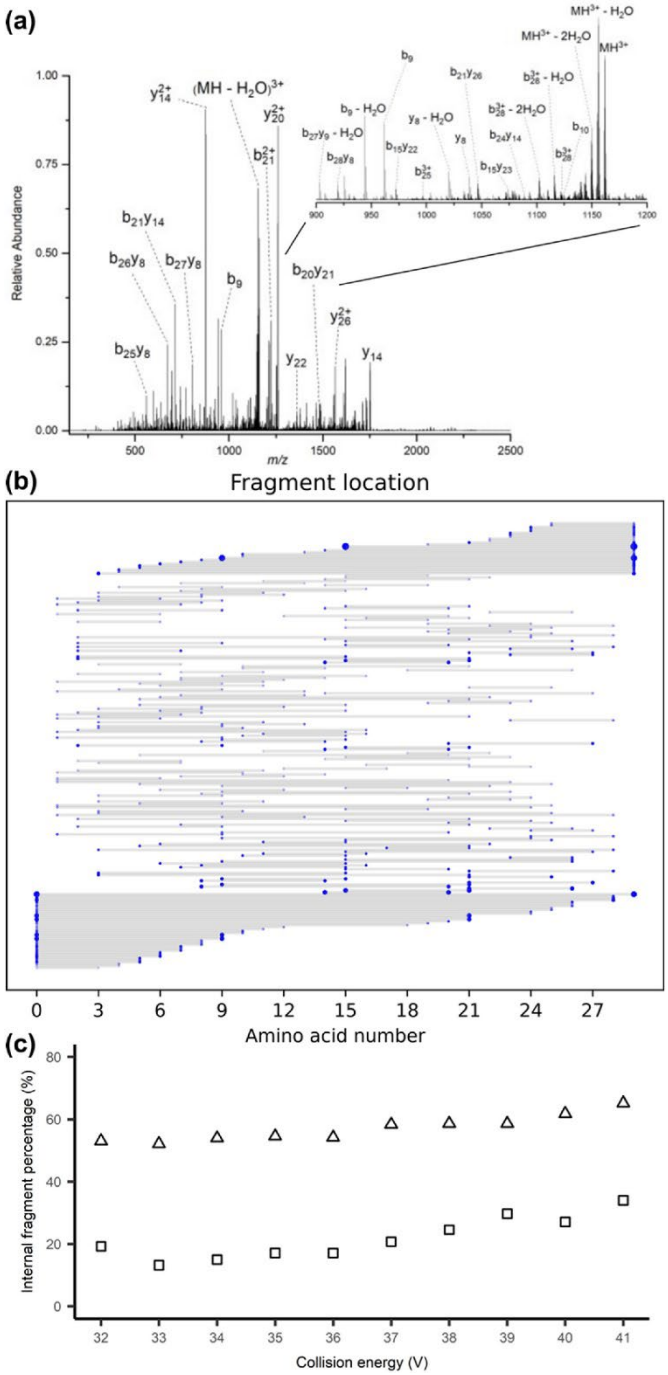


Figure 1. Fragments and sequence coverage analysis of glucagon.

(a) CAD MS/MS spectrum of glucagon in acidic denaturing solution conditions. (b) Fragment location map indicating the region of the protein sequence covered by terminal and internal fragments. (c) A comparison between the number and abundance percentage of internal fragments formed by CAD of glucagon. Open triangles indicate the number percentage of internal fragments while open squares indicate the abundance percentage of internal fragments. Internal fragment percentage is calculated by the internal fragment metric (number or abundance) divided by the sum of the internal and terminal fragment metric (number or

abundance).

To compare the number and abundances between assigned terminal and internal fragments in a larger scale, 42 peptides and proteins ranging from 1.5 kDa to 8.8 kDa were fragmented by CAD (Table S1). The number of assigned fragments listed in Table S1 for each peptide was documented from the experiment generating the most fragments for that specific peptide, regardless of the precursor charge state. Figure. 2 summarizes internal fragment abundances and number percentages for all peptides analyzed in Table S1. This analysis includes all precursor charge states shown in Table S1 and spans collision energies that range from where the precursor ion signal is approximately 90% relative abundance to where almost no precursor ion signal (~5%) can be observed. For all peptides analyzed, the relative abundances of internal fragments are mostly below 20%, indicating that internal fragments usually only account for a small portion of the mass spectral signals in a single spectrum (Figure. 2a). Nonetheless, as shown in Figure. 2b, the percentage of assignable mass spectral signals explained by internal fragments for most peptides lies between 40% and 60%. This can account for up to 100% of the peptide amino acid sequence depending on the collision energy applied. As a result, including internal fragments in a MS/MS analysis can provide valuable information on the polypeptide sequence despite accounting for a small proportion of the ion signal and can be beneficial for protein characterization.

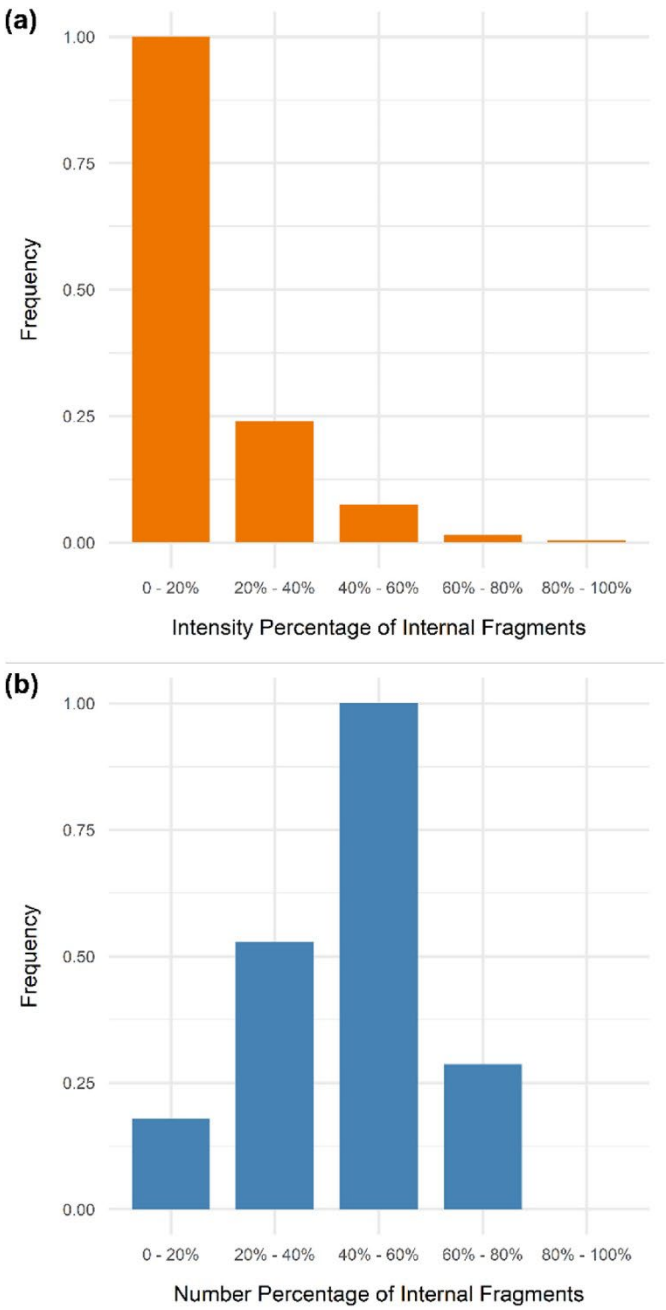


Figure 2. Internal fragment number and abundance percentage analysis for all peptides. The distribution indicating (a) the frequency of number percentage of internal fragments in every 20% and (b) the frequency of abundance percentage of internal fragments in every 20%. The internal fragment percentage is calculated by the internal fragment metric (number or abundance) divided by the sum of the internal and terminal fragment metric (number or abundance).

3.2. Internal fragments share similar fragmentation propensities with terminal fragments

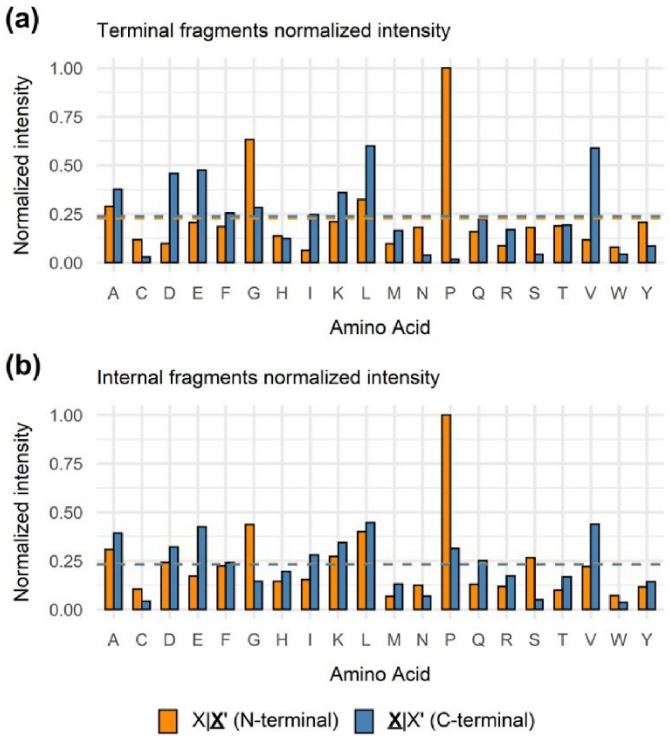


Figure 3. Fragmentation propensity analysis based on normalized abundance for all peptides.

The distribution of normalized abundance adjacent to each amino acid residue of (a) terminal fragments and (b) internal fragments. For each residue, X|**X'** (orange) refers to the cleavage occurring N-terminal to the amino acid residue whereas **X**|X' (blue) refers to the cleavage occurring C-terminal to the amino acid residue. Orange and blue dashed lines indicate average normalized abundance N-terminal and C-terminal to all 20 residues, respectively.

We applied a statistical approach consisting of the dataset utilized in Figure. 2 to compare fragmentation propensities occurring either N- or C-terminal to a specific residue between terminal and internal fragments. The counts of every amino acid residue used in our dataset are listed in Table S2. Fragmentation propensities between two adjacent amino acids can be affected by many factors (*e.g.*, amino acid basicity, secondary structure, precursor charge states), and thus fragmentation events are not evenly distributed across all amino acid residues in CAD experiments. The fragmentation propensity describes the likelihood of cleavages occurring

adjacent to an amino acid residue (Figure. 3) or between a specific residue pair (Figure. 4). For terminal fragments, cleavages N-terminal to proline have a normalized abundance of 1.00, well above the average of 0.23 for all N-terminal fragments that retain the N-terminus (Figure. 3a). In contrast, cleavages C-terminal to proline to generate b- and y-type terminal fragments is highly unlikely, as a particularly low normalized abundance (0.02) is observed (Figure. 3a). This observation of enhanced cleavages N-terminal to proline, termed the “proline effect”, is due to the rigid cyclic structure of the y terminal ion's leaving group (C-terminal cleavages) for proline and has been reported in many studies [24-29]. Similarly, as shown in Figure. 3a, other notable selective cleavages to generate terminal fragments have also been observed in our dataset, *e.g.*, enhanced cleavages C-terminal to aspartic acid and glutamic acid, N-terminal to glycine and tyrosine, and C-terminal to leucine and valine. These preferred fragmentation pathways by CAD have been previously reported in both small and large-scale studies and can be explained by the mobile proton model [21,32,34-38]. Our data agrees well with these well understood fragmentation pathways and demonstrate that sequence strongly impacts the CAD fragmentation propensities.

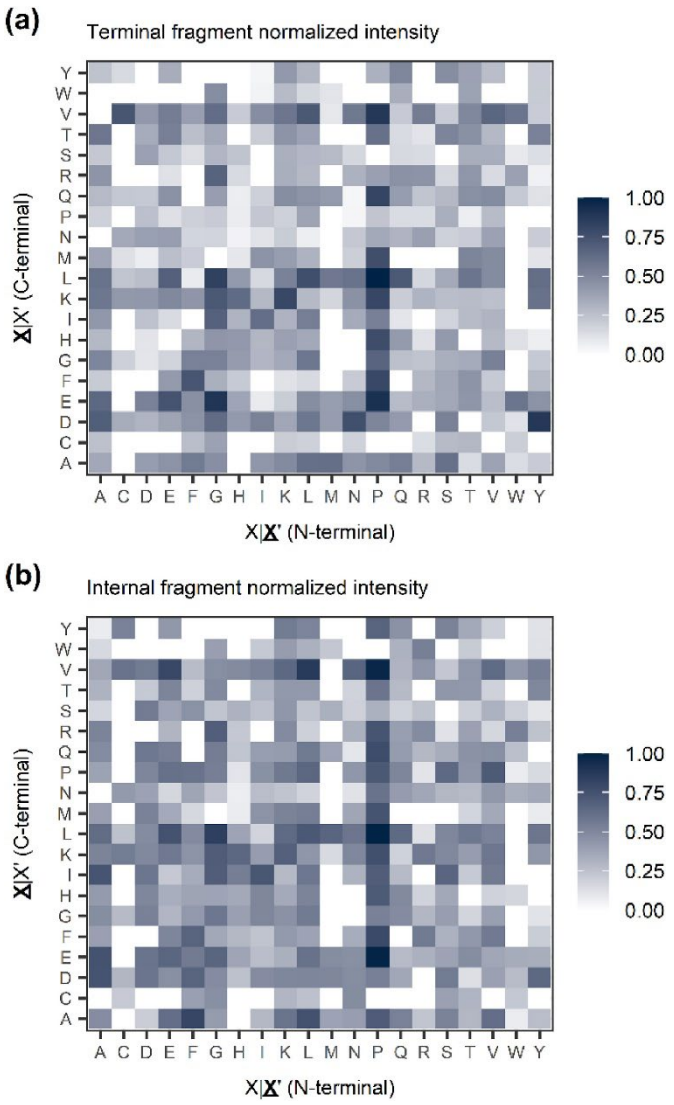


Figure 4. Fragmentation propensity analysis based on normalized abundance deconstructed by residue pair for all peptides.

Heatmaps depicting the normalized abundance deconstructed by residue pair for (a) terminal fragments and (b) internal fragments. For all cells, X|**X'** (x-axis) refers to the cleavage occurring N-terminal to the amino acid residue while **X**|X' (y-axis) refers to the cleavage occurring C-terminal to the amino acid residue. Darker color indicates higher normalized abundance.

The CAD fragmentation propensities to generate *internal* fragments were compared to those for terminal fragments. Similar fragmentation propensities were observed for internal and terminal fragments (Figure. 3a vs. Figure. 3b). For example, cleavages N-terminal to proline

remain the most prominent fragmentation events across all internal fragments (Figure. 3b). Additionally, although to a lesser extent compared to terminal fragments, enhanced cleavages C-terminal to aspartic and glutamic acid residues, N-terminal to glycine, and C-terminal to leucine and valine were observed for internal fragments (Figure. 3b). Overall, the fragmentation propensities for each residue appear to be slightly more evenly distributed for internal fragments, with N- and C-terminal fragments having closer normalized abundances (Figure. 3b). Despite this difference, internal fragments share similar fragmentation propensities with terminal fragments.

To further corroborate this idea, we deconstructed fragmentation propensities by specific amino acid residue pairs [35,36,38] to investigate selective cleavages among the 400 residue combinations (Figure. 4). The prominent proline effect and the enhanced fragmentation C-terminal to valine, leucine, aspartic acid, and glutamic acid for both terminal and internal fragments are featured. In addition to a specific column or row that shows the fragmentation propensity adjacent to a single residue, Figure. 4 also displays the fragmentation propensity *between* a specific residue pair. For both terminal and internal fragments, L|P, E|P, K|P, V|P, L|G, and D|Y are all notable preferred cleavage sites (Figure. 4). The number of cleavages that occur between adjacent amino acid sites are shown in Figs. S2 and S3. Similar selective cleavages between amino acid residue pairs for both terminal and internal fragments can be rationalized as the same ion activation method (CAD) is utilized.

3.3. Internal fragments are generated by subsequent cleavages of terminal fragments

Although internal fragments share similar fragmentation propensities with terminal

fragments, a few dramatic differences between terminal and internal fragments are observed when comparing fragmentation pathways for each residue. Delta normalized abundances (internal fragments minus terminal fragments) for each residue were summarized in Figure. 5. Surprisingly, fragmentation occurring N-terminal to aspartic acid increased by approximately 2.5-fold for internal fragments compared with terminal fragments and increased >10-fold C-terminal to proline for internal fragments compared to terminal fragments (Figure. 5a). These observations are further confirmed by the heatmap shown in Figure. 5b. For example, the red colored proline row indicates that C-terminal cleavages to proline were enhanced for internal fragments compared to terminal fragments, and this increase could largely be explained by cleavages between P|P and P|V pairs (Figure. 5b). In addition, diminished fragmentation occurring C-terminal to aspartic acid, enhanced cleavages N-terminal to leucine and valine, and diminished fragmentation N-terminal to tyrosine, whereas enhanced fragmentation C-terminal to tyrosine (Figure. 5a and b) were all observed for internal fragments compared to terminal fragments.

Enhanced fragmentation N-terminal to proline and tyrosine, C-terminal to aspartic acid, leucine and valine are prominent selective cleavages for terminal fragments (Figure. 3). However, for internal fragments, these enhanced fragmentation events are mostly suppressed (Figure. 5). In contrast, the suppressed cleavages for terminal fragments such as C-terminal to proline and N-terminal to aspartic acid are otherwise enhanced for internal fragments (Figure. 5). To further confirm this observation, the “N-bias” that describes the preference of fragmentation occurring N-terminal to a specific residue was calculated (Figure. S4). The N-bias value of

aspartic acid increases while the N-bias value of proline decreases for internal fragments compared with terminal fragments, which suggests that the amino acid backbone is cleaved preferentially similar to that observed in terminal fragments albeit at a lower propensity. The generation of internal fragments reduces the impact of specific amino acid residues to fragmentation pathways and makes the fragmentation propensities for each residue more evenly distributed. This would require more energy to be distributed along the peptide backbone to overcome certain structural barriers for specific residues, *e.g.*, the bulky cyclic structure of proline to induce more C-terminal cleavages. As a result, it is likely that internal fragments are generated by subsequent cleavage(s) of terminal fragments.

3.4. CAD-generated internal fragments can be explained by the mobile proton model

In the mobile proton model, the probability of protonation on peptide sites depends on the internal energy content of the peptide and gas-phase basicities of different residues of the peptide [23]. In general, peptide ions leaving the electrospray source have protons residing on residues with the largest proton affinities (arginine, histidine, lysine, N-terminal α-amino group) [18]. Energy will be required to move these protons to the peptide backbone to produce a population of ions with protons locating at various amide bond positions, inducing charge-directed fragmentation to generate band y-type terminal fragments [18,23]. If insufficient energy is deposited onto the peptide backbone, no mobile proton will be readily available and selective cleavages will be observed. For example, charge-remote fragmentation pathways that do not require intramolecular proton mobilization to the amide bond can occur. In this case, protons are usually sequestered on arginines and the hydrogen in the side chain of acidic residues will

serve as the proton source to initiate amide bond cleavages. This is typically observed as enhanced fragmentation C-terminal to aspartic acid and glutamic acid, which agrees well with our data for terminal fragments (Figures. 3a and 4a).

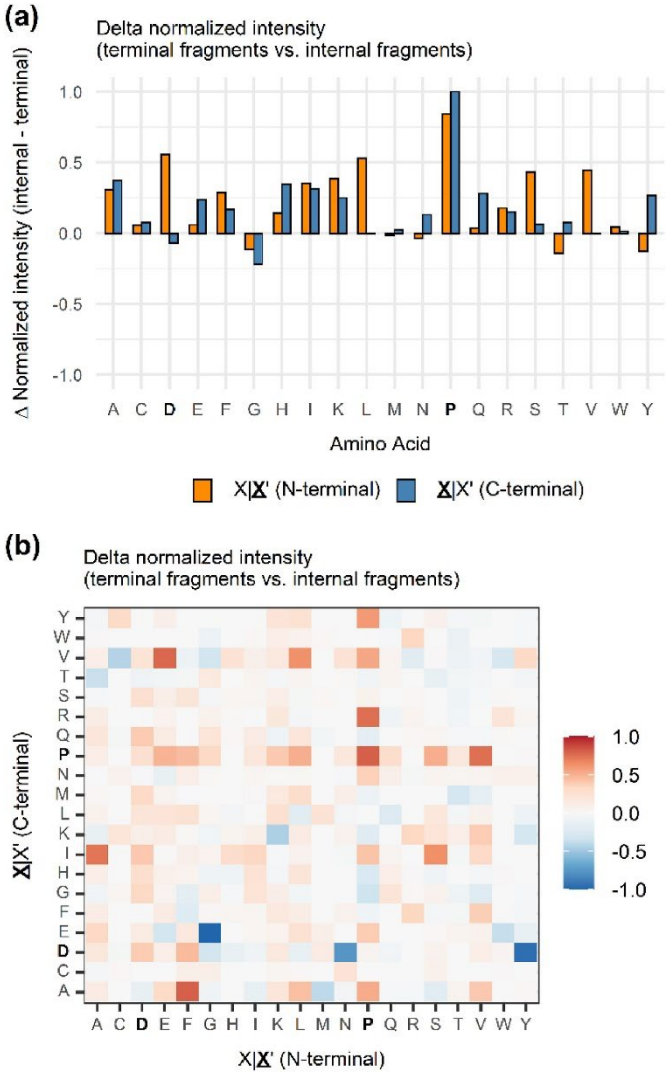


Figure 5. Fragmentation propensity difference analysis between terminal and internal fragments.

(a) Bar graph depicting the difference of normalized abundance between internal and terminal fragments (internal - terminal) for each amino acid residue. (b) Heatmap depicting the difference of normalized abundance between internal and terminal fragments (internal - terminal) deconstructed by residue pair. For both (a) and (b), blue indicates a decrease in fragmentation propensity for internal fragments compared with terminal fragments, while orange/red indicates an increase for internal fragments. For each residue, X|**X'** refers to the cleavage occurring N-terminal to the amino acid residue while **X**|X' refers to the cleavage occurring C-terminal to the

amino acid residue.

To generate a terminal fragment, only a single cleavage event is required; energy deposited onto the peptide backbone moves protons initially residing on basic residues to amide bonds to induce fragmentation. However, this energy accumulation is not sufficient to move protons across all amide bonds to initiate evenly distributed charge-directed fragmentation. This will result in enhanced cleavages observed for terminal fragments (Figures. 3a and 4a). In contrast, multiple cleavage events are required to generate an internal fragment, allowing multiple energy accumulation events to occur that can enhance proton mobility. Therefore, more mobile protons should be available to generate internal fragments than terminal fragments. As a result, charge-remote fragmentation should be suppressed for internal fragments so less enhanced cleavages C-terminal to acidic residues should be observed, which is consistent with our data (Figure. 3a vs. 3b, 4a vs. 4b, and Figure. 5). The suppressed proline effect of internal fragments can also be explained, as well as the enhanced C-terminal cleavages to proline (Figure. 5). Multiple cleavage events result in more energy accessible to the peptide, which can be utilized to overcome the unstable strained 5-5 bicyclic ring in the transition state to initiate more C-terminal fragmentation events to proline. Further evidence to support multiple cleavages is the more evenly distributed fragmentation propensities across all residues for internal fragments compared with terminal fragments (Figure. 3b). More energy deposition results in more mobile protons accessible at various amide bonds; thus, a greater variety of residue pairs can be cleaved by charge-directed fragmentation to generate internal fragments. As a result, for internal

fragments, the preference for specific fragmentation pathways, which are extremely prominent for terminal fragments, are largely diminished. This leads to the more evenly distributed fragmentation propensities across all residues for internal fragments.

3.5. Internal fragments enhance sequence coverage for TD-MS

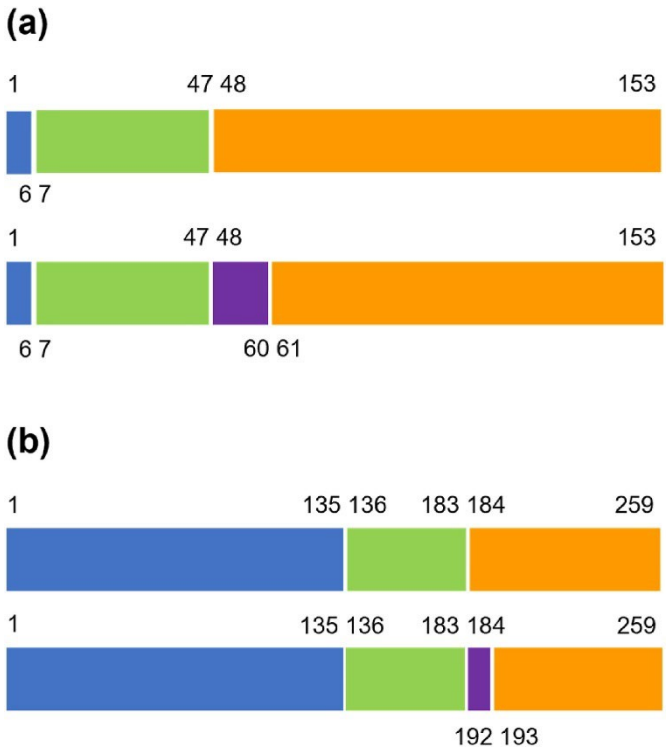


Figure 6. Complementary fragments analysis of apomyoglobin and carbonic anhydrase II. Examples of complementary product ions that include terminal and internal fragments to cover the entire protein sequence for (a) apomyoglobin, $[apoMb + 17H]^{17+}$ and (b) carbonic anhydrase II, $[CAII + 32H]^{32+}$. Numbers above and below each bar indicate the amino acid residue number at the cleavage site and the N- or C-termini. (Blue color indicates N-terminal fragments, orange color indicates C-terminal fragments, and green and purple colors indicate their complementary internal fragments.).

Previous reports by our group [14,43,48,50,51] and by others [39,44-47,49] have established the value of increasing sequence coverage by including internal fragment assignments. However, what has not been extensively discussed to date is the extent for which

information has not been considered in the past in TD-MS experiments. For example, CAD of the 17+ charge state of apoMb (16.9 kDa) yielded 492 peaks in the deconvoluted mass spectrum; of the 492 peaks, 74 were assigned as unique terminal fragments, or an assigned peak percentage (APP) of 15.0% (and yielding a sequence coverage of 46.1%). However, by including the 174 peaks assigned as unique internal fragments, the APP increases to 50.4% (and 80.3% sequence coverage). Similarly, for the CAD mass spectrum of the 32+ charge state of CAII (29 kDa), 55 of the 349 total deconvoluted peaks were assigned as terminal fragments, or an APP of 15.8% (22.1% sequence coverage). Including the 121 peaks assigned as internal fragments increases the APP to 50.4% (50.8% sequence coverage). A large fraction of the remaining ca. 50% of the unassigned peaks are likely due to neutral losses (*e.g.*, loss of H_2O, NH_3, etc.) common to polypeptide MS/MS experiments. Work is ongoing to accurately account for these unassigned peaks.

In principle, the presence of internal fragments may enhance the protein identification process in TD proteomics. Since the early days of tandem MS of multiply charged polypeptides, complementary ion pairs, which in sum account for the entire precursor molecule, have been observed in TD-MS of proteins [26,53,54]; later, complementary ion pairs have been observed in native TD-MS of protein complexes [55]. Complementary ion pairs often result from cleavage of the N-terminal bond to a proline residue, but it can be found from fragmentation of other residues. The inclusion of internal fragments, that when combined with terminal fragments span the entire polypeptide sequence, can result in complementary product ions. For example, CAD of $[\text{apoMb} + 17\text{H}]^{17+}$ yields terminal fragments b_6 (1+) and y_{106} (10+); when joined by internal

fragment $by_{7\text{-}47}$ (4+, 5+), these 3 product ions form complementary product ions that cover the entire sequence (Figure. 6a). Similarly, CAD of $[CAII + 32H]^{32+}$ generates b_{135} (15+ - 21+) and y_{76} (7+ - 9+) terminal fragments and the $by_{136\text{-}183}$ (4+ - 7+) internal fragment that can be combined to form complementary product ions (Figure. 6b). (Tables S3 and S4 list many examples of complementary product ions for CAD of apoMb and CAII.) Nielsen et al. suggested the inclusion of complementary ion pairs to improve the protein identification process for bottom-up proteomics [56]. We posit that complementary product ions that include internal fragments could improve the protein identification for top-down proteomics.

4. Conclusions

Here we report the first extensive investigation of fragmentation propensities of internal fragments generated by CAD of peptides and small proteins. We demonstrate that although internal fragments only account for a small portion of mass spectral signals in a single spectrum, they can explain a large number of fragments generated overall. Therefore, many previously unassigned signals can be explained as internal fragments and provide additional sequence information to enhance peptide and protein sequence coverage. By applying a statistical approach, we have shown that internal fragments share similar fragmentation propensities with terminal fragments as similar selective cleavages are observed. Importantly, the abundance of terminal fragments and their sequentially generated internal fragments agree well with each other. This suggests that internal fragments generated by CAD follow the same fragmentation pattern as terminal fragments and can be explained by the mobile proton model. However,

these enhanced cleavages are slightly suppressed, causing a more evenly distributed fragmentation propensities across all residues for internal fragments compared with terminal fragments. This is likely due to more mobile protons readily available to generate internal fragments, providing evidence that internal fragments are generated by subsequent cleavages of terminal fragments.

The gas-phase fragmentation propensity of internal fragments presented here improves our understanding of the formation of internal fragments. This knowledge, along with the assignment of complementary product ions that account for the total polypeptide sequence, could be beneficial for the development of sequencing algorithms to assign internal fragments more accurately and reliably, as well as providing a new strategy for protein identification and validation in top-down proteomics. By assigning internal fragments, it is possible to gain more insight into protein sequence, leading to more efficient TD-MS analysis of proteins and proteoforms. Notably, the results presented here can be expanded to bottom-up and middle-down MS experiments, potentially improving the confidence and efficiency of protein identification in these MS techniques as well.

CRediT authorship contribution statement

Benqian Wei: Investigation, Data acquisition, Formal analysis, Writing - original draft, Writing - review & editing. **Muhammad A. Zenaidee:** Investigation, Formal analysis, Writing - review & editing. **Carter Lantz:** Investigation, Formal analysis, Writing - review& editing. **Rachel R. Ogorzalek Loo:** Supervision, Resources, Writing - review & editing. **Joseph A. Loo:**

Supervision, Funding acquisition, Resources, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Support from the US National Institutes of Health (R01GM103479, S10RR028893), the US National Science Foundation (NSF) (CHE1808492), and the US Department of Energy (DEFC02-02ER63421) are gratefully acknowledged. C. L. acknowledges support from the Ruth L. Kirschstein National Research Service Award program (GM007185).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2021.339400.

Chapter 2: Supporting Information

Supplementary Figures

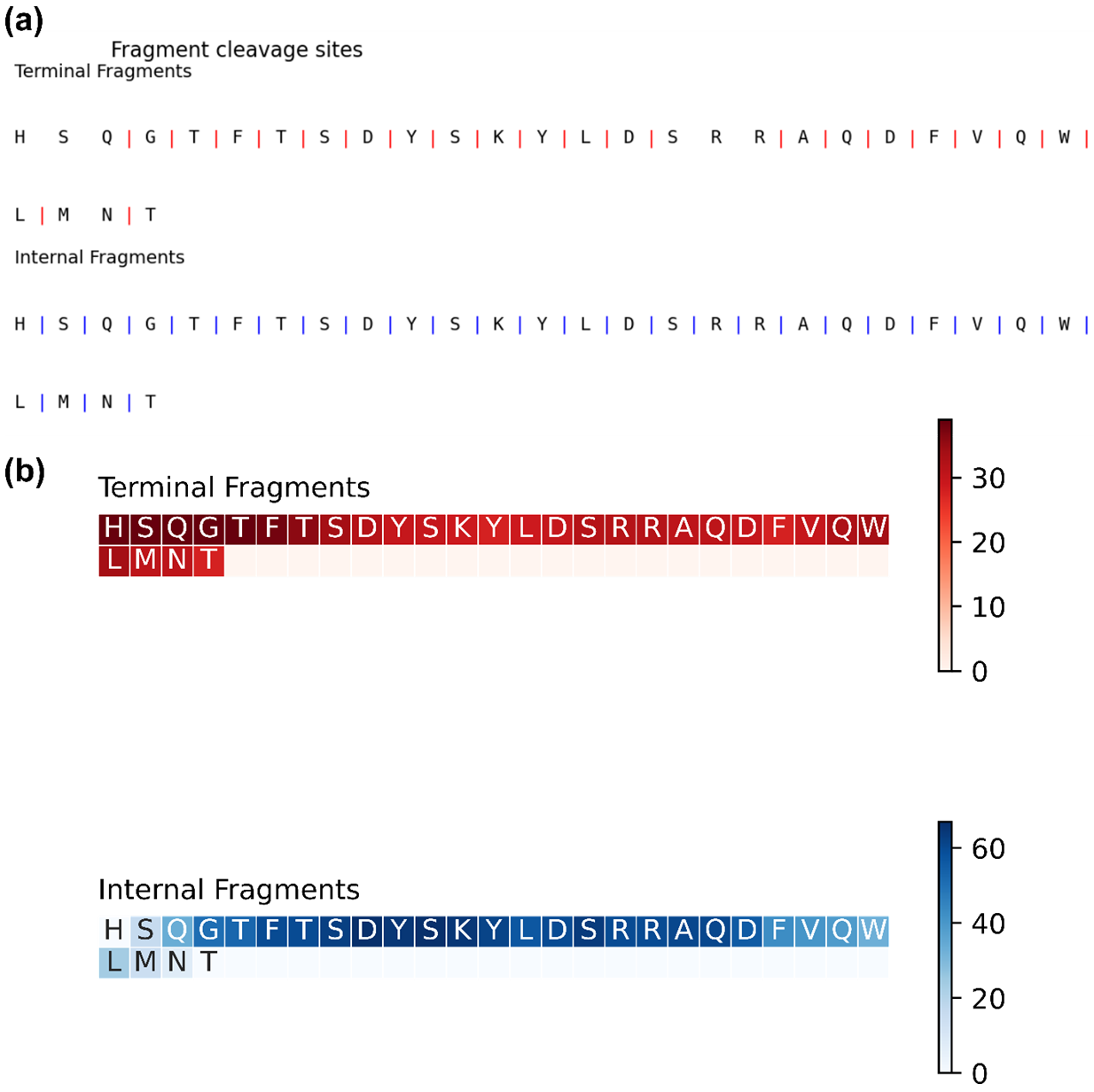


Figure S1. Additional sequence coverage analysis of glucagon.
For CAD MS/MS of [glucagon + 3H]$^{3+}$, (a) A fragment cleavage map indicating the location of inter-amino acid cleavage sites for terminal and internal fragments. (b) A sequence coverage map for terminal and internal fragments. Darker regions indicate more fragments covering those amino acid residues.

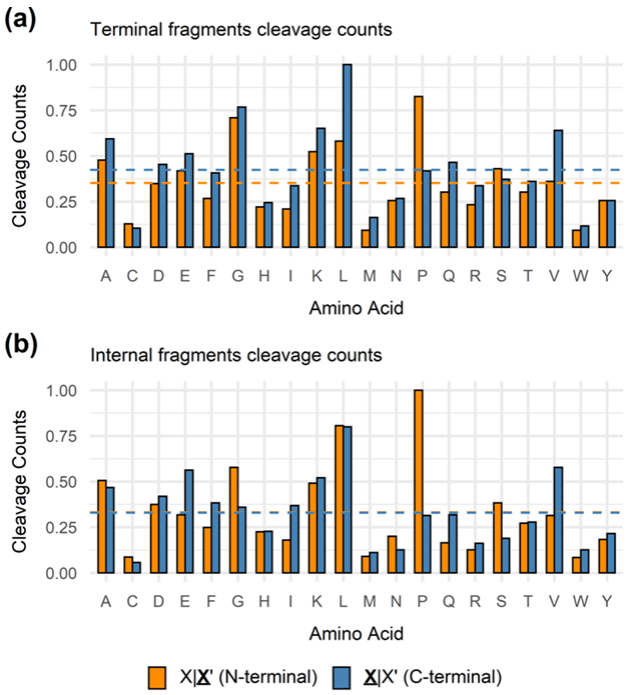


Figure S2. Fragmentation propensity analysis based on number of cleavages for all peptides.

The distribution of the number of cleavage events adjacent to each amino acid residue of (a) terminal fragments and (b) internal fragments. For each residue, X|**X'** (orange) refers to the cleavage occurring N-terminal to the amino acid residue while **X**|X' (blue) refers to the cleavage occurring C-terminal to the amino acid residue. Orange and blue dashed lines indicate average cleavage counts N-terminal or C-terminal to all 20 residues.

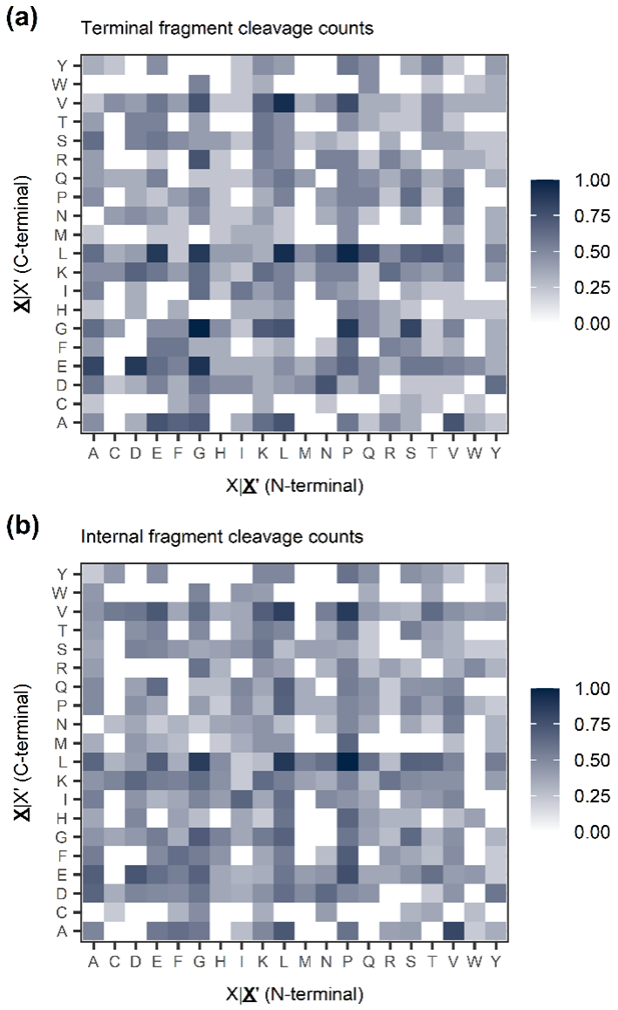


Figure S3. Fragmentation propensity analysis based on normalized abundance deconstructed by residue pair for all peptides.

A heatmap depicting the number of cleavage events deconstructed by residue pair for (a) terminal fragments and (b) internal fragments. For each cell, X|**X'** (x-axis) refers to the cleavage occurring N-terminal to the amino acid residue while **X**|X' (y-axis) refers to the cleavage occurring C-terminal to the amino acid residue.

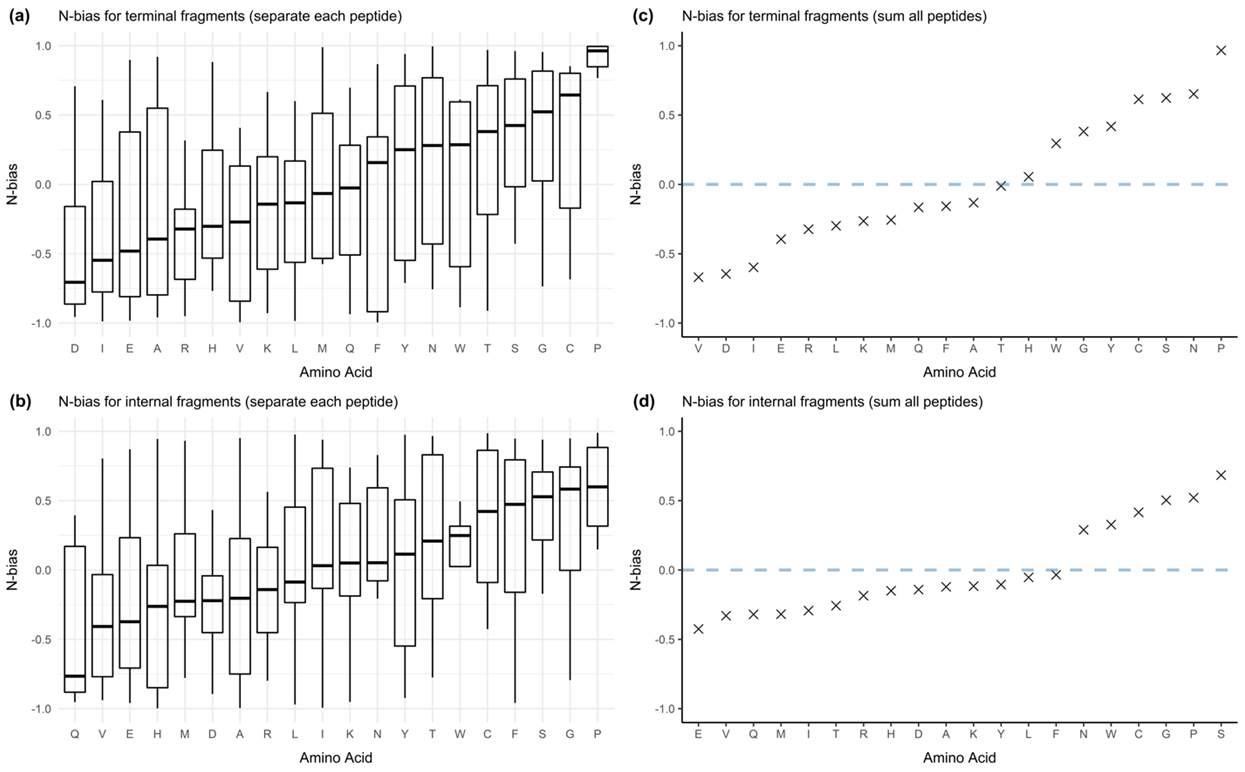


Figure S4. "N-bias" analysis for all peptides.

N-bias for all amino acids calculated for each peptide separately for (a) terminal fragments and (b) internal fragments. The boxplot shows variance for different peptides within each residue. The black line within each box indicates the median value of that residue. N-bias for all amino acids after summing normalized abundance of all peptides for (c) terminal fragments and (d) internal fragments. N-bias is calculated by the difference between N-terminal normalized abundance and C-terminal normalized abundance divided by sum of the two abundances (see Experimental Section).

Supplementary Tables

Table S1. Molecular weight, isoelectric point [215], precursor charge states and number of assigned terminal and internal fragments for each peptide and protein analyzed.

Peptide	Molecular Weight (Da)	Isoelectric point	Precursor charge state	Number of assigned terminal fragments	Number of assigned internal fragments
Apomyoglobin (137-148)	1467.69	8.5	2	19	16
Glu-fib	1570.68	4	2	20	17
UOM-6 peptide	1574.84	6.23	3	12	7
Alpha casein (126-141)	1778.89	5.4	2	19	22
LARL peptide	2014.3	4.68	3	20	16
Apomyoglobin (1-18)	2044.25	4.14	2	23	46
Tummino peptide	2068.39	7.94	3	28	41
ACTH (1-17)	2093.41	10.45	3, 4	28	41
Apo-myoglobin (86-105)	2315.79	9.83	3	27	59
ACTH (18-39)	2465.7	4.25	3	28	36
Beta casein (101-121)	2474.92	6.81	3	27	49
Insulin Chain A	2531.64	5.66	2	29	30
Fibronectin Type III	2732.05	4.3	3, 4	26	31
Carbonic anhydrase (237-259)	2734.31	12.31	3	18	25
Melittin	2846.76	12.02	4	31	24
Gastrin releasing peptide	2859.4	9.99	3, 4	31	69
3X FLAG	2861.91	4.16	3, 4	28	53
Xenin	2972.6	11.17	3, 4	34	67
Tau peptide (45-73)	2977.97	3.67	3	42	56
Peripheral Myelin protein P2 (53-78)	3019.44	4.42	4	26	30
C-peptide (3-33)	3020.26	3.45	3, 4	40	39
CGRP (8-37)	3125.6	12.02	3, 4	31	35
GALP	3166.6	10.74	3, 4	44	30
Tau peptide (275-305)	3264.75	9.52	3, 4	48	78
Synthetic 1	3271.88	10.65	4, 5	33	28
Apomyoglobin (106-136)	3277.7	5.74	4	36	65
VIP	3325.9	9.82	3, 4	31	49
Apomyoglobin (55-85)	3326.91	9.3	4	47	65
Proinsulin C-peptide	3340.7	6.29	3, 4	49	40
OVA (241-270)	3421.9	3.83	3	46	49
Glucagon	3482.79	6.75	3	36	58

Insulin Chain B	3495.89	6.9	3, 4	38	50
ACTH (7-38)	3659.17	9.4	3, 4, 5	35	25
Alpha casein (158-192)	3793.72	3.43	3	39	25
Apelin-36	4195.87	12.85	6	36	12
Neuropeptide Y	4272.72	6.76	5, 6	32	25
Beta-amyloid (1-42)	4514.08	5.31	5	31	18
Anti-BetaGamma	4514.1	5.31	4	46	167
Beta casein (45-91)	5202.04	5.32	4	43	49
Alpha-synuclein (96-140)	5217.56	3.76	4	50	61
Ubiquitin	8565.91	6.56	6	35	53
Beta casein (132-209)	8791.51	9.52	6	70	105

Table S2. Counts of each amino acid residue analyzed for all peptides and proteins.

Amino acid residue	Counts
A	76
C	12
D	59
E	93
F	45
G	104
H	40
I	49
K	84
L	104
M	26
N	41
P	98
Q	62
R	54
S	66
T	48
V	73
W	15
Y	39

Table S3. All complementary product ions covering the entire protein sequence including one or two internal fragments for CAD of apomyoglobin, $[apoMb + 17H]^{17+}$. Observed charge states of every fragment ion are shown in the paratheses.

Apomyoglobin		
N-terminal fragment	Internal fragment	C-terminal fragment
b6 (1+)	by7-47 (4+, 5+)	y106 (10+)
b6 (1+)	by7-13 (1+), by14-20 (1+)	y133 (14+, 15+)
b6 (1+)	by7-47 (4+, 5+), by48-54 (1+)	y99 (11+)
b6 (1+)	by7-47 (4+, 5+), by48-60 (1+, 2+)	y93 (9+, 10+, 11+)
b7 (1+)	by8-20 (1+)	y133 (14+, 15+)
b7 (1+)	by8-20 (1+), by21-36 (2+)	y117 (13+)
b7 (1+)	by8-20 (1+), by21-47 (4+)	y106 (10+)
b8 (1+)	by9-20 (2+)	y133 (14+, 15+)
b8 (1+)	by9-20 (2+), by21-36 (2+)	y117 (13+)
b8 (1+)	by9-20 (2+), by21-47 (4+)	y106 (10+)
b8 (1+)	by9-60 (5+), by61-72 (2+)	y81 (9+, 10+)
b8 (1+)	by9-60 (5+), by61-83 (3+)	y70 (6+, 7+, 8+)
b13 (1+, 2+)	by14-20 (1+)	y133 (14+, 15+)
b13 (1+, 2+)	by14-19 (1+), bx20-47 (2+)	y106 (10+)
b13 (1+, 2+)	by14-20 (1+), by21-36 (2+)	y117 (13+)
b13 (1+, 2+)	by14-20 (1+), by21-47 (4+)	y106 (10+)
b14 (1+)	by15-20 (1+)	y133 (14+, 15+)
b14 (1+)	by15-22 (1+)	y131 (13+, 14+)
b14 (1+)	by15-20 (1+), by21-36 (2+)	y117 (13+)
b14 (1+)	by15-20 (1+), by21-47 (4+)	y106 (10+)
b14 (1+)	by15-22 (1+), by23-47 (2+)	y106 (10+)
b14 (1+)	by15-22 (1+), by23-54 (3+)	y99 (11+)
b20 (1+)	by21-36 (2+)	y117 (13+)
b20 (1+)	by21-47 (4+)	y106 (10+)
b20 (1+)	by21-36 (2+), by37-47 (2+)	y106 (10+)
b20 (1+)	by21-36 (2+), by37-60 (2+, 3+)	y93 (9+, 10+, 11+)
b20 (1+)	by21-47 (4+), by48-54 (1+)	y99 (11+)
b20 (1+)	by21-47 (4+), by48-60 (1+, 2+)	y93 (9+, 10+, 11+)
b36 (3+)	by37-47 (2+)	y106 (10+, 12+)
b36 (3+)	by37-60 (2+, 3+)	y93 (9+, 10+, 11+)
b36 (3+)	by37-47 (2+), by48-54 (1+)	y99 (11+)
b36 (3+)	by37-60 (2+, 3+), by61-72 (2+)	y81 (9+, 10+)
b36 (3+)	by37-60 (2+, 3+), by61-83 (3+)	y70 (6+, 7+, 8+)
a55 (5+)	bx56-148 (10+)	y5 (1+)

Table S4. All complementary product ions covering the entire protein sequence including one or two internal fragments for CAD of carbonic anhydrase II, $[CAII + 32H]^{32+}$. Observed charge states of every fragment ion are shown in the paratheses.

Carbonic anhydrase II

N-terminal fragment	Internal fragment	C-terminal fragment
b135 (15+ - 21+)	by136-183 (4+ - 7+)	y76 (7+ - 9+)
b135 (15+ - 21+)	by136-186 (5+, 6+)	y73 (7+, 8+)
b135 (15+ - 21+)	by136-191 (5+)	y68 (6+, 7+)
b135 (15+ - 21+)	by136-198 (5+)	y61 (5+ - 8+)
b135 (15+ - 21+)	by136-182 (5+), by183-195 (1+)	y64 (6+)
b135 (15+ - 21+)	by136-183 (4+ - 7+), by184-189 (1+)	y70 (6+, 7+)
b135 (15+ - 21+)	by136-183 (4+ - 7+), by184-190 (1+)	y69 (6+, 7+)
b135 (15+ - 21+)	by136-183 (4+ - 7+), by184-191 (2+)	y68 (6+, 7+)
b135 (15+ - 21+)	by136-183 (4+ - 7+), by184-192 (1+, 2+)	y67 (5+ - 8+)
b135 (15+ - 21+)	by136-183 (4+ - 7+), by184-194 (1+, 2+)	y65 (6+)
b135 (15+ - 21+)	by136-183 (4+ - 7+), by184-195 (1+)	y64 (6+)
b135 (15+ - 21+)	by136-183 (4+ - 7+), by184-234 (4+)	y25 (3+ - 5+)
b135 (15+ - 21+)	by136-191 (5+), by192-196 (1+)	y63 (5+ - 8+)
b135 (15+ - 21+)	by136-198 (5+), by199-231 (3+)	y28 (4+, 5+)
b138 (18+)	ay139-195 (4+ - 6+)	y64 (6+)
b138 (18+)	by139-182 (4+ - 6+), by183-195 (1+)	y64 (6+)
b140 (17+)	by141-183 (4+ - 6+)	y76 (7+ - 9+)
b140 (17+)	by141-178 (6+), by179-186 (1+)	y73 (7+, 8+)
b140 (17+)	by141-178 (6+), by179-189 (1+)	y70 (6+, 7+)
b140 (17+)	by141-178 (6+), by179-190 (2+)	y69 (6+, 7+)
b140 (17+)	by141-178 (6+), by179-191 (1+, 2+)	y68 (6+, 7+)
b140 (17+)	by141-183 (4+ - 6+), by184-189 (1+)	y70 (6+, 7+)
b140 (17+)	by141-183 (4+ - 6+), by184-190 (1+)	y69 (6+, 7+)
b140 (17+)	by141-183 (4+ - 6+), by184-191 (2+)	y68 (6+, 7+)
b140 (17+)	by141-183 (4+ - 6+), by184-192 (1+, 2+)	y67 (5+ - 8+)
b140 (17+)	by141-183 (4+ - 6+), by184-194 (1+, 2+)	y65 (6+)
b140 (17+)	by141-183 (4+ - 6+), by184-195 (1+)	y64 (6+)
b140 (17+)	by141-183 (4+ - 6+), by184-234 (4+)	y25 (3+ - 5+)
b178 (19+)	by179-186 (1+)	y73 (7+, 8+)
b178 (19+)	by179-189 (1+)	y70 (6+, 7+)
b178 (19+)	by179-190 (2+)	y69 (6+, 7+)
b178 (19+)	by179-191 (1+, 2+)	y68 (6+, 7+)
b178 (19+)	by179-187 (2+), by188-192 (1+)	y67 (5+ - 8+)
b178 (19+)	by179-187 (2+), by188-198 (1+)	y61 (5+ - 8+)

b178 (19+)	by179-189 (1+), by190-194 (1+)	y65 (6+)
b178 (19+)	by179-189 (1+), by190-195 (1+)	y64 (6+)
b178 (19+)	by179-189 (1+), by190-234 (4+)	y25 (3+ - 5+)
b178 (19+)	by179-190 (2+), by191-196 (1+)	y63 (5+ - 8+)
b178 (19+)	by179-190 (2+), by191-197 (1+)	y62 (6+, 7+)
b178 (19+)	by179-191, by192-196 (1+)	y63 (5+ - 8+)
b183 (19+ - 25+)	by184-189 (1+)	y70 (6+, 7+)
b183 (19+ - 25+)	by184-190 (1+, 2+)	y69 (6+, 7+)
b183 (19+ - 25+)	by184-191 (2+)	y68 (6+, 7+)
b183 (19+ - 25+)	by184-192 (1+, 2+)	y67 (5+ - 8+)
b183 (19+ - 25+)	by184-194 (1+, 2+)	y65 (6+)
b183 (19+ - 25+)	by184-195 (1+)	y64 (6+)
b183 (19+ - 25+)	by184-234 (4+)	y25 (3+ - 5+)
b183 (19+ - 25+)	by184-189 (1+), by190-194 (1+)	y65 (6+)
b183 (19+ - 25+)	by184-189 (1+), by190-195 (1+)	y64 (6+)
b183 (19+ - 25+)	by184-189 (1+), by190-234 (4+)	y25 (3+ - 5+)
b183 (19+ - 25+)	by184-190 (1+, 2+), by191-196 (1+)	y63 (5+ - 8+)
b183 (19+ - 25+)	by184-190 (1+, 2+), by191-197 (1+)	y62 (6+, 7+)
b183 (19+ - 25+)	by184-191 (2+), by192-196 (1+)	y63 (5+ - 8+)
b183 (19+ - 25+)	by184-192 (1+, 2+), by193-201 (1+)	y58 (7+)
b183 (19+ - 25+)	by184-192 (1+, 2+), by193-230 (4+)	y29 (5+)
b183 (19+ - 25+)	by184-192 (1+, 2+), by193-231 (5+)	y28 (4+, 5+)
b183 (19+ - 25+)	by184-192 (1+, 2+), by193-232 (4+, 5+)	y27 (3+ - 5+)
b183 (19+ - 25+)	by184-192 (1+, 2+), by193-234 (3+, 4+)	y25 (3+ - 5+)

References

[1] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server, in: J.M. Walker (Ed.), The Proteomics Protocols Handbook, Springer, Humana Press Inc., Totowa, NJ, 2005, pp. 571-607.

References

[1] S.E. Ong, M. Mann, Mass spectrometry-based proteomics turns quantitative, Nat. Chem. Biol. 1 (5) (2005) 252–262.

[2] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels, Anal. Chem. 68 (5) (1996) 850–858.

[3] A.D. Catherman, O.S. Skinner, N.L. Kelleher, Top down proteomics: facts and perspectives, Biochem. Biophys. Res. Commun. 445 (4) (2014) 683–693.

[4] N.L. Kelleher, H.Y. Lin, G.A. Valaskovic, D.J. Aaserud, E.K. Fridriksson, F.W. McLafferty, Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry, J. Am. Chem. Soc. 121 (4) (1999) 806–812.

[5] F. Lermyte, Y.O. Tsybin, P.B. O'Connor, J.A. Loo, Top or middle? Up or down? Toward a standard lexicon for protein top-down and allied mass spectrometry approaches, J. Am. Soc. Mass Spectrom. 30 (7) (2019) 1149–1157.

[6] T.K. Toby, L. Fornelli, N.L. Kelleher, Progress in top-down proteomics and the analysis of proteoforms, Annu. Rev. Anal. Chem. 9 (1) (2016) 499–519.

[7] L.A. Macias, I.C. Santos, J.S. Brodbelt, Ion activation methods for peptides and proteins, Anal. Chem. 92 (1) (2020) 227–251.

[8] L. Sleno, D.A. Volmer, Ion activation methods for tandem mass spectrometry, J. Mass Spectrom. 39 (10) (2004) 1091–1112.

[9] S.A. Mcluckey, Principles of collisional activation in analytical mass spectrometry, J. Am. Soc. Mass Spectrom. 3 (6) (1992) 599–614.

