Title

Mass Spectrometry Analysis Reveals Sequence and Higher-Order Structure Information for Proteins and Protein Complexes

Author

Lantz, Carter

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Mass Spectrometry Analysis Reveals Sequence and Higher-Order Structure Information for Proteins and Protein Complexes

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Biochemistry, Molecular and Structural Biology

by

Carter Lantz

2022

ABSTRACT OF THE DISSERTATION

Mass Spectrometry Analysis Reveals Sequence and
Higher-Order Structure Information for Proteins and
Protein Complexes

by

Carter Lantz
Doctor of Philosophy in Biochemistry, Molecular and Structural Biology
University of California, Los Angeles, 2022
Professor Joseph Ambrose Loo, Chair

Mass spectrometry (MS) has been found to be a useful technique for the study of various compounds. Fundamentally, MS is a way to measure the masses of compounds for the purpose of identifying and quantifying those compounds. Protein mass spectrometry, where proteins are the analyte of interest, has been found to return relevant information on those proteins including the mass, the identity and location of modifications, and even aspects of protein higher-order structure. The work here aims to develop mass spectrometry methods for the study of proteins and to use those methods to investigate amyloid proteins common in neurodegenerative diseases. Herein, it is described how ClipsMS, a program that assigns internal fragments resulting from top-down mass spectrometry (TD-MS), can be utilized to increase the sequence coverage of proteins and locate modifications. This work also illustrates how native TD-MS of large protein complexes with high-energy C-trap dissociation (HCD) can release covalent fragments that reveal aspects of higher-order structure. Furthermore, it is described how TD-MS of proteins with electron capture dissociation (ECD) on an orbitrap-based mass spectrometer can return relevant information on proteins including sequence information, the location of modifications, and higher-order structure information on protein complexes. In this work, MS techniques were also utilized to characterize

neurodegenerative disease protein monomers and oligomers. The research conducted reveals the location of phosphorylation sites on commonly phosphorylated amyloid proteins and that phosphorylation compacts the gas-phase structure of those proteins. It is possible that structure alteration due to phosphorylation could modulate the aggregation potential of amyloid proteins. In addition, it was found that CLR01, a molecular tweezer compound that has been found to inhibit amyloid protein aggregation, directly interacts with the N-terminus of multiple proteoforms of the amyloid protein α-synuclein and that the molecule compacts the gas-phase structure of the protein. It is possible that compaction of the N-terminal region of α-synuclein by CLR01 could prevent monomers from interacting with one another and forming oligomers and fibrils of the protein. Lastly, MS was utilized to determine size information and aggregation interface information for various amyloid protein oligomers. Characterization of these small aggregates could provide information on how they become toxic in brain neurons. The data presented here aims to further mass spectrometry methods for the characterization of proteins and to use those methods to reveal protein aggregation mechanisms and discover possible therapies for neurodegenerative diseases.

The dissertation of Carter Lantz is approved.

Jose Alfonso Rodriguez

Steven G. Clarke

Gal Bitan

Joseph Ambrose Loo, Committee Chair

University of California, Los Angeles

2022

To my supportive parents, siblings

and

to Grandpa (Leon) Magner,

for being my inspiration

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1:

CHAPTER 2:

CHAPTER 6:

CHAPTER 7:

LIST OF TABLES

CHAPTER 2:

CHAPTER 3:

LIST OF SCHEMES

CHAPTER 3:

CHAPTER 4:

ACKNOWLEDGEMENTS

I'd like to thank my advisors Dr. Joseph Loo and Dr. Rachel Loo for their support during my Ph.D. The completion of this Ph.D. would not have been possible without their support and guidance. It has been a joy to work with them and I have learned much from them. I would also like to thank Dr. Gal Bitan, Dr. Jose Rodriguez, Dr. David Eisenberg, and Dr. Steven Clarke for all the conversations that furthered my research and the samples I received over the years.

I would also like to thank all the people who have helped with many of my projects. The development of ClipsMS would not have been possible without the help of Dr. Muhammad Zenaidee and Benqian Wei. The study of large protein complexes with HCD was efficiently completed with the help of Dr. Wonhyeuk Jung, Benqian Wei, Boyu Zhao, Andrew Goring, and Jessie Le. In addition, I would like to thank Dr. Yu Chen and Dr. Greg Khitrov for their help on many of my projects.

I would also like to thank the engineers and scientists at Thermo and especially the help of Rosa Viner and Weijing Liu for allowing us to collect data on their instrument. Finally, I would like to thank Iain Campuzano for collaborating with us on many projects in the lab.

CURRICULUM VITAE

EDUCATION

Baylor University, Waco, TX
B.A. in University Scholars **May 2017**
Honors Program
Concentrations in Biochemistry and Mathematics

PEER REVIEWED PUBLICATIONS

Samanta, N. et. al., Superoxide dismutase 1 folding stability as a target for molecular *Tweezers in SOD1-related amyotrophic lateral sclerosis.* ChemBioChem. **2022** **September 2022**

Wei, B. et. al., *Towards understanding the formation of internal fragments generated by collisionally activated dissociation for top-down mass spectrometry.* Analytica Chimica Acta. **2021**, 1194, 15. **December 2021**

Shahpasand-Kroner, H. et. al., *Three-repeat and four-repeat Tau isoforms form different oligomers.* Protein Science. **2021**, 31, 3, 613-627. **December 2021**

Zenaidee, M. et. al., *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. **June 2021**

Lantz, C. et. al., *ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting From Top-Down Mass Spectrometry.* Journal of Proteome Research, **2021**, 20, 4, 1928-1935. **March 2021**

Zhou, M. et. al., *Higher-order structural characterisation of native proteins and complexes by top-down mass spectrometry.* Chemical Science, **2020**, 11, 12918-12936. **October 2020**

Zenaidee, M. et. al., *Internal Fragments Generated by Electron Ionization Dissociation Enhances Top-down Mass Spectrometry*. Journal of The American Society for Mass Spectrometry, **2020**, 31, 9, 1896-1902. **August 2020**

Campuzano, I. et. al., *High Mass Analysis with a Fourier Transform Ion Cyclotron Resonance Mass spectrometer: From Inorganic Salt Cluster to Antibody Conjugates and Beyond.* Journal of The American Society for Mass Spectrometry, **2020**, 31, 5, 1155-1162. **March 2020**

Donnelly, D. et. al., *Best practices and benchmarks for intact protein analysis for top-down mass spectrometry.* Nature Methods, **2019**, 16, 587-594. **July 2019**

Nshanian, M. et. al., *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, **2019,** *30*, 1, 16-23. **January 2019**

Klyczek, K.K., et al., *Tales of diversity: Genomic and morphological characteristics of forty-six Arthrobacter phages.* PloS One, **2017**, 12(7). **July 2017**

AWARDS AND FELLOWSHIPS

UCLA Chemists Association Dissertation Award
UCLA Department of Chemistry and Biochemistry, Los Angeles, CA **May 2022**

Dissertation Year Fellowship
UCLA Graduate Division, Los Angeles, CA **Jan 2022- Dec 2022**

ASMS Graduate Student Travel Award
The American Society for Mass Spectrometry, Philadelphia, PA **November 2021**

Fowler Fellowship for Protein Science
UCLA Molecular Biology Institute, Los Angeles, CA **October 2021**

John M. Jordan Memorial Award for Excellence in Research
UCLA Department of Chemistry and Biochemistry, Los Angeles, CA **May 2021**

Agilent Technologies Student Travel Award
Advancing Mass Spectrometry Conference, Amhurst, MA **July 2019**

Chemical and Molecular Biology NIH Fellowship
University of California Los Angeles, Los Angeles, CA **July 2018 – June 2020**

Undergraduate Poster Award
The American Society for Mass Spectrometry, Indianapolis, IN **June 2017**

Outstanding Presentation in Lipid Chemistry
Gulf Coast Undergraduate Research Symposium, Houston, TX **October 2016**

OTHER WORKS

Advanced Fragmentation Methods in Biomolecular Mass Spectrometry: Probing Primary and Higher Order Structure with Electrons, Photons and Surfaces
Royal Society of Chemistry. Editor: Frederik Lermyte. Published 12/15/2020 **December 2020**

A mass spectrometry view of the interaction of Alzheimer's disease proteins with an anti-aggregation compound
Atlas of Science, Joseph A. Loo, Posted 1/30/2020 **January 2020**

The Mass Spectrometry Analysis of Clamydomonas reinhardtii During Nitrogen Starvation
Baylor University's Digital Repository, Carter Lantz, Posted 5/23/2017 **May 2017**

CODE REPOSITORIES

Comprehensive Localization of Internal Protein Sequences (ClipsMS)
This program will generate terminal and internal masses of a protein sequence and match those masses to a deconvoluted mass list from a top-down mass spectrum. It can be utilized to increase sequence coverage of proteins and locate various modifications on those proteins. (https://github.com/loolab2020/ClipsMS/releases) **December 2020**

CHAPTER 1

Introduction

PROTEIN STRUCTURE AND FUNCTION

Proteins are the molecular machines that drive the processes necessary to sustain life. These proteins facilitate various reactions in the cell that would otherwise occur infrequency. A common example of a protein facilitating an important reaction is glucose 6-phosphatase which converts glucose 6-phosphoate (G6P) to glucose.[1, 2] The conversion from G6P to glucose is highly endergonic, so the liver utilizes glucose 6-phosphatase to catalyze the reaction converting G6P to glucose. In this example, glucose 6-phosphatase is utilized to overcome a large energy barrier to perform a reaction that would not otherwise be possible. Proteins like glucose 6-phosphatase are imperative for sustaining life, so characterization of these molecular machines is quite important. Proteins are synthesized by the central dogma of biology, a term that was first coined by Francis Crick in 1958.[3] This term came to represent the idea that DNA in the nucleus of cell is transcribed to RNA which is then translated into a protein sequence.[4] This is the fundamental concept for protein synthesis and the basis for molecular and structural biology.

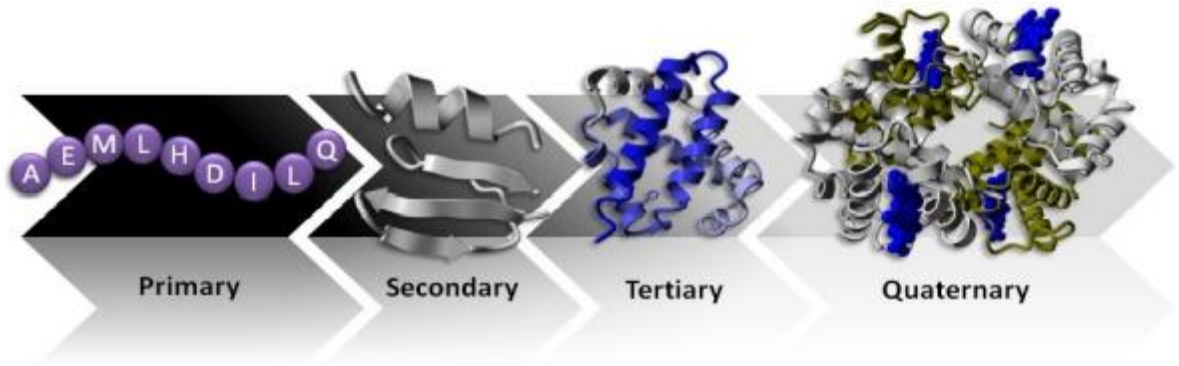


Figure 1: A graphic portraying the layers of protein structure from primary structure (left) to quaternary structure (top left).[5]

To understand how proteins function, it is important to understand how the protein is structured. (Fig. 1) The first layer of protein structure is the primary structure or the sequence of the protein. The first protein sequence to be identified was insulin discovered by Frederik Sanger in 1953.[6, 7] This study established that proteins are distinguished by unique covalent sequences of amino acids. Since then there

have been over 8 million unique sequences reported[8] the largest of which is β-galactosidase which was sequenced by Dr. Audree Fowler in the UCLA Department of Chemistry and Biochemistry in 1978.[9] Protein primary structure has been shown to be essential for correct protein folding and proper function of the protein. This concept was discovered by experimenting with heat denaturation of bacteriophage tails.[10, 11] In these experiments, it was found that heat prevented mutant bacteriophage tails from forming. However, at lower temperatures, the tails of the bacteriophages were formed correctly, and the tails became functional. Notice in this example how the native structure of the bacteriophage tails was formed from the primary structure of the 76kDa monomers. This experiment demonstrated that protein primary structure is necessary but not solely sufficient for proteins to function. This concept has driven the scientific community to investigate how higher-order protein structure is necessary for the function of proteins.

Protein higher-order structure includes intramolecular interactions within the chain of a protein monomer that does not involve covalent peptide bonds. These interactions can be facilitated by hydrogen bonds, disulfide bonds, or salt bridge interactions between amino acid residues. Secondary structures of proteins, which includes α-helical and β-sheet structures, constitute rigid localized structures that are utilized to facilitate protein function. Aquaporin Z, for example, contains numerous α-helices which create a channel across membranes that allow for the passage of water.[12] The concanavalin A protein, on the other hand, utilizes an arrangement of β-sheets to bind metal ions and carbohydrates.[13] In these examples, localized α-helical and β-sheet structures of amino acids facilitate protein function. The global structure of a protein monomer, known as the tertiary structure, provides the necessary arrangement of atoms in space for the protein to perform its intended function. The retinol binding protein arranges into a beta-barrel structure that allows for capture and transport of the hydrophobic molecule vitamin A through the bloodstream.[14] In another example, the structure of myoglobin allows for binding of the heme group which allows for oxygen binding to the protein monomer.[15] In these examples, higher-order structure characteristics within a monomer provides the monomer with the ability to perform its necessary function.

Protein higher-order structure can also refer to established intermolecular interactions between different subunits or the transitory interactions between subunits and other ligands. The arrangements of

subunits in a defined protein complex, also known as quaternary structure, can play a role in the function of protein complexes. The quaternary structure of the hemoglobin tetramer,[16] for example, transitions between the R state and the T state allowing for efficient uptake of oxygen in the lungs and release of oxygen in tissues.[17] The quaternary structure of the pyruvate dehydrogenase complex allows for multiple reactions to occur to efficiently convert pyruvate to acetyl-CoA.[18] Quinary structure, a relatively understudied concept, describes the transient interaction between the target protein and other surrounding proteins and ligands.[19, 20] Although the quinary structure of proteins is difficult to decipher, there are a few studies that have attempted to uncover the quinary structure of key proteins such as ribosomes.[21] Characterizing protein structure at all levels provides explanations for protein function and illuminates how cells regulate key metabolic processes necessary to support life.

Correct primary sequence and higher-order structure is necessary for the proper functioning of proteins; however, it has also been found that modifications to the protein sequence can affect the function of proteins. It has been found that covalent post-translational modifications (PTMs) can modulate the function of many proteins and therefore various cellular processes. For example, it has been found that methylation of histones has been found to regulate gene expression[22] and methylation of isomerized proteins mitigates the process of aging.[23] Acetylation modulates the function of p53 and thus the regulation of the cell cycle and cellular apoptosis,[24] and maintains the translational rate of proteins synthesis when added to ribosomes.[25] Phosphorylation has been shown to be the most prevalent PTM and has been found to have a profound effect on numerous cellular processes. For example, the peripheral membrane protein c-Ras regulates differentiation and cell growth by phosphorylating various proteins[26, 27] and those proteins have been the target of many cancer therapies.[28] Crosstalk between PTMs is common.[29] In fact, crosstalk between acetylation and methylation regulates p53 function[30] and crosstalk between acetylation and phosphorylation regulates the microtubule binding affinity and disease pathology of the amyloid protein tau.[31] In addition, non-covalent modifications has also been revealed to affect the function of proteins. For example, the catalytic zinc metal ion binding to alcohol dehydrogenase facilitates the reaction reducing acetaldehyde to ethanol.[32] Drug binding to HIV-1 reverse transcriptase alters the mobilities of certain domains preventing

the protein from functioning.[33] These examples indicate that modifications can affect the function of proteins which can regulate their activity in the cell.

Numerous studies point to the fact that modifications affect the function of proteins; however, determining how modifications affect the structure of proteins is more difficult. Some studies show that structural changes due to the addition of PTMs can alter protein function. For example, phosphorylation of the N-terminus of glycogen phosphorylase induces the formation of a helical structure necessary for stable formation of the dimer.[34] Recently, it was found that phosphorylation induces a structural change in the mitochondrial calcium uniporter affecting dimerization of tetramers.[35] Acetylation has also been found to alter the conformational dynamics of the αK40 loop of α-tubulin which restricts its interaction with acetylase.[36] Even methylation has been found to stabilize a domain of a chemotaxis receptor which controls the binding of the CheA kinase.[37] Non-covalent modifications can also affect the structure of proteins. Zinc binding to alcohol dehydrogenase provides structure to the protein complex monomers.[32] Ion-mobility mass spectrometry has also revealed that non covalent salt adducts can even shift the structure of protein ions.[38] These examples reveal that understanding the effect modifications have on protein structure can provide information on protein dynamics and may provide explanations on how their function is altered.

PROTEIN MISFOLDING AND AMYLOID PROTEIN AGGREGATION

Protein misfolding can result in the loss of protein function and in some cases can lead to protein aggregation. Protein misfolding diseases such as sickle cell disease characterized by the misfolding of the protein hemoglobin[39] and cystic fibrosis characterized by the misfolding of the cystic fibrosis transmembrane conductance regulator[40] occur due to various mutations to the protein sequence. Misfolded proteins are normally labeled by ubiquitin for degradation by proteosomes[41] or decomposed with lysozymes.[42] When lysosomes fail to function properly, various disorders are common.[43] Misfolding and insufficient degradation of certain proteins known as amyloid proteins lead to the aggregation of these proteins and the onset of neurodegenerative diseases.[44] (Fig. 2) These diseases affect over 50 million people worldwide,[45] and they were first reported by Alois Alzheimer in the early 20th century.[46] These neurodegenerative diseases which include Alzheimer's disease, Parkinson's disease, Machado-Joseph's

disease, and Huntington's disease among others are characterized by protein deposits in the brain. Although the existence of these diseases has been known for quite some time and much research has been performed on these proteins, a complete understanding of amyloid protein aggregation as well as a cure remain elusive.

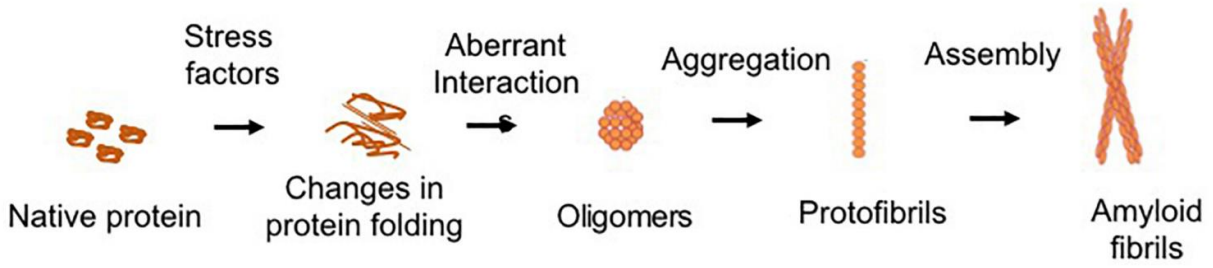


Figure 2: The process of misfolding and aggregation of amyloid proteins which cause the death of brain neurons.[47]

Tau is an amyloid protein that has been studied extensively due to its involvement in various neurodegenerative diseases, coined tauopathies.[48] Tau's intended function is to stabilize microtubules allowing for the polymerization of tubulin, and it is found that tubulin does not assemble into microtubules without tau.[49] Tau is an intrinsically disordered protein that has 6 isoforms ranging from 352 to 441 amino acids and contains 4 broad domains including the N-terminal domain, the proline rich domain, the microtubule binding domain, and the C-terminal domain.[50] Aggregates of tau have been studied extensively. It has been reported that the 3R isoforms of tau aggregate more aggressively than the 4R isoforms.[51] Aggregation prone peptides based off of tau's sequence have been reported to exist in the microtubule binding region of tau including the sequence VQIVYK.[52] These sequences have been found to bind together by a network of hydrogen bonds creating steric zippers of amino acids. Since then, it has been found that fibrils of tau also form these steric zipper structures and have multiple regions along the protein sequence that are prone to aggregate.[53] In Alzheimer's disease, tau aggregates manifest as neurofibrillary tangles in brain neurons.[54] It is possible that these aggregates of tau could induce cellular degeneration and the death of neurons in the brain.

α-Synuclein is another intrinsically disordered protein that has also been studied extensively due to its involvement in various neurodegenerative diseases, coined synucleinopathies.[55] α-Synuclein is an

intrinsically disordered protein with 120 amino acids with 3 broad domains including the amphipathic lipid binding N-terminal domain, the NAC-core, and the acidic C-terminus.[56] The intended function of α-synuclein is not clear; however, there some studies suggests that it helps recruit vesicles to the membrane of neurons[57] and this seems to be modulated by calcium binding.[58] Aggregates of α-synuclein have also been studied extensively. Peptide regions in the NACore and the preNAC core have been shown to form tight hydrogen bond interactions that resemble steric zipper structures.[59] Cryo-EM structures of α-synuclein fibrils also show tight interactions between protein monomers that form these structures.[60] In Parkinson's disease, α-synuclein aggregates manifest as Lewy bodies in brain neurons.[61] These large aggregates of α-synuclein could induce cellular degeneration and the death of neurons in the brain.

There is no consensus on the exact reason these proteins aggregate; however, there are factors which seem to modulate their rate of aggregation. Post-translational modifications seem to be one of those factors. For example, threonine 231 seems to be correlated with progression of AD[62] while acetylation at lysine 321 and 353 has been found to decrease the rate of tau aggregation.[63] In addition, phosphorylation of α-synuclein at serine 129 has been linked to increased α-synuclein,[64] while acetylation of the N-terminus has been found to decrease aggregation.[65] It has been reported that crosstalk between PTMs such as Lys-321 and phosphorylation at Serine 324 on tau can also modulate aggregation.[31] It is clear that modifications affect amyloid protein aggregation potential; however, the reason why these modifications affect the aggregation rate is still not clear.

Mutations can also affect the aggregation of amyloid proteins. Tau contains at least 37 known mutations with many resulting in increased aggregation of tau.[66] α-Synuclein also contains various mutations, some forming increased amounts of inclusions and others forming more oligomeric species.[67] Amyloid-β, a small 40 amino acid peptide linked to Alzheimer's disease, contains a proreform with 2 extra hydrophobic amino acids on its tail which lowers the barrier for nucleation.[68] The research in these studies suggest that mutations can reduce the protein's ability to perform its original function, decrease solubility of the protein, or lower the barrier of nucleation which would increase the aggregation rate of these amyloid

proteins. Although much research has been done on this subject, more research is needed to determine the mechanism of aggregation for these mutated amyloid proteins.

Metal binding has also been shown to modulate the rate of aggregation of amyloid proteins. In 1973 it was reported that high levels of aluminum were present in the brains of Alzheimer's disease patients.[69] Since then elevated metal levels have been found to correlate with other neurodegenerative diseases such as Parkinson's disease[70] and amyotrophic lateral sclerosis.[71] The reason accumulation of metals is correlated with neurodegenerative diseases is not abundantly clear. However, it is thought that metals tend to oxidize amyloid proteins causing various complications including apoptosis.[72, 73] Other studies have found that metal ions directly interact with amyloid proteins which could alter the structure and thus the aggregation propensity of amyloid proteins.[74-76] The research provided so far has given us many leads; however, more research is needed to pinpoint the reason metals correlate with increased levels of amyloid protein aggregation.

Small molecule inhibitors of amyloid proteins have also been shown to modulate amyloid protein aggregation. These aggregation inhibiting compounds can be in the form of small aromatic compounds, small peptides, or antibodies. Some small aromatic compounds such as CLR01,[77] EGCG,[78] and rifamycin[79] have been shown to slow the aggregation of amyloid proteins or to disaggregate existing fibrils. In some cases, these molecules have been found to directly interact with these amyloid proteins and bind in aggregation prone regions of the protein.[77, 80, 81] Small peptides have been found to inhibit aggregation of amyloid proteins. The first aggregation inhibiting peptides were targeted at inhibiting amyloid-β aggregation[82] and some peptides seemed to be stable in monkey cerebral fluid.[83] This logic was further utilized to form inhibitors for α-synuclein[84] and the islet amyloid polypeptide.[85] Since then, structure based peptide inhibitors created with help of computer algorithms have also been found to inhibit aggregation of amyloid proteins.[86, 87] Lastly, antibodies have also been found to have some inhibitor effect on amyloid proteins. Structure and electron microscopy analysis shows that a M204 antibody binds toxic tau oligomers.[88] Recently, an APOE antibody inhibits tau seeding by amyloid-β in a mouse model.[89] In fact, the only approved drug for Alzheimer's disease is aducanumab which is an antibody that has a slight inhibitory

effect.[90] These compounds have proven to inhibit or disaggregate fibril formation to some extent but more importantly study of these compounds may provide information that could lead to more effective inhibitors of aggregation.

MASS SPECTROMETRY INSTRUMENTATION UTILIZED FOR PROTEIN ANALYSIS

Mass spectrometry (MS) involves the mass measurement of ions in the gas-phase. Over the past few decades, MS has become a useful technique for the analysis of organic compounds and biopolymers necessary for life including small molecules,[91] carbohydrates,[92] lipids,[93] and proteins. [94, 95] MS of peptides and proteins has been shown to reveal molecular weight,[96] sequence,[97] modification,[98] and in some cases, higher-order structure information.[99] MS has been utilized to solve a variety of questions involving proteins such as the identification of all expressed proteins in a culture of yeast[100] and the characterization of all proteoforms of native chicken ovalbumin.[101] In many cases, MS can obtain information about proteins that supplement information from other techniques. In one study, MS was utilized to determine that the presence of N-terminal acetylation and the deletion of two amino acids on the C-terminus of the 20S proteosome which could not resolved by the crystal structure of the protein.[102] MS has been found to be a useful technique for the characterization of peptides and proteins largely due to the development of instrumentation allowing for analysis of these biopolymers.

Protein and peptide analysis with mass spectrometry begins by transferring proteins into the gas-phase which can be facilitated with matrix assisted laser desorption ionization (MALDI)[103] and electrospray ionization (ESI).[104] Protein ionization with MALDI involves the use of a laser to ablate proteins mixed with a matrix compound from a plate.[105] The matrix mixed into the protein solution absorbs some of the energy from the laser allowing for the ionization of protein ions without excessive fragmentation. ESI transfers proteins and peptides directly from the solution into the gas-phase by applying either a positive or negative voltage to a small tip.[106] The voltage applied to the tip creates a Taylor cone of solution which eventually releases droplets of solution containing proteins and peptides. These droplets are desolvated until proteins and peptides are released into the gas-phase. There are multiple theories on how this occurs including the charge residue model,[107] ion evaporation model,[108] and the chain ejection model.[109] Since their development

in the late 1980s, MALDI and ESI have been utilized to ionize peptides,[110, 111] intact proteins,[105, 112] and even native protein complexes.[96, 113] Both techniques won the Nobel prize for chemistry in 2002 for their impact on protein mass spectrometry and the larger scientific community.

Once proteins are ionized into the gas-phase, protein ions must be transferred to a detector which involves the use of ion funnels, quadrupoles, and ion activation. Once proteins are ionized, they are transferred from higher pressure to lower pressure with facilitates the scattering of ions. To combat this phenomenon, ion funnels,[114, 115] a series of electrodes with decreasing diameter, focus ions which prevents scattering.[116] This greatly increases the sensitivity of mass spectrometers and allows for sufficient signal to noise of protein signal. Quadrupole mass selectors are made of 4 poles that were first developed to separate ions by their mass to charge ratio when an electric fields are applied to the rods.[117] The motion of ions in a quadrupole can be described by Mathieu's second order differential equation.[118, 119] These mass selectors have been found to efficiently transfer ions including proteins to detectors and have even been optimized for the transmission of large protein complexes such as viruses.[120] Ion activation has been found to be important for the study of proteins. The process of ionization and ion transmission contribute to the desolvation of protein ions; however, solvent, metal ion adducts, and other small molecule ligands can cling to proteins which can lead to peak broadening and loss of sensitivity. To combat this, proteins can be activated with collision activated dissociation[121] or infrared multiple photon dissociation.[122] The use of ion activation reveals highly resolved protein peaks and leads to more accurate analysis of the proteins that are ionized. The development of ion funnels, quadrupoles, and ion activation have significantly forwarded the field of protein mass spectrometry and allow for the collection of highly resolved and accurate mass spectra.

Protein and peptide mass spectrometry utilize time of flight (TOF) and Fourier transform mass spectrometry (FTMS) for the detection of ions. A TOF detector pushes ions from a trap to an electron multiplier tube.[123] The time it takes for the ions to reach the detector is dependent on the mass of the ion.[124] This technique benefits from speed and sensitivity and has been utilized for a variety of protein analyses such as intact mass analysis,[125] ion-mobility mass spectrometry,[80] and top-down mass spectrometry.[126] Fourier transform MS is performed by exciting ions into an orbit and recording the cyclotron frequency or

axial frequency with which they oscillate. Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry measures the cyclotron frequency of ions[127, 128] while orbitrap instruments measure the axial frequency.[129] A Fourier transform is then performed on the resulting waveform to obtain peaks for individual ions. Traditionally, FTMS benefits from higher resolution and accuracy and can be utilized to obtain the mass of peptides and intact proteins.[130] Additionally, collecting FT-MS spectra in absorption mode[131] or enhanced FT mode[132] can increase the resolution of resulting peaks leading to more resolved and accurate data. FTMS techniques have been found to acquire high resolution intact protein mass spectra[101] and top-down mass spectra.[133] Accurate detection of proteins and peptides with these detection systems can reveal relevant information about the identity of the proteins and peptides in the sample.

PEPTIDE AND PROTEIN SEPARATION TECHNIQUIES

For complex samples with multiple species, direct MS analysis of the entire sample is not sufficient for the detection of all species in the sample. In these samples, low abundance species will be lost as our current instrumentation does not have the sensitivity or dynamic range to accurately detect all proteins in a complex sample. To reduce sample complexity of the sample, it is common for peptides and proteins to be separated before analysis with MS. This way the maximum number of proteins and peptides in the solution can be detected without saturating the detector. Common separation techniques before MS analysis include gel separation, liquid chromatography, and capillary electrophoresis.

Performing gel electrophoresis can reduce the complexity of samples by separating out analytes by size before analysis with MS. In this workflow, a protein mixtures are separated out with 1D or 2D gel electrophoresis and the gel bands are then cut from the gel and the proteins in the gel are digested.[134] The digested peptide solution can then be analyzed with MALDI or ESI MS. 2D gel electrophoresis involves the separation of peptides by size and isoelectric point. This has been performed on the proteome of lung cancer patients for the identification of biomarkers[135] and to identify modifications on proteins.[136] Gel-based separation has also been found to benefit the analysis of intact proteins. This can be done by directly ionizing the proteins in the gel by MALDI,[136] or by first extracting proteins from the gel before analysis.[137]

Gel separation before MS analysis yields greater identification of proteins compared to direct analysis of the protein solution.

Other common separation techniques that are utilized in tandem with mass spectrometry include reverse-phase liquid chromatography, strong cation exchange chromatography, and size exclusion chromatography. Reverse-phase liquid chromatography separates out peptides by hydrophobicity and is commonly utilize to separate out peptides for bottom up proteomics.[138, 139] Reverse-phase liquid chromatography MS has been utilized for a variety of applications including localizing phosphorylation sites on proteins[140] and identifying biomarkers in pathogenic bacteria[141] and irradiated mice.[142] In addition, reverse-phase liquid chromatography MS has been found to separate out intact proteins proteins.[143] Strong cation exchange mass spectrometry has also been utilized as a robust method to separate out peptides before analysis.[144] It has also been adapted for the targeted analysis of protein modifications such as acetylation and phosphorylation.[145] Size exclusion chromatography separates out proteins by size and is normally coupled to MS instruments for the analysis of large biomolecules.[146, 147] Size exclusion chromatography has been found to efficiently desalts solutions of large biomolecules making it easier to ionize with subsequent mass spectrometry. These chromatography techniques have proven useful to effectively characterize complex samples and efficiently analyze large biomolecules.

Capillary zone electrophoresis has recently been found to efficiently separate out proteins and peptides before MS analysis. This technique separates proteins out based on their electrostatic interaction with a capillary.[148, 149] Capillary zone electrophoresis has been shown to separate out peptides for glycopeptide analysis.[150] This technique has also been utilized to obtain a mass measurement for intact proteins[151] and to perform top-down fragmentation on intact proteins.[152] The benefits of performing these techniques is the sensitive nature of the technique and the speed with which the technique can be performed. Like other chromatography techniques, Capillary zone electrophoresis can significantly reduce the complexity of a sample which allows for more complete analysis of the proteins in the sample.

Many times, combinations of these techniques can be utilized to obtain more separation. In one example, strong cation exchange separation with subsequent reverse-phase liquid chromatography

separation was utilized to separate out peptides of the yeast proteome.[153] In another example, size exclusion chromatography in combination with capillary zone electrophoresis was utilized to separate native proteins from a lysate of Ecoli before being analyzed with MS.[154] In this study they were able to identify 144 proteins and 23 protein complexes in the sample. Combination of separation techniques can lead to more separation and a greater number of proteins and peptides identified. However, the drawback of combining separation techniques is the speed of analysis decreases considerably.

GAS-PHASE SEPARATION OF PEPTIDES AND PROTEINS

It is also possible to separate ions in the gas-phase in with a technique known as ion mobility (IM). Ion mobility separates ions by their overall shape known as the collisional cross section (CCS) in a pressurized cell before analyzing those ions by MS. It is possible to perform ion mobility combined with mass spectrometry due to the relatively lengthy time scale of ion mobility compared with MS analysis. IM-MS was first developed in the early 1960s to study ion-molecule reactions in the gas-phase.[155] In 1995, a study by Clemmer et. al. found that protein conformations could be resolved when taking IM-MS of the 7+ charge state of cytochrome C.[156] Since then, various ion mobility techniques have been developed and utilized for the analysis of peptide and protein characterization in the gas-phase.

There are three main types of ion mobility that can be utilized to separate gas-phase peptide and protein ions. (Fig. 3) Drift tube ion mobility spectrometry (DTIMS) utilizes a uniform electric field to pull ions through a region with neutral gas molecules.[157, 158] Smaller ions will travel through the gas region at a faster rate than larger molecules due to less collisions with neutral gas molecules. The time it takes for ions to tumble through the neutral gas corresponds to its CCS and their relationship is described by the Mason-Schamp equation[159] which was developed to explain the mobility of ions in a weak electric field. This type of IM-MS can separate ions with high resolution; however, the duty cycle of performing this type of analysis is quite long.[159] Traveling wave ion mobility spectrometry (TWIMS) guides ions through a neutral gas by applying a DC voltage along ring electrodes while a RF frequency confines the ions in the region.[157, 160] The pulsed wave pushes the ions through the ion mobility cell allowing ions to be separated by their CCS. It is possible to determine the exact CCS of ions with TWIMS, although, calibration of the TWIMS device is

needed before measurement.[160] TWIMS also suffers from low resolving power and potential heating of ions injected into the cell.[157] The last type of mobility separation of gas-phase ions is known as field asymmetric ion mobility spectrometry (FAIMS). FAIMS operates by changing the path of ions with an electric field as a neutral gas carries ions through the regions with changing electric fields. [161, 162] The changing electric fields will manipulate the path of the ions which separates out ions by CCS. FAIMS can separate ions with relatively high resolving power; however, the device reduces the amount of ions that reach the detector decreasing sensitivity.[157] In addition, the precise CCS value can not be determined with FAIMS.[157]

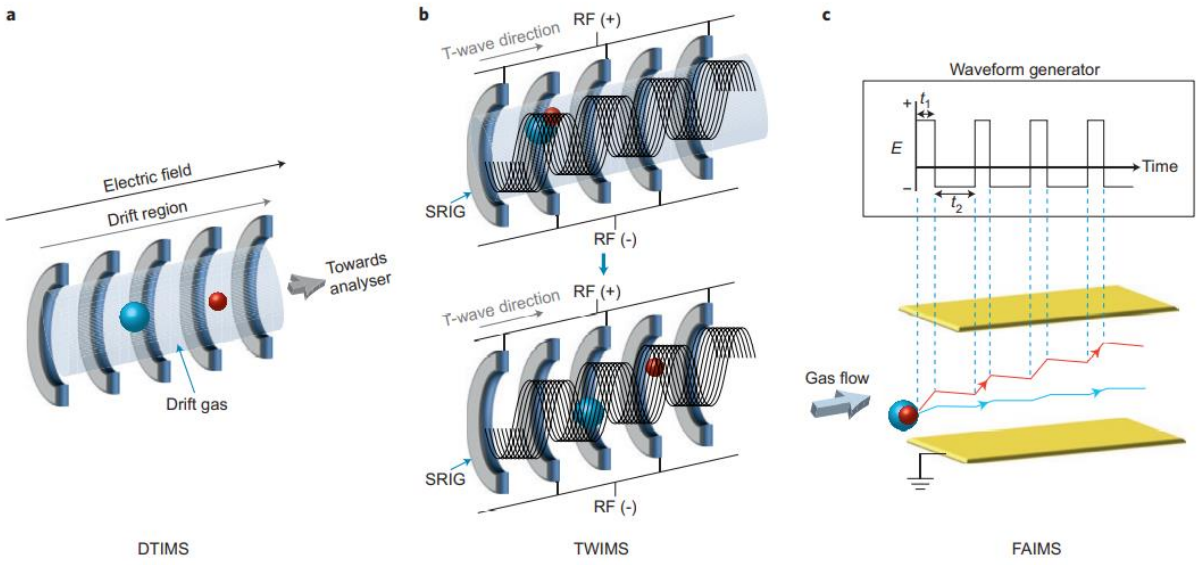


Figure 3: Ion mobility techniques and their general principles including a.) drift time ion mobility, b.) traveling wave ion mobility, and c.) field asymmetric ion mobility.[157]

Ion mobility has been utilized to separate out peptide and proteins for various applications. FAIMS is often utilized in bottom-up proteomics workflows to reduce background noise in the mass spectrum for more accurate analysis of peptide digests.[163] DTIMS and TWIMS have been utilized to separate out peptides due to modifications on the protein sequence. For example, phosphorylated peptides have been found to have a different mobility than their non-phosphorylated counterparts allowing for better characterization of the peptides.[164] In addition, ETD fragments from a peptide bound with a metallic drug could be separated

based on whether the fragments contained that drug.[165] Ion mobility analysis have also been utilized to calculate the CCS of protein ions in the gas-phase giving information on the size of protein in the gas-phase.[166] [167] Ion mobility analysis can also be utilized to separate out subcomplexes protein complexes as many times overlapping charge states make it difficult to assign peaks.[102] Collision induced unfolding (CIU) of proteins was developed to determine regions of relative structure of a protein monomer. With this technique proteins are activated with collisional energy and ion mobility of the ions is take after activation.[168] From this data, regions of relative order can become unfolded which is reflected in the mobiligram. Top-down mass spectrometry along with ion mobility can reveal regions of protein complexes that are unfolded when activated.[126] Ion mobility analysis have become an integral part of peptide and protein analysis with mass spectrometry providing us with relevant information on protein identification and analysis of protein structure.

PEPTIDE AND PROTEIN DISSOCIATION TECHNIQUIES

Mass spectrometers can be modified to obtain additional information on peptides and proteins. Some instrument modifications aim to dissociate peptide bonds resulting in the protein being fragmented into daughter ions.[169] The mass of these daughter ions are measured by the mass analyzer, and can be mapped back to the protein sequence. Fragments containing the N-terminus can either be *a*, *b*, or *c* fragment ions, and fragments that contain the C-terminus can either be *x*, *y*, or *z* fragment ions.[170] In addition, internal fragment ions which contain neither terminus, can also form if two cleavages occur.[171] Since the early days of protein MS, various methods have been developed utilizing collision-, electron-, and photon-based fragmentation techniques to reveal information on the protein sequence.

Collision techniques such as collisionally activated dissociation (CAD) and surface induced dissociation (SID) are common collision-based techniques. CAD increases the internal energy of a protein ion by colliding the peptide or protein of interest with numerous neutral molecules (e.g., nitrogen, argon, and xenon) which eventually break peptide bonds.[172] This technique is quite efficient at fragmenting proteins, although the timescale for fragmentation is quite long which tends to unfold protein ions before fragmenting them.[172] This technique is commonly used for protein and peptide sequencing and locating of

covalent modifications.[173] High energy C-trap dissociation (HCD), a collisionally activated technique that occurs on a faster timescale with higher energy, has also been utilized for these purposes as well.[174] CAD of protein complexes tends to release highly charged monomer products in addition to covalent products.[175] SID collides proteins with a surface to obtain information about proteins.[176] SID happens on a much faster timescale than CAD because dissociation is obtained with 1 collision.[177] SID has been found to release some *b*- and *y*- ions; however, its intended use is to symmetrically dissociate protein complexes for the determination of complex connectivity.[178, 179] Collision based techniques have proven a useful tool for protein sequencing and determination of protein complex quaternary structure.

TD-MS can also be performed with electron-based techniques including electron capture dissociation (ECD) and electron transfer dissociation (ETD). Electron capture dissociation (ECD) was developed in 1998 and is performed by placing protein ions in the path of an electron beam.[180] The electrons are captured and cause a reaction that dissociates peptide bonds.[181] Because the electron beam needs to be focused to interact with protein ions ECD has historically been performed on FT-ICR instruments; however, ECD has recently been designed to perform on orbitrap instruments as well.[182] Electron ionization dissociation (EID) is a variant of ECD that utilized high energy electrons (25V-30V) to fragment proteins. In some circumstances EID has been found to obtain higher sequence coverage of proteins.[183] Electron transfer dissociation (ETD) is performed by transferring an electron from a donating compound to a protein facilitating covalent cleavage.[184] These electron-based techniques are known as soft fragmentation techniques because they can dissociate covalent bonds without excessive heating of protein ions.[185] This feature makes electron-based techniques ideal for obtaining information from native protein complexes[186] and protein ligand complexes.[80] Electron-based techniques are ideal for obtaining sequence as well as tertiary and quaternary structure information on protein and protein complexes.

Additionally, TD-MS can also be performed with photon-based techniques including infrared multiple-photon dissociation (IRMPD) and ultraviolet photodissociation (UVPD). IRMPD utilizes low frequency photons to slowly heat proteins until peptide bonds break[187] yielding similar products as CAD.[185] In addition, IRMPD has also been utilized to dissociate fragments that are noncovalently attached to

proteins due to hydrogen bond or salt bridge interactions.[186] Ultraviolet photodissociation (UVPD) utilizes photons in the ultraviolet range (e.g., 157nm,[188] 213nm,[189] and 266nm[190]) of the electromagnetic spectrum to break peptide bonds. UVPD has been found to efficiently fragment peptides[191, 192] and large proteins,[133, 193] and in some cases they can reveal the location of covalent[194] and non-covalent modifications[195] on proteins. Other uses for UVPD includes the determination of isomeric residues on peptides.[196] These dissociation techniques are frequently utilized to perform proteomics as characterizing proteins with them can provide information about the sequence and structure of a protein.

BOTTOM-UP VS. TOP-DOWN PROTEOMICS

Over the past few decades, bottom-up proteomics (BUP) has proven a useful technique for the identification of proteins and the location of covalent modifications.[197] BUP involves digestion of all proteins in a sample with cleavage enzymes and performing liquid chromatography on the peptides before the peptides are analyzed with MS.[139, 198] This type of analysis can determine proteins present in cell lysates[153] and to interrogate PTMs on a specific protein of interest.[140] To further sequence peptides when they elute off the column, BUP mainly utilizes CAD and HCD[199] however other techniques such as UVPD[200] have also been implemented into this workflow. Although BUP can provide information on a wide variety of subjects there are some drawbacks to bottom-up proteomics including incomplete digestion and chemicals that are not suitable for MS which cause the loss of some protein and PTM information.[201] Nevertheless, BUP has proven a useful tool for the study of proteins.

Analysis of BUP data has been developed for the analysis of large datasets. Tools such as proteome discoverer,[202] mascot,[203] skyline,[204] among others will sift through peptide spectra and match their sequence to a specific protein in a database. The peptide match indicates that that protein was present in the sample. BUP can also search for covalent modifications such as PTMs on peptide sequences and some workflows have been implemented to determine the amount of modified proteins in a sample.[205] These software packages also calculate statistics to measure the validity of a match which are based off the quality of the match and the number of sequences in a database.[203] These software packages have made analysis of large

proteomics datasets efficient over the past few decades and have provides us with much data on complex protein mixtures.

On the other hand, top-down mass spectrometry forgoes the digestion process and aims to directly fragment the intact protein in the gas-phase.[206] In this workflow intact protein ions are ionized with either ESI[207] or MALDI[208] and those intact ions are fragmented with collision-, electron-, or photon-based fragmentation techniques. The resulting daughter ions from the intact protein or protein complex are measured by the mass analyzer and can be mapped back to regions of the protein sequence. This analysis is normally aimed at the characterization of a single protein species and its proteoforms;[209, 210] however, top down proteomics of samples with multiple species has also been attempted.[154, 211] TD-MS has no theoretical limit on its sequencing capability, although, with the current instrumentation, the soft limit for TD-MS is 30kDa.[212] Recently, there have been attempts to obtain sequence coverage from larger proteins with techniques such as UVPD.[193] TD-MS has been found to be useful for returning information on the sequence of intact proteins and locating modifications along the protein sequence without additional processing of digestion of the proteins.

Analysis of TD-MS data is less developed than analysis of BUP data; however, there have been great strides in data analysis over the past couple of years. TD-MS data analysis is split into two steps including a deconvolution step and an assignment step. Deconvolution involves the deisotoping of fragment ions and consolidation of charge states into one mass. Various methods of deconvolution have been put fourth including MsDeconv,[213] FLASHDeconv,[214] and TopPIC.[215] In addition, there are software tools that assign MS deconvolution values to the protein sequence. These include PrositeLite,[216] PrositePTM,[217] and the recently developed ClipsMS.[218] These programs can also search for modifications on the protein sequence. Deconvolution and assignment of mass spectra has simplified analysis of MS data; however, there are still issues with false positive deconvolution as well as assignment. More research is needed to perfect and standardize this process.

NATIVE MASS SPECTROMETRY

Native mass spectrometry (nMS) involves the ionization and analysis of the physiologically relevant structure of proteins and protein complexes.[96] To obtain a native mass spectrum, the protein must be dissolved in a solution that maintains its native structure, gently ionized into the gas-phase, and minimally activated so the structure of the protein remains intact. To perform nMS proteins are dissolved in a solution at neutral pH and that solvent must be volatile enough so the protein can be desolvated during the ionization process. This is normally done with 5mM-1M ammonium acetate which has been found to have a buffering capacity at pH 7 and be volatile enough to allow for the desolvation of proteins.[95] It has been found that ESI from ammonium acetate is able to ionize protein ions from water droplets and minimizes perturbation of the protein and protein complex structure.[219] These ions are transferred to the detector with minimal protein activation. Excessive activation can increase the internal energy of a protein or protein complex which could undermine the integrity of protein or protein complex structure. However, tunning the instrument to provide the least activation possible can preserve protein structure as well as weakly bound ligands that are also located on native proteins. This workflow analysis has become a valuable tool for the interrogation of intact proteins in their biologically active state.

Native mass spectrometry has been utilized to conserve the structure of monomer ions, the integrity of protein complexes, and the interaction between proteins and small ligands. The first reports of native MS being utilized to preserve non-covelent complexes were done with heme binding of myoglobin[220] and the macrolide FK506 1 binding to human FKBP.[221] These studies indicated that non-covalent interactions can survive the electrospray ionization process into the gas-phase and be transferred to a detector. Since then, it has been found that nMS can conserve other non-covalent protein ligand interactions including the interaction between a small aggregation inhibiting compound with amyloid proteins[77] and a zinc metal ion bound to carbonic anhydrase.[222] Furthermore, it has been found that nMS can conserve quaternary structure of non-covalent protein complexes including the alcohol dehydrogenase tetramer (150kDa),[223] the GroEL 14mer (800kDa),[224] and even heterogenous megadalton viral capsids (1.2MDa).[225] Native MS of protein

complexes has been found to reveal various ligands bound to proteins and reveal the makeup of protein complexes.

Over the past few decades there has been attempts to deconvolute intact native mass spectra. Deconvolution of native mass spectra involves the consolidation of protein charge states to a neutral mass peak. This can give the exact molecular weight of the protein and their proteoforms in a native mass spectrum. A couple of algorithms have been developed to deconvolute native mass spectra including UniDec,[226] ReSpect, and PMI.[227] These algorithms can subtract excessive baseline noise and consolidate all charge states to a single mass. There have been attempts to deconvolute all types of spectra including spectra of ADH,[228] membrane protein lipid complexes,[229] and intact antibodies.[227] These algorithms have been found to provide accurate molecular weights for native compounds in mass spectra.

COMPLEX-DOWN FRAGMENTATION OF NONCOVELENT PROTEIN COMPLEXES

Native mass spectrometry has been found to reveal overall protein or protein complex mass and the identify of modifications on the native protein; however, for some applications it is necessary to obtain sequence information of native protein complexes. Fragmentation of denatured proteins can sequence proteins and locate covalent modifications; however, information about protein structure and information on non-covalent protein/ligand interactions is lost. To obtain select information on protein structure and noncovalent interactions, MS^3 analysis of native complexes is often performed. This workflow is known as "complex-down"[197] MS or "nativeomics"[230] and has been found to provide a greater understanding of protein complex function.

The first step of these experiments is to transfer an intact native protein complex into the gas-phase with minimal disruption of protein structure. Next the internal energy of the protein complex is increased (normally with CAD or SID) resulting in the release of protein monomers or non-covalent ligands from the protein complex. In the case of CAD, the monomer ions accumulate more charges relative to the number of charges on each monomer in the intact complex. These monomers are subsequently isolated and fragmented to reveal covalent fragments. This workflow has be utilized to fragment large protein complexes including the GroEL 14mer[231] and a virus-like nanocontainer.[232] In these examples, complex monomers

could be fragmented and sequence information could be obtained after they had been ejected from the complex. Complex-down analysis of protein complexes can reveal relevant information about protein complexes including modification and structure information.

Complex-down analysis can reveal abnormalities such as mutations and deletions as well as modifications on protein complexes. In one study, complex-down fragmentation of the 20S proteosome found that the N-terminus was acetylated and the last two amino acids on the C-terminus were absent.[102] This data complemented cryo-EM data which could not resolve the N- or C-terminus of the protein. Complex-down analysis can also pinpoint modification sites that serve functional roles. It was found that different proteoforms of the triosephosphate isomerase complex have different affinities for dimerization.[233] A proteoform phosphorylated at serine 20 and a proteoform that was N-terminally acetylated inhibited dimerization of the complex, but the proteoform that was not modified did dimerize. In another example, complex-down analysis showed that phosphorylation of fructose-1,6-bisphosphatase 1 in yeast differed depending on the media that the yeast was grown in.[234] The data showed that Ser12 or Thr13 was highly phosphorylated when grown on glucose. Ser12 is known to deactivate the complex, this data indicates that when grown on glucose, the cells switch from gluconeogenesis to glycolysis. MS^3 analysis was also able to determine the identity of peptide ligands bound to a human E3 ligase.[235] Complex-down has been shown to provide useful information on protein complexes and reveal key regulatory modifications that modulate protein function.

Lastly, complex-down analysis has been found to study membrane proteins. Membrane proteins are normally dissolved in solutions with detergents or lipids so that their native structure can be preserved. During the ESI processes those molecules stick to the protein after ionization. Because detergents stick to the protein, it is necessary to provide activation to remove those detergents from the protein complex.[236] Increasing the activation of membrane protein complexes can result in the ejection of intact membrane protein complexes or membrane protein monomers. It was found that complex down with UVPD fragmentation increased sequence coverage by 21% compared to native fragmentation of the tetramer.[237] In another case, complex-down fragmentation with CAD of the mechanosensitive ion channel of large

conductance (MscL) was found to release *b*- and *y*- ions that stem from transmembrane helix regions of the protein.[238] This example indicates that structure information can be obtained with this workflow in some instances. These studies indicate that complex down analysis can reveal sequence and structure information for various membrane protein complexes.

NATIVE FRAGMENTATION OF PROTEINS AND PROTEIN COMPLEXES

The complex-down workflow disassembles the complex before covalent bonds are cleaved. However, to obtain information on the native structure of protein complexes, direct fragmentation of proteins or complexes in their native state must be performed. In this workflow, known as native top-down mass spectrometry (nTD-MS), the least amount of activation is given to the protein complex until the protein is ready for fragmentation. nTD-MS is normally performed with ExD and UVPD fragmentation techniques as these techniques have been shown to release covalent products that reflect higher-order structure characteristics during fragmentation. Data from this workflow can reveal the locations of covalent and non-covalent modifications along the protein sequence and reveal aspects of higher-order structure.

nTD-MS can determine the location of non-covalent ligands present on protein and protein complexes. Covalent fragments that result from nTD-MS can retain the ligand that bound to the native protein complex. In 2006 it was found that ECD of the α-synuclein/spermine complex revealed fragments that contained the spermine ligand and that the ligand could be localized to the C-terminus of the protein.[239] A couple of years later it was found the drug cisplatin could be located on a peptide with ETD fragmentation.[165] Since then ECD has been utilized to localize NAD^+ on alcohol dehydrogenase[133] and a small aggregation inhibiting compound on amyloid proteins such as α-synuclein[240] and tau.[80] Similarly, UVPD fragmentation has been found to determine the location of heme on myoglobin[241] and GTP on eIF4E.[242] Metal ions can also be localized along a protein sequence. Early studies utilized CAD to localize metal ions;[243] however, it is possible that fragmentation with CAD can eject metal ions from the binding site. More recently, ECD has been found to localize zinc metal ions on carbonic anhydrase[222] and the alcohol dehydrogenase tetramer.[133] In addition, it was found that UVPD is able to localize metal ions on various metaloproteins[195] and even an insulin pentamer.[242] Information from nTD-MS can be utilized to infer ligand

binding effects. For example, localization of CLR01 along the sequence of tau with nTD-MS revealed that CLR01 binding in the microtubule binding region could disrupt the interaction between monomers and prevent their aggregation into oligomers and fibrils.[80] In another example, localization of CLR01 binding to α-synuclein could be utilized to guide a molecular dynamic simulation indicating that CLR01 shifts the structure of α-synuclein.[240] nTD-MS of protein/ligand complexes can utilized to reveal the location of ligands and that information can be used to determine the function of those ligands.

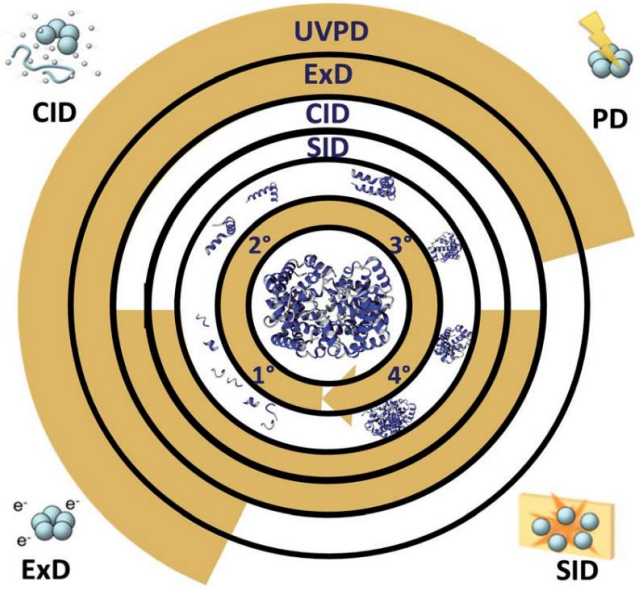


Figure 4: A graphic revealing the levels of protein structure accessible by the different top-down mass spectrometry methods. The circles represent the fragmentation techniques and gold filling represents the levels that the technique is reported to probe.[99]

nTD-MS has also been found to reveal higher-order structure of proteins and protein complexes. (Fig. 4) It is known that interactions besides covalent peptide bonds such as salt bridges[244] and disulfide bonds[245] can hold protein fragments together even after covalent bonds have been dissociated. This phenomenon is known as ExnoD. These regions of relative stability and instability can be probed by nTDMS fragmentation. Early experiments utilized ECD to probe changes in the structure of ubiquitin[246] and cytochrome C[247] due to an increase in source temperature and infrared activation. ExD of protein complexes reveals fragments that reflect the structure of the intact complex. For example, ExD of the ADH homotetramer revealed numerous N-terminal fragments and few C-terminal fragments.[248, 249] The relative

amount of solvent exposure of the N-terminus compared to the C-terminus may explain the greater number of fragments containing the N-terminus in the spectrum. Furthermore, nTD-MS of protein complexes such as hemoglobin[250] and glutamate dehydrogenase[186] reveals regions of stability of those protein complexes. UVPD has also been found to preferentially fragment more flexible regions of protein complexes;[242, 251] however, the vibrational excitement of the protein allows for deeper sequence coverage into the protein. A unique feature of UVPD fragmentation is that a+1 and a+2 ions can be utilized to probe secondary structure of proteins. For example, these ions have been found to occur in proteins with hydrogen bonding motifs[252] and turn structures.[251] nTDMS can probe structures of proteins and protein complexes which can give information on the structure and function of protein complexes.

Lastly, nTDMS can locate interfaces between peptides and proteins. It was found that ExD techniques could be utilized to determine the binding sites between small acidic and basic peptides[253] and a hexapeptide on a protein/peptide complex.[254] CAD has also been used to locate the binding site of RNA on proteins.[255] The electrostatic interactions holding the RNA and protein together were able to keep the RNA from dissociating from the complex during CAD fragmentation. In another example, ECD of amylin oligomers revealed the interface holding monomers of the amyloid protein together was located on the C-terminus.[256] These results indicate that nTD-MS is able to reveal information on important interactions between proteins and peptides.

ION MOBILITY UTILIZED FOR NATIVE PROTEIN ANALYSIS

Ion mobility analysis has been shown to provide useful information on native proteins and protein complexes. IM-MS can determine the relative size of proteins and protein complexes in the gas-phase.[166, 257] The gas-phase structures of the ions would presumably differ slightly from their solution phase structures; however, this generalized size information can give us an estimation of the size of the protein or protein complex.[257] IM-MS has also been found to separate subcomplexes of protein complexes and oligomer states of proteins. Using IM to separate subcomplexes from their intact precursors is especially helpful when dissociating complexes with SID, as each species can be separated by IM for more accurate quantification.[102] In addition, it has been found that IM-MS of amyloid protein oligomers can separate out

oligomeric states for accurate analysis of oligomer size.[258, 259] This information can reveal possible mechanisms of amyloid protein aggregation.[260] There have been multiple studies that find that small molecule ligands and metal ions are able to shift protein distributions. Recently, zinc ions have been found to shift the structure of transthyretin tetramer[261] and manganese and cobalt ions have been found to shift the structure of the amyloid protein α-synuclein.[75] In addition, it has been found that a conformational shift of the c-MYC:MAX complex is observed when 10058-F4 compound is added[262] and a conformation shift of α-synuclein is observed when small molecules such as EGCG[81] and CLR01[80] bind to the protein. This analysis can give valuable information on the size of protein ions and changes in structure due to ligand binding.

In some cases, IM has been utilized in combination with TD-MS to study native protein complexes. An early study utilized FAIMS to determine that ECD fragmentation was more efficient when the ubiquitin structure was more extended.[263] A couple other studies found that IM-MS with subsequent ECD can probe structural dynamics of lymphotactin.[264, 265] IM-MS of hemoglobin revealed that activation of the hemoglobin tetramer complex showed more fragmentation coming from the B-chain indicating that the B-chain was becoming unfolded.[126] It was also found that IM-MS can reveal ETnoD can be utilized to study structured regions of proteins such as ubiquitin.[266] Activation post IM revealed that fragments are retained on protein complexes after a covalent bond was broken with ETD. This can also be performed with protein complexes such as ADH, hemoglobin, and concanavalin A.[250, 267] nTD-MS fragmentation along with IM-MS has been able to reveal more information about protein complexes and probe their structure for ordered regions. This has given us valuable information on the dynamics of protein complexes and the regions that formed their ordered structure.

MASS SPECTROMETRY ANALYSIS OF AMYLOID PROTEINS

Protein misfolding and aggregation has been of significant interest over the past century. These amyloid proteins have been found to create aggregates in the brain and have been studied through a variety of techniques including Cryo-electron microscopy, X-ray crystallography, and NMR. Mass spectrometry has also been utilized for the studying various aspects of amyloid proteins. In many cases, MS techniques

can provide relevant information on amyloid protein monomers and their proteoforms as well as amyloid protein oligomers.

MS has been utilized for the determination of amyloid protein identify and the location of PTMs. As was discussed previously, PTMs of amyloid proteins can trigger aggregation. BUP has been found to be an invaluable resource for determining the location of modifications sites on amyloid proteins. In one study tau protein from cerebrospinal fluid and soluble brain fractions was digested and the phosphorylation landscape was mapped over the entire sequence.[268] This study gave valuable insight into the location of phosphorylation sites common to Alzheimer's disease patients. Intact protein mass spectrometry has also been utilized to categorize the proteoforms of amyloid proteins in solution. Intact protein mass spectrometry has been utilized to determine the extent of acetylation[269] and phosphorylation [270] of α-synuclein and phosphorylation of tau.[80] Top-down mass spectrometry can be utilized to locate PTMs such as acetylation,[269] and phosphorylation[80] on these proteins. These analyses can give insight into mechanisms that PTMs utilized to alter aggregation potentials of amyloid proteins.