[10] R.A. Zubarev, D.M. Horn, E.K. Fridriksson, N.L. Kelleher, N.A. Kruger, M.A. Lewis, B.K. Carpenter, F.W. McLafferty, Electron capture dissociation for structural characterization of multiply charged protein cations, Anal. Chem. 72 (3) (2000) 563─573.

[11] R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, Electron capture dissociation of multiply charged protein cations. A nonergodic process, J. Am. Chem. Soc. 120 (13) (1998) 3265─3266.

[12] J.E.P. Syka, J.J. Coon, M.J. Schroeder, J. Shabanowitz, D.F. Hunt, Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Proc. Natl. Acad. Sci. U.S.A. 101 (26) (2004) 9528─9533.

[13] Y.M.E. Fung, C.M. Adams, R.A. Zubarev, Electron ionization dissociation of singly and multiply charged peptides, J. Am. Chem. Soc. 131 (2009) 9977─9985.

[14] H. Li, Y. Sheng, W. McGee, M. Cammarata, D. Holden, J.A. Loo, Structural characterization of native proteins and protein complexes by electron ionization dissociation-mass spectrometry, Anal. Chem. 89 (2017) 2731─2738.

[15] J.S. Brodbelt, Photodissociation mass spectrometry: new tools for characterization of biological molecules, Chem. Soc. Rev. 43 (8) (2014) 2757─2783.

[16] J.P. Reilly, Ultraviolet photofragmentation of biomolecular ions, Mass Spectrom, Rev. 28 (3) (2009) 425─447.

[17] J.S. Brodbelt, Ion activation methods for peptides and proteins, Anal. Chem. 88 (1) (2016) 30─51.

[18] A.R. Dongre, J.L. Jones, A. Somogyi, V.H. Wysocki, Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: evidence for the mobile

proton model, J. Am. Chem. Soc. 118 (35) (1996) 8365─8374.

[19] B. Paizs, S. Suhai, Fragmentation pathways of protonated peptides, Mass Spectrom, Rev. 24 (4) (2005) 508─548.

[20] V.H. Wysocki, G. Tsaprailis, L.L. Smith, L.A. Breci, Mobile and localized protons: a framework for understanding peptide dissociation, J. Mass Spectrom. 35 (12) (2000) 1399─1406.

[21] C. Gu, A. Somogyi, V.H. Wysocki, K.F. Medzihradszky, Fragmentation of protonated oligopeptides XLDVLQ (X=L, H, K or R) by surface induced dissociation: additional evidence for the ‘mobile proton’ model, Anal. Chim. Acta 397 (1─3) (1999) 247─256.

[22] S.G. Summerfield, S.J. Gaskell, Fragmentation efficiencies of peptide ions following low energy collisional activation, Int. J. Mass Spectrom. Ion Process. 165─166 (1997) 509─521.

[23] G. Tsaprailis, A. Somogyi, E.N. Nikolaev, V.H. Wysocki, Refining the model for selective cleavage at acidic residues in arginine-containing protonated peptides, Int. J. Mass Spectrom. 195─196 (2000) 467─479.

[24] L.A. Breci, D.L. Tabb, J.R. Yates, V.H. Wysocki, Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra, Anal. Chem. 75 (9) (2003) 1963─1971.

[25] D.F. Hunt, J.R. Yates, J. Shabanowitz, S. Winston, C.R. Hauer, Protein sequencing by tandem mass spectrometry, Proc. Natl. Acad. Sci. U.S.A. 83 (17) (1986) 6233─6237.

[26] J.A. Loo, C.G. Edmonds, R.D. Smith, Tandem mass spectrometry of very large molecules. 2. Dissociation of multiply charged proline-containing proteins from electrospray ionization, Anal. Chem. 65 (4) (1993) 425─438.

[27] X.J. Tang, P. Thibault, R.K. Boyd, Fragmentation reactions of multiply protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation, Anal. Chem. 65 (20) (1993) 2824─2834.

[28] T. Vaisar, J. Urban, Probing the proline effect in CID of protonated peptides, J. Mass Spectrom. 31 (10) (1996) 1185─1187.

[29] C. Bleiholder, S. Suhai, A.G. Harrison, B. Paizs, Towards understanding the tandem mass spectra of protonated oligopeptides. 2: the proline effect in collision-induced dissociation of protonated Ala-Ala-Xxx-Pro-Ala (Xxx = Ala, Ser, Leu, Val, Phe, and Trp), J. Am. Soc. Mass Spectrom. 22 (6) (2011) 1032─1039.

[30] Y. Huang, V.H. Wysocki, D.L. Tabb, J.R. Yates, The influence of histidine on cleavage C-terminal to acidic residues in doubly protonated tryptic peptides, Int. J. Mass Spectrom. 219 (1) (2002) 233─244.

[31] G. Tsaprailis, H. Nair, W. Zhong, K. Kuppannan, J.H. Futrell, V.H. Wysocki, A mechanistic investigation of the enhanced cleavage at histidine in the gas-phase dissociation of protonated peptides, Anal. Chem. 76 (7) (2004) 2083─2094.

[32] C. Gu, G. Tsaprailis, L. Breci, V.H. Wysocki, Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of asp-containing peptides, Anal. Chem. 72 (23) (2000) 5804─5813.

[33] D.L. Tabb, Y. Huang, V.H. Wysocki, J.R. Yates, Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides, Anal. Chem. 76 (5) (2004) 1243─1248.

[34] G. Tsaprailis, H. Nair, A. Somogyi, V.H. Wysocki, W. Zhong, J.H. Futrell, S.G. Summerfield, S.J. Gaskell, Influence of secondary structure on the fragmentation of protonated peptides, J. Am. Chem. Soc. 121 (22) (1999) 5142─5154.

[35] Y. Huang, J.M. Triscari, G.C. Tseng, L. Pasa-Tolic, M.S. Lipton, R.D. Smith, V.H. Wysocki, Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns, Anal. Chem. 77 (18) (2005) 5800─5813.

[36] Y. Huang, G.C. Tseng, S. Yuan, L. Pasa-Tolic, M.S. Lipton, R.D. Smith, V.H. Wysocki, A data-mining scheme for identifying peptide structural motifs responsible for different MS/MS fragmentation intensity patterns, J. Proteome Res. 7 (1) (2008) 70─79.

[37] D.L. Tabb, L.L. Smith, L.A. Breci, V.H. Wysocki, D. Lin, J.R. Yates, Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides, Anal. Chem. 75 (5) (2003) 1155─1163.

[38] N.A. Haverland, O.S. Skinner, R.T. Fellers, A.A. Tariq, B.P. Early, R.D. LeDuc, L. Fornelli, P.D. Compton, N.L. Kelleher, Defining gas-phase fragmentation propensities of intact proteins during native top-down mass spectrometry, J. Am. Soc. Mass Spectrom. 28 (6) (2017) 1203─1215.

[39] K.R. Durbin, O.S. Skinner, R.T. Fellers, N.L. Kelleher, Analyzing internal fragmentation of electrosprayed ubiquitin ions during beam-type collisional dissociation, J. Am. Soc. Mass Spectrom. 26 (5) (2015) 782─787.

[40] A. Michalski, N. Neuhauser, J. Cox, M. Mann, A systematic investigation into the nature of tryptic HCD spectra, J. Proteome Res. 11 (11) (2012) 5479─5491.

[41] N.F. Zinnel, P.-J. Pai, D.H. Russell, Ion mobility-mass spectrometry (IM-MS) for top-down proteomics: increased dynamic range affords increased sequence coverage, Anal. Chem. 84 (7) (2012) 3390─3397.

[42] Y.A. Lyon, D. Riggs, L. Fornelli, P.D. Compton, R.R. Julian, The ups and downs of repeated cleavage and internal fragment production in top-down proteomics, J. Am. Soc. Mass Spectrom. 29 (1) (2018) 150─157.

[43] M.A. Zenaidee, C. Lantz, T. Perkins, W. Jung, R.R. Ogorzalek Loo, J.A. Loo, Internal fragments generated by electron ionization dissociation enhance protein top-down mass spectrometry, J. Am. Soc. Mass Spectrom. 31 (9) (2020), 1896─1902.

[44] K.D. Ballard, S.J. Gaskell, Sequential mass spectrometry applied to the study of the formation of “internal” fragment ions of protonated peptides, Int. J. Mass Spectrom. Ion Process. 111 (1991) 173─189.

[45] P.E. Barran, N.C. Polfer, D.J. Campopiano, D.J. Clarke, P.R.R. Langridge-Smith, R.J. Langley, J.R.W. Govan, A. Maxwell, J.R. Dorin, R.P. Millar, M.T. Bowers, Is it biologically relevant to measure the structures of small peptides in the gas-phase? Int. J. Mass Spectrom. 240 (3) (2005) 273─284.

[46] J. Chen, P. Shiyanov, K.B. Green, Top-down mass spectrometry of intact phosphorylated b-casein: correlation between the precursor charge state and internal fragments, J. Mass Spectrom. 54 (6) (2019) 527─539.

[47] J.S. Cobb, M.L. Easterling, J.N. Agar, Structural characterization of intact proteins is enhanced by prevalent fragmentation pathways rarely observed for peptides, J. Am. Soc. Mass

Spectrom. 21 (6) (2010) 949–959.

[48] M.A. Zenaidee, B. Wei, C. Lantz, H.T. Wu, T.R. Lambeth, J.K. Diedrich, R.R. Ogorzalek Loo, R.R. Julian, J.A. Loo, Internal fragments generated from different top-down mass spectrometry fragmentation methods extend protein sequence coverage, J. Am. Soc. Mass Spectrom. 32 (7) (2021) 1752–1758.

[49] N.D. Schmitt, J.M. Berger, J.B. Conway, J.N. Agar, Increasing top-down mass spectrometry sequence coverage by an order of magnitude through optimized internal fragment generation and assignment, Anal. Chem. 93 (16) (2021) 6355–6362.

[50] H. Li, H.H. Nguyen, R.R. Ogorzalek Loo, I.D.G. Campuzano, J.A. Loo, An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes, Nat. Chem. 10 (2) (2018) 139–148.

[51] C. Lantz, M.A. Zenaidee, B. Wei, Z. Hemminger, R.R. Ogorzalek Loo, J.A. Loo, Clips MS: an algorithm for analyzing internal fragments resulting from topdown mass spectrometry, J. Proteome Res. 20 (4) (2021), 1928–1935.

[52] M. Strohalm, D. Kavan, P. NovaK, M. Volný, V.R. HavlíCEk, mMass3: a crossplatform software environment for precise analysis of mass spectrometric data, Anal. Chem. 82 (11) (2010) 4648–4651.

[53] J.A. Loo, C.G. Edmonds, R.D. Smith, Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry, Science 248 (4952) (1990) 201.

[54] D.J. Aaserud, D.P. Little, P.B. O'Connor, F.W. McLafferty, Distinguishing N- and C-terminus ions for mass spectrometry sequencing of proteins without prior degradation, Rapid Commun.

Mass Spectrom. 9 (10) (1995) 871─876.

[55] H. Li, J.J. Wolff, S.L. Van Orden, J.A. Loo, Native top-down electrospray ionization-mass spectrometry of 158 kDa protein complex by high-resolution fourier transform ion cyclotron resonance mass spectrometry, Anal. Chem. 86 (2014) 317─320.

[56] M.L. Nielsen, M.M. Savitski, R.A. Zubarev, Improving protein identification using complementary fragmentation techniques in fourier transform mass spectrometry, Mol. Cell. Proteomics 4 (6) (2005) 835─845.

Chapter 3: Top-down mass spectrometry and assigning internal fragments for determining disulfide bond positions in proteins

Reprinted with permission from

Wei, B.; Zenaidee, M. A.; Lantz, C.; Williams, B. J.; Totten, S.; Ogorzalek Loo, R. R.; Loo, J. A., Top-down mass spectrometry and assigning internal fragments for determining disulfide bond positions in proteins. *The Analyst* **2023,** *148* (1), 26-37. DOI: https://doi.org/10.1039/D2AN01517J

Benqian Wei[a], Muhammad A. Zenaidee[a, c], Carter Lantz[a], Brad J. Williams[d], Sarah Totten[d], Rachel R. Ogorzalek Loo[a], Joseph A. Loo[a, b, *]

[a] Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA; [b] Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA; [c] Australian Proteome Analysis Facility, Macquarie University, NSW, Australia; [d] Waters Corporation, Milford, MA, USA.

* Corresponding Author: Joseph A. Loo, Department of Chemistry and Biochemistry, Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA, E-mail: jloo@chem.ucla.edu.

Abstract

Disulfide bonds in proteins have a substantial impact on protein structure, stability, and biological activity. Localizing disulfide bonds is critical for understanding protein folding and higher-order structure. Conventional top-down mass spectrometry (TD-MS), where only terminal fragments are assigned for disulfide-intact proteins, can access disulfide information, but suffers from low fragmentation efficiency, thereby limiting sequence coverage. Here, we show that assigning internal fragments generated from TD-MS enhances the sequence coverage of disulfide-intact proteins by 20–60% by returning information from the interior of the protein sequence, which cannot be obtained by terminal fragments alone. The inclusion of internal fragments can extend the sequence information of disulfide-intact proteins to near complete sequence coverage. Importantly, the enhanced sequence information that arise from the assignment of internal fragments can be used to determine the relative position of disulfide bonds and the exact disulfide connectivity between cysteines. The data presented here demonstrates the benefits of incorporating internal fragment analysis into the TD-MS workflow for analyzing disulfide-intact proteins, which would be valuable for characterizing biotherapeutic proteins such as monoclonal antibodies and antibody–drug conjugates.

1. Introduction

Disulfide bonds are among the most important posttranslational modifications (PTMs) in proteins, as they have a substantial impact on protein structure, stability, and biological activity.[1–4] Determining disulfide bonding patterns is critical for understanding protein folding and higher-order structure as non-native disulfide bridges and aggregates can have detrimental effects on a protein's three-dimensional structure and consequently their function.[5,6] The advancement of biotherapeutics such as monoclonal antibodies and antibody–drug conjugates have further driven the development of more efficient and accurate experimental strategies including mass spectrometry (MS) and ion mobility-MS to characterize disulfide bond linkages,[7–13] as disulfide connectivity, which ensures its proper folding and consequently biological function and immunogenicity, is considered as a critical quality attribute during antibody manufacturing.[14,15] Mass spectrometry has established itself as a frontrunner for these characterizations owing to its exceptional sensitivity, low sample requirements, and the ability to be coupled with chromatographic separations to generate and detect diagnostic fragment ions possessing various disulfide connectivities,[16–19] which cannot be achieved easily by conventional methods such as nuclear magnetic resonance (NMR) and X-ray crystallography.[20,21]

Conventional "bottom-up" MS approaches employ chemical reduction and alkylation to cleave disulfide bonds and cap the free cysteines, followed by enzymatic digestion of the protein prior to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.[22,23] Although protein sequence usually can be unambiguously determined using this approach, information on disulfide bond locations and connectivities can be lost.[18,19] To compensate for this limitation,

alternative strategies including proteolysis without prior reduction or with partial reduction have been utilized to generate disulfide-linked peptides for LC-MS/MS measurements.[24–29] This allows for the elucidation of disulfide bonding patterns by comparing the peptides resulting from the reduced regions with the peptides from constrained regions to identify disulfide-linked peptides. However, it is difficult to control the amount of disulfide reduction using this approach, which results in complex mixtures of peptides with differing amounts of capped cysteines, making data analysis challenging.[30] Moreover, with limited disulfide reduction, protein sequence coverage may not be sufficient to capture all disulfide linkage information. This problem will be exacerbated with increasing protein size and/ or proteins that contain a large number of disulfide bonds.[31]

Top-down mass spectrometry (TD-MS), where direct mass measurement and subsequent fragmentation of intact gas-phase protein ions in the mass spectrometer to obtain the primary sequence information, has gained in popularity in recent years for interrogating proteins with various PTMs, including but not limited to disulfide bonds.[32–36] TD-MS bypasses the time-consuming digestion and separation steps, allowing for all disulfide information to be preserved. By comparing the accurate measured mass with the theoretical sequence mass of disulfide intact proteins, the number of disulfide bonds can be readily determined. The modification sites can be further identified by subsequent fragmentation of the intact protein ions with high sequence coverage. However, challenges still remain. Accessing disulfide bond information usually requires concurrent fragmentation of the protein backbone and disulfide bonds to gain extensive sequence coverage, which is important for localizing disulfide bridges, whereas TD-

MS suffers from low relative fragmentation efficiency, limiting sequence coverage.[37–39] To increase sequence coverage, various fragmentation methods (alternative to the traditionally employed collision-based techniques) have been employed to characterize disulfide-intact peptides and proteins including electron-based dissociation (ExD),[9,31,40–43] photon-based dissociation (PD),[19,44–48] and their hybrid methods with varying success.[12,30,49,50] An additional approach to increase TD-MS sequencing efficiency is to incorporate the assignment of internal fragments,[51,52] generated by multiple gas-phase cleavages of the polypeptide backbone, into the data analysis workflow.[53]

While the analysis of internal fragment ions has been largely ignored by the TD-MS community due to the general lack of software tools to accurately and reliably assign them, the concept of the formation of internal fragment ions in TD-MS spectra is not novel. Previous studies have shown that the inclusion of internal fragments results in much richer sequence information of small peptides,[51,54–57] intact proteins,[58–64] protein complexes,[65,66] and aid the identification of ambiguous proteoforms in mammalian cell lysates by top-down ptoteomics.[67] In addition, a recent study by Chin et al. demonstrated the utility of internal fragments to enhance sequence coverage and to decipher disulfide bonds of disulfide-rich peptides.[68] Schmitt et al. also applied internal fragments to determine sequence motifs located within a disulfide constrained loop of SOD1 protein that could not be achieved by terminal fragments alone.[63] The benefits of including internal fragments for characterizing disulfide-intact proteins are two-fold. Identifiable internal fragment ions within disulfide constrained regions can be generated without the need to cleave the disulfide bond,[68] lowering the barrier to obtaining more sequence

information. Second, by including internal fragments, the chance of identifying product ions that result from cleavage of disulfide bonds to access disulfide linkage information is higher than analyzing terminal fragments alone.

Here, we show that assigning internal fragments generated from collisionally activated dissociation (CAD) and ExD can increase the sequence coverage of disulfide-intact proteins by accessing the interior of the protein sequence constrained by multiple disulfide bonds. Importantly, by correlating the number of disulfide bonds cleaved by internal fragments to their sequence positions, the relative locations of disulfide bonds can be determined. By specifically analyzing internal fragments with disulfide bonds remaining intact, disulfide connectivity can be determined. This study demonstrates the benefits of considering internal fragments when analyzing these heavily constrained proteins, which would be valuable for characterizing biotherapeutic proteins that contain a large number of disulfide bond.

2. Experimental

2.1. Materials and sample preparation.

The proteins β-lactoglobulin from bovine milk, ribonuclease A from bovine pancreas, α-lactalbumin from bovine milk, trypsin inhibitor from glycine max, and m-nitrobenzyl alcohol (mNBA) were purchased from Sigma-Aldrich (St Louis, MO, USA). Lysozyme from chicken white egg was acquired from EMD Millipore (Darmstadt, Germany). LC/MS-grade water, methanol and formic acid were obtained from Fisher Chemical (Hampton, NH). All proteins were used without further purification. Protein samples were prepared in 49.5:49.5:1 water/

methanol/formic acid to a final concentration of 10 or 20 μM. Supercharging agent mNBA was added to the ribonuclease A and α-lactalbumin solutions at a 0.25% (v/v) concentration.

2.2. Mass Spectrometry.

All samples were measured with a 15-Tesla SolariX Fourier transform ion cyclotron resonance (FTICR)-MS instrument equipped with an infinity ICR cell (Bruker Daltonics, Billerica, MA, USA). The protein solutions were loaded into in-house pulled capillaries coated with gold, and electrosprayed by applying a voltage between 0.7 and 1.5 kV on the electrospray ionization capillary. Individual charge states of each multiply protonated protein (11+ to 15+ for β-lactoglobulin, 8+ to 12+ for lysozyme, 8+ to 14+ for ribonuclease A, and 11+ to 14+ for α-lactalbumin) were isolated in the quadrupole, with an isolation window of 10 m/z before fragmentation. Three fragmentation methods including CAD, electron capture dissociation (ECD), and electron induced dissociation (EID) were applied to each isolated ion. For CAD fragmentation, collision energies were adjusted to achieve the same lab-frame energy for different charge states of each protein. The lab-frame energy is defined as the multiplication product of charge state and collision energy. The lab-frame energies used for each protein are: β-lactoglobulin, 336 V; lysozyme, 438 V; ribonuclease A, 330 V; α-lactalbumin, 286 V to achieve optimal fragmentation. For ECD fragmentation, the pulse length was set at 0.02 s, with a lens voltage at 50 V and bias voltage at 2 V. For EID fragmentation, the pulse length was set at 0.02 s, with a lens voltage at 50 V and bias voltage ranging from 26 to 30 V.

CAD-MS/MS of trypsin inhibitor (TI) was done by isolating $[TI + 17H]^{17+}$ with an isolation window of 10 *m/z*. The CAD energy was set at 20 V, which reduced the precursor ion signal to

~40% of the mass spectral level.

ECD-MS/MS of β-lactoglobulin and lysozyme were also performed on a Waters SELECT SERIES™ Cyclic IMS Q-ToF mass spectrometer (Waters, Milford, MA, USA) with an electromagnetostatic ExD cell (e-MSion Inc., Corvallis, OR) mounted before the cyclic ion mobility cell to allow for pre-IMS ECD fragmentation. All ECD parameters were optimized to achieve the best fragmentation.

2.3. Data analysis.

2.3.1. Data processing and fragment assignment.

Raw MS/MS spectra acquired on FTICR were deconvoluted using Bruker Data Analysis software (SNAP algorithm). Mass spectra acquired on the Waters Cyclic IMS Q-ToF instrument was deconvoluted using Waters' BayesSpray algorithm. Deconvoluted mass lists were uploaded into the ClipsMS (2.0) program[53] for fragment ion matching. The mass tolerance was set at 2 ppm for FTICR data and 5 ppm for Waters Q-ToF data and the smallest internal fragment size was set at 5 amino acids. For sequence coverage and disulfide bond cleavage analyses, to account for all disulfide-containing fragment ions, modifications considering all possible disulfide cleavage positions (S–S and C–S cleavage) were imported as an unlocalized modification file for fragment matching. Up to 2 water and ammonia losses were included in the unlocalized modification file for CAD fragmentation. No localized modifications were imported for these analyses. For disulfide connectivity analysis, modifications applying one hydrogen loss on each cysteine to suggest the integrity of the disulfide bond were imported as a localized modification file for fragment matching. No unlocalized modifications were imported for this analysis. All

localized and unlocalized modification files for fragment matching are available in the ESI (Tables S1–S7†). All six terminal fragment types including *a*, *b*, *c*, *x*, *y*, *z* were searched for all three fragmentation methods, while only *by* internal fragments were searched for CAD and *cz* internal fragments for ECD/EID spectra. All terminal fragments were assigned first (*i.e.*, given priority) before considering internal fragments, and all overlapping internal fragments due to the arrangement and/or frameshift ambiguity[63] were removed. After fragment matching and duplicates removal, all assigned internal fragments were further verified by manually examining their isotopic profiles against the raw MS/MS spectra to eliminate uncertain assignments.

2.3.2 Protein Sequence Coverage.

Protein sequence coverage is calculated by the number of observed inter-residue cleavage sites divided by the total number of possible inter-residue cleavage sites on the protein backbone.

3. Results and discussion

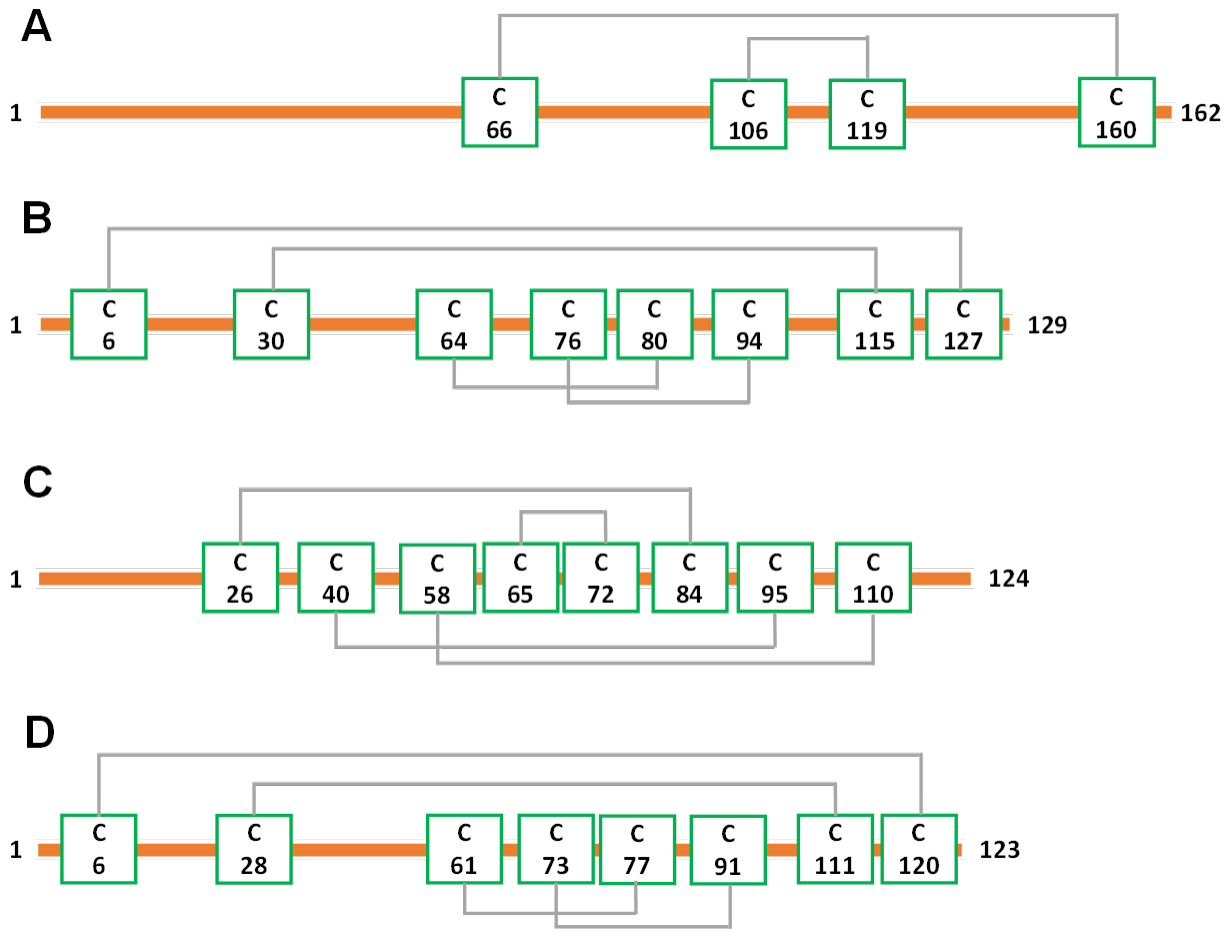
3.1. Internal fragments can access the interior protein sequence constrained by multiple disulfide bonds.

To demonstrate that internal fragments can enhance sequence information of disulfide intact proteins, three fragmentation methods were applied, CAD, ECD, and EID on various isolated precursor charge states of four disulfide-intact proteins, including β-lactoglobulin (2 disulfide bonds), lysozyme (4 disulfide bonds), ribonuclease A (4 disulfide bonds), and α-lactalbumin (4 disulfide bonds). The disulfide connectivity of these proteins is shown in Scheme 1. EID

fragmentation of β-lactoglobulin, $[B\text{-lac} + 14H]^{14+}$ generated rich mass spectra filled with informative peaks (Figure. 1A). Many of the peaks in the spectra that were not assigned as terminal fragments can be assigned as internal fragments (Figure. 1A inset), demonstrating that more information can be extracted from a single MS/ MS spectrum when considering internal fragments. Importantly, the location of all the assigned fragments for B-lac demonstrates that internal fragments span much of the interior sequence enclosed by multiple disulfide bonds, providing complementary sequence information to terminal fragments (Figure. 1B). Similar results were also observed for EID of lysozyme, $[Lys + 10H]^{10+}$ (Figure. 1C and D). In both cases, the extent of information extracted from a single mass spectrum can be enhanced significantly when including internal fragments. Further, ECD and CAD of the same isolated precursor ions show similar fragmentation patterns, although ECD is less energetic than EID and CAD, and generated significantly fewer internal fragments (Figure. S1 and S2†).

To compare sequence information obtained from terminal fragments with internal fragments, all assigned unique fragments generated from every charge state for each protein were integrated. Assigning internal fragments generated from CAD, ECD, and EID increases the sequence coverage by 20–60% for all proteins examined. For example, sequence coverage increases from 43% to 83% for EID of β-lactoglobulin (Figure. 2D), 37% to 84% for EID of lysozyme (Figure. 3F), 40% to 87% for EID of ribonuclease A (Figure. S3F†), and 36% to 90% for EID of α-lactalbumin (Figure. S4F†) after including internal fragments. Incorporating internal fragments can cover almost every single inter-residue site to achieve near complete sequence coverage (99%) for CAD of lysozyme (Figure. 3F), with CAD of α-lactalbumin also close to 100%

sequence coverage (96%, Figure. S4F†). This is primarily due to the fact that the generation of terminal fragments beyond regions enclosed by disulfide bonds is difficult (*vide infra*); most often, an S–S bond would need to be cleaved in order to release the terminal fragment. This is further discussed below.



Scheme 1. Disulfide bond connectivities of the four proteins examined.
(A) β-lactoglobulin (2 disulfide bonds), (B) lysozyme (4 disulfide bonds), (C) ribonuclease A (4 disulfide bonds), (D) α-lactalbumin (4 disulfide bonds).

The sequence of these proteins can be classified into different regions depending on the number of disulfide bonds enclosed. For example, β-lactoglobulin has two disulfide bonds with a connectivity of Cys66–Cys160 and Cys106–Cys119 (Scheme 1A), thus the β-lactoglobulin

sequence can be classified into three regions: (i) sequence not enclosed by disulfide bond (residues 1–66, 160–162), (ii) sequence enclosed by one disulfide bond (residues 66–106, 119–160), and (iii) sequence enclosed by two disulfide bonds (residues 106–119). Similarly, the sequence of lysozyme, which possesses four disulfide bonds (Scheme 1B) can be classified into five regions including sequence not enclosed by a disulfide bond (residues 1–6, 127–129), sequence enclosed by one disulfide bond (residues 6–30, 115–127), sequence enclosed by two disulfide bonds (residues 30–64, 94–115), sequence enclosed by three disulfide bonds (residues 64–76, 80–94), and sequence enclosed by four disulfide bonds (residues 76–80). For the other two proteins with four disulfide bonds, the primary protein sequence can also be separated into specific regions (ribonuclease A, Scheme 1C, and α-lactalbumin, Scheme 1D). To investigate the utility of internal fragments for accessing highly disulfide constrained regions, the extent of sequence information obtained from terminal and internal fragments at different sequence regions were compared and a clear trend can be observed. Generally, most internal fragments originate from the interior of the sequence within disulfide bonded regions, while terminal fragments originate from the outermost sequence. For example, for CAD of β-lactoglobulin, terminal fragments cover more sequence not enclosed by disulfide bond than internal fragments (64% vs. 60%, Figure. 2A), corresponding to a change of +4%, while no terminal fragments and only internal fragments cover the sequence enclosed by two disulfide bonds (0% vs. 54%, Figure. 2C), corresponding to a change of −54%. Similarly, for CAD of lysozyme with four disulfide bonds and five distinct sequence regions, the sequence coverage change when comparing terminal vs. internal fragments are +43%, −14%, −47%, −65%, and

−20%, respectively, when going deeper into the middle of the sequence (Fig. 3A–E). This data clearly demonstrates that internal fragments significantly enhance sequence information of the regions constrained by multiple disulfide bonds. A similar trend was observed for ECD and EID of these two proteins (Figures. 2 and 3) and the other two proteins possessing four disulfide bonds (ribonuclease A, Figure. S3, and α-lactalbumin, Figure. S4†), with the relative sequence coverage decreasing for terminal fragments while increasing for internal fragments when reaching the interior protein sequence (Figure. 2A–C, 3A–E, S3A–E, and S4A–E†). Notably, some specific sequence regions can only be accessed by internal fragments, such as the sequence enclosed by two disulfide bonds of β-lactoglobulin (Figure. 2C) and sequence enclosed by four disulfide bonds of ribonuclease A (Figure. S3E†) and α-lactalbumin (Figure. S4E†), highlighting the ability of internal fragments to cover regions that cannot be reached by terminal fragments. The data shown here shows promise for the inclusion of internal fragments in obtaining more comprehensive sequence information for disulfide-intact proteins.

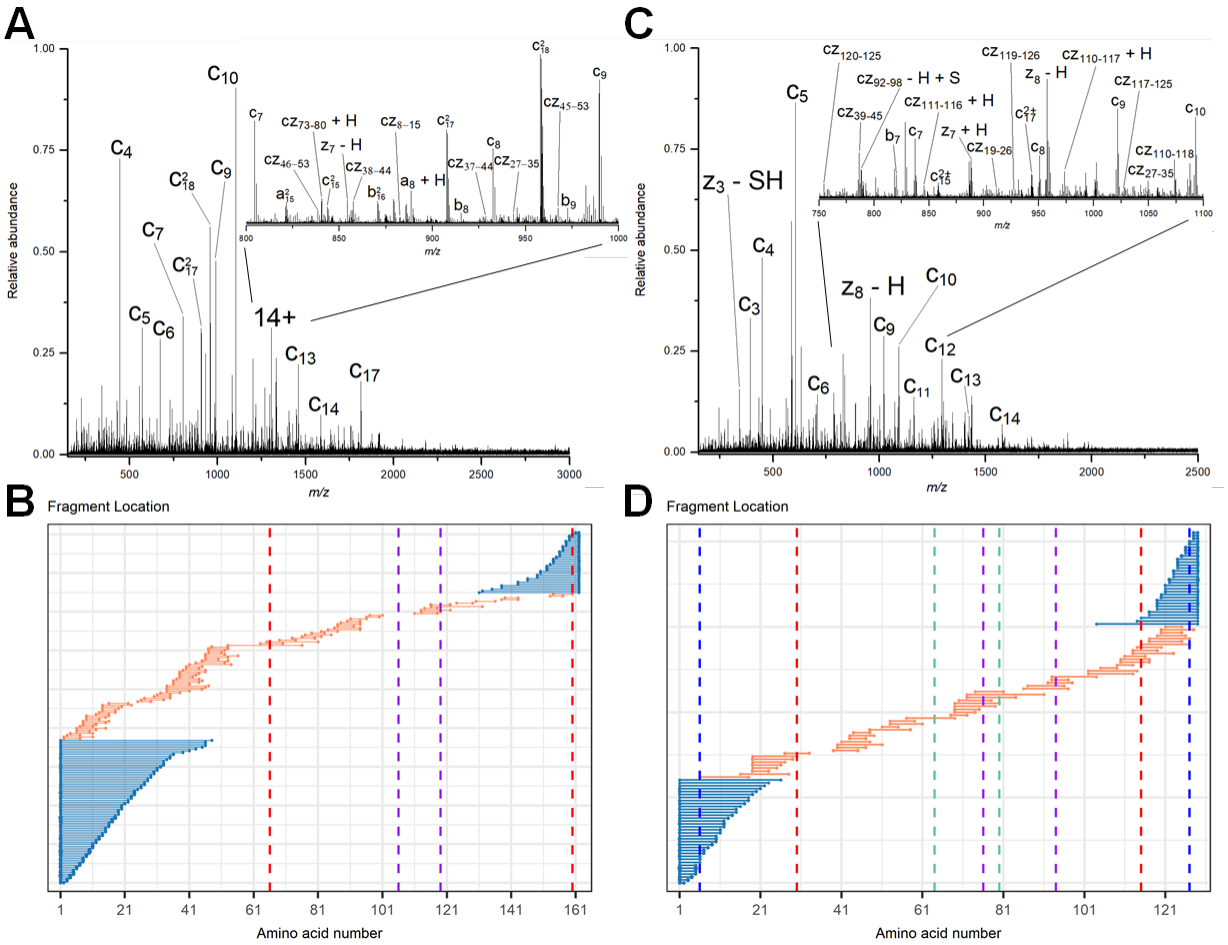


Figure 1. Data analysis of EID MS/MS of β-lactoglobulin and lysozyme.
Representative EID MS/MS spectra of (A) β-lactoglobulin, $[B\text{-lac} + 14H]^{14+}$ and (C) lysozyme, $[Lys + 10H]^{10+}$. Fragment location maps indicating the region of the protein sequence covered by terminal fragments (blue) and internal fragments (orange) for (B) EID of β-lactoglobulin, $[B\text{-lac} + 14H]^{14+}$ (spectrum in A) and (D) EID of lysozyme, $[Lys + 10H]^{10+}$ (spectrum in C). Vertical dashed lines in panels B and D represent cysteines positions, with the same color indicating a disulfide bond is formed between those two cysteines.

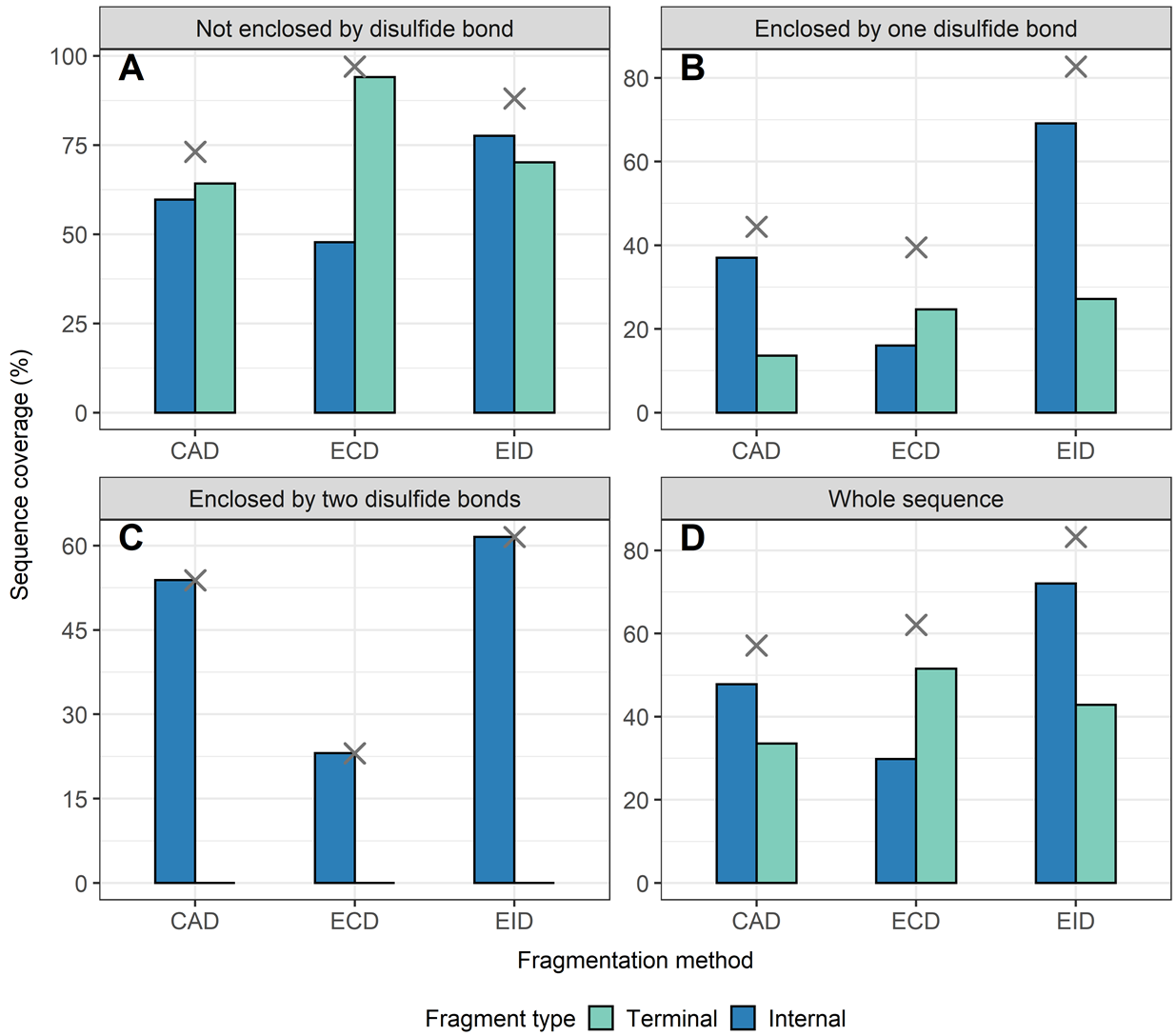


Figure 2. Sequence coverage analysis of β-lactoglobulin.
The extent of sequence information obtained by terminal and internal fragments for β-lactoglobulin at different sequence regions after integrating data from all five charge states (11+ to 15+) and for all three fragmentation methods (CAD, ECD, and EID) examined, (A) sequence not enclosed by disulfide bond, (B) sequence enclosed by one disulfide bond, (C) sequence enclosed by two disulfide bonds, (D) whole sequence. Cross marks in each panel indicate the sequence coverage after combing terminal and internal fragments.

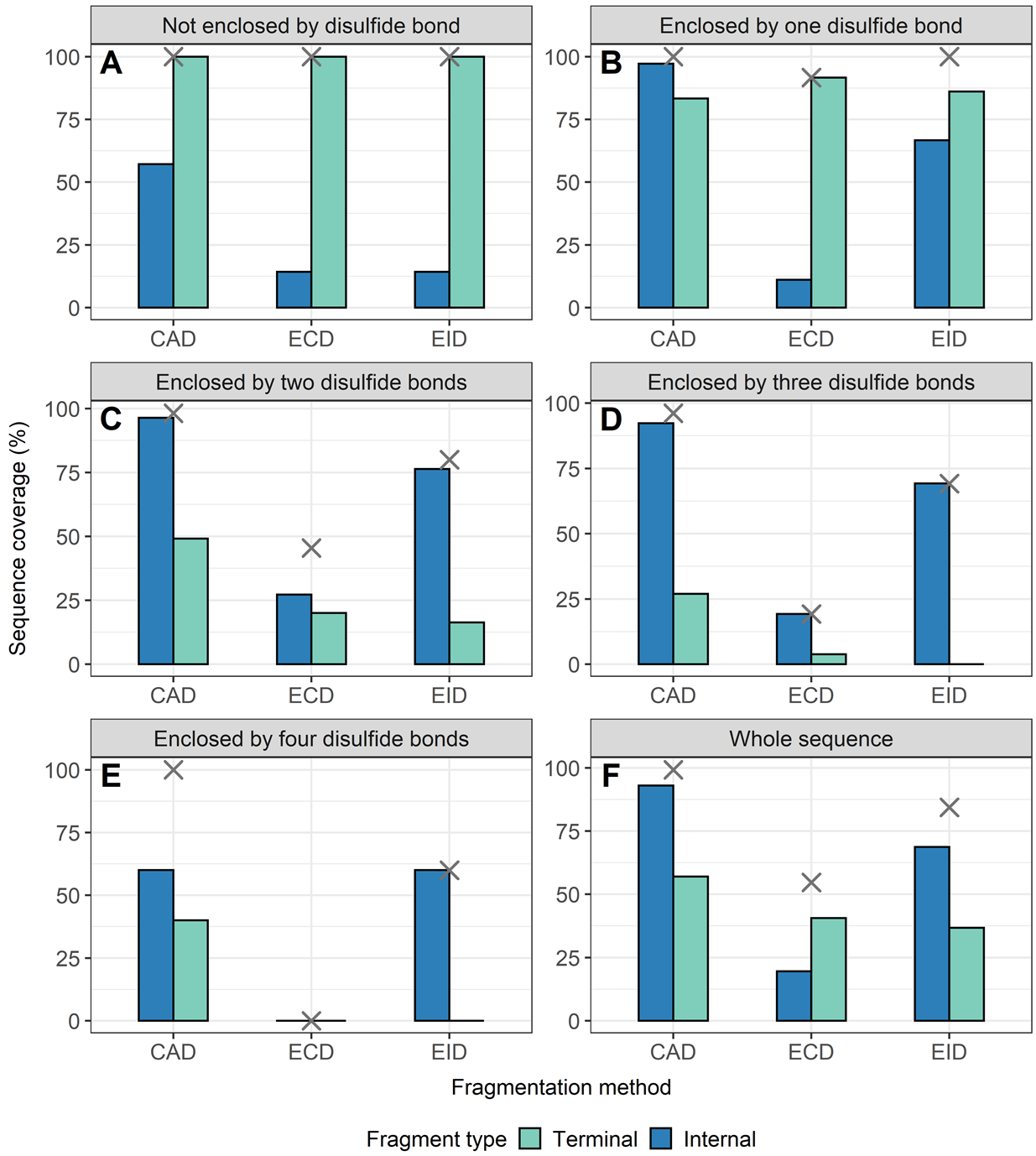


Figure 3. Sequence coverage analysis of lysozyme.

The extent of sequence information obtained by terminal and internal fragments for lysozyme at different sequence regions after combining data from all five charge states (8+ to 12+) and for all three fragmentation methods (CAD, ECD, and EID) examined, (A) sequence not enclosed by disulfide bond, (B) sequence enclosed by one disulfide bond, (C) sequence enclosed by two disulfide bonds, (D) sequence enclosed by three disulfide bonds. (E) Sequence enclosed by four disulfide bonds. (F) Whole sequence. Cross marks in each panel indicate the sequence

coverage after combing terminal and internal fragments.

3.2. Internal fragments can determine the relative position of disulfide bonds.

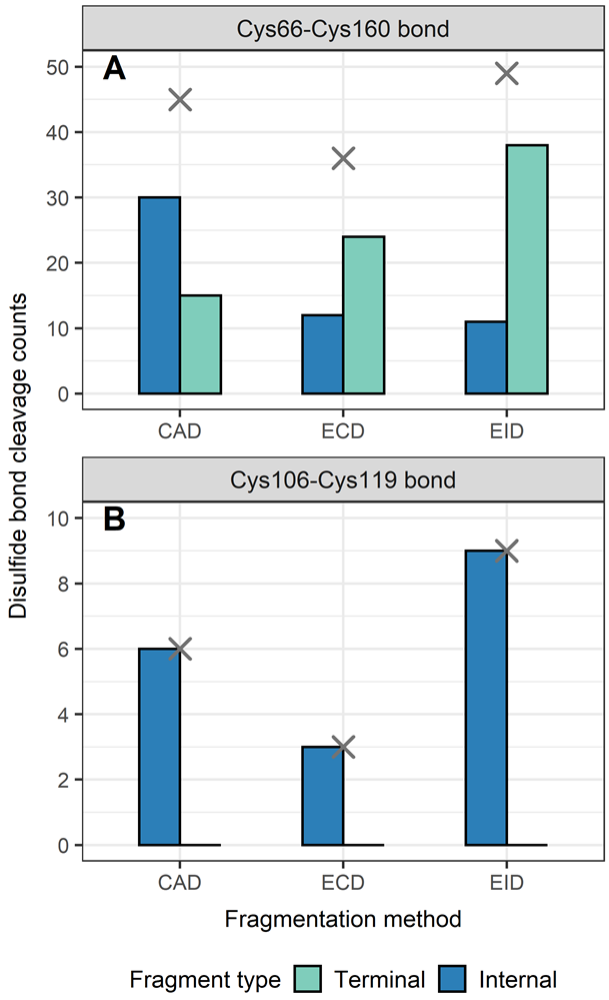


Figure 4. Disulfide bond cleavage analysis of β-lactoglobulin.
Number of disulfide bonds cleaved by terminal and internal fragments for β-lactoglobulin after integrating data from all five charge states (11+ to 15+) and for all three fragmentation methods (CAD, ECD, and EID) examined, (A) Cys66–Cys160 bond, (B) Cys106–Cys119 bond. Cross marks in each panel indicate the disulfide bond cleavage counts after combing terminal and internal fragments.

To determine the position of disulfide bonds for these proteins, the number of disulfide bond cleavages were analyzed. We show here that terminal fragments result from cleavage of

disulfide bonds located on the exterior of the protein, while internal fragments can result from cleavage of disulfide bonds within the interior of the protein. For example, terminal fragments generated by EID of β-lactoglobulin (2 disulfide bonds) resulted from more cleavages at the outermost disulfide bond (Cys66–Cys160) than internal fragments (38 vs. 11, Figure. 4A), while only internal fragments originated from the cleavage of the interior disulfide bond (9 times at the Cys106–Cys119 bond, Figure. 4B). This trend is more pronounced for proteins with a greater number of disulfide bonds. For example, EID of lysozyme (4 disulfide bonds) showed that the Cys6–Cys127 bond was cleaved 62 times by terminal fragments but only 6 times by internal fragments (Figure. 5A). For the Cys30–Cys115 bond, located more interior of the protein sequence, the difference between disulfide cleavages from terminal and internal fragments was reversed, 10 vs. 16, respectively (Figure. 5B). For the Cys64–Cys80bondand theCys76–Cys94 bond, the disulfide cleavages comparison is 0 vs. 17 and 0 vs. 19 (terminal vs. internal, Figure. 5C and D). This trend was also observed for CAD and ECD of β-lactoglobulin and lysozyme, and the other two disulfide bonded proteins (ribonuclease A, Figure. S5, and α-lactalbumin, Figure. S6†). Surprisingly, for disulfide bonds buried within the protein, their cleavages were only explained by internal fragments (Figure.4B, 5C and D, S5D, S6C and D†), highlighting the use of internal fragments to access disulfide bond information that cannot be obtained by terminal fragments.

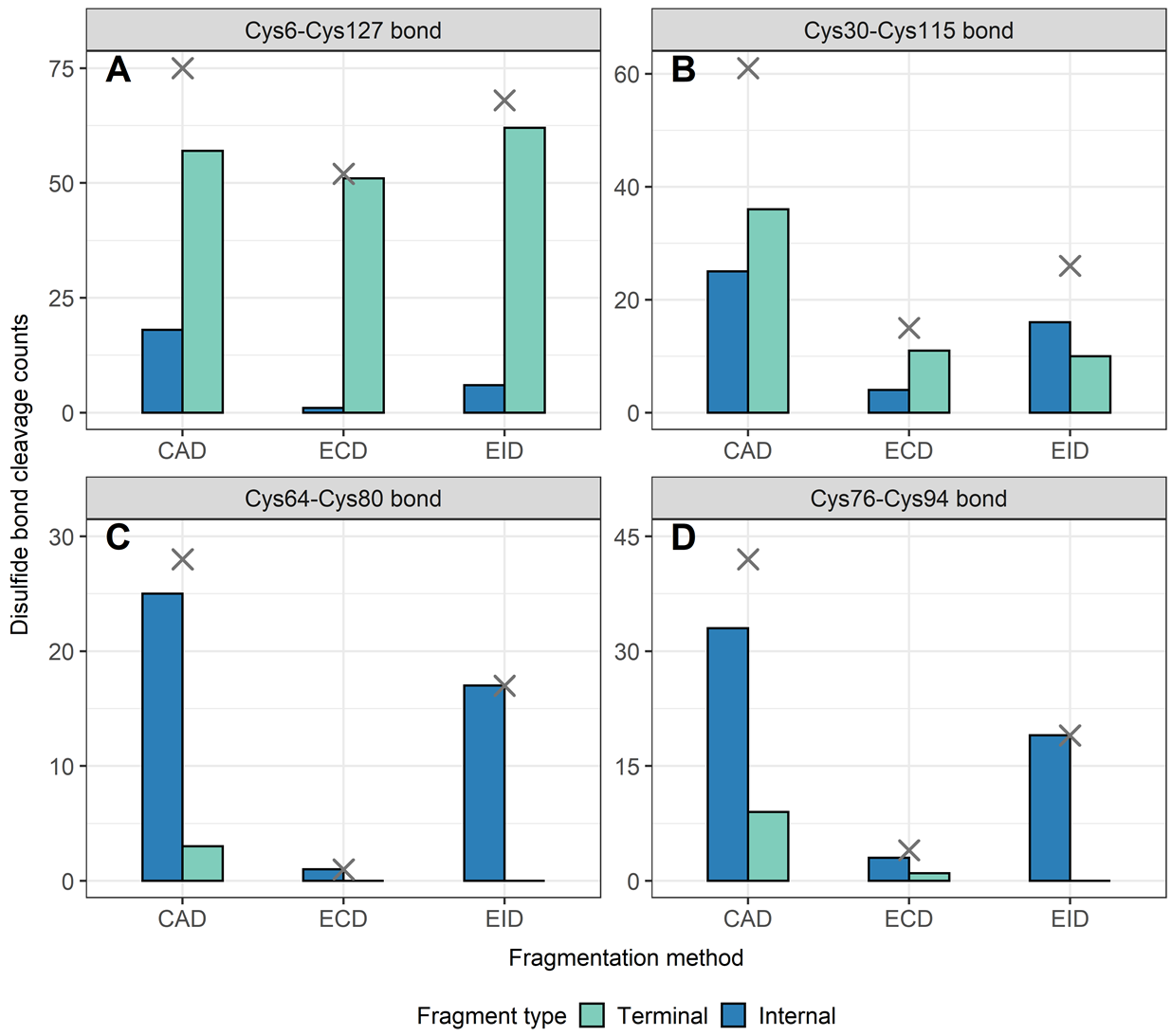


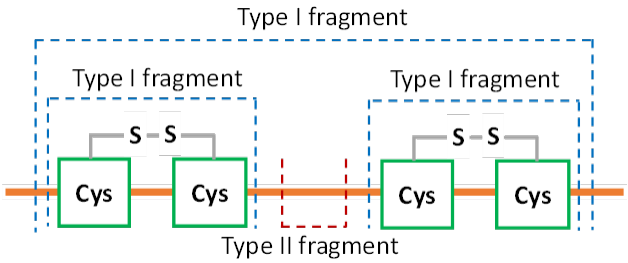
Figure 5. Disulfide bond cleavage analysis of lysozyme.
Number of disulfide bonds cleaved by terminal and internal fragments for lysozyme after combining data from all five charge states (11+ to 15+) for all three fragmentation methods (CAD, ECD, and EID) examined, (A) Cys6–Cys127 bond, (B) Cys30–Cys115 bond, (C) Cys64–Cys80 bond, (D) Cys76–Cys94 bond. Cross marks in each panel indicate the disulfide bond cleavage counts after combing terminal and internal fragments.

These data indicate that by correlating the relative number of disulfide cleavages resulting from internal fragments to their sequence positions, the relative locations of disulfide bonds can be determined. The outermost disulfide bonds are explained more by terminal fragments, as their

formation usually only requires one backbone cleavage in addition to one disulfide bond cleavage. In contrast, in order for internal fragments to explain these outermost disulfide bond cleavages, simultaneous cleavages of one disulfide bond and multiple protein backbone bonds are required, raising the energy barrier compared to terminal fragments. When going deeper into the protein sequence, more internal fragments result from cleavage of innermost disulfide bonds. In these highly constrained regions, simultaneous cleavages of multiple disulfide bonds and one protein backbone bond are needed to generate terminal fragments, while the formation of internal fragments still only require one disulfide bond cleavage in addition to multiple protein backbone cleavages. These results can be rationalized by considering the relative energies required to cleave the protein backbone (~10–15 kcal mol^{-1}) compared to the disulfide bond (~45–60 kcal mol^{-1}).[38,69] Because the energy barrier of cleaving a disulfide bond is higher than cleaving a protein backbone bond, the energy requirement of forming internal fragments in the interior protein sequence could be lower than for terminal fragments, and thus internal fragments could more easily result from cleavage of disulfide bonds buried within the protein. To support our data, ECD of β-lactoglobulin and lysozyme were conducted using a different mass spectrometry system (Waters Select Series Cyclic IMS Q-TOF). Similar trends for both sequence coverages and disulfide bond cleavages were observed (Figure. S7 and S8†), further demonstrating the utility of internal fragments to cover the interior protein sequence and determine the relative positions of disulfide bonds.

3.3. Internal fragments retaining intact disulfide bonds can determine disulfide connectivity.

To determine the disulfide connectivity between cysteines, we focus on fragments that only result from protein backbone cleavages and retain the intact disulfide bonds. Fragments that arise from these types of cleavages can be divided into type I fragments and type II fragments (Scheme 2). Type I fragments correspond to fragments (terminal and internal) that traverse an even number of dehydrocysteine residues (e.g.,2,4, 6) and contain mass shifts associated with the multiplication product of the number of disulfide bonds and dehydrocysteines (no. of disulfide bonds × −2 Da, Scheme 2). Type II fragments correspond to internal fragments formed between adjacent cysteine residues; thus, no disulfide bonds are involved (Scheme 2). Type I fragments suggest that intact disulfide bonds are maintained within the cysteines involved, while type II fragments suggest that those two adjacent cysteines are highly unlikely to be connected.