MS has also been utilized to characterize the binding of non-covalent modifications that modulate aggregation including metal ions and aggregation modulating compounds. It has been found that native mass spectrometry can determine the stoichiometry of metal ions. nMS has categorized the binding of manganese and cobalt to α-synuclein.[75] In another study calcium was also found to bind α-synuclein and modulate the aggregation of various proteoforms.[271] nMS has also determined the binding of small molecules to amyloid proteins. MS was able to characterize the binding of EGCG, a compound most associated with green tea, to α-synuclein.[81] In addition, MS was also utilized to characterize the binding of CLR01, a molecular tweezer compound, to tau,[80] α-synuclein,[240] amyloid-β,[77] SOD1,[272] and the N17 domain of huntingtin.[273] These small compounds tend to bind amyloid protein with a K_d in the micromolar range. Top-down mass spectrometry can locate these modifications on the protein sequence. To determine the location of these noncovalent modifications along the protein sequence, electron-based techniques are normally utilized so the interaction between the compound and the protein is not disturbed. ExD techniques have been utilized to determine the binding location of metal ions such as manganese and cobalt on α-

synuclein. [75] In addition, ExD techniques have also been utilized to determine to binding location of small molecules such as CLR01 on tau,[80] α-synuclein,[240] amyloid-β,[77] SOD1,[272] and the N17 domain of huntingtin[273] and EGCG on α-synuclein.[81]

MS techniques have also been utilized to determine the size of oligomers of amyloid proteins. It is thought that oligomers of amyloid proteins become toxic by interesting themselves into membranes and disrupting other cellular processes; however, due to their transitory nature they are difficult to study. Mass spectrometry is uniquely qualified for measuring the molecular weight of these oligomers, but the heterogeneity of the oligomer samples complicates oligomer analysis. Most amyloid protein oligomer research has been performed with ion mobility from the Mike Bowers lab. By measuring collisional cross section of peaks of amyloid proteins, they can determine oligomeric states. They have demonstrated that IM-MS is able to detect various oligomeric states of amyloid-β[259] human IAPP peptides[274] and TDP-43 mutant peptides.[275] This information can provide information on the size of toxic amyloid protein species and provide mechanisms for amyloid protein aggregation. If the oligomers are small, they can be characterized directly with mass spectrometry. This has been done with α-synuclein dimers,[276] tau dimers,[277] as well as small IAPP oligomers.[256] This analysis has proven useful for the identification of toxic species, however, better instrumentation is necessary for accurate characterization of larger amyloid protein oligomers.

Lastly, mass spectrometry can determine the binding interface of amyloid protein oligomers. This analysis follows the same workflow as nTD-MS of protein complexes described earlier. The fragments that result from TD-MS can be correlated back to the protein sequence and inferences can be made based off the fragmentation pattern. Direct nTD-MS on amyloid protein oligomers has been reported once on small oligomers of amylin. In this experiment amylin oligomers were fragmented with ECD.[256] The resulting fragments that were mapped to the sequence showed the aggregation interface to the on the C-terminus of the amylin peptide. This technique has not been utilized extensively; however, it does show promise for the determination of amyloid protein oligomer interfaces and could provide information on these toxic species in neurodegenerative diseases.

REFERENCES

1. Ashmore, J.; Weber, G., The role of hepatic glucose-6-phosphatase in the regulation of carbohydrate metabolism. In *Vitamins & Hormones*, Elsevier: 1959; Vol. 17, pp 91-132.

2. Ginsburg, V.; Hers, H., On the conversion of fructose to glucose by guinea pig intestine. *Biochimica et biophysica acta* **1960,** *38*, 427-434.

3. Crick, F. H. In *On protein synthesis*, Symp Soc Exp Biol, 1958; p 8.

4. Crick, F., Central dogma of molecular biology. *Nature* **1970,** *227* (5258), 561-563.

5. Joosten, R. P., *X-Ray structure re-refinement. Combining old data with new methods for better structural bioinformatics*. 2010. https://hdl.handle.net/2066/76549

6. Sanger, F.; Tuppy, H., The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochemical journal* **1951,** *49* (4), 463.

7. Sanger, F.; Thompson, E., The amino-acid sequence in the glycyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochemical Journal* **1953,** *53* (3), 353.

8. Levitt, M., Nature of the protein universe. *Proceedings of the National Academy of Sciences* **2009,** *106* (27), 11079-11084.

9. Fowler, A. V.; Zabin, I., Amino acid sequence of beta-galactosidase. XI. Peptide ordering procedures and the complete sequence. *Journal of Biological Chemistry* **1978,** *253* (15), 5521-5525.

10. Goldenberg, D. P.; King, J., Temperature-sensitive mutants blocked in the folding or subunit assembly of the bacteriophage P22 tail spike protein: II. Active mutant proteins matured at 30° C. *Journal of Molecular Biology* **1981,** *145* (4), 633-651.

11. Smith, D. H.; King, J., Temperature-sensitive mutants blocked in the folding or subunit assembly of the bacteriophage P22 tail spike protein: III. Inactive polypeptide chains synthesized at 39° C. *Journal of molecular biology* **1981,** *145* (4), 653-676.

12. Savage, D. F.; Egea, P. F.; Robles-Colmenares, Y.; III, J. D. O. C.; Stroud, R. M.; Simon, S., Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin Z. *PLoS biology* **2003,** *1* (3), e72.

13. Hardman, K. D.; Ainsworth, C. F., Structure of concanavalin A at 2.4-Ang resolution. *Biochemistry* **1972,** *11* (26), 4910-4919.

14. Newcomer, M. E.; Ong, D. E., Plasma retinol binding protein: structure and function of the prototypic lipocalin. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **2000,** *1482* (1-2), 57-64.

15. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R.; Wyckoff, H.; Phillips, D. C., A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **1958,** *181* (4610), 662-666.

16. Friedman, J. M., Structure, dynamics, and reactivity in hemoglobin. *Science* **1985,** *228* (4705), 1273-1280.

17. Ackers, G. K.; Doyle, M. L.; Myers, D.; Daugherty, M. A., Molecular code for cooperativity in hemoglobin. *Science* **1992,** *255* (5040), 54-63.

18. Patel, M. S.; Nemeria, N. S.; Furey, W.; Jordan, F., The pyruvate dehydrogenase complexes: structure-based function and regulation. *Journal of Biological Chemistry* **2014,** *289* (24), 16615-16623.

19. Edelstein, S., Patterns in the quinary structures of proteins. Plasticity and inequivalence of individual molecules in helical arrays of sickle cell hemoglobin and tubulin. *Biophysical Journal* **1980,** *32* (1), 347-360.

20. Cohen, R. D.; Pielak, G. J., Electrostatic contributions to protein quinary structure. *Journal of the American Chemical Society* **2016,** *138* (40), 13139-13142.

21. Breindel, L.; DeMott, C.; Burz, D. S.; Shekhtman, A., Real-time in-cell nuclear magnetic resonance: ribosome-targeted antibiotics modulate quinary protein interactions. *Biochemistry* **2018,** *57* (5), 540-546.

22. Kouzarides, T., Histone methylation in transcriptional control. *Current opinion in genetics & development* **2002,** *12* (2), 198-209.

23. Clarke, S., Aging as war between chemical and biochemical processes: protein methylation and the recognition of age-damaged proteins for repair. *Ageing research reviews* **2003,** *2* (3), 263-285.

24. Tang, Y.; Zhao, W.; Chen, Y.; Zhao, Y.; Gu, W., Acetylation is indispensable for p53 activation. *Cell* **2008,** *133* (4), 612-626.

25. Kamita, M.; Kimura, Y.; Ino, Y.; Kamp, R. M.; Polevoda, B.; Sherman, F.; Hirano, H., Nα-Acetylation of yeast ribosomal proteins and its effect on protein synthesis. *Journal of proteomics* **2011,** *74* (4), 431-441.

26. Goodsell, D. S., The molecular perspective: the ras oncogene. *Stem cells* **1999,** *17* (4), 235-236.

27. Lowy, D. R.; Willumsen, B. M., Function and regulation of ras. *Annual review of biochemistry* **1993,** *62* (1), 851-891.

28. Downward, J., Targeting RAS signalling pathways in cancer therapy. *Nature reviews cancer* **2003,** *3* (1), 11-22.

29. Yang, X.-J.; Seto, E., Lysine acetylation: codified crosstalk with other posttranslational modifications. *Molecular cell* **2008,** *31* (4), 449-461.

30. Kurash, J. K.; Lei, H.; Shen, Q.; Marston, W. L.; Granda, B. W.; Fan, H.; Wall, D.; Li, E.; Gaudet, F., Methylation of p53 by Set7/9 mediates p53 acetylation and activity in vivo. *Molecular cell* **2008,** *29* (3), 392-400.

31. Carlomagno, Y.; Chung, D.-e. C.; Yue, M.; Castanedes-Casey, M.; Madden, B. J.; Dunmore, J.; Tong, J.; DeTure, M.; Dickson, D. W.; Petrucelli, L., An acetylation–phosphorylation switch that regulates tau aggregation propensity and function. *Journal of Biological Chemistry* **2017,** *292* (37), 15277-15286.

32. Raj, S. B.; Ramaswamy, S.; Plapp, B. V., Yeast alcohol dehydrogenase structure and catalysis. *Biochemistry* **2014,** *53* (36), 5791-5803.

33. Temiz, N. A.; Bahar, I., Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase. *Proteins: Structure, Function, and Bioinformatics* **2002,** *49* (1), 61-70.

34. Sprang, S.; Acharya, K.; Goldsmith, E.; Stuart, D.; Varvill, K.; Fletterick, R.; Madsen, N.; Johnson, L., Structural changes in glycogen phosphorylase induced by phosphorylation. *Nature* **1988,** *336* (6196), 215-221.

35. Lee, Y.; Park, J.; Lee, G.; Yoon, S.; Min, C. K.; Kim, T. G.; Yamamoto, T.; Kim, D. H.; Lee, K. W.; Eom, S. H., S92 phosphorylation induces structural changes in the N-terminus domain of human mitochondrial calcium uniporter. *Scientific reports* **2020,** *10* (1), 1-13.

36. Eshun-Wilson, L.; Zhang, R.; Portran, D.; Nachury, M. V.; Toso, D. B.; Löhr, T.; Vendruscolo, M.; Bonomi, M.; Fraser, J. S.; Nogales, E., Effects of α-tubulin acetylation on microtubule structure and stability. *Proceedings of the National Academy of Sciences* **2019,** *116* (21), 10366-10371.

37. Malik, N.; Wahlbeck, K. A.; Thompson, L. K., Strategies for identifying dynamic regions in protein complexes: flexibility changes accompany methylation in chemotaxis receptor signaling states. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2020,** *1862* (9), 183312.

38. Berezovskaya, Y.; Porrini, M.; Barran, P. E., The effect of salt on the conformations of three model proteins is revealed by variable temperature ion mobility mass spectrometry. *International Journal of Mass Spectrometry* **2013,** *345*, 8-18.

39. Rees, D. C.; Williams, T. N.; Gladwin, M. T., Sickle-cell disease. *The Lancet* **2010,** *376* (9757), 2018-2031.

40. Cheng, S. H.; Gregory, R. J.; Marshall, J.; Paul, S.; Souza, D. W.; White, G. A.; O'Riordan, C. R.; Smith, A. E., Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* **1990,** *63* (4), 827-834.

41. Komander, D.; Rape, M., The ubiquitin code. *Annual review of biochemistry* **2012,** *81*, 203-229.

42. Settembre, C.; Fraldi, A.; Medina, D. L.; Ballabio, A., Signals from the lysosome: a control centre for cellular clearance and energy metabolism. *Nature reviews Molecular cell biology* **2013,** *14* (5), 283-296.

43. Platt, F. M.; Boland, B.; van der Spoel, A. C., Lysosomal storage disorders: The cellular impact of lysosomal dysfunction. *Journal of Cell Biology* **2012,** *199* (5), 723-734.

44. Soto, C.; Pritzkow, S., Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nature neuroscience* **2018,** *21* (10), 1332-1340.

45. Livingston, G.; Huntley, J.; Sommerlad, A.; Ames, D.; Ballard, C.; Banerjee, S.; Brayne, C.; Burns, A.; Cohen-Mansfield, J.; Cooper, C., Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet* **2020,** *396* (10248), 413-446.

46. Schroeder, P., Alois Alzheimer†. *European Neurology* **1916,** *39* (1), 61-64.

47. Malik, R.; Wiedau, M., Therapeutic approaches targeting protein aggregation in amyotrophic lateral sclerosis. *Frontiers in Molecular Neuroscience* **2020,** *13*, 98.

48. Kovacs, G. G., Invited review: neuropathology of tauopathies: principles and practice. *Neuropathology and applied neurobiology* **2015,** *41* (1), 3-23.

49. Weingarten, M. D.; Lockwood, A. H.; Hwo, S.-Y.; Kirschner, M. W., A protein factor essential for microtubule assembly. *Proceedings of the National Academy of Sciences* **1975,** *72* (5), 1858-1862.

50. Feinstein, H. E.; Benbow, S. J.; LaPointe, N. E.; Patel, N.; Ramachandran, S.; Do, T. D.; Gaylord, M. R.; Huskey, N. E.; Dressler, N.; Korff, M., Oligomerization of the microtubule-associated protein tau is mediated by its N-terminal sequences: implications for normal and pathological tau action. *Journal of neurochemistry* **2016,** *137* (6), 939-954.

51. Zhong, Q.; Congdon, E. E.; Nagaraja, H. N.; Kuret, J., Tau isoform composition influences rate and extent of filament formation. *Journal of Biological Chemistry* **2012,** *287* (24), 20711-20719.

52. Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J.; McFarlane, H. T., Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **2007,** *447* (7143), 453-457.

53. Fitzpatrick, A. W.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H., Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017,** *547* (7662), 185-190.

54. Gotz, J.; Chen, F. v.; Van Dorpe, J.; Nitsch, R., Formation of neurofibrillary tangles in P301L tau transgenic mice induced by Aβ42 fibrils. *Science* **2001,** *293* (5534), 1491-1495.

55. Peelaerts, W.; Bousset, L.; Van der Perren, A.; Moskalyuk, A.; Pulizzi, R.; Giugliano, M.; Van den Haute, C.; Melki, R.; Baekelandt, V., α-Synuclein strains cause distinct synucleinopathies after local and systemic administration. *Nature* **2015,** *522* (7556), 340-344.

56. Emamzadeh, F. N., Alpha-synuclein structure, functions, and interactions. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences* **2016,** *21*.

57. Diao, J.; Burré, J.; Vivona, S.; Cipriano, D. J.; Sharma, M.; Kyoung, M.; Südhof, T. C.; Brunger, A. T., Native α-synuclein induces clustering of synaptic-vesicle mimics via binding to phospholipids and synaptobrevin-2/VAMP2. *elife* **2013,** *2*, e00592.

58. Lautenschläger, J.; Stephens, A. D.; Fusco, G.; Ströhl, F.; Curry, N.; Zacharopoulou, M.; Michel, C. H.; Laine, R.; Nespovitaya, N.; Fantham, M., C-terminal calcium binding of α-synuclein modulates synaptic vesicle interaction. *Nature communications* **2018,** *9* (1), 1-13.

59. Rodriguez, J. A.; Ivanova, M. I.; Sawaya, M. R.; Cascio, D.; Reyes, F. E.; Shi, D.; Sangwan, S.; Guenther, E. L.; Johnson, L. M.; Zhang, M., Structure of the toxic core of α-synuclein from invisible crystals. *Nature* **2015,** *525* (7570), 486-490.

60. Li, B.; Ge, P.; Murray, K. A.; Sheth, P.; Zhang, M.; Nair, G.; Sawaya, M. R.; Shin, W. S.; Boyer, D. R.; Ye, S., Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nature communications* **2018,** *9* (1), 1-10.

61. Baba, M.; Nakajo, S.; Tu, P.-H.; Tomita, T.; Nakaya, K.; Lee, V.; Trojanowski, J. Q.; Iwatsubo, T., Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies. *The American journal of pathology* **1998,** *152* (4), 879.

62. Hampel, H.; Bürger, K.; Pruessner, J. C.; Zinkowski, R.; DeBernardis, J.; Kerkman, D.; Leinsinger, G.; Evans, A. C.; Davies, P.; Möller, H.-J., Correlation of cerebrospinal fluid levels of tau protein phosphorylated at threonine 231 with rates of hippocampal atrophy in Alzheimer disease. *Archives of neurology* **2005,** *62* (5), 770-773.

63. Xia, Y.; Bell, B. M.; Giasson, B. I., Tau K321/K353 pseudoacetylation within KXGS motifs regulates tau–microtubule interactions and inhibits aggregation. *Scientific reports* **2021,** *11* (1), 1-9.

64. Fujiwara, H.; Hasegawa, M.; Dohmae, N.; Kawashima, A.; Masliah, E.; Goldberg, M. S.; Shen, J.; Takio, K.; Iwatsubo, T., α-Synuclein is phosphorylated in synucleinopathy lesions. *Nature cell biology* **2002,** *4* (2), 160-164.

65. Kang, L.; Moriarty, G. M.; Woods, L. A.; Ashcroft, A. E.; Radford, S. E.; Baum, J., N-Terminal acetylation of α-synuclein induces increased transient helical propensity and decreased aggregation rates in the intrinsically disordered monomer. *Protein Science* **2012,** *21* (7), 911-917.

66. Wolfe, M. S., Tau mutations in neurodegenerative diseases. *Journal of Biological Chemistry* **2009,** *284* (10), 6021-6025.

67. Lázaro, D. F.; Rodrigues, E. F.; Langohr, R.; Shahpasandzadeh, H.; Ribeiro, T.; Guerreiro, P.; Gerhardt, E.; Kröhnert, K.; Klucken, J.; Pereira, M. D., Systematic comparison of the effects of alpha-synuclein mutations on its oligomerization and aggregation. *PLoS genetics* **2014,** *10* (11), e1004741.

68. Zheng, W.; Tsai, M.-Y.; Wolynes, P. G., Comparing the aggregation free energy landscapes of amyloid beta (1–42) and amyloid beta (1–40). *Journal of the American Chemical Society* **2017,** *139* (46), 16666-16676.

69. Crapper, D.; Krishnan, S.; Dalton, A., Brain aluminum distribution in Alzheimer's disease and experimental neurofibrillary degeneration. *Science* **1973,** *180* (4085), 511-513.

70. Dexter, D.; Wells, F.; Lee, A.; Agid, F.; Agid, Y.; Jenner, P.; Marsden, C., Increased nigral iron content and alterations in other metal ions occurring in brain in Parkinson's disease. *Journal of neurochemistry* **1989,** *52* (6), 1830-1836.

71. Kasarskis, E. J.; Tandon, L.; Lovell, M. A.; Ehmann, W. D., Aluminum, calcium, and iron in the spinal cord of patients with sporadic amyotrophic lateral sclerosis using laser microprobe mass spectroscopy: a preliminary study. *Journal of the neurological sciences* **1995,** *130* (2), 203-208.

72. Jomova, K.; Vondrakova, D.; Lawson, M.; Valko, M., Metals, oxidative stress and neurodegenerative disorders. *Molecular and cellular biochemistry* **2010,** *345* (1), 91-104.

73. Bush, A. I.; Curtain, C. C., Twenty years of metallo-neurobiology: where to now? *European Biophysics Journal* **2008,** *37* (3), 241-245.

74. Danielsson, J.; Pierattelli, R.; Banci, L.; Gräslund, A., High-resolution NMR studies of the zinc-binding site of the Alzheimer's amyloid β-peptide. *The FEBS journal* **2007,** *274* (1), 46-59.

75. Wongkongkathep, P.; Han, J. Y.; Choi, T. S.; Yin, S.; Kim, H. I.; Loo, J. A., Native top-down mass spectrometry and ion mobility MS for characterizing the cobalt and manganese metal binding of α-synuclein protein. *Journal of the American Society for Mass Spectrometry* **2018,** *29* (9), 1870-1880.

76. Choi, T. S.; Lee, J.; Han, J. Y.; Jung, B. C.; Wongkongkathep, P.; Loo, J. A.; Lee, M. J.; Kim, H. I., Supramolecular modulation of structural polymorphism in pathogenic α-synuclein fibrils using copper (II) coordination. *Angewandte Chemie International Edition* **2018,** *57* (12), 3099-3103.

77. Sinha, S.; Lopes, D. H.; Du, Z.; Pang, E. S.; Shanmugam, A.; Lomakin, A.; Talbiersky, P.; Tennstaedt, A.; McDaniel, K.; Bakshi, R., Lysine-specific molecular tweezers are broad-spectrum inhibitors of assembly and toxicity of amyloid proteins. *Journal of the American Chemical Society* **2011,** *133* (42), 16958-16969.

78. Ehrnhoefer, D. E.; Bieschke, J.; Boeddrich, A.; Herbst, M.; Masino, L.; Lurz, R.; Engemann, S.; Pastore, A.; Wanker, E. E., EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nature structural & molecular biology* **2008,** *15* (6), 558-566.

79. Li, J.; Zhu, M.; Rajamani, S.; Uversky, V. N.; Fink, A. L., Rifampicin inhibits α-synuclein fibrillation and disaggregates fibrils. *Chemistry & biology* **2004,** *11* (11), 1513-1521.

80. 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.

81. Liu, Y.; Ho, L. H.; Carver, J. A.; Pukala, T. L., Ion mobility mass spectrometry studies of the inhibition of alpha synuclein amyloid fibril formation by (–)-epigallocatechin-3-gallate. *Australian Journal of Chemistry* **2011,** *64* (1), 36-40.

82. Tjernberg, L. O.; Näslund, J.; Lindqvist, F.; Johansson, J.; Karlström, A. R.; Thyberg, J.; Terenius, L.; Nordstedt, C., Arrest of-Amyloid Fibril Formation by a Pentapeptide Ligand (∗). *Journal of Biological Chemistry* **1996,** *271* (15), 8545-8548.

83. Findeis, M. A.; Musso, G. M.; Arico-Muendel, C. C.; Benjamin, H. W.; Hundal, A. M.; Lee, J.-J.; Chin, J.; Kelley, M.; Wakefield, J.; Hayward, N. J., Modified-peptide inhibitors of amyloid β-peptide polymerization. *Biochemistry* **1999,** *38* (21), 6791-6800.

84. El-Agnaf, O. M.; Paleologou, K. E.; Greer, B.; Abogrein, A. M.; King, J. E.; Salem, S. A.; Fullwood, N. J.; Benson, F. E.; Hewitt, R.; Ford, K. J., A strategy for designing inhibitors of α-synuclein aggregation and toxicity as a novel treatment for Parkinson's disease and related disorders. *The FASEB journal* **2004,** *18* (11), 1315-1317.

85. Porat, Y.; Mazor, Y.; Efrat, S.; Gazit, E., Inhibition of islet amyloid polypeptide fibril formation: a potential role for heteroaromatic interactions. *Biochemistry* **2004,** *43* (45), 14454-14462.

86. Sievers, S. A.; Karanicolas, J.; Chang, H. W.; Zhao, A.; Jiang, L.; Zirafi, O.; Stevens, J. T.; Münch, J.; Baker, D.; Eisenberg, D., Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* **2011,** *475* (7354), 96-100.

87. Seidler, P.; Boyer, D.; Rodriguez, J.; Sawaya, M.; Cascio, D.; Murray, K.; Gonen, T.; Eisenberg, D., Structure-based inhibitors of tau aggregation. *Nature chemistry* **2018,** *10* (2), 170-176.

88. Abskharon, R.; Seidler, P. M.; Sawaya, M. R.; Cascio, D.; Yang, T. P.; Philipp, S.; Williams, C. K.; Newell, K. L.; Ghetti, B.; DeTure, M. A., Crystal structure of a conformational antibody that binds tau oligomers and inhibits pathological seeding by extracts from donors with Alzheimer's disease. *Journal of Biological Chemistry* **2020,** *295* (31), 10662-10676.

89. Gratuze, M.; Jiang, H.; Wang, C.; Xiong, M.; Bao, X.; Holtzman, D. M., APOE antibody inhibits Aβ-associated tau seeding and spreading in a mouse model. *Annals of Neurology* **2022**.

90. Lalli, G.; Schott, J. M.; Hardy, J.; De Strooper, B., Aducanumab: a new phase in therapeutic development for Alzheimer's disease? *EMBO Molecular Medicine* **2021,** *13* (8), e14781.

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

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

93. 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.

94. Domon, B.; Aebersold, R., Mass spectrometry and protein analysis. *science* **2006,** *312* (5771), 212-217.

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

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

97. Kelleher, N. L., Peer reviewed: top-down proteomics. *Analytical chemistry* **2004,** *76* (11), 196 A-203 A.

98. Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A., How many human proteoforms are there? *Nature chemical biology* **2018,** *14* (3), 206-214.

99. 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.

100. Kolkman, A.; Daran-Lapujade, P.; Fullaondo, A.; Olsthoorn, M. M.; Pronk, J. T.; Slijper, M.; Heck, A. J., Proteome analysis of yeast response to various nutrient limitations. *Molecular systems biology* **2006,** *2* (1), 2006.0026.

101. Yang, Y.; Barendregt, A.; Kamerling, J. P.; Heck, A. J., Analyzing protein micro-heterogeneity in chicken ovalbumin by high-resolution native mass spectrometry exposes qualitatively and semi-quantitatively 59 proteoforms. *Analytical chemistry* **2013,** *85* (24), 12037-12045.

102. 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 central science* **2020,** *6* (4), 573-588.

103. 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.

104. 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.

105. 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.

106. 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.

107. De La Mora, J. F., Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica chimica acta* **2000,** *406* (1), 93-104.

108. Hogan Jr, C. J.; 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.

109. 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.

110. 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.

111. 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.

112. 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.

113. Cohen, L. R.; 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.

114. 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.

115. 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.

116. 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.

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

118. Mathieu, É., Mémoire sur le mouvement vibratoire d'une membrane de forme elliptique. *Journal de mathématiques pures et appliquées* **1868,** *13*, 137-203.

119. March, R. E., An introduction to quadrupole ion trap mass spectrometry. *Journal of mass spectrometry* **1997,** *32* (4), 351-369.

120. Snijder, J.; Rose, R. J.; Veesler, D.; Johnson, J. E.; Heck, A. J., Studying 18 MDa virus assemblies with native mass spectrometry. *Angewandte Chemie International Edition* **2013,** *52* (14), 4020-4023.

121. Campuzano, I. D.; 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.

122. Mikhailov, V. A.; Liko, I.; Mize, T. H.; Bush, M. F.; Benesch, J. L.; Robinson, C. V., Infrared laser activation of soluble and membrane protein assemblies in the gas phase. *Analytical chemistry* **2016,** *88* (14), 7060-7067.

123. Wolff, M.; Stephens, W., A pulsed mass spectrometer with time dispersion. *Review of Scientific Instruments* **1953,** *24* (8), 616-617.

124. 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.

125. Wang, L.; Amphlett, G.; Lambert, J. M.; Blättler, W.; Zhang, W., Structural characterization of a recombinant monoclonal antibody by electrospray time-of-flight mass spectrometry. *Pharmaceutical research* **2005,** *22* (8), 1338-1349.

126. 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.

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

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

129. 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.

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

131. Kilgour, D. P.; Nagornov, K. O.; Kozhinov, A. N.; Zhurov, K. O.; Tsybin, Y. O., Producing absorption mode Fourier transform ion cyclotron resonance mass spectra with non-quadratic phase correction functions. *Rapid communications in mass spectrometry* **2015,** *29* (11), 1087-1093.

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

133. 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.

134. Shevchenko, A.; Tomas, H.; Havli, J.; Olsen, J. V.; Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols* **2006,** *1* (6), 2856-2860.

135. Xiao, H.; Zhang, L.; Zhou, H.; Lee, J. M.; Garon, E. B.; Wong, D. T., Proteomic analysis of human saliva from lung cancer patients using two-dimensional difference gel electrophoresis and mass spectrometry. *Molecular & Cellular Proteomics* **2012,** *11* (2).

136. Lohnes, K.; Quebbemann, N. R.; Liu, K.; Kobzeff, F.; Loo, J. A.; Loo, R. R. O., Combining high-throughput MALDI-TOF mass spectrometry and isoelectric focusing gel electrophoresis for virtual 2D gel-based proteomics. *Methods* **2016,** *104*, 163-169.

137. Takemori, A.; Butcher, D. S.; Harman, V. M.; Brownridge, P.; Shima, K.; Higo, D.; Ishizaki, J.; Hasegawa, H.; Suzuki, J.; Yamashita, M., PEPPI-MS: Polyacrylamide-gel-based prefractionation for analysis of intact proteoforms and protein complexes by mass spectrometry. *Journal of proteome research* **2020,** *19* (9), 3779-3791.

138. Ducret, A.; Oostveen, I. V.; Eng, J. K.; Yates III, J. R.; Aebersold, R., High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Science* **1998,** *7* (3), 706-719.

139. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates III, J. R., Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews* **2013,** *113* (4), 2343-2394.

140. Morris, M.; Knudsen, G. M.; Maeda, S.; Trinidad, J. C.; Ioanoviciu, A.; Burlingame, A. L.; Mucke, L., Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. *Nature neuroscience* **2015,** *18* (8), 1183-1189.

141. Karlsson, R.; Thorsell, A.; Gomila, M.; Salvà-Serra, F.; Jakobsson, H. E.; Gonzales-Siles, L.; Jaén-Luchoro, D.; Skovbjerg, S.; Fuchs, J.; Karlsson, A., Discovery of species-unique peptide biomarkers of bacterial pathogens by tandem mass spectrometry-based proteotyping. *Molecular & Cellular Proteomics* **2020,** *19* (3), 518-528.

142. Liu, K.; Singer, E.; Cohn, W.; Micewicz, E. D.; McBride, W. H.; Whitelegge, J. P.; Loo, J. A., Time-Dependent Measurement of Nrf2-Regulated Antioxidant Response to Ionizing Radiation Toward Identifying Potential Protein Biomarkers for Acute Radiation Injury. *PROTEOMICS–Clinical Applications* **2019,** *13* (6), 1900035.

143. Nshanian, M.; Lakshmanan, R.; Chen, H.; Loo, R. R. O.; Loo, J. A., Enhancing sensitivity of liquid chromatography–mass spectrometry of peptides and proteins using supercharging agents. *International journal of mass spectrometry* **2018,** *427*, 157-164.

144. Slebos, R. J.; Brock, J. W.; Winters, N. F.; Stuart, S. R.; Martinez, M. A.; Li, M.; Chambers, M. C.; Zimmerman, L. J.; Ham, A. J.; Tabb, D. L., Evaluation of strong cation exchange versus isoelectric focusing of peptides for multidimensional liquid chromatography-tandem mass spectrometry. *Journal of proteome research* **2008,** *7* (12), 5286-5294.

145. Mohammed, S.; Heck, A. J., Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. *Current opinion in biotechnology* **2011,** *22* (1), 9-16.

146. Lazar, A. C.; Wang, L.; Blättler, W. A.; Amphlett, G.; Lambert, J. M.; Zhang, W., Analysis of the composition of immunoconjugates using size-exclusion chromatography coupled to mass spectrometry. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* **2005,** *19* (13), 1806-1814.

147. Haberger, M.; Leiss, M.; Heidenreich, A.-K.; Pester, O.; Hafenmair, G.; Hook, M.; Bonnington, L.; Wegele, H.; Haindl, M.; Reusch, D. In *Rapid characterization of biotherapeutic proteins by size-exclusion chromatography coupled to native mass spectrometry*, MAbs, Taylor & Francis: 2016; pp 331-339.

148. Smith, R. D.; Olivares, J. A.; Nguyen, N. T.; Udseth, H. R., Capillary zone electrophoresis-mass spectrometry using an electrospray ionization interface. *analytical chemistry* **1988,** *60* (5), 436-441.

149. Smith, R. D.; Barinaga, C. J.; Udseth, H. R., Improved electrospray ionization interface for capillary zone electrophoresis-mass spectrometry. *Analytical chemistry* **1988,** *60* (18), 1948-1952.

150. Qu, Y.; Sun, L.; Zhang, Z.; Dovichi, N. J., Site-specific glycan heterogeneity characterization by hydrophilic interaction liquid chromatography solid-phase extraction, reversed-phase liquid chromatography fractionation, and capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry. *Analytical chemistry* **2018,** *90* (2), 1223-1233.

151. Loo, J. A.; Udseth, H. R.; Smith, R. D., Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry. *Analytical biochemistry* **1989,** *179* (2), 404-412.

152. Shen, X.; Yang, Z.; McCool, E. N.; Lubeckyj, R. A.; Chen, D.; Sun, L., Capillary zone electrophoresis-mass spectrometry for top-down proteomics. *TrAC Trends in Analytical Chemistry* **2019,** *120*, 115644.

153. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology* **2001,** *19* (3), 242-247.

154. Shen, X.; Kou, Q.; Guo, R.; Yang, Z.; Chen, D.; Liu, X.; Hong, H.; Sun, L., Native proteomics in discovery mode using size-exclusion chromatography–capillary zone electrophoresis–tandem mass spectrometry. *Analytical chemistry* **2018,** *90* (17), 10095-10099.

155. McDaniel, E.; Martin, D.; Barnes, W., Drift tube-mass spectrometer for studies of low-energy ion-molecule reactions. *Review of Scientific Instruments* **1962,** *33* (1), 2-7.

156. Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., Naked protein conformations: cytochrome c in the gas phase. *Journal of the American Chemical Society* **1995,** *117* (40), 10141-10142.

157. Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature chemistry* **2014,** *6* (4), 281-294.

158. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill Jr, H. H., Ion mobility–mass spectrometry. *Journal of mass spectrometry* **2008,** *43* (1), 1-22.

159. Mason, E. A.; Schamp Jr, H. W., Mobility of gaseous lons in weak electric fields. *Annals of physics* **1958,** *4* (3), 233-270.

160. Giles, K.; Williams, J. P.; Campuzano, I., Enhancements in travelling wave ion mobility resolution. *Rapid Communications in Mass Spectrometry* **2011,** *25* (11), 1559-1566.

161. Purves, R. W.; Guevremont, R., Electrospray ionization high-field asymmetric waveform ion mobility spectrometry− mass spectrometry. *Analytical chemistry* **1999,** *71* (13), 2346-2357.

162. Kolakowski, B. M.; Mester, Z., Review of applications of high-field asymmetric waveform ion mobility spectrometry (FAIMS) and differential mobility spectrometry (DMS). *Analyst* **2007,** *132* (9), 842-864.

163. Barnett, D. A.; Ells, B.; Guevremont, R.; Purves, R. W., Application of ESI-FAIMS-MS to the analysis of tryptic peptides. *Journal of the American Society for Mass Spectrometry* **2002,** *13* (11), 1282-1291.

164. Ruotolo, B. T.; Gillig, K. J.; Woods, A. S.; Egan, T. F.; Ugarov, M. V.; Schultz, J. A.; Russell, D. H., Analysis of phosphorylated peptides by ion mobility-mass spectrometry. *Analytical chemistry* **2004,** *76* (22), 6727-6733.

165. Williams, J. P.; Brown, J. M.; Campuzano, I.; Sadler, P. J., Identifying drug metallation sites on peptides using electron transfer dissociation (ETD), collision induced dissociation (CID) and ion mobility-mass spectrometry (IM-MS). *Chemical communications* **2010,** *46* (30), 5458-5460.

166. Ruotolo, B. T.; Benesch, J. L.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V., Ion mobility–mass spectrometry analysis of large protein complexes. *Nature protocols* **2008,** *3* (7), 1139-1152.

167. Myung, S.; Wiseman, J. M.; Valentine, S. J.; Takats, Z.; Cooks, R. G.; Clemmer, D. E., Coupling desorption electrospray ionization with ion mobility/mass spectrometry for analysis of protein structure: evidence for desorption of folded and denatured states. *The Journal of Physical Chemistry B* **2006,** *110* (10), 5045-5051.

168. Eschweiler, J. D.; Rabuck-Gibbons, J. N.; Tian, Y.; Ruotolo, B. T., CIUSuite: a quantitative analysis package for collision induced unfolding measurements of gas-phase protein ions. *Analytical chemistry* **2015,** *87* (22), 11516-11522.

169. Wysocki, V. H.; Resing, K. A.; Zhang, Q.; Cheng, G., Mass spectrometry of peptides and proteins. *Methods* **2005,** *35* (3), 211-222.

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

171. Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L., Analyzing internal fragmentation of electrosprayed ubiquitin ions during beam-type collisional dissociation. *Journal of the American Society for Mass Spectrometry* **2015,** *26* (5), 782-787.

172. Wells, J. M.; McLuckey, S. A., Collision-induced dissociation (CID) of peptides and proteins. *Methods in enzymology* **2005,** *402*, 148-185.

173. Smith, R. D.; Loa, J. A.; Barinaga, C. J.; Edmonds, C. G.; Udseth, H. R., Collisional activation and collision-activated dissociation of large multiply charged polypeptides and proteins produced by electrospray ionization. *Journal of the American Society for Mass Spectrometry* **1990,** *1* (1), 53-65.

174. 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.

175. Zhou, M.; Dagan, S.; Wysocki, V. H., Impact of charge state on gas-phase behaviors of noncovalent protein complexes in collision induced dissociation and surface induced dissociation. *Analyst* **2013,** *138* (5), 1353-1362.

176. Zhou, M.; Wysocki, V. H., Surface induced dissociation: dissecting noncovalent protein complexes in the gas phase. *Accounts of chemical research* **2014,** *47* (4), 1010-1018.

177. 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* **2018,** *91* (1), 190-209.

178. Quintyn, R. S.; Yan, J.; Wysocki, V. H., Surface-induced dissociation of homotetramers with D2 symmetry yields their assembly pathways and characterizes the effect of ligand binding. *Chemistry & biology* **2015,** *22* (5), 583-592.

179. Harvey, S. R.; Seffernick, J. T.; Quintyn, R. S.; Song, Y.; Ju, Y.; Yan, J.; Sahasrabuddhe, A. N.; Norris, A.; Zhou, M.; Behrman, E. J., Relative interfacial cleavage energetics of protein complexes revealed by surface collisions. *Proceedings of the National Academy of Sciences* **2019,** *116* (17), 8143-8148.

180. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *Journal of the American Chemical Society* **1998,** *120* (13), 3265-3266.

181. 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* (12), 5014-5030.

182. Shaw, J. B.; Malhan, N.; Vasil'ev, Y. V.; Lopez, N. I.; Makarov, A.; Beckman, J. S.; Voinov, V. G., Sequencing Grade Tandem Mass Spectrometry for Top–Down Proteomics Using Hybrid Electron Capture Dissociation Methods in a Benchtop Orbitrap Mass Spectrometer. *Analytical chemistry* **2018,** *90* (18), 10819-10827.

183. 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. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (9), 1896-1902.

184. Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences* **2004,** *101* (26), 9528-9533.

185. 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.

186. 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.

187. 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.

188. Fung, Y. E.; Kjeldsen, F.; Silivra, O. A.; Chan, T. D.; Zubarev, R. A., Facile disulfide bond cleavage in gaseous peptide and protein cations by ultraviolet photodissociation at 157 nm. *Angewandte Chemie International Edition* **2005,** *44* (39), 6399-6403.

189. 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* **2018,** *30* (1), 24-33.

190. 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.

191. Greer, S. M.; Bern, M.; Becker, C.; Brodbelt, J. S., Extending proteome coverage by combining MS/MS methods and a modified bioinformatics platform adapted for database searching of positive and negative polarity 193 nm ultraviolet photodissociation mass spectra. *Journal of proteome research* **2018,** *17* (4), 1340-1347.

192. 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.

193. Dunham, S. D.; Sanders, J. D.; Holden, D. D.; Brodbelt, J. S., Improving the Center Section Sequence Coverage of Large Proteins Using Stepped-Fragment Ion Protection Ultraviolet Photodissociation. *Journal of the American Society for Mass Spectrometry* **2022,** *33* (3), 446-456.

194. 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.

195. 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.

196. Van Orman, B. L.; Wu, H.-T.; Julian, R. R., Differentiation of peptide isomers by excited-state photodissociation and ion–molecule interactions. *Physical Chemistry Chemical Physics* **2020,** *22* (41), 23678-23685.

197. Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A., Top or middle? Up or down? Toward a standard lexicon for protein top-down and allied mass spectrometry approaches. *Journal of The American Society for Mass Spectrometry* **2019,** *30* (7), 1149-1157.

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

199. Zhang, Y.; Ficarro, S. B.; Li, S.; Marto, J. A., Optimized Orbitrap HCD for quantitative analysis of phosphopeptides. *Journal of the American Society for Mass Spectrometry* **2009,** *20* (8), 1425-1434.

200. Greer, S. M.; Parker, W. R.; Brodbelt, J. S., Impact of protease on ultraviolet photodissociation mass spectrometry for bottom-up proteomics. *Journal of proteome research* **2015,** *14* (6), 2626-2632.

201. Dupree, E. J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B. A.; Darie, C. C., A critical review of bottom-up proteomics: The good, the bad, and the future of this field. *Proteomes* **2020,** *8* (3), 14.

202. Orsburn, B. C., Proteome Discoverer—A Community Enhanced Data Processing Suite for Protein Informatics. *Proteomes* **2021,** *9* (1), 15.

203. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS: An International Journal* **1999,** *20* (18), 3551-3567.

204. MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010,** *26* (7), 966-968.

205. Schlaffner, C. N.; Kahnert, K.; Muntel, J.; Chauhan, R.; Renard, B. Y.; Steen, J. A.; Steen, H., FLEXIQuant-LF to quantify protein modification extent in label-free proteomics data. *Elife* **2020,** *9*, e58783.

206. Chait, B. T., Mass spectrometry: bottom-up or top-down? *Science* **2006,** *314* (5796), 65-66.

207. Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W., Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proceedings of the National Academy of Sciences* **2002,** *99* (4), 1774-1779.

208. Nicolardi, S.; Switzar, L.; Deelder, A. M.; Palmblad, M.; van der Burgt, Y. E., Top-down MALDI-in-source decay-FTICR mass spectrometry of isotopically resolved proteins. *Analytical chemistry* **2015,** *87* (6), 3429-3437.

209. Zabrouskov, V.; Ge, Y.; Schwartz, J.; Walker, J. W., Unraveling molecular complexity of phosphorylated human cardiac troponin I by top down electron capture dissociation/electron transfer dissociation mass spectrometry. *Molecular & Cellular Proteomics* **2008,** *7* (10), 1838-1849.

210. Shi, S. D.-H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W., Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Analytical Chemistry* **2001,** *73* (1), 19-22.

211. Chen, B.; Lin, Z.; Alpert, A. J.; Fu, C.; Zhang, Q.; Pritts, W. A.; Ge, Y., Online hydrophobic interaction chromatography–mass spectrometry for the analysis of intact monoclonal antibodies. *Analytical chemistry* **2018,** *90* (12), 7135-7138.

212. Weisbrod, C. R.; Kaiser, N. K.; Syka, J. E.; Early, L.; Mullen, C.; Dunyach, J.-J.; English, A. M.; Anderson, L. C.; Blakney, G. T.; Shabanowitz, J., Front-end electron transfer dissociation coupled to a 21 Tesla FT-ICR mass spectrometer for intact protein sequence analysis. *Journal of The American Society for Mass Spectrometry* **2017,** *28* (9), 1787-1795.

213. Liu, X.; Inbar, Y.; Dorrestein, P. C.; Wynne, C.; Edwards, N.; Souda, P.; Whitelegge, J. P.; Bafna, V.; Pevzner, P. A., Deconvolution and database search of complex tandem mass spectra of intact proteins. *Molecular & Cellular Proteomics* **2010,** *9* (12), 2772-2782.

214. Jeong, K.; Kim, J.; Gaikwad, M.; Hidayah, S. N.; Heikaus, L.; Schlüter, H.; Kohlbacher, O., FLASHDeconv: Ultrafast, high-quality feature deconvolution for top-down proteomics. *Cell Systems* **2020,** *10* (2), 213-218. e6.

215. Kou, Q.; Xun, L.; Liu, X., TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **2016,** *32* (22), 3495-3497.

216. Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M., ProSight Lite: graphical software to analyze top-down mass spectrometry data. *Proteomics* **2015,** *15* (7), 1235-1238.

217. Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y.-B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L., ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic acids research* **2007,** *35* (suppl_2), W701-W706.

218. 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.

219. Leney, A. C.; Heck, A. J., Native mass spectrometry: what is in the name? *Journal of the American Society for Mass Spectrometry* **2016,** *28* (1), 5-13.

220. Katta, V.; Chait, B. T., Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry. *Journal of the American Chemical Society* **1991,** *113* (22), 8534-8535.

221. Ganem, B.; Li, Y. T.; Henion, J. D., Detection of noncovalent receptor-ligand complexes by mass spectrometry. *Journal of the American Chemical Society* **1991,** *113* (16), 6294-6296.

222. Yin, S.; Loo, J. A., Top-down mass spectrometry of supercharged native protein–ligand complexes. *International journal of mass spectrometry* **2011,** *300* (2-3), 118-122.

223. Loo, J. A., Observation of large subunit protein complexes by electrospray ionization mass spectrometry. *Journal of Mass Spectrometry* **1995,** *30* (1), 180-183.

224. Sobott, F.; Robinson, C. V., Characterising electrosprayed biomolecules using tandem-MS—the noncovalent GroEL chaperonin assembly. *International Journal of Mass Spectrometry* **2004,** *236* (1-3), 25-32.

225. van de Waterbeemd, M.; Snijder, J.; Tsvetkova, I. B.; Dragnea, B. G.; Cornelissen, J. J.; Heck, A. J., Examining the heterogeneous genome content of multipartite viruses BMV and CCMV by native mass spectrometry. *Journal of The American Society for Mass Spectrometry* **2016,** *27* (6), 1000-1009.

226. 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.

227. Bern, M.; Caval, T.; Kil, Y. J.; Tang, W.; Becker, C.; Carlson, E.; Kletter, D.; Sen, K. I.; Galy, N.; Hagemans, D., Parsimonious charge deconvolution for native mass spectrometry. *Journal of proteome research* **2018,** *17* (3), 1216-1226.

228. Kostelic, M.; Marty, M., Deconvolving Native and Intact Protein Mass Spectra with UniDec. **2020**.

229. Marty, M. T.; Hoi, K. K.; Gault, J.; Robinson, C. V., Probing the Lipid Annular Belt by Gas-Phase Dissociation of Membrane Proteins in Nanodiscs. *Angewandte Chemie International Edition* **2016,** *55* (2), 550-554.

230. Gault, J.; Liko, I.; Landreh, M.; Shutin, D.; Bolla, J. R.; Jefferies, D.; Agasid, M.; Yen, H.-Y.; Ladds, M. J.; Lane, D. P., Combining native and 'omics' mass spectrometry to identify endogenous ligands bound to membrane proteins. *Nature methods* **2020,** *17* (5), 505-508.

231. Belov, M. E.; Damoc, E.; Denisov, E.; Compton, P. D.; Horning, S.; Makarov, A. A.; Kelleher, N. L., From protein complexes to subunit backbone fragments: a multi-stage approach to native mass spectrometry. *Analytical chemistry* **2013,** *85* (23), 11163-11173.

232. Greisch, J.-F.; Tamara, S.; Scheltema, R. A.; Maxwell, H. W.; Fagerlund, R. D.; Fineran, P. C.; Tetter, S.; Hilvert, D.; Heck, A. J., Expanding the mass range for UVPD-based native top-down mass spectrometry. *Chemical science* **2019,** *10* (30), 7163-7171.

233. Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentić, K.; Kelleher, N. L., Top-down characterization of endogenous protein complexes with native proteomics. *Nature chemical biology* **2018,** *14* (1), 36-41.

234. Ben-Nissan, G.; Belov, M. E.; Morgenstern, D.; Levin, Y.; Dym, O.; Arkind, G.; Lipson, C.; Makarov, A. A.; Sharon, M., Triple-stage mass spectrometry unravels the heterogeneity of an endogenous protein complex. *Analytical chemistry* **2017,** *89* (8), 4708-4715.

235. Canzani, D.; Rusnac, D. a.-V.; Zheng, N.; Bush, M. F., Degronomics: Mapping the Interacting Peptidome of a Ubiquitin Ligase Using an Integrative Mass Spectrometry Strategy. *Analytical chemistry* **2019,** *91* (20), 12775-12783.

236. Laganowsky, A.; Reading, E.; Hopper, J. T.; Robinson, C. V., Mass spectrometry of intact membrane protein complexes. *Nature protocols* **2013,** *8* (4), 639-651.

237. Sipe, S. N.; Patrick, J. W.; Laganowsky, A.; Brodbelt, J. S., Enhanced characterization of membrane protein complexes by ultraviolet photodissociation mass spectrometry. *Analytical chemistry* **2019,** *92* (1), 899-907.

238. Konijnenberg, A.; Bannwarth, L.; Yilmaz, D.; Koçer, A.; Venien-Bryan, C.; Sobott, F., Top-down mass spectrometry of intact membrane protein complexes reveals oligomeric state and sequence information in a single experiment. *Protein Science* **2015,** *24* (8), 1292-1300.

239. 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.

240. Acharya, S.; Safaie, B. M.; Wongkongkathep, P.; Ivanova, M. I.; Attar, A.; Klärner, F.-G.; Schrader, T.; Loo, J. A.; Bitan, G.; Lapidus, L. J., Molecular basis for preventing α-synuclein aggregation by a molecular tweezer. *Journal of Biological Chemistry* **2014,** *289* (15), 10727-10737.

241. 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.

242. 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.

243. Loo, J. A.; Hu, P.; Smith, R. D., Interaction of angiotensin peptides and zinc metal ions probed by electrospray ionization mass spectrometry. *Journal of the American Society for Mass Spectrometry* **1994,** *5* (11), 959-965.

244. Zhang, Z.; Browne, S. J.; Vachet, R. W., Exploring salt bridge structures of gas-phase protein ions using multiple stages of electron transfer and collision induced dissociation. *Journal of The American Society for Mass Spectrometry* **2014,** *25* (4), 604-613.

245. Zhang, J.; Loo, R. R. O.; Loo, J. A., Increasing fragmentation of disulfide-bonded proteins for top–down mass spectrometry by supercharging. *International journal of mass spectrometry* **2015,** *377*, 546-556.

246. Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W., Detailed unfolding and folding of gaseous ubiquitin ions characterized by electron capture dissociation. *Journal of the American Chemical Society* **2002,** *124* (22), 6407-6420.

247. Horn, D. M.; Breuker, K.; Frank, A. J.; McLafferty, F. W., Kinetic intermediates in the folding of gaseous protein ions characterized by electron capture dissociation mass spectrometry. *Journal of the American Chemical Society* **2001,** *123* (40), 9792-9799.

248. Zhang, H.; Cui, W.; Wen, J.; Blankenship, R. E.; Gross, M. L., Native electrospray and electron-capture dissociation FTICR mass spectrometry for top-down studies of protein assemblies. *Analytical chemistry* **2011,** *83* (14), 5598-5606.

249. Lermyte, F.; Konijnenberg, A.; Williams, J. P.; Brown, J. M.; Valkenborg, D.; Sobott, F., ETD allows for native surface mapping of a 150 kDa noncovalent complex on a commercial Q-TWIMS-TOF instrument. *Journal of the American Society for Mass Spectrometry* **2014,** *25* (3), 343-350.

250. 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.

251. 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.

252. Morrison, L. J.; Chai, W.; Rosenberg, J. A.; Henkelman, G.; Brodbelt, J. S., Characterization of hydrogen bonding motifs in proteins: hydrogen elimination monitoring by ultraviolet photodissociation mass spectrometry. *Physical Chemistry Chemical Physics* **2017,** *19* (30), 20057-20074.

253. Jackson, S. N.; Dutta, S.; Woods, A. S., The use of ECD/ETD to identify the site of electrostatic interaction in noncovalent complexes. *Journal of the American Society for Mass Spectrometry* **2009,** *20* (2), 176-179.

254. Clarke, D. J.; Murray, E.; Hupp, T.; Mackay, C. L.; Langridge-Smith, P. R., Mapping a noncovalent protein–peptide interface by top-down FTICR mass spectrometry using electron capture dissociation. *Journal of The American Society for Mass Spectrometry* **2011,** *22* (8), 1432-1440.

255. Schneeberger, E. M.; Breuker, K., Native top-down mass spectrometry of TAR RNA in complexes with a wild-type tat peptide for binding site mapping. *Angewandte Chemie* **2017,** *129* (5), 1274-1278.

256. Lam, Y. P.; Wootton, C. A.; Hands-Portman, I.; Wei, J.; Chiu, C. K.; Romero-Canelon, I.; Lermyte, F.; Barrow, M. P.; O'Connor, P. B., Determination of the aggregate binding site of amyloid protofibrils using electron capture dissociation tandem mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (2), 267-276.

257. Hall, Z.; Politis, A.; Robinson, C. V., Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. *Structure* **2012,** *20* (9), 1596-1609.

258. Bleiholder, C.; Do, T. D.; Wu, C.; Economou, N. J.; Bernstein, S. S.; Buratto, S. K.; Shea, J.-E.; Bowers, M. T., Ion mobility spectrometry reveals the mechanism of amyloid formation of Aβ (25–35) and its modulation by inhibitors at the molecular level: epigallocatechin gallate and scyllo-inositol. *Journal of the American Chemical Society* **2013,** *135* (45), 16926-16937.

259. Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V., Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nature chemistry* **2009,** *1* (4), 326-331.

260. Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T., Ion mobility–mass spectrometry reveals a conformational conversion from random assembly to β-sheet in amyloid fibril formation. *Nature chemistry* **2011,** *3* (2), 172-177.

261. Poltash, M. L.; McCabe, J. W.; Shirzadeh, M.; Laganowsky, A.; Clowers, B. H.; Russell, D. H., Fourier transform-ion mobility-orbitrap mass spectrometer: a next-generation instrument for native mass spectrometry. *Analytical chemistry* **2018,** *90* (17), 10472-10478.

262. Harvey, S. R.; Porrini, M.; Stachl, C.; MacMillan, D.; Zinzalla, G.; Barran, P. E., Small-molecule inhibition of c-MYC: MAX leucine zipper formation is revealed by ion mobility mass spectrometry. *Journal of the American Chemical Society* **2012,** *134* (47), 19384-19392.

263. Robinson, E. W.; Leib, R. D.; Williams, E. R., The role of conformation on electron capture dissociation of ubiquitin. *Journal of the American Society for Mass Spectrometry* **2006,** *17* (10), 1470-1479.

264. 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.

265. Harvey, S. R.; Porrini, M.; Tyler, R. C.; MacPhee, C. E.; Volkman, B. F.; Barran, P. E., Electron capture dissociation and drift tube ion mobility-mass spectrometry coupled with site directed mutations provide insights into the conformational diversity of a metamorphic protein. *Physical Chemistry Chemical Physics* **2015,** *17* (16), 10538-10550.

266. Lermyte, F.; Verschueren, T.; Brown, J. M.; Williams, J. P.; Valkenborg, D.; Sobott, F., Characterization of top-down ETD in a travelling-wave ion guide. *Methods* **2015,** *89*, 22-29.

267. Lermyte, F.; Williams, J. P.; Brown, J. M.; Martin, E. M.; Sobott, F., Extensive charge reduction and dissociation of intact protein complexes following electron transfer on a quadrupole-ion mobility-time-of-flight MS. *Journal of The American Society for Mass Spectrometry* **2015,** *26* (7), 1068-1076.

268. Barthélemy, N. R.; Mallipeddi, N.; Moiseyev, P.; Sato, C.; Bateman, R. J., Tau phosphorylation rates measured by mass spectrometry differ in the intracellular brain vs. extracellular cerebrospinal fluid compartments and are differentially affected by Alzheimer's disease. *Frontiers in aging neuroscience* **2019,** *11*, 121.

269. Sarafian, T. A.; Ryan, C. M.; Souda, P.; Masliah, E.; Kar, U. K.; Vinters, H. V.; Mathern, G. W.; Faull, K. F.; Whitelegge, J. P.; Watson, J. B., Impairment of mitochondria in adult mouse brain overexpressing predominantly full-length, N-terminally acetylated human α-synuclein. *PloS one* **2013,** *8* (5), e63557.

270. Kellie, J. F.; Higgs, R. E.; Ryder, J. W.; Major, A.; Beach, T. G.; Adler, C. H.; Merchant, K.; Knierman, M. D., Quantitative measurement of intact alpha-synuclein proteoforms from post-mortem control and Parkinson's disease brain tissue by intact protein mass spectrometry. *Scientific reports* **2014,** *4* (1), 1-10.

271. Stephens, A. D.; Zacharopoulou, M.; Moons, R.; Fusco, G.; Seetaloo, N.; Chiki, A.; Woodhams, P. J.; Mela, I.; Lashuel, H. A.; Phillips, J. J., Extent of N-terminus exposure of monomeric alpha-synuclein determines its aggregation propensity. *Nature communications* **2020,** *11* (1), 1-15.

272. Malik, R.; Meng, H.; Wongkongkathep, P.; Corrales, C. I.; Sepanj, N.; Atlasi, R. S.; Klärner, F.-G.; Schrader, T.; Spencer, M. J.; Loo, J. A., The molecular tweezer CLR01 inhibits aberrant superoxide dismutase 1 (SOD1) self-assembly in vitro and in the G93A-SOD1 mouse model of ALS. *Journal of Biological Chemistry* **2019,** *294* (10), 3501-3513.

273. Vöpel, T.; Bravo-Rodriguez, K.; Mittal, S.; Vachharajani, S.; Gnutt, D.; Sharma, A.; Steinhof, A.; Fatoba, O.; Ellrichmann, G.; Nshanian, M., Inhibition of huntingtin exon-1 aggregation by the molecular tweezer CLR01. *Journal of the American Chemical Society* **2017,** *139* (16), 5640-5643.

274. Hoffmann, W.; Folmert, K.; Moschner, J.; Huang, X.; von Berlepsch, H.; Koksch, B.; Bowers, M. T.; von Helden, G.; Pagel, K., NFGAIL amyloid oligomers: the onset of beta-sheet formation and the mechanism for fibril formation. *Journal of the American Chemical Society* **2018,** *140* (1), 244-249.

275. Laos, V.; Do, T. D.; Bishop, D.; Jin, Y.; Marsh, N. M.; Quon, B.; Fetters, M.; Cantrell, K. L.; Buratto, S. K.; Bowers, M. T., Characterizing TDP-43307–319 Oligomeric Assembly: Mechanistic and Structural Implications Involved in the Etiology of Amyotrophic Lateral Sclerosis. *ACS Chemical Neuroscience* **2019,** *10* (9), 4112-4123.

276. Bernstein, S. L.; Liu, D.; Wyttenbach, T.; Bowers, M. T.; Lee, J. C.; Gray, H. B.; Winkler, J. R., α-Synuclein: stable compact and extended monomeric structures and pH dependence of dimer formation. *Journal of the American Society for Mass Spectrometry* **2004,** *15* (10), 1435-1443.

277. Shahpasand-Kroner, H.; Portillo, J.; Lantz, C.; Seidler, P. M.; Sarafian, N.; Loo, J. A.; Bitan, G., Three-repeat and four-repeat tau isoforms form different oligomers. *Protein Science* **2022**.

CHAPTER 2

ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry

Adapted with permission from *J. Proteome Res.* 2021, 20, 4, 1928–1935. 2022 American Chemical Society

Journal of Proteome Research

March 2nd 2021

DOI: 10.1021/acs.jproteome.0c00952

ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry

Carter Lantz[1], Muhammad A. Zenaidee[1], Benqian Wei[1], Zachary Hemminger[1], Rachel R. Ogorzalek Loo[2], Joseph A. Loo[1,2]

[1]Department of Chemistry and Biochemistry, [2]Department of Biological Chemistry, University of California Los Angeles, Los Angeles, CA

ABSTRACT: Top-down mass spectrometry (TD-MS) of peptides and proteins results in product ions that can be correlated to polypeptide sequence. Fragments can either be terminal fragments, which contain either the N- or the C-terminus, or internal fragments that contain neither termini. Normally, only terminal fragments are assigned due to the computational difficulties of assigning internal fragments. Here we describe ClipsMS, an algorithm that can assign both terminal and internal fragments generated by top-down MS fragmentation. Further, ClipsMS can be used to locate various modifications on the protein sequence. Using ClipsMS to assign TD-MS generated product ions, we demonstrate that for apo-myoglobin, the inclusion of internal fragments increases the sequence coverage up to 78%. Interestingly, many internal fragments cover complimentary regions to the terminal fragments that enhance the information that is extracted from a single top-down mass spectrum. Analysis of oxidized apo-myoglobin using terminal and internal fragment matching by ClipsMS confirmed the locations of oxidation sites on the two methionine residues. Internal fragments can be beneficial for top-down protein fragmentation analysis, and ClipsMS can be a valuable tool for assigning both terminal and internal fragments present in a top-down mass spectrum

Keywords: Top-Down Mass Spectrometry (TD-MS), Terminal Fragment, Internal Fragment, Electron Capture Dissociation (ECD)

INTRODUCTION

Top-down mass spectrometry (TD-MS) has become a prominent tool for the analysis and characterization of *intact* proteins and protein complexes.[1, 2] TD-MS analysis of proteins and protein complexes has many advantages, including the ability to detect and identify degradation products, sequence

variations, post-translational modifications (PTMs), and other proteoforms.[3] TD-MS has progressed significantly in the last decade owing to advances in instrumentation and associated technologies.[4-6] For example, TD-MS has recently been utilized for the characterization and analysis of heterogeneous samples, large noncovalent protein complexes, and intact monoclonal antibodies.[7] Despite these advances, however, the application of TD-MS for profiling PTMs and proteoforms is limited in sensitivity and scope, as data produced by top-down MS methods are not as easily analyzed compared to bottom-up proteomics.

TD-MS analysis of intact proteins typically starts by forming multiply charged gas-phase proteins using electrospray ionization (ESI).[8, 9] The protein ions can then be activated and fragmented by collision-,[10, 11] photon-,[12, 13] or electron-based dissociation methods[14-16] to generate product ions that can be assigned to the protein primary sequence.[17] Product ions formed by top-down MS can either be i) a terminal fragment ion, which includes the N-terminus (*a, b,* or *c* fragment) or the C-terminus (*x, y,* or *z* fragment) of the polypeptide sequence,[18] or ii) an internal fragment ion that results from multiple cleavage events of the protein backbone to generate *ax, ay, az, bx, by, bz, cx, cy,* and *cz* fragment ions, with the first letter designating the cleavage site on the N-terminal side of the fragment and the second letter designating the cleavage on the C-terminal side of the fragment.[19-21] The isotopically resolved mass spectral signals[22] can be matched to regions of the sequence to return information about the protein's primary structure. Within a single protein fragmentation mass spectrum, there can be hundreds of product ion signals that could be assignable.