Scheme 2. Two types of fragments that retain intact disulfide bonds to determine the disulfide connectivity.
The two types of fragments retaining intact disulfide bonds to determine disulfide connectivity. A hydrogen loss (−1 Da) was applied on every cysteine residue to suggest the integrity of disulfide bonds involved. Type I fragment traverses an even number of dehydrocysteines (2, 4, 6 *etc.*), suggesting that intact disulfide bonds are formed within the cysteines involved. Type II fragment is generated between adjacent cysteines with no disulfide bonds involved, suggesting that those two adjacent cysteines are highly unlikely to be connected.

To determine disulfide connectivity using type I and type II fragments, CAD fragmentation of trypsin inhibitor (181 residues, 20.1 kDa, 2 disulfide bonds, Figure. S9A†), $[TI + 17H]^{17+}$ (Figure. 6A) was investigated, as the non-overlapping feature of the two disulfide bonds of trypsin inhibitor makes it a good test example. Type I fragments can be used to determine the disulfide connectivity of the two disulfide bonds of trypsin inhibitor. For example, the two dehydrocysteines (Cys39 and Cys86) located close to the N-terminus were traversed by 9 type I terminal fragments and 70 type I internal fragments, and the two dehydrocysteines (Cys136 and Cys145) located closer to the C-terminus were traversed by 8 type I terminal fragments and 7 type I internal fragments, which strongly suggests that the connectivity between these cysteines should be "Cys39–Cys86" and "Cys136–Cys145" for these two disulfide bonds (Figure. 6A). Four examples of type I internal fragments traversing these two disulfide bonds are shown (Figure. S9†). It should be noted that fragments traversing an even number of dehydrocysteines do not guarantee the integrity of disulfide bonds involved; however, the likelihood of them being cleaved is much lower. For example, only one internal fragment (*by*42–137) traversed the middle two dehydrocysteines (Cys86 and Cys136), whereas the formation of 3 type II fragments between Cys86 and Cys136 (*by*96–115, *by*100–115, *by*125–132) indicates that these two cysteines are not likely to be connected.

Similar results could also be gleaned when α-lactalbumin (123 residues, 14.2 kDa, 4 disulfide bonds, Scheme 1D), which possesses a more complicated disulfide linkage, was analyzed (Figure. 6B). Disulfide connectivity of α-lactalbumin was determined by interrogating the innermost disulfide bonds, and expanding to the outermost disulfide bonds. The middle four

dehydrocysteines (Cys61, Cys73, Cys77, Cys91) were traversed by 4 type I internal fragments (*by*50–97, *by*51–106, *by*53–97, *by*60–106, Figure. S10†), indicating that two disulfide bonds are formed within these four cysteines (Figure. 6B). Type II internal fragments were then used to aid the assignment of the exact connectivity within these four cysteines. The formation of 8 type II internal fragments between Cys61 and Cys73, and 12 type II internal fragments between Cys77 and Cys91 strongly suggests that the connectivity of “Cys61–Cys73” and “Cys77– Cys91” is not likely. In addition, the lack of type I internal fragments traversing the middle two dehydrocysteines (Cys73 and Cys77) indicates that the “Cys73–Cys77” connectivity is not likely either. Should Cys73 and Cys77 be connected, type I internal fragments traversing the dehydro form of these two cysteines would have been generated, as demonstrated by CAD of trypsin inhibitor (Figure. 6A). Therefore, the only possible connectivity of these four cysteines is “Cys61–Cys77” and “Cys73– Cys91”. Expanding to the outermost cysteines, the formation of 1 type I internal fragment traversing the middle six dehydrocysteines (*by*20–113), and 8 type I terminal fragments traversing all eight dehydrocysteines indicates that two more disulfide bonds are formed between the four cysteines located on the exterior protein sequence. The presence of type I internal fragment by20–113 determines the connectivity of “Cys28–Cys111”, provided that the middle four cysteines are associated with each other. This is further supported by the fact that 28 type II internal fragments are formed between Cys6 and Cys28, 40 type II internal fragments are formed between Cys28 and Cys61, 28 type II internal fragments are formed between Cys91 and Cys111, and 1 type II internal fragment is formed between Cys111 and Cys120. These type II internal fragments rule out the possibility of “Cys6–Cys28”, “Cys28–Cys61”, “Cys91–Cys111”, and “Cys111– Cys120”

connectivities. Therefore, the two outermost disulfide bond connectivities can be determined as “Cys28–Cys111” and “Cys6–Cys120”. It is noteworthy that only internal fragments can access the middle four cysteines, demonstrating again the value of analyzing internal fragments to obtain comprehensive disulfide bond information. The disulfide connectivity of lysozyme can be elucidated and determined in a similar way using these two types of fragments (Figure. S11†).

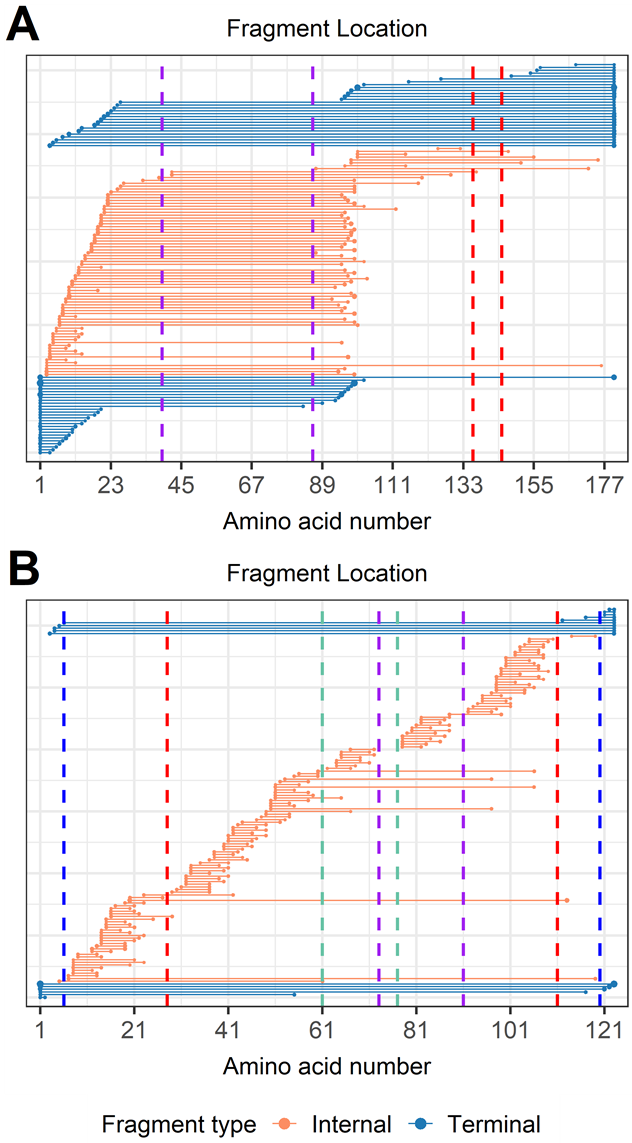


Figure 6. Analysis of fragments that retain intact disulfide bonds to determine disulfide connectivities of trypsin inhibitor and α-lactalbumin.

Fragment location maps after importing a hydrogen loss localized modification on every

cysteine. (A) CAD of trypsin inhibitor, $[TI + 17H]^{17+}$, and (B) CAD of α-lactalbumin after integrating data from all four charge states examined (11+ to 14+). Vertical dashed lines represent cysteines positions, with the same color indicating a disulfide bond is formed between those two cysteines. Internal fragments traversing an even number of dehydrocysteines (type I fragments) suggest that intact disulfide bonds are formed within those cysteines, while internal fragments formed between adjacent cysteines (type II fragments) suggest that those two cysteines are not likely to relate to each other.

4. Conclusions

Here we report the utility of internal fragments to enhance information obtained from disulfide-intact proteins. We demonstrate that internal fragments can access the interior protein sequence constrained by multiple disulfide bonds that cannot be reached by terminal fragments, resulting in a sequence coverage increase of 20–60% to cover nearly the complete sequence of disulfide-intact proteins. We show that terminal fragments result from cleavage of disulfide bonds located on the exterior of the protein while internal fragments represent cleavage of more disulfide bonds buried within the interior of the protein. By correlating the relative number of internal fragments that result in disulfide cleavages to their sequence positions, the relative positions of disulfide bonds can be determined. Lastly, we show that internal fragments retaining intact disulfide bonds, which are traditionally overlooked, can be used to determine the disulfide connectivity. By analyzing internal fragments, it is possible to gain more sequence information and elucidate disulfide linkage patterns for proteins with unknown disulfide connectivities, which would be valuable for characterizing biotherapeutic proteins that contain many disulfide bonds.

Data Availability

All experimental supporting data associated with this article are available in the main manuscript and in the ESI.†

Author contributions

Benqian Wei: investigation, formal analysis, writing – original draft, writing – review & editing. Muhammad A. Zenaidee: investigation, supervision, writing – review & editing. Carter Lantz: investigation, writing – review & editing. Brad J. Williams and Sarah Totten: investigation, writing – review & editing. Rachel R. Ogorzalek Loo and Joseph A. Loo: supervision, funding acquisition, writing – review & editing.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Support from the US National Institutes of Health (R01GM103479, R35GM145286, S10RR028893), the US National Science Foundation (NSF) (CHE1808492), and the US Department of Energy (DEFC02-02ER63421) are gratefully acknowledged. C. L. acknowledges support from the Ruth L. Kirschstein National Research Service Award Program (GM007185).

Chapter 3: Supporting Information

Supplementary Figures

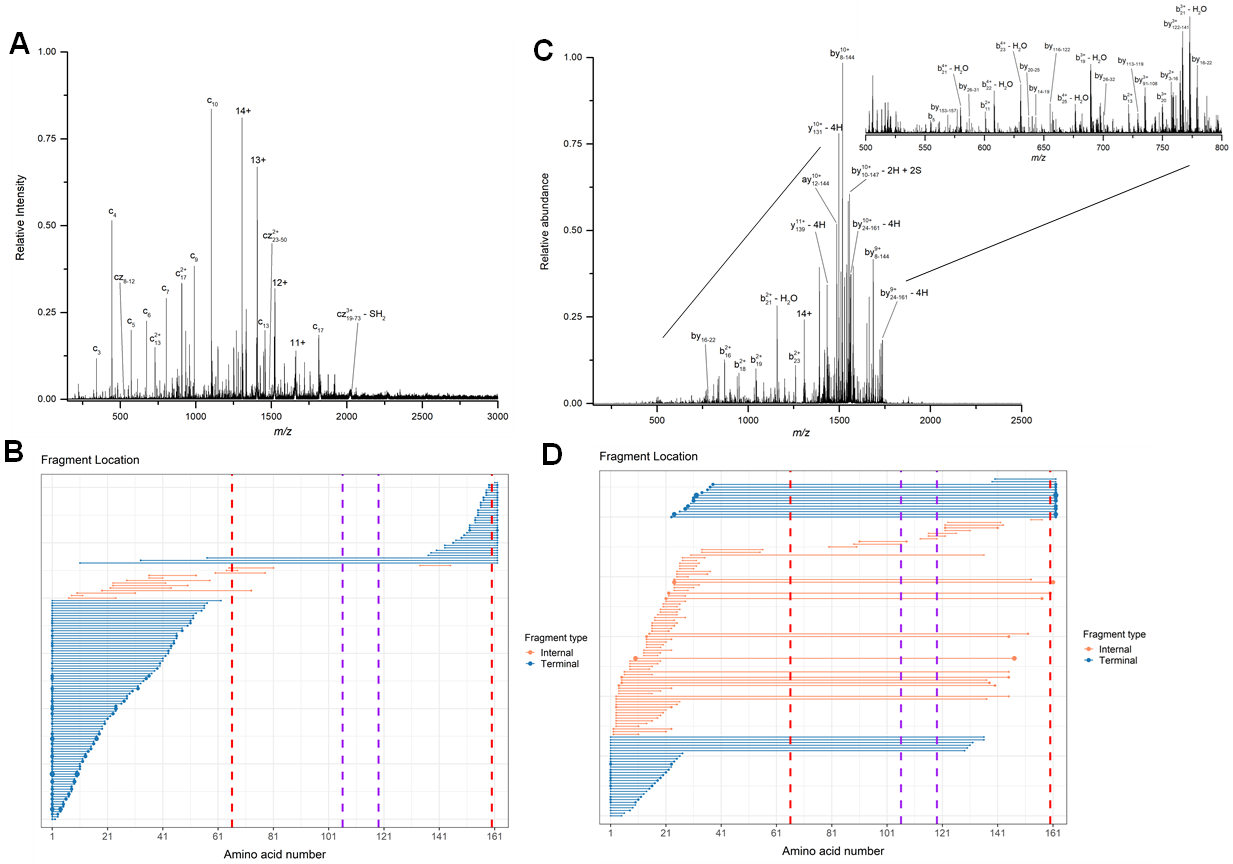


Figure S1. Data analysis of ECD and CAD MS/MS of β-lactoglobulin.
A. Representative ECD MS/MS spectrum of beta-lactoglobulin, $[Blac + 14H]^{14+}$. **B.** Fragment location map indicating the region of the protein sequence covered by terminal and internal fragments for spectrum in panel A. **C.** Representative CAD MS/MS spectrum of beta-lactoglobulin, $[Blac + 14H]^{14+}$. **D.** Fragment location map indicating the region of the protein sequence covered by terminal and internal fragments for spectrum in panel C.

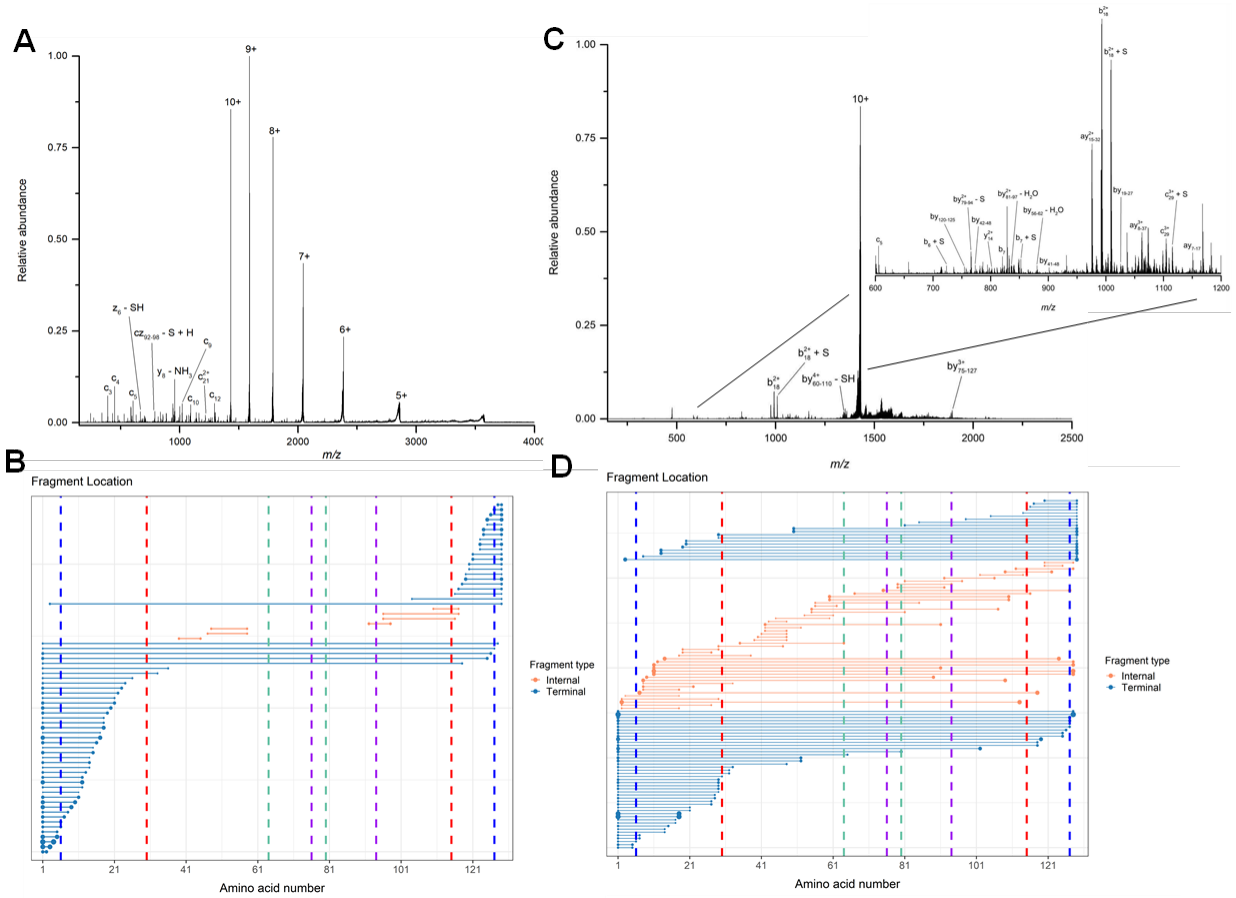


Figure S2. Data analysis of ECD and CAD MS/MS of lysozyme.

A. Representative ECD MS/MS spectrum of lysozyme, $[Lys + 10H]^{10+}$. **B.** Fragment location map indicating the region of the protein sequence covered by terminal and internal fragments for spectrum in panel A. **C.** Representative CAD MS/MS spectrum of lysozyme, $[Lys + 10H]^{10+}$. **D.** Fragment location map indicating the region of the protein sequence covered by terminal and internal fragments for spectrum in panel C.

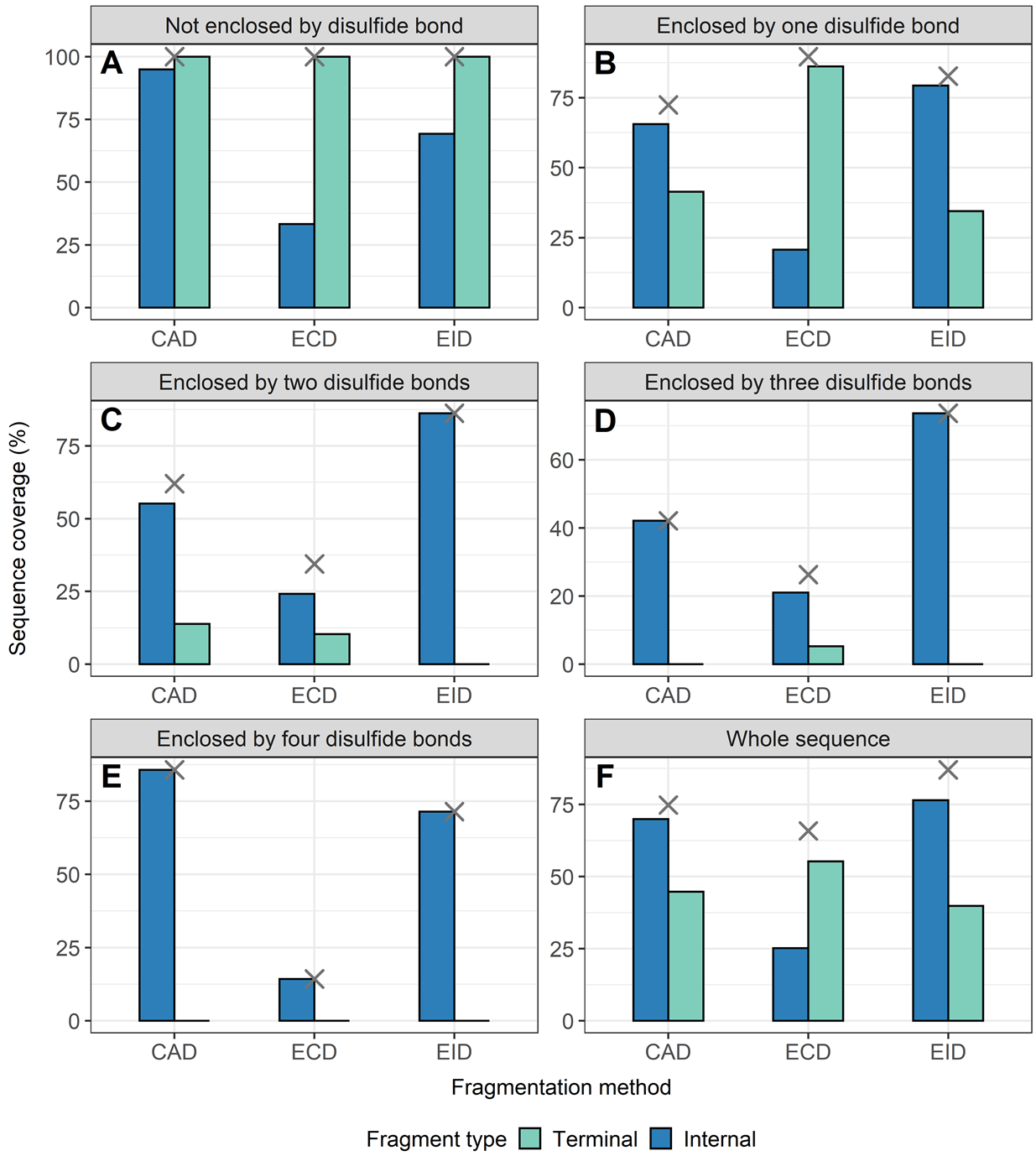


Figure S3. Sequence coverage analysis of ribonuclease A.

The extent of sequence information obtained by terminal and internal fragments for ribonuclease A at different sequence regions after integrating data from all seven charge states (8+ to 14+) for all three fragmentation methods (CAD, ECD, and EID) examined, **A.** sequence not enclosed by disulfide bond, **B.** sequence enclosed by one disulfide bond, **C.** sequence enclosed by two disulfide bonds, **D.** sequence enclosed by three disulfide bonds, **E.** sequence enclosed by four disulfide bonds, **F.** whole sequence.

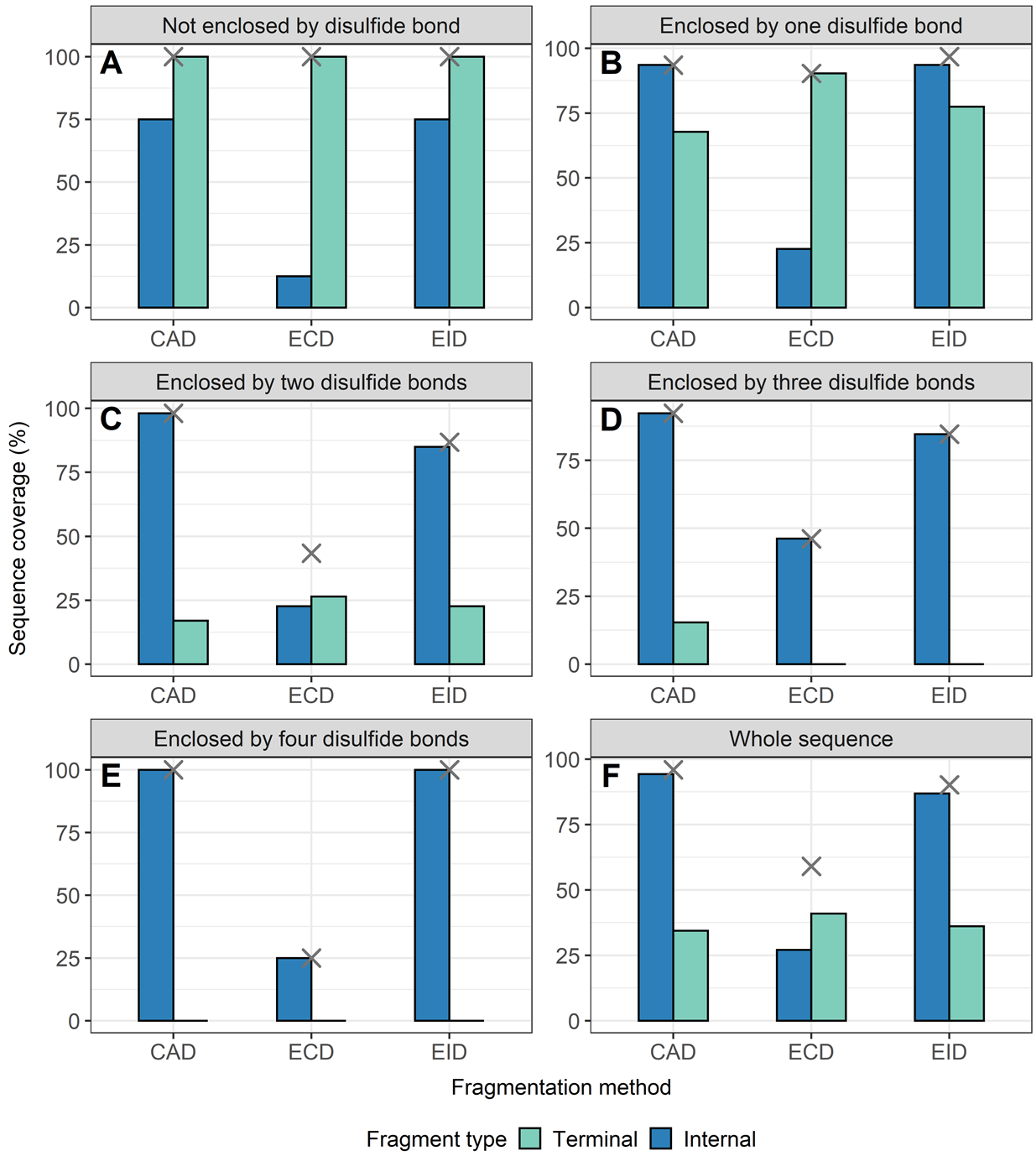


Figure S4. Sequence coverage analysis of α-lactalbumin.

The extent of sequence information obtained by terminal and internal fragments for alpha-lactalbumin at different sequence regions after integrating data from all four charge states (11+ to 14+) for all three fragmentation methods (CAD, ECD, and EID) examined, **A.** sequence not enclosed by disulfide bond, **B.** sequence enclosed by one disulfide bond, **C.** sequence enclosed by two disulfide bonds, **D.** sequence enclosed by three disulfide bonds, **E.** sequence enclosed by four disulfide bonds, **F.** whole sequence.

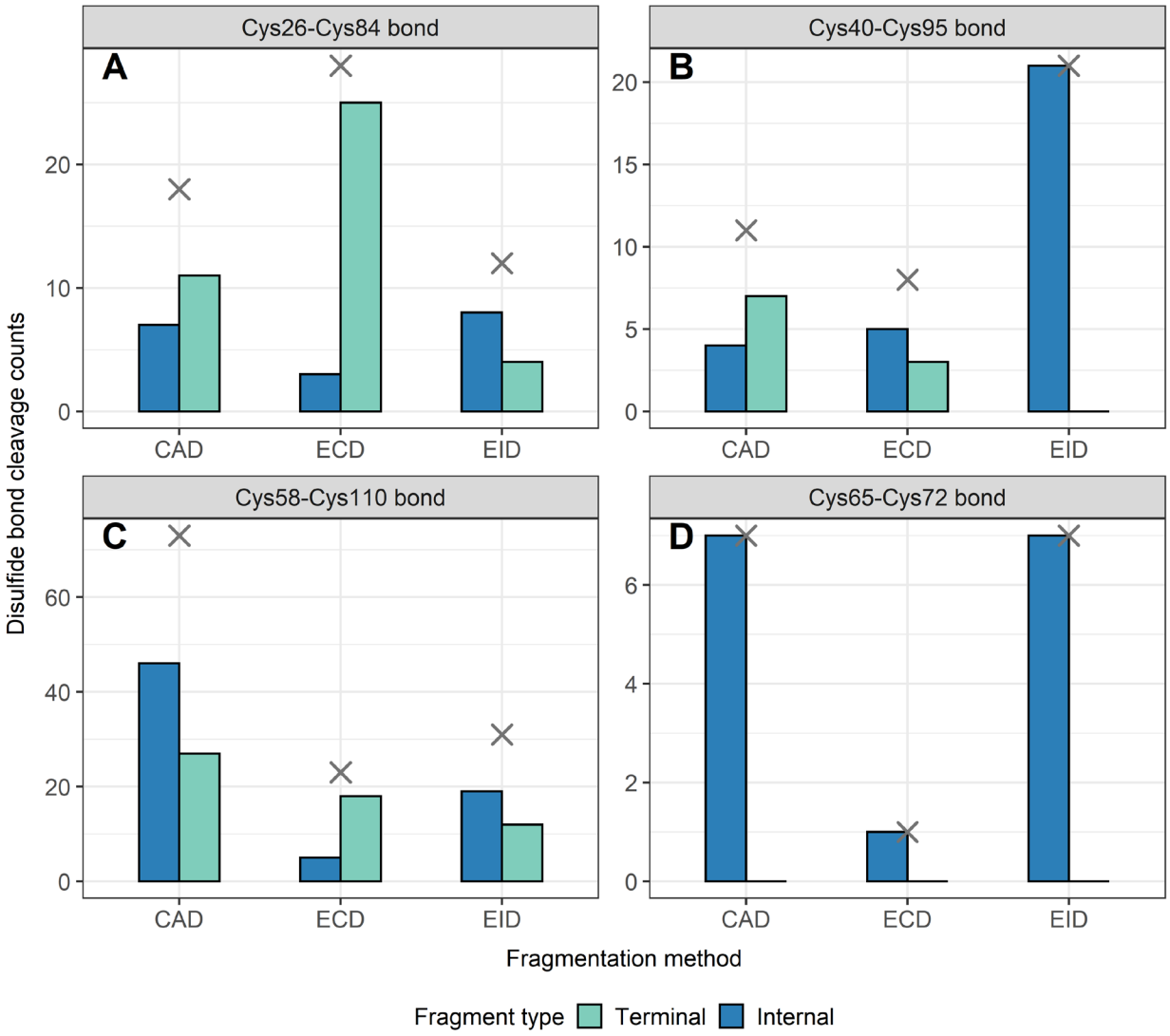


Figure S5. Disulfide bond cleavage analysis of ribonuclease A.
Number of disulfide bonds cleaved by terminal and internal fragments for ribonuclease A after integrating data from all seven charge states (8+ to 14+) for all three fragmentation methods (CAD, ECD, and EID) examined, **A.** Cys26-Cys84 bond, **B.** Cys40-Cys95 bond, **C.** Cys58-Cys110 bond, **D.** Cys65-Cys72 bond.

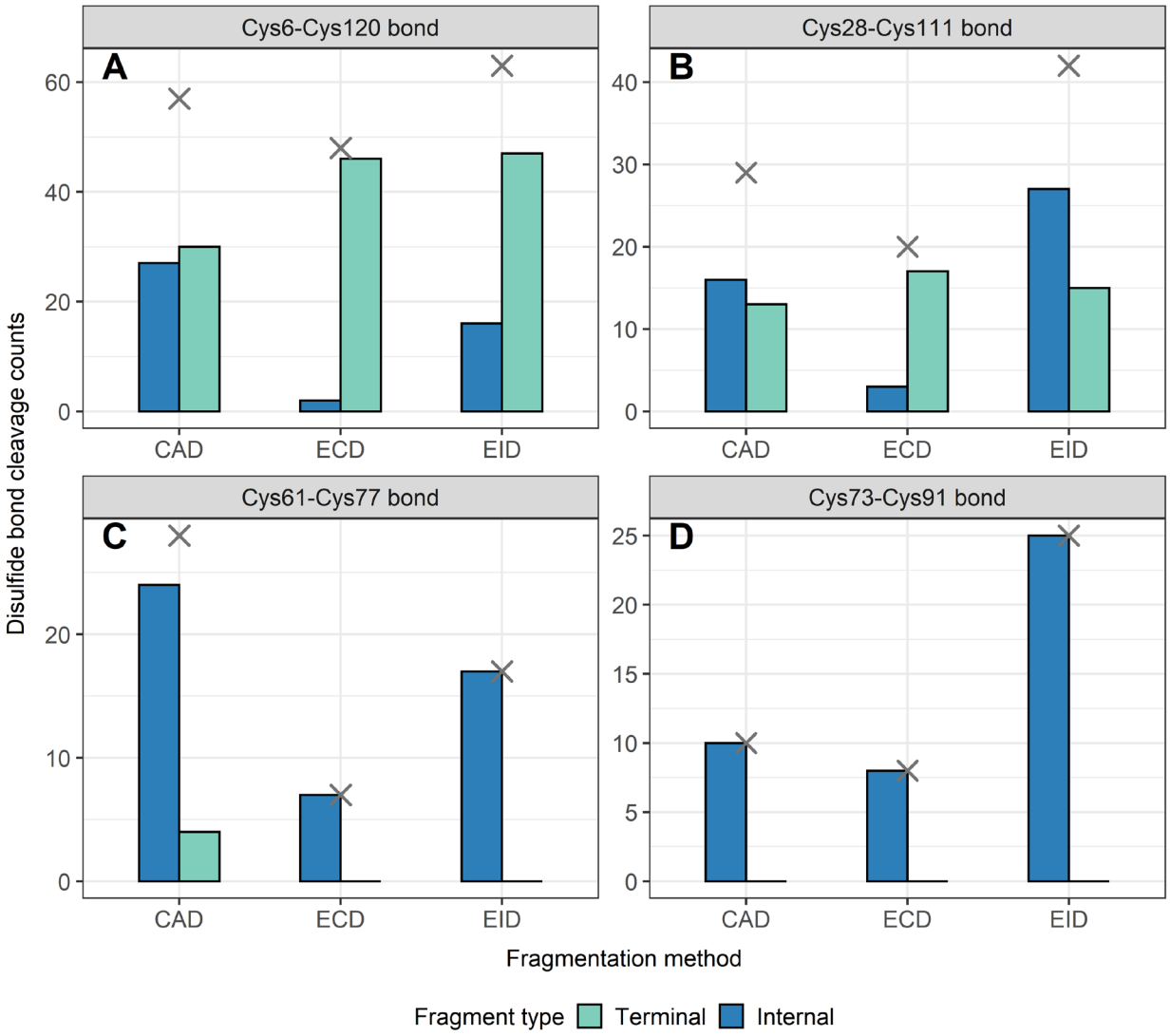


Figure S6. Disulfide bond cleavage analysis of α-lactalbumin.
Number of disulfide bonds cleaved by terminal and internal fragments for alpha-lactalbumin after integrating data from all four charge states (11+ to 14+) for all three fragmentation methods (CAD, ECD, and EID) examined, **A.** Cys6-Cys120 bond, **B.** Cys28-Cys111 bond, **C.** Cys61-Cys77 bond, **D.** Cys73-Cys91 bond.

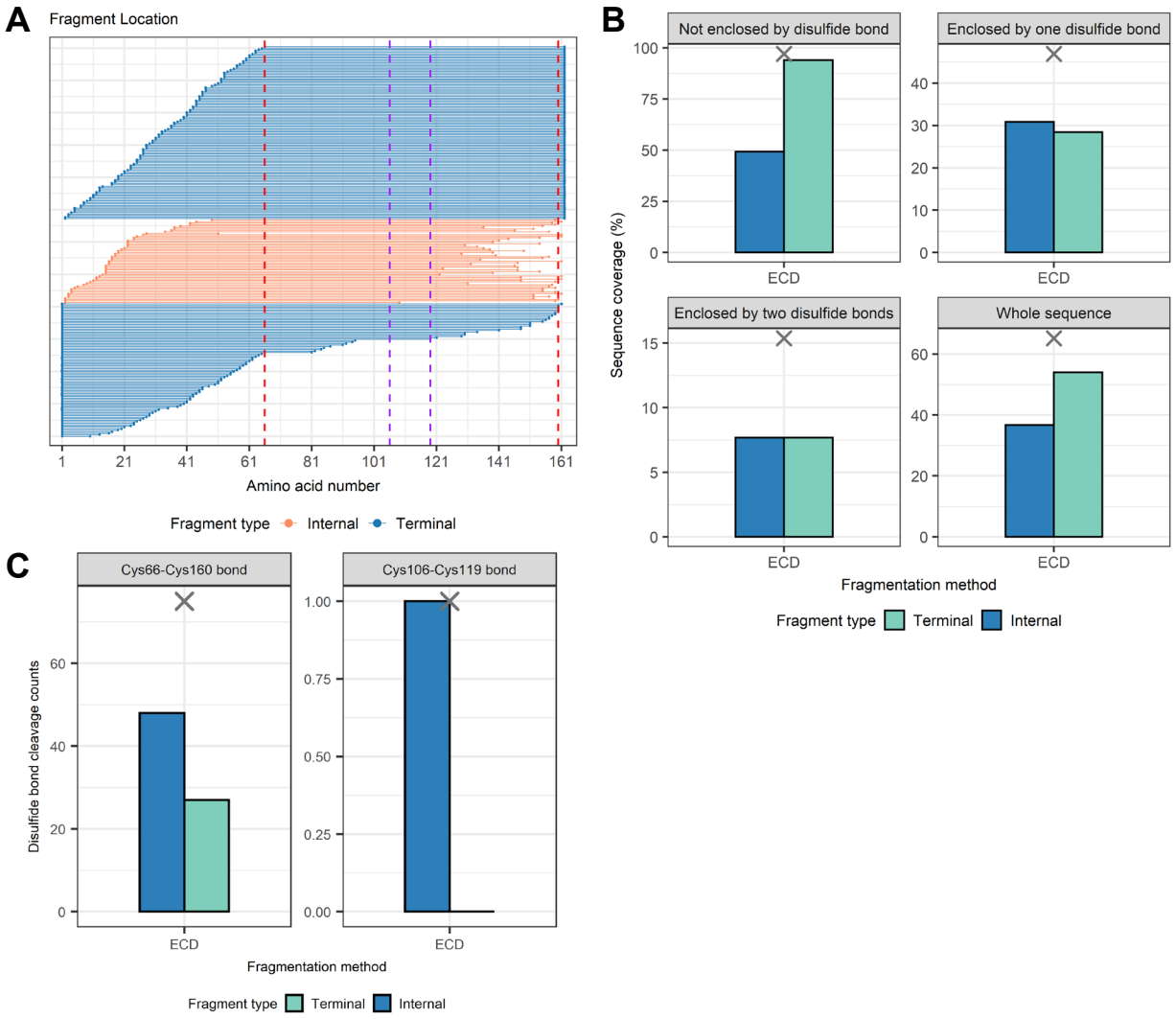


Figure S7. ECD MS/MS data of β-lactoglobulin, [Blac + 15H]$^{15+}$ acquired from Waters Select Series Cyclic IMS mass spectrometer.

A. fragment location map indicating the region of the protein sequence covered by terminal and internal fragments, **B.** the extent of sequence information obtained at different sequence regions by terminal and internal fragments, **C.** the number of disulfide bonds cleaved by terminal and internal fragments. Cross markers in panels B and C indicate the metric (sequence coverage or disulfide bond cleavage counts) after combining terminal and internal fragments.

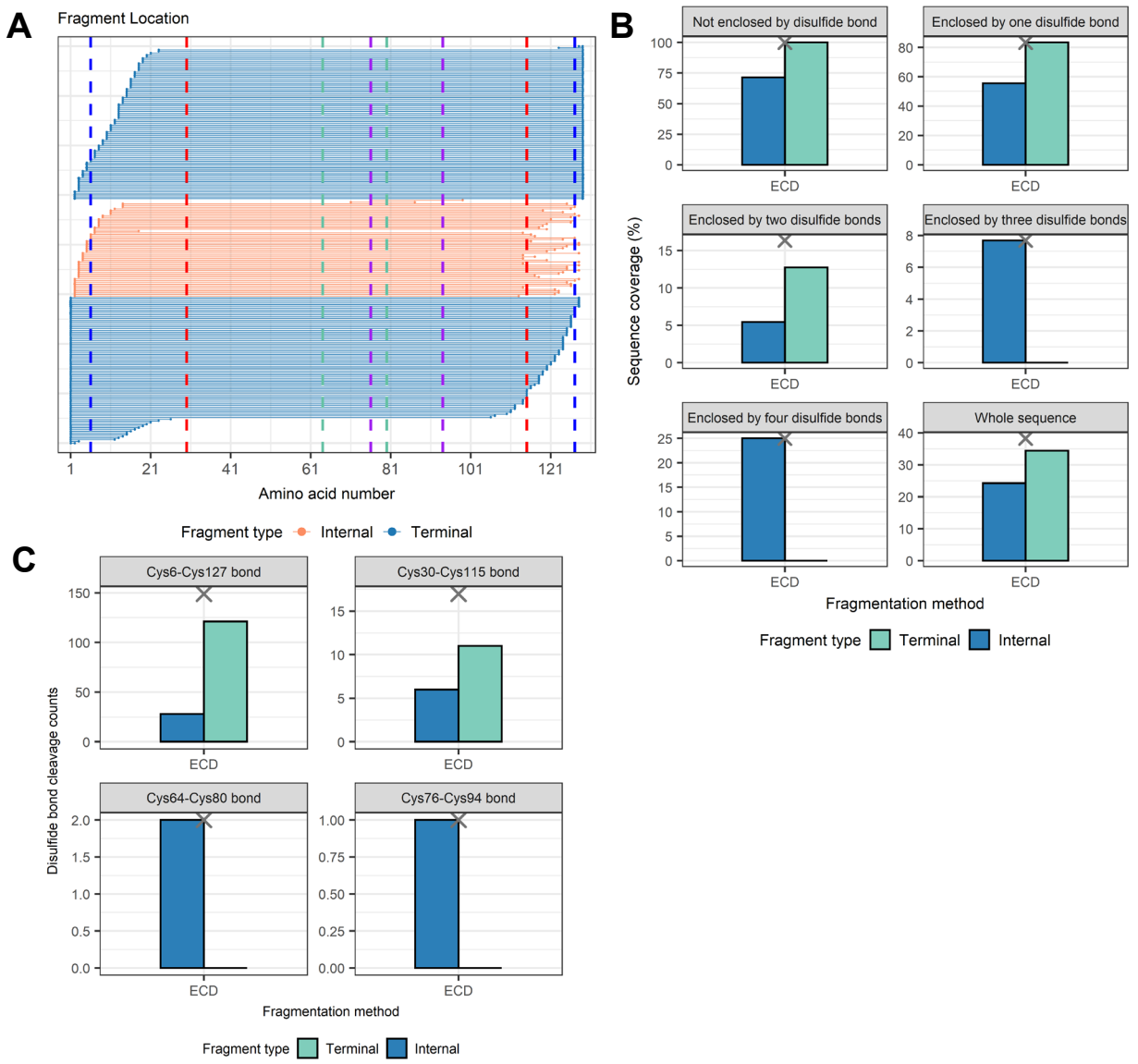


Figure S8. ECD MS/MS data of lysozyme, $[Lys + 9H]^{9+}$ acquired from Waters Select Series Cyclic IMS mass spectrometer.

A. fragment location map indicating the region of the protein sequence covered by terminal and internal fragments, **B.** the extent of sequence information obtained at different sequence regions by terminal and internal fragments, **C.** the number of disulfide bonds cleaved by terminal and internal fragments. Cross markers in panels B and C indicate the metric after combining terminal and internal fragments.

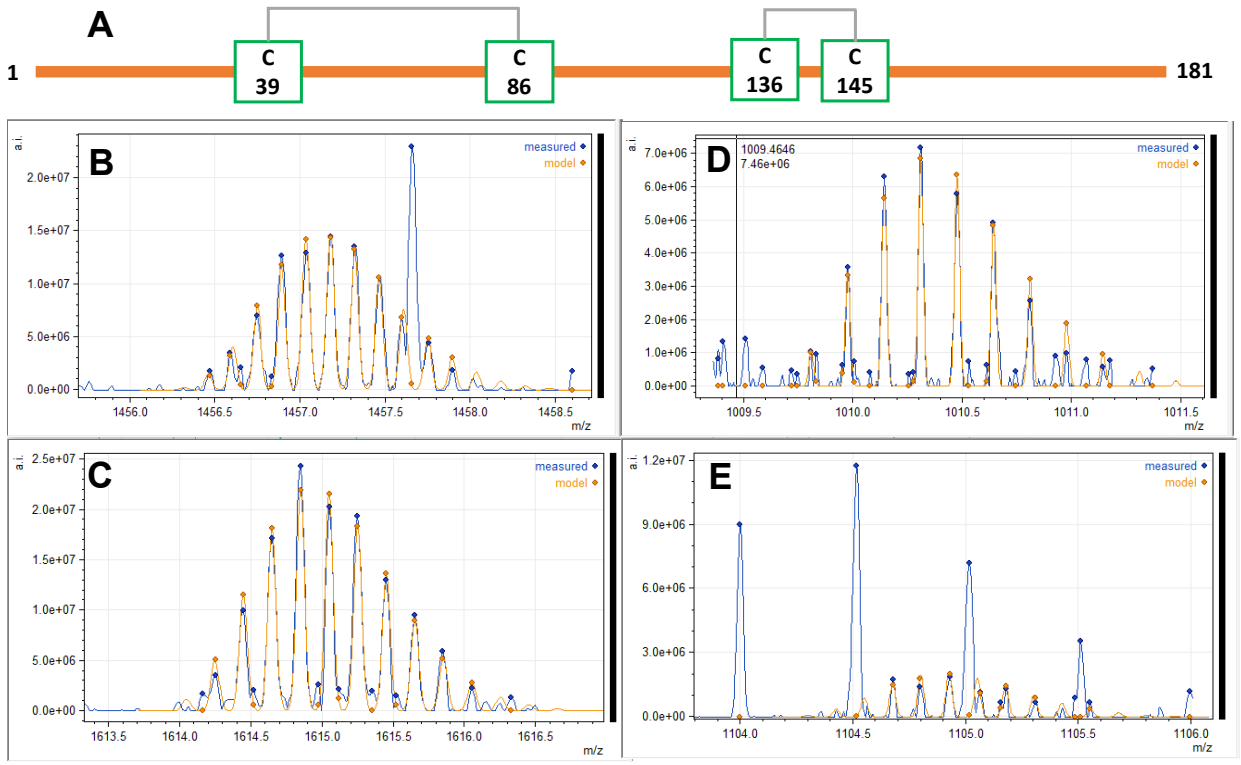


Figure S9. Disulfide connectivity analysis of trypsin inhibitor.

A. Disulfide bond connectivities of trypsin inhibitor (2 disulfide bonds). Isotope envelope fits of representative internal fragments traversing intact disulfide bonds generated by CAD of trypsin inhibitor, $[TI + 17H]^{17+}$ (Figure 6, panel A), **B.** $[by_{4\text{-}97} + 7H]^{7+}$ (0.332 ppm), **C.** $[by_{26\text{-}99} + 5H]^{5+}$ (-0.001 ppm), **D.** $[by_{100\text{-}153} + 6H]^{6+}$ (-0.112 ppm), **E.** $[by_{98\text{-}175} + 8H]^{8+}$ (-0.189 ppm). Internal fragments shown in panels B and C traverse the Cys39-Cys86 bond (the "purple" disulfide bond), while internal fragments shown in panels D and E traverse the Cys136-Cys145 bond (the "red" disulfide bond).

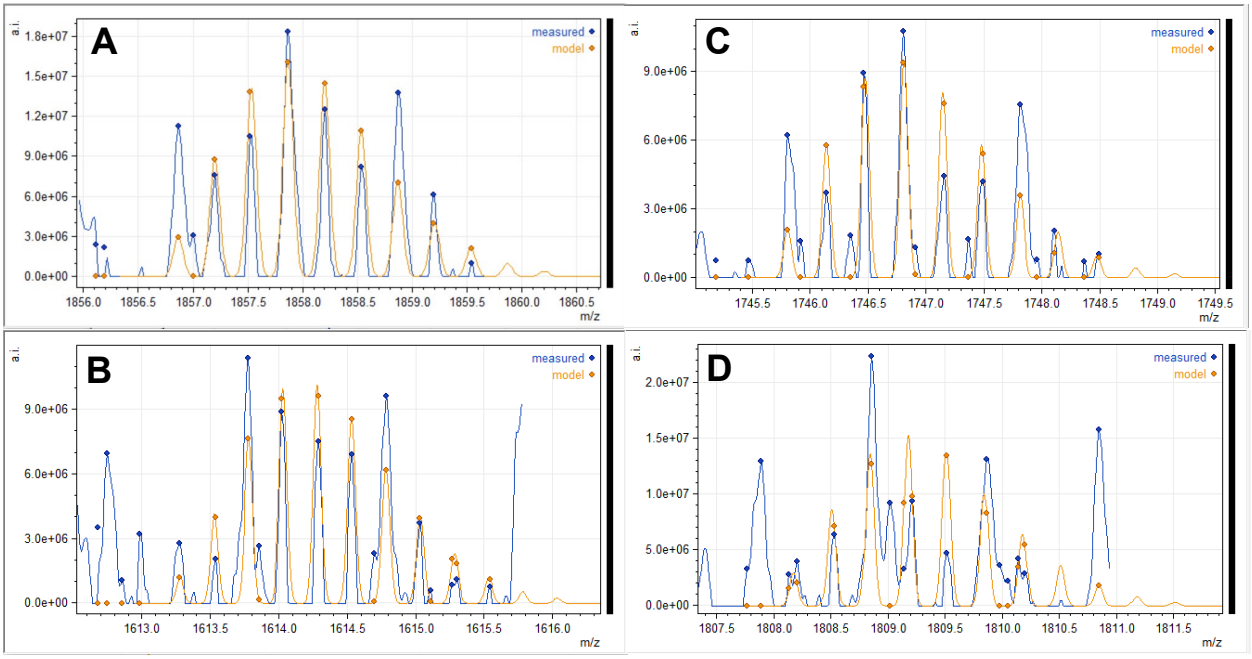


Figure S10. Isotope fits of selected internal fragments of α-lactalbumin that can determine its disulfide connectivity.

Isotope envelope fits of internal fragments traversing the two intact disulfide bonds in the interior of alpha-lactalbumin (Cys61-Cys77 bond and Cys73-Cys91 bond, the "green" and "purple" disulfide bonds) generated by CAD (Figure 6, panel C), **A.** $[by_{50\text{-}97} + 3H]^{3+}$ (-1.205 ppm), **B.** $[by_{51\text{-}106} + 4H]^{4+}$ (-1.969 ppm), **C.** $[by_{53\text{-}97} + 3H]^{3+}$ (1.634ppm), **D.** $[by_{60\text{-}106} + 3H]^{3+}$ (1.666ppm).

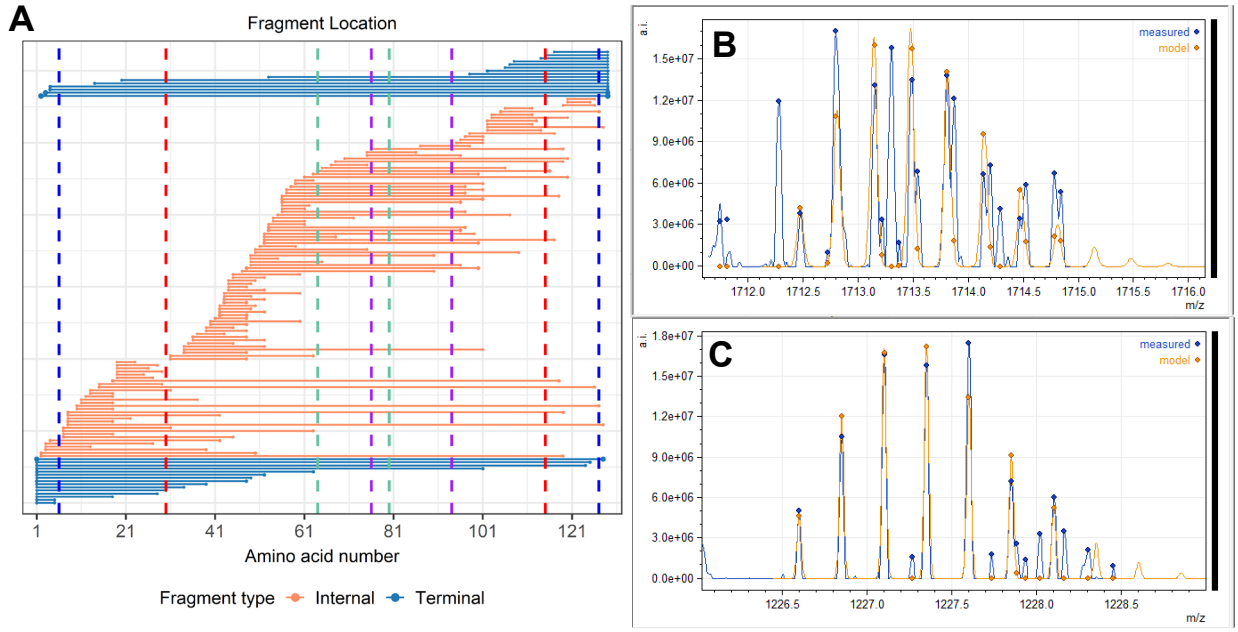


Figure S11. Disulfide connectivity analysis of lysozyme.

A. Fragment location map after importing a hydrogen loss localized modification on every cysteine, suggesting the integrity of every disulfide bond of CAD of lysozyme after integrating data from all five charge states examined (8+ to 12+). Isotope envelope fits of representative internal fragments traversing the two intact disulfide bonds located in the interior of lysozyme (Cys64-Cys80 bond and Cys76-Cys94 bond, the "green" and "purple" disulfide bonds) generated by CAD (panel A), **B.** $[by_{49-96} + 3H]^{3+}$ (-1.579 ppm), **C.** $[by_{53-97} + 4H]^{4+}$ (-0.064 ppm).

Supplementary Tables

Table S1. Unlocalized modifications for fragment matching of ExD of beta-lactoglobulin (2 disulfide bonds) to account for all disulfide-containing fragment ions.

Modification type	Mass shift (Da)	Amino acid residue	Number of residues required
H+	1.00783	Any	1
H-	-1.00783	C	1
2H-	-2.01566	C	2
3H-	-3.02349	C	4
4H-	-4.03132	C	5
1S	31.9721	C	1
2S	63.9442	C	2
SH-	-32.9799	C	1
2SH-	-65.9598	C	2
S-	-31.9721	C	1
2S-	-63.9442	C	2
H-S+	30.9643	C	1
2(H-S+)	61.9286	C	2
SH2-	-33.9878	C	1
2(SH2-)	-67.9755	C	2

Table S2. Unlocalized modifications for fragment matching of CAD of beta-lactoglobulin (2 disulfide bonds) to account for all disulfide-containing fragment ions.

Modification type	Mass shift (Da)	Amino acid residue	Number of residues required
2H2O-	-36.02113	Any	0
H2O-	-18.010565	Any	0
2NH3-	-34.053098	Any	0
NH3-	-17.026549	Any	0
H-	-1.00783	C	1
2H-	-2.01566	C	2
3H-	-3.02349	C	4
4H-	-4.03132	C	5
1S	31.9721	C	1
2S	63.9442	C	2
SH-	-32.9799	C	1
2SH-	-65.9598	C	2
S-	-31.9721	C	1
2S-	-63.9442	C	2
H-S+	30.9643	C	1
2(H-S+)	61.9286	C	2
SH2-	-33.9878	C	1
2(SH2-)	-67.9755	C	2

Table S3. Unlocalized modifications for fragment matching of ExD of lysozyme, ribonuclease A, and alpha-lactalbumin (all with 4 disulfide bonds) to account for all disulfide-containing fragment ions.

Modification type	Mass shift (Da)	Amino acid residue	Number of residues required
H+	1.00783	Any	1
H-	-1.00783	C	1
2H-	-2.01566	C	2
3H-	-3.02349	C	3
4H-	-4.03132	C	4
5H-	-5.03915	C	4
6H-	-6.04698	C	4
7H-	-7.05481	C	4
8H-	-8.06264	C	4
1S	31.9721	C	1
2S	63.9442	C	2
3S	95.9163	C	3
4S	127.8884	C	4
SH-	-32.9799	C	1
2SH-	-65.9598	C	2
3SH-	-98.9397	C	3
4SH-	-131.92	C	4
S-	-31.9721	C	1
2S-	-63.9442	C	2
3S-	-95.9163	C	3
4S-	-127.888	C	4
H-S+	30.9643	C	1
2(H-S+)	61.9286	C	2
3(H-S+)	92.8929	C	3
4(H-S+)	123.8572	C	4

Table S4. Unlocalized modification for fragment matching of CAD of lysozyme, ribonuclease A, and alpha-lactalbumin (all with 4 disulfide bonds) to account for all disulfide-containing fragment ions.

Modification type	Mass shift (Da)	Amino acid residue	Number of residues required
2H2O-	-36.0211	Any	0
H2O-	-18.0106	Any	0
2NH3-	-34.0531	Any	0
NH3-	-17.0265	Any	0
H-	-1.00783	C	1
2H-	-2.01566	C	2
3H-	-3.02349	C	3
4H-	-4.03132	C	4
5H-	-5.03915	C	4
6H-	-6.04698	C	4
7H-	-7.05481	C	4
8H-	-8.06264	C	4
1S	31.9721	C	1
2S	63.9442	C	2
3S	95.9163	C	3
4S	127.8884	C	4
SH-	-32.9799	C	1
2SH-	-65.9598	C	2
3SH-	-98.9397	C	3
4SH-	-131.92	C	4
S-	-31.9721	C	1
2S-	-63.9442	C	2
3S-	-95.9163	C	3
4S-	-127.888	C	4
H-S+	30.9643	C	1
2(H-S+)	61.9286	C	2
3(H-S+)	92.8929	C	3
4(H-S+)	123.8572	C	4

Table S5. Localized modifications for fragment matching of CAD of trypsin inhibitor, which applies one hydrogen loss on every cysteine to suggest the integrity of the disulfide bonds.