Assignment of mass spectral signals within a mass spectrum can be a long and arduous task that can require manual comparisons of experimentally measured masses to lists of computed theoretical masses to return putative information of the protein sequence. There has been significant development of software tools to aid the deconvolution and automated assignment of TD mass spectral signals. Multiple algorithms have been developed to deconvolute MS spectra including MSDeconv,[23] YADA,[24] and mMass.[25] Other programs such as ProSight PTM 2.0[26] and ProSightLite[27] match deconvoluted mass lists to a theoretical mass list from a given protein sequence. Other assignment algorithms include TopPIC[28] and MS-Align+.[29] Ge and co-workers have developed MASH Explorer, which allows the user to load raw data from a top-down mass spectrum, deconvolute the peaks present in that spectrum, and match the resulting values to

theoretical masses from a given protein sequence.[30] Although these and other tools are potentially powerful, they largely consider only the assignment of terminal fragments, which could leave many peaks in a fragmentation mass spectrum, including those representing internal fragments, to be unassigned. These unassigned signals, that we colloquially term as "dark matter" of a fragmentation mass spectrum, could provide valuable information if assigned correctly to the protein sequence.

However, internal fragment ions have been largely ignored due to the difficulty of accurately and efficiently assigning these mass spectral signals. As the size of the protein increases, the number of internal fragments that can be generated increases exponentially, hence increasing the false discovery rates limiting the accuracy of these assignments.[31, 32] Due to this, internal fragment ion analysis of top-down mass spectra has been limited to peptides and small proteins. Despite the limitations associated with extending the use of internal fragment ion analysis on larger proteins, the inclusion of accurately assigned internal fragment ions could offer richer sequence and structural information.

Recently, the assignment of internal fragment signals for the analysis of protein ions in TD-MS experiments has been reported. Kelleher and co-workers demonstrated that for collision induced dissociation of ubiquitin (8.6 kDa), the inclusion of internal fragments resulted in a greater fraction of the fragmentation spectrum to be explained.[21] Loo and co-workers recently demonstrated that internal fragments can be formed by electron-based fragmentation of ubiquitin and carbonic anhydrase II (29 kDa).[32] Although these reports suggest that internal fragments can significantly enhance the information obtained from a TD-MS experiment, which could be beneficial for localizing sites of protein modifications, to date there have been few readily available computational methods that can be utilized to assign internal fragment ions.

Here, we describe an algorithm developed in Python, coined ClipsMS (**C**omprehensive **L**ocalization of **I**nternal **P**rotein **S**equences), that can be utilized to assign both terminal and internal fragments resulting from a top-down mass spectrometry experiment. This algorithm generates every possible terminal and internal fragment, compares those fragments against a deconvoluted mass list, and graphically displays the data. We demonstrate the use of ClipsMS for the analysis of top-down mass spectra

of wild type (wt) and oxidized apo-myoglobin. Assigning internal fragment masses is shown to increase sequence coverage of the protein sequence and confidence in the location of modified sites present on the protein.

MATERIALS AND METHODS

Apo-myoglobin from equine skeletal muscle was purchased from Sigma-Aldrich (St. Louis, MO, USA) and used without further purification. LC/MS grade water and methanol were obtained from Fisher Chemical (Hampton, NH, USA). Hydrogen peroxide (H_2O_2) was obtained from Sigma-Aldrich (St. Louis, MO, USA).

The oxidized form of apo-myoglobin was prepared by reaction with hydrogen peroxide at a 1:10 ratio of molar concentration (H_2O_2/ apo-myoglobin = 1:10) at 37°C for 30 min. Both wild type and oxidized apo-myoglobin were dissolved in water/ methanol/ formic acid (49.5:49.5:1, v/v/v) at a concentration of 20 μM.

All experiments were conducted on a 15-T Bruker SolariX Fourier transform ion cyclotron resonance (FTICR)-MS instrument equipped with an infinity ICR cell (Bruker Daltonics, Billerica, MA, USA). Protein solutions were loaded into in-house pulled capillaries coated with gold, and electrosprayed by applying a voltage between 0.8 and 1.2 kV on the ESI capillary. MS1 spectra were collected of wildtype and oxidized myoglobin and the spectra were deconvoluted with UniDec.[33] Broadband ECD experiments were conducted without precursor isolation. For ECD fragmentation of wild type apo-myoglobin, the pulse length was set at 0.1s, with a lens voltage at 50 V and an ECD bias voltage at 2V. For ECD fragmentation of the oxidized form, the lens and bias voltage were kept the same, while the pulse length was set at 0.025s to obtain an optimized ECD fragmentation. For each spectrum, 200 scans were obtained. The data was deconvoluted with the SNAP^{TM} 2.0 algorithm from the Bruker Daltronics DataAnalysis software and internally calibrated against theoretical terminal fragments of apo myoglobin.

The algorithm and GUI were designed in PyCharm 2020.2. The program runs on at least python 3.7 and is available on GitHub (https://github.com/loolab2020/ClipsMS-Version-1.0.0). Because ECD was

performed on myoglobin, *c*, *z*, and *cz* fragments were searched for both wildtype and oxidized myoglobin. At the end of every N-terminal fragment a hydrogen atom (1.00783) is added to complete the amino group, and at the end of every C-terminal fragment a hydroxyl group (17.00274) is added to complete the carboxyl group (Figure S1A). The error given to the program was 2 ppm, the smallest internal fragment size is 5, and any fragments containing n-terminal cuts at proline residues were disregarded as false positives. In both instances, the biased version of the algorithm was run. For the oxidized version of the protein, localized modifications corresponding to oxidation (15.99491Da) were added to methionine 55 and methionine 131. An unlocalized modification corresponding to a hydrogen atom (1.00783Da) was also added to each fragment.

Sequence coverage was calculated as number of inter-residue cleavages divided by the total number of inter-residue sites. The number of cleavage site was calculated as the number of unique cleavage sites not already cut by another fragment. The coverage for a PTM site was designated as the number of times that the modified amino acid was covered by a unique fragment.

RESULTS

Graphical user interface of ClipsMS

A graphical user interface (GUI) was designed for ClipsMS so the user can easily compare theoretical fragments of a peptide or protein sequence against a user specified deconvoluted mass list from a top-down mass spectrum (Figure 1A). The algorithm can generate any theoretical terminal and internal fragment from a given amino acid sequence with a user defined minimum sequence length for internal fragments. Users can select the fragment types that can be formed by their experiments and set the mass error tolerance for matching. In addition, modifications can be accounted for: to include localized modifications, where a single amino acid site has been modified, and unlocalized modifications in which modifications can occur on any amino acid site. Once the user inputs the information required, the user can run either a biased version of the algorithm where terminal fragments are weighted higher than internal

fragments, or an unbiased version of the algorithm where both terminal and internal fragments are weighted the same.

Processing and generation of theoretical fragments and fragment matching using ClipsMS

The algorithm calculates a molecular weight based on the amino acid sequence (Table S1) plus a H+ (1.00728 amu) to return the monoisotopic $[M+H]^+$ mass (Figure 1B). Next, all possible *a*, *b*, *c*, *x*, *y*, and *z* terminal fragments of the protein are calculated and stored as a list (Figure 1B, Figure S1A). After terminal fragments are calculated, those fragments are used to calculate the mass of all possible internal fragments: *ax*, *ay*, *az*, *bx*, *by*, *bz*, *cx*, *cy*, and *cz* fragments of a protein (Figure 1B, Figure S1B). All fragment masses are calculated as monoisotopic $[M+H]^+$ masses.

After the base theoretical fragment masses of the amino acid sequence have been generated, these masses are compared against a given deconvoluted mass list (Figure 1B). Each observed mass in the deconvoluted mass list is compared against every theoretical terminal and internal mass for completeness. If modifications have been imported, the shift in mass will be accounted for by the algorithm. The modifications the algorithm accounts for are: (i) localized modifications, (ii) unlocalized modifications, and (iii) terminal modifications. (i) Localized modifications are treated as static modifications that occur on a single amino acid and will not detach from the protein. This would include previously located PTMs, nonstandard mutations (*e.g.*, selenocysteine), the absence of hydrogen atoms from oxidized cysteine residues, and/or user modified proteins. These modifications are added to every terminal and internal fragment that contains that amino acid at a given site before the fragment is compared against the deconvoluted values (Figure 1B). (ii) Unlocalized modifications include modifications that are not attributed to a specific amino acid. This may include addition or subtraction of hydrogen atoms or water molecules, PTMs where the location is not known, or ligands where the location is not known. Unlocalized modifications are added to each mass after the unmodified theoretical masses have been searched (Figure 1B). These modifications were designed to allow for both unmodified and modified peaks to be analyzed. (iii) Terminal modifications are added to the end of every C or N terminal fragment before comparing that fragment against the deconvoluted values (Figure 1B). The terminal modifications do not affect internal

fragments matching as they do not contain either terminus. If the measured mass error of an observed fragment compared to a theoretical fragment is within the error tolerance set by the user, the fragment is counted as a match.

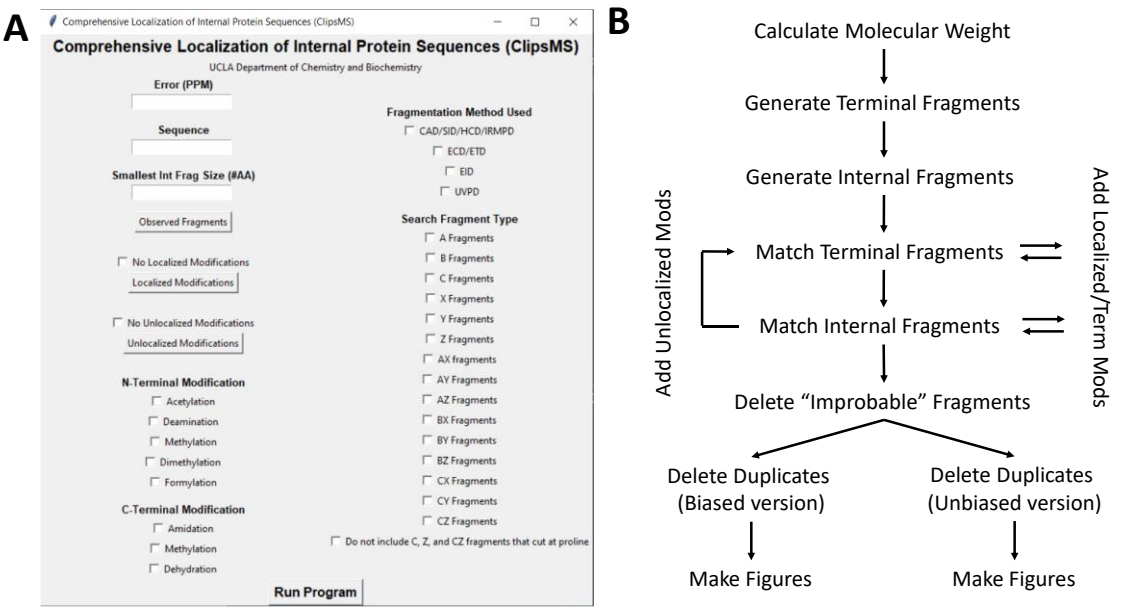


Figure 1: A. The graphical user interface (GUI) for ClipsMS. The user can input several key parameters including the error allowed, the smallest internal fragment size, the sequence, the observed fragments, any modifications on the sequence and the type of fragments to search. B. The workflow of the algorithm and how it matches peaks input by the user. The algorithm calculates all theoretical terminal and internal fragments, matches all peaks, makes decisions on which assignments to keep, and automatically generates figures.

After deconvoluted masses are matched with the theoretical masses, matches that cannot occur are automatically deleted from the matched list (Figure 1B). This can include matches that contain *c* and *z* fragmentation at proline residues in electron-based fragmentation or matches containing improbable unlocalized modifications. For electron-based fragmentation methods, the α-amino group on proline contains two bonds making it an imino acid, *c* or *z* fragmentation at these residues are indicated as improbable[34] and removed from the matched list. There is an option to use this feature in the GUI. (Figure 1A) An example of an improbable unlocalized modification could be a match that contains a mass shift equal to phosphorylation on a fragment not containing a serine, threonine, or tyrosine. If a fragment contains a modification but does not have the amino acids required, the match is designated as a false positive and removed from the list. The masses of unlocalized modifications and the amino acids on which they occur

can be input by the user with the GUI (Figure 1A). These safeguards help reduce false positives and generate more accurate results.

After all possible fragments have been matched and improbable fragments have been removed, the algorithm makes decisions on duplicate fragment assignments. For example, a single deconvoluted mass can be assigned to multiple theoretical masses provided these masses fall within the error tolerance set by the user. The algorithm decides which fragment to keep depending on the version run. The unbiased version of the algorithm assigns the fragment with a lower mass error (Figure 1B). If two fragments have exactly the same chemical formula (and as a result the same mass), both fragments are kept and displayed. All matched fragments are output in a table for the user to review. The biased version of the algorithm favors the terminal fragments (Figure 1B). If a terminal fragment and an internal fragment are matched to a single deconvoluted mass, the terminal fragment is chosen as the assignment. If two terminal or two internal fragments are matched to the same deconvoluted mass, the fragment with the lower mass error is assigned. However, two terminal or two internal fragments with exactly the same chemical formula and mass are retained and output in a table for the user to review. These features were incorporated to limit false positive matches.

The output table contains all the information for the observed fragments that were matched with the theoretical fragment list. It includes the fragment type, any localized and unlocalized modifications, any terminal modifications, the observed mass, the theoretical mass calculated by ClipsMS, the starting and ending amino acids, the error of the observed mass compared to the theoretical mass, the sequence of the fragment, the intensity given by the user, and the molecular formula of the fragment (Table S2). This table allows for manual interpretation of the data if desired.

The time it takes for the algorithm to run depends on the length of the protein sequence, the number of fragments uploaded, and the number of modifications on the protein sequence. Peptides and small proteins (< 30 kDa) with a few hundred deconvoluted peaks take less than 1 minute to analyze on a laptop with 12GB of ram with an Intel Core i7-2760QM processor (4 cores @ 2.40 GHz). Larger sequences or sequences with more modifications can take up to a few minutes to complete. At the end of every run, a

.csv document with all matched fragments is exported (Table S2) and 3 figures are output representing the fragments that are matched.

Top-down fragmentation data analysis using ClipsMS

To test ClipsMS for top-down fragmentation analysis, apo-myoglobin was prepared in denaturing conditions and electrosprayed on a Bruker 15T FT-ICR MS (Figure S2A). Apo-myoglobin was fragmented with broadband electron capture dissociation (ECD) MS (Figure 2A). In the resulting spectrum, charge reduced precursor ions are present as well as fragment ions (Figure 2A). The data deconvoluted from the SNAP^{TM} algorithm was input to the algorithm along with the intensity values and the biased version of the algorithm was run.

Unlike terminal fragments where one end is fixed, internal fragments contain neither the N nor C terminus. Because neither terminus is fixed, it is difficult to represent internal fragments with the conventional top-down fragmentation map such as the one used by Prosight Lite[27] and MASH Explorer.[30] To represent both terminal and internal fragments, the algorithm outputs 3 figures: (i) a fragment location map, (ii) a sequence coverage map, and (iii) a fragmentation cleavage site map. Each of these figures displays a key piece of information to describe the data analyzed by the algorithm.

The first figure displays the data in a way so that the number of internal fragments identified is easily determined and the coverage of the protein sequence can be easily shown (Figure 2B). In addition, the size of the dots indicates the relative intensity of the matched fragments. The myoglobin data indicated that 98 *c* and *z* terminal fragments were assigned, and 15 *cz* internal fragments were assigned (Figure 2B). Furthermore, the data indicates that the internal fragments are normally lower in abundance than many of the terminal fragments (Figure 2B). This data also shows that the *c* and *z* terminal fragments assigned heavily cover both the N and C terminal regions of the sequence and the *cz* internal fragments cover the interior of the protein.

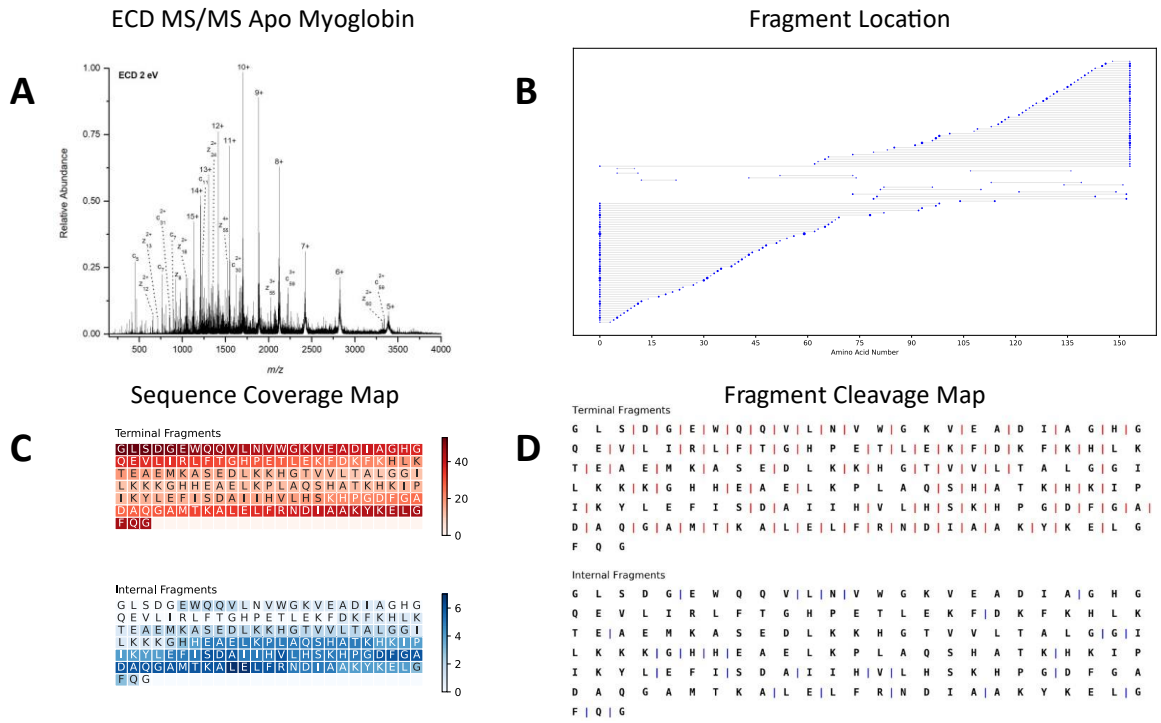


Figure 2: A. Broadband ECD MS of 20 μM apo-myoglobin formed from acidic denaturing conditions. B. A fragment location map indicating the region of the protein sequence covered by terminal and internal fragments. C. A sequence coverage map for the terminal and internal fragments. Darker regions indicate more coverage. D. A fragment cleavage map indicating the location of inter-amino acid cleavage sites for terminal and internal fragments.

The second figure includes the sequence information map of a protein. This map is based off a figure in a paper published by the Kelleher lab.[21] The figure represents the areas of the sequence that are covered by the product ions. Darker regions of the sequence indicate regions of the protein that are covered by many fragments while lighter colored regions of the sequence indicate regions of the protein that are covered by fewer fragments. This data indicates regions of the protein that terminal and internal fragments cover. Terminal fragments heavily cover the ends of protein sequences and internal fragments cover the interior of the protein. Internal fragments increase the amount of sequence information of the protein. In the apo-myoglobin data, the sequence coverage is 59% when only terminal fragments are considered; however, by including internal product ions in the search, the sequence coverage was increased to 66% (Figure 2C). Including internal fragments tends to enhance the sequence coverage on proteins by giving

information of the amino acid sequence in the center of the protein sequence. This observation has been reported before by our lab for carbonic anhydrase II,[32] and an updated fragmentation location map is included (Figure S3).

Figure 2D shows the cleavage sites of terminal and terminal fragments. Analyzing internal fragments can increase the number of fragmentation sites that occur in the protein sequence. The cleavage sites of terminal fragments and the cleavage sites of internal fragments are displayed. In the apo-myoglobin data, it is shown that the terminal fragments identified cleave at 90 of the 152 inter-residue cleavage sites of the protein (Figure 2D). Inclusion of *cz* internal fragments increased the number of cleavage sites on apo myoglobin to 101 (Figure 2D). Increasing the number of cleavage sites can increase confidence in the sequence identity as more protein information can be extracted.

Unlocalized modification feature of ClipsMS

ClipsMS can also be utilized to investigate unlocalized modifications on protein fragment ions. When unlocalized modifications are considered, ClipsMS will generate a list of fragments with the addition and/or subtraction of a user defined mass after it has searched for the unmodified masses. This list of fragments will then be matched to fragments that the user has input. For wt apo-myoglobin, we investigated the unlocalized modification feature by examining *c*, *z*, and *cz* fragments from wt apo-myoglobin that can be matched with the loss/gain of a hydrogen atom (1.00783Da), which are prevalent in ECD.[35] Including the addition of a hydrogen atom increased the number of assignments to 158 (Figure S4B). In addition, the inclusion of internal fragments increased the sequence coverage from 63% to 78% when fragments with an extra hydrogen were searched, (Figure S4C) and the number of cut sites increased from 95 to 118 (Figure S4D). This data indicates that ClipsMS's unlocalized modification feature can be utilized to increase the number of assignments and the sequence coverage of proteins. In addition to analyzing fragments with neutral mass losses, this feature can also be used to pinpoint PTMs where the location is not specified and/or ligand binding sites of a protein as the inclusion of internal fragments could give more informative data on where PTMs and ligand binding occurs.

Confirmation of localized modification sites using ClipsMS

To test ClipsMS's localized modification feature, apo-myoglobin was oxidized and fragmented with ECD. An MS1 spectrum of oxidized apo-myoglobin showed a mass difference of 32Da compared to wt apo-myoglobin, suggesting that 2 oxidation sites were present (Figure S2B). Oxidation of intact proteins can occur on methionine residues,[36] and apo-myoglobin contains two methionine residues at positions 55 and 131. To confirm that oxidation occurred on these two residues, broadband ECD was performed on the oxidized myoglobin sample (Figure 3A). For oxidized apo-myoglobin, 73 terminal fragments and 20 internal fragments were identified, with 39 of those fragments containing a single oxidation site and 1 fragment containing both oxidation sites (Figure 3B). Modifications on apo-myoglobin were confirmed to be on methionine 55 and 131 and these sites are indicated by a dashed line (Figure 3B). This data indicates that the inclusion of internal fragments enhances the confidence of the location of both oxidation sites. Terminal fragments cover the residues near N and C terminus while the internal fragments cover the residues interior of the protein (Figure 3C). For Met-55, terminal fragments cover the residue 11 times, and 4 additional internal fragments cover Met-55. Similarly, Met-131 was covered by a terminal fragment 23 times, and 3 additional internal fragments cover the modified residue.

When the inter-amino acid cleavage sites of apo-myoglobin are considered, terminal fragments account for 67 inter-amino cleavage sites and inclusion of internal fragments increased the number of inter-amino acid cleavage sites to 84 (Figure 3D). This included a cut site between position 53 and 54, which narrowed down the location of oxidation on Met-55. Increased coverage of residues with PTMs increases confidence that a modification occurs on a particular residue.

DISCUSSION

ClipsMS efficiently assigns both terminal and internal fragments present in top-down mass spectra. Assigning internal fragments can enhance top-down mass spectrometry analysis (Figure 2B). For apo-myoglobin, the sequence coverage obtained was enhanced from 59% to 66%. This data agrees well with previous reports obtained for proteins and peptides.[32, 37, 38] Conventionally, internal fragment analysis has been ignored; including internal fragments may help to increase the molecular weight limit of 30 kDa that

is often observed for high sequence coverage TD-MS.[17, 39, 40] Here, we demonstrate that internal fragments can provide more information of the interior of the protein (Figure 2C). In addition, a plethora of PTMs are located within the interior of protein sequences,[41] hence internal fragment assignments can aid in the localization of these PTMs.

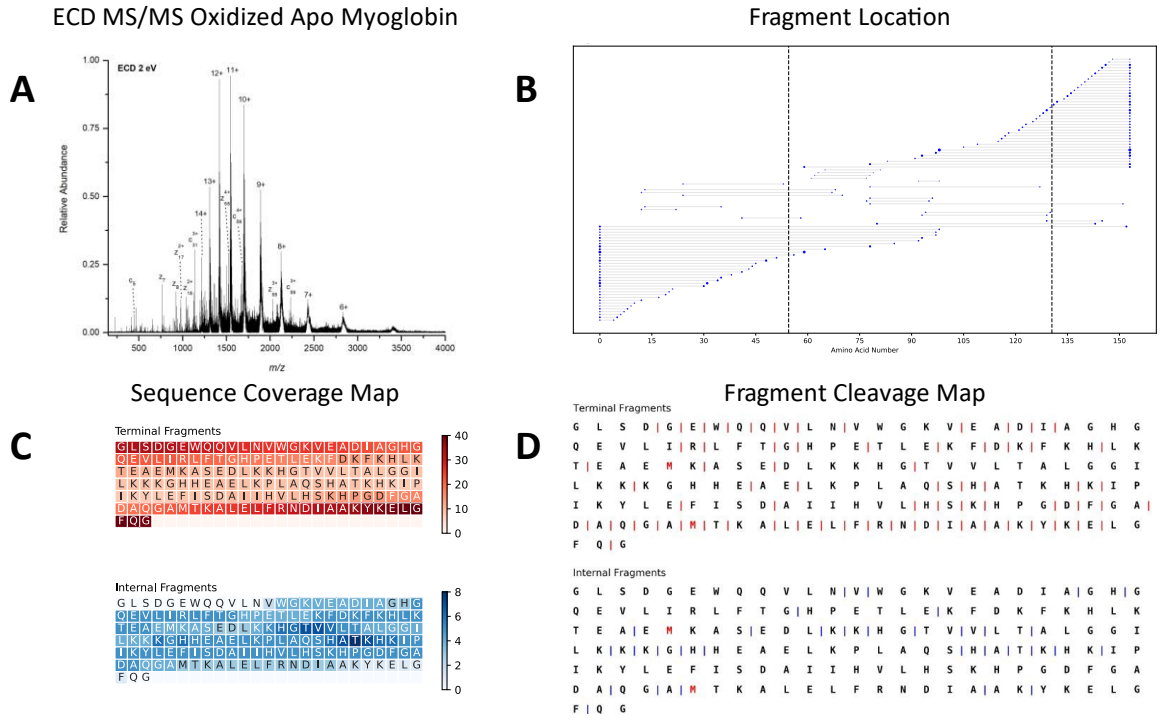


Figure 3: A. Broadband ECD MS of 20 μM oxidized apo-myoglobin formed from acidic denaturing conditions. B. A fragment location map indicating the region of the protein sequence covered by terminal and internal fragments. Dashed lines indicate sites of oxidation. C. A sequence coverage map for the terminal and internal fragments assigned indicating terminal and internal fragments cover both oxidation sites. Darker regions indicate more coverage. D. A fragment cleavage map indicating the location of inter-amino acid cleavage sites for terminal and internal fragments. Red amino acids indicate sites of oxidation.

Although ClipsMS has been shown to be a powerful tool for top-down fragmentation assignments, there are a few limitations of the algorithm. Duplicated fragments pose a problem for internal fragment analysis. A single deconvoluted mass can be matched to multiple theoretical masses due to those fragments having the same elemental composition. It is possible that a better understanding of top-down fragmentation mechanisms and/or ion mobility analysis of top-down fragments can help overcome this issue. Another current limitation of ClipsMS is that some protein modification types are not considered.[35, 42, 43] In the future,

the algorithm will include the capability to specifically search for neutral losses and more diverse fragment types such as *c*+1 fragments, *z*+1 fragments, and *z*· fragments. In addition, the ability to search for labile modifications on specific amino acids will be incorporated into the search algorithm. Lastly, larger sequences with more fragments or data containing many modifications can take up to a few minutes to complete. For the myoglobin data, processing times ranged from ~30 seconds to ~11 minutes on the same device (Table S3). A dummy 1023 amino acid sequence (116.3kDa) with 250 fragments took approximately an hour to run on a laptop with 12GB of ram with an Intel Core i7-2760QM processor (4 cores @ 2.40 GHz). Currently however, the architecture of the algorithm is such that it only uses a single core, which limits the processing times. By allowing access to more cores and/or ram the processing time can be significantly reduced. Despite all these limitations, the information obtained from ClipsMS can still be beneficial for top-down protein fragmentation analysis.

As top-down proteomics becomes more mainstream in proteome research, it is becoming increasingly important to efficiently analyze top-down mass spectrometry data. The top-down community for the most part has disregarded internal fragments and opted to only analyze terminal fragments. By analyzing internal fragments, it is possible to gain more insight into the protein sequence. We hope this algorithm will aid researchers to mine some of the previously unknown "dark matter" in top-down mass spectra and will spur research in proteomics and the proteoforms that exist in nature.

ClipsMS (https://github.com/loolab2020/ClipsMS-Version-1.0.0)

REFERENCES

1. Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; Mclafferty, F. W., Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *Journal of the American Chemical Society* **1999,** *121* (4), 806-812.

2. Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A., Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *Journal of the American Society for Mass Spectrometry* **2019,** *30* (7), 1149-1157.

3. Durbin, K. R.; Fornelli, L.; Fellers, R. T.; Doubleday, P. F.; Narita, M.; Kelleher, N. L., Quantitation and Identification of Thousands of Human Proteoforms below 30 kDa. *Journal of Proteome Research* **2016,** *15* (3), 976-982.

4. Denisov, E.; Damoc, E.; Lange, O.; Makarov, A., Orbitrap mass spectrometry with resolving powers above 1,000,000. *International Journal of Mass Spectrometry* **2012,** *325-327*, 80-85.

5. Kelly, R. T.; Tolmachev, A. V.; Page, J. S.; Tang, K.; Smith, R. D., The ion funnel: Theory, implementations, and applications. *Mass Spectrometry Reviews* **2009,** *29* (2), 294-312.

6. Shaw, J. B.; Lin, T. Y.; Leach, F. E., 3rd; Tolmachev, A. V.; Tolic, N.; Robinson, E. W.; Koppenaal, D. W.; Pasa-Tolic, L., 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer Greatly Expands Mass Spectrometry Toolbox. *J Am Soc Mass Spectrom* **2016,** *27* (12), 1929-1936.

7. 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.

8. Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246* (4926), 64-71.

9. Donnelly, D. P.; Rawlins, C. M.; Dehart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; Lantz, C.; Jung, W.; Johnson, K. R.; Koller, A.; Wolff, J. J.; Campuzano, I. D. G.; Auclair, J. R.; Ivanov, A. R.; Whitelegge, J. P.; Paša-Tolić, L.; Chamot-Rooke, J.; Danis, P. O.; Smith, L. M.; Tsybin, Y. O.; Loo, J. A.; Ge, Y.; Kelleher, N. L.; Agar, J. N., Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nature Methods* **2019,** *16* (7), 587-594.

10. Katta, V.; Chowdhury, S. K.; Chait, B. T., Use of a single-quadrupole mass spectrometer for collision-induced dissociation studies of multiply charged peptide ions produced by electrospray ionization. *Analytical chemistry* **1991,** *63* (2), 174-178.

11. McCormack, A. L.; Jones, J. L.; Wysocki, V. H., Surface-induced dissociation of multiply protonated peptides. *Journal of the American Society for Mass Spectrometry* **1992,** *3* (8), 859-862.

12. Brodbelt, J. S., Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chemical Society Reviews* **2014,** *43* (8), 2757-2783.

13. Li, H.; Nguyen, H. H.; Loo, R. R. O.; 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.

14. Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences* **2004,** *101* (26), 9528-9533.

15. 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. *Analytical chemistry* **2000,** *72* (3), 563-573.

16. 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. *Analytical chemistry* **2017,** *89* (5), 2731-2738.

17. Catherman, A. D.; Skinner, O. S.; Kelleher, N. L., Top Down proteomics: Facts and perspectives. *Biochemical and Biophysical Research Communications* **2014,** *445* (4), 683-693.

18. Zubarev, R., Protein primary structure using orthogonal fragmentation techniques in Fourier transform mass spectrometry. *Expert Review of Proteomics* **2006,** *3* (2), 251-261.

19. Zinnel, N. F.; Pai, P.-J.; Russell, D. H., Ion Mobility-Mass Spectrometry (IM-MS) for Top-Down Proteomics: Increased Dynamic Range Affords Increased Sequence Coverage. *Analytical Chemistry* **2012,** *84* (7), 3390-3397.

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

21. Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L., Analyzing internal fragmentation of electrosprayed ubiquitin ions during beam-type collisional dissociation. *Journal of the American Society for Mass Spectrometry* **2015,** *26* (5), 782-787.

22. Haverland, N. A.; Skinner, O. S.; Fellers, R. T.; Tariq, A. A.; Early, B. P.; LeDuc, R. D.; Fornelli, L.; Compton, P. D.; Kelleher, N. L., Defining Gas-Phase Fragmentation Propensities of Intact Proteins During Native Top-Down Mass Spectrometry. *J Am Soc Mass Spectrom* **2017,** *28* (6), 1203-1215.

23. Liu, X.; Inbar, Y.; Dorrestein, P. C.; Wynne, C.; Edwards, N.; Souda, P.; Whitelegge, J. P.; Bafna, V.; Pevzner, P. A., Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Molecular & Cellular Proteomics* **2010,** *9* (12), 2772-2782.

24. Carvalho, P. C.; Xu, T.; Han, X.; Cociorva, D.; Barbosa, V. C.; Yates III, J. R., YADA: a tool for taking the most out of high-resolution spectra. *Bioinformatics* **2009,** *25* (20), 2734-2736.

25. Strohalm, M.; Hassman, M.; Košata, B.; Kodíček, M., mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* **2008,** *22* (6), 905-908.

26. Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y.-B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L., ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic acids research* **2007,** *35* (suppl_2), W701-W706.

27. Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; Leduc, R. D.; Kelleher, N. L.; Thomas, P. M., ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *PROTEOMICS* **2015,** *15* (7), 1235-1238.

28. Kou, Q.; Xun, L.; Liu, X., TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **2016,** *32* (22), 3495-3497.

29. Liu, X.; Sirotkin, Y.; Shen, Y.; Anderson, G.; Tsai, Y. S.; Ting, Y. S.; Goodlett, D. R.; Smith, R. D.; Bafna, V.; Pevzner, P. A., Protein identification using top-down spectra. *Molecular & cellular proteomics* **2012,** *11* (6).

30. Wu, Z.; Roberts, D. S.; Melby, J. A.; Wenger, K.; Wetzel, M.; Gu, Y.; Ramanathan, S. G.; Bayne, E. F.; Liu, X.; Sun, R.; Ong, I. M.; McIlwain, S. J.; Ge, Y., MASH Explorer: A Universal Software Environment for Top-Down Proteomics. *J Proteome Res* **2020,** *19* (9), 3867-3876.

31. 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.

32. 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. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (9), 1896-1902.

33. 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.

34. Axelsson, J.; Palmblad, M.; Håkansson, K.; Håkansson, P., Electron capture dissociation of substance P using a commercially available Fourier transform ion cyclotron resonance mass spectrometer. *Rapid communications in mass spectrometry* **1999,** *13* (6), 474-477.

35. 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* (12), 5014-5030.

36. Levine, R. L.; Mosoni, L.; Berlett, B. S.; Stadtman, E. R., Methionine residues as endogenous antioxidants in proteins. *Proceedings of the National Academy of Sciences* **1996,** *93* (26), 15036-15040.

37. 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? *International Journal of Mass Spectrometry* **2005,** *240* (3), 273-284.

38. 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.

39. Kelleher, N. L., Peer Reviewed: Top-Down Proteomics. *Analytical Chemistry* **2004,** *76* (11), 196 A-203 A.

40. Chen, B.; Brown, K. A.; Lin, Z.; Ge, Y., Top-Down Proteomics: Ready for Prime Time? *Analytical Chemistry* **2018,** *90* (1), 110-127.

41. Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A., How many human proteoforms are there? *Nature chemical biology* **2018,** *14* (3), 206.

42. TurečEk, F., N-Cα Bond Dissociation Energies and Kinetics in Amide and Peptide Radicals. Is the Dissociation a Non-ergodic Process? *Journal of the American Chemical Society* **2003,** *125* (19), 5954-5963.

43. Zubarev, R. A., Electron-capture dissociation tandem mass spectrometry. *Current Opinion in Biotechnology* **2004,** *15* (1), 12-16.

SUPPORTING INFORMATION

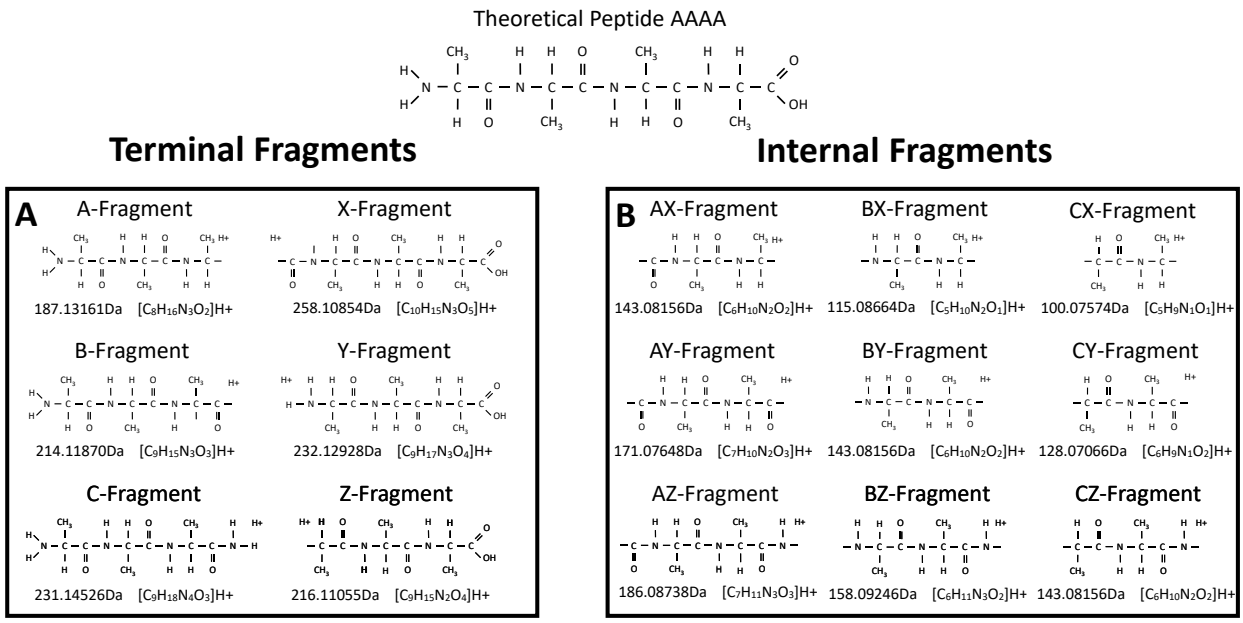


Figure S1: Examples of theoretical (A) terminal and (B) internal fragments and masses that are searched in ClipsMS.

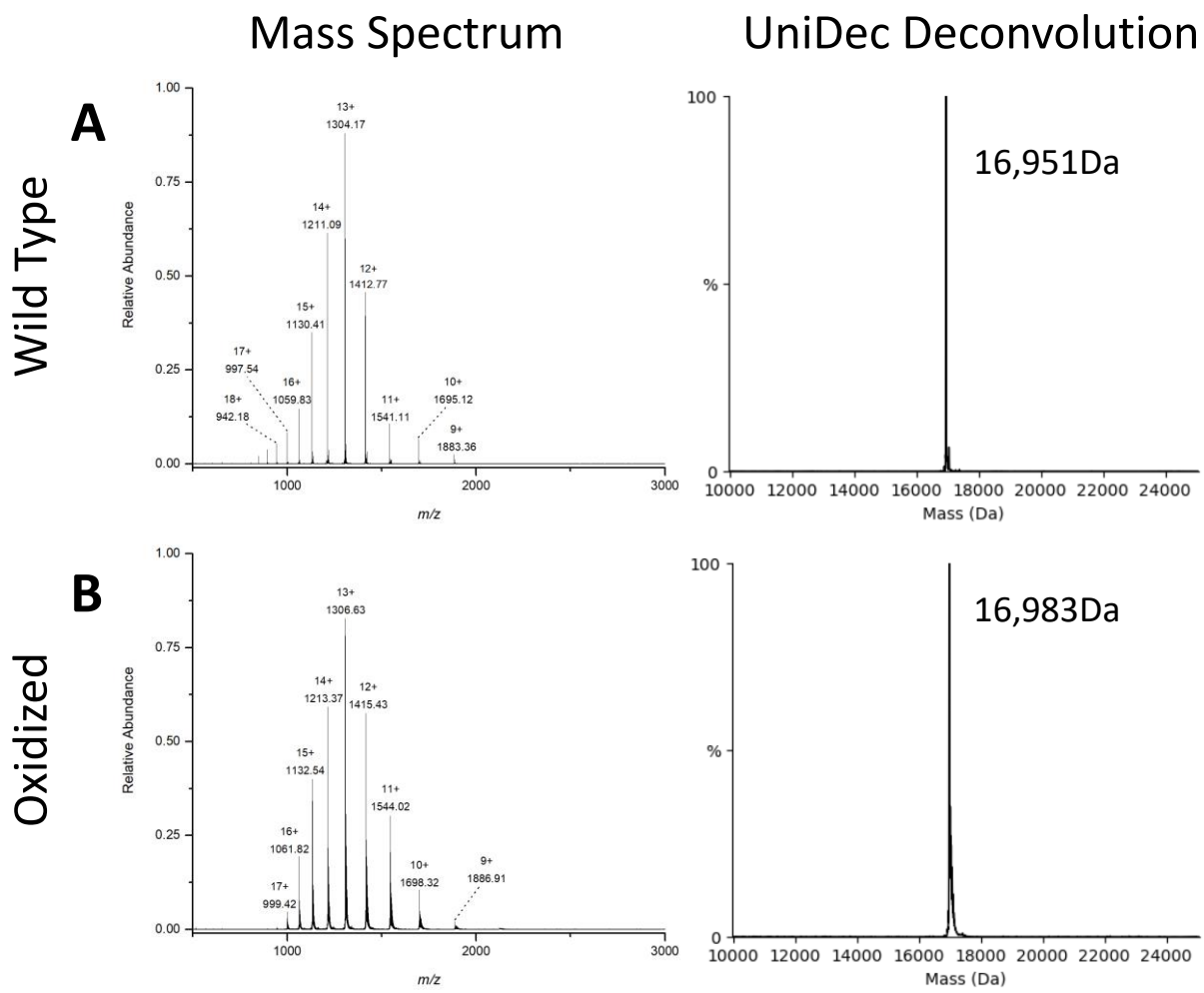


Figure S2: (A) MS1 spectrum of apo-myoglobin with the corresponding UniDec spectrum confirming the mass of wt apo-myoglobin (16,951Da), and (B) a spectrum of oxidized apo-myoglobin with the corresponding UniDec spectrum confirming the mass of oxidized apo-myoglobin (16,983 kDa).[1] The mass shift of 32 Da on oxidized apo-myoglobin indicates the presence of two oxidation sites on the protein.

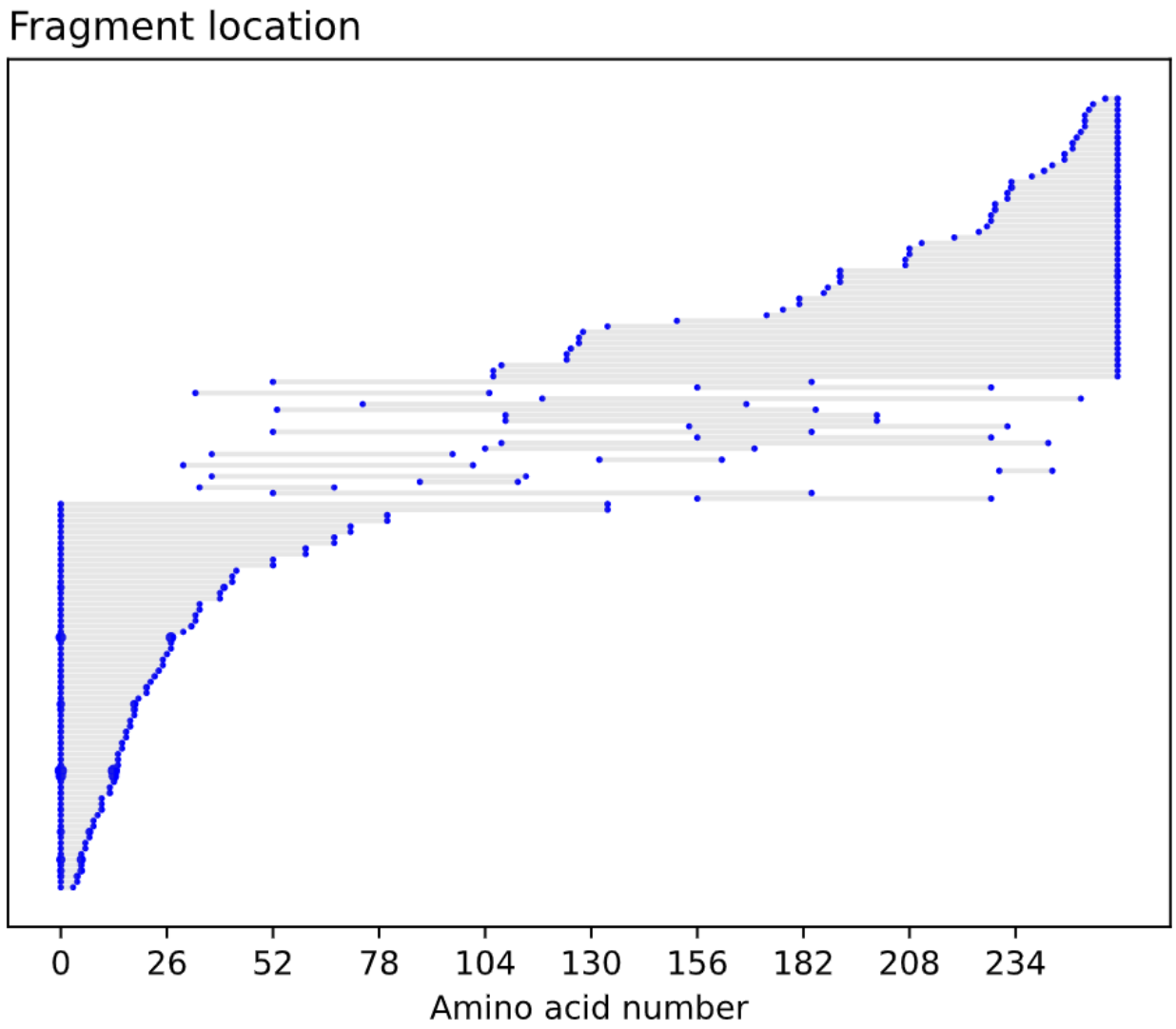


Figure S3: The fragment location figure for a top-down mass spectrum of carbonic anhydrase II. Internal fragments are shown to increase sequence coverage in the center of the protein sequence complementing the terminal fragments.

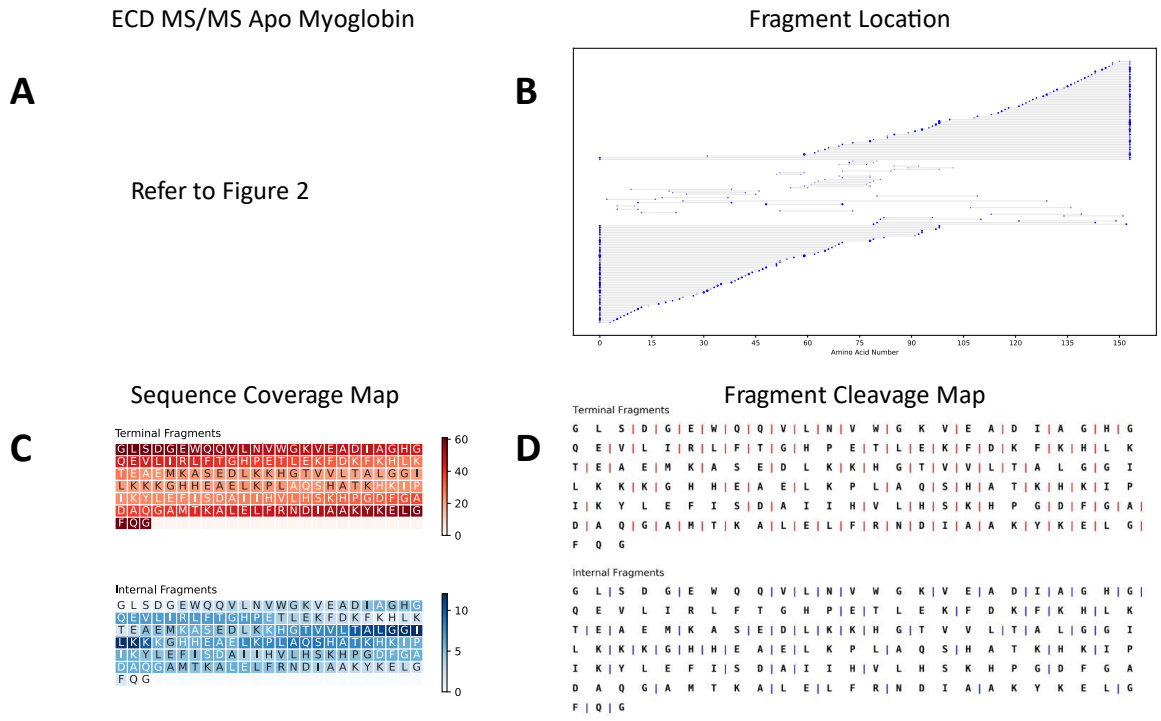


Figure S4: A. Broadband ECD MS of 20 uM apo-myoglobin formed from acidic denaturing conditions. B. A fragment location map indicating the region of the protein sequence covered by terminal and internal fragments including fragments with the addition of a hydrogen atom (1.00783Da). C. A sequence coverage map for the terminal and internal fragments also including fragments with the addition of a hydrogen atom (1.00783Da). Darker regions indicate more coverage. D. A fragment cleavage map indicating the location of inter-amino acid cleavage sites for terminal and internal fragments including fragments containing an additional hydrogen atom (1.00783Da).

Table S1. Amino acid residue masses used to calculate theoretical fragments in ClipsMS.[2]

Amino Acid	Molecular Weight
A	71.037114
R	156.101111
N	114.042928
D	115.026944
C	103.009185
E	129.042594
Q	128.058578
G	57.021464
H	137.058912
I	113.084064
L	113.084064
K	128.094963
M	131.040485
F	147.068414
P	97.052764
S	87.032029
T	101.047679
W	186.079313
Y	163.063329
V	99.068414

Table S2. Sample output of ClipsMS including the fragment type, localized modifications, unlocalized modifications, terminal modifications, observed mass, theoretical mass, start amino acid, end amino acid, error, sequence, intensity, and molecular formula.

Frag Type	Localized Mod	Unlocalized Mod	Term Mod	Observed Mass	Theoredical Mass	Start AA	End AA	Error
C Fragment	0	0	0	6675.415628	6675.409316	1	59	0.94556
C Fragment	0	0	0	3402.755447	3402.755322	1	31	0.03673
C Fragment	0	0	0	3246.652926	3246.654211	1	30	-0.39579
C Fragment	0	0	0	8647.638721	8647.629036	1	78	1.11996
C Fragment	0	0	0	4802.465087	4802.468927	1	43	-0.79959
C Fragment	0	0	0	1230.609567	1230.611331	1	11	-1.43343
C Fragment	0	0	0	447.2195446	447.21979	1	5	-0.54872
C Fragment	0	0	0	3820.973598	3820.976943	1	35	-0.87543
C Fragment	0	0	0	7353.79476	7353.790626	1	65	0.56216
C Fragment	0	0	0	3763.952	3763.955479	1	34	-0.92429
C Fragment	0	0	0	5457.813483	5457.813123	1	48	0.06596
C Fragment	0	0	0	4184.12768	4184.131213	1	38	-0.84438
C Fragment	0	0	0	4527.301635	4527.30555	1	41	-0.86475
C Fragment	0	0	0	5320.755681	5320.754211	1	47	0.27628
C Fragment	0	0	0	1344.654492	1344.654259	1	12	0.17328
C Fragment	0	0	0	4398.256538	4398.262956	1	40	-1.45921
C Fragment	0	0	0	5800.036508	5800.039829	1	51	-0.57258
C Fragment	0	0	0	890.3996115	890.400275	1	8	-0.74517
C Fragment	0	0	0	2114.062323	2114.066535	1	19	-1.99237
C Fragment	0	0	0	4917.496342	4917.495871	1	44	0.09578
C Fragment	0	0	0	10310.49998	10310.48808	1	93	1.15446
C Fragment	0	0	0	10875.83815	10875.82171	1	98	1.51179
C Fragment	0	0	0	7766.059605	7766.059197	1	69	0.05254
C Fragment	0	0	0	7454.839683	7454.838305	1	66	0.18485
C Fragment	0	0	0	2470.235109	2470.236121	1	23	-0.40968
C Fragment	0	0	0	10747.73932	10747.72675	1	97	1.17001
C Fragment	0	0	0	1913.985537	1913.986827	1	17	-0.67399
C Fragment	0	0	0	3662.904928	3662.9078	1	33	-0.78408
C Fragment	0	0	0	6388.294862	6388.297579	1	56	-0.42531
C Fragment	0	0	0	10173.43828	10173.42917	1	92	0.89596
C Fragment	0	0	0	5192.649935	5192.659248	1	46	-1.79349
C Fragment	0	0	0	2921.416012	2921.417669	1	27	-0.56719
C Fragment	0	0	0	7553.906906	7553.906719	1	67	0.02476
C Fragment	0	0	0	2342.174462	2342.177543	1	21	-1.31544
C Fragment	0	0	0	1117.526602	1117.527267	1	10	-0.59506
C Fragment	0	0	0	7031.625444	7031.615287	1	62	1.44448
C Fragment	0	0	0	762.3415546	762.341697	1	7	-0.18679
C Fragment	0	0	0	390.1985072	390.198326	1	4	0.46438
C Fragment	0	0	0	576.2620209	576.262384	1	6	-0.63009
C Fragment	0	0	0	4285.174033	4285.178892	1	39	-1.13391
C Fragment	0	0	0	9106.849359	9106.863287	1	82	-1.5294
C Fragment	0	0	0	1018.458448	1018.458853	1	9	-0.39766
C Fragment	0	0	0	5929.083304	5929.082423	1	52	0.14859
C Fragment	0	0	0	6129.17301	6129.162131	1	54	1.77496
C Fragment	0	0	0	7159.709381	7159.71025	1	63	-0.12137
C Fragment	0	0	0	2607.293471	2607.295033	1	24	-0.59909
C Fragment	0	0	0	275.1719229	275.171382	1	3	1.96568
C Fragment	0	0	0	4655.396281	4655.400513	1	42	-0.90905
C Fragment	0	0	0	7652.967559	7652.975133	1	68	-0.98968
C Fragment	0	0	0	3020.485035	3020.486083	1	28	-0.34696
C Fragment	0	0	0	3515.837402	3515.839386	1	32	-0.5643
C Fragment	0	0	0	1629.799519	1629.801986	1	14	-1.51368
Z Fragment	0	0	0	6633.499172	6633.499352	94	153	-0.02713
Z Fragment	0	0	0	6196.264101	6196.260684	98	153	0.55146
Z Fragment	0	0	0	2698.415629	2698.414716	130	153	0.33835

Z Fragment	0	0	0	6068.163995	6068.165721	99	153	-0.28444
Z Fragment	0	0	0	2083.07356	2083.073297	136	153	0.12626
Z Fragment	0	0	0	762.3913146	762.390658	147	153	0.86124
Z Fragment	0	0	0	925.4541015	925.453987	146	153	0.12372
Z Fragment	0	0	0	2395.290142	2395.289438	133	153	0.29391
Z Fragment	0	0	0	2755.438676	2755.43618	129	153	0.90585
Z Fragment	0	0	0	3965.966715	3965.972884	117	153	-1.55548
Z Fragment	0	0	0	1954.030364	1954.030703	137	153	-0.17349
Z Fragment	0	0	0	6857.599054	6857.590293	92	153	1.27756
Z Fragment	0	0	0	7508.010042	7508.00184	86	153	1.09243
Z Fragment	0	0	0	1423.733137	1423.734186	141	153	-0.73679
Z Fragment	0	0	0	3750.842668	3750.845892	119	153	-0.85954
Z Fragment	0	0	0	4103.026667	4103.031796	116	153	-1.25005
Z Fragment	0	0	0	9590.203303	9590.196803	66	153	0.67778
Z Fragment	0	0	0	2496.341564	2496.337117	132	153	1.78141
Z Fragment	0	0	0	3459.709664	3459.712752	122	153	-0.89256
Z Fragment	0	0	0	5744.936969	5744.944829	102	153	-1.36816
Z Fragment	0	0	0	3344.68812	3344.685808	123	153	0.69125
Z Fragment	0	0	0	3878.938215	3878.940855	118	153	-0.6806
Z Fragment	0	0	0	9912.384769	9912.372142	63	153	1.27386
Z Fragment	0	0	0	3069.556017	3069.558816	126	153	-0.91186
Z Fragment	0	0	0	1195.620887	1195.623178	143	153	-1.91616
Z Fragment	0	0	0	2954.528513	2954.531872	127	153	-1.1369
Z Fragment	0	0	0	1693.877693	1693.878225	139	153	-0.31407
Z Fragment	0	0	0	4749.445478	4749.448428	110	153	-0.62112
Z Fragment	0	0	0	3197.611844	3197.617394	124	153	-1.73567
Z Fragment	0	0	0	8835.752994	8835.737911	74	153	1.70704
Z Fragment	0	0	0	1308.706673	1308.707242	142	153	-0.43478
Z Fragment	0	0	0	9489.159251	9489.149124	67	153	1.06722
Z Fragment	0	0	0	16925.95941	16925.9536	1	153	0.34302
Z Fragment	0	0	0	6333.310883	6333.319596	97	153	-1.37574
Z Fragment	0	0	0	1537.776592	1537.777114	140	153	-0.33945
Z Fragment	0	0	0	1840.945689	1840.946639	138	153	-0.51604
Z Fragment	0	0	0	1124.585088	1124.586064	144	153	-0.86787
Z Fragment	0	0	0	3140.595896	3140.59593	125	153	-0.01083
Z Fragment	0	0	0	4315.183216	4315.184274	114	153	-0.24518
Z Fragment	0	0	0	9784.285342	9784.277179	64	153	0.8343
Z Fragment	0	0	0	7708.078568	7708.081548	84	153	-0.38661
Z Fragment	0	0	0	8168.267095	8168.26343	80	153	0.44869
Z Fragment	0	0	0	2196.154896	2196.157361	135	153	-1.12242
Z Fragment	0	0	0	4864.470212	4864.475372	109	153	-1.06075
Z Fragment	0	0	0	2627.37608	2627.377602	131	153	-0.57928
Z Fragment	0	0	0	505.2525515	505.253101	149	153	-1.08757
CZ Int Fragment	0	0	0	1125.592217	1125.593889	105	114	-1.48544
CZ Int Fragment	0	0	0	8109.234979	8109.250125	80	152	-1.86774
CZ Int Fragment	0	0	0	6987.649429	6987.663178	81	143	-1.96761
CZ Int Fragment	0	0	0	8776.734388	8776.724606	74	152	1.11454
CZ Int Fragment	0	0	0	3068.549188	3068.550991	122	149	-0.58757
CZ Int Fragment	0	0	0	3296.772817	3296.779016	82	110	-1.88032
CZ Int Fragment	0	0	0	1504.813291	1504.811822	83	96	0.9762
CZ Int Fragment	0	0	0	2009.089113	2009.085478	135	151	1.80928
CZ Int Fragment	0	0	0	2778.412683	2778.414436	114	139	-0.63094
CZ Int Fragment	0	0	0	1069.567559	1069.567674	13	22	-0.10752
CZ Int Fragment	0	0	0	3363.806619	3363.809331	44	74	-0.80623
CZ Int Fragment	0	0	0	2180.174515	2180.174371	53	73	0.06605
CZ Int Fragment	0	0	0	784.398022	784.398817	6	11	-1.01352
CZ Int Fragment	0	0	0	2998.489568	2998.483974	108	136	1.86561
CZ Int Fragment	0	0	0	671.3144366	671.314753	6	10	-0.47131

Sequence	Intensity	Formula
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASE	398886720	C300 H466 N81 O90 S1
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIR	277860608	C151 H237 N44 O46 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLI	198142896	C145 H225 N40 O45 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKK	194605216	C391 H627 N106 O113 S1
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKF	191463008	C218 H334 N59 O64 S0
GLSDGEWQQVL	177542192	C54 H84 N15 O18 S0
GLSDG	165365232	C17 H31 N6 O8 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTG	160233120	C172 H267 N48 O51 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHG	148447040	C330 H516 N91 O98 S1
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFT	140507136	C170 H264 N47 O50 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKH	139385312	C249 H379 N68 O71 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPE	135383504	C188 H288 N53 O56 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLE	122917280	C203 H313 N56 O62 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFK	104063632	C243 H372 N65 O70 S0
GLSDGEWQQVLN	89008488	C58 H90 N17 O20 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETL	86289096	C198 H306 N55 O59 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKT	77548496	C265 H409 N72 O75 S0
GLSDGEWQ	74813128	C38 H56 N11 O14 S0
GLSDGEWQQVLNVWGKVEA	72110784	C95 H145 N26 O29 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFD	72101728	C222 H339 N60 O67 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKGHHEAELK PLAQSH	71896952	C464 H741 N130 O134 S1
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKGHHEAELK PLAQSHATKHK	71311528	C489 H784 N139 O140 S1
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVL	69863056	C350 H552 N95 O103 S1
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGT	66992788	C334 H523 N92 O100 S1
GLSDGEWQQVLNVWGKVEADIAG	66749252	C110 H169 N30 O35 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKGHHEAELK PLAQSHATKH	63984388	C483 H772 N137 O139 S1
GLSDGEWQQVLNVWGKV	63687832	C87 H133 N24 O25 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLF	60314392	C166 H257 N46 O48 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMK	58699376	C289 H449 N78 O84 S1
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKGHHEAELK PLAQS	57872672	C458 H734 N127 O133 S1
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKF	46574104	C237 H360 N63 O69 S0
GLSDGEWQQVLNVWGKVEADIAGHGQE	46013272	C128 H194 N37 O42 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTV	44378872	C339 H532 N93 O101 S1
GLSDGEWQQVLNVWGKVEADI	40127252	C105 H161 N28 O33 S0
GLSDGEWQQV	38401180	C48 H73 N14 O17 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLK	38126776	C316 H494 N85 O95 S1
GLSDGEW	36999452	C33 H48 N9 O12 S0
GLSD	33560796	C15 H28 N5 O7 S0
GLSDGE	30180068	C22 H38 N7 O11 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPET	29675962	C192 H295 N54 O58 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKGHH	28001424	C411 H656 N115 O117 S1
GLSDGEWQQ	25162260	C43 H64 N13 O16 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTE	20696572	C270 H416 N73 O78 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAE	20345026	C278 H428 N75 O82 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKK	19655936	C322 H506 N87 O96 S1
GLSDGEWQQVLNVWGKVEADIAGH	13978781	C116 H176 N33 O36 S0
GLS	10444626	C11 H23 N4 O4 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEK	10153317	C209 H325 N58 O63 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVV	8005857	C344 H541 N94 O102 S1
GLSDGEWQQVLNVWGKVEADIAGHGQEV	7187482	C133 H203 N38 O43 S0
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRL	6456200	C157 H248 N45 O47 S0
GLSDGEWQQVLNVW	2774463	C74 H109 N20 O22 S0
ATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	245264208	C305 H474 N80 O84 S1
KIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	199922288	C286 H443 N73 O79 S1
AMTKALELFRNDIAAKYKELGFQG	199080704	C123 H195 N31 O35 S1
IPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	193644096	C280 H431 N71 O78 S1
ELFRNDIAAKYKELGFQG	150563984	C96 H146 N24 O28 S0
KELGFQG	140075296	C35 H54 N8 O11 S0