Modification type	Mass shift (Da)	Position of modified amino acid residue
H-	-1.00783	39
H-	-1.00783	86
H-	-1.00783	136
H-	-1.00783	145

Table S6. Localized modifications for fragment matching of CAD of alpha lactalbumin, which applies one hydrogen loss on every cysteine to suggest the integrity of the disulfide bonds.

Modification type	Mass shift (Da)	Position of modified amino acid residue
H-	-1.00783	6
H-	-1.00783	120
H-	-1.00783	28
H-	-1.00783	111
H-	-1.00783	61
H-	-1.00783	77
H-	-1.00783	73
H-	-1.00783	91

Table S7. Localized modifications for fragment matching of CAD of lysozyme, which applies one hydrogen loss on every cysteine to suggest the integrity of the disulfide bonds.

Modification type	Mass shift (Da)	Position of modified amino acid residue
H-	-1.00783	6
H-	-1.00783	127
H-	-1.00783	30
H-	-1.00783	115
H-	-1.00783	64
H-	-1.00783	80
H-	-1.00783	76
H-	-1.00783	94

References

1. M. Matsumura, G. Signor and B. W. Matthews, *Nature*, 1989, **342**, 291-293.

2. W. J. Wedemeyer, E. Welker, M. Narayan and H. A. Scheraga, *Biochem.*, 2000, **39**, 4207-4216.

3. P. J. Hogg, *Trends Biochem. Sci.*, 2003, **28**, 210-214.

4. M. V. Trivedi, J. S. Laurence and T. J. Siahaan, *Curr. Protein Pept. Sci.*, 2009, **10**, 614-625.

5. M. O. Glocker, B. Arbogast and M. L. Deinzer, *J. Am. Soc. Mass Spectrom.*, 1995, **6**, 638-643.

6. M. J. Feige, I. Braakman and L. M. Hendershot, in *Oxidative Folding of Proteins: Basic Principles, Cellular Regulation and Engineering*, The Royal Society of Chemistry, 2018, DOI: 10.1039/9781788013253-00001, pp. 1-33.

7. W. Zhang, L. A. Marzilli, J. C. Rouse and M. J. Czupryn, *Anal. Biochem.*, 2002, **311**, 1-9.

8. J. Wypych, M. Li, A. Guo, Z. Zhang, T. Martinez, M. J. Allen, S. Fodor, D. N. Kelner, G. C. Flynn, Y. D. Liu, P. V. Bondarenko, M. S. Ricci, T. M. Dillon and A. Balland, *J. Biol. Chem.*, 2008, **283**, 16194-16205.

9. S.-L. Wu, H. Jiang, Q. Lu, S. Dai, W. S. Hancock and B. L. Karger, *Anal. Chem.*, 2009, **81**, 112-122.

10. D. Bagal, J. F. Valliere-Douglass, A. Balland and P. D. Schnier, *Anal. Chem.*, 2010, **82**, 6751-6755.

11. G. Badescu, P. Bryant, M. Bird, K. Henseleit, J. Swierkosz, V. Parekh, R. Tommasi, E. Pawlisz, K. Jurlewicz, M. Farys, N. Camper, X. Sheng, M. Fisher, R. Grygorash, A. Kyle, A.

Abhilash, M. Frigerio, J. Edwards and A. Godwin, *Bioconjug. Chem.*, 2014, **25**, 1124-1136.

12. J. B. Shaw, W. Liu, Y. V. Vasil Ev, C. C. Bracken, N. Malhan, A. Guthals, J. S. Beckman and V. G. Voinov, *Anal. Chem.*, 2020, **92**, 766-773.

13. E. Deslignière, T. Botzanowski, H. Diemer, D. A. Cooper-Shepherd, E. Wagner-Rousset, O. Colas, G. Béchade, K. Giles, O. Hernandez-Alba, A. Beck and S. Cianférani, *J. Am. Soc. Mass Spectrom.*, 2021, DOI: 10.1021/jasms.1c00151.

14. L. Zhang, C. P. Chou and M. Moo-Young, *Biotechnol. Adv.*, 2011, **29**, 923-929.

15. M. Góngora-Benítez, J. Tulla-Puche and F. Albericio, *Chem. Rev.*, 2014, **114**, 901-926.

16. J. J. Gorman, T. P. Wallis and J. J. Pitt, *Mass Spectrom. Rev.*, 2002, **21**, 183-216.

17. P. L. Tsai, S.-F. Chen and S. Y. Huang, *Rev Anal Chem*, 2013, **32**.

18. J. Wiesner, A. Resemann, C. Evans, D. Suckau and W. Jabs, *Expert Rev. Proteom.*, 2015, **12**, 115-123.

19. M. M. Quick, C. M. Crittenden, J. A. Rosenberg and J. S. Brodbelt, *Anal. Chem.*, 2018, **90**, 8523-8530.

20. A. Walewska, J. J. Skalicky, D. R. Davis, M.-M. Zhang, E. Lopez-Vera, M. Watkins, T. S. Han, D. Yoshikami, B. M. Olivera and G. Bulaj, *J. Am. Chem. Soc.*, 2008, **130**, 14280-14286.

21. L. Poppe, J. O. Hui, J. Ligutti, J. K. Murray and P. D. Schnier, *Anal. Chem.*, 2012, **84**, 262-266.

22. A. Shevchenko, M. Wilm, O. Vorm and M. Mann, *Anal. Chem.*, 1996, **68**, 850-858.

23. S. E. Ong and M. Mann, *Nat. Chem. Biol.*, 2005, **1**, 252-262.

24. S. F. Foley, Y. Sun, T. S. Zheng and D. Wen, *Anal. Biochem.*, 2008, **377**, 95-104.

25. Y. Zhang, H. D. Dewald and H. Chen, *J. Proteome Res.*, 2011, **10**, 1293-1304.

26. D. S. Zhao, Z. R. Gregorich and Y. Ge, *Proteomics*, 2013, **13**, 3256-3260.

27. S. Nicolardi, M. Giera, P. Kooijman, A. Kraj, J.-P. Chervet, A. M. Deelder and Y. E. M. van der Burgt, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 1980-1987.

28. C. N. Cramer, K. F. Haselmann, J. V. Olsen and P. K. Nielsen, *Anal. Chem.*, 2016, **88**, 1585-1592.

29. C. N. Cramer, C. D. Kelstrup, J. V. Olsen, K. F. Haselmann and P. K. Nielsen, *Anal. Chem.*, 2017, **89**, 5949-5957.

30. P. Wongkongkathep, H. Li, X. Zhang, R. R. Ogorzalek Loo, R. R. Julian and J. A. Loo, *Int. J. Mass Spectrom.*, 2015, **390**, 137-145.

31. J. Zhang, R. R. O. Loo and J. A. Loo, *Int. J. Mass Spectrom.*, 2015, **377**, 546-556.

32. N. Siuti and N. L. Kelleher, *Nat. Methods*, 2007, **4**, 817-821.

33. Y. Zhang, W. Cui, H. Zhang, H. D. Dewald and H. Chen, *Anal. Chem.*, 2012, **84**, 3838-3842.

34. A. D. Catherman, O. S. Skinner and N. L. Kelleher, *Biochem. Biophys. Res. Commun.*, 2014, **445**, 683-693.

35. F. Lermyte, Y. O. Tsybin, P. B. O'Connor and J. A. Loo, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 1149-1157.

36. T. K. Toby, L. Fornelli and N. L. Kelleher, *Annu Rev Anal Chem (Palo Alto Calif)*, 2016, **9**, 499-519.

37. A. J. Kleinnijenhuis, M. C. Duursma, E. Breukink, R. M. A. Heeren and A. J. R. Heck, *Anal. Chem.*, 2003, **75**, 3219-3225.

38. H. Lioe and R. A. J. OrsHair, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 1109-1123.

39. J. Chen, P. Shiyanov, L. Zhang, J. J. Schlager and K. B. Green-Church, *Anal. Chem.*, 2010, **82**, 6079-6089.

40. N. A. K. Roman A. Zubarev, Einar K. Fridriksson, Mark A. Lewis,David M. Horn, Barry K. Carpenter, and Fred W. McLafferty, *J. Am. Chem. Soc.*, 1999, **121**, 2857-2862.

41. B. Ganisl and K. Breuker, *ChemistryOpen*, 2012, **1**, 260-268.

42. S. R. Cole, X. Ma, X. Zhang and Y. Xia, *J. Am. Soc. Mass Spectrom.*, 2012, **23**, 310-320.

43. D. Wen, Y. Xiao, M. M. Vecchi, B. J. Gong, J. Dolnikova and R. B. Pepinsky, *Anal. Chem.*, 2017, **89**, 4021-4030.

44. Y. M. E. Fung, F. Kjeldsen, O. A. Silivra, T. W. D. Chan and R. A. Zubarev, *Angew. Chem. Int. Ed.*, 2005, **44**, 6399-6403.

45. A. Agarwal, J. K. Diedrich and R. R. Julian, *Anal. Chem.*, 2011, **83**, 6455-6458.

46. J. Bonner, L. E. Talbert, N. Akkawi and R. R. Julian, *Analyst*, 2018, **143**, 5176-5184.

47. L. E. Talbert and R. R. Julian, *J. Am. Soc. Mass Spectrom.*, 2018, **29**, 1760-1767.

48. S. K. Gammelgaard, S. B. Petersen, K. F. Haselmann and P. K. Nielsen, *ACS Omega*, 2020, **5**, 7962-7968.

49. M. J. P. Rush, N. M. Riley, M. S. Westphall and J. J. Coon, *Anal. Chem.*, 2018, **90**, 8946-8953.

50. S. K. Gammelgaard, S. B. Petersen, K. F. Haselmann and P. K. Nielsen, *J. Am. Soc. Mass Spectrom.*, 2020, DOI: 10.1021/jasms.0c00257.

51. K. D. Ballard and S. J. Gaskell, *Int. J. Mass Spectrom. Ion Processes*, 1991, **111**, 173-189.

52. V. H. Wysocki, K. A. Resing, Q. Zhang and G. Cheng, *Methods*, 2005, **35**, 211-222.

53. C. Lantz, M. A. Zenaidee, B. Wei, Z. Hemminger, R. R. Ogorzalek Loo and J. A. Loo, *J. Proteome Res.*, 2021, **20**, 1928–1935.

54. P. E. Barran, N. C. Polfer, D. J. Campopiano, D. J. Clarke, P. R. R. Langridge-Smith, R. J. Langley, J. R. W. Govan, A. Maxwell, J. R. Dorin, R. P. Millar and M. T. Bowers, *Int. J. Mass Spectrom.*, 2005, **240**, 273-284.

55. A. Michalski, N. Neuhauser, J. Cox and M. Mann, *J. Proteome Res.*, 2012, **11**, 5479-5491.

56. Y. A. Lyon, D. Riggs, L. Fornelli, P. D. Compton and R. R. Julian, *J. Am. Soc. Mass Spectrom.*, 2018, **29**, 150-157.

57. B. Wei, M. A. Zenaidee, C. Lantz, R. R. Ogorzalek Loo and J. A. Loo, *Anal. Chim. Acta*, 2022, **1194**.

58. J. S. Cobb, M. L. Easterling and J. N. Agar, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 949-959.

59. K. R. Durbin, O. S. Skinner, R. T. Fellers and N. L. Kelleher, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 782-787.

60. J. Chen, P. Shiyanov and K. B. Green, *J. Mass Spectrom.*, 2019, **54**, 527-539.

61. F. Griaud, B. Denefeld, C.-Y. Kao-Scharf, J. Dayer, M. Lang, J.-Y. Chen and M. Berg, *Anal. Chem.*, 2019, **91**, 8845-8852.

62. M. A. Zenaidee, C. Lantz, T. Perkins, W. Jung, R. R. O. Loo and J. A. Loo, *J. Am. Soc. Mass Spectrom.*, 2020, **31**, 1896–1902.

63. N. D. Schmitt, J. M. Berger, J. B. Conway and J. N. Agar, *Anal. Chem.*, 2021, **93**, 6355-6362.

64. M. A. Zenaidee, B. Wei, C. Lantz, H. T. Wu, T. R. Lambeth, J. K. Diedrich, R. R. Ogorzalek Loo, R. R. Julian and J. A. Loo, *J. Am. Soc. Mass Spectrom.*, 2021, **32**, 1752–1758.

65. H. Li, Y. Sheng, W. Mcgee, M. Cammarata, D. Holden and J. A. Loo, *Anal. Chem.*, 2017, **89**, 2731-2738.

66. H. Li, H. H. Nguyen, R. R. Ogorzalek Loo, I. D. G. Campuzano and J. A. Loo, *Nat. Chem.*, 2018, **10**, 139-148.

67. Z. Rolfs and L. M. Smith, *J. Proteome Res.*, 2021, DOI: 10.1021/acs.jproteome.1c00599.

68. S. Chin, T. Chen, R. N. Hannoush and C. M. Crittenden, *J. Pharm. Biomed. Anal.*, 2021, **195**, 113893.

69. B. Paizs and S. Suhai, *J. Am. Soc. Mass Spectrom.*, 2004, **15**, 103–113.

Chapter 4: Added Value of Internal Fragments for Top-Down Mass Spectrometry of Intact Monoclonal Antibodies and Antibody−Drug Conjugates

Reprinted with permission from

Wei, B.; Lantz, C.; Liu, W.; Viner, R.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A., Added Value of Internal Fragments for Top-Down Mass Spectrometry of Intact Monoclonal Antibodies and Antibody–Drug Conjugates. *Analytical Chemistry* **2023,** *95* (24), 9347–9356. DOI: https://doi.org/10.1021/acs.analchem.3c01426.

Benqian Wei[a], Carter Lantz[a], Weijing Liu[c], Rosa Viner[c], Rachel R. Ogorzalek Loo[a], Iain D.G. Campuzano[d], Joseph A. Loo[a, b, *]

[a] Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA; [b] Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA; [c] Thermo Fisher Scientific, San Jose, CA, USA; [d] Center for Research Acceleration by Digital Innovation, Molecular Analytics, Amgen Research, Thousand Oaks, CA, USA.

* Corresponding Author: Joseph A. Loo, Department of Chemistry and Biochemistry, Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA, E-mail: jloo@chem.ucla.edu.

Abstract

Monoclonal antibodies (mAbs) and antibody−drug conjugates (ADCs) are two of the most important therapeutic drug classes that require extensive characterization, whereas their large size and structural complexity make them challenging to characterize and demand the use of advanced analytical methods. Top-down mass spectrometry (TD-MS) is an emerging technique that minimizes sample preparation and preserves endogenous post-translational modifications (PTMs); however, TD-MS of large proteins suffers from low fragmentation efficiency, limiting the sequence and structure information that can be obtained. Here, we show that including the assignment of internal fragments in native TD-MS of an intact mAb and an ADC can improve their molecular characterization. For the NIST mAb, internal fragments can access the sequence region constrained by disulfide bonds to increase the TD-MS sequence coverage to over 75%. Important PTM information, including intrachain disulfide connectivity and N-glycosylation sites, can be revealed after including internal fragments. For a heterogeneous lysine-linked ADC, we show that assigning internal fragments improves the identification of drug conjugation sites to achieve a coverage of 58% of all putative conjugation sites. This proof-of principle study demonstrates the potential value of including internal fragments in native TD-MS of intact mAbs and ADCs, and this analytical strategy can be extended to bottom-up and middle-down MS approaches to achieve even more comprehensive characterization of important therapeutic molecules.

1. Introduction

Monoclonal antibody (mAb) therapeutics have become increasingly important for the diagnosis

and treatment of a host of diseases including cancer and viral infections. They can achieve targeted tumor cell elimination with high specificity and desirable pharmacokinetics properties.[1−7] MAbs are highly complex molecules with large size (∼150 kDa) and have a series of intra- and interchain disulfide bridges and numerous post-translational modifications (PTMs), with the most common ones being N-glycosylation, N-terminal pyroglutamine cyclization, oxidation, C-terminal lysine processing, and deamidation.[8,9] This high molecular complexity can impact critical quality attributes (CQAs)[10] of mAb products including stability, solubility, and pharmacokinetics/pharmacodynamics properties; thus, extensive sequence and structure characterization is required to produce high quality mAb products.[1]

Antibody−drug conjugates (ADCs), which arm the antibodies with highly potent cytotoxic payloads via a linker to improve its antitumor efficacy, have emerged as a promising therapeutic drug class.[12−16] The conjugation of a linker and a payload onto antibodies introduces an additional dimension of heterogeneity to ADCs, increasing the challenge of their complete characterization. This is particularly true for nonspecific lysine-linked ADCs, in which payloads are conjugated with primary amines (lysines and N-termini) of the antibody, resulting in a highly heterogeneous molecule with various numbers of payloads binding to a large array of locations.[12,17−21] Comprehensive analytical profiling of ADCs include evaluating CQAs such as drug-to-antibody ratio (DAR), drug distribution, and drug conjugation sites.[22−30] In contrast to the routinely characterized DAR and drug distribution, relatively few studies have focused on determining drug conjugation sites, despite their important role in affecting the physical and pharmaceutical properties of ADCs.[17,26,31,32] For example, the binding specificity of lysine-linked

ADCs to the target antigen can be affected if the conjugation occurs in the complementarity-determining regions (CDRs).[12,17,33] Such instances necessitate the determination of drug conjugation sites, particularly for nonspecific lysine-linked ADCs.

Mass spectrometry (MS) based techniques, such as bottom-up MS (BU-MS) and middle-down MS (MD-MS), are powerful analytical tools routinely used for the characterization of mAbs and ADCs. BU-MS, or peptide mapping, analyzes enzymatically digested peptides of mAbs/ADCs using liquid chromatography−tandem mass spectrometry (LC−MS/ MS).[12,34−39] While BU-MS can provide high sequence coverage with amino acid resolution and pinpoint ADC drug conjugation sites (by identifying payload-bound peptide ions), it comes at a cost of relatively extensive sample preparation and the possibility of introducing artificial PTMs.[40−42] MD-MS, which analyzes ∼25 kDa subunits of mAbs and ADCs by reducing disulfide bonds and cleaving the hinge region of the antibody heavy chain, has become a promising complementary approach to BU-MS.[43−50] MD-MS does not reach the extensiveness of BU-MS in terms of sequence and drug conjugation site coverage; it avoids the digestion step required by BU-MS, although enzymatic and chemical reduction and chromatographic separation are still necessary.

Top-down MS (TD-MS), where intact gas-phase protein ions are measured and fragmented, has gained popularity in recent years for the characterization of mAbs.[43,51−56] Compared to BU-MS and MD-MS, TD-MS holds the advantages of minimal sample preparation and preserving endogenous modifications of mAbs. However, TD-MS suffers from low fragmentation efficiency for proteins of the size of mAbs and for proteins with significant disulfide bond compositions.[53,54,57,58] Recently, the Coon lab utilized activated ion electron transfer dissociation

(AI-ETD), a novel fragmentation technique that combines the advantage of electron- and photon-based fragmentation, to achieve over 60% sequence coverage on intact NIST mAb by TD-MS alone.[54] Although this is a substantial improvement over previous TD-MS studies, comprehensive TD-MS sequence coverage on intact mAbs is still challenging. Studies utilizing TD-MS to identify drug conjugation sites of intact ADCs are even more sparse. The Ge lab applied a three-tier TD-MS strategy for multiattribute analysis of a site-specific cysteine-linked ADC; however, limited sequence and drug conjugation site information is obtained due to relatively low fragmentation efficiency.[59] Thus far, BU-MS and MD-MS are still the preferred methods for the characterization of either cysteine-linked ADCs[22,48−50,60] or lysine-linked ADCs.[12,17,35,37,47]

A strategy to improve the apparent fragmentation efficiency of TD-MS, and thereby increase sequence information content, is to incorporate noncanonical internal fragments, which contain neither the N- nor C-terminus of the protein sequence, into data analysis workflow.[61] Previous studies have shown that including internal fragment assignments in TD-MS can enhance the sequence information obtained from intact proteins,[62−67] from protein complexes,[68,69] and on a proteome-wide scale.[70] Specifically, internal fragments have been demonstrated to aid the TD-MS characterization of disulfide-intact proteins,[64,67,71] inspiring us to investigate the application of internal fragments in TD-MS of mAbs and ADCs, which contain a large number of disulfide bonds.

Here, we show that assigning internal fragments in TD-MS increases mAb sequence coverage to over 75% and allows the determination of intrachain disulfide connectivity and

various N-glycosylation types. For a therapeutic nonspecific lysine- linked ADC, identification of nearly 60% of all putative drug conjugation sites was achieved. To our knowledge, the sequence coverage reported here on intact mAbs is the highest achieved by TD-MS alone, and this is the first report utilizing TD-MS to characterize lysine-linked ADCs.

2. Experimental Section

2.1. Materials and sample preparation.

The humanized IgG1k monoclonal antibody reference material 8671 was purchased from the National Institute of Standards and Technology (NIST, Gaithersburg, MD). IgG1-DM1 ADC was supplied by Amgen; the mAb is comprised of two human heavy chains and two human λ light chains, and was expressed in Chinese hamster ovary cells and purified by chromatographic procedures developed at Amgen.[72] Detailed procedures for the preparation of the ADC have been described previously.[12] Briefly, a maytansinoid DM1 payload was conjugated onto primary amines of the naked IgG1 mAb through a noncleavable linker, N-succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) (Figure S1). Ammonium acetate solution (7.5 M) was purchased from Sigma-Aldrich (St. Louis, MO, USA) and diluted to 200 mM. The NIST mAb and ADC samples were buffer exchanged into 200 mM ammonium acetate using Biospin 6 columns (BioRad) and diluted to a final concentration of 5 μM prior to native mass spectrometry measurements.

2.2. Native Top-Down Mass Spectrometry.

All samples were measured using a Thermo Q Exactive Plus UHMR Orbitrap instrument

(Thermo Fisher Scientific, Bremen, Germany) modified with an electromagnetostatic ExD cell (e-MSion Inc., Corvallis, OR) located between the quadrupole and C-Trap. All protein solutions were loaded into in-house pulled capillaries coated with platinum and electrosprayed by applying a capillary voltage between 1.1 and 1.7 kV on the NSI source. The source temperature was set at 250 °C, and the S-lens RF level was set at 200. Other crucial instrument parameters corresponding to ion transmission were listed in Table S1. For HCD fragmentation, five individual charge states of NIST mAb (22+ to 26+) and four individual ions of ADC (DAR 1 ions at charge states 23+ and 24+, DAR 2 ions at charge states 23+ and 24+) were isolated in the quadrupole, with an isolation window of 40 *m/z* before fragmentation. HCD was performed by applying CE ranging from 190 to 240 V to achieve optimal fragmentation. For ECD fragmentation, the aforementioned five NIST mAb ions and four ADC ions were still isolated with an isolation window of 40 *m/z* in the quadrupole, while two additional ADC ions, DAR 1 and DAR 2 ions grouped together at charge states 22+ and 25+, were isolated with an isolation window of 80 *m/z* in the quadrupole due to lower signal level of these two charge states. After isolation, these ions were transmitted through the ExD cell into the HCD cell in the absence of electrons, where ECD was occurring. A set of seven voltage parameters of the ExD cell controlling the emitting and confinement of electrons were optimized to ensure efficient electron capture by the protein ions in the HCD cell (Table S1). Post-ECD collisional activation was applied by setting CE values between 150 and 200 V to minimize the effect of electron capture without dissociation (ECnoD).57 All HCD and ECD MS/MS spectra were collected with the noise threshold set at 3, a resolution of 200 000 at m/z 400, AGC target of 1e6, and maximum inject

time of 200 ms. Between 100 and 200 scans were averaged for each spectrum.

2.3. Data Analysis.

2.3.1. Data Processing and Fragment Assignment.

Raw MS/MS spectra were deconvoluted using Thermo BioPharma Finder 5.0 (Xtract algorithm), and deconvoluted mass lists were exported as .csv files for fragment assignment in ClipsMS.61 The mass tolerance was set at 3 ppm, and the smallest internal fragment size was set at 5 amino acids. For sequence coverage analysis of NIST mAb, modifications relating to disulfide bond cleavages and hydrogen gains expected for ECD were considered. Similarly, for disulfide connectivity analysis of NIST mAb, modifications applying one hydrogen loss on each cysteine forming intrachain disulfide bonds to suggest the integrity of the disulfide bond were included. For drug conjugation sites analysis of ADC, modifications considering one or two intact DM1 conjugations were searched for, depending on the identity of the isolated precursor ion (DAR 1 or DAR 2 ions). Water and ammonia losses were considered for HCD fragmentation, and glycosylations including G0F, G1F, and G2F were considered for heavy chain fragments. Four terminal fragment types including *b*, *c*, *y*, *z* were searched for ECD, while only b and y terminal fragments were searched for HCD. As for internal fragment matching, only *by* internal fragments were searched for HCD and *cz* internal fragments for ECD. All terminal fragments were assigned before considering internal fragments, and all overlapping internal fragments due to the arrangement and/ or frameshift ambiguity[64] were removed. The fragment search was done separately for light chain and heavy chain for a single TD-MS data set, i.e., one TD-MS spectrum was searched once against the light chain sequence and once against the heavy

chain sequence. For overlapping assignments shared by the light chain and heavy chain, two factors were considered to remove these duplicates: (1) If a deconvoluted mass is assigned as a terminal fragment for one chain while as an internal fragment for the other chain (both within 3 ppm), the terminal fragment assignment was retained. (2) If a deconvoluted mass is assigned as an internal fragment for either chain, the one with the lower ppm error was retained. After fragment matching and duplicates removal, all assigned internal fragments were further refined by a two-step manual validation process: (1) Examine the isotopic profile of every assigned internal fragment to eliminate poorly fitted uncertain assignments. (2) Compare the assignment results with the theoretical fragment lists generated by ProteinProspector v 6.4.273 to eliminate any possible overlap between assigned internal fragments and theoretical terminal fragments together with their possible modifications such as neutral losses. The assignment results for every isolated charge state of NIST mAb were combined, as well as for every isolated ion of ADC after manual validation.

2.3.2. Protein Sequence Coverage.

Protein sequence coverage is calculated by the number of observed inter-residue cleavage sites divided by the total number of possible inter-residue cleavage sites on the protein backbone.

2.3.3. Connectivity Mismatch Rate.

S−S connectivity mismatch rate is calculated by dividing the number of fragments that cause mismatched S−S connectivity (orange-colored fragments in Figures 4 and S3) by the total number of fragments that can determine S−S connectivity (green-colored fragments in Figures 4

and S3) and fragments that cause mismatched S−S connectivity.

3. Results and Discussion

3.1. Internal Fragments Increase the Sequence Coverage of Intact mAbs.

Native MS provides a global overview of the composition of intact mAbs, including major glycoforms (Figure 1A and B); however, their characterization requires comprehensive sequence analysis. Assigning internal fragments can significantly enhance the sequence information obtained from intact mAbs. To demonstrate this, two fragmentation methods, ECD and HCD, were applied on the five most abundant charge states of intact NIST mAb (22+ to 26+). An example ECD spectrum of the intact NIST mAb is shown in Figure S2. ECD of [NIST + 25H]$^{25+}$ precursor ion generated many informative product ions along with charge reduced precursor ions (Figure S2A and B). Terminal and internal fragments from both heavy and light chains are observed, demonstrating that more information can be obtained by assigning these previously ignored signals (Figure S2B). Importantly, internal fragments provide complementary sequence information to terminal fragments, spanning most of the interior sequence of both chains that cannot be reached by terminal fragments (Figure S2C and D).

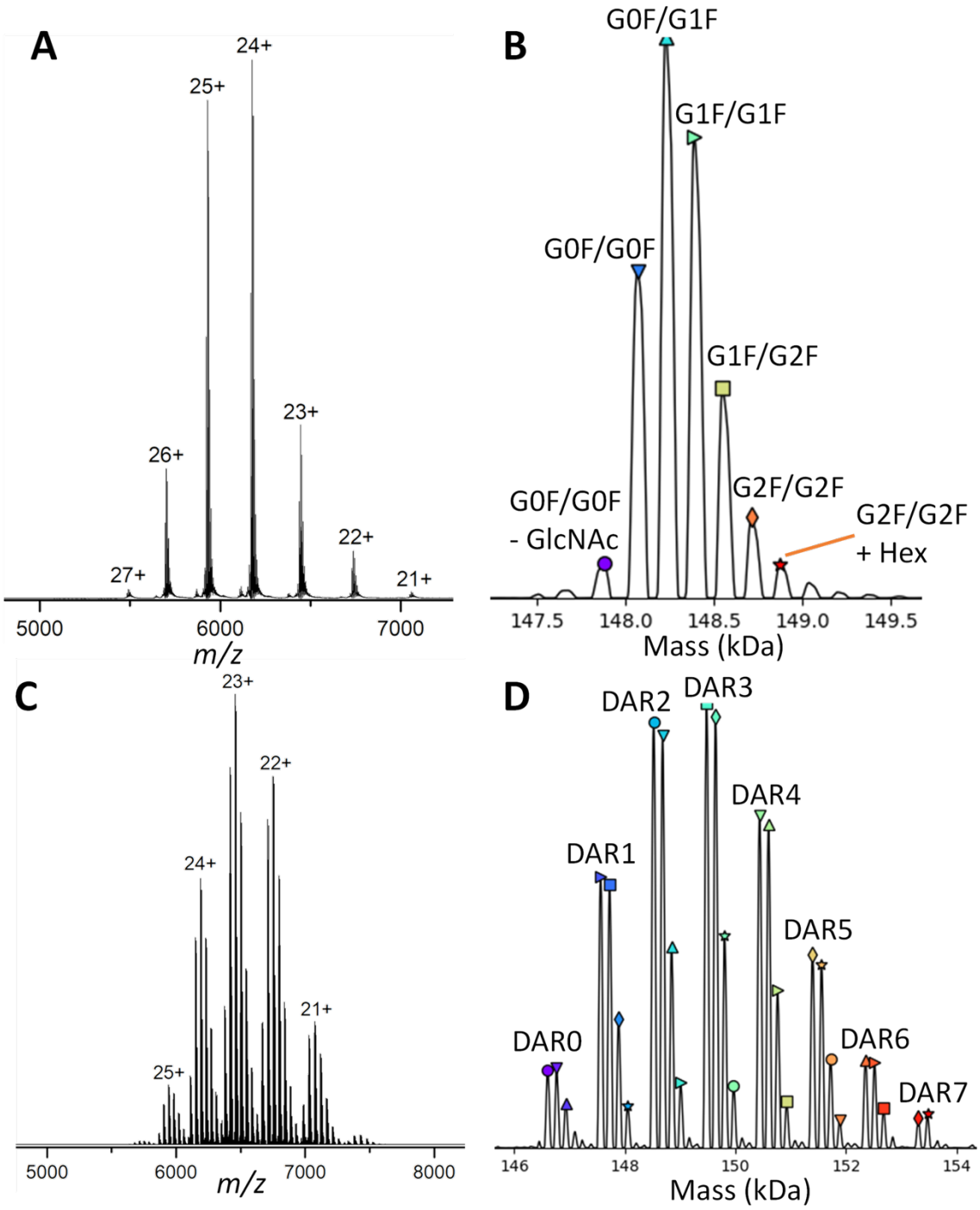


Figure 1. Native MS analysis of intact NIST mAb.

(A) Native MS spectrum of the intact NIST mAb. (B) Deconvoluted zero-charged spectrum[74] of the intact NIST mAb showing its major glycoforms. (C) Native MS spectrum of the intact IgG1-DM1 ADC. (D) Deconvoluted zero-charged spectrum of the intact IgG1-DM1 ADC showing its drug distribution profile.

The major obstacle that prevents comprehensive sequence coverage of intact mAbs is the presence of numerous disulfide bonds that contribute to maintaining protein structure and stability. For terminal fragments to cover the disulfide bonded sequence region, cleavages of both protein backbone and disulfide bonds are required. However, internal fragments can access these highly constrained regions without the need of breaking disulfide bonds, thus having the potential to substantially enhance sequence information on intact mAbs.[67] Additionally, internal fragments contain two cleavage sites while terminal fragments only contain one, making internal fragments naturally more information-rich than terminal fragments. This also leads to an increase in the sequence coverage obtained when considering internal fragments. To investigate the additional sequence information that can be obtained by assigning internal fragments, we combined the ECD and HCD TD-MS results from each isolated charge state of intact NIST mAb. Inclusion of internal fragments increases the sequence coverage of the light chain from 54% to 83% and the heavy chain from 28% to 72%, which when combined shows an increase from 36% to 76% for the whole NIST mAb (Figure 2); to our knowledge is the highest sequence coverage of an intact mAb achieved by TD-MS. A more significant increase is observed for the heavy chain (44%) than the light chain (29%), demonstrating that assigning internal fragments becomes more valuable with increasing protein size.

As expected, the significant increase in sequence coverage is largely due to improved access by internal fragments of the highly disulfide constrained regions, which cannot be reached by terminal fragments (Figure 2 and Table S2 and Table S4). For example, the disulfide constrained sequence between Cys133 and Cys193 of the light chain is almost exclusively

accessed by internal fragments (Figure 3A). The same is true for sequence regions Cys147-Cys203 and Cys264-Cys324 of the heavy chain (Figure 3B). For the intact NIST mAb, assigning internal fragments increases the coverage of the non-disulfide constrained sequence by 34%, compared to a more significant increase of 44% in the disulfide constrained sequence.

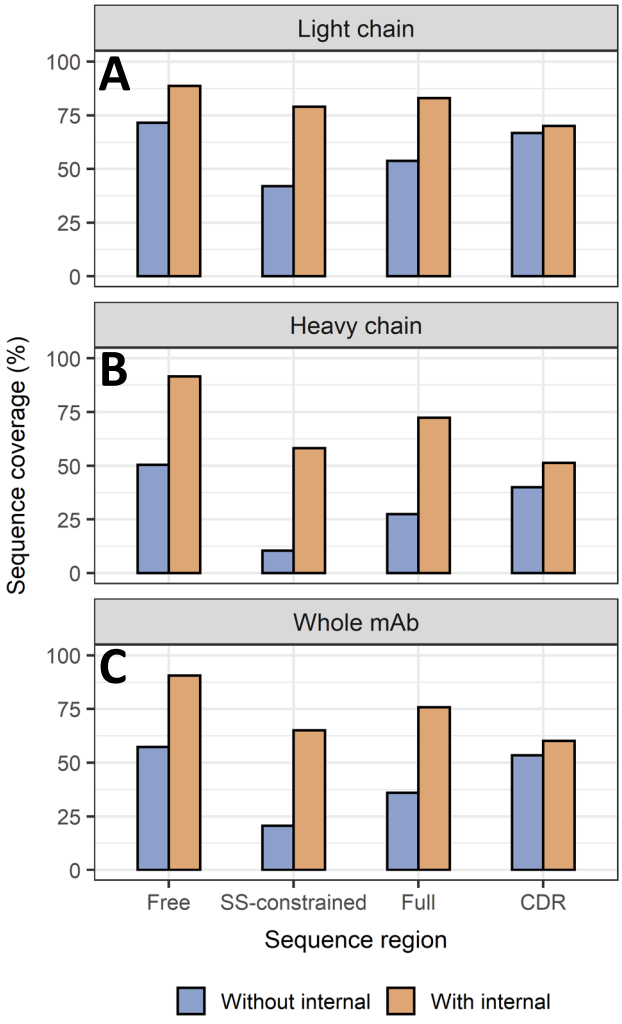


Figure 2. Sequence coverage analysis of intact NIST mAb.
Sequence coverage of different sequence regions including non-disulfide constrained sequence ("Free"), disulfide constrained sequence ("SS-constrained"), whole sequence ("Full"), and CDR sequence ("CDR") before and after considering internal fragments of (A) light chain, (B) heavy chain, and (C) whole mAb.

The sequence of the CDRs needs to be unambiguously determined, as it is responsible for antigen target specificity of mAbs. Additionally, chemical liabilities such as deamidation, isomerization, and oxidation in the CDRs are problematic; therefore, they need to be fully characterized.75 In total, an increase from 53% to 60% of amino acids in the CDRs are confirmed after assigning internal fragments, demonstrating that more insight of this critical region can be gleaned with the inclusion of internal fragments. Despite the improvement, the sequence coverage of CDRs is still far from optimal, particularly for the longer heavy chain (70% for light chain vs 51% for heavy chain). Therefore, complementary techniques such as middle-down and bottom-up approaches are needed to unambiguously determine the CDRs sequence, and internal fragments can play a pivotal role in these techniques.

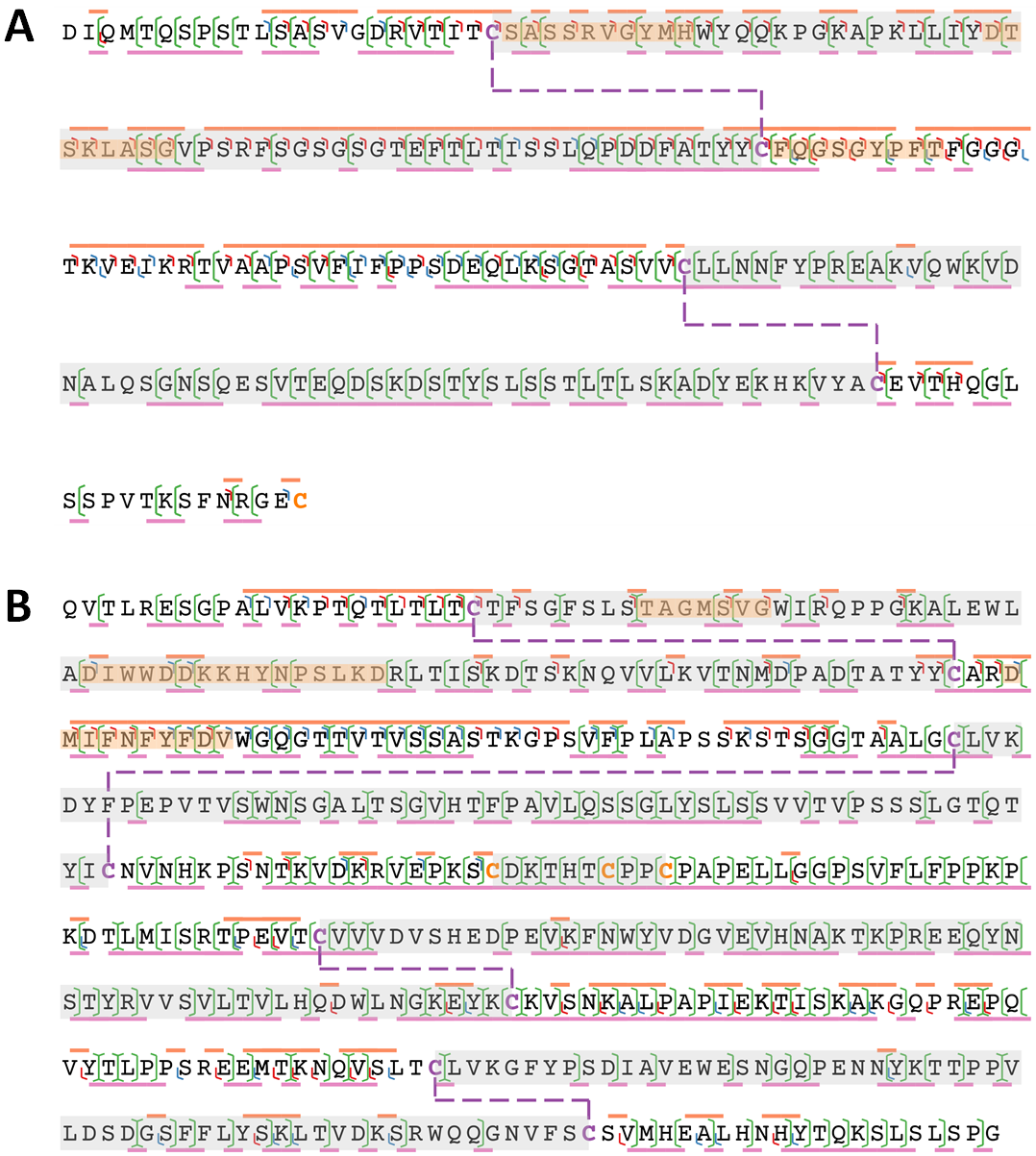


Figure 3. Cleavage site analysis of the light and heavy chain of NIST mAb.
NIST mAb sequence coverage maps for (A) light chain and (B) heavy chain. Blue, red, and green cleavages on the protein backbone represent *b*/*y*, *c*/*z*·, and *by*/*cz*· fragments, respectively. The solid line above the sequence represents terminal fragment sequence coverage, while the solid line beneath the sequence represents internal fragment sequence coverage. The purple dashed lines represent intrachain disulfide bonds, with the sequence region constrained by disulfide bonds covered in light gray, and complementarity-determining regions (CDRs) covered in orange.

3.2. Internal Fragments Can Identify mAb PTMs, Including Intrachain S−S Connectivity and N-Glycosylations.

Previously, our group demonstrated that internal fragments that retain intact disulfide bonds can be used to determine S−S connectivity of intact proteins (Figure S3).[67] This encourages us to explore the utility of such fragments to determine intrachain S−S connectivity of intact NIST mAb, which is comprised of 16 disulfide bonds. HCD is known to only cleave the protein backbone while maintaining the integrity of disulfide bonds; therefore, we applied HCD on the intact NIST mAb to generate such fragments to determine S−S connectivity.

A hydrogen loss was applied on every cysteine forming intrachain disulfide bonds to indicate their integrity after HCD fragmentation, and fragments that traverse an intact disulfide bond can determine S−S connectivity (green fragments in Figures 4 and S3). For example, for the light chain, 52 terminal fragments and 12 internal fragments traverse S−S bond I, 17 terminal fragments traverse S−S bond II, and 6 terminal fragments traverse both disulfide bonds, clearly demonstrating the S−S bonding pattern of these two disulfide bonds (Figure 4A). The value of analyzing internal fragments to determine intrachain S−S connectivity is exhibited more clearly in the case of the heavy chain. The two disulfide bonds close to either terminus of the heavy chain, S−S bond I and S−S bond IV, are traversed by 89 terminal fragments and 9 internal fragments. However, the two middle disulfide bonds, S−S bond II and S− S bond III, are only traversed by 24 internal fragments but no terminal fragments (Figure 4B and C). Fragments traversing two intact disulfide bonds were also observed. For example, S−S bonds I and II are traversed by 10 terminal fragments and 1 internal fragment, and S−S bonds III and IV are

traversed by 3 terminal fragments and 9 internal fragments. In contrast, no terminal fragments but 2 internal fragments traverse S−S bonds II and III, the middle two disulfide bonds. This corroborates the ability of internal fragments to access the interior regions of the protein sequence and disulfide bonds, particularly for larger proteins. These results demonstrate the intrachain S−S connectivity of the NIST mAb heavy chain, and importantly, the two middle S−S bonding patterns can only be determined by internal fragments.

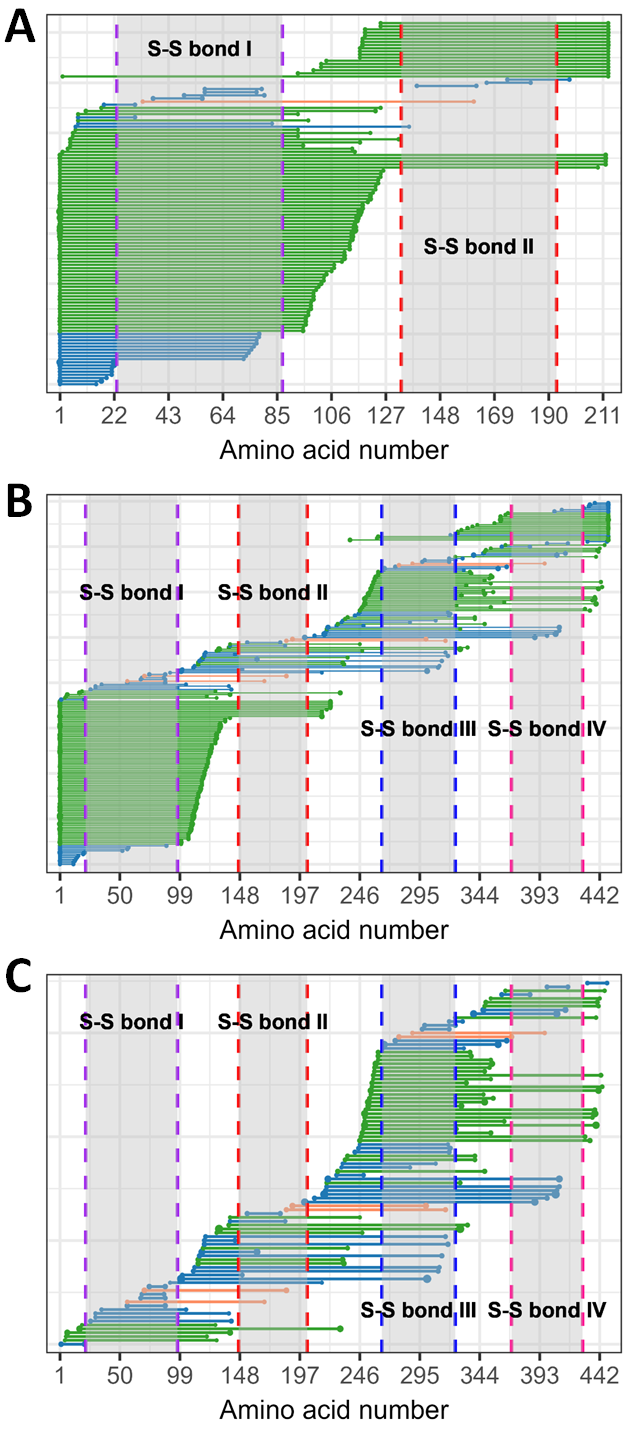


Figure 4. Disulfide connectivity analysis of the light and heavy chain of NIST mAb.
Fragment location maps generated by HCD TD-MS of (A) light chain, (B) heavy chain, and (C) heavy chain with only internal fragments of intact NIST mAb after applying one hydrogen loss on each cysteine forming intrachain disulfide bonds to indicate the integrity of the disulfide bond. Vertical dashed lines represent cysteine positions, with the same color corresponding to an

intrachain disulfide bond formed between those two cysteines. Green horizontal lines indicate fragments that can determine S−S connectivity, orange horizontal lines indicate fragments that cause mismatched dehydrocysteines, and blue horizontal lines indicate fragments that are irrelevant of S−S connectivity determination.

Although much rarer, HCD can cleave both the protein backbone and disulfide bonds simultaneously, generating two types of fragments that have differing ramifications in the determination of S−S connectivity. These fragments are represented as blue and orange fragments in Figure 4 and Figure S3. The blue fragments do not bridge any dehydrocysteines, or they traverse through intact disulfide bonds and from inside a disulfide bond simultaneously. These fragments do not directly contribute to the determination of S−S connectivity. The orange fragments, on the other hand, do not traverse through existing disulfide bonds, but rather they only cross through nonconnected dehydrocysteines from inside a disulfide bond, thus leading to mismatched S−S connectivity. For example, one such fragment was observed for the light chain (Figure 4A, Table S2) and six fragments for the heavy chain (Figure 4B and C, Table S3), resulting in S−S connectivity mismatch rates of 1.1% for the light chain and 3.9% for the heavy chain. The low mismatch rates support the concept of using fragments retaining intact disulfide bonds to determine intrachain S−S connectivity of intact mAbs. Alternatively, a small portion of disulfide bonds of NIST mAb ions may have been present in a reduced form prior to HCD fragmentation due to high source temperature and high in-source energy used, potentially leading to disulfide scrambling. This could also result in the formation of “S−S mismatched fragments” (orange fragments in Figure 4 and Figure S3). However, it is difficult to distinguish between these two possibilities by HCD TD-MS alone. A study comparing the HCD

fragmentation pattern of mAbs in both native and heat-stressed conditions could shed light on the contribution of disulfide scrambling.

In addition to determining intrachain S−S connectivity, assigning internal fragments also contributes to identifying N-glycosylations, a ubiquitous PTM class of mAbs. After combining data from five isolated precursor charge states, ECD and HCD TD-MS of intact NIST mAb generates only 9 unique C-terminal fragments containing N-glycosylations, with 2 containing G0F, 5 containing G1F, and 2 containing G2F (Figure S4). However, when internal fragments were considered, an additional 25 fragments containing G0F, 42 fragments containing G1F, and 34 fragments containing G2F were assigned, demonstrating the power of analyzing internal fragments for N-glycosylation identification (Figure S4). This is mainly attributed to the ability of internal fragments to access interior protein sequence constrained by disulfide bonds that are typically inaccessible to terminal fragments. The inclusion of internal fragments can potentially also contribute to identifying other common PTMs of mAbs such as oxidation and deamidation, improving the accuracy and consistency of mAb production.

3.3. Internal Fragments Can Determine Drug Conjugation Sites of Lysine-Linked ADCs.

The promising results obtained from TD-MS of intact NIST mAb inspired us to push one step further to explore the utility of internal fragments for the determination of drug conjugation sites of ADCs, an even more heterogeneous drug class. To achieve this, we took a similar TD-MS approach by applying both ECD and HCD on a previously well characterized nonspecific lysine-linked ADC.[12]

Native MS reveals the maytansinoid DM1 distribution profile of the intact ADC. Seven major DAR species were observed (DAR0−DAR7), confirming the highly heterogeneous nature of this ADC (Figure 1C and D). DAR 1 and DAR 2 species of the major four charge states (22+ to 25+) were isolated and fragmented with ECD and HCD (see Experimental Section). Similar to the NIST mAb, ECD fragmentation of the intact ADC generated both terminal and internal fragments from both chains of the antibody (Figure S5). Importantly, DM1-bound fragments were also observed, providing direct evidence to determine drug conjugation sites (Figures S5 and S6). The inherently random lysine conjugation makes the determination of drug conjugation sites of lysine-linked ADC challenging. For example, a total of 90 potential conjugation sites exist on the ADC used in this study after excluding the clipped C-terminal lysine on the heavy chain, including 11 from the light chain and 34 from the heavy chain. Nevertheless, applying two fragmentation methods, ECD and HCD, on the intact ADC allowed us to unambiguously determine a large fraction of potential conjugation sites, in which internal fragments played a critical role.

Here, we define *localizing a conjugation site* as when the conjugation can be specified on an exact lysine residue, while *identifying a conjugation site* is defined as when the conjugation can only be confirmed on several possible lysine residues. TDMS of the ADC generated only 8 DM1-bound terminal fragments on the light chain, including 5 one-DM1-bound and 3 two-DM1-bound fragments (Figure 5A). The two-DM1-bound terminal fragment c_{85} conclusively localized the conjugation sites to K46 and K67, whereas all other assigned terminal fragments could not localize any additional conjugation sites but only identify 2 other conjugation sites on the light

chain. Unsurprisingly, assigning internal fragments significantly improved the determination of DM1 conjugation sites. TD-MS of the ADC generated 61 one-DM1-bound and 15 two-DM1-bound internal fragments on the light chain, which localized 3 more conjugation sites (K106, K114, K133) and narrowed down the identified two conjugation sites to 4 lysine residues (K153, K160, K170, K175) (Figure 5A). For example, the assignment of one-DM1-bound internal fragments cz_{96-113}, cz_{97-113}, cz_{98-111}, and $by_{104-112}$ localized the conjugation site to K106, $cz_{112-125}$ and $cz_{114-124}$ localized the conjugation site to K114, $cz_{119-152}$, $cz_{128-140}$, and $by_{133-141}$ localized the conjugation site to K133 (Figure 5A). Similar results were observed for the heavy chain. TD-MS of the ADC generated only 11 DM1-bound terminal fragments (6 one-DM1-bound and 5 two-DM1-bound) but 167 DM1-bound internal fragments (107 one-DM1-bound and 60 two-DM1bound) on the heavy chain. With terminal fragments alone, no conjugation sites could be localized, but 4 were identified; however, after considering internal fragments, 9 conjugations sites were localized (K13, K43, K89, K127, K139, K153, K252, K254, K323), and 10 additional conjugation sites were identified. In summary, for the intact ADC, only 16 conjugation sites were confirmed (4 localized and 12 identified) with terminal fragments alone, whereas this number was increased to 52 (28 localized, 24 identified) upon inclusion of internal fragments, covering approximately 58% of all putative conjugation sites of the antibody (Table S5). Although lower than the previously reported 83% coverage achieved by peptide mapping,12 this result still demonstrates the value of analyzing TD-MS internal fragments to determine drug conjugation sites of ADCs.

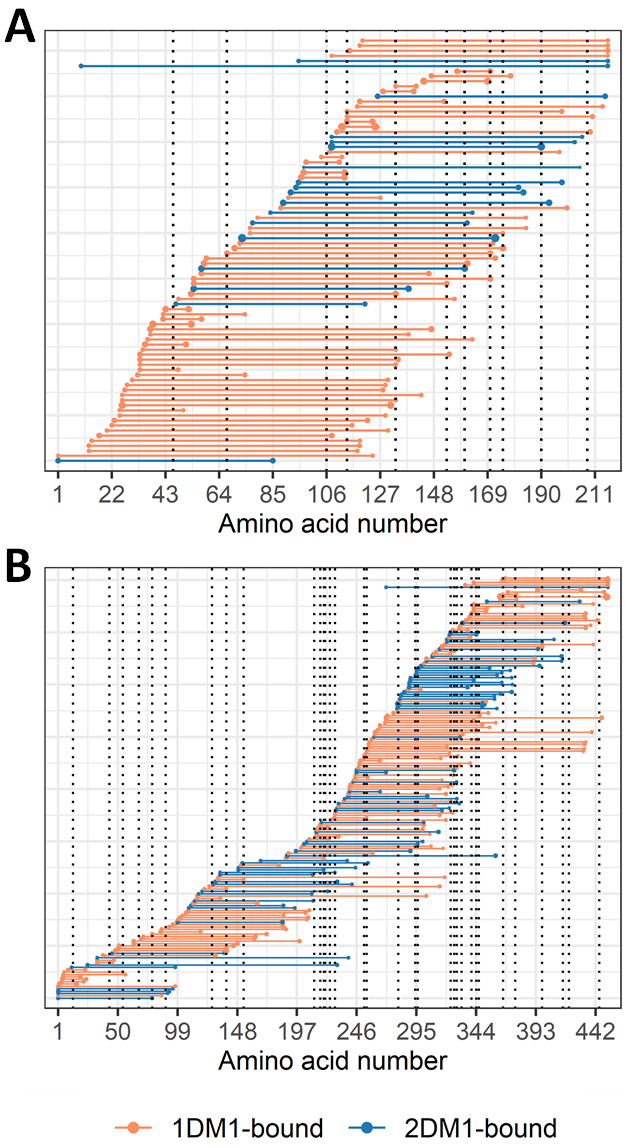


Figure 5. Drug binding analysis of the light and heavy chain of the IgG1-DM1 ADC. Fragment location maps generated by ECD and HCD TDMS of (A) light chain and (B) heavy chain of the intact IgG1-DM1 ADC. Black vertical dotted lines represent lysine positions. Orange and blue horizontal solid lines represent DM1-bound fragments.

4. Conclusions

Here we report for the first time the benefits of analyzing internal fragments in the TD-MS characterization of intact NIST mAb and a heterogeneous lysine-linked ADC. Inclusion of

internal fragments significantly increases the sequence coverage of intact mAbs to over 75% by accessing the disulfide constrained regions that are hardly accessed by terminal fragments, particularly for the larger heavy chain. Important PTM information, including disulfide linkage patterns and N-glycosylations, can be obtained by including internal fragments. Internal fragments retaining intact disulfide bonds were used to determine the intrachain S−S connectivity of an intact mAb, an important CQA required for accurate determination during antibody production. And importantly, we show that internal fragments can help pinpoint drug conjugation sites of a highly heterogeneous lysine-linked ADC, an attribute that is as important as, but not as well evaluated as, characteristics like DAR and drug distribution.

It should be noted that limitations still exist for TD-MS of mAbs despite the added benefit of assigning internal fragments. For example, only ~60% sequence coverage on the CDRs of an intact mAb and ~58% drug conjugation site coverage of an intact lysine-linked ADC were achieved, which do not meet the requirement of biologics development in the pharmaceutical industry. Identifying uncommon, low intensity glycoforms also remains challenging. Furthermore, the heavy reliance on manual inspection during data analysis likely prevents the widespread adoption of this approach. Therefore, complementary techniques such as MD-MS and BU-MS approaches are currently still required to achieve more comprehensive and efficient therapeutics characterization. For instance, denatured reversed-phase LC−MS of intact and FabRICATOR/reduced treated mAbs results in smaller protein subunits with higher charge states, thus enabling more efficient MS/MS fragmentation. Nevertheless, the results presented here demonstrate the multiple benefits of assigning internal fragments to obtain critical structural

information on intact mAbs and ADCs.

From an analytical viewpoint, the raw data for internal fragments are present in the TD mass spectra, but assigning the peaks in the spectra to uncover this *hidden treasure* can be a fruitful endeavor, especially for characterizing therapeutic proteins. Although bottom-up MS is firmly entrenched in pharmaceutical industry workflows, TD-MS offers potential benefits if robust automation and computational support can be established. Once this native TD-MS method has fully evolved and matured to a point where it is comparable to the already-established bottom-up approaches, one can realize the significantly reduced amount of sample handling (no denaturation, reduction, or alkylation); therefore, there is less opportunity to introduce sample handling-related artifacts. This study should also suggest that the incorporation of internal fragments can be applied to bottom-up and middle-down MS analysis of mAbs and ADCs, potentially extending their characterization to a near complete level on a routine basis.