YKELGFQG	131402608	C44 H63 N9 O13 S0
KALELFRNDIAAKYKELGFQG	123460160	C111 H174 N28 O31 S0
GAMTKALELFRNDIAAKYKELGFQG	112769520	C125 H198 N32 O36 S1
SKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	100691576	C177 H272 N48 O54 S1
LFRNDIAAKYKELGFQG	92797088	C91 H139 N23 O25 S0
SHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	86512464	C314 H486 N84 O87 S1
LKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	86432040	C345 H540 N92 O94 S1
DIAAKYKELGFQG	80702824	C66 H101 N15 O20 S0
HPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	59094216	C168 H255 N45 O51 S1
HSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	58875760	C183 H279 N51 O55 S1
TVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	56931200	C439 H699 N119 O120 S1
TKALELFRNDIAAKYKELGFQG	55520912	C115 H181 N29 O33 S0
DFGADAQGAMTKALELFRNDIAAKYKELGFQG	55503328	C155 H238 N40 O48 S1
KYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	52084328	C263 H402 N68 O75 S1
FGADAQGAMTKALELFRNDIAAKYKELGFQG	49979400	C151 H233 N39 O45 S1
KHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	45241468	C174 H267 N47 O52 S1
KHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	42212112	C453 H721 N125 O123 S1
DAQGAMTKALELFRNDIAAKYKELGFQG	40466532	C137 H216 N36 O42 S1
AAKYKELGFQG	38369120	C56 H85 N13 O16 S0
AQGAMTKALELFRNDIAAKYKELGFQG	38083344	C133 H211 N35 O39 S1
RNDIAAKYKELGFQG	38052780	C76 H119 N21 O23 S0
AIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	37808856	C215 H333 N59 O61 S1
GADAQGAMTKALELFRNDIAAKYKELGFQG	36089288	C142 H224 N38 O44 S1
GILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	35859152	C404 H637 N111 O110 S1
IAAKYKELGFQG	31220382	C62 H96 N14 O17 S0
VVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	30659956	C435 H692 N118 O118 S1
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	29789670	C769 H1211 N209 O218 S2
HKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	28632288	C292 H450 N76 O80 S1
NDIAAKYKELGFQG	22523802	C70 H107 N17 O22 S0
FRNDIAAKYKELGFQG	22165450	C85 H128 N22 O24 S0
AKYKELGFQG	22041524	C53 H80 N12 O15 S0
ADAQGAMTKALELFRNDIAAKYKELGFQG	20005334	C140 H221 N37 O43 S1
VLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	18964876	C194 H299 N53 O57 S1
HGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	12819766	C447 H709 N123 O122 S1
AELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	12603535	C353 H552 N94 O98 S1
GHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	12411123	C372 H576 N102 O104 S1
LELFRNDIAAKYKELGFQG	11926676	C102 H157 N25 O29 S0
DAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG	7441562	C219 H338 N60 O64 S1
MTKALELFRNDIAAKYKELGFQG	6890706	C120 H190 N30 O34 S1
LGFQG	6823282	C24 H35 N5 O7 S0
EFISDAIIHV	86469552	C53 H81 N12 O15 S0
GHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQ	77332336	C370 H573 N102 O102 S1
HHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIA	58335840	C317 H494 N89 O88 S1
GILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQ	55233628	C402 H634 N111 O108 S1
DFGADAQGAMTKALELFRNDIAAKYKEL	31020380	C137 H215 N36 O42 S1
HEAELKPLAQSHATKHKIPIKYLEFISDA	25602240	C152 H239 N40 O42 S0
EAELKPLAQSHATK	15867519	C66 H110 N19 O21 S0
LELFRNDIAAKYKELGF	15034973	C95 H146 N23 O25 S0
VLHSKHPGDFGADAQGAMTKALELFR	14319574	C124 H193 N36 O35 S1
VWGKVEADIA	13795592	C50 H77 N12 O14 S0
DKFKHLKTEAEMKASEDLKKHGTVVLTALGG	12841460	C149 H248 N41 O45 S1
AEMKASEDLKKHGTVVLTALG	12616580	C95 H163 N26 O30 S1
EWQQVL	12085659	C37 H54 N9 O10 S0
SDAIIHVLHSKHPGDFGADAQGAMTKALE	7827239	C131 H205 N38 O41 S1
EWQQV	4593338	C31 H43 N8 O9 S0

Table S3: ClipsMS Processing Times for Myoglobin with Different Parameters						
Protein	Fragments Searched	PPM Error	Minimum Internal AA length	Localized Modification	Unlocalized Modification	Time(sec)
Myoglobin	c/z/cz	2	5	None	None	26.31
Myoglobin	c/z/cz	5	5	None	None	29.02
Myoglobin	c/z/cz	2	3	None	None	27.22
Myoglobin	c/z/cz	2	5	None	H+	49.6
Myoglobin	c/z/cz	5	5	None	H+	57.61
Myoglobin	c/z/cz	2	3	None	H+	51.84
Myoglobin	All Fragments	2	5	None	H+	427.95
Myoglobin	All Fragments	5	5	None	H+	636.12
Myoglobin	All Fragments	2	3	None	H+	483.52
Myoglobin	c/z/cz	2	5	2 Oxidation sites	H+	80.09
Myoglobin	All Fragments	2	5	2 Oxidation Sites	H+	657.99

REFERENCES

1. 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.

2. Haynes, W. M., *CRC handbook of chemistry and physics*. CRC press: 2014.

CHAPTER 3

Native Top-Down Mass Spectrometry with Collisionally Activated Dissociation Yields Higher-Order Structure Information for Protein Complexes

This work has been submitted for publication.

Carter Lantz[1], Benqian Wei[1], Boyu Zhao[1], Wonhyeuk Jung[1], Andrew K. Goring[1], Jessie Le[1], Justin Miller,[4] Rachel R. Ogorzalek Loo[1,3,4], Joseph A. Loo[1,2,3,4]

[1]Department of Chemistry and Biochemistry, University of California-Los Angeles, Los Angeles, CA, 90095; [2]Department of Biological Chemistry, University of California-Los Angeles, Los Angeles, CA, 90095; [3]UCLA-DOE Institute, University of California-Los Angeles, Los Angeles, CA, 90095; and [4]Molecular Biology Institute, University of California-Los Angeles, Los Angeles, CA, 90095

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

Keywords: Native Mass Spectrometry, Top-Down Mass Spectrometry (TDMS), Collisionally Activated Dissociation (CAD), High-Energy C-Trap Dissociation (HCD), Higher-Order Structure

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 and a faster timescale 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 (Fig. S1A); 24 *b*-fragments and 18 *y*-fragments resulted in 11.8% total sequence coverage. (Fig. 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 structure monomer such that subsequent HCD products yield little information about the 3D structure of the native tetramer.

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 (Fig. 1B, Fig. 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. (Fig. S2) nTDMS of ADH yielded 60 N-terminal *b*-fragments, but only 3 C-terminal *y*-fragments (17.6% sequence coverage). (Fig. 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. (Fig. 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). (Fig. 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 (Fig. S5 and Table S1), indicating direct fragmentation of the tetramer. nTDMS yielded 35 C-terminal *y*-fragments but only 8 N-terminal *b*-fragments (11.0% sequence coverage). (Fig. 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). (Fig. S6)

The HCD fragments from the aldolase tetramer mainly cover the solvent exposed C-terminus and are absent from the interface forming N-terminus. (Fig. 2) Our analysis indicates that fragments that cut at the interface of the tetramer (AA 110-224) only 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.

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. (Fig. S7A) In contrast, the native fragmentation spectrum of GSTA1 reveals 5 N-terminal *b*-fragments and 19 C-terminal *y*-fragments, (Fig. S7B) consistent with the GSTA1 crystal structure showing that the C-terminus is more solvent exposed than the N-terminus. (Fig. S7B) For the yeast enolase dimer, 27 *b*-fragments along with 18 *y*-fragments were measured by complex-down MS. (Fig. S8A) nTDMS revealed 48 N-terminal *b*-fragments along with 51 C-terminal *y*-fragments without the appearance of abundant monomer ions. (Fig. S8B) The crystal structure of enolase (Fig. 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 9 *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. (Fig. S9) Similarly, HCD of 6-phosphogluconate dehydrogenase (GND1) dimer generated 23 *b*-fragments but only 6 *y*-fragments, consistent with the GND1 crystal structure showing the N-termini to be solvent exposed and C-termini forming the dimer interface. (Fig. 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. (Fig. 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. (Fig. 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; Fig. 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; Fig. 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.

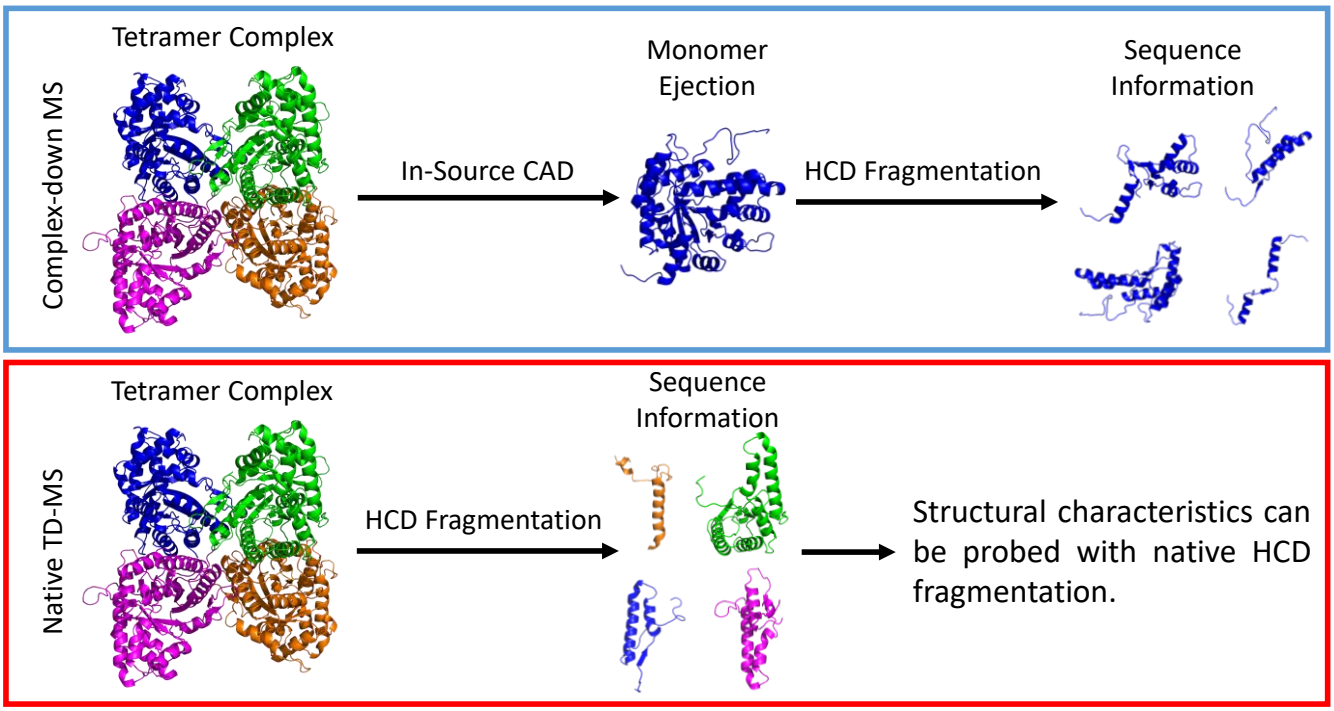
Monomer releases during HCD are not limited to membrane protein complexes. HCD fragmentation of the hemoglobin (Hb) tetramer revealed monomer and trimer peaks in addition to 10 *b*-fragments and 8 *y*-fragments from the α-subunit and 7 *b*-fragments and 7 *y*-fragments from the β-subunit. (Fig. S13A) 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 4 *y*-fragments from the β-subunit. (Fig. S13B) A similar HCD fragmentation pattern can be observed from complex-down MS of individual subunits (9 *b*- and *y*-fragments, and 6 *b*- and 5 *y*-fragments from the α- and β-subunits, respectively). (Fig. S13C) Similarly, nTDMS of human transthyretin (TTR) tetramers by HCD releases monomer products in addition to 2 *b*- and 38 *y*-fragments. (Fig. 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 3 *b*-fragments and 41 *y*-fragments. (Fig. 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 shows that HCD fragmentation of ADH tetramers reveals numerous internal fragments spanning residues 178-236, (Fig. S15A) which correspond to a solvent exposed region. (Fig. 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 timescale, 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 timeframe, 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.

FIGURES



Scheme 1: Complex-down MS and nTDMS workflows used in this study.

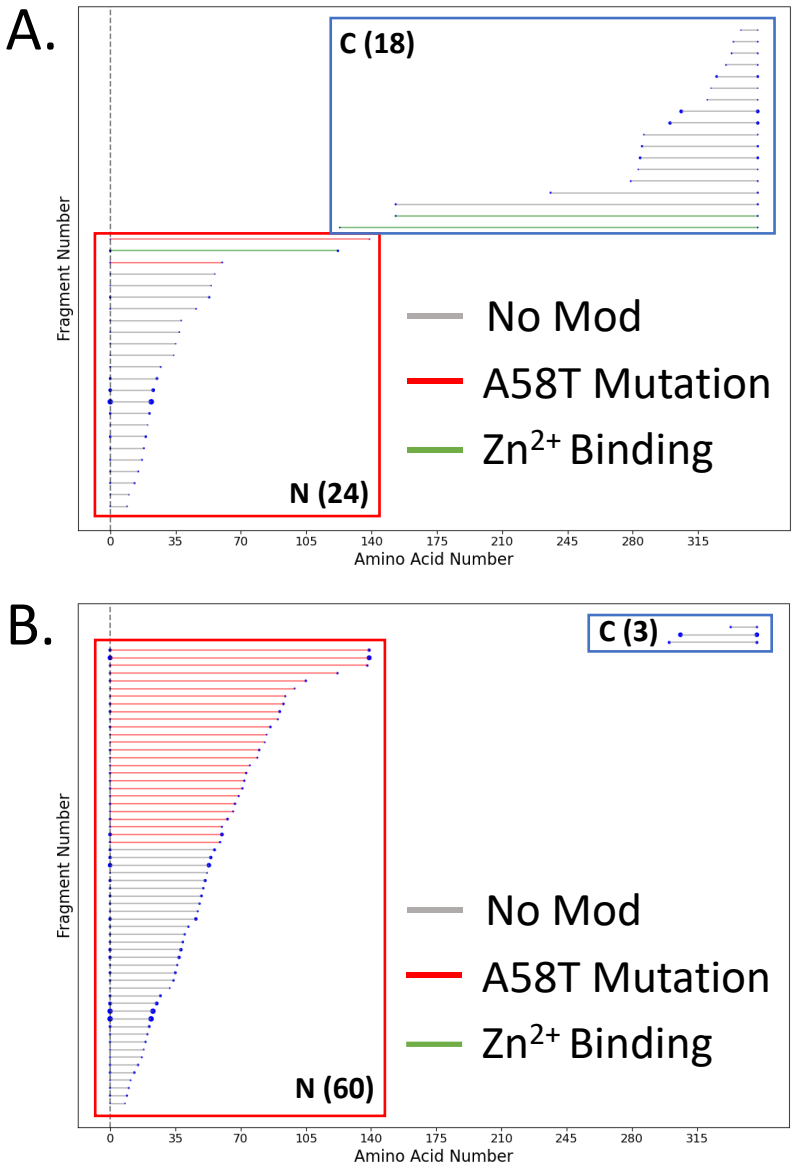


Figure 1: Fragment location maps for ADH representing *b*-/*y*-product ions measured by (A) complex-down MS and (B) nTDMS with HCD. Red lines indicate V58T mutation, green lines indicate Zn^{2+} binding, the vertical dotted line indicates N-terminal acetylation, and the size of the blue dots indicates the relative intensity of each fragment. Numbers in parentheses indicate the number of product ions detected.

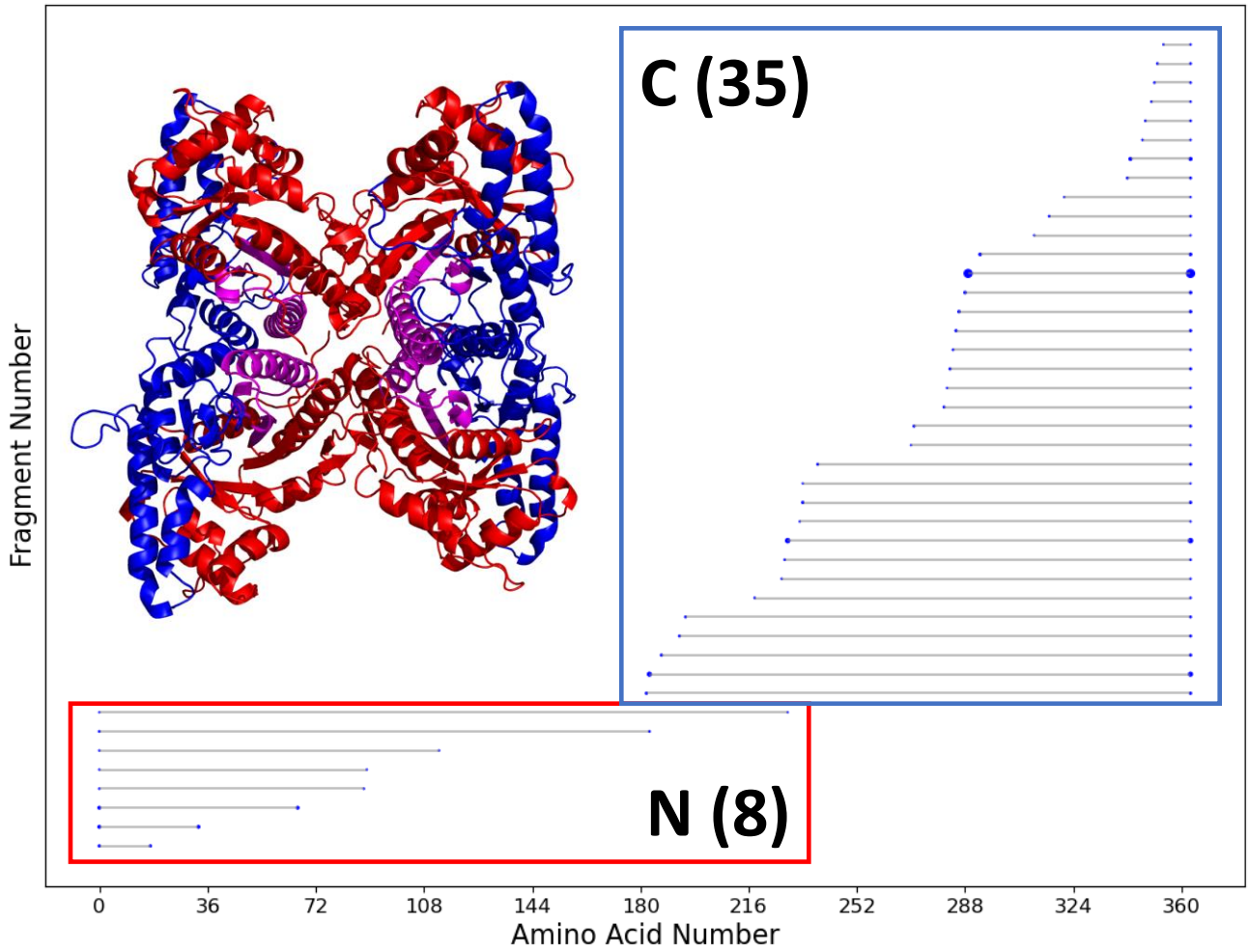


Figure 2: Fragment location map for nTDMS products of the 25+ charged precursor of aldolase homotetramer, with the size of the blue dots corresponding to the relative intensity of the fragments The crystal structure shows that most cleavage sites lie on the solvent exposed C-terminus (blue), rather than the interface forming N-terminus (red). The purple region is covered by both N-terminal and C-terminal fragments.

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.

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.

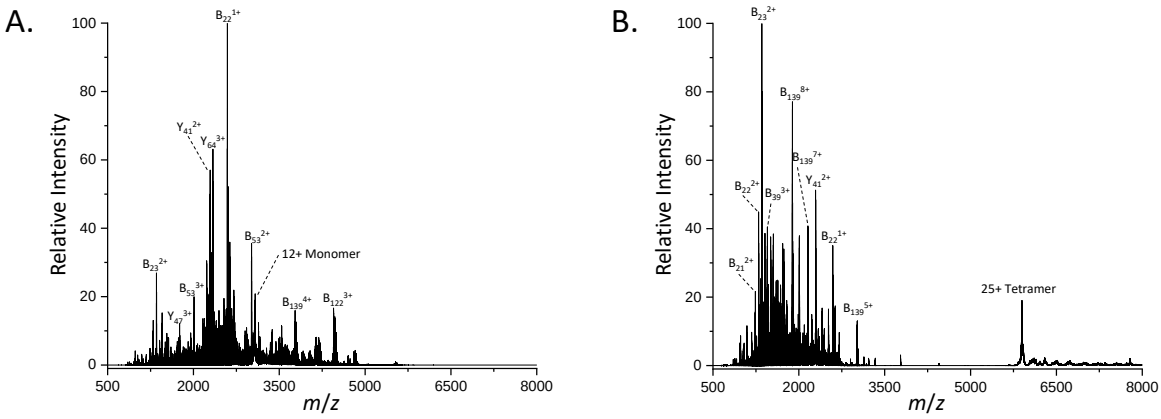


Figure S1: (A) Complex-down fragmentation mass spectrum and (B) native TDMS spectrum of ADH.

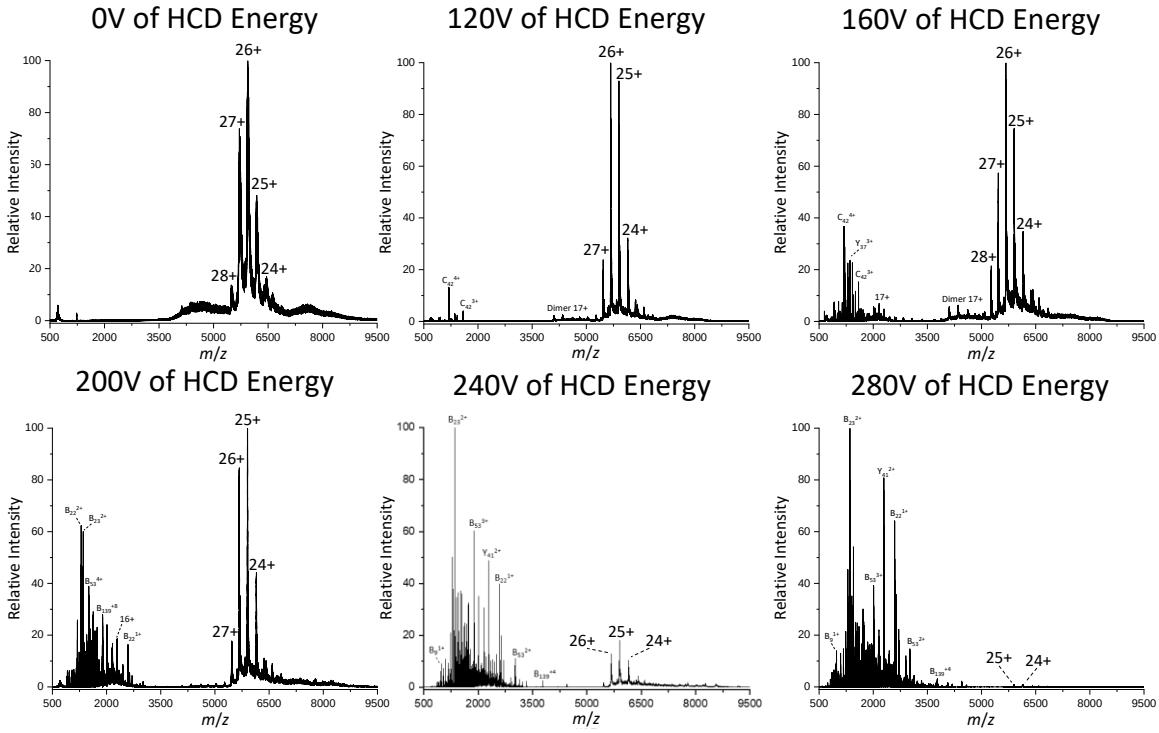


Figure S2: Broadband nTDMS spectra of ADH at various HCD voltages.

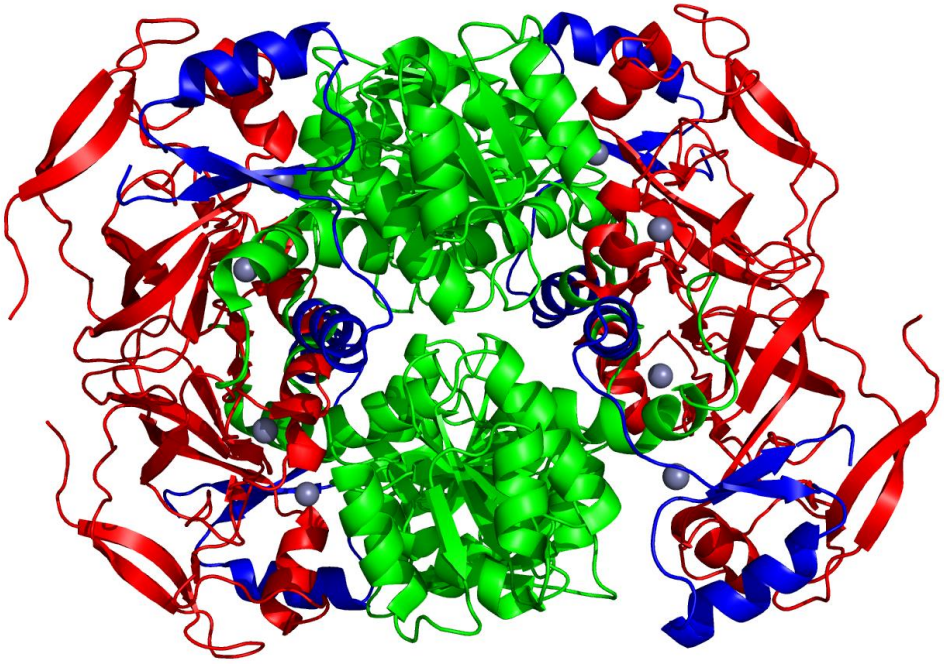


Figure S3: The structure of ADH with the region covered by the N-terminal fragments labeled in red and the region covered by the C-terminal fragments labeled in blue. Fragmentation occurs in the solvent exposed regions and does not occur in the subunit-subunit interface region (green).

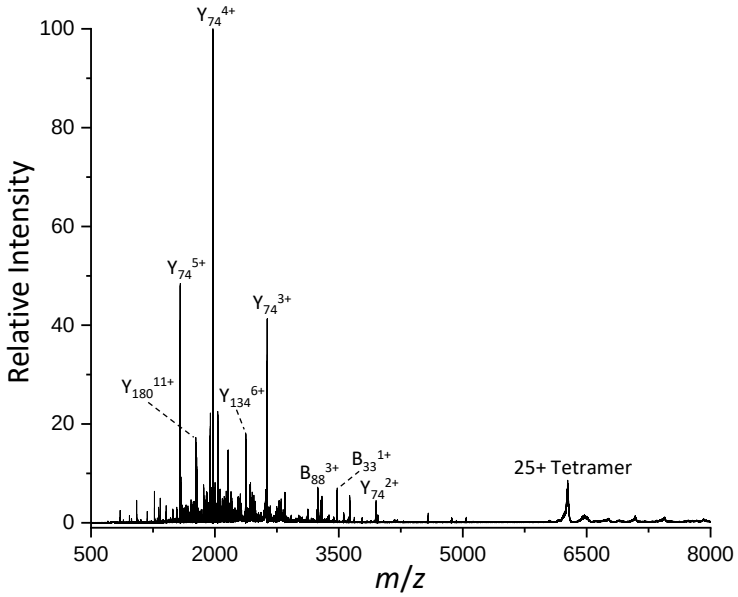


Figure S4: Native top-down mass spectrum of the 25+ charged aldolase homotetramer.

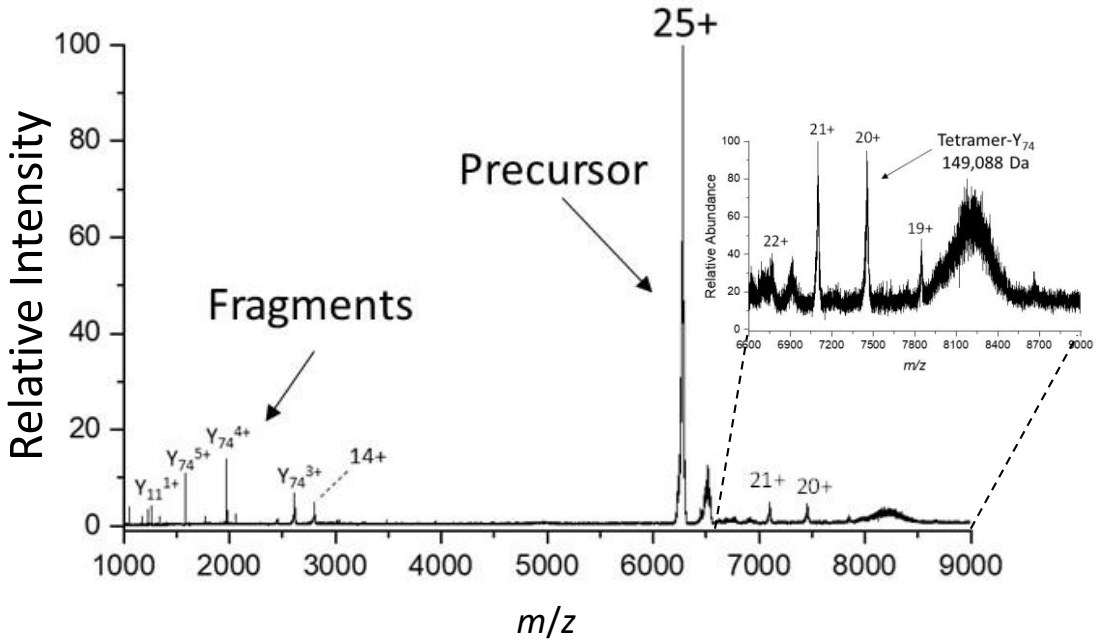


Figure S5: Native top-down mass spectrum of the 25+ charge state of the aldolase homotetramer showing multiple charge states of an abundant y_{74} fragment and high *m/z* peaks corresponding to charge states of the (4M - y_{74}) product ion.