Associate Content

*sı Supporting Information

The Supporting Information is available free of charge at:

https://pubs.acs.org/doi/10.1021/acs.analchem.3c01426.

Conjugation process of the ADC; additional ECD MS/ MS spectra; additional fragment location maps showing the identification of N-glycosylation identification; details of internal fragments generated by HCD TDMS of intact NIST mAb; additional data showing the exact values of sequence coverage and number of drug conjugation sites; and instrument and ExD cell

parameters (PDF).

Author Information

Corresponding Author

Joseph A. Loo − Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States; Department of Biological Chemistry, UCLA-DOE Institute, and Molecular Biology Institute, University of California—Los Angeles, Los Angeles, California 90095, United States; orcid.org/0000-0001-9989-1437; Email: jloo@chem.ucla.edu

Authors

Benqian Wei − Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States; orcid.org/0000-00034853-4848.

Carter Lantz − Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States.

Weijing Liu − Thermo Fisher Scientific, San Jose, California 95134, United States; orcid.org/0000-0003-2322-0500.

Rosa Viner − Thermo Fisher Scientific, San Jose, California 95134, United States.

Rachel R. Ogorzalek Loo − Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States; UCLA-DOE Institute and Molecular Biology Institute, University of California—Los Angeles, Los Angeles, California 90095, United States; orcid.org/0000-0002-0580-2833.

Iain D. G. Campuzano − Center for Research Acceleration and Digital Innovation, Molecular Analytics, Amgen Research, Thousand Oaks, California 91320, United States; orcid.org/0000-0003-4310-8540.

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c01426.

Notes:

The authors declare no competing financial interest.

Acknowledgements

J.A.L. and R.R.O.L. acknowledge support from the U.S. National Institutes of Health (Grants R01GM103479 and R35GM145286) and the U.S. Department of Energy (Grant DE-FC02-02ER63421). C.L. acknowledges support from the Ruth L. Kirschstein National Research Service Award program (Grant GM007185).

Chapter 4: Supporting Information

Supplementary Figures

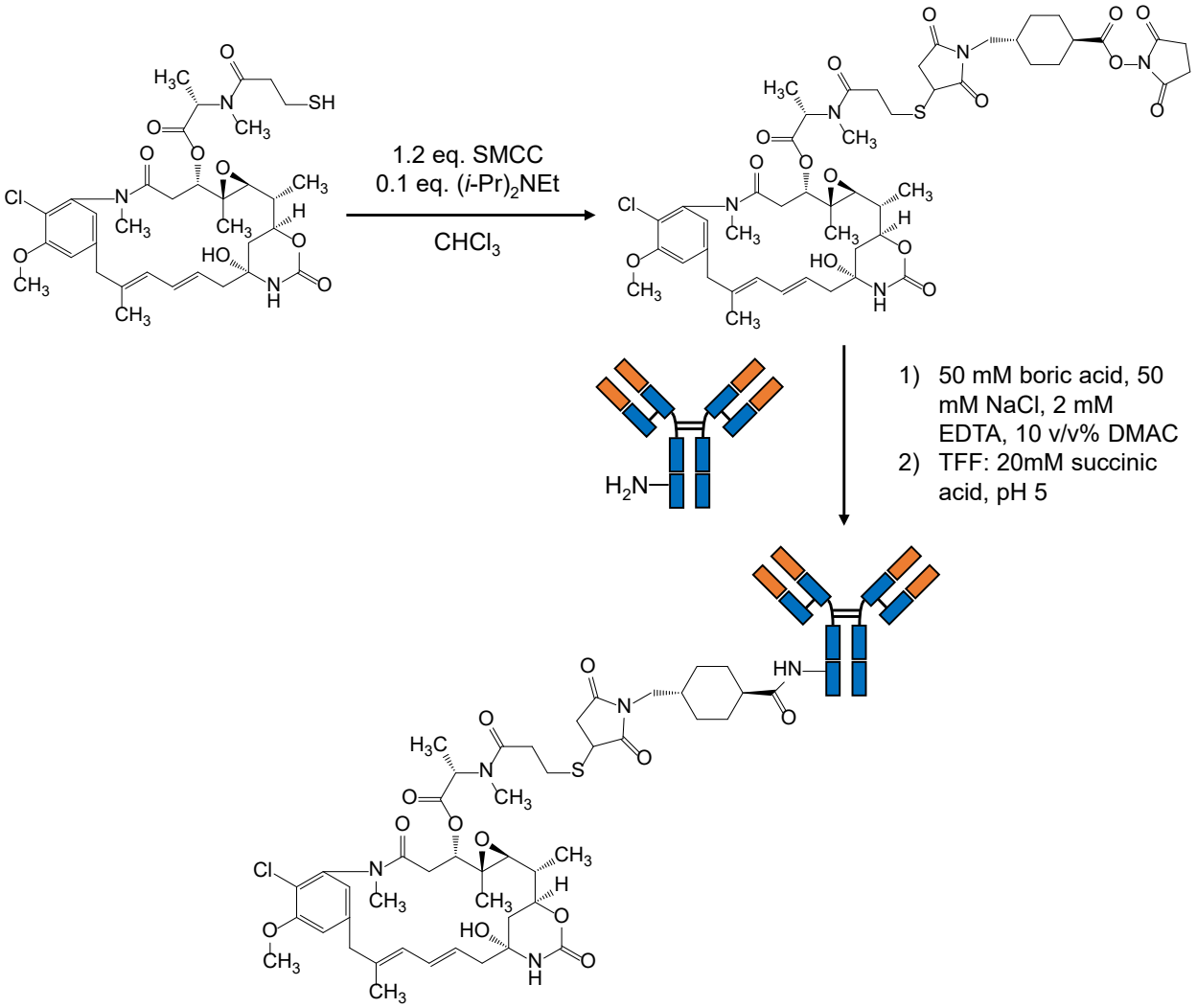


Figure S1. Conjugation process of ADC.
Reproduced with permission from Ref.[199]. Copyright 2016 American Chemical Society.

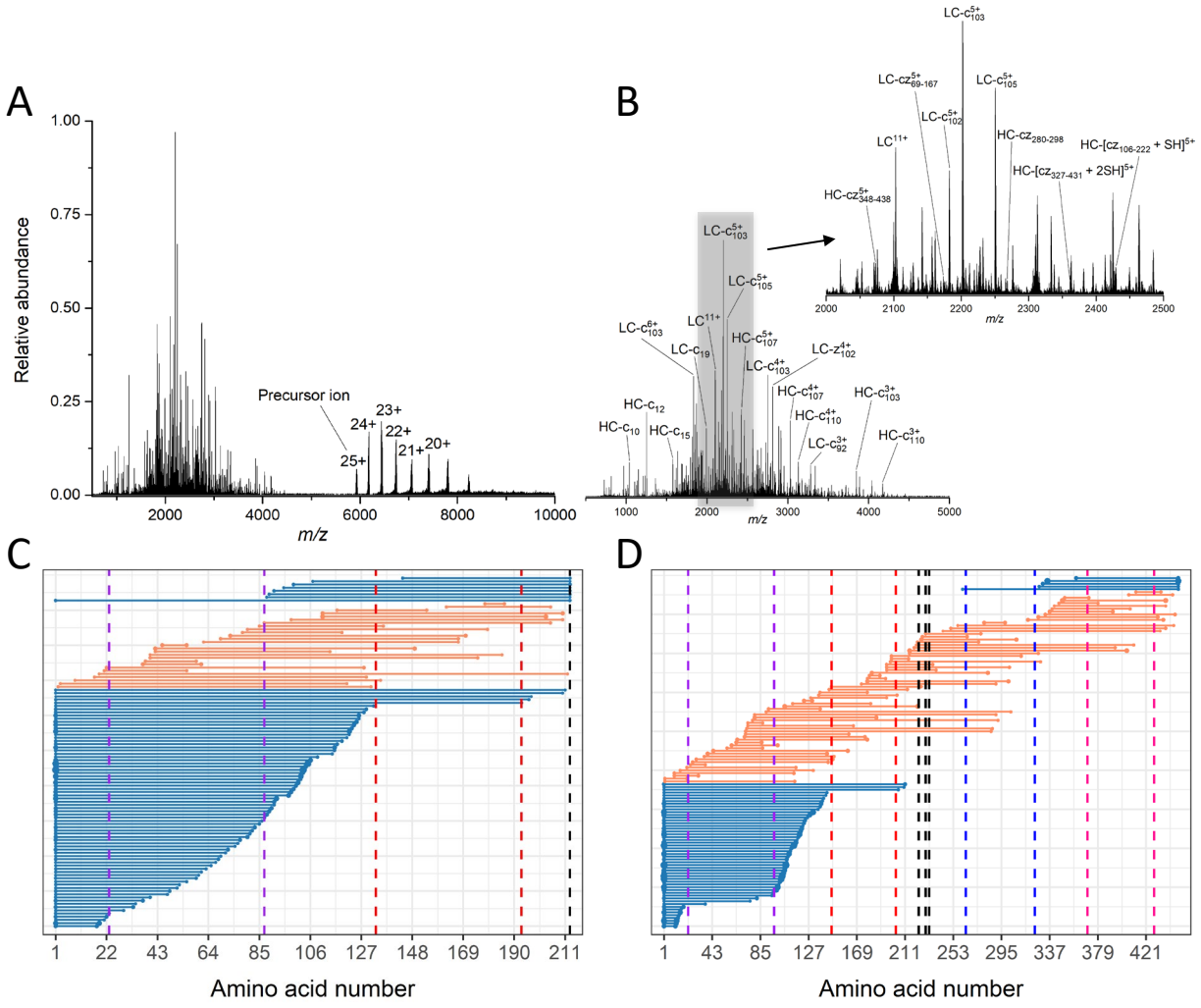


Figure S2. Data analysis of ECD MS/MS of intact NIST mAb.

(A) Representative ECD MS/MS spectrum of intact NIST mAb, $[NIST + 25H]^{25+}$. (B) Zoomed-in of spectrum in panel A in the range from 500 to 5000 *m/z* and 2000 to 2500 *m/z* showing terminal and internal fragments from both heavy and light chains are generated. Fragment location maps of the (C) light chain and (D) heavy chain indicating the region of the protein sequence covered by terminal and internal fragments for spectrum in panel A. The black vertical dashed lines indicate the inter-chain disulfide bond position, while the dashed lines with the same color indicate the intra-chain disulfide bond position. Blue horizontal lines represent terminal fragments, while orange horizontal lines represent internal fragments.

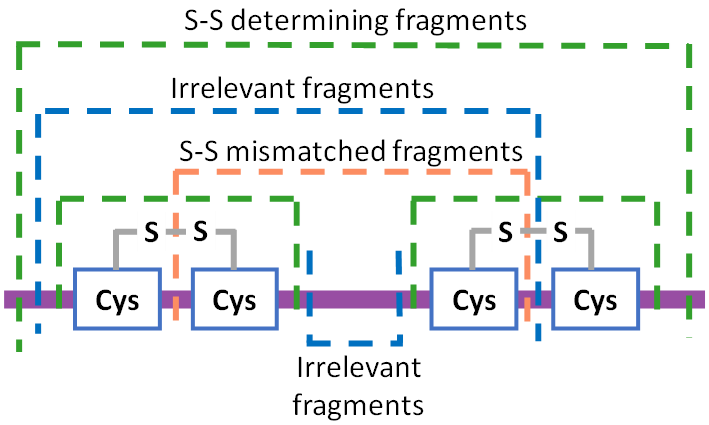


Figure S3. Elucidation of the three types of fragments for disulfide connectivity determination.

The three types of fragments generated by HCD TD-MS used for S-S connectivity determination. A hydrogen loss (-1 Da) was applied on every cysteine residue to suggest the integrity of disulfide bonds involved. Green fragments traverse an even number of dehydrocysteines (2, 4, 6 etc.), suggesting that intact disulfide bonds are formed within the cysteines involved. Blue fragments either do not bridge any dehydrocysteines, or they traverse though intact disulfide bonds and from inside a disulfide bond simultaneously. These fragments are not involved in the determination of S-S connectivity. Orange fragments cross through non-connected dehydrocysteines from inside a disulfide bond, leading to mismatched S-S connectivity.

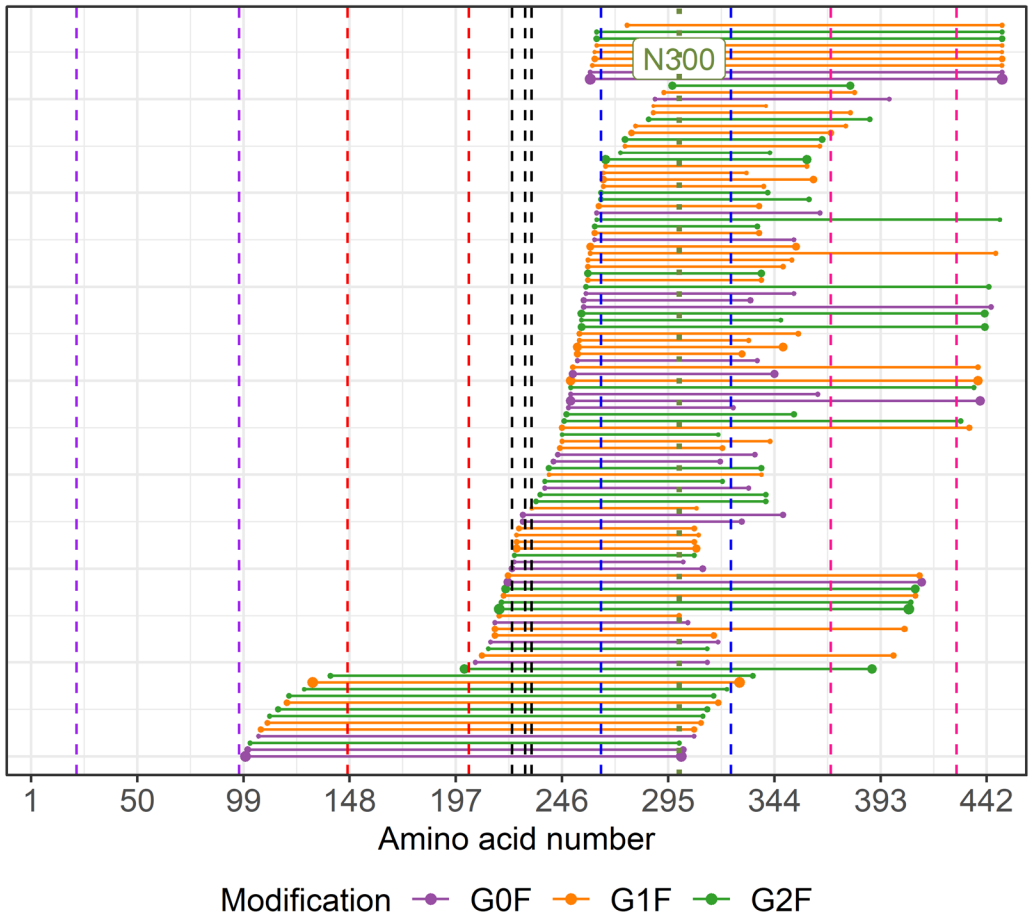


Figure S4. Fragment location map of NIST mAb heavy chain showing fragments containing N-glycosylations.

Each horizontal solid line represents a fragment with an N-glycosylation modification. Colored vertical dashed lines represent cysteine positions, with the same color representing an intra-chain disulfide bond formed between those two cysteines. Black vertical dashed lines indicate the position of cysteines forming inter-chain disulfide bonds. The olive-colored vertical dotted line indicates the position of asparagine (N) 300, the residue that is known to be highly glycosylated.

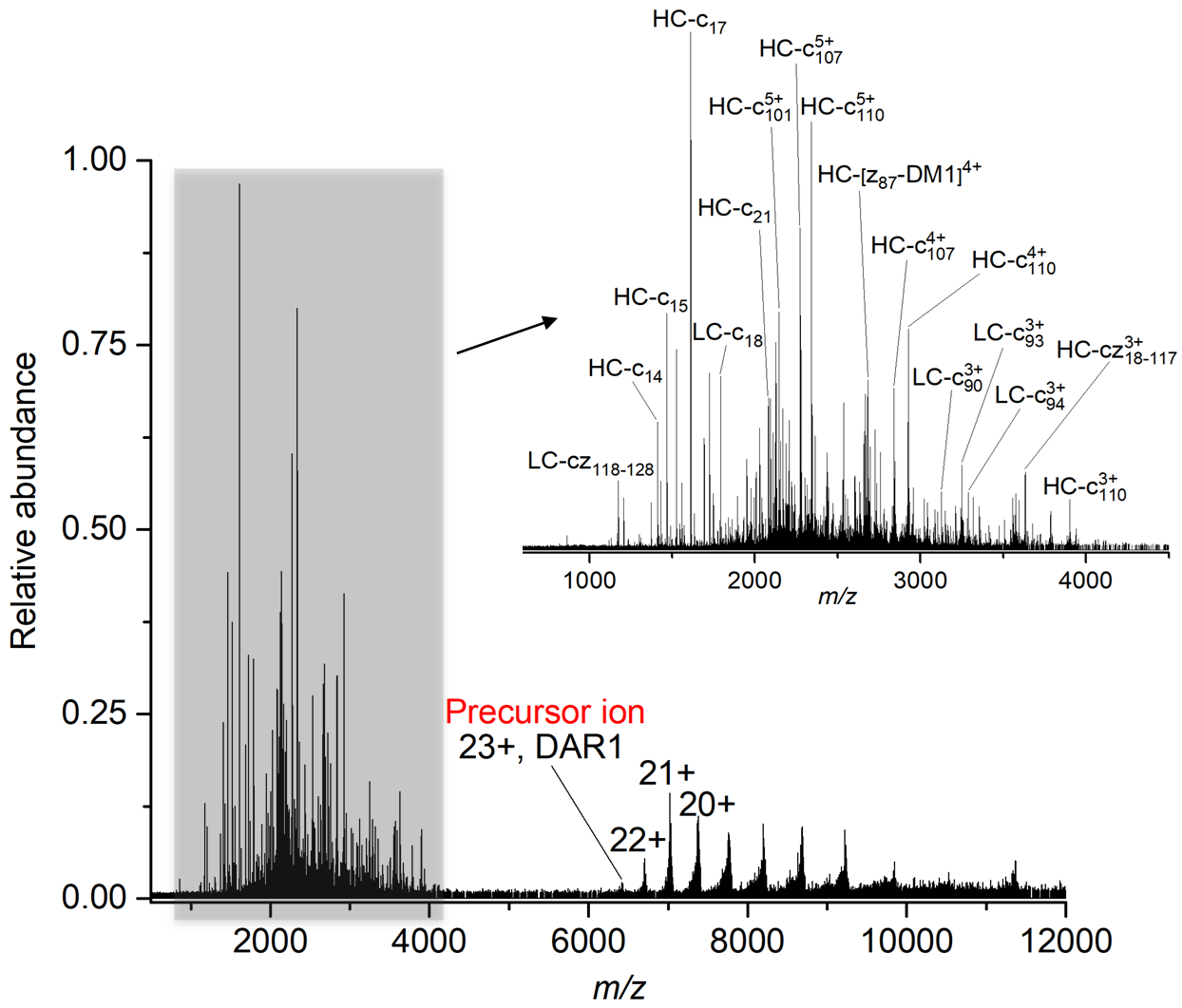


Figure S5. An ECD MS/MS spectrum of the intact IgG1-DM1 ADC.
Representative ECD MS/MS spectrum of DAR1 species of intact ADC, $[ADC\text{-}DAR1 + 23H]^{23+}$, with a zoomed-in spectrum in the range from 1000 to 4000 *m/z* showing terminal and internal fragments, and their DM1-bound forms from both heavy and light chains.

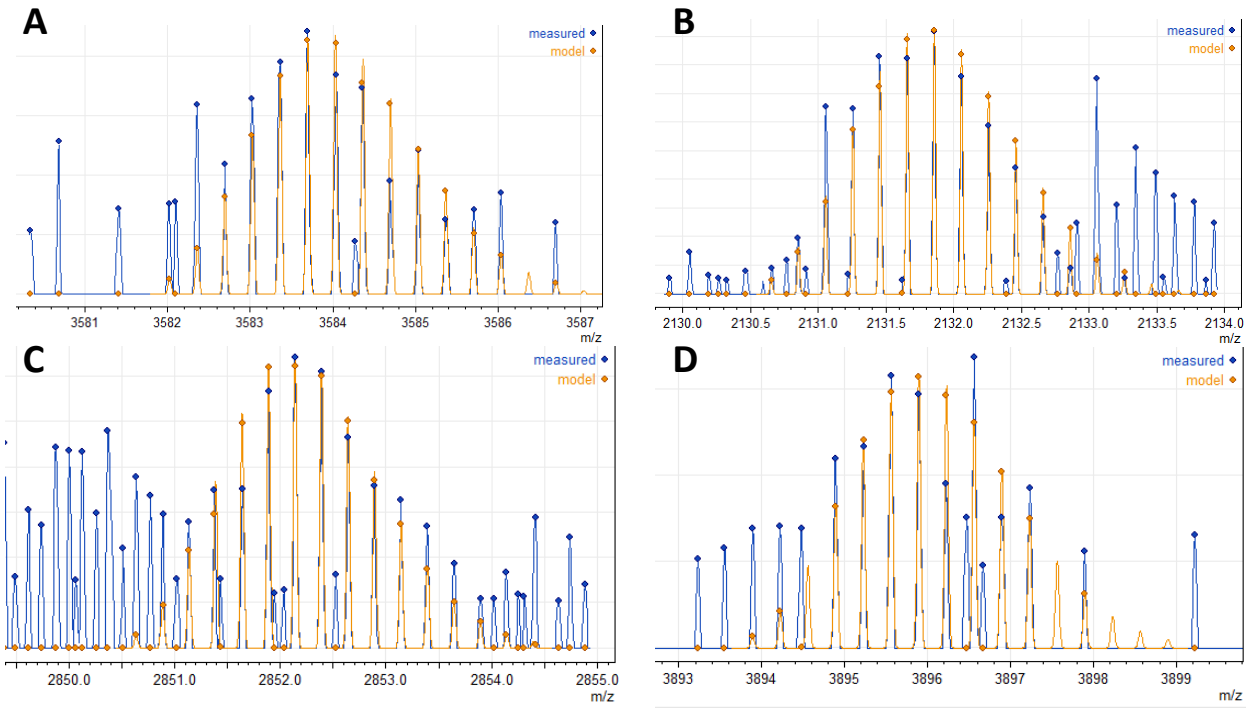


Figure S6. Isotope envelope fits of representative DM1-bound internal fragments generated by ECD of intact IgG1-DM1 ADC.

(A) HC-[$CZ_{351-436}$-DM1]$^{3+}$ (10742.04 Da), (B) HC-[$CZ_{214-300}$-DM1]$^{5+}$ (10648.23 Da), (C) LC-[CZ_{34-131}-DM1]$^{4+}$ (11398.50 Da), (D) LC-[CZ_{58-161}-DM1]$^{4+}$ (11677.63 Da).

Supplementary Tables

Table S1. Instrument parameters crucial for ion transmission and ExD cell parameters for NIST mAb and ADC fragmentation.

Instrument					
Detector *m/z* Optimization	Low *m/z*	Source DC Offset (V)	21	In-source Trapping (V)	-125
Ion Transfer Target *m/z*	Low *m/z*	Injection Flatapole DC (V)	5.8	Inter Flatapole Lens (V)	2
Bent Flatapole DC (V)	1	Transfer Multipole DC (V)	0	C-Trap Entrance Lens Inject (V)	3
ExD Cell for NIST mAb fragmentation (V)					
L1	1.48	L2	-50.0	LM3	4.2
L4	14.7	FB	-1.7	LM5	4.2
L6	-50.0	L7	1.3		
ExD Cell for ADC fragmentation (V)					
L1	-1.27	L2	-45.0	LM3	5.7
L4	8.8	FB	0.9	LM5	5.7
L6	-45.0	L7	0.0		

Table S2. All internal fragments generated by HCD TD-MS of intact NIST mAb used to determine the S-S connectivity of the light chain.

Start Amino Acid Number	End Amino Acid Number	Unlocalized Modification	Observed Mass	Theoretical Mass	Mass Error (ppm)	S-S Connectivity Determination
2	115	None	12168.01	12168.01	-0.3841	Yes
4	114	None	11779.8	11779.8	-0.6467	Yes
5	95	None	9749.697	9749.704	-0.7516	Yes
5	117	None	12055.97	12055.98	-1.2902	Yes
5	132	None	13551.72	13551.73	-1.0233	Yes
6	93	None	9404.524	9404.536	-1.2202	Yes
6	121	None	12351.12	12351.1	1.9936	Yes
7	93	None	9276.479	9276.477	0.225	Yes
8	83	None	8021.005	8021.004	0.138	Yes
8	97	None	9681.677	9681.682	-0.5308	Yes
11	123	None	12107.97	12107.98	-0.3254	Yes
17	125	None	11834.87	11834.88	-1.3801	Yes
33	161	None	14028.98	14028.98	-0.1664	Mismatch
7	136	None	13764.84	13764.84	-0.2076	Irrelevant
8	30	None	2232.117	2232.116	0.3123	Irrelevant
8	93	None	9189.437	9189.445	-0.8112	Irrelevant
18	30	None	1317.684	1317.682	1.6901	Irrelevant
37	56	None	2097.208	2097.207	0.6409	Irrelevant
49	80	None	3211.571	3211.575	-1.3557	Irrelevant
57	77	None	2112.059	2112.061	-0.8437	Irrelevant
57	79	None	2337.176	2337.172	1.6584	Irrelevant
139	162	None	2715.352	2715.349	1.2794	Irrelevant
166	183	None	1885.942	1885.939	1.6247	Irrelevant
174	198	None	2832.411	2832.411	-0.0314	Irrelevant

Table S3. All internal fragments generated by HCD TD-MS of intact NIST mAb used to determine the S-S connectivity of the heavy chain.

Start Amino Acid Number	End Amino Acid Number	Unlocalized Modification	Observed Mass	Theoretical Mass	Mass Error (ppm)	S-S Connectivity Determination
5	129	None	13864.9	13864.9	-0.3325	Yes
6	121	None	12905.4	12905.4	0.5623	Yes
7	140	None	14592.3	14592.3	0.217	Yes
17	230	None	23086.4	23086.4	1.059	Yes
19	128	None	12240	12240	2.5396	Yes
114	232	None	12236.1	12236.1	-0.1224	Yes
114	233	None	12333.1	12333.1	0.0598	Yes
118	236	None	12228.1	12228.1	0.2222	Yes
129	248	None	12452.2	12452.2	-1.9304	Yes
131	328	G1F	23097.3	23097.3	-0.3464	Yes
139	334	G2F	23065.3	23065.3	-2.2507	Yes
140	246	None	11155.6	11155.6	2.1844	Yes
219	409	G1F	23022.3	23022.3	-0.2252	Yes
228	348	G0F	14991.5	14991.5	-0.4563	Yes
234	340	G2F	13822.9	13822.9	-0.4578	Yes
236	340	G2F	13654.8	13654.8	-0.6349	Yes
246	434	G1F	23098.3	23098.3	-1.6343	Yes
247	430	G2F	22645.1	22645.1	1.4118	Yes
248	353	G2F	13880.9	13880.9	-0.1061	Yes
249	325	G0F	10428.1	10428	0.9334	Yes
250	364	G0F	14614.3	14614.3	-0.1314	Yes
250	436	G2F	23041.2	23041.2	-0.4336	Yes
250	438	G1F	23130.3	23130.3	-1.412	Yes
250	439	G0F	23131.3	23131.3	-0.8296	Yes
251	438	G1F	23033.2	23033.2	0.8294	Yes
253	329	G1F	10550	10550.1	-1.5181	Yes
253	348	G1F	12565.2	12565.2	-0.2793	Yes
254	355	G1F	13262.6	13262.6	-1.5399	Yes
255	347	G2F	12384.1	12384.1	0.9712	Yes
255	441	G2F	23130.2	23130.2	1.0535	Yes
256	333	G0F	10395	10395	-1.5175	Yes
256	444	G0F	23003.3	23003.3	0.0173	Yes
257	353	G0F	12533.2	12533.2	-0.0887	Yes
257	443	G2F	23101.2	23101.2	1.2677	Yes
258	348	G1F	12019.9	12019.9	0.6498	Yes
258	352	G1F	12507.2	12507.2	0.8576	Yes
259	354	G1F	12565.2	12565.2	-0.9261	Yes

261	337	G1F	10470.1	10470.1	-0.3163	Yes
261	336	G2F	10504	10504	0.3666	Yes
261	353	G0F	12092	12092	0.8891	Yes
324	439	None	13062.5	13062.5	0.3556	Yes
348	441	None	10747.1	10747.1	0.6574	Yes
349	440	None	10490	10490	2.3105	Yes
353	442	None	10258.9	10258.9	-1.2885	Yes
365	446	None	9275.48	9275.47	0.6737	Yes
56	168	None	12170	12170	-0.0371	Mismatch
70	186	None	12304	12304	-0.0311	Mismatch
186	316	None	14592.3	14592.3	0.2982	Mismatch
191	300	None	12235.1	12235	1.4143	Mismatch
278	370	G1F	12303.1	12303.1	-0.5663	Mismatch
289	397	G0F	13862.9	13862.9	1.6621	Mismatch
2	21	None	2107.21	2107.21	-0.5391	Irrelevant
26	141	None	12704.3	12704.3	-0.143	Irrelevant
30	97	None	7733.85	7733.85	0.0948	Irrelevant
30	139	None	12186	12186	-0.1098	Irrelevant
35	104	None	8134.08	8134.08	0.9152	Irrelevant
56	87	None	3697.97	3697.97	-0.9235	Irrelevant
67	87	None	2372.3	2372.3	-0.258	Irrelevant
68	86	None	2126.23	2126.23	-1.9673	Irrelevant
74	87	None	1558.83	1558.83	1.4383	Irrelevant
91	215	None	13089.4	13089.4	1.8925	Irrelevant
100	301	G0F	23142.3	23142.2	1.2717	Irrelevant
101	150	None	5080.52	5080.53	-2.7403	Irrelevant
110	310	G1F	23031.3	23031.3	0.1236	Irrelevant
111	311	G2F	23106.4	23106.4	0.1708	Irrelevant
115	313	G2F	23013.3	23013.3	-2.2528	Irrelevant
117	162	None	4609.3	4609.3	-1.1956	Irrelevant
119	144	None	2362.19	2362.19	-1.2048	Irrelevant
119	318	G1F	23107.4	23107.4	0.1434	Irrelevant
120	316	G2F	22955.2	22955.2	-0.5408	Irrelevant
140	185	None	4678.34	4678.34	-0.4767	Irrelevant
154	181	None	2807.41	2807.4	2.1753	Irrelevant
201	389	G2F	23057.3	23057.4	-2.3385	Irrelevant
209	399	G1F	23066.4	23066.4	-1.2839	Irrelevant
217	406	G2F	23032.3	23032.3	0.172	Irrelevant
218	407	G2F	23023.3	23023.3	0.6827	Irrelevant
219	328	None	12429.2	12429.2	-0.1003	Irrelevant
220	409	G2F	23055.3	23055.3	-0.4073	Irrelevant
227	249	None	2386.21	2386.21	-1.1093	Irrelevant
231	287	None	6260.12	6260.13	-2.6744	Irrelevant
232	308	G1F	10256.9	10256.9	-1.9064	Irrelevant
242	319	G0F	10437.1	10437.1	-0.1454	Irrelevant

245	320	G1F	10394	10394.1	-1.7186	Irrelevant
246	318	G2F	10257.9	10257.9	0.8626	Irrelevant
265	331	G1F	9403.53	9403.53	-0.4256	Irrelevant
266	359	G2F	12565.2	12565.2	-0.0847	Irrelevant
275	366	G2F	12418.2	12418.2	0.1686	Irrelevant
297	319	None	2703.35	2703.35	-0.9425	Irrelevant
299	324	None	3096.55	3096.55	-0.8348	Irrelevant
319	330	None	1335.7	1335.7	-0.2695	Irrelevant
339	407	None	7678.76	7678.77	-0.5149	Irrelevant
347	414	None	7721.75	7721.76	-1.6262	Irrelevant
362	385	None	2708.36	2708.35	2.7884	Irrelevant
399	416	None	1985	1984.99	2.397	Irrelevant
432	448	None	1943.97	1943.97	-0.1235	Irrelevant

Table S4. Sequence coverage values of different sequence regions of light chain, heavy chain, and whole mAb of intact NIST mAb.

	Light chain (%)		Heavy chain (%)		Whole mAb (%)	
	Without internal	With internal	Without internal	With internal	Without internal	With internal
Free	71.6	88.6	50.5	91.6	57.2	90.7
SS-constrained	41.9	79.0	10.5	58.1	20.7	64.9
Full	53.8	83.0	27.5	72.3	35.9	75.8
CDR	66.7	70.0	40.0	51.4	53.3	60.0

Table S5. Number of potential, localized, and identified DM1 conjugation sites of light chain, heavy chain, and whole antibody of the intact IgG1-DM1 ADC.

	Light chain		Heavy chain		Whole ADC	
	Without internal	With internal	Without internal	With internal	Without internal	With internal
Number of potential conjugation sites	11		34		90	
Number of localized conjugations sites	2	5	0	9	4	28
Number of identified conjugations sites	2	2	4	10	12	24

References

1. Luo, Q.; Chung, H. H.; Borths, C.; Janson, M.; Wen, J.; Joubert, M. K.; Wypych, J., Structural Characterization of a Monoclonal Antibody–Maytansinoid Immunoconjugate. *Analytical Chemistry* **2016**, *88* (1), 695-702.

References

1. Kisalu, N. K.; Idris, A. H.; Weidle, C.; Flores-Garcia, Y.; Flynn, B. J.; Sack, B. K.; Murphy, S.; Schön, A.; Freire, E.; Francica, J. R.; Miller, A. B.; Gregory, J.; March, S.; Liao, H.-X.; Haynes, B. F.; Wiehe, K.; Trama, A. M.; Saunders, K. O.; Gladden, M. A.; Monroe, A.; Bonsignori, M.; Kanekiyo, M.; Wheatley, A. K.; McDermott, A. B.; Farney, S. K.; Chuang, G.-Y.; Zhang, B.; Kc, N.; Chakravarty, S.; Kwong, P. D.; Sinnis, P.; Bhatia, S. N.; Kappe, S. H. I.; Sim, B. K. L.; Hoffman, S. L.; Zavala, F.; Pancera, M.; Seder, R. A., A human monoclonal antibody prevents malaria infection by targeting a new site of vulnerability on the parasite. *Nature Medicine* **2018,** *24* (4), 408-416.

2. London, M.; Gallo, E., Epidermal growth factor receptor (EGFR) involvement in epithelial-derived cancers and its current antibody-based immunotherapies. *Cell Biology International* **2020,** *44* (6), 1267-1282.

3. Schrama, D.; Reisfeld, R. A.; Becker, J. C., Antibody targeted drugs as cancer therapeutics. *Nature Reviews Drug Discovery* **2006,** *5* (2), 147-159.

4. Bebbington, C.; Yarranton, G., Antibodies for the treatment of bacterial infections: current experience and future prospects. *Current Opinion in Biotechnology* **2008,** *19* (6), 613-619.

5. Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N., Strategies and challenges for the next generation of therapeutic antibodies. *Nature Reviews Immunology* **2010,** *10* (5), 345-352.

6. Mould, D. R.; Meibohm, B., Drug Development of Therapeutic Monoclonal Antibodies. *BioDrugs* **2016,** *30* (4), 275-293.

7. Marston, H. D.; Paules, C. I.; Fauci, A. S., Monoclonal Antibodies for Emerging Infectious Diseases — Borrowing from History. *New England Journal of Medicine* **2018,** *378* (16), 1469-1472.

8. Frenzel, A.; Hust, M.; Schirrmann, T., Expression of recombinant antibodies. *Front Immunol* **2013,** *4*, 217.

9. Sissolak, B.; Lingg, N.; Sommeregger, W.; Striedner, G.; Vorauer-Uhl, K., Impact of mammalian cell culture conditions on monoclonal antibody charge heterogeneity: an accessory monitoring tool for process development. *Journal of Industrial Microbiology and Biotechnology* **2019,** *46* (8), 1167-1178.

10. Rathore, A. S.; Winkle, H., Quality by design for biopharmaceuticals. *Nature Biotechnology* **2009,** *27* (1), 26-34.

11. Meibohm, B., Pharmacokinetics and Pharmacodynamics of Therapeutic Peptides and Proteins. In *Pharmaceutical Biotechnology: Fundamentals and Applications*, Crommelin, D. J. A.; Sindelar, R. D.; Meibohm, B., Eds. Springer International Publishing: Cham, 2019; pp 105-137.

12. Luo, Q.; Chung, H. H.; Borths, C.; Janson, M.; Wen, J.; Joubert, M. K.; Wypych, J., Structural Characterization of a Monoclonal Antibody–Maytansinoid Immunoconjugate. *Analytical Chemistry* **2016,** *88* (1), 695-702.

13. Chari, R. V. J.; Miller, M. L.; Widdison, W. C., Antibody–Drug Conjugates: An Emerging Concept in Cancer Therapy. *Angewandte Chemie International Edition* **2014,** *53* (15), 3796-3827.

14. Peters, C.; Brown, S., Antibody–drug conjugates as novel anti-cancer chemotherapeutics. *Bioscience Reports* **2015,** *35* (4), e00225.

15. Khongorzul, P.; Ling, C. J.; Khan, F. U.; Ihsan, A. U.; Zhang, J., Antibody–Drug Conjugates: A Comprehensive Review. *Molecular Cancer Research* **2020,** *18* (1), 3-19.

16. 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. *Journal of Cellular Physiology* **2019,** *234* (5), 5628-5642.

17. 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. *Analytica Chimica Acta* **2017,** *955*, 67-78.

18. Ricart, A. D., Antibody-Drug Conjugates of Calicheamicin Derivative: Gemtuzumab Ozogamicin and Inotuzumab Ozogamicin. *Clinical Cancer Research* **2011,** *17* (20), 6417-6427.

19. Lambert, J. M.; Chari, R. V. J., Ado-trastuzumab Emtansine (T-DM1): An Antibody–Drug Conjugate (ADC) for HER2-Positive Breast Cancer. *Journal of Medicinal Chemistry* **2014,** *57* (16), 6949-6964.

20. Yoder, N. C.; Bai, C.; Tavares, D.; Widdison, W. C.; Whiteman, K. R.; Wilhelm, A.; Wilhelm, S. D.; Mcshea, M. A.; Maloney, E. K.; Ab, O.; Wang, L.; Jin, S.; Erickson, H. K.; Keating, T. A.; Lambert, J. M., A Case Study Comparing Heterogeneous Lysine- and Site-Specific Cysteine-

Conjugated Maytansinoid Antibody-Drug Conjugates (ADCs) Illustrates the Benefits of Lysine Conjugation. *Molecular Pharmaceutics* **2019,** *16* (9), 3926-3937.

21. Haque, M.; Forte, N.; Baker, J. R., Site-selective lysine conjugation methods and applications towards antibody–drug conjugates. *Chemical Communications* **2021,** *57* (82), 10689-10702.

22. 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. *Analytical Chemistry* **2021,** *93* (15), 6135-6144.

23. Zhu, X.; Huo, S.; Xue, C.; An, B.; Qu, J., Current LC-MS-based strategies for characterization and quantification of antibody-drug conjugates. *Journal of Pharmaceutical Analysis* **2020,** *10* (3), 209-220.

24. Bobály, B.; Fleury-Souverain, S.; Beck, A.; Veuthey, J.-L.; Guillarme, D.; Fekete, S., Current possibilities of liquid chromatography for the characterization of antibody-drug conjugates. *Journal of Pharmaceutical and Biomedical Analysis* **2018,** *147*, 493-505.

25. Agarwal, P.; Bertozzi, C. R., Site-Specific Antibody–Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. *Bioconjugate Chemistry* **2015,** *26* (2), 176-192.

26. 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. *Nature Biotechnology* **2012,** *30* (2), 184-189.

27. Wakankar, A.; Chen, Y.; Gokarn, Y.; Jacobson, F. S., Analytical methods for physicochemical characterization of antibody drug conjugates. *mAbs* **2011,** *3* (2), 161-172.

28. Huang, R. Y. C.; Chen, G., Characterization of antibody–drug conjugates by mass spectrometry: advances and future trends. *Drug Discovery Today* **2016,** *21* (5), 850-855.

29. Campuzano, I. D. G.; Netirojjanakul, C.; Nshanian, M.; Lippens, J. L.; Kilgour, D. P. A.; Van Orden, S.; Loo, J. A., Native-MS Analysis of Monoclonal Antibody Conjugates by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **2018,** *90* (1), 745-751.

30. Campuzano, I. D. G.; Nshanian, M.; Spahr, C.; Lantz, C.; Netirojjanakul, C.; Li, H.; Wongkongkathep, P.; Wolff, J. J.; Loo, J. A., High Mass Analysis with a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: From Inorganic Salt Clusters to Antibody Conjugates and Beyond. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (5), 1155-1162.

31. Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S.; Lu, Y.; Meng, Y. G.; Ng, C.; Yang, J.; Lee, C. C.; Duenas, E.; Gorrell, J.; Katta, V.; Kim, A.; McDorman, K.; Flagella, K.; Venook, R.; Ross, S.; Spencer, S. D.; Lee Wong, W.; Lowman, H. B.; Vandlen, R.; Sliwkowski, M. X.; Scheller, R. H.; Polakis, P.; Mallet, W., Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nature Biotechnology* **2008,** *26* (8), 925-932.

32. Song, Y.; Gao, J.; Meng, Q.; Tang, F.; Wang, Y.; Zeng, Y.; Huang, W.; Shao, H.; Zhou, H., Conjugation site characterization of antibody–drug conjugates using electron-transfer/higher-energy collision dissociation (EThcD). *Analytica Chimica Acta* **2023**, 340978.

33. Afar, D. E. H.; Bhaskar, V.; Ibsen, E.; Breinberg, D.; Henshall, S. M.; Kench, J. G.; Drobnjak, M.; Powers, R.; Wong, M.; Evangelista, F.; O'Hara, C.; Powers, D.; DuBridge, R. B.; Caras, I.; Winter, R.; Anderson, T.; Solvason, N.; Stricker, P. D.; Cordon-Cardo, C.; Scher, H. I.; Grygiel, J. J.; Sutherland, R. L.; Murray, R.; Ramakrishnan, V.; Law, D. A., Preclinical validation of anti-TMEFF2-auristatin E–conjugated antibodies in the treatment of prostate cancer. *Molecular Cancer Therapeutics* **2004,** *3* (8), 921-932.

34. Bongers, J.; Cummings, J. J.; Ebert, M. B.; Federici, M. M.; Gledhill, L.; Gulati, D.; Hilliard, G. M.; Jones, B. H.; Lee, K. R.; Mozdzanowski, J.; Naimoli, M.; Burman, S., Validation of a peptide mapping method for a therapeutic monoclonal antibody: what could we possibly learn about a method we have run 100 times? *Journal of Pharmaceutical and Biomedical Analysis* **2000,** *21* (6), 1099-1128.

35. Fleming, M. S.; Zhang, W.; Lambert, J. M.; Amphlett, G., A reversed-phase high-performance liquid chromatography method for analysis of monoclonal antibody–maytansinoid immunoconjugates. *Analytical Biochemistry* **2005,** *340* (2), 272-278.

36. Ayoub, D.; Bertaccini, D.; Diemer, H.; Wagner-Rousset, E.; Colas, O.; Cianférani, S.; Van Dorsselaer, A.; Beck, A.; Schaeffer-Reiss, C., Characterization of the N-Terminal Heterogeneities of Monoclonal Antibodies Using In-Gel Charge Derivatization of α-Amines and LC-MS/MS. *Analytical Chemistry* **2015,** *87* (7), 3784-3790.

37. Wang, L.; Amphlett, G.; Blättler, W. A.; Lambert, J. M.; Zhang, W., Structural characterization of the maytansinoid–monoclonal antibody immunoconjugate, huN901–DM1, by mass spectrometry. *Protein Science* **2005,** *14* (9), 2436-2446.

38. 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. *Journal of Chromatography B* **2015,** *981-982*, 9-13.

39. 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. *Analytica Chimica Acta* **2016,** *918*, 50-59.

40. Song, Y.; Schowen, R. L.; Borchardt, R. T.; Topp, E. M., Effect of 'pH' on the rate of asparagine deamidation in polymeric formulations: 'pH'–rate profile. *Journal of Pharmaceutical Sciences* **2001,** *90* (2), 141-156.

41. Nielsen, M. L.; Vermeulen, M.; Bonaldi, T.; Cox, J.; Moroder, L.; Mann, M., Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nature Methods* **2008,** *5* (6), 459-460.

42. Lippincott, J.; Apostol, I., Carbamylation of Cysteine: A Potential Artifact in Peptide Mapping of Hemoglobins in the Presence of Urea. *Analytical Biochemistry* **1999,** *267* (1), 57-64.

43. 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. *Analytical Chemistry* **2018,** *90* (14), 8421-8429.

44. 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. *Analytical Chemistry* **2018,** *90* (21), 12527-12535.

45. 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. *Analytical Chemistry* **2014,** *86* (6), 3005-3012.

46. Cejkov, M.; Greer, T.; Johnson, R. O. B.; Zheng, X.; Li, N., Electron Transfer Dissociation Parameter Optimization Using Design of Experiments Increases Sequence Coverage of Monoclonal Antibodies. *Journal of the American Society for Mass Spectrometry* **2021,** *32* (3), 762-771.

47. 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* **2019,** *91* (18), 11661-11669.

48. 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. *Journal of the American Society for Mass Spectrometry* **2019,** *30* (11), 2419-2429.

49. 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. *Analytical Chemistry* **2020,** *92* (22), 15096-15103.

50. 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. *Analytical Chemistry* **2020,** *92* (14), 9790-9798.

51. Fornelli, L.; Damoc, E.; Thomas, P. M.; Kelleher, N. L.; Aizikov, K.; Denisov, E.; Makarov, A.; Tsybin, Y. O., Analysis of Intact Monoclonal Antibody IgG1 by Electron Transfer Dissociation Orbitrap FTMS. *Molecular & Cellular Proteomics* **2012,** *11* (12), 1758-1767.

52. 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. *Analytical Chemistry* **2013,** *85* (9), 4239-4246.

53. 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* (1), 106-115.

54. 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. *Analytical Chemistry* **2020,** *92* (15), 10246-10251.

55. 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* (1), 766-773.

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

57. 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.

58. Fekete, S.; Guillarme, D.; Sandra, P.; Sandra, K., Chromatographic, Electrophoretic, and Mass Spectrometric Methods for the Analytical Characterization of Protein Biopharmaceuticals. *Analytical Chemistry* **2016,** *88* (1), 480-507.

59. 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. *Analytical Chemistry* **2021,** *93* (29), 10013-10021.

60. Davis, T. K.; Jennings, M. E., Site-Specific Conjugation Quantitation of a Cysteine-Conjugated Antibody–Drug Conjugate Using Stable Isotope Labeling Peptide Mapping LC–MS/MS Analysis. *Analytical Chemistry* **2022,** *94* (6), 2772-2778.

61. Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A., ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* **2021,** *20* (4), 1928–1935.

62. Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L., Analyzing internal fragmentation of electrosprayed ubiquitin ions during beam-type collisional dissociation. *J. Am. Soc. Mass Spectrom.* **2015,** *26* (5), 782-787.

63. Zenaidee, M. A.; Lantz, C.; Perkins, T.; Jung, W.; Loo, R. R. O.; Loo, J. A., Internal Fragments Generated by Electron Ionization Dissociation Enhance Protein Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020,** *31* (9), 1896–1902.

64. Schmitt, N. D.; Berger, J. M.; Conway, J. B.; Agar, J. N., Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem.* **2021,** *93* (16), 6355-6362.

65. Zenaidee, M. A.; Wei, B.; Lantz, C.; Wu, H. T.; Lambeth, T. R.; Diedrich, J. K.; Ogorzalek Loo, R. R.; Julian, R. R.; Loo, J. A., Internal Fragments Generated from Different Top-Down Mass Spectrometry Fragmentation Methods Extend Protein Sequence Coverage. *J. Am. Soc. Mass Spectrom.* **2021,** *32* (7), 1752–1758.

66. Wei, B.; Zenaidee, M. A.; Lantz, C.; Ogorzalek Loo, R. R.; Loo, J. A., Towards understanding the formation of internal fragments generated by collisionally activated dissociation for top-down mass spectrometry. *Anal. Chim. Acta* **2022,** *1194*, 339400.

67. Wei, B.; Zenaidee, M. A.; Lantz, C.; Williams, B. J.; Totten, S.; Ogorzalek Loo, R. R.; Loo, J. A., Top-down mass spectrometry and assigning internal fragments for determining disulfide bond positions in proteins. *The Analyst* **2023,** *148* (1), 26-37.

68. Li, H.; Sheng, Y.; Mcgee, W.; Cammarata, M.; Holden, D.; Loo, J. A., Structural Characterization of Native Proteins and Protein Complexes by Electron Ionization Dissociation-Mass Spectrometry. *Anal. Chem.* **2017,** *89* (5), 2731-2738.

69. Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A., An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* **2018,** *10* (2), 139-148.

70. Rolfs, Z.; Smith, L. M., Internal Fragment Ions Disambiguate and Increase Identifications in Top-Down Proteomics. *J. Proteome Res.* **2021,** *20* (12), 5412-5418.

71. 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. *J. Pharm. Biomed. Anal.* **2021,** *195*, 113893.

72. Shukla, A. A.; Hubbard, B.; Tressel, T.; Guhan, S.; Low, D., Downstream processing of monoclonal antibodies—Application of platform approaches. *Journal of Chromatography B* **2007,** *848* (1), 28-39.

73. Clauser, K. R.; Baker, P.; Burlingame, A. L., Role of Accurate Mass Measurement (±10 ppm) in Protein Identification Strategies Employing MS or MS/MS and Database Searching. *Analytical Chemistry* **1999,** *71* (14), 2871-2882.

74. Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V., Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Analytical Chemistry* **2015,** *87* (8), 4370-4376.

75. 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* (5), 291-305.

Chapter 5: Internal Fragments Enhance Middle-down Mass Spectrometry Structural Characterization of Monoclonal Antibodies and Antibody-drug Conjugates

Manuscript submitted to Analytical Chemistry

Benqian Wei[a], Carter Lantz[a], Rachel R. Ogorzalek Loo[a], Iain D.G. Campuzano[d], Joseph A. Loo[a, b, *]

[a] Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA; [b] Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA; [c] Thermo Fisher Scientific, San Jose, CA, USA; [d] Center for Research Acceleration by Digital Innovation, Molecular Analytics, Amgen Research, Thousand Oaks, CA, USA.

* Corresponding Author: Joseph A. Loo, Department of Chemistry and Biochemistry, Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA, E-mail: jloo@chem.ucla.edu.

Abstract

Monoclonal antibodies (mAbs) and antibody-drug conjugates (ADCs) are important biotherapeutics with large size (~150 kDa) and high structural complexity that require extensive sequence and structure characterization. Middle-down mass spectrometry (MD-MS) is an emerging technique that sequences and maps subunits larger than those released by trypsinolysis. It avoids introducing artifactual modifications that may occur in bottom-up MS, while achieving higher sequence coverage compared to top-down MS. However, returning complete sequence information by MD-MS is still challenging. Here, we show that assigning internal fragments in direct infusion MD-MS of a mAb and an ADC substantially improves their structural characterization. For MD-MS of the reduced NIST mAb, including internal fragments recovers nearly 100% of the sequence by accessing the middle sequence region that is inaccessible by terminal fragments. The identification of important glycosylations can also be improved after including internal fragments. For the reduced lysine-linked IgG1-DM1 ADC, we show that considering internal fragments increases the DM1 conjugation sites coverage to 80%, comparable to the reported 83% coverage achieved by peptide mapping on the same ADC.[1] This study expands our work on the application of internal fragment assignments in top-down MS of mAbs and ADCs, and can be extended to other heterogeneous therapeutic molecules such as multispecifics and fusion proteins for more widespread applications.

1. Introduction

The first monoclonal antibody (mAb) drug was approved in 1986;[2] since then, mAb-based therapeutics have become increasingly important for the treatment of a host of human diseases including cancer, metabolic disorders, and viral infections.[3-6] The success of mAbs stems from their unique pharmacological properties such as target specificity and affinity, long circulating half-life, and extraordinary safety profiles.[7, 8] MAbs possess high molecular complexity due to their large size (~150 kDa), multiple disulfide bonds within and between light and heavy chains, and a series of post-translational modifications (PTMs)[9-11] that could impact their critical quality attributes (CQAs).[12] Therefore, comprehensive sequence and structure characterization of these intricate molecules as a function of manufacturing and accelerated stability[13-15] is imperative for the production of high quality mAb therapeutics.

In recent years, new mAb formats including nanobodies, fusion proteins, multispecific antibodies, and antibody-drug conjugates (ADCs) have been evolving.[16-18] These formats have enabled new immunotherapy approaches through multitargeting and enhanced antitumor efficacy, of which ADCs have emerged as a promising therapeutic drug classes.[19-23] ADCs couple the target specificity of mAbs with the toxicity of small molecule payloads to enable their "magic bullets" feature, which allows them to selectively kill antigen-expressing targets with higher potency than their mAb counterparts.[24, 25] ADCs are even more heterogeneous molecules than mAbs due to the conjugation of payloads at varying sites, depending on the type of the linker that bridges the payload to the antibody.[22, 26] One common linker strategy targets primary amines (lysine side chains or N-termini), which produces highly heterogeneous

nonspecific lysine-linked ADCs in which a large array of locations are conjugated with differing number of payloads.[27-29] Drug conjugation sites are one of the most important CQAs of ADCs because they play a significant role in affecting the physical and pharmaceutical properties of ADCs.[30-32] Therefore, they need to be unambiguously determined to avoid the instances that the conjugation occurs in complementarity-determining regions (CDRs), which could impact the target specificity of lysine-linked ADCs.[1, 33, 34] Robust and reliable analytical techniques need to be established to resolve such molecular heterogeneity.

Among various available analytical techniques, liquid chromatography coupled with mass spectrometry (LC-MS) has been the most popular method routinely used for the characterization of mAbs and ADCs.[35-42] In particular, peptide mapping measures trypsin or Lys-C digested peptides of mAbs/ADCs by reversed phase LC-MS (RPLC-MS).[1, 33, 43-46] This bottom-up approach offers high sequence coverage with amino acid resolution and can identify drug conjugation sites of ADCs; however, it is prone to the introduction of artifactual modifications due to relatively prolonged sample preparation.[47, 48] On the other hand, top-down MS (TD-MS) measures intact gas-phase mAb/ADC ions, which minimizes sample preparation and preserves endogenous modifications, but suffers from relatively low fragmentation efficiency for proteins of this large size and high complexity.[34, 49-54]

Middle-down MS (MD-MS) is a promising technique for the characterization of mAbs/ADCs that sequences and maps subunits larger than those released by trypsinolysis.[52, 54-61] It avoids introducing artifactual modifications such as asparagine deamidation and methionine oxidation that may occur during digestion in bottom-up MS,[62-64] while achieving higher sequence coverage

compared to top-down MS; however, obtaining the same level of sequence and drug conjugation information as peptide mapping remains challenging. Various fragmentation methods have been applied to improve the MD-MS fragmentation efficiency of mAbs and ADCs including collision-,[57, 59, 60] electron-,[52, 54-59, 61] and photon-based dissociation,[52, 54, 57, 59, 61] among which electron-based dissociation (ExD) has shown promising results, particularly with the aid of collisional activation.[52, 54, 57, 61] Typically, such MD-MS experiments involve analyzing ~25 kDa subunits produced by FabRICATOR/reduced treated mAbs/ADCs using online denaturing RPLCMS. However, ExD MD-MS can require meticulous parameter tuning to achieve optimal fragmentation, rendering online RPLC-MS relatively inefficient and time-consuming due to the need to consider RPLC elution time. An alternative approach is direct infusion MD-MS, which offers higher flexibility in adjusting ExD parameters to maximize fragmentation efficiency.

In addition to applying multiple fragmentation methods, incorporating internal fragments into the data analysis workflow represents a viable strategy to enhance sequence information. These noncanonical internal fragments, which arise from multiple cleavage events on the protein backbone,[65] have been demonstrated in previous studies to significantly improve the characterization of proteins,[34, 66-72] protein complexes,[73, 74] and even proteome-wide analysis.[75] Specifically, the inclusion of internal fragment assignments has shown to be valuable in the TDMS analysis of mAbs and ADCs,[34] which motivates us to explore the employment of internal fragments in MD-MS of mAbs and ADCs to obtain more comprehensive characterization results.