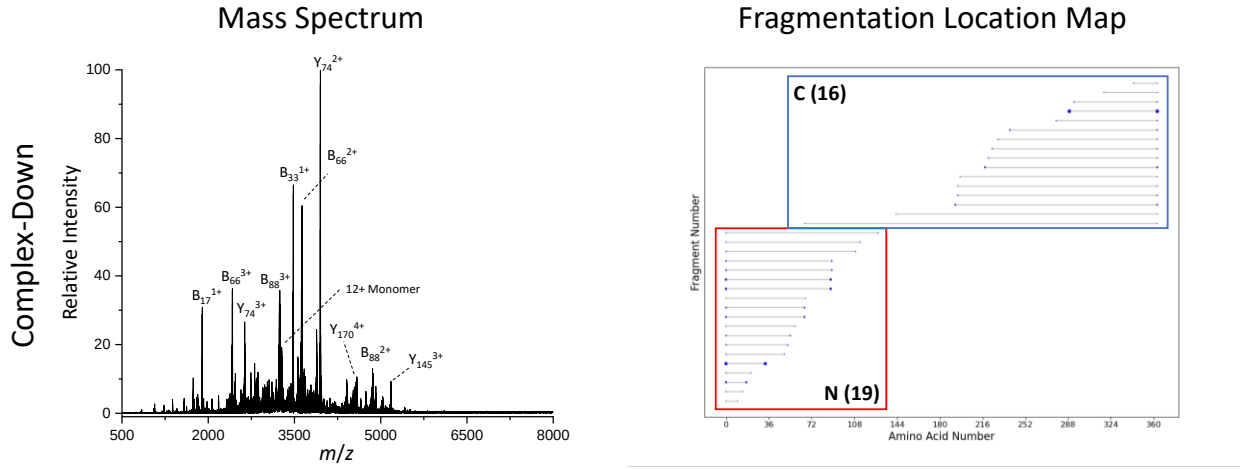


Figure S6: A complex-down mass spectrum (12+ monomer) and the corresponding fragment location map for aldolase.

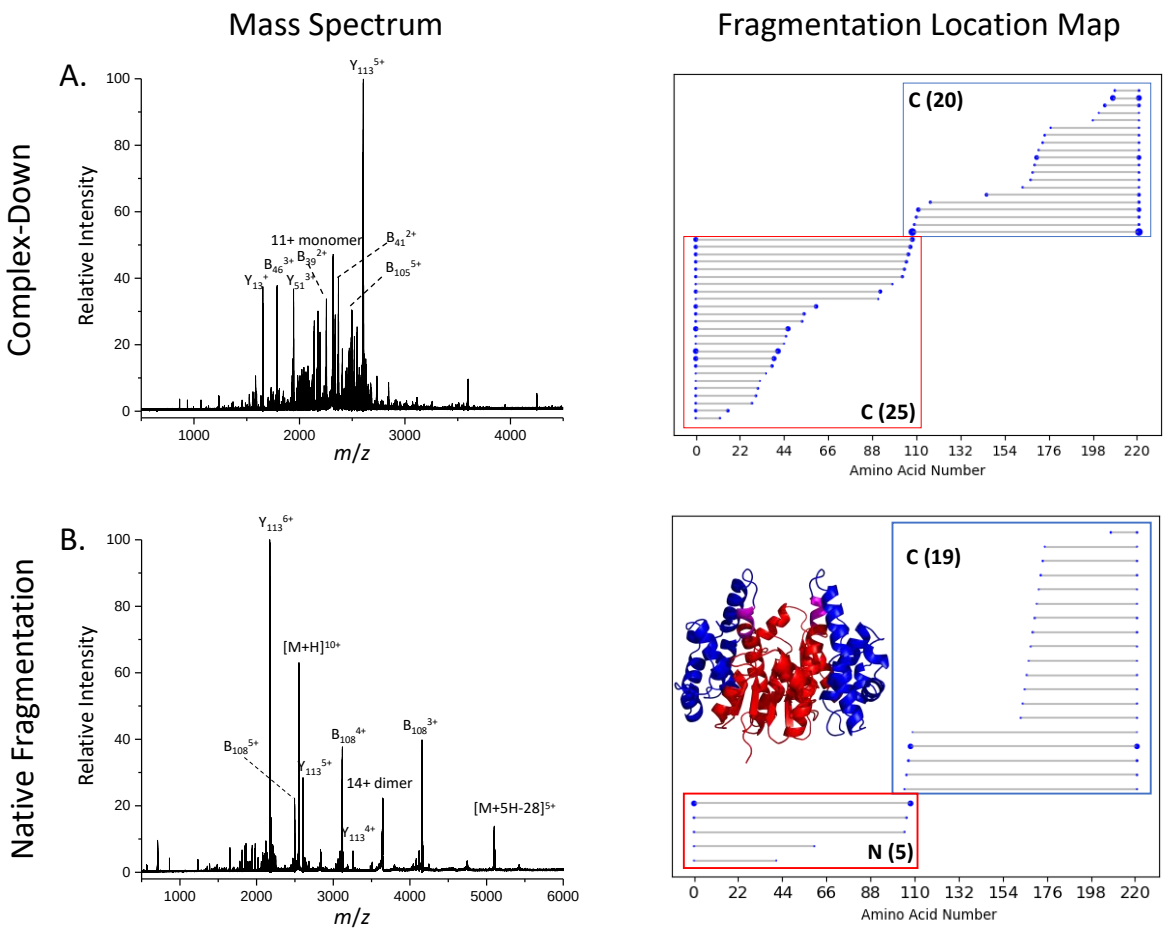


Figure S7: (A) A complex-down mass spectrum (11+ monomer) with the corresponding fragment location map and (B) a native top-down mass spectrum with the corresponding fragmentation location map for the 12+ charged human GST A1 dimer. The inset shows the structure of GST A1 with the region covered by N-terminal 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.

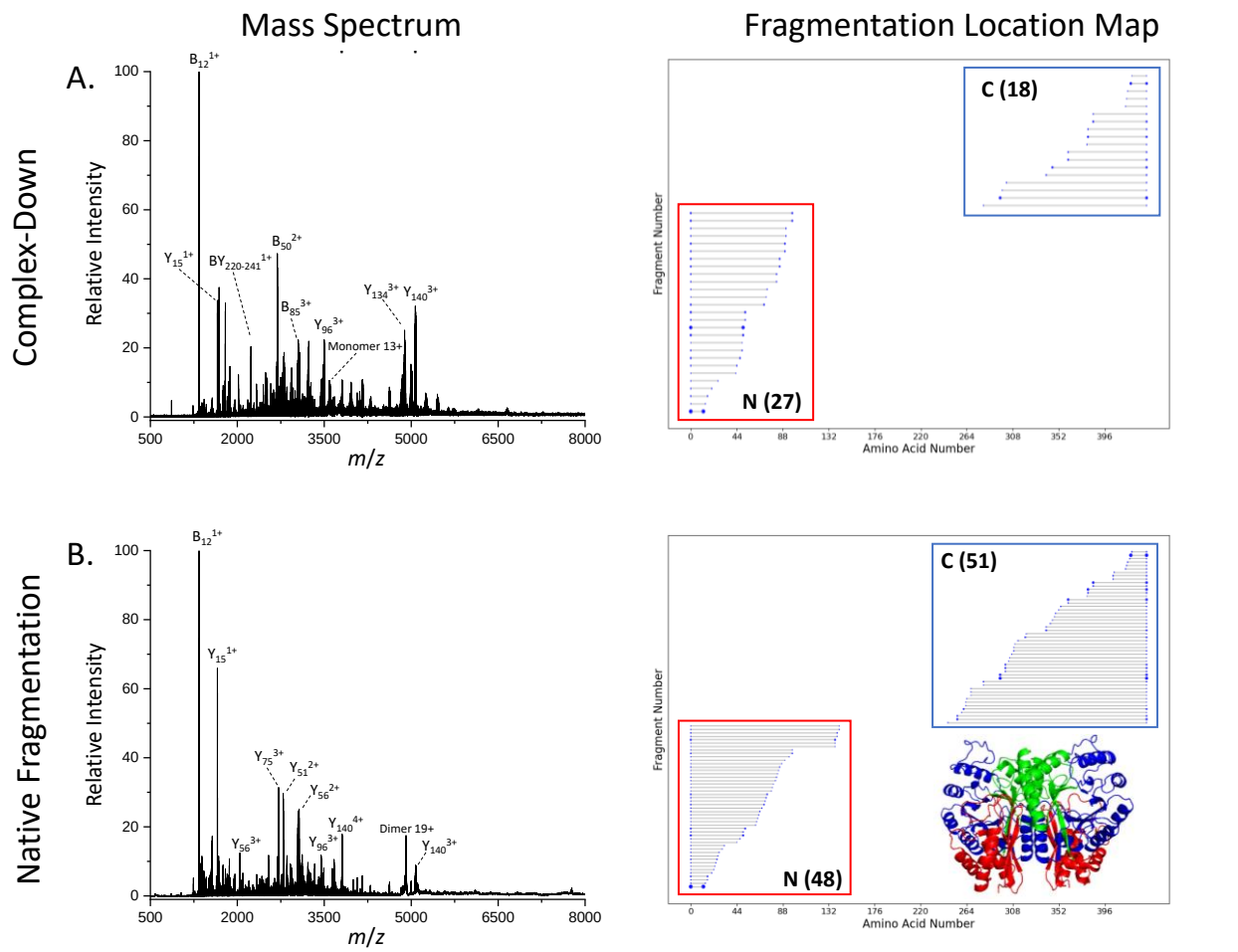


Figure S8: (A) A complex-down mass spectrum (13+ monomer) with the corresponding fragment location map and (B) a native top-down mass spectrum with the corresponding fragmentation location map for the 19+ charged enolase dimer. The inset shows the structure of enolase with the region covered by N-terminal fragments colored in red and the region covered by C-terminal fragments colored in blue (and green indicates no fragment coverage).

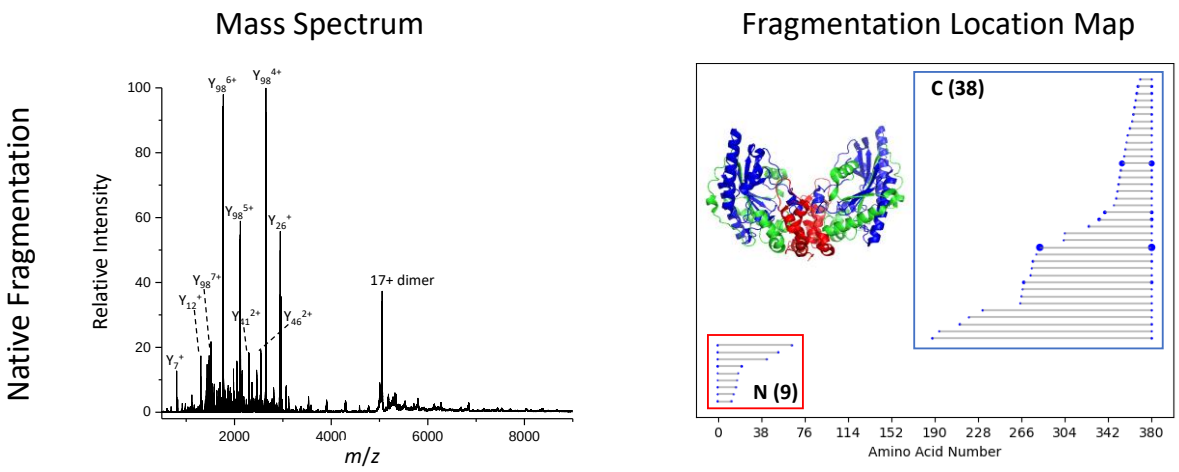


Figure S9: A native top-down mass spectrum with the corresponding fragmentation location map for 17+ charged creatine kinase dimer. The inset shows the structure of creatine kinase with the region covered by N-terminal fragments colored in red and the region covered by C-terminal fragments colored in blue (and green indicates no fragment coverage).

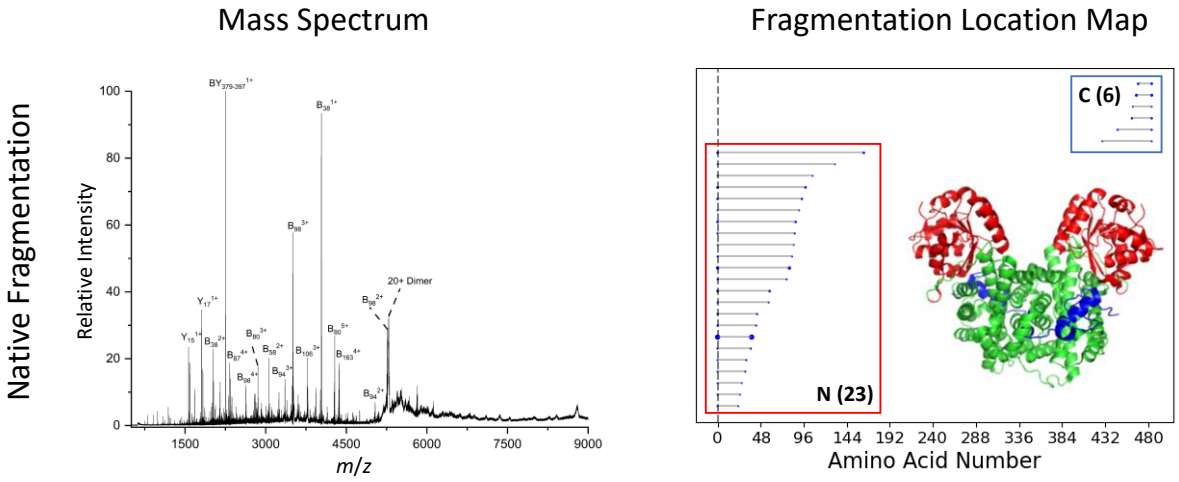


Figure S10: A native top-down mass spectrum with the corresponding fragmentation location map for the 20+ charged GND1 dimer with the vertical dotted line representing N-terminal acetylation. The inset shows the structure of GND1 with the region covered by N-terminal fragments colored in red and the region covered by C-terminal fragments colored in blue (and green indicates no fragment coverage).

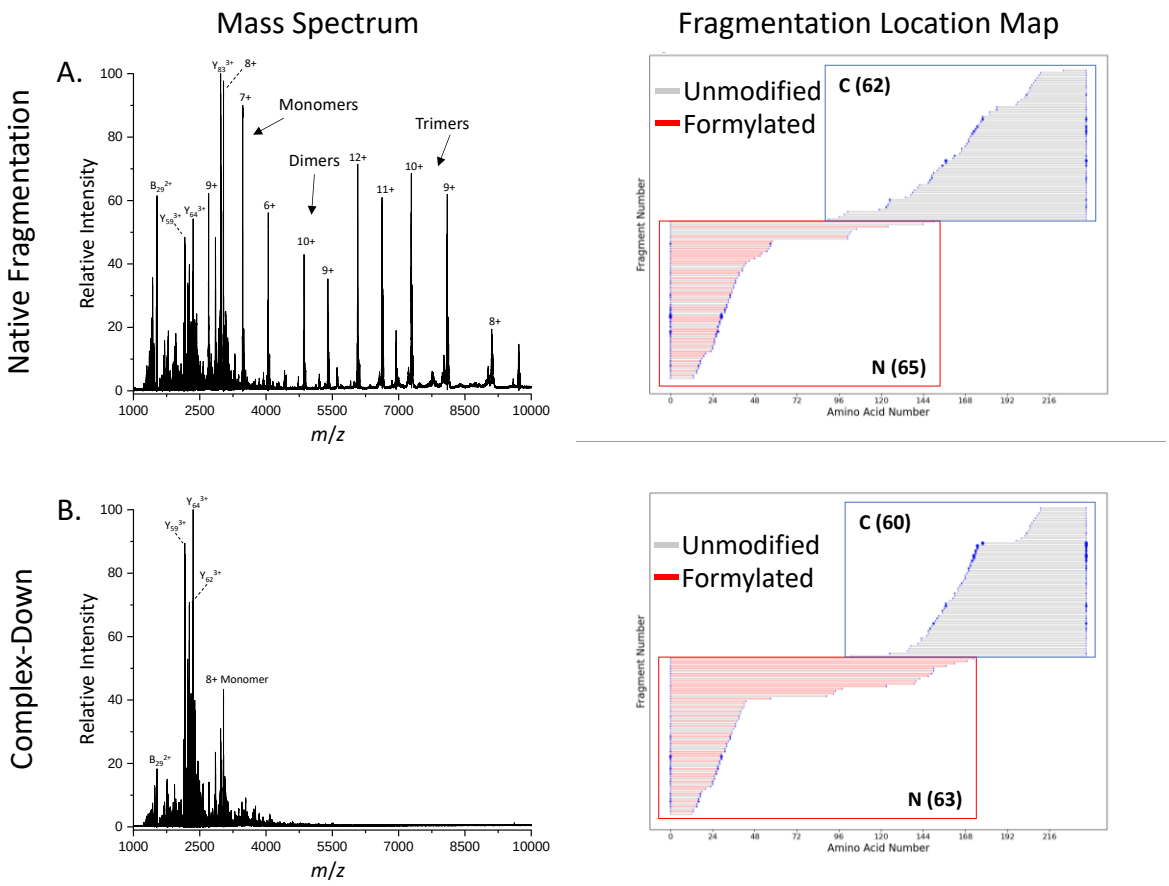


Figure S11: (A) A native top-down mass spectrum (18+ charged tetramer) with the corresponding fragment location map and (B) a complex-down mass spectrum (8+ monomer) with the corresponding fragmentation location map for AqpZ.

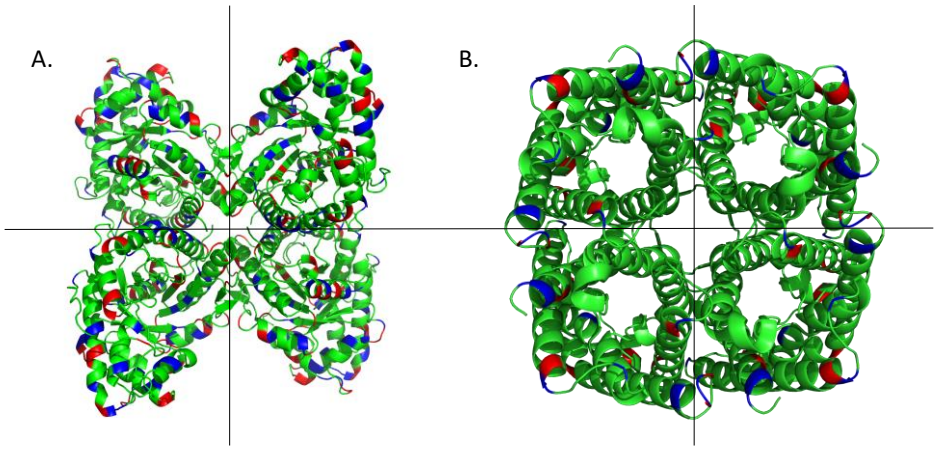


Figure S12: (A) The crystal structures of rabbit aldolase and (B) aquaporin Z with positively charged amino acids (Lys and Arg) labeled in blue and negatively charged amino acids (Glu and Asp) labeled in red. The black lines indicate the complex interface. Aldolase contains many charged residues at the 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.

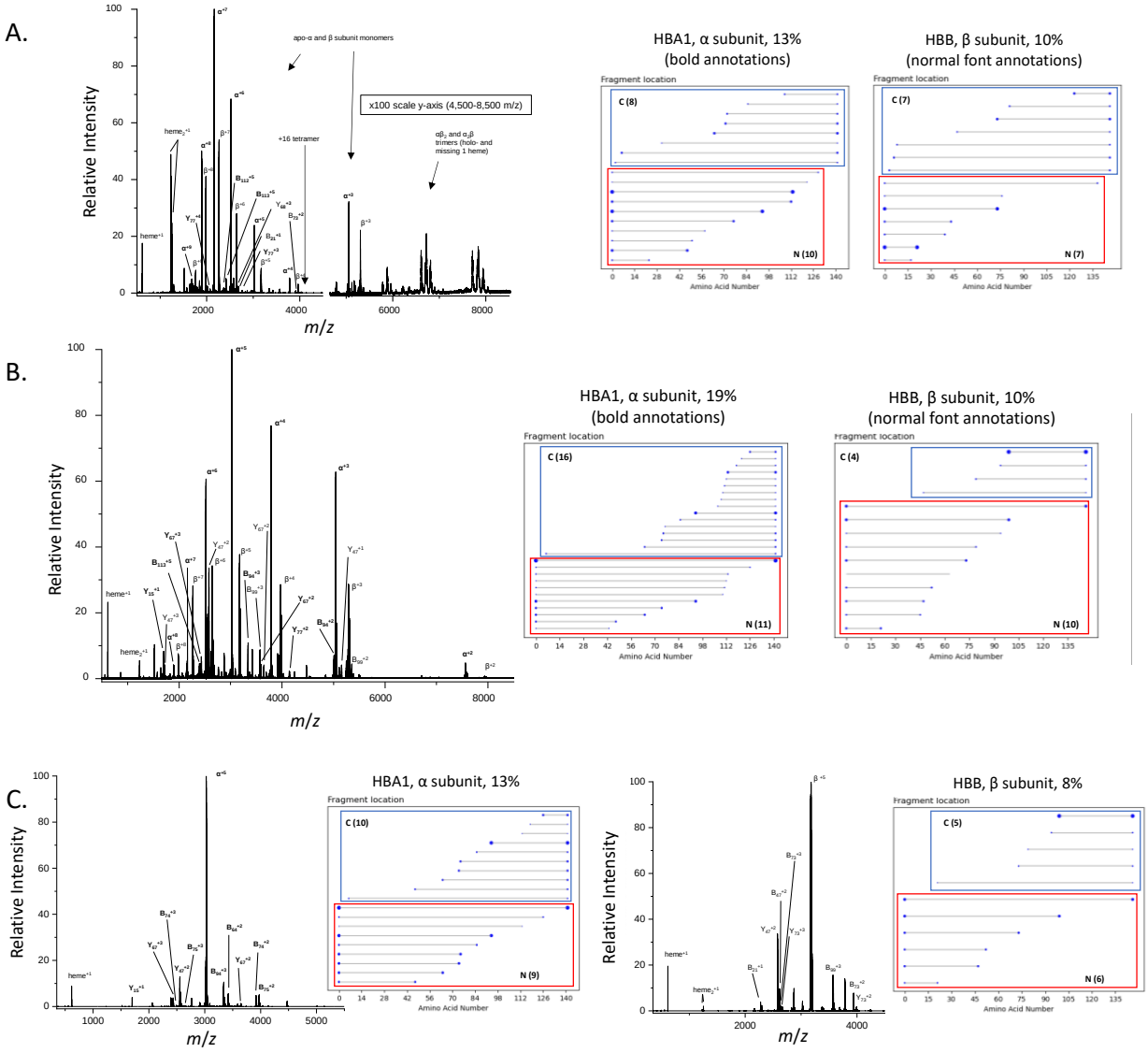


Figure S13: (A) nTDMS spectrum of the 16+ charged hemoglobin tetramer with the corresponding fragment location maps for the α-subunit and β-subunit, (B) a nTDMS mass spectrum of the 12+ charged hemoglobin dimer with the corresponding fragmentation location maps for the α-subunit and β-subunit, and (C) complex-down fragmentation mass spectra and the corresponding fragment location maps for the 6+ charged α-subunit and β-subunit.

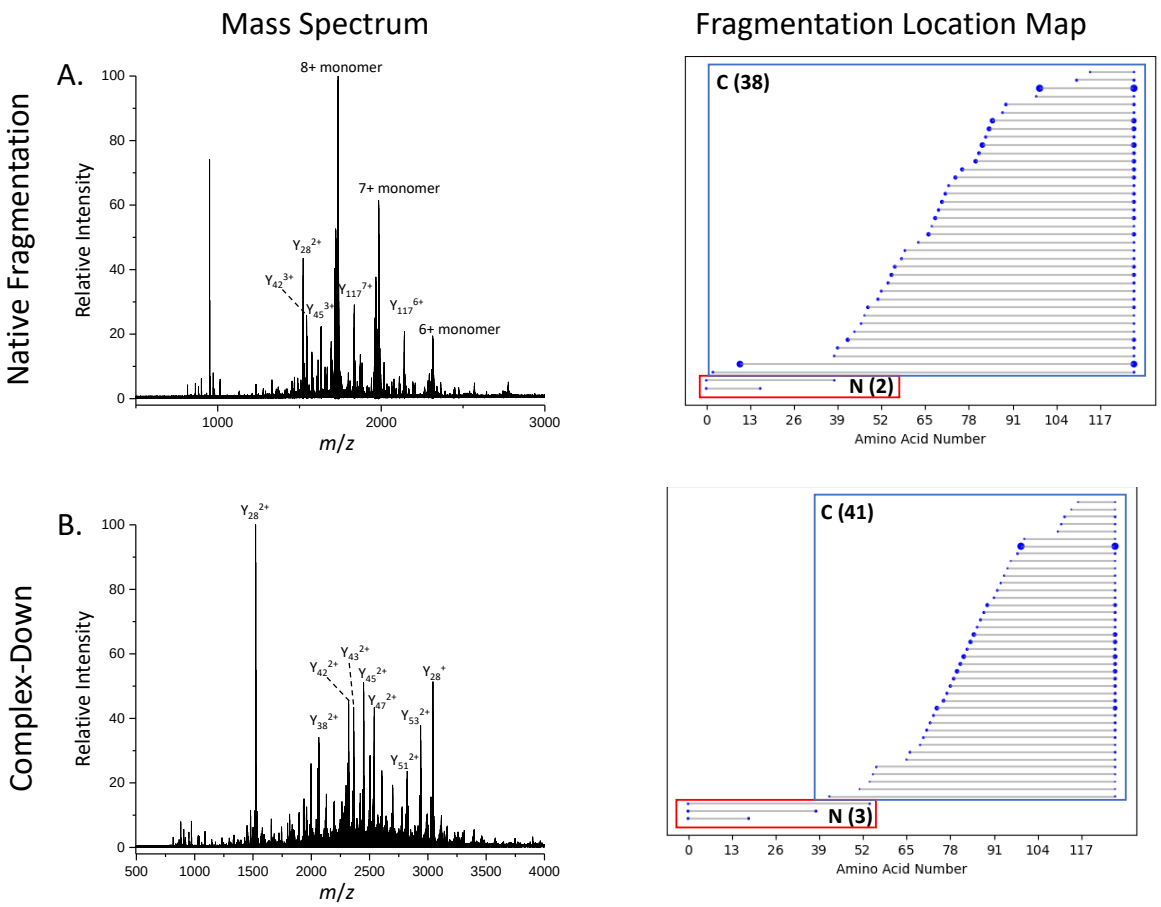


Figure S14: (A) Native top-down mass spectrum (15+ tetramer) with the corresponding fragment location map and (B) a complex-down mass spectrum (6+ monomer) with the corresponding fragmentation location map for the TTR tetramer.

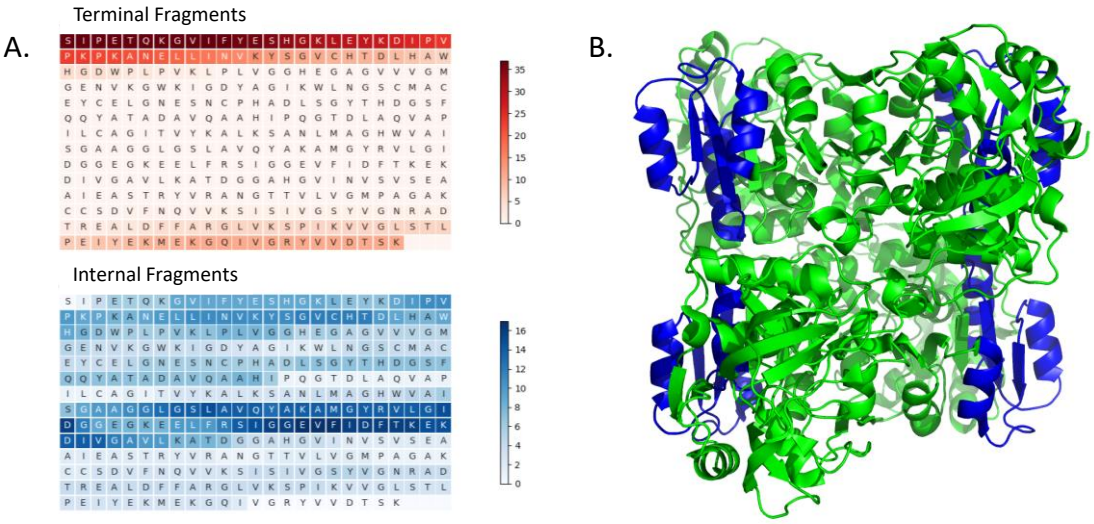


Figure S15: (A) A heatmap representing terminal and internal fragment analysis of the 24+ charged ADH tetramer and (B) the structure of the ADH tetramer with an internal fragment hotspot (residues 178-236) highlighted in blue. Notice how this region of the ADH tetramer is solvent exposed.

Supplementary Tables:

Table S1: Molecular weights of species present in a low HCD 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.			
General Inforamtion			
Name of Complex	Type of Complex	MW Complex (Da)	MW Monomer (Da)
ADH	Tetramer	147,472	36,738
Aldolase	Tetramer	156,748	39,187
GSTA1	Dimer	51,000	25,500
Enolase	Dimer	93,312	46,656
Creatine Kinase	Dimer	86,224	43,112
GND1	Dimer	106,003	52,957
AqpZ	Tetramer	24,269	97,076
Hemoglobin	Tetramer	61,986	α=15,126 β=15,867
Hemoglobin	Dimer	30,993	α=15,126 β=15,867
Transthyretin	Tetramer	55,044	13,761

	Native TD-MS Fragmentation			
Name of Complex	Monomer Release with HCD	# N-Term Frags	# C-Term Frags	Sequence Coveage
ADH	No	60	3	18%
Aldolase	No	8	35	11%
GSTA1	No	5	19	11%
Enolase	No	48	51	18%
Creatine Kinase	No	9	38	13%
GND1	No	23	6	6%
AqpZ	Yes	65	62	38%
Hemoglobin	Yes	α=10 β=7	α=8 β=7	α=13% β=10%
Hemoglobin	Yes	α=11 β=10	α=16 β=4	α=19% β=10%
Transthyretin	Yes	2	38	33%

	Complex-Down Fragmentation			
Name of Complex	Monomer Release with CAD	# N-Term Frags