In this study, we show that assigning internal fragments in native direct infusion MD-MS recovers nearly 100% of the mAb sequence and facilitates the elucidation of various types of N-

glycosylations. Notably, this represents the highest sequence coverage of mAbs achieved by methods other than peptide mapping reported to date. For a therapeutic IgG1-DM1 lysine-linked ADC, we successfully determined 80% of all putative DM1 conjugation sites, comparable to the reported 83% coverage achieved by bottom-up peptide mapping on the same ADC.[1] These results highlight the added benefits of analyzing internal fragments in MD-MS and establish MD-MS as a valuable complementary technique to the conventional peptide mapping method for characterizing a variety of therapeutic proteins.

2. Experimental Section

2.1. Materials and Reagents.

The humanized IgG1k monoclonal antibody reference material 8671 was purchased from the National Institute of Standards and Technology (NIST, Gaithersburg, MD). The therapeutic ADC supplied by Amgen is an IgG1 covalently conjugated with maytansinoid DM1 payloads on native lysine residues. Details on its preparation and production has been described previously.1 FabRICATOR (IdeS) protease was purchased from Genovis (Lund, Sweden). Tris-HCl buffer solution (pH 7.5) and tris(2-carboxyethyl)phosphine hydrochloride (TCEP) were acquired from Thermo Fisher Scientific (Bremen, Germany). Ammonium acetate solution (7.5 M) was purchased from Sigma-Aldrich (St. Louis, MO, USA) and diluted to 200 mM.

2.2. Sample Preparation.

The mAb and ADC stock samples (10 mg/ml) were diluted in 50 mM Tris-HCl buffer, followed by IdeS digestion (1 unit per µg of mAb/ADC) at 37°C for 1 hour. Subsequently, the IdeS-digested

mAb and ADC samples were reduced with 25 mM TCEP in 50 mM Tris-HCl buffer at 37°C for 90 minutes. Additionally, ADC samples were prepared with IdeS digestion alone by treating diluted ADC samples with IdeS protease at a ratio of 1 unit per µg of ADC, with incubation at 37°C for 1 hour. All reduced mAb and ADC samples were buffer exchanged into 200 mM ammonium acetate using Amicon ultra centrifugal filters (10k MWCO) and diluted to a final concentration of ~5-20 µM prior to mass spectrometry measurements.

2.3. Native Middle-down Mass Spectrometry.

All reduced samples were directly infused into a Thermo Q Exactive Plus UHMR Orbitrap (Thermo Fisher Scientific, Bremen, Germany) modified with an electromagnetostatic ExD cell (e-MSion Inc., Corvallis, OR) by nano-electrospray ionization (nESI) using Pt-coated, in-house pulled borosilicate capillaries. The capillary voltage on the nESI source was set between 1.1 and 1.7 kV. The source temperature was set at 250 °C, and the S-lens RF level was set at 200. Other crucial instrument parameters corresponding to ion transmission are listed in Table S1. In the case of NIST mAb subunit fragmentation, individual charge states of the LC subunit (ranging from 6+ to 11+), the Fd' subunit (ranging from 5+ to 9+), and the Fc/2 subunit (ranging from 5+ to 10+) were isolated in the quadrupole using a 20 m/z isolation window. For ADC subunit fragmentation, individual charge states of the LC-DM1 subunit (ranging from 7+ to 10+), the Fd'-DM1 subunit (ranging from 7+ to 9+), and the Fc/2- DM1 subunit (ranging from 6+ to 9+) were isolated in the quadrupole using a 60 m/z isolation window. Following isolation, the ions were directed from the ExD cell into the HCD cell, where electron capture dissociation (ECD) took place. Optimization of a set of seven voltage parameters in the ExD cell (details in Figure S1)

controlling the emitting and confinement of electrons was performed to facilitate efficient electron capture by the protein ions in the HCD cell. Post-ECD collisional activation was applied to minimize the impact of electron capture without dissociation (ECnoD)50 with collision energy values ranging from 70 to 150V, depending on the isolated charge state. All ECD MS/MS spectra were collected with a noise threshold set at 3, a resolution of 200,000 at m/z 400, AGC target of 1e6, and maximum inject time of 200 ms. Each spectrum was averaged over 100 to 200 scans.

2.4. Data Analysis.

2.4.1. Peak Assignments.

All raw MS/MS spectra were deconvoluted using Thermo BioPharma Finder 5.0 (Xtract algorithm). Signal-to-noise ratio (S/N) was varied between 5 to 20 to maximize the ratio of the number of matched fragments to the number of deconvoluted peaks. The resulting deconvoluted mass lists were exported as .csv files for terminal and internal fragment assignment in ClipsMS.[68] In fragment matching, the mass accuracy tolerance was set at 3 ppm, and the smallest internal fragment size was set at 5 amino acids. For NIST mAb subunit fragmentation, modifications associated with hydrogen transfer expected for ECD and neutral losses (water and ammonia) resulting from post-ECD collisional activation were considered. In the case of ADC subunit fragmentation, the modification of DM1 conjugation (956.36 Da, formula $C_{47}H_{61}ClN_4O_{13}S$) was included in the search parameters. Furthermore, for fragment assignments of the Fc subunit of both mAb and ADC, the three predominant types of N-glycosylations (G0F, G1F, and G2F) were considered. All six types of terminal fragments (*a*, *x*, *b*,

y, *c*, *z*) and only *cz* internal fragment were considered. Priority was given to the assignment of terminal fragments including their modified forms (hydrogen transfer, neutral losses, etc) *before* considering internal fragments. Any overlapping internal fragments resulting from arrangement and/or frameshift ambiguity[69] were subsequently eliminated. Following fragment assignments and duplicates removal, a two-step manual validation process was applied to further refine all matched internal fragments. First, the isotopic profile of each assigned internal fragment was examined to eliminate uncertain assignments with poor fit. Second, the assignment results were cross-referenced with theoretical fragment lists generated by ProteinProspector v6.4.2.[76] This comparison aimed to further identify and eliminate any instances of potential overlap between assigned internal fragments and theoretical terminal fragments as well as all of their possible modified forms. The assignments result of an ECD data of NIST mAb LC subunit is shown as an example in Table S2. After manual validation, the assignment results for each isolated charge state of NIST mAb subunits were combined. Similarly, the assignment results for each isolated charge state of IgG1-DM1 ADC subunits were combined.

2.4.2. Protein Sequence Coverage.

Protein sequence coverage is determined by dividing the number of observed inter-residue cleavage sites by the total number of inter-residue sites on the protein backbone.

3. Results and Discussion

3.1. A Native Direct Infusion MD-MS Platform for the Characterization of mAbs and ADCs.

We have developed an innovative native direct infusion MD-MS platform that integrates IdeS

digestion and TCEP reduction of mAbs and ADCs with nontraditional ExD fragmentation of direct-infused mAb/ADC subunit ions under native condition (Figure 1). The IdeS protease cleaves antibodies below the hinge region followed by disulfide bond reduction to produce approximately 25 kDa subunits (Figure 1A). Traditionally, these subunits are separated by RPLC, and subsequent MS/MS analysis generates sequence-informative fragment ions.[52, 55, 58, 61] This well-established workflow can be operated with high automation, achieving efficient separation of all distinct reduced subunits and their variants of mAbs or ADCs. However, the use of ECD in RPLC-MS/MS can lead to decreased experiment efficiency due to the need to consider the chromatography run time during the delicate ECD parameter tuning step.

As a result, optimizing ECD fragmentation becomes a time-consuming process. Therefore, we took an alternative approach to directly infuse reduced mAb/ADC subunits under native condition (Figure 1B). Contrary to RPLC-MS/MS, a long-lasting nano-electrospray provides increased flexibility and efficiency in fine-tuning ECD parameters to maximize fragmentation efficiency. The native condition also ensures that fewer charge states were generated for each subunit, enabling their separation in the m/z dimension in a single MS spectrum. Finally, by consolidating all validated ECD data from each isolated charge state (Figure 1C), we aim to design a comprehensive and fast workflow to obtain optimized fragmentation results.

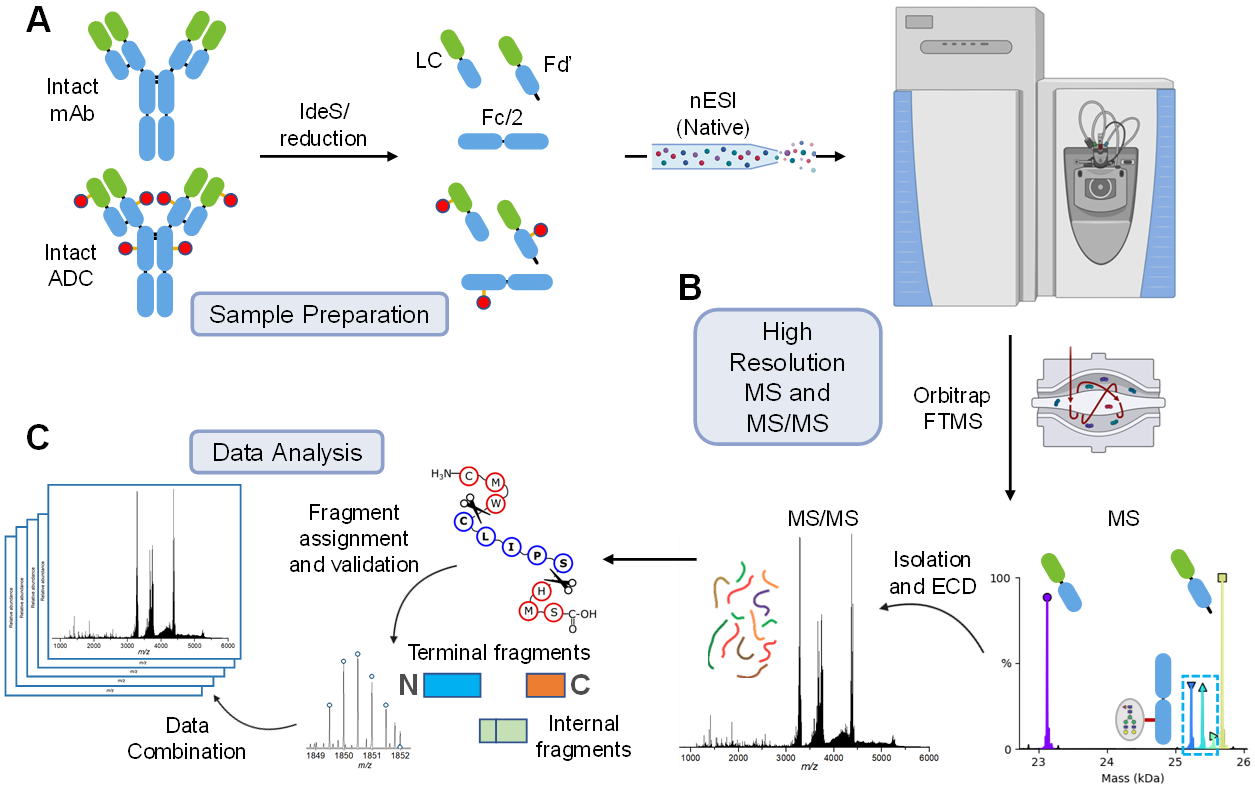


Figure 1. A schematic showing the MD-MS workflow for the characterization of mAbs and ADCs. (A) Sample preparation. The intact NIST mAb and IgG1-DM1 ADC samples were digested by IdeS FabRICATOR and reduced by TCEP into LC, Fd' and Fc/2 subunits. (B) High resolution MS and MS/MS after direct infusion by nano-electrospray under native condition. All subunits were separated in the m/z dimension followed by precursor isolation and ECD fragmentation. (C) Data analysis. Terminal and internal fragments were assigned using ClipsMS[68] and manually validated for each ECD spectrum, followed by data integration from all isolated charge states of each subunit.

3.2. Characterization of NIST mAb subunits.

All three NIST mAb subunits, LC, Fd', and Fc/2, can be effectively separated by native MS in the *m/z* dimension, allowing their subsequent isolation and ECD fragmentation (Figures 2A, 2B). Importantly, the major glycosylations occurring on the Fc/2 subunit were identified through native MS (Figure 2B). While these measurements offered a bird's-eye view of the attributes of NIST mAb, it is necessary to unambiguously determine the sequence and PTMs of the antibody

by MS/MS, in which internal fragment analysis can play a significant role. To maximize sequence information obtained by MD-MS, ECD was applied on the most abundant charge states of all three subunits (Figures 3, S2, S3). Both terminal and internal fragments were generated by ECD on the NIST LC subunit, with numerous previously unassigned signals now assigned as internal fragment ions (Figure 3A). Notably, internal fragments largely complement the sequence information obtained by terminal fragments, covering the interior sequence that is largely inaccessible to terminal fragments (Figures 3B, 3C). Similar results were observed for the Fd' and Fc/2 subunits, in which the incorporation of internal fragments substantially enhanced sequence information obtained for these subunits (Figures S2, S3).

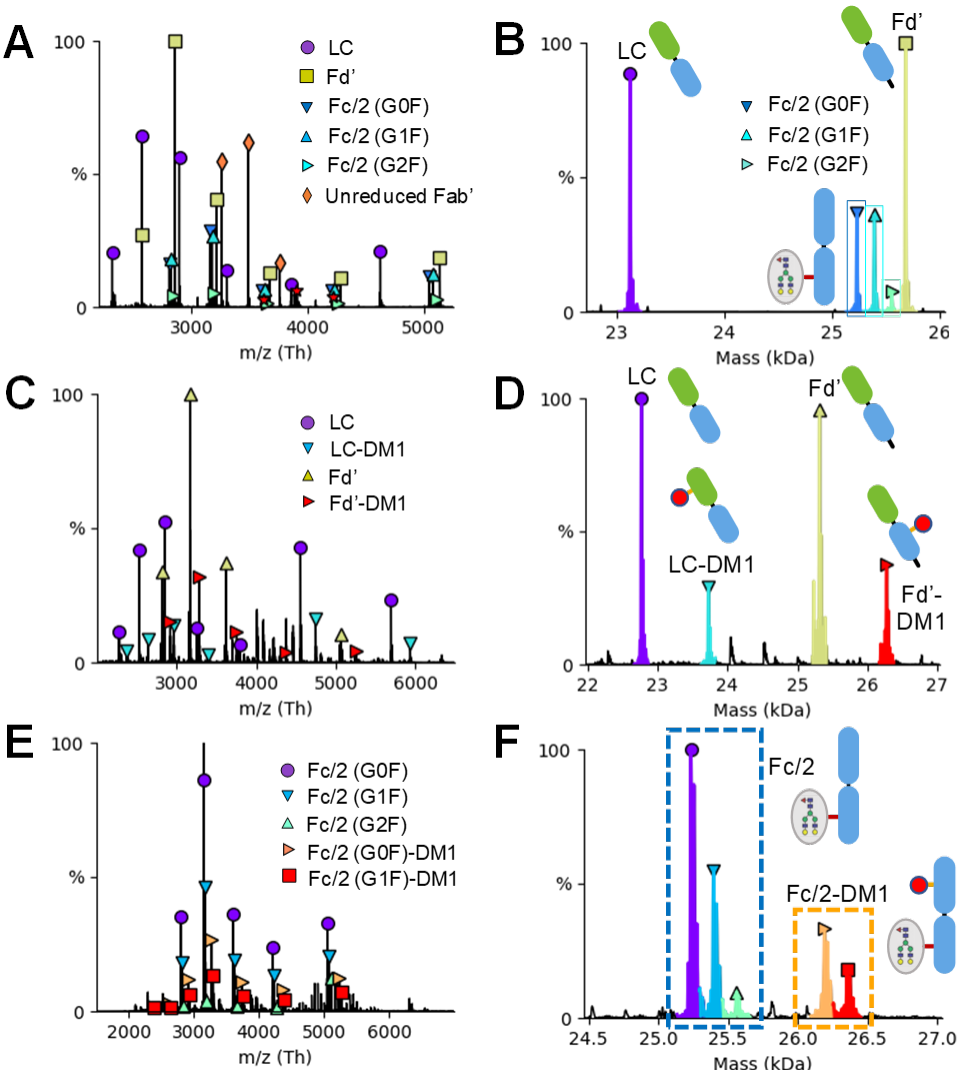


Figure 2. Native MS and deconvoluted spectra of the reduced NIST mAb and IgG1-DM1 ADC. Native MS spectra of (A) the IdeS/TCEP reduced NIST mAb, (C) the IdeS/TCEP reduced IgG1-DM1 ADC, and (E) the IdeS digested IgG1-DM1 ADC. Deconvoluted zero-charged spectra[77] of (B) the IdeS/TCEP reduced NIST mAb, (D) the IdeS/TCEP reduced IgG1-DM1 ADC, and (F) the IdeS digested IgG1-DM1 ADC. All three subunits of the NIST mAb, all three subunits and their DM1-bound forms of the ADC can be separated and observed from our native direct infusion middle-down experiments.

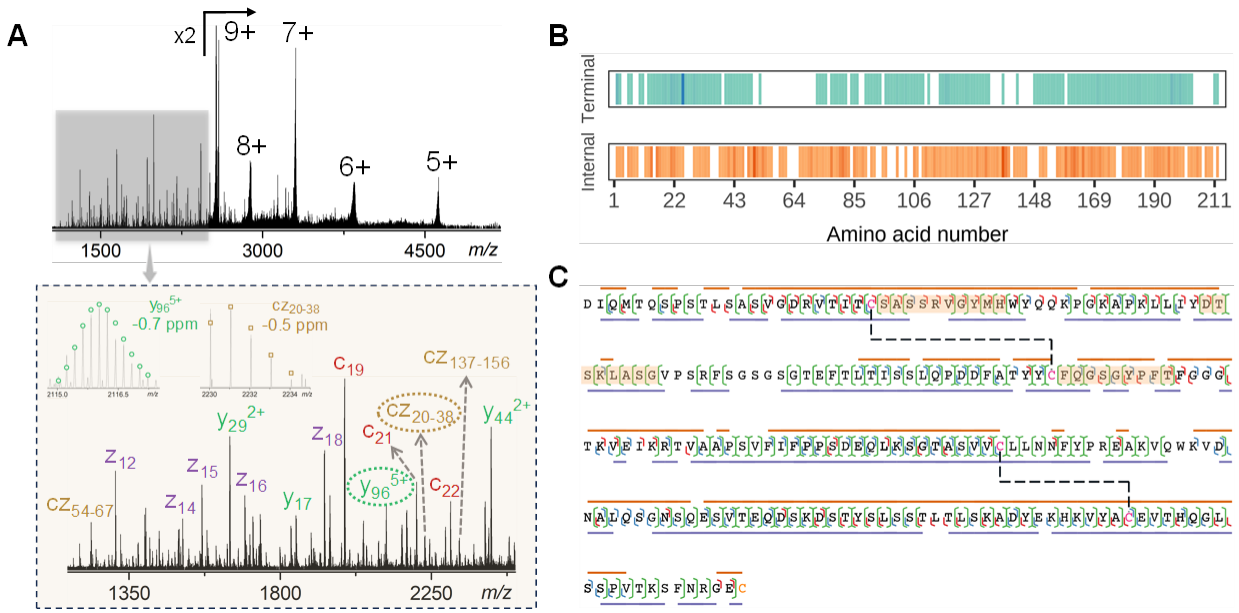


Figure 3. MD-MS characterization of the NIST mAb LC subunit.
(A) A representative ECD MS/MS spectrum of the NIST mAb LC subunit, [NIST-LC + 9H]$^{9+}$, with a zoomed-in spectrum in the range from 1200 to 2500 *m/z* showing both terminal and internal fragments are generated. Theoretical isotope distributions are overlaid on representative terminal and internal fragment ions to confirm the assignments. (B) A sequence map showing the sequence coverage achieved by terminal (top panel) and internal (bottom panel) fragments. Deeper color indicates higher fragment intensity on the cleavage site. (C) A fragmentation map showing cleavage sites by terminal and internal fragments. Blue, red, and green cleavages on the protein backbone represent *a* or *b*/*x* or *y* terminal, *c*/*z*· terminal, and *cz*· internal fragments, respectively. The solid line above the sequence represents terminal fragment sequence coverage, while the solid line beneath the sequence represents internal fragment sequence coverage. The black dashed lines represent intrachain disulfide bonds, with the complementarity-determining regions (CDRs) covered in orange.

By incorporating internal fragment assignments, we achieved near complete sequence coverage of NIST mAb LC, Fd', and Fc/2 subunits. Through the integration of one optimized ECD data per charge state for each subunit, the sequence coverage increased from 74% to 95% for the LC subunit, 58% to 92% for the Fd' subunit, and 55% to 92% for the Fc/2 subunit after considering internal fragments (Figure 4, Table S3). Unsurprisingly, this significant improvement is largely due to the ability of internal fragments to access the interior protein sequence. The

coverage increased from 76% to 100% for the LC subunit middle region (residues 71-142), 37% to 91% for the Fd' subunit middle region (residues 79-158), and 37% to 89% for the Fc/2 subunit middle region (residues 70-140) with the inclusion of internal fragments (Figure 4). Furthermore, internal fragments possess two cleavage sites on the protein backbone, whereas terminal fragments only cleave the protein once. This difference enables internal fragments to inherently carry more sequence information than terminal fragments, contributing to the enhancement in sequence coverage.

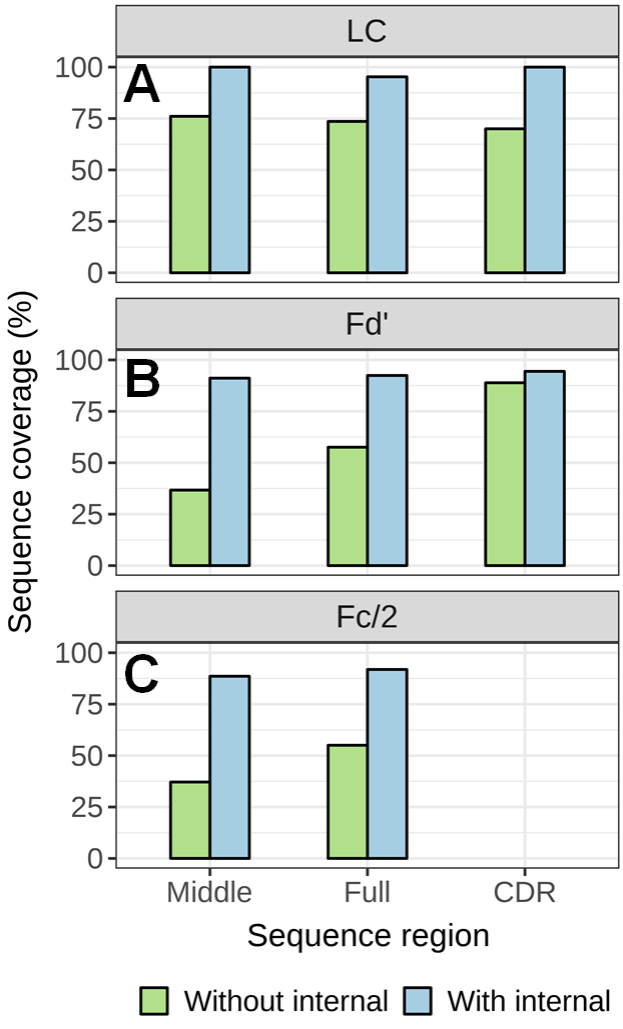


Figure 4. Sequence coverage analysis of NIST mAb subunits.

Sequence coverages of different sequence regions including the middle sequence (“Middle”), whole sequence (“Full”), and CDR sequence (“CDR”) before and after considering internal fragments of the (A) LC, (B) Fd’, and (C) Fc/2 subunits.

The complementarity-determining regions (CDRs) play a vital role in defining the antigen specificity of mAbs; thus unambiguous determination of their sequences is necessary. In addition, the possible presence of problematic chemical liabilities such as deamidation, isomerization, and oxidation within the CDRs highlights the need to thoroughly characterize the CDRs sequences.[78] Overall, the assignment of internal fragments improved sequence coverages of both LC CDRs, increasing from 70% to 100% (Figure 4A), and Fd' CDRs, increasing from 89% to 94% (Figure 4B). Only two missed cleavages, including K66|D67 within CDR2 of the Fd’ subunit, and R99|D100 at the beginning of CDR3 of the Fd’ subunit, were observed across all CDRs of the NIST mAb after including internal fragments. However, the first missed cleavage (K66|D67) can be covered by incorporating adjusted ECD data (vide infra), and the cleavage occurring at D100|M101 captures any potential chemical liability that may occur at residue D100 as R99 is not part of CDR3 of the Fd’ subunit. This demonstrates the power of incorporating internal fragments to comprehensively and unambiguously determine the CDRs sequence. Lastly, although with little automation and relatively low throughput, the high flexibility of our native direct infusion MDMS platform allowed for the integration of one or two additional adjusted ECD datasets per charge state of each subunit with minimal time loss. This further increased the sequence coverage of all three subunits even closer to 100% (98% for the LC subunit, 97% for the Fd’ subunit, and 97% for the Fc/2 subunit), with only one single missed

cleavage observed in the CDRs sequence (Table S4).

In addition to enhancing sequence information, the assignment of internal fragments also contributes to identifying various types of prevalent N-glycosylations. By consolidating one ECD data per charge state, MD-MS of the NIST mAb Fc subunit generated a total of 21 terminal fragments containing G0F (11), G1F (4), and G2F (6) N-glycans, while 33 internal fragments were generated containing these three predominant N-glycosylations, with 19 containing G0F, 8 containing G1F, and 6 containing G2F (Figure S4A). This highlights the value of analyzing internal fragments for N-glycosylation identification. The flexibility of tuning ECD parameters within our native direct infusion MD-MS platform further improved the detection of N-glycosylations. By integrating one or two additional adjusted ECD datasets per charge state, we observed an increase of the number of assigned G0F/G1F/G2F-bound terminal and internal fragments to 42 and 50, respectively (Figure S4B). Internal fragments can potentially play a bigger role in identifying low-level problematic PTMs such as oxidation or deamidation, as these modifications may not be accessible by terminal fragments.

3.3. Characterization of IgG1-DM1 ADC Subunits.

The naked antibody of the IgG1-DM1 ADC used in this study contains a total of 90 putative conjugation sites, including 11 from the LC subunit, 16 from the Fd' subunit, and 18 from the Fc/2 subunit. The large number of potential conjugation sites and the inherent random nature of lysine conjugation result in the high heterogeneity of this ADC. To determine the DM1 conjugation sites of the ADC, we took a similar native direct infusion MD-MS approach by applying ECD on isolated DM1-bound subunit ions of the ADC. However, the close mass

similarity between the Fd' and Fc/2 subunits of the antibody used in this ADC makes it challenging to isolate each subunit individually within a single spectrum due to potential peak overlaps. Therefore, we performed two reduction experiments, one involving both IdeS digestion and TCEP reduction (Figure 2C and D), and the other with IdeS digestion alone (Figure 2E and F), to produce Fd' and Fc/2 subunits separately. Native MS of the reduced ADC offers a global overview of the conjugation level and reveals that DM1 is conjugated to all three subunits of the antibody, confirming the high heterogeneity of this ADC (Figures 2D, 2F). ECD on all three DM1-bound subunits generated both terminal and internal fragments as well as their DM1-bound forms, providing direct evidence to determine the DM1 conjugation sites (Figures 5A, S5A, S6A).

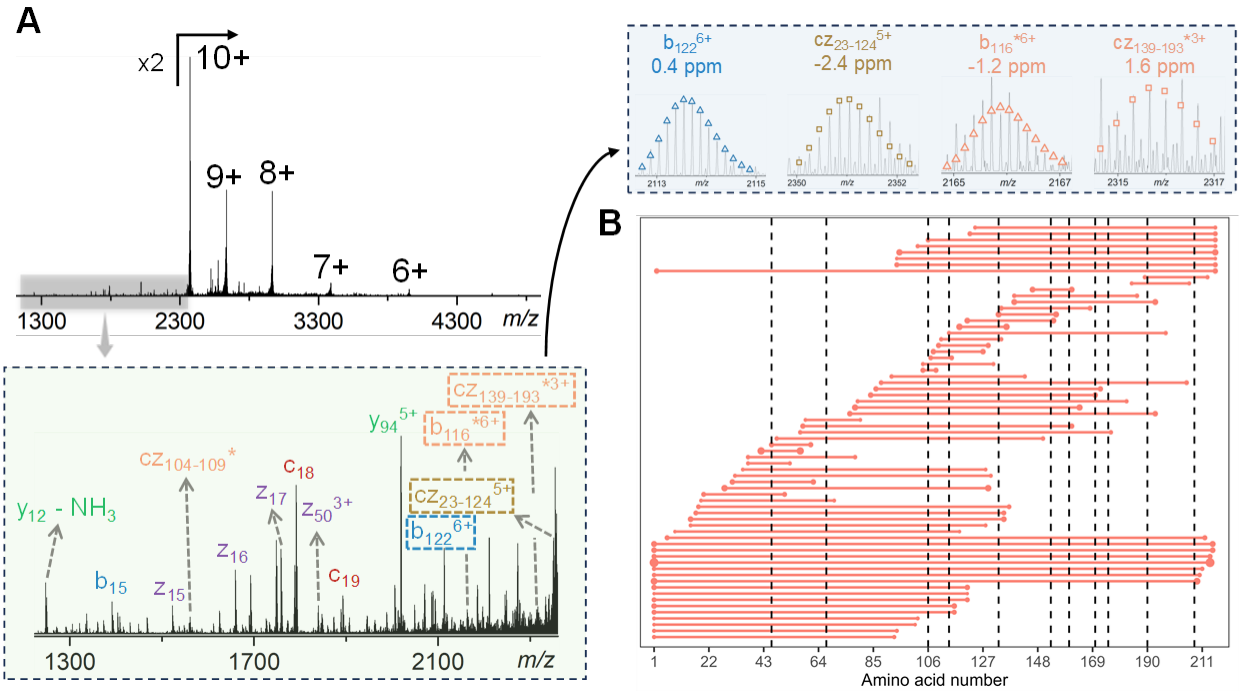


Figure 5. Middle-down characterization of the IgG1-DM1 ADC LC subunit.
(A) A representative ECD MS/MS spectrum of the ADC LC subunit, $[ADC\text{-}LC + 10H]^{10+}$, with a zoomed-in spectrum in the range from 1100 to 2400 *m/z* showing both terminal and internal fragments, and their DM1-bound forms are generated. A superscripted asterisk (*) indicates fragments that contain a DM1 payload. Theoretical isotope distributions are overlaid on representative terminal and internal fragments to confirm the assignments. (B) A fragment

location map of the ADC LC subunit showing all DM1-bound fragments after combining ECD data from all isolated charge states. Black vertical dashed lines represent lysine positions.

Similar to our report on ADC characterization using TD-MS,[34] here we define *localizing a conjugation site* as being able to specify the exact lysine residue where the conjugation occurs, while *identifying a conjugation site* refers to confirming the conjugation on several possible lysine residues without pinpointing the exact one. ECD analysis of the LC subunit generated 24 DM1-bound terminal fragments and 43 DM1-bound internal fragments, enabling the localization of 6 conjugation sites and identification of 3 additional conjugation sites (Figure 5B). Assigning DM1-bound terminal fragments could only identify 2 conjugation sites without localizing any (Figure 5B). In contrast, adding DM1-bound internal fragments allowed us to localize 6 conjugation sites (K46, K67, K106, K114, K133, K190) and identify 3 extra conjugation sites (Figure 5B). Specifically, the conjugation site at K46 was localized by 5 DM1-bound internal fragments (cz_{20-51}, cz_{31-62}, cz_{37-53}, cz_{42-57}, cz_{46-61}), K67 by DM1-bound cz_{59-80}, K106 by DM1-bound $cz_{104-109}$, K114 by 3 DM1-bound internal fragments ($cz_{107-115}$, $cz_{108-127}$, $cz_{110-129}$), K133 by DM1-bound $cz_{118-136}$, and K190 by DM1-bound $cz_{184-206}$ (Figure 5B). Similar results were observed for the Fd' and Fc/2 subunits (Figure S5 and S6). ECD of the Fd' subunit generated 35 DM1-bound terminal fragments and 30 DM1-bound internal fragments (Figure S5). Only 2 conjugations sites were identified with terminal fragments alone, whereas 4 conjugation sites were localized (K13, K43, K67, K153) and 5 additional conjugation sites were identified with the inclusion of internal fragments (Figure S5). In the case of Fc subunit, 43 DM1-bound terminal fragments and 72 DM1-bound internal fragments were generated by ECD (Figure S6). The

assignment of internal fragments localized 5 conjugations sites (K98, K124, K134, K178, K203) and identified 8 other conjugation sites, significantly improved from 2 identified conjugation sites with terminal fragments alone (Figure S6).

The determination status of each potential conjugation site across all three subunits is summarized in Figure 6. The *identified* but not *localized* conjugation sites were shown as blue-colored residues (*determined*) on the lysine site closest to either terminus for illustration purposes. Terminal fragments mainly determined conjugations sites close to the termini while internal fragments largely improved the determination of interior conjugations sites (Figure 6). In total, the incorporation of internal fragments resulted in the determination of 62 conjugation sites (30 *localized*, 32 *identified*), covering 69% of all putative conjugation sites of the antibody (Table S5). Importantly, the flexibility of our native direct infusion MD-MS system shows promising value for characterizing the IgG1-DM1 ADC. By adding one or two ECD datasets per charge state for each subunit, we increased the number of determined conjugation sites from 62 to 72 (44 *localized*, 28 *identified*), resulting in an 80% DM1 conjugation site coverage (Table S6). This result is comparable to the reported 83% coverage obtained from peptide mapping on the same ADC,1 demonstrating the immense value of analyzing internal fragments for characterizing ADCs by MD-MS, boosting its performance to a level close to that of peptide mapping.

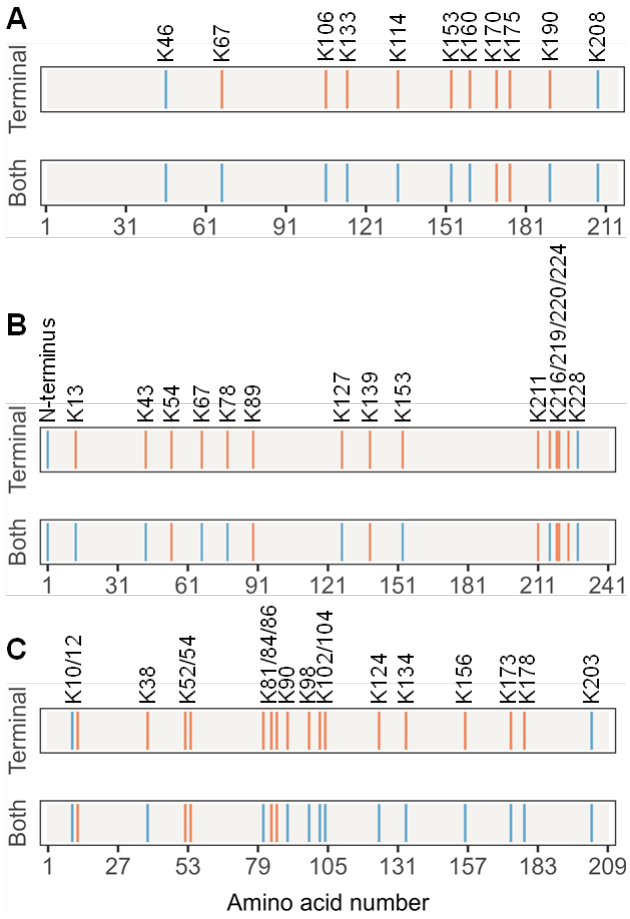


Figure 6. A sequence map that shows the conjugation determination status of all potential conjugation sites.

Conjugation determination status of each putative conjugation site for the (A) LC, (B) Fd', and (C) Fc/2 subunit of the IgG1-DM1 ADC with only terminal fragments considered (top panel) and both terminal and internal fragments considered (bottom panel). All potential conjugation sites (all K residues and the N-terminus of the Fd' subunit) are highlighted. Orange-colored residues represent *undetermined* conjugation sites and blue-colored residues represent *determined* conjugation sites.

4. Conclusions

Here we report the primary benefits of including internal fragments for the MD-MS characterization of the NIST mAb and a heterogeneous lysine-linked IgG1-DM1 ADC. We developed a native direct infusion MD-MS platform that provides high flexibility to maximize

ECD performance with high efficiency, which is difficult to achieve by traditional MD-MS methods using RPLC. The assignment of internal fragments increases the sequence coverage of all three NIST mAb subunits to nearly 100% by accessing the interior protein sequence that is challenging to probe by terminal fragments. Important N-glycosylation information can be elucidated by analyzing internal fragments, which opens the potential of applying internal fragments to identify other low-level PTMs, such as deamidation, oxidation, isomerization, and other unexpected PTMs. In addition, we show that assigning internal fragments significantly improves the determination of drug conjugation sites of a IgG1-DM1 ADC to achieve an 80% drug conjugation site coverage, comparable to the 83% coverage obtained from the routinely utilized bottom-up RPLC-MS/MS peptide mapping approach.[1]

The results presented here build upon and demonstrate improvements from our previous work on applying internal fragments for native TD-MS characterization of intact mAbs and ADCs.[34] Although TD-MS offers easier sample preparation, it cannot reach the extensiveness of MD-MS characterization in terms of sequence and drug conjugation coverage. Nonetheless, our TD-MS platform possesses a unique advantage in determining intra-chain disulfide connectivity which cannot be achieved by MD-MS.[34] While bottom-up MS remains a well-established approach in the pharmaceutical industry, the MD-MS platform described in this study achieves comparable results and holds great potential if supported by robust automation (e.g., sample processing, nESI) and informatics tools. The reduced sample handling of MD-MS compared to bottom-up MS decreases the risk for introducing sample preparation-related artifacts. These findings highlight the potential of MD-MS and internal fragment analysis to enhance the

analytical capabilities of MS-based methodologies in characterizing biotherapeutic proteins. It also features the increasing role of native MS could play in therapeutic protein analysis. Furthermore, this study suggests that incorporating internal fragments into the bottom-up MS workflow could enable comprehensive characterization of mAbs and ADCs on a routine basis.

Associate Content

*sı Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Instrument parameters and ECD cell parameters; middle-down characterization of the NIST mAb and IgG1-DM1 ADC Fd' and Fc/2 subunits; assignment result of an ECD data of NIST mAb LC subunit; measured sequence coverage of subunits.

Author Information

Corresponding Author

Joseph A. Loo − Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States; Department of Biological Chemistry, UCLA-DOE Institute, and Molecular Biology Institute, University of California—Los Angeles, Los Angeles, California 90095, United States; orcid.org/0000-0001-9989-1437; Email: jloo@chem.ucla.edu

Authors

Benqian Wei − Department of Chemistry and Biochemistry, University of California—Los Angeles,

Los Angeles, California 90095, United States; orcid.org/0000-00034853-4848.

Carter Lantz − Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States.

Rachel R. Ogorzalek Loo − Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States; UCLA-DOE Institute and Molecular Biology Institute, University of California—Los Angeles, Los Angeles, California 90095, United States; orcid.org/0000-0002-0580-2833.

Iain D. G. Campuzano − Center for Research Acceleration and Digital Innovation, Molecular Analytics, Amgen Research, Thousand Oaks, California 91320, United States; orcid.org/0000-0003-4310-8540.

Notes:

The authors declare no competing financial interest.

Acknowledgements

J.A.L. and R.R.O.L. acknowledge support from the US National Institutes of Health (R35GM145286) and the US Department of Energy (DE-FC02-02ER63421). C.L. acknowledges support from the Ruth L. Kirschstein National Research Service Award program (GM007185). B.W acknowledge support from the Agilent/ACS Analytical Graduate Fellowship.

Chapter 5: Supporting Information

Supplementary Figures

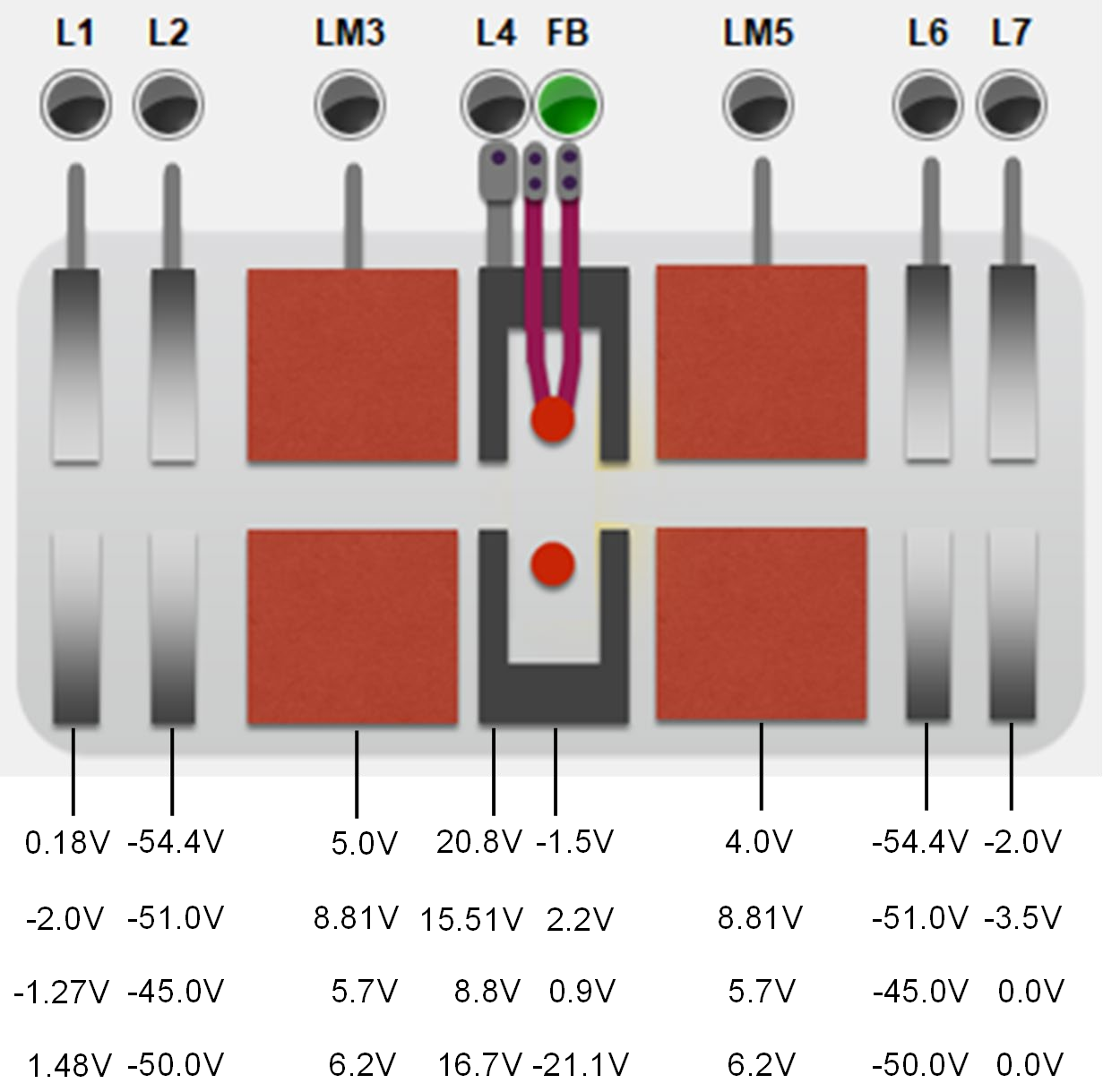


Figure S1. ExD cell parameters.
Four major sets of ExD cell parameters used for ExD fragmentation of LC, Fd', and Fc/2 ions of reduced NIST mAb and IgG1-DM1 ADC.

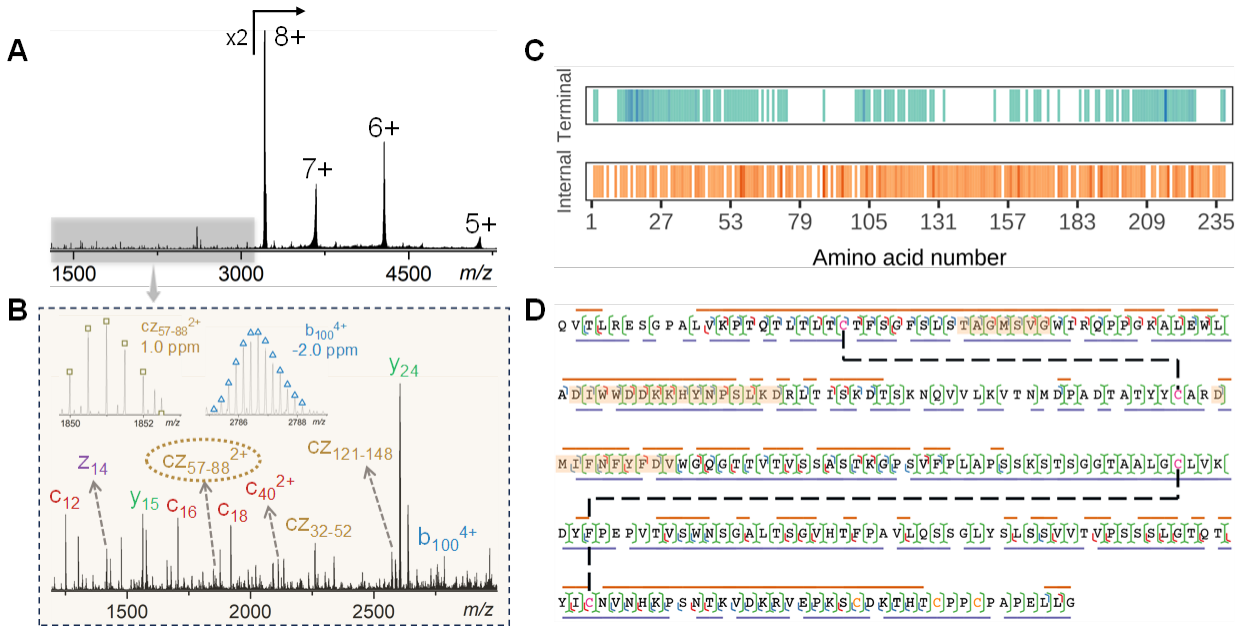


Figure S2. Middle-down characterization of the NIST mAb Fd' subunit.

(A) A representative ECD MS/MS spectrum of the NIST mAb Fd' subunit, $[NIST\text{-}Fd' + 8H]^{8+}$, with a zoomed-in spectrum in the range from 1100 to 3000 *m/z* showing both terminal and internal fragments are generated. Theoretical isotope distributions are overlaid on representative terminal and internal fragment ions to confirm the assignments. (B) A sequence map showing the sequence coverage achieved by terminal (top panel) and internal (bottom panel) fragments. Deeper color indicates higher fragment intensity on the cleavage site. (C) A fragmentation map showing cleavage sites by terminal and internal fragments. Blue, red, and green cleavages on the protein backbone represent *a* or *b*/*x* or *y* terminal, *c*/*z*· terminal, and *by*/*cz*· internal fragments, respectively. The solid line above the sequence represents terminal fragment sequence coverage, while the solid line beneath the sequence represents internal fragment sequence coverage. The black dashed lines represent intrachain disulfide bonds, with the complementarity-determining regions (CDRs) covered in orange.

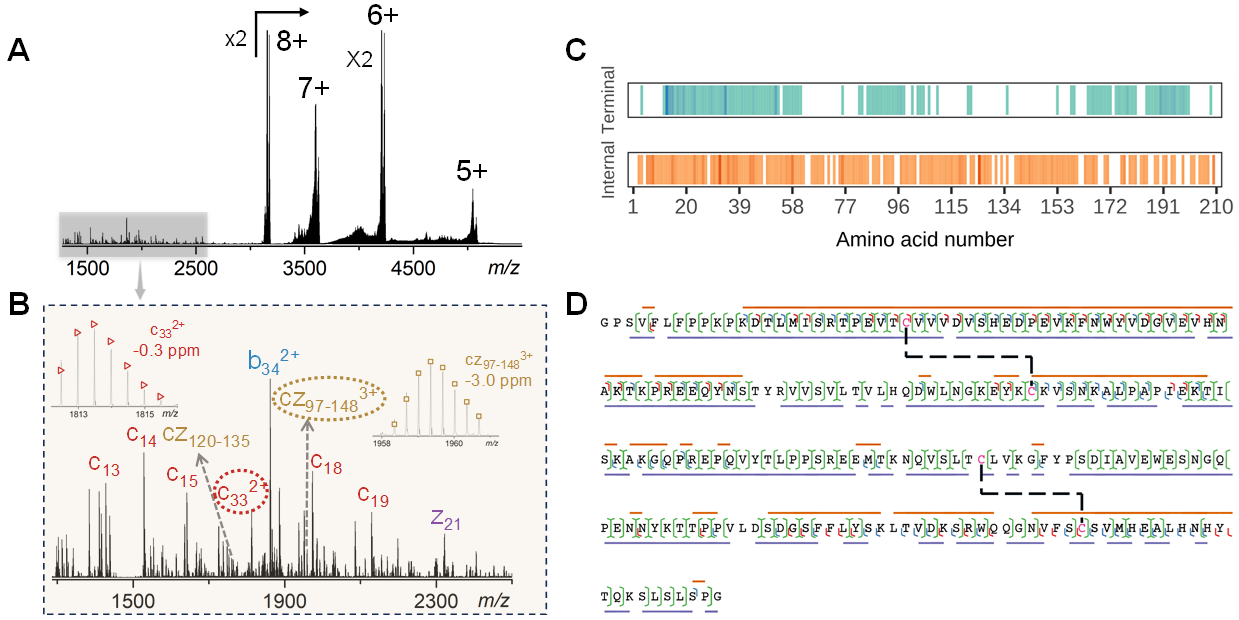


Figure S3. Middle-down characterization of the NIST mAb Fc/2 subunit.

(A) A representative ECD MS/MS spectrum of the NIST mAb Fc/2 subunit, $[NIST\text{-}Fc/2 + 8H]^{8+}$, with a zoomed-in spectrum in the range from 1100 to 2500 *m/z* showing both terminal and internal fragments are generated. Theoretical isotope distributions are overlaid on representative terminal and internal fragment ions to confirm the assignments. (B) A sequence map showing the sequence coverage achieved by terminal (top panel) and internal (bottom panel) fragments. Deeper color indicates higher fragment intensity on the cleavage site. (C) A fragmentation map showing cleavage sites by terminal and internal fragments. Blue, red, and green cleavages on the protein backbone represent *a* or *b*/*x* or *y* terminal, *c*/*z*· terminal, and *by*/*cz*· internal fragments, respectively. The solid line above the sequence represents terminal fragment sequence coverage, while the solid line beneath the sequence represents internal fragment sequence coverage. The black dashed lines represent intrachain disulfide bonds, with the complementarity-determining regions (CDRs) covered in orange.

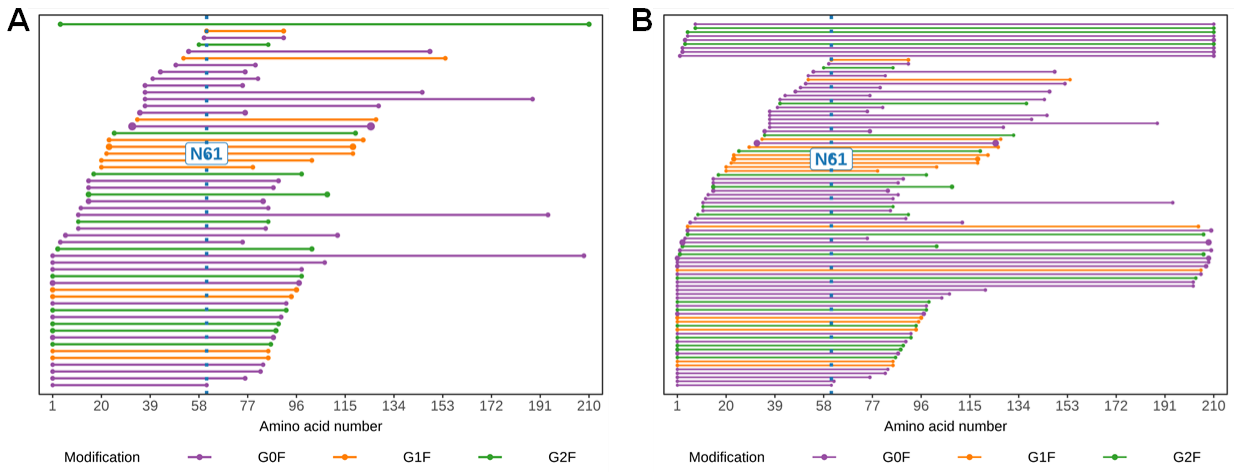


Figure S4. Fragment location maps of NIST mAb Fc/2 subunit showing fragments containing N-glycosylations.

(A) One ECD data per charge state, (B) when one or two additional adjusted ECD datasets per charge state were added to the data shown in panel A. Each horizontal solid line represents a fragment containing an N-glycosylation modification. The blue-colored vertical dotted line indicates the position of asparagine (N) 61, the residue that is known to be highly glycosylated for the NIST mAb.

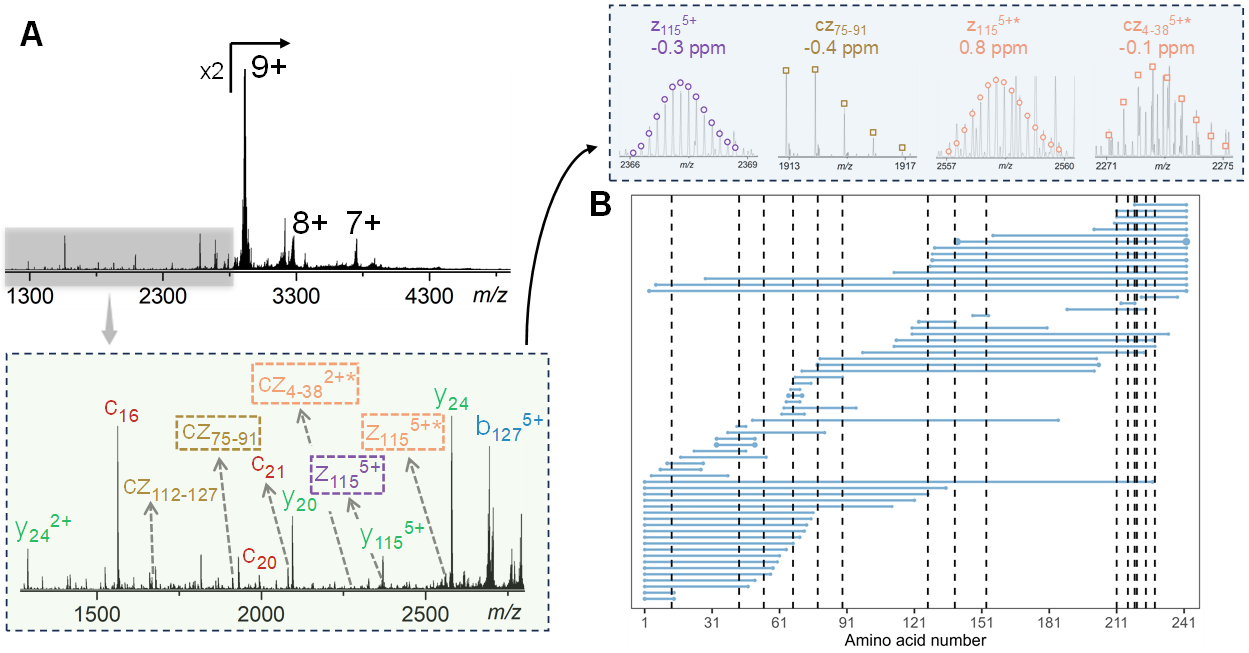


Figure S5. Middle-down characterization of the IgG1-DM1 ADC Fd' subunit.

(A) A representative ECD MS/MS spectrum of the ADC mAb Fd' subunit, [ADC-Fd' + 9H]9+, with a zoomed-in spectrum in the range from 1100 to 2800 m/z showing both terminal and internal fragments, and their DM1-bound forms are generated. Theoretical isotope distributions are overlaid on representative terminal and internal fragments to confirm the assignments. A superscripted asterisk (*) indicates fragments that contain a DM1 molecule. (B) A fragment location map of the ADC Fd' subunit showing all DM1-bound fragments after combining ECD data from all isolated charge states. Black vertical dashed lines represent lysine positions.

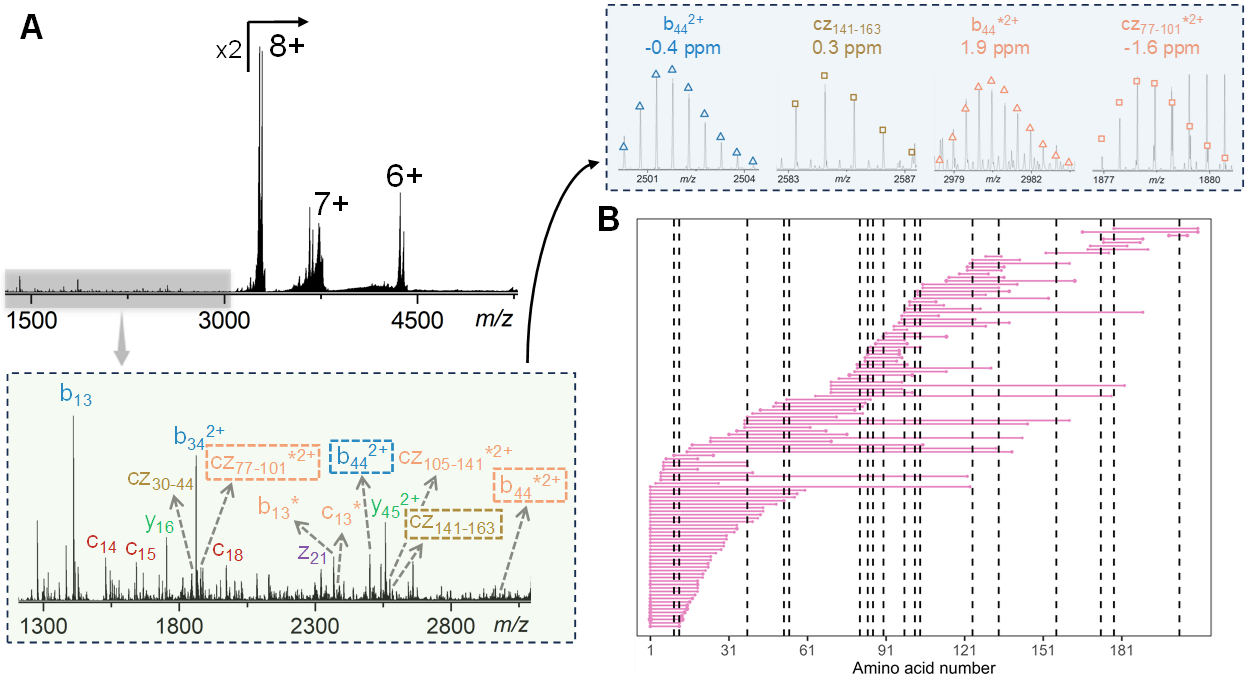


Figure S6. Middle-down characterization of the IgG1-DM1 ADC Fc/2 subunit.
(A) A representative ECD MS/MS spectrum of the ADC mAb Fc subunit, $[ADC\text{-}Fc/2 + 8H]^{8+}$, with a zoomed-in spectrum in the range from 1100 to 3000 *m/z* showing both terminal and internal fragments, and their DM1-bound forms are generated. Theoretical isotope distributions are overlaid on representative terminal and internal fragments to corroborate the assignments. A superscripted asterisk (*) indicates fragments that contain a DM1 molecule. (B) A fragment location map of the ADC Fc/2 subunit showing all DM1-bound fragments after combining ECD data from all isolated charge states. Black vertical dashed lines represent lysine positions.

Supplementary Tables

Table S1. Instrument parameters crucial for ion transmission of reduced NIST mAb and IgG1-DM1 ADC ions.

UHMR Instrument

Detector *m/z* Optimization	Low *m/z*	Source DC Offset (V)	21	In-source Trapping (V)	-100
Ion Transfer Target *m/z*	High *m/z*	Injection Flatapole DC (V)	5	Inter Flatapole Lens (V)	4
Bent Flatapole DC (V)	2	Transfer Multipole DC (V)	0	C-Trap Entrance Lens Inject (V)	1.8
Extended Trapping (eV)	10				

Table S2. The assignment result of an ECD data of NIST mAb LC subunit.

Frag Type	Unlocalized Mod	Observed Mass	Theoretical Mass	Start AA	End AA	Error
A	2H+	1947.977	1947.980	1	19	-1.757
A	2H+	2162.107	2162.112	1	21	-2.389
A	None	2261.141	2261.144	1	22	-1.413
A	2H+	2524.245	2524.238	1	25	2.705
A	Water	2933.443	2933.446	1	29	-0.888
A	None	2951.462	2951.456	1	29	1.997
A	None	3008.478	3008.478	1	30	0.143
A	None	3439.637	3439.640	1	33	-0.958
A	None	3625.715	3625.720	1	34	-1.271
B	None	1546.746	1546.742	1	15	2.744
B	Water	1585.751	1585.753	1	16	-1.084
B	None	1718.791	1718.790	1	17	0.487
B	Water	2862.376	2862.372	1	28	1.360
B	None	8280.151	8280.148	1	78	0.375
B	H-	8491.206	8491.220	1	80	-1.623
B	2H-	12396.122	12396.125	1	116	-0.216
B	2H-	12543.181	12543.193	1	117	-0.964
B	2H-	12939.346	12939.358	1	121	-0.896
B	2H-	13437.62	13437.638	1	125	-1.324
B	2H-	16243.068	16243.065	1	150	0.183
B	2H-	17930.783	17930.794	1	166	-0.589
B	None	17932.776	17932.809	1	166	-1.852
B	2H-	18260.937	18260.948	1	169	-0.575
B	H+	22994.261	22994.300	1	212	-1.680
C	None	1563.769	1563.768	1	15	0.445
C	None	1620.792	1620.790	1	16	1.377
C	None	1735.818	1735.817	1	17	0.742
C	None	1891.916	1891.918	1	18	-0.964
C	None	1990.987	1990.986	1	19	0.383
C	None	2092.034	2092.034	1	20	0.040

C	None	2205.117	2205.118	1	21	-0.444
C	None	2306.165	2306.166	1	22	-0.286
C	None	2409.174	2409.175	1	23	-0.350
C	None	2496.204	2496.207	1	24	-1.151
C	None	2567.245	2567.244	1	25	0.395
C	None	2654.273	2654.276	1	26	-1.136
C	None	2897.411	2897.409	1	28	0.636
C	None	2996.473	2996.478	1	29	-1.525
C	None	3053.498	3053.499	1	30	-0.339
C	None	3216.563	3216.562	1	31	0.198
C	None	3347.608	3347.603	1	32	1.539
C	None	3484.656	3484.662	1	33	-1.653
C	H-	3669.736	3669.733	1	34	0.750
C	None	3833.8	3833.804	1	35	-1.148
C	None	3961.865	3961.863	1	36	0.510
C	None	4372.082	4372.091	1	40	-2.001
C	None	4500.187	4500.186	1	41	0.286
C	2H-	10354.99	10354.989	1	96	0.132
C	2H-	10502.054	10502.057	1	97	-0.290
C	2H-	10673.1	10673.121	1	100	-2.009
C	H-	10903.282	10903.272	1	102	0.926
C	H-	11002.325	11002.340	1	103	-1.392
C	2H-	11243.464	11243.459	1	105	0.431
C	H-	11372.548	11372.562	1	106	-1.226
C	2H-	12841.367	12841.357	1	120	0.763
C	2H-	13326.56	13326.569	1	124	-0.704
C	H-	14056.959	14056.979	1	132	-1.445
C	H-	14614.233	14614.242	1	137	-0.649
C	2H-	15305.577	15305.563	1	142	0.923
C	None	16946.395	16946.426	1	157	-1.849
CZ	H+	1235.589	1235.589	54	67	0.066
CZ	None	1885.939	1885.939	167	184	0.034
CZ	2H+	2230.052	2230.053	20	38	-0.543
CZ	None	2332.184	2332.183	137	156	0.302

CZ	H-	2532.254	2532.250	137	158	1.411
CZ	H+	2553.268	2553.271	54	79	-1.001
CZ	None	2689.335	2689.336	169	192	-0.209
CZ	H+	2817.396	2817.404	139	163	-2.850
CZ	H+	2894.419	2894.427	117	142	-2.674
CZ	H+	2977.466	2977.468	137	162	-0.569
CZ	H+	3035.503	3035.508	48	77	-1.720
CZ	None	3076.518	3076.511	165	191	2.288
CZ	None	3110.53	3110.528	51	81	0.747
CZ	None	3121.543	3121.541	181	208	0.591
CZ	H+	3124.534	3124.533	14	41	0.335
CZ	None	3203.581	3203.587	135	162	-1.829
CZ	None	3211.581	3211.575	50	81	1.758
CZ	H-	3248.59	3248.591	168	196	-0.176
CZ	H-	3264.591	3264.592	13	42	-0.162
CZ	None	3328.637	3328.633	51	83	1.141
CZ	None	3563.749	3563.739	161	192	2.864
CZ	None	3690.828	3690.833	128	160	-1.440
CZ	H+	4593.243	4593.235	82	123	1.748
CZ	2H-	4794.317	4794.307	68	111	2.033
CZ	2H-	4866.337	4866.328	128	171	1.942
CZ	2H+	5150.522	5150.519	137	182	0.599
CZ	2H-	6292.094	6292.093	112	168	0.129
CZ	H+	7247.563	7247.575	46	114	-1.640
CZ	None	9026.355	9026.381	48	133	-2.864
CZ	2H-	11447.581	11447.585	85	188	-0.310
CZ	2H-	12558.247	12558.257	95	209	-0.766
CZ	2H-	12639.24	12639.243	12	130	-0.255
CZ	H+	15375.621	15375.665	2	144	-2.857
X	Water	1419.636	1419.632	201	213	2.670
X	2H+	1737.824	1737.822	198	213	0.883
X	2H+	1874.881	1874.881	197	213	-0.201
X	2H-	13307.533	13307.537	93	213	-0.293
X	2H-	13451.619	13451.590	91	213	2.127

X	2H-	22910.203	22910.164	3	213	1.695
X	2H+	22914.197	22914.195	3	213	0.067
X	Water	23007.293	23007.253	2	213	1.722
X	2H-	23023.245	23023.248	2	213	-0.140
Y	None	1411.667	1411.663	201	213	2.519
Y	None	1581.772	1581.769	199	213	1.914
Y	2H-	1707.812	1707.812	198	213	0.059
Y	None	1709.828	1709.828	198	213	0.263
Y	None	1846.887	1846.886	197	213	0.291
Y	None	2047.006	2047.003	195	213	1.683
Y	None	2176.045	2176.045	194	213	-0.068
Y	None	2350.091	2350.091	192	213	-0.191
Y	None	2513.155	2513.155	191	213	0.089
Y	None	2877.381	2877.377	188	213	1.367
Y	None	3005.472	3005.472	187	213	-0.010
Y	None	3297.581	3297.578	185	213	0.924
Y	None	3611.734	3611.737	182	213	-0.823
Y	None	4852.355	4852.356	170	213	-0.139
Y	None	5095.481	5095.478	168	213	0.671
Y	None	5182.507	5182.510	167	213	-0.504
Y	None	6870.233	6870.238	151	213	-0.750
Y	None	9675.677	9675.665	126	213	1.201
Y	None	10173.95	10173.946	122	213	0.434
Y	Water	10552.088	10552.100	118	213	-1.098
Z	H+	1509.74	1509.737	200	213	2.246
Z	H+	1566.761	1566.758	199	213	1.868
Z	H-	1692.799	1692.801	198	213	-1.182
Z	H+	1694.817	1694.817	198	213	0.206
Z	H+	1831.876	1831.876	197	213	0.239
Z	H-	1930.907	1930.908	196	213	-0.307
Z	H+	2498.146	2498.144	191	213	0.849
Z	H+	3282.567	3282.567	185	213	-0.016
Z	H+	3397.596	3397.594	184	213	0.590
Z	H+	3596.72	3596.726	182	213	-1.689

Z	H+	3683.759	3683.758	181	213	0.244
Z	H+	4010.966	4010.974	178	213	-1.972
Z	H+	4399.166	4399.170	174	213	-0.844
Z	H+	4486.197	4486.202	173	213	-1.057
Z	None	4648.261	4648.257	172	213	0.808
Z	None	4749.293	4749.305	171	213	-2.510
Z	None	4836.341	4836.337	170	213	0.837
Z	H-	5078.449	5078.451	168	213	-0.401
Z	None	6854.216	6854.219	151	213	-0.501
Z	H-	12758.313	12758.315	97	213	-0.121
Z	None	12860.337	12860.370	96	213	-2.570
Z	None	22983.256	22983.266	2	213	-0.430

Table S3. Sequence coverage values of different sequence regions of the NIST LC, Fd', and Fc/2 subunits when combining one ECD data per charge state.

	LC (%)		Fd' (%)		Fc/2 (%)	
	Without internal	With internal	Without internal	With internal	Without internal	With internal
Full sequence	74	95	58	92	55	92
Middle sequence	76	100	37	91	37	89
CDR sequence	70	100	89	94	/	/

Table S4. Sequence coverage values of different sequence regions of the NIST LC, Fd', and Fc/2 subunits when one or two additional adjusted ECD datasets per charge state of each subunit were added to the data shown in Table S2.

	LC (%)		Fd' (%)		Fc/2 (%)	
	Without internal	With internal	Without internal	With internal	Without internal	With internal
Full sequence	79	98	72	97	64	97
Middle sequence	79	100	54	95	39	93
CDR sequence	73	100	94	97	/	/

Table S5. The number of potential, localized, and identified DM1 conjugation sites of the LC, Fd', and Fc/2 subunits of the IgG1-DM1 ADC when combining one ECD data per charge state for each subunit.

	LC		Fd'		Fc/2	
	Without internal	With internal	Without internal	With internal	Without internal	With internal
Number of potential conjugation sites	11		16		18	
Number of localized conjugations sites	0	6	0	4	0	5
Number of identified conjugations sites	2	3	2	5	2	8

Table S6. The number of potential, localized, and identified DM1 conjugation sites of the LC, Fd', and Fc/2 subunits of the IgG1-DM1 ADC when one or two additional adjusted ECD datasets per charge state of each subunit were added to the data shown in Table S4.

	LC		Fd'		Fc/2	
	Without internal	With internal	Without internal	With internal	Without internal	With internal
Number of potential conjugation sites	11		16		18	
Number of localized conjugations sites	1	7	0	6	1	9
Number of identified conjugations sites	2	2	2	7	2	5

References

1. Luo, Q.; Chung, H. H.; Borths, C.; Janson, M.; Wen, J.; Joubert, M. K.; Wypych, J., Structural Characterization of a Monoclonal Antibody–Maytansinoid Immunoconjugate. *Analytical Chemistry* **2016,** *88* (1), 695-702.

2. Goldstein, G. J. S.; Tsai, H. Z.; Cosimi, B. A.; Russell, P. S.; Norman, D.; Barry, J.; Shield, C. F.; Cho, S. I.; Levey, A. S.; Burdick, J. F.; Williams, G. M.; Stuart, F. P.; Alexander, J. W.; First, R.; Helderman, J. H.; Wathen, R. L.; Lordon, R. E.; Sampson, D.; Levin, B. S., A Randomized Clinical Trial of OKT3 Monoclonal Antibody for Acute Rejection of Cadaveric Renal Transplants. *New England Journal of Medicine* **1985,** *313* (6), 337-342.

3. Schrama, D.; Reisfeld, R. A.; Becker, J. C., Antibody targeted drugs as cancer therapeutics. *Nature Reviews Drug Discovery* **2006,** *5* (2), 147-159.

4. Bebbington, C.; Yarranton, G., Antibodies for the treatment of bacterial infections: current experience and future prospects. *Current Opinion in Biotechnology* **2008,** *19* (6), 613-619.

5. Buss, N. A. P. S.; Henderson, S. J.; McFarlane, M.; Shenton, J. M.; de Haan, L., Monoclonal antibody therapeutics: history and future. *Current Opinion in Pharmacology* **2012,** *12* (5), 615-622.

6. Weiner, G. J., Building better monoclonal antibody-based therapeutics. *Nature Reviews Cancer* **2015,** *15* (6), 361-370.

7. Kisalu, N. K.; Idris, A. H.; Weidle, C.; Flores-Garcia, Y.; Flynn, B. J.; Sack, B. K.; Murphy, S.; Schön, A.; Freire, E.; Francica, J. R.; Miller, A. B.; Gregory, J.; March, S.; Liao, H.-X.; Haynes, B. F.; Wiehe, K.; Trama, A. M.; Saunders, K. O.; Gladden, M. A.; Monroe, A.;

Bonsignori, M.; Kanekiyo, M.; Wheatley, A. K.; McDermott, A. B.; Farney, S. K.; Chuang, G.-Y.; Zhang, B.; Kc, N.; Chakravarty, S.; Kwong, P. D.; Sinnis, P.; Bhatia, S. N.; Kappe, S. H. I.; Sim, B. K. L.; Hoffman, S. L.; Zavala, F.; Pancera, M.; Seder, R. A., A human monoclonal antibody prevents malaria infection by targeting a new site of vulnerability on the parasite. *Nature Medicine* **2018,** *24* (4), 408-416.

8. Walsh, G.; Walsh, E., Biopharmaceutical benchmarks 2022. *Nature Biotechnology* **2022,** *40* (12), 1722-1760.

9. Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N., Strategies and challenges for the next generation of therapeutic antibodies. *Nature Reviews Immunology* **2010,** *10* (5), 345-352.

10. Liu, H.; May, K., Disulfide bond structures of IgG molecules. *mAbs* **2012,** *4* (1), 17-23.

11. Nimmerjahn, F.; Vidarsson, G.; Cragg, M. S., Effect of posttranslational modifications and subclass on IgG activity: from immunity to immunotherapy. *Nature Immunology* **2023**.

12. Rathore, A. S.; Winkle, H., Quality by design for biopharmaceuticals. *Nature Biotechnology* **2009,** *27* (1), 26-34.

13. Lowe, D.; Dudgeon, K.; Rouet, R.; Schofield, P.; Jermutus, L.; Christ, D., Aggregation, stability, and formulation of human antibody therapeutics. In *Advances in Protein Chemistry and Structural Biology*, Donev, R., Ed. Academic Press: 2011; Vol. 84, pp 41-61.

14. Kerr, R. A.; Keire, D. A.; Ye, H., The impact of standard accelerated stability conditions on antibody higher order structure as assessed by mass spectrometry. *mAbs* **2019,** *11* (5), 930-941.

15. Kaur, H., Stability testing in monoclonal antibodies. *Critical Reviews in Biotechnology* **2021,** *41* (5), 692-714.

16. Husain, B.; Ellerman, D., Expanding the Boundaries of Biotherapeutics with Bispecific Antibodies. *BioDrugs* **2018,** *32* (5), 441-464.

17. Grilo, A. L.; Mantalaris, A., The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends in Biotechnology* **2019,** *37* (1), 9-16.

18. Khongorzul, P.; Ling, C. J.; Khan, F. U.; Ihsan, A. U.; Zhang, J., Antibody–Drug Conjugates: A Comprehensive Review. *Molecular Cancer Research* **2020,** *18* (1), 3-19.

19. Sliwkowski, M. X.; Mellman, I., Antibody Therapeutics in Cancer. *Science* **2013,** *341* (6151), 1192-1198.

20. Kaplon, H.; Crescioli, S.; Chenoweth, A.; Visweswaraiah, J.; Reichert, J. M., Antibodies to watch in 2023. *mAbs* **2023,** *15* (1), 2153410.

21. Peters, C.; Brown, S., Antibody–drug conjugates as novel anti-cancer chemotherapeutics. *Bioscience Reports* **2015,** *35* (4), e00225.

22. 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. *Journal of Cellular Physiology* **2019,** *234* (5), 5628-5642.

23. Maecker, H.; Jonnalagadda, V.; Bhakta, S.; Jammalamadaka, V.; Junutula, J. R., Exploration of the antibody–drug conjugate clinical landscape. *mAbs* **2023,** *15* (1), 2229101.

24. Chari, R. V. J.; Miller, M. L.; Widdison, W. C., Antibody–Drug Conjugates: An Emerging Concept in Cancer Therapy. *Angewandte Chemie International Edition* **2014,** *53* (15), 3796-3827.

25. Tsuchikama, K.; An, Z., Antibody-drug conjugates: recent advances in conjugation and linker chemistries. *Protein & Cell* **2018,** *9* (1), 33-46.

26. Beck, A.; Goetsch, L.; Dumontet, C.; Corvaïa, N., Strategies and challenges for the next generation of antibody–drug conjugates. *Nature Reviews Drug Discovery* **2017,** *16* (5), 315-337.

27. Ricart, A. D., Antibody-Drug Conjugates of Calicheamicin Derivative: Gemtuzumab Ozogamicin and Inotuzumab Ozogamicin. *Clinical Cancer Research* **2011,** *17* (20), 6417-6427.

28. Lambert, J. M.; Chari, R. V. J., Ado-trastuzumab Emtansine (T-DM1): An Antibody–Drug Conjugate (ADC) for HER2-Positive Breast Cancer. *Journal of Medicinal Chemistry* **2014,** *57* (16), 6949-6964.

29. Haque, M.; Forte, N.; Baker, J. R., Site-selective lysine conjugation methods and applications towards antibody–drug conjugates. *Chemical Communications* **2021,** *57* (82), 10689-10702.

30. Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S.; Lu, Y.; Meng, Y. G.; Ng, C.; Yang, J.; Lee, C. C.; Duenas, E.; Gorrell, J.; Katta, V.; Kim, A.; McDorman, K.; Flagella, K.; Venook, R.; Ross, S.; Spencer, S. D.; Lee Wong, W.; Lowman, H. B.; Vandlen, R.; Sliwkowski, M. X.; Scheller, R. H.; Polakis, P.; Mallet, W., Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nature Biotechnology* **2008,** *26* (8), 925-932.

31. 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. *Nature Biotechnology* **2012,** *30* (2), 184-189.

32. Afar, D. E. H.; Bhaskar, V.; Ibsen, E.; Breinberg, D.; Henshall, S. M.; Kench, J. G.; Drobnjak, M.; Powers, R.; Wong, M.; Evangelista, F.; O'Hara, C.; Powers, D.; DuBridge, R. B.; Caras, I.; Winter, R.; Anderson, T.; Solvason, N.; Stricker, P. D.; Cordon-Cardo, C.; Scher, H. I.; Grygiel, J. J.; Sutherland, R. L.; Murray, R.; Ramakrishnan, V.; Law, D. A., Preclinical validation of anti-TMEFF2-auristatin E–conjugated antibodies in the treatment of prostate cancer. *Molecular Cancer Therapeutics* **2004,** *3* (8), 921-932.

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. *Analytica Chimica Acta* **2017,** *955*, 67-78.

34. Wei, B.; Lantz, C.; Liu, W.; Viner, R.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A., Added Value of Internal Fragments for Top-Down Mass Spectrometry of Intact Monoclonal Antibodies and Antibody–Drug Conjugates. *Analytical Chemistry* **2023,** *95* (24), 9347–9356.

35. Bobály, B.; Fleury-Souverain, S.; Beck, A.; Veuthey, J.-L.; Guillarme, D.; Fekete, S., Current possibilities of liquid chromatography for the characterization of antibody-drug conjugates. *Journal of Pharmaceutical and Biomedical Analysis* **2018,** *147*, 493-505.

36. Wakankar, A.; Chen, Y.; Gokarn, Y.; Jacobson, F. S., Analytical methods for physicochemical characterization of antibody drug conjugates. *mAbs* **2011,** *3* (2), 161-172.

37. Zhu, X.; Huo, S.; Xue, C.; An, B.; Qu, J., Current LC-MS-based strategies for characterization and quantification of antibody-drug conjugates. *Journal of Pharmaceutical Analysis* **2020,** *10* (3), 209-220.

38. Fekete, S.; Guillarme, D.; Sandra, P.; Sandra, K., Chromatographic, Electrophoretic, and Mass Spectrometric Methods for the Analytical Characterization of Protein Biopharmaceuticals. *Analytical Chemistry* **2016,** *88* (1), 480-507.

39. Graf, T.; Heinrich, K.; Grunert, I.; Wegele, H.; Haberger, M.; Bulau, P.; Leiss, M., Recent advances in LC–MS based characterization of protein-based bio-therapeutics – mastering analytical challenges posed by the increasing format complexity. *Journal of Pharmaceutical and Biomedical Analysis* **2020,** *186*, 113251.

40. Yang, F.; Zhang, J.; Buettner, A.; Vosika, E.; Sadek, M.; Hao, Z.; Reusch, D.; Koenig, M.; Chan, W.; Bathke, A.; Pallat, H.; Lundin, V.; Kepert, J. F.; Bulau, P.; Deperalta, G.; Yu, C.; Beardsley, R.; Camilli, T.; Harris, R.; Stults, J., Mass spectrometry-based multi-attribute method in protein therapeutics product quality monitoring and quality control. *mAbs* **2023,** *15* (1).

41. Dillon, T. M.; Bondarenko, P. V.; Rehder, D. S.; Pipes, G. D.; Kleemann, G. R.; Ricci, M. S., Optimization of a reversed-phase high-performance liquid chromatography/mass

spectrometry method for characterizing recombinant antibody heterogeneity and stability. *Journal of Chromatography A* **2006,** *1120* (1), 112-120.

42. Campuzano, I. D. G.; Pelegri-O'Day, E. M.; Srinivasan, N.; Lippens, J. L.; Egea, P.; Umeda, A.; Aral, J.; Zhang, T.; Laganowsky, A.; Netirojjanakul, C., High-Throughput Mass Spectrometry for Biopharma: A Universal Modality and Target Independent Analytical Method for Accurate Biomolecule Characterization. *Journal of the American Society for Mass Spectrometry* **2022,** *33* (11), 2191-2198.

43. Bongers, J.; Cummings, J. J.; Ebert, M. B.; Federici, M. M.; Gledhill, L.; Gulati, D.; Hilliard, G. M.; Jones, B. H.; Lee, K. R.; Mozdzanowski, J.; Naimoli, M.; Burman, S., Validation of a peptide mapping method for a therapeutic monoclonal antibody: what could we possibly learn about a method we have run 100 times? *Journal of Pharmaceutical and Biomedical Analysis* **2000,** *21* (6), 1099-1128.

44. Song, Y.; Gao, J.; Meng, Q.; Tang, F.; Wang, Y.; Zeng, Y.; Huang, W.; Shao, H.; Zhou, H., Conjugation site characterization of antibody–drug conjugates using electron-transfer/higher-energy collision dissociation (EThcD). *Analytica Chimica Acta* **2023,** *1251*, 340978.

45. Ayoub, D.; Bertaccini, D.; Diemer, H.; Wagner-Rousset, E.; Colas, O.; Cianférani, S.; Van Dorsselaer, A.; Beck, A.; Schaeffer-Reiss, C., Characterization of the N-Terminal Heterogeneities of Monoclonal Antibodies Using In-Gel Charge Derivatization of α-Amines and LC-MS/MS. *Analytical Chemistry* **2015,** *87* (7), 3784-3790.

46. 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. *Journal of Chromatography B* **2015,** *981-982*, 9-13.

47. Krokhin, O. V.; Antonovici, M.; Ens, W.; Wilkins, J. A.; Standing, K. G., Deamidation of -Asn-Gly- Sequences during Sample Preparation for Proteomics: Consequences for MALDI and HPLC-MALDI Analysis. *Analytical Chemistry* **2006,** *78* (18), 6645-6650.

48. Nielsen, M. L.; Vermeulen, M.; Bonaldi, T.; Cox, J.; Moroder, L.; Mann, M., Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nature Methods* **2008,** *5* (6), 459-460.

49. 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. *Analytical Chemistry* **2013,** *85* (9), 4239-4246.

50. 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.

51. 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. *Analytical Chemistry* **2020,** *92* (15), 10246-10251.

52. 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. *Analytical Chemistry* **2018,** *90* (14), 8421-8429.

53. 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* (1), 106-115.

54. 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* (1), 766-773.

55. 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. *Analytical Chemistry* **2014,** *86* (6), 3005-3012.

56. 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. *Analytical Chemistry* **2018,** *90* (21), 12527-12535.

57. Srzentić, K.; Fornelli, L.; Tsybin, Y. O.; Loo, J. A.; Seckler, H.; Agar, J. N.; Anderson, L. C.; Bai, D. L.; Beck, A.; Brodbelt, J. S.; Van Der Burgt, Y. E. M.; Chamot-Rooke, J.; Chatterjee, S.; Chen, Y.; Clarke, D. J.; Danis, P. O.; Diedrich, J. K.; D'Ippolito, R. A.; Dupré,

M.; Gasilova, N.; Ge, Y.; Goo, Y. A.; Goodlett, D. R.; Greer, S.; Haselmann, K. F.; He, L.; Hendrickson, C. L.; Hinkle, J. D.; Holt, M. V.; Hughes, S.; Hunt, D. F.; Kelleher, N. L.; Kozhinov, A. N.; Lin, Z.; Malosse, C.; Marshall, A. G.; Menin, L.; Millikin, R. J.; Nagornov, K. O.; Nicolardi, S.; Paša-Tolić, L.; Pengelley, S.; Quebbemann, N. R.; Resemann, A.; Sandoval, W.; Sarin, R.; Schmitt, N. D.; Shabanowitz, J.; Shaw, J. B.; Shortreed, M. R.; Smith, L. M.; Sobott, F.; Suckau, D.; Toby, T.; Weisbrod, C. R.; Wildburger, N. C.; Yates, J. R.; Yoon, S. H.; Young, N. L.; Zhou, M., Interlaboratory Study for Characterizing Monoclonal Antibodies by Top-Down and Middle-Down Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (9), 1783-1802.

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. *Analytical Chemistry* **2019,** *91* (18), 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. *Journal of the American Society for Mass Spectrometry* **2019,** *30* (11), 2419-2429.

60. 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. *Analytical Chemistry* **2020,** *92* (22), 15096-15103.

61. 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. *Analytical Chemistry* **2020,** *92* (14), 9790-9798.

62. Ren, D.; Pipes, G. D.; Liu, D.; Shih, L.-Y.; Nichols, A. C.; Treuheit, M. J.; Brems, D. N.; Bondarenko, P. V., An improved trypsin digestion method minimizes digestion-induced modifications on proteins. *Analytical Biochemistry* **2009,** *392* (1), 12-21.

63. Formolo, T.; Heckert, A.; Phinney, K. W., Analysis of deamidation artifacts induced by microwave-assisted tryptic digestion of a monoclonal antibody. *Analytical and Bioanalytical Chemistry* **2014,** *406* (26), 6587-6598.

64. Hains, P. G.; Robinson, P. J., The Impact of Commonly Used Alkylating Agents on Artifactual Peptide Modification. *Journal of Proteome Research* **2017,** *16* (9), 3443-3447.

65. Wei, B.; Zenaidee, M. A.; Lantz, C.; Ogorzalek Loo, R. R.; Loo, J. A., Towards understanding the formation of internal fragments generated by collisionally activated dissociation for top-down mass spectrometry. *Anal. Chim. Acta* **2022,** *1194*, 339400.

66. Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L., Analyzing internal fragmentation of electrosprayed ubiquitin ions during beam-type collisional dissociation. *J. Am. Soc. Mass Spectrom.* **2015,** *26* (5), 782-787.

67. Zenaidee, M. A.; Lantz, C.; Perkins, T.; Jung, W.; Loo, R. R. O.; Loo, J. A., Internal Fragments Generated by Electron Ionization Dissociation Enhance Protein Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020,** *31* (9), 1896–1902.

68. Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A., ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* **2021,** *20* (4), 1928–1935.

69. Schmitt, N. D.; Berger, J. M.; Conway, J. B.; Agar, J. N., Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem.* **2021,** *93* (16), 6355-6362.

70. Zenaidee, M. A.; Wei, B.; Lantz, C.; Wu, H. T.; Lambeth, T. R.; Diedrich, J. K.; Ogorzalek Loo, R. R.; Julian, R. R.; Loo, J. A., Internal Fragments Generated from Different Top-Down Mass Spectrometry Fragmentation Methods Extend Protein Sequence Coverage. *J. Am. Soc. Mass Spectrom.* **2021,** *32* (7), 1752–1758.

71. Wei, B.; Zenaidee, M. A.; Lantz, C.; Williams, B. J.; Totten, S.; Ogorzalek Loo, R. R.; Loo, J. A., Top-down mass spectrometry and assigning internal fragments for determining disulfide bond positions in proteins. *The Analyst* **2023,** *148* (1), 26-37.

72. Dunham, S. D.; Wei, B.; Lantz, C.; Loo, J. A.; Brodbelt, J. S., Impact of Internal Fragments on Top-Down Analysis of Intact Proteins by 193 nm UVPD. *Journal of Proteome Research* **2023,** *22* (1), 170-181.

73. Li, H.; Sheng, Y.; Mcgee, W.; Cammarata, M.; Holden, D.; Loo, J. A., Structural Characterization of Native Proteins and Protein Complexes by Electron Ionization Dissociation-Mass Spectrometry. *Anal. Chem.* **2017,** *89* (5), 2731-2738.

74. Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A., An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nat. Chem.* **2018,** *10* (2), 139-148.

75. Rolfs, Z.; Smith, L. M., Internal Fragment Ions Disambiguate and Increase Identifications in Top-Down Proteomics. *J. Proteome Res.* **2021,** *20* (12), 5412-5418.

76. Clauser, K. R.; Baker, P.; Burlingame, A. L., Role of Accurate Mass Measurement (±10 ppm) in Protein Identification Strategies Employing MS or MS/MS and Database Searching. *Analytical Chemistry* **1999,** *71* (14), 2871-2882.

77. Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V., Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Analytical Chemistry* **2015,** *87* (8), 4370-4376.

78. 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* (5), 291-305.

Chapter 6: Native Top-Down Mass Spectrometry with Collisionally Activated Dissociation Yields Higher-Order Structure Information for Protein Complexes

Reprinted with permission from

Lantz, C.; Wei, B.; Zhao, B.; Jung, W.; Goring, A. K.; Le, J.; Miller, J.; Loo, R. R. O.; Loo, J. A., Native Top-Down Mass Spectrometry with Collisionally Activated Dissociation Yields Higher-Order Structure Information for Protein Complexes. *Journal of the American Chemical Society* **2022,** *144* (48), 21826-21830. DOI: https://doi.org/10.1021/jacs.2c06726.

Carter Lantz[a], Benqian Wei[a], Boyu Zhao[a], Wonhyeuk Jung[a], Andrew K. Goring[a], Jessie Le[a], Justin Miller[b], Rachel R. Ogorzalek Loo[a,b], Joseph A. Loo[a, b, c *]

[a] Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA, USA; [b] Molecular Biology Institute, University of California−Los Angeles, Los Angeles, CA, United States; [c] Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA.

* Corresponding Author: Joseph A. Loo, Department of Chemistry and Biochemistry, Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA, E-mail: jloo@chem.ucla.edu.

Abstract

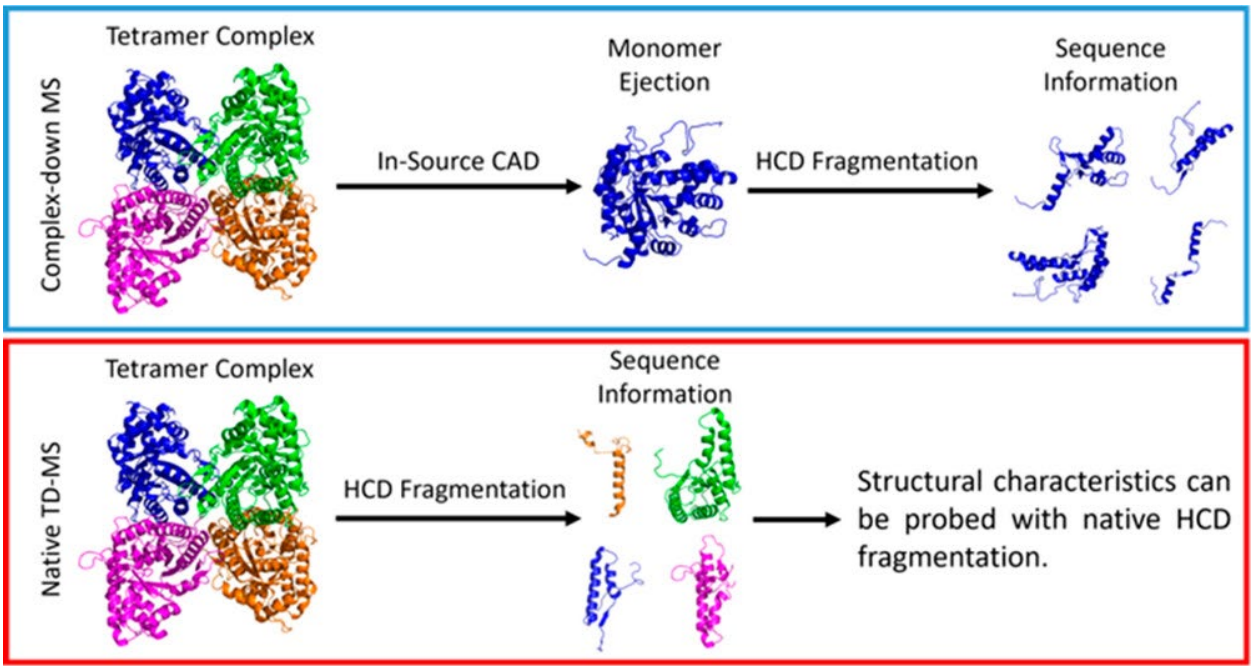
Native mass spectrometry (MS) of proteins and protein assemblies reveals size and binding stoichiometry, but elucidating structures to understand their function is more challenging. Native top-down MS (nTDMS), i.e., fragmentation of the gas-phase protein, is conventionally used to derive sequence information, locate post-translational modifications (PTMs), and pinpoint ligand binding sites. nTDMS also endeavors to dissociate covalent bonds in a conformation-sensitive manner, such that information about higher-order structure can be inferred from the fragmentation pattern. However, the activation/dissociation method used can greatly affect the resulting information on protein higher-order structure. Methods such as electron capture/ transfer dissociation (ECD and ETD, or ExD) and ultraviolet photodissociation (UVPD) can produce product ions that are sensitive to structural features of protein complexes. For multi-subunit complexes, a long-held belief is that collisionally activated dissociation (CAD) induces unfolding and release of a subunit, and thus is not useful for higher-order structure characterization. Here we show not only that sequence information can be obtained directly from CAD of native protein complexes but that the fragmentation pattern can deliver higher-order structural information about their gas- and solution-phase structures. Moreover, CAD-generated internal fragments (i.e., fragments containing neither N-/C-termini) reveal structural aspects of protein complexes.

Main Text

Native top-down mass spectrometry (nTDMS) of gas-phase proteins yields product ions that can provide information on amino acid sequence,[1,2] sites of modifications,[3–5] and even higher-order structure.[6] Performing nTDMS with electron-based techniques such as electron capture dissociation (ECD) and electron transfer dissociation (ETD)[7–11] and photon-based techniques such as infrared multiphoton dissociation (IRMPD) and ultraviolet photodissociation (UVPD)[8,12–14] is generally favored, as it fragments the complex directly without disrupting the overall complex structure. In contrast, it has been generally assumed that collision-based fragmentation does not reveal higher-order structural information, as unfolding and ejection of monomer subunits (and ligands) occurs. However, we have found that direct fragmentation of native protein complexes with Orbitrap-based high-energy C-trap dissociation (HCD),[15] a collision-based fragmentation technique performed with higher energy on a faster time scale than conventional collisionally activated dissociation (CAD), can uncover aspects of protein higher-order structure. For a variety of protein complexes, we show here that HCD can generate *b*-/*y*-type product ions that provide information on solvent-exposed regions and subunit interfaces.

To investigate HCD fragmentation of protein complexes,[16] complex-down MS (pseudo-MS^3)[17,18] and nTDMS (Scheme 1) of yeast alcohol dehydrogenase (ADH) homotetramer (147 kDa) were compared. Complex-down MS was performed by using in-source CAD to detach a monomer from the tetramer and to subsequently activate the 12+ charged monomer with HCD. The resultant MS/MS spectrum revealed both N-terminal *b*-fragments and C-terminal *y*-

fragments of ADH (Figure S1A); 24 *b*-fragments and 18 *y*-fragments resulted in 11.8% total sequence coverage (Figure 1A). The fragmentation pattern also revealed the presence of N-terminal acetylation, a V58T proteoform, and Zn^{2+} binding. The presence of near equal numbers of abundant *b*- and *y*-fragments from the complex-down MS workflow suggests that both termini of the ADH monomer subunit are easily accessed by HCD fragmentation, i.e., the in-source CAD process releases a low-structured monomer such that subsequent HCD products yield little information about the 3D structure of the native tetramer.



Scheme 1. Complex-Down MS and nTDMS Workflows Used in This Study.

For comparison, nTDMS results from HCD of the 25+ charged ADH tetramer were examined. Primarily *b*-products and surprisingly few peaks corresponding to released ADH monomers (Figure 1B, Figure S1B) were detected. We speculate that monomers were not ejected from the tetramer complex prior to covalent bond cleavage, i.e., the tetramer fragmented directly. To further support this claim, broadband fragmentation (of all ADH tetramer charge

states) with a range of HCD energies did not yield significant levels of released monomer signals (Figure S2). nTDMS of ADH yielded 60 N-terminal *b*-fragments, but only three C-terminal *y*-fragments (17.6% sequence coverage) (Figure 1B). Numerous abundant N-terminal fragments produced by HCD resemble nTDMS products from electron-based[7,8] and photodissociation techniques.[8,12] Mapping the fragments onto the crystal structure of ADH shows that the N-terminal region is more solvent exposed than the C-terminal region, with the latter forming subunit−subunit interfaces of the complex (Figure S3). Our analysis indicates that fragments that cut at the interface of the tetramer (residues 240−310) accounted for only 8% of the fragment ion current.

To further examine how collision-based fragmentation can reveal structural information from protein complexes, intact (rabbit) aldolase homotetramer (157 kDa) was fragmented with HCD. Much like ADH, aldolase did not release monomers upon HCD, but rather y-fragments including an especially abundant y_{74} ion (2+ to 5+ charged) (Figure S4). At low HCD energies, a large complementary fragment corresponding to the mass of the intact tetramer losing a y_{74} fragment, i.e., (4M − y_{74}), was observed (Figure S5 and Table S1), indicating direct fragmentation of the tetramer. nTDMS yielded 35 C-terminal *y*-fragments but only eight N-terminal *b*-fragments (11.0% sequence coverage) (Figure 2). This result differs from the complex-down mass spectrum of aldolase, which shows a nearly equal proportion of N-terminal *b*-fragments (19) and C-terminal *y*-fragments (16) (Figure S6).

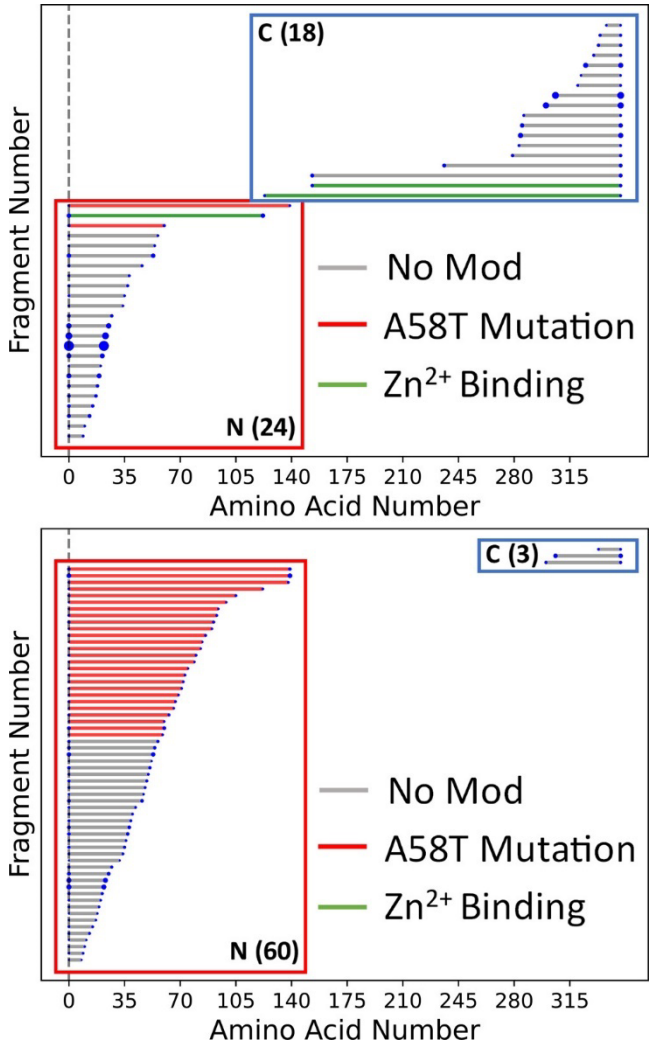


Figure 1. Fragment location maps for ADH representing *b*-/*y*-product ions measured by (top) complex-down MS and (bottom) nTDMS with HCD.
Red lines indicate V58T mutation, green lines indicate Zn2+ binding, the vertical dotted line indicates N-terminal acetylation, and the size of the blue dots indicates the relative intensity of each fragment. Numbers in parentheses indicate the number of product ions detected.

The HCD fragments from the aldolase tetramer mainly cover the solvent-exposed C-terminus and are absent from the interface forming N-terminus (Figure 2). Our analysis indicates that fragments that cut at the interface of the tetramer (residues 110−224) accounted for only 1% of the fragment ion current. The relatively high proportion of C-terminal fragments present in the native HCD spectrum of aldolase is similar to that measured by ECD previously[19]

and further suggests that direct HCD fragmentation of some protein complexes can reveal regions of solvent accessibility.

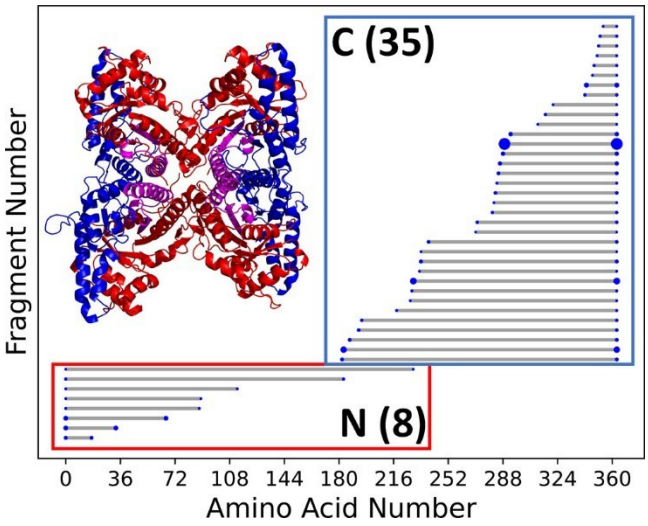


Figure 2. Fragment location map for nTDMS products of the 25+ charged precursor of aldolase homotetramer and its crystal structure.
The size of the blue dots corresponds to the relative intensity of the fragments. The crystal structure shows that most cleavage sites lie on the solvent-exposed C-terminus (blue), rather than the interface forming N-terminus (red). The purple region is covered by both N-terminal and C-terminal fragments.

NTDMS with HCD was performed on several other protein complexes. Complex-down fragmentation of the glutathione S-transferase A1 (GSTA1) dimer revealed 25 N-terminal *b*-fragments and 20 C-terminal *y*-fragments (Figure S7A). In contrast, the native fragmentation spectrum of GSTA1 reveals five N-terminal *b*-fragments and 19 C-terminal *y*-fragments (Figure S7B), consistent with the GSTA1 crystal structure showing that the C-terminus is more solvent exposed than the N-terminus (Figure S7B). For the yeast enolase dimer, 27 *b*-fragments along with 18 *y*-fragments were measured by complex-down MS (Figure S8A). nTDMS revealed 48 N-terminal *b*-fragments along with 51 C-terminal *y*-fragments without the appearance of abundant

monomer ions (Figure S8B). The crystal structure of enolase (Figure S8B) indicates that both N-/C-termini are solvent exposed and are not involved in forming the dimer interface, consistent with the near equal proportion of *b*-/*y*-products measured by nTDMS.

Some complexes did not release monomers from in-source CAD for complex-down fragmentation; however, HCD of the native complexes still returned structural information. Native HCD of the creatine kinase dimer revealed nine *b*- and 38 *y*-fragments, which suggests that the C-terminus is solvent exposed and the N-terminus forms the interface of the dimer; this aligns well with the crystal structure of creatine kinase (Figure S9). Similarly, HCD of 6-phosphogluconate dehydrogenase (GND1) dimer generated 23 *b*-fragments but only six *y*-fragments, consistent with the GND1 crystal structure showing the N-termini to be solvent exposed and the C-termini forming the dimer interface (Figure S10).

There are some exceptions to this pattern of *b*-/*y*-product formation directly from intact native complexes under HCD. For example, HCD of the native membrane protein, aquaporin Z (AqpZ) homotetramer,[20–22] yielded abundant monomer, dimer, and trimer products released from the intact complex (Figure S11A). This observation can be attributed to the weak hydrophobic binding interface between the monomer subunits of the AqpZ tetramer. Complexes such as aldolase and ADH are stabilized somewhat by salt bridges that strengthen greatly in the gas phase,[23,24] potentially preventing monomer ejection during HCD (Figure S12). That monomer products are released when HCD is applied to native AqpZ complexes suggests that structural information (such as the locations of solvent-exposed regions and the tetramer interface) cannot be inferred from the resulting *b*-/*y*-fragments, at least assuming that the monomers likely eject

before covalent bonds cleave. This suggestion is supported by the fact that the nTDMS fragmentation pattern of AqpZ tetramers (65 *b*-fragments, 62 *y*-fragments, 38.4% sequence coverage; Figure S11A) does not differ significantly from the complex-down fragmentation pattern of isolated monomers (63 *b*-fragments, 60 C-terminal *y*-fragments, 34.6% sequence coverage; Figure S11B). Although HCD fragmentation of native AqpZ does not reveal significant higher-order structural information, it does suggest that the interaction between complex monomers in the gas phase is relatively weak.

Monomer releases during HCD are not limited to membrane protein complexes. HCD fragmentation of the hemoglobin (Hb) tetramer revealed monomer and trimer peaks in addition to 10 *b*-fragments and eight *y*-fragments from the α-subunit and seven *b*-fragments and seven y-fragments from the β-subunit (Figure S13A). Fragmentation of the Hb dimer also revealed released monomer peaks in addition to 11 *b*-fragments and 16 *y*-fragments from the α-subunit and 10 *b*-fragments and four *y*-fragments from the β-subunit (Figure S13B). A similar HCD fragmentation pattern can be observed from complex-down MS of individual subunits (nine *b*- and 10 *y*-fragments and six *b*- and five *y*-fragments from the α- and β-subunits, respectively) (Figure S13C). Similarly, nTDMS of human transthyretin (TTR) tetramers by HCD releases monomer products in addition to two *b*- and 38 *y*-fragments (Figure S14A). The relative proportion of *b*-/*y*-product ions between the tetramer and monomer TTR is similar, with complex-down of the TTR monomer yielding three *b*-fragments and 41 *y*-fragments (Figure S14B). The HCD results for all of the complexes included in the study are listed in Table S2.

Lastly, we investigated the utility of internal fragments (i.e., product ions containing neither

N-/C-termini that result from at least two bond cleavage events)[1,25–30] for structure determination of protein complexes. Preliminary data show that HCD fragmentation of ADH tetramers reveals numerous internal fragments spanning residues 178−236 (Figure S15A), which correspond to a solvent-exposed region (Figure S15B). More work will extend this concept further, but it demonstrates that HCD-derived internal fragments can deliver structural information on protein assemblies.

Although other studies have noted the detection of *b*-/*y*-products with concurrent subunit release from CAD[31] and HCD[16] of protein complexes, we have found that collision-based fragmentation with HCD can reveal higher-order structure information for several multi-subunit protein complexes that appear to be stabilized through the presence of salt bridges.[23] These complexes fragment directly by HCD without significant monomer release. The resulting products map to solvent-exposed areas, while regions delivering fewer fragments likely comprise subunit interfaces. Other weak gas-phase complexes eject monomers upon HCD. Nonetheless, it is currently unclear what differences between HCD and other beam-type CAD experiments are responsible for the unique fragmentation behavior.

An assumption carried over from small-molecule dissociation studies to macroion decompositions is that, on the experimental time scale, activation from collisions always randomizes fully to steer collision-induced decompositions along the lowest energy pathways. However, those assumptions fail to consider that entropically demanding, slow rearrangements might be essential to releasing a subunit, e.g., to reposition salt bridges tethering one subunit to others.[23] In cases where the number of collisions and/or energy per collision are insufficient to

stumble on the rare configuration ejecting a subunit within the experimental time frame, alternative rearrangements to eject smaller polypeptide fragments (with fewer tethers) may be competitive. Nevertheless, we show that HCD can be a powerful biophysical tool to probe the structure of proteins without the need for other electron- and photon-based activation/dissociation methods.

Associate Content

*sı Supporting Information

The Supporting Information is available free of charge at:

https://doi.org/10.1021/jacs.2c06726.

Materials and methods; supplementary figures (Figures S1−S15) showing nTDMS and complex-down MS spectra, data, and protein structures of those studied; Table S1, listing information for all proteins studied (PDF).

Author Information

Corresponding Author

Joseph A. Loo − Department of Chemistry and Biochemistry, University of California─Los Angeles, Los Angeles, California 90095, United States; Department of Biological Chemistry, UCLA-DOE Institute, and Molecular Biology Institute, University of California─Los Angeles, Los Angeles, California 90095, United States; orcid.org/0000-0001-9989-1437; Email: jloo@chem.ucla.edu

Authors

Carter Lantz − Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States.

Benqian Wei − Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States; orcid.org/0000-00034853-4848.

Boyu Zhao − Department of Chemistry and Biochemistry, University of California−Los Angeles, Los Angeles, California 90095, United States.

Wonhyeuk Jung − Department of Chemistry and Biochemistry, University of California−Los Angeles, Los Angeles, California 90095, United States.

Andrew K. Goring − Department of Chemistry and Biochemistry, University of California−Los Angeles, Los Angeles, California 90095, United States.

Jessie Le − Department of Chemistry and Biochemistry, University of California−Los Angeles, Los Angeles, California 90095, United States.

Justin Miller − Molecular Biology Institute, University of California−Los Angeles, Los Angeles, California 90095, United States.

Rachel R. Ogorzalek Loo − Department of Chemistry and Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States; UCLA-DOE Institute and Molecular Biology Institute, University of California—Los Angeles, Los Angeles, California 90095, United States; orcid.org/0000-0002-0580-2833.

Complete contact information is available at:

https://doi.org/10.1021/jacs.2c06726.

Notes:

The authors declare no competing financial interest.

Acknowledgements

J.A.L. and R.R.O.L. acknowledge support from the US National Institutes of Health (R01GM103479, R35GM145286), the US National Science Foundation (NSF) (CHE1808492), and the US Department of Energy (DE-FC02-02ER63421). C.L. acknowledges support from the Ruth L. Kirschstein National Research Service Award program (GM007185). A.G. acknowledges support from the National Institute of Dental & Craniofacial Research (T90DE030860).

Chapter 6: Supporting Information

Materials and Methods

Proteins samples were obtained from Sigma Aldrich (St. Louis, MO, USA), dissolved in 200mM ammonium acetate, and desalted with 10K Amicon filters from Sigma Aldrich. The samples were then diluted to 10μM and sprayed on a Thermo UHMR (Thermo Fisher Scientific, San Jose, CA) with voltages of 1-2kV. To fragment native complexes, HCD energies of 125-280V were applied. Lower voltages were applied for select applications. Complex-down MS experiments were performed by applying 5-150V of in-source CAD or -60V of desolvation voltage to eject monomers and then applying 100V-177V of HCD energy to subsequently activate those monomers. For internal fragment analysis, the ADH the tetramer was fragmented with 215V of collision energy with argon as the collision gas.

Deconvolution was performed with BioPharma Finder 3.2 and the resulting deconvoluted peak list was run through ClipsMS.[1] *b*- and *y*-fragments were matched to protein sequences with an error tolerance of 5ppm and unlocalized modifications included the addition of a hydrogen atom and the abstraction of a water molecule were added to the theoretical masses. For ADH, additional modifications including an N-terminal acetylation, a V58T mutation, and a Zn^{2+} ion were added to theoretical fragments. *by* internal fragments of ADH were searched with ClipsMS with an error tolerance of 5ppm. To deconvolute large complementary fragments, UniDec was used.[2] Fragments were mapped onto crystal structures of protein complexes using Pymol 2.5.4. The ADH pymol code used was 4W6Z, the aldolase pymol code used was 1ADO, the enolase pymol code used was 1EGB, the GSTA1 pymol code used was 1GSD, and the

creatine kinase pymol code used was 1U6R, and the aquaporin Z pymol code used was 1RC2.

References

1. Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A., ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *Journal of Proteome Research* **2021,** *20* (4), 1928-1935.

2. Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K.; Benesch, J. L.; Robinson, C. V., Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Analytical Chemistry* **2015,** *87* (8), 4370-4376.

Supplementary Figures

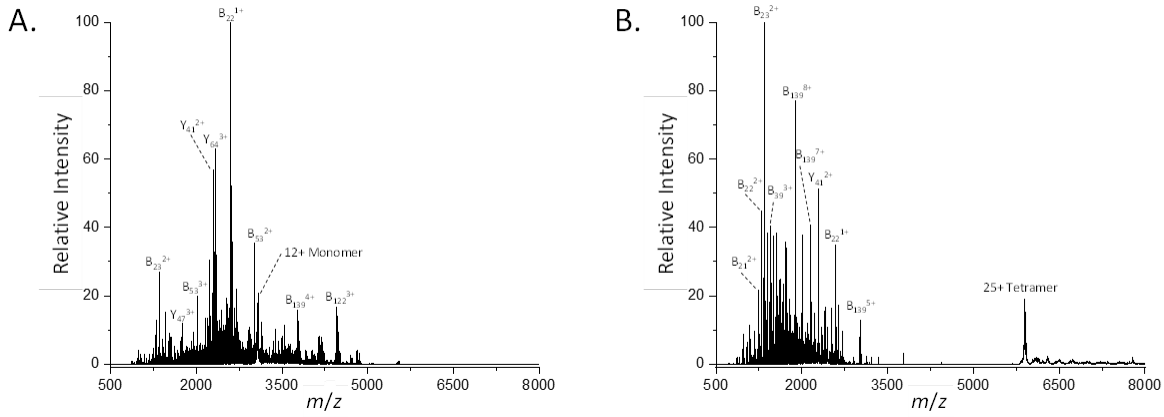


Figure S1. (A) Complex-down fragmentation mass spectrum and (B) native TDMS spectrum of ADH.

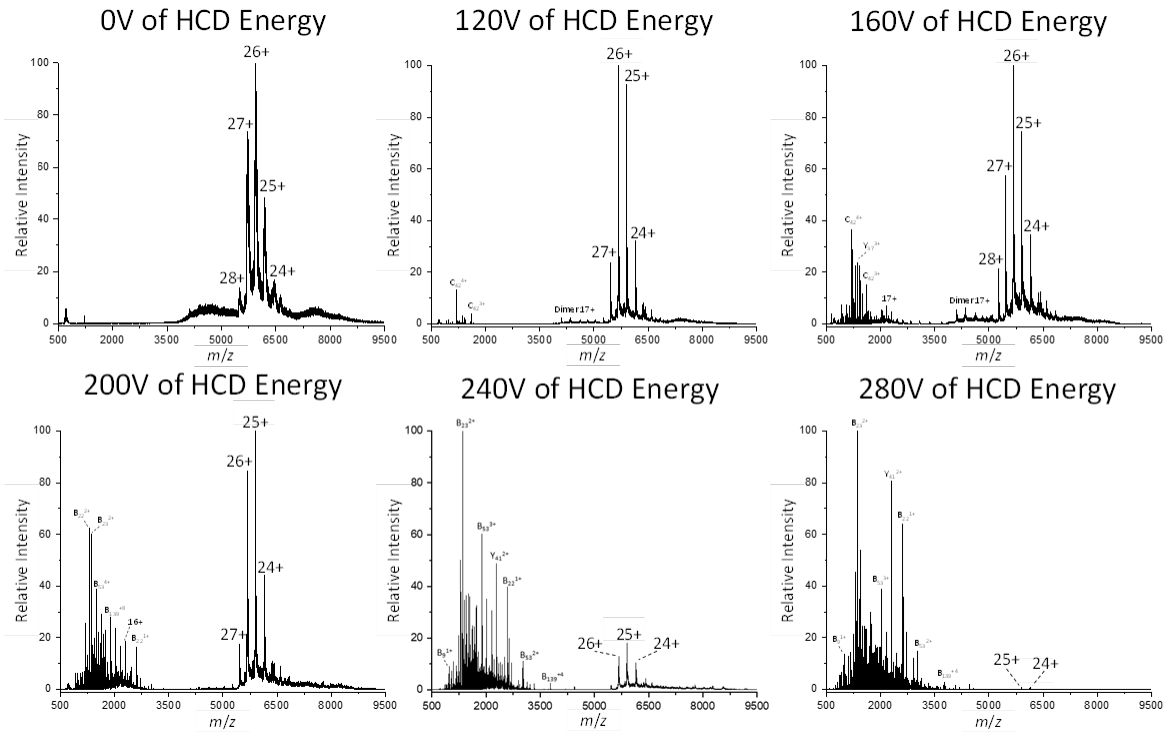


Figure S2. Broadband nTDMS spectra of ADH at various HCD voltages.

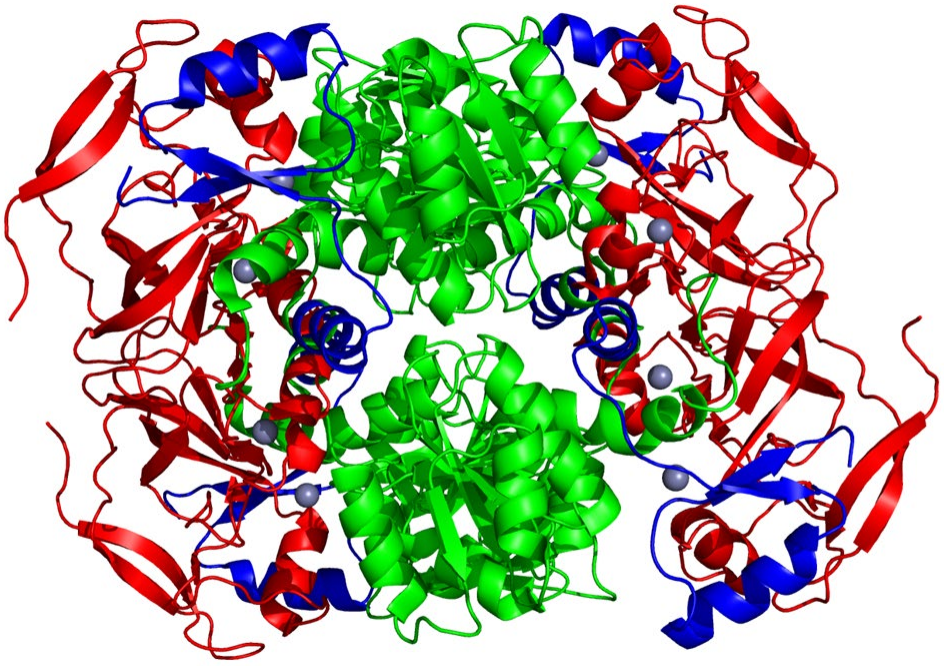


Figure S3: The structure of ADH with the region covered by the N-terminal fragments labeled in red and the region covered by the C-terminal fragments labeled in blue.
Fragmentation occurs in the solvent exposed regions and does not occur in the subunit-subunit interface region (green).

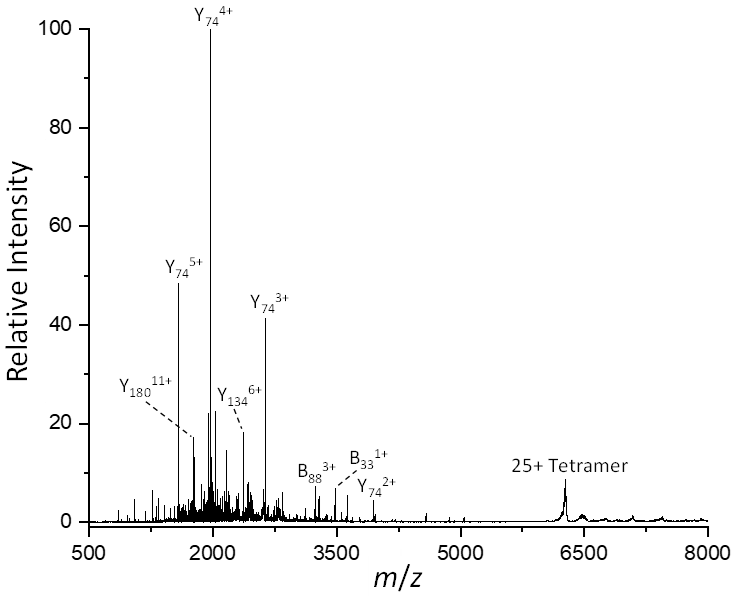


Figure S4: Native top-down mass spectrum of the 25+ charged aldolase homotetramer.

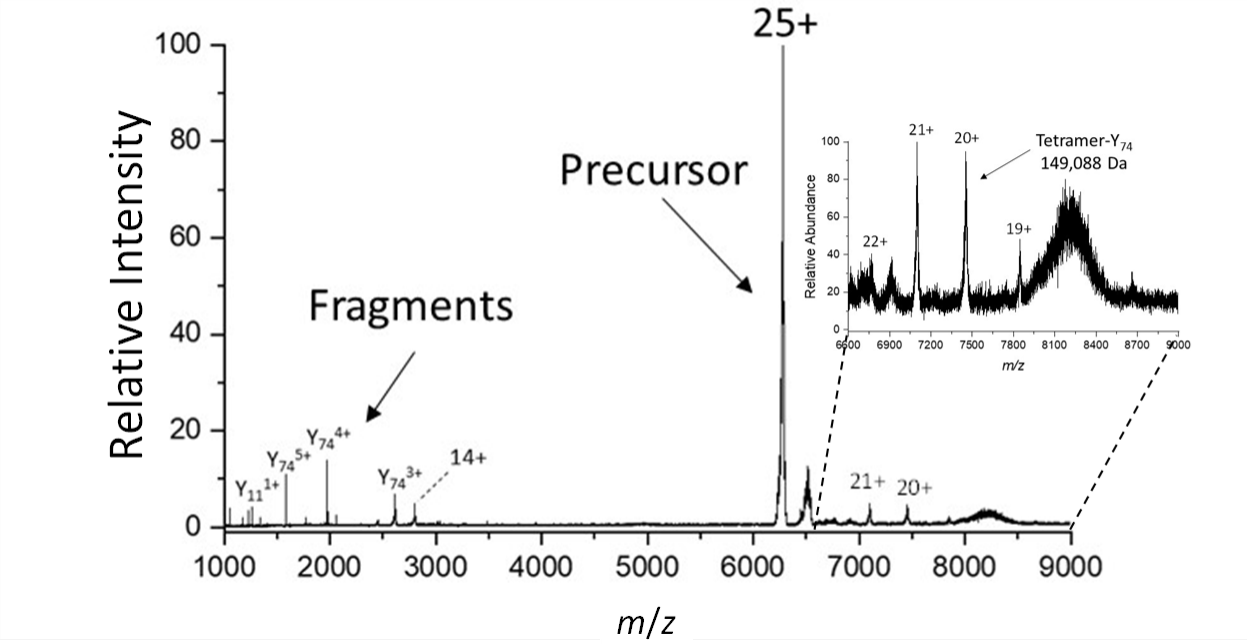


Figure S5: Native top-down mass spectrum of the 25+ charge state of the aldolase homotetramer showing multiple charge states of an abundant y_{74} fragment and high *m/z* peaks corresponding to charge states of the (4M - y_{74}) product ion.

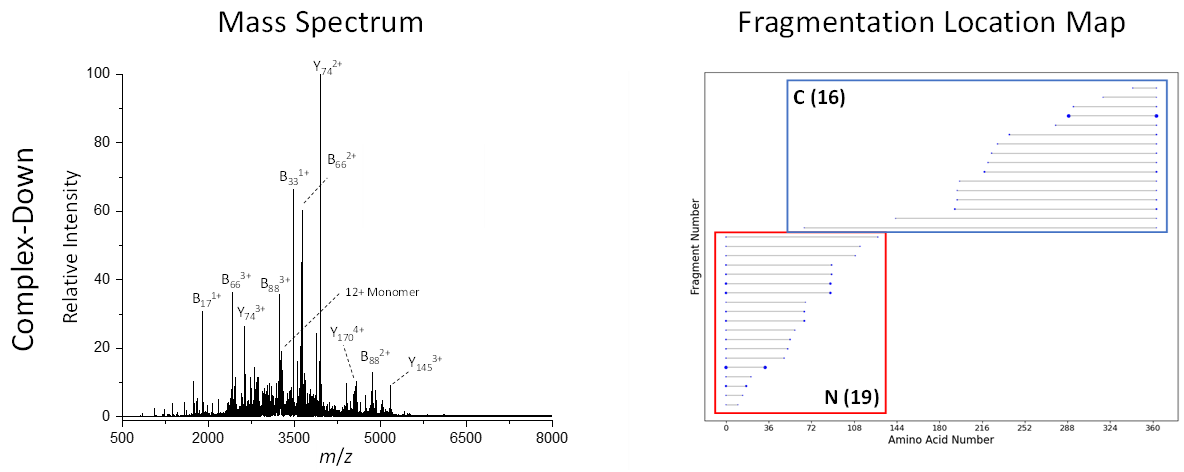


Figure S6: A complex-down mass spectrum (12+ monomer) and the corresponding fragment location map for aldolase.

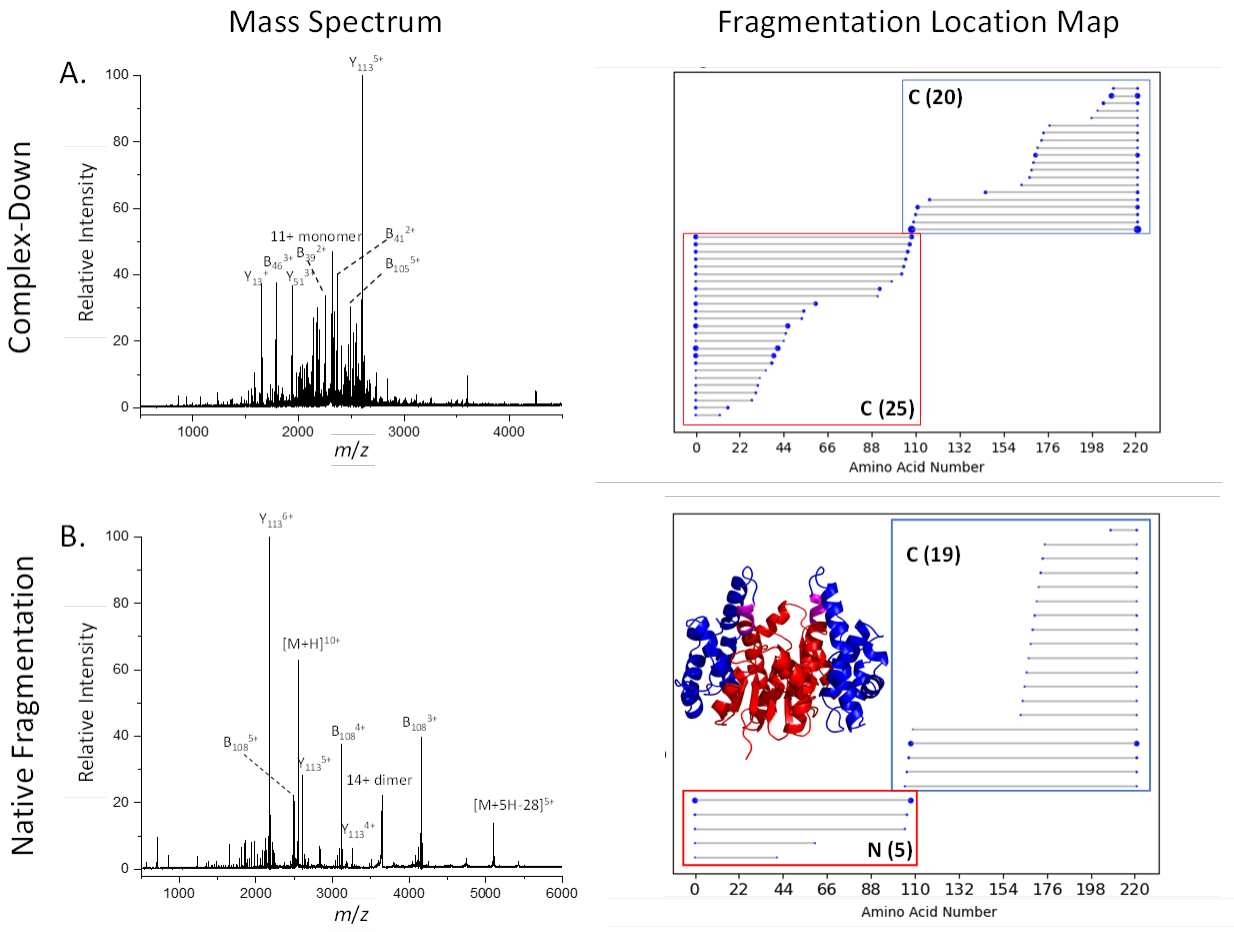


Figure S7: Complex-down vs. native top-down of the human GST A1 dimer.
(A) A complex-down mass spectrum (11+ monomer) with the corresponding fragment location map and (B) a native top-down mass spectrum with the corresponding fragmentation location map for the 12+ charged human GST A1 dimer. The inset shows the structure of GST A1 with the region covered by N-terminal fragments labeled in red, the region covered by C-terminal fragments labeled in blue, and the region covered by N- and C-terminal fragments labeled in purple.

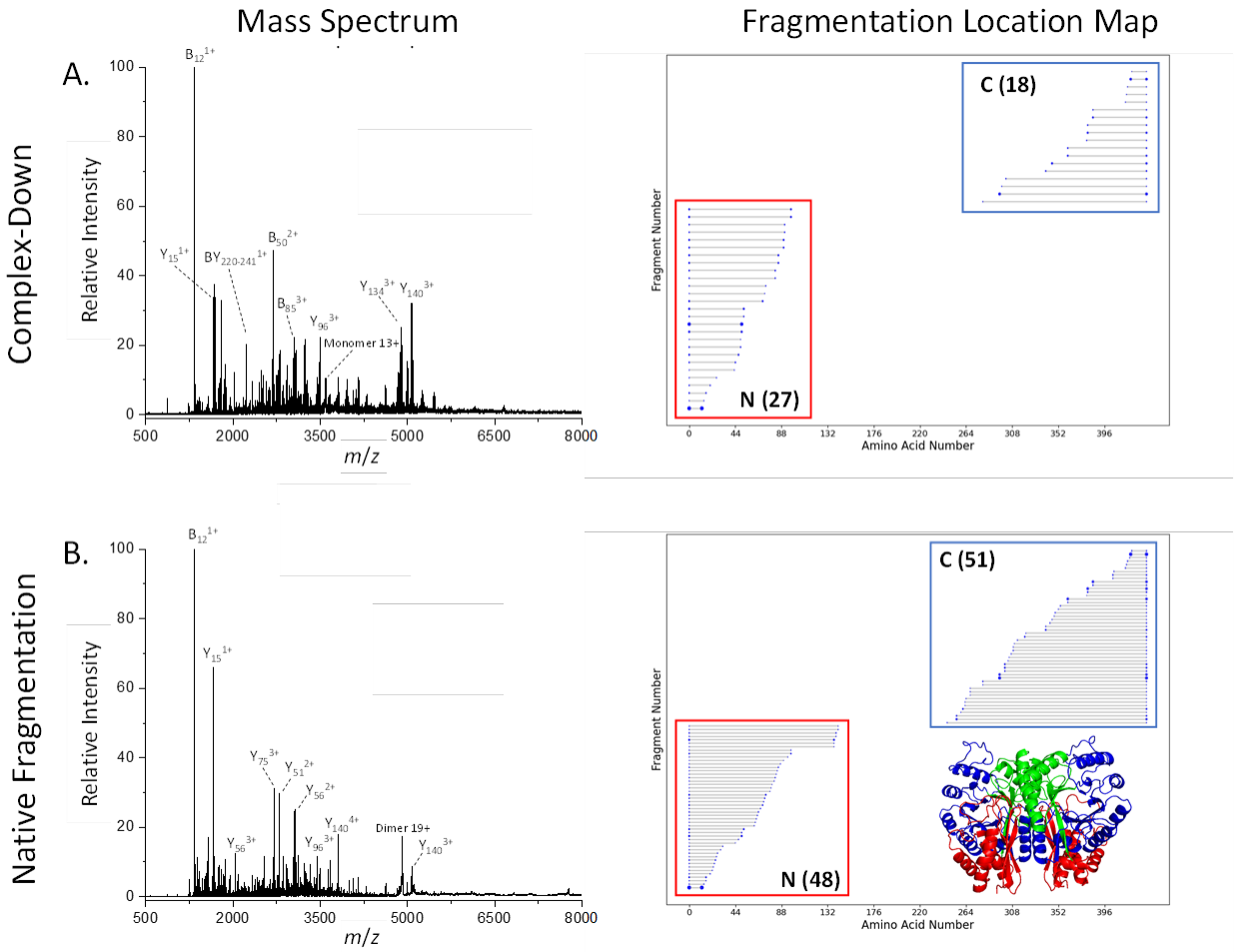


Figure S8: Complex-down vs. native top-down of the enolase dimer.
(A) A complex-down mass spectrum (13+ monomer) with the corresponding fragment location map and (B) a native top-down mass spectrum with the corresponding fragmentation location map for the 19+ charged enolase dimer. The inset shows the structure of enolase with the region covered by N-terminal fragments colored in red and the region covered by C-terminal fragments colored in blue (and green indicates no fragment coverage).

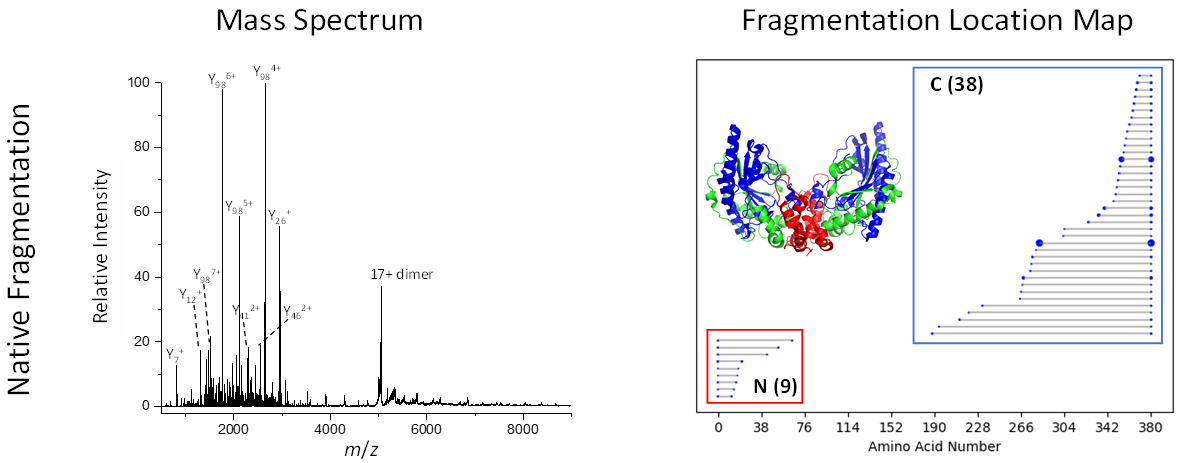


Figure S9: A native top-down mass spectrum with the corresponding fragmentation location map for 17+ charged creatine kinase dimer.

The inset shows the structure of creatine kinase with the region covered by N-terminal fragments colored in red and the region covered by C-terminal fragments colored in blue (and green indicates no fragment coverage).

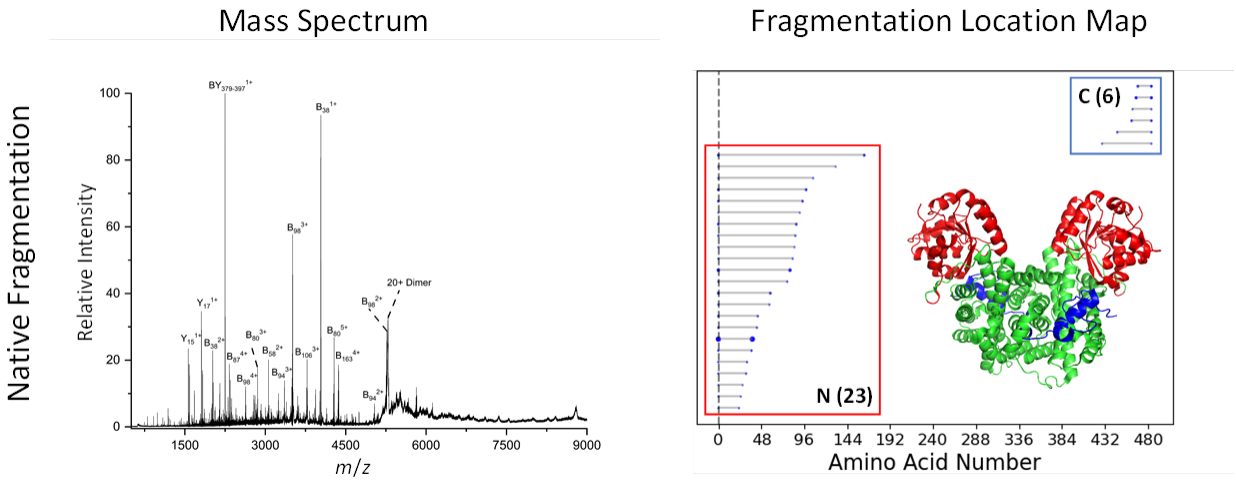


Figure S10: A native top-down mass spectrum with the corresponding fragmentation location map for the 20+ charged GND1 dimer with the vertical dotted line representing N-terminal acetylation.

The inset shows the structure of GND1 with the region covered by N-terminal fragments colored in red and the region covered by C-terminal fragments colored in blue (and green indicates no fragment coverage).

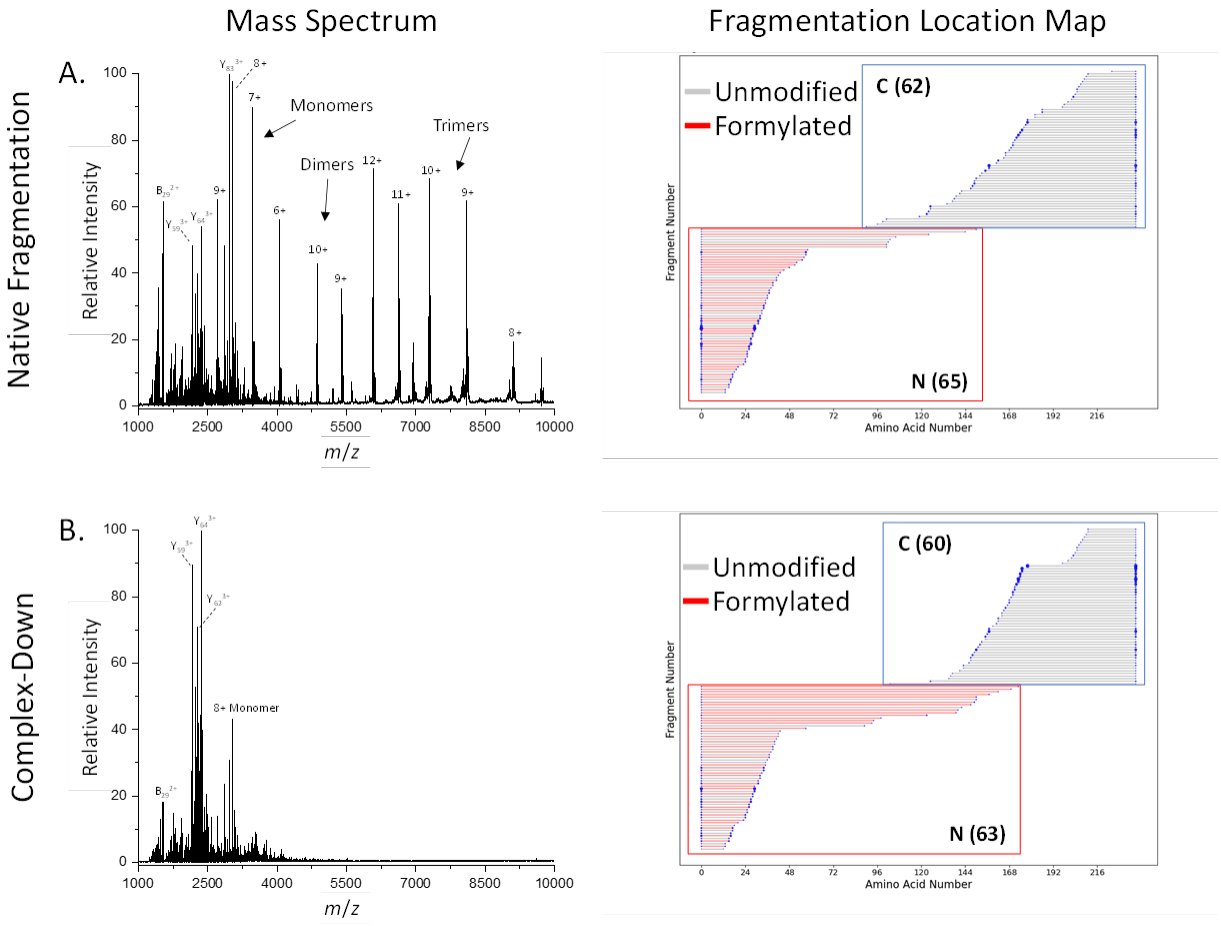


Figure S11: Complex-down vs. native top-down of the AqpZ tetramer.
(A) A native top-down mass spectrum (18+ charged tetramer) with the corresponding fragment location map and (B) a complex-down mass spectrum (8+ monomer) with the corresponding fragmentation location map for AqpZ.

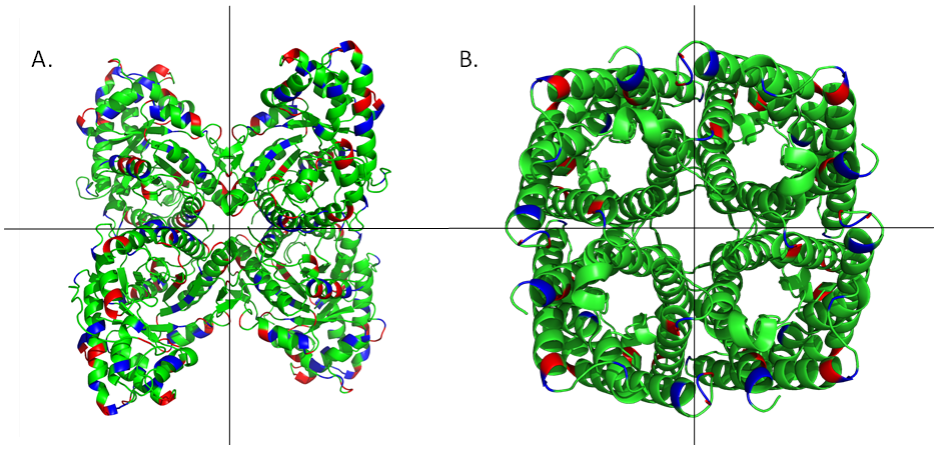


Figure S12: Crystal structures of rabbit aldolase and AqpZ tetramers.
(A) The crystal structures of rabbit aldolase and (B) aquaporin Z with positively charged amino acids (Lys and Arg) labeled in blue and negatively charged amino acids (Glu and Asp) labeled in red. The black lines indicate the complex interface. Aldolase contains many charged residues at the interface of the protein complex and aquaporin Z does not, which may explain why aquaporin Z releases monomers and aldolase does not when HCD is applied to the intact complex.

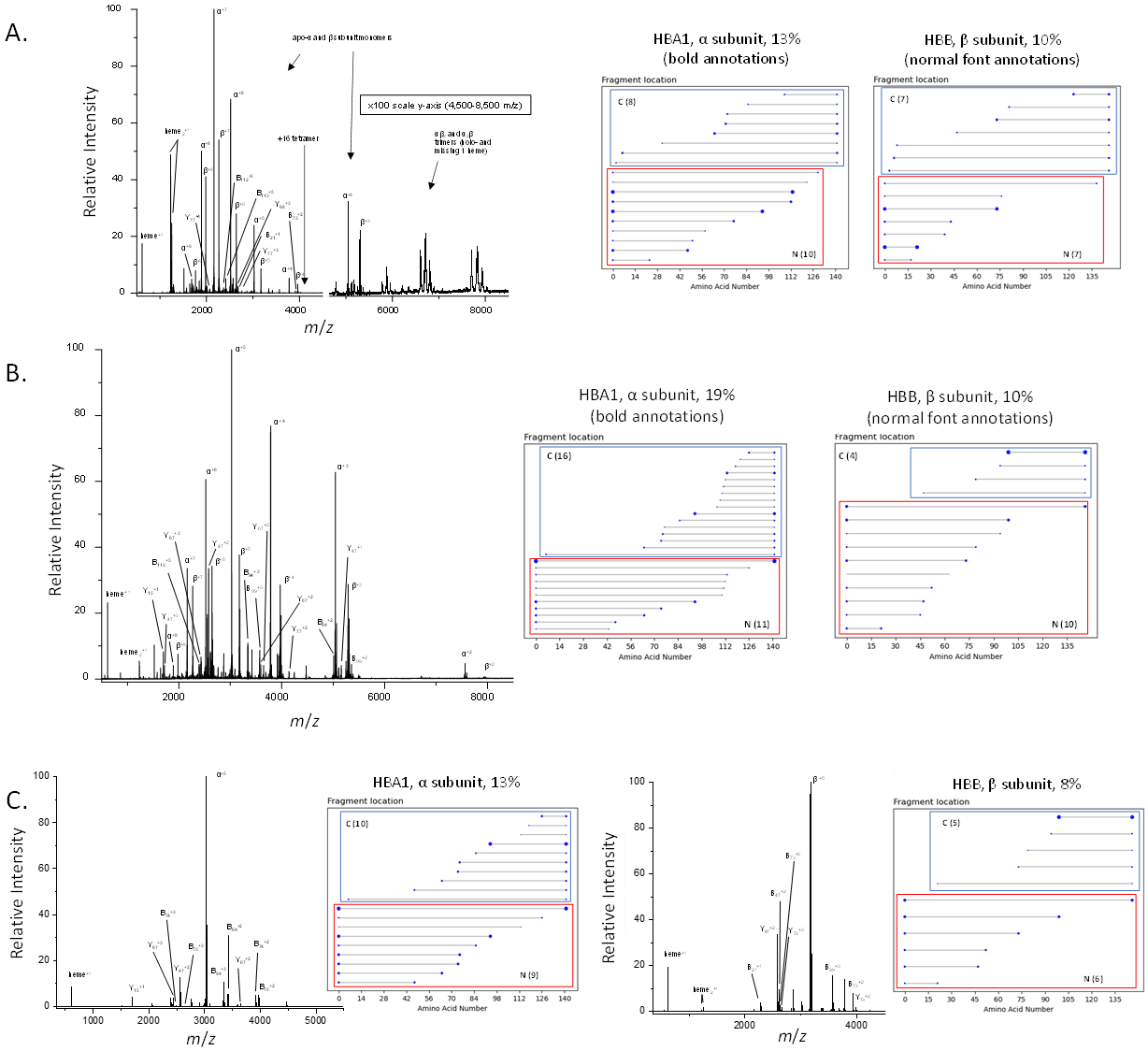


Figure S13: Complex-down vs. native top-down of the hemoglobin tetramer and dimer. (A) nTDMS spectrum of the 16+ charged hemoglobin tetramer with the corresponding fragment location maps for the α-subunit and β-subunit, (B) a nTDMS mass spectrum of the 12+ charged hemoglobin dimer with the corresponding fragmentation location maps for the α-subunit and β-subunit, and (C) complex-down fragmentation mass spectra and the corresponding fragment location maps for the 6+ charged α-subunit and β-subunit.

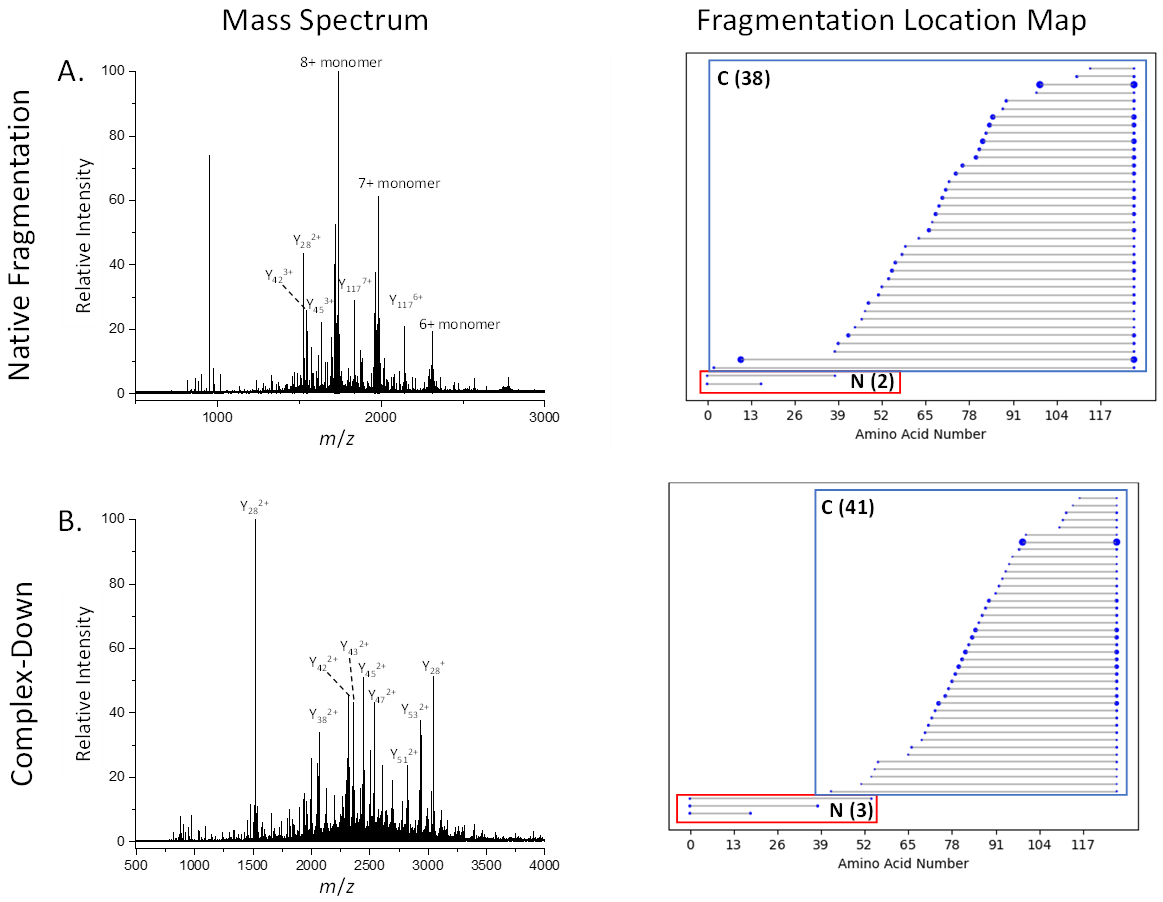


Figure S14: Complex-down vs. native top-down of the TTR tetramer.
(A) Native top-down mass spectrum (15+ tetramer) with the corresponding fragment location map and (B) a complex-down mass spectrum (6+ monomer) with the corresponding fragmentation location map for the TTR tetramer.

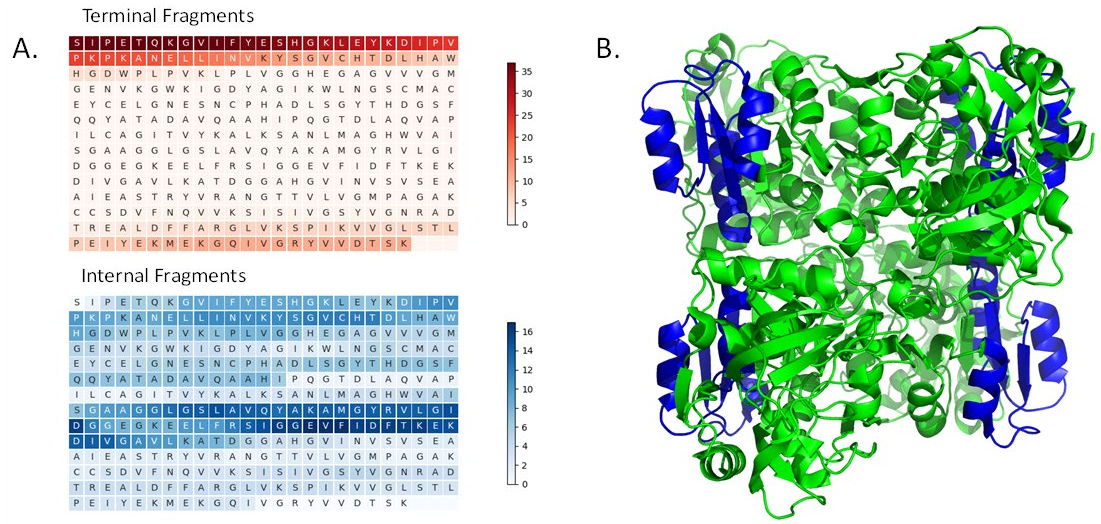


Figure S15: Terminal and internal fragment analysis of nTDMS of the ADH tetramer.
(A) A heatmap representing terminal and internal fragment analysis of the 24+ charged ADH tetramer and (B) the structure of the ADH tetramer with an internal fragment hotspot (residues 178-236) highlighted in blue. Notice how this region of the ADH tetramer is solvent exposed.

Supplementary Tables

Table S1. Molecular weights of species present in a low HCD spectrum of aldolase tetramer.

Table S1: Molecular weights of species present in a low HCD energy spectrum of aldolase. The high m/z ions in the spectrum correspond to the aldolase tetramer-y74.

Species	MW (Da)
Measured Molecular weight of Aldolase	156,982
Measured Molecular Weight Higher m/z Peaks	149,088
Molecular Weight Difference	7,894
Theoredical Mass of Y74	7,896

Table S2. Information on the complexes analyzed in this study.

Table S2: Information on the complexes analyzed in this study.

General Inforamtion				Native TD-MS Fragmentation				Complex-Down Fragmentation			
Name of Complex	Type of Complex	MW Complex (Da)	MW Monomer (Da)	Monomer Release with HCD	# N-Term Frags	# C-Term Frags	Sequence Coveage	Monomer Release with CAD	# N-Term Frags	# C-Term Frags	Sequence Coverage
ADH	Tetramer	147,472	36,738	No	60	3	18%	Yes	25	21	12%
Aldolase	Tetramer	156,748	39,187	No	8	35	11%	Yes	19	16	8%
GSTA1	Dimer	51,000	25,500	No	5	19	11%	Yes	25	20	21%
Enolase	Dimer	93,312	46,656	No	48	51	18%	Yes	27	18	8%
Creatine Kinase	Dimer	86,224	43,112	No	9	38	13%	No	NA	NA	NA
GND1	Dimer	106,003	52,957	No	23	6	6%	No	NA	NA	NA
AqpZ	Tetramer	24,269	97,076	Yes	65	62	38%	Yes	63	60	35%
Hemoglobin	Tetramer	61,986	α=15,126 β=15,867	Yes	α=10 β=7	α=8 β=7	α=13% β=10%	Yes	α=9 β=6	α=10 β=6	α=13% β=8%
Hemoglobin	Dimer	30,993	α=15,126 β=15,867	Yes	α=11 β=10	α=16 β=4	α=19% β=10%	Refer to row above	Refer to row above	Refer to row above	Refer to row above
Transthyretin	Tetramer	55,044	13,761	Yes	2	38	33%	Yes	3	41	36%

References

1. Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A., ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* **2021,** *20* (4), 1928-1935.

2. Vimer, S.; Ben-Nissan, G.; Morgenstern, D.; Kumar-Deshmukh, F.; Polkinghorn, C.; Quintyn, R. S.; Vasil'ev, Y. V.; Beckman, J. S.; Elad, N.; Wysocki, V. H., Comparative structural analysis of 20S proteasome ortholog protein complexes by native mass spectrometry. *ACS Cent. Sci.* **2020,** *6* (4), 573-588.

3. Nshanian, M.; Lantz, C.; Wongkongkathep, P.; Schrader, T.; Klärner, F.-G.; Blümke, A.; Despres, C.; Ehrmann, M.; Smet-Nocca, C.; Bitan, G., Native top-down mass spectrometry and ion mobility spectrometry of the interaction of tau protein with a molecular tweezer assembly modulator. *Journal of The American Society for Mass Spectrometry* **2018,** *30* (1), 16-23.

4. Xie, Y.; Zhang, J.; Yin, S.; Loo, J. A., Top-down ESI-ECD-FT-ICR mass spectrometry localizes noncovalent protein-ligand binding sites. *Journal of the American Chemical Society* **2006,** *128* (45), 14432-14433.

5. O'Brien, J. P.; Li, W.; Zhang, Y.; Brodbelt, J. S., Characterization of native protein complexes using ultraviolet photodissociation mass spectrometry. *Journal of the American Chemical Society* **2014,** *136* (37), 12920-12928.

6. Zhou, M.; Lantz, C.; Brown, K. A.; Ge, Y.; Paša-Tolić, L.; Loo, J. A.; Lermyte, F., Higher-order structural characterisation of native proteins and complexes by top-down mass spectrometry. *Chemical science* **2020,** *11* (48), 12918-12936.

7. Zhang, H.; Cui, W.; Wen, J.; Blankenship, R. E.; Gross, M. L., Native electrospray and electron-capture dissociation in FTICR mass spectrometry provide top-down sequencing of a protein component in an intact protein assembly. *Journal of the American Society for Mass Spectrometry* **2011,** *21* (12), 1966-1968.

8. Li, H.; Wongkongkathep, P.; Van Orden, S. L.; Ogorzalek Loo, R. R.; Loo, J. A., Revealing ligand binding sites and quantifying subunit variants of noncovalent protein complexes in a single native top-down FTICR MS experiment. *Journal of The American Society for Mass Spectrometry* **2014,** *25* (12), 2060-2068.

9. Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D.; Loo, J. A., An integrated native mass spectrometry and top-down proteomics method that connects sequence to structure and function of macromolecular complexes. *Nature chemistry* **2018,** *10* (2), 139-148.

10. Lermyte, F.; Sobott, F., Electron transfer dissociation provides higher-order structural information of native and partially unfolded protein complexes. *Proteomics* **2015,** *15* (16), 2813-2822.

11. Williams, J. P.; Morrison, L. J.; Brown, J. M.; Beckman, J. S.; Voinov, V. G.; Lermyte, F., Top-down characterization of denatured proteins and native protein complexes using electron capture dissociation implemented within a modified ion mobility-Mass spectrometer. *Analytical chemistry* **2020,** *92* (5), 3674-3681.

12. Zhou, M.; Liu, W.; Shaw, J. B., Charge Movement and Structural Changes in the Gas-Phase Unfolding of Multimeric Protein Complexes Captured by Native Top-Down Mass Spectrometry. *Analytical Chemistry* **2019,** *92* (2), 1788-1795.

13. Cammarata, M. B.; Thyer, R.; Rosenberg, J.; Ellington, A.; Brodbelt, J. S., Structural characterization of dihydrofolate reductase complexes by top-down ultraviolet photodissociation mass spectrometry. *Journal of the American Chemical Society* **2015,** *137* (28), 9128-9135.

14. Cammarata, M. B.; Brodbelt, J. S., Structural characterization of holo-and apo-myoglobin in the gas phase by ultraviolet photodissociation mass spectrometry. *Chemical Science* **2015,** *6* (2), 1324-1333.

15. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **2007,** *4*, 709-712.

16. Hale, O. J.; Cooper, H. J., Native Mass Spectrometry Imaging of Proteins and Protein Complexes by Nano-DESI. *Anal. Chem.* **2021,** *93*, 4619−4627.

17. Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A., Top or middle? Up or down? A standard lexicon for protein top-down and allied mass spectrometry approaches. *J. Am. Soc. Mass Spectrom.* **2019,** *30*, 1149-1157.

18. Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Vale, L. H. F. D.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentic´, K.; Kelleher, N. L.; Compton, P. D., Top-down characterization of endogenous protein complexes with native proteomics. *Nature Chem. Biol.* **2018,** *14*, 36-41.

19. Li, H.; Wolff, J. J.; Van Orden, S. L.; Loo, J. A., Native top-down electrospray ionization-mass spectrometry of 158 kDa protein complex by high-resolution Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2014,** *86* (1), 317-20.

20. Campuzano, I. D. G.; Nshanian, M.; Spahr, C.; Lantz, C.; Netirojjanakul, C.; Li, H.; Wongkongkathep, P.; Wolff, J. J.; Loo, J. A., High Mass Analysis with a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: From Inorganic Salt Clusters to Antibody Conjugates and Beyond. *J. Am. Soc. Mass Spectrom.* **2020,** *31*, 1155−1162.

21. Laganowsky, A.; Reading, E.; Hopper, J. T.; Robinson, C. V., Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.* **2013,** *8*, 639-651.

22. Lippens, J. L.; Nshanian, M.; Spahr, C.; Egea, P. F.; Loo, J. A.; Campuzano, I. D. G., Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry as a Platform for Characterizing Multimeric Membrane Protein Complexes. *J. Am. Soc. Mass Spectrom.* **2018,** *29*, 183-193.

23. Ogorzalek Loo, R. R.; Loo, J. A., Salt Bridge Rearrangement (SaBRe) Explains the Dissociation Behavior of Noncovalent Complexes. *J. Am. Soc. Mass Spectrom.* **2016,** *27*, 975-990.

24. Yin, S.; Loo, J. A., Elucidating the Site of Protein-ATP Binding by Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 899-907.

25. Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L., Analyzing Internal Fragmentation of Electrosprayed Ubiquitin Ions During Beam-Type Collisional Dissociation. *J. Am. Soc. Mass Spectrom.* **2015,** *26*, 782-787.

26. Schmitt, N. D.; Berger, J. M.; Conway, J. B.; Agar, J. N., Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem.* **2021,** *93*, 6355–6362.

27. Wei, B.; Zenaidee, M. A.; Lantz, C.; Ogorzalek Loo, R. R.; Loo, J. A., Towards Understanding the Formation of Internal Fragments Generated by Collisionally Activated Dissociation for Top-Down Mass Spectrometry. *Anal. Chim. Acta* **2022,** *1194*, 339400.

28. Zenaidee, M. A.; Lantz, C.; Perkins, T.; Jung, W.; Ogorzalek Loo, R. R.; Loo, J. A., Internal fragments generated by electron ionization dissociation enhance protein top-down mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2020,** *31*, 1896–1902.

29. Zenaidee, M. A.; Wei, B.; Lantz, C.; Wu, H. T.; Lambeth, T. R.; Diedrich, J. K.; Loo, R. R. O.; Julian, R. R.; Loo, J. A., Internal Fragments Generated from Different Top-Down Mass Spectrometry Fragmentation Methods Extend Protein Sequence Coverage. *J. Am. Soc. Mass Spectrom.* **2021,** *32*, 1752-1758.

30. Harvey, S. R.; Porrini, M.; Konijnenberg, A.; Clarke, D. J.; Tyler, R. C.; Langridge-Smith, P. R.; MacPhee, C. E.; Volkman, B. F.; Barran, P. E., Dissecting the dynamic conformations of the metamorphic protein lymphotactin. *The Journal of Physical Chemistry B* **2014,** *118* (43), 12348-12359.

31. Pagel, K.; Hyung, S.-J.; Ruotolo, B. T.; Robinson, C. V., Alternate Dissociation Pathways Identified in Charge-Reduced Protein Complex Ions. *Anal. Chem.* **2010,** *82*, 5363–5372.

Chapter 7: Conclusions and Outlook

1. Conclusions

This work explores the utility of protein internal fragment ions generated by top-down and middle-down mass spectrometry (TD-MS and MD-MS), from both fundamental and application perspectives, for more efficient and comprehensive sequence, structure, and PTM characterization of proteins of biological and therapeutic significance.

In chapter 2, the fundamental formation mechanism of internal fragments was established. We first demonstrated that many previously unassigned mass spectral signals can be assigned as internal fragments and can provide additional and complementary protein sequence information. We then summarized and compared the fragmentation propensities of over 3200 terminal and internal fragments generated from TD-MS of a series of polypeptides. The similar, yet slightly suppressed preferential cleavages at certain amino acid residues of internal fragments suggest that they should come from subsequent cleavages of terminal fragments and the formation of internal fragments can be correlated to the well-known mobile proton model. This sequential fragmentation event results in more readily available mobile protons to generate internal fragments, leading to a more evenly distributed fragmentation propensities across all residues for internal fragments compared with terminal fragments. This chapter enhances our understanding of the formation of internal fragments and is beneficial for the development of protein sequencing algorithms to assign internal fragments more accurately and reliably. As a direct result, it provided valuable insights into the development of ClipsMS, a python-based program designed for internal fragment analysis developed by my co-worker.

We delved into the application of internal fragments for disulfide-intact protein characterization in chapter 3. By analyzing four standard disulfide-intact proteins including β-lactoglobulin (2 disulfide bonds), lysozyme (4 disulfide bonds), ribonuclease A (4 disulfide bonds), and α-lactalbumin (4 disulfide bonds), we demonstrated that internal fragments could access the interior protein sequence constrained by multiple disulfide bonds that cannot be reached by terminal fragments to increase the protein sequence coverage by 20-60% to achieve almost 100% sequence coverage. In addition, internal fragments mainly cleave interior disulfide bonds while terminal fragments cleave more disulfide bonds located on the exterior of the protein, establishing a correlation between the relative position of disulfide bonds and the number of internal fragments responsible for each disulfide cleavage. Importantly, we show that internal fragments that retain intact disulfide bonds can be used to directly determine the disulfide connectivity. This chapter serves as a proof-of-concept for applying internal fragments to elucidate disulfide bridges and pave the way for utilizing internal fragments for the characterization of proteins with more complex disulfide linkage patterns such as mAb based biotherapeutics.

Inspired by the work presented in chapter 3, we applied TD-MS with the analysis of internal fragments to characterize highly heterogeneous intact mAbs and ADCs in chapter 4. The assignment of internal fragments significantly increases the sequence coverage of intact NIST mAb to over 75% by accessing the highly disulfide-bonded sequence regions that are hardly covered by terminal fragments. The sequence coverage reported here is the highest value achieved by TD-MS alone on an intact mAb. In addition, important PTM information including

disulfide connectivity patterns and N-glycosylations, can be obtained by analyzing internal fragments. This shows the promise of applying internal fragments to better identify and localize potential modification liabilities such as oxidation and deamidation. By applying this workflow to characterize an IgG1-DM1 lysine linked ADC, we demonstrated that the assignment of internal fragments is beneficial for pinpointing drug conjugation sites which cannot be achieved by analyzing terminal fragments alone. The analytical workflow presented in this chapter promises to be incorporated into the biopharma industry once robust automation and computational support can be established. This chapter also suggests the potential of integrating internal fragment analysis into bottom-up and middle-down workflows to push the characterization of mAb based therapeutics to a near complete level on a routine basis.

The TD-MS work presented in chapter 4 shows great potential, yet the results were still far from ideal and has plenty of room to improve. The MD-MS work presented in chapter 5 built upon and demonstrated significant improvement from the TD-MS project presented in chapter 4. We developed an innovative MD-MS platform using direct infusion instead of traditional online rpLC for sample introduction. This allows us to take advantage of the high flexibility of direct infusion to maximize the ECD fragmentation efficiency to achieve nearly 100% sequence coverage of all three subunits of NIST mAb by accessing the interior protein sequence. Again, important N-glycosylation information can be elucidated by assigning internal fragments, laying the foundations for applying internal fragments to identify low-level PTMs. The same IgG1-DM1 ADC used in chapter 4 was analyzed using this novel native direct infusion MD-MS workflow and 80% of DM1 conjugation sites were confirmed, comparable to the 83% achieved by the

peptide mapping method. Overall, we demonstrated that using our MD-MS workflow, similar characterization results on both mAbs and ADCs can be achieved compared to the bottom-up peptide mapping method. This chapter highlights the potential of MD-MS and internal fragment analysis to advance the MS technology for the characterization of mAb based therapeutics.

In chapter 6, we demonstrated that in contrary to the long-standing belief that collision-based fragmentation cannot produce conformation-sensitive fragments to reveal higher-order structure of protein assemblies, HCD on an orbitrap instrument yields fragmentation patterns that agree well with the solvent-accessible or interface regions of several multi-subunit protein complexes. HCD-derived internal fragments played a critical role in this aspect by delivering structural information of alcohol dehydrogenase tetramers. This chapter proves that HCD can be a powerful biophysics tool to probe protein structures without the need of less accessible electron- and photon-based fragmentation techniques.

This work shows the potential and sets the stage for incorporating internal fragment analysis into the mainstream MS technology for the interrogation of protein primary sequence, important PTMs, ligand binding, and higher order structure which is crucial for understanding protein function in biological systems.

2. What's next for research on internal fragments?

Further developments in internal fragments research mainly lies on informatics to improve the confidence and reduce the false positive rate of assigning these informative yet hard-to-assign mass spectral signals. There are a few steps within the workflow that need to be improved. Firstly, deconvolution remains problematic and a more robust and reliable deconvolution tool is

needed to reduce the possibility of falsely assigning noise as real signals. This has been an issue in TD-MS for a long time but is specifically jeopardizing the accuracy of assigning internal fragments as the theoretical possibility of internal fragments is much higher than terminal fragments, and all the wrongly deconvoluted noise could be more likely assigned as internal fragments. In addition, building a mature scoring system like what we have in the well-established bottom-up workflow to assess the assignment result is another potential development. Currently, all internal fragment assignments need to be manually validated which is not only time-consuming and manual heavy, but also subjective and can cause plenty of inter-laboratory variations. Therefore, a robust scoring system to streamline and automate the data analysis workflow is needed. This also significantly increases the throughput of the workflow. Lastly, as mentioned in the introduction, a universal nomenclature system of internal fragments needs to be established. This would first require a full understanding of the formation mechanism of all types of internal fragments generated from various common fragmentation techniques; our TD-MS community need to discuss and agree upon a widely accepted notation of all types of internal fragments afterwards. Our group is actively working with other labs to achieve this goal.

Internal fragments research can also benefit from the advancements in mass spectrometry instrumentation. For example, higher mass accuracy and resolution can significantly improve the assignment confidence. Furthermore, adapting more complicated MS workflows on instruments like Thermo tribrids which has the capability of multistage tandem mass spectrometry and Waters Cyclic IM-MS which has high resolution ion mobility can also further

internal fragments research. I believe the future of the application of internal fragments is bright and can possibly bloom once the informatics basics can be established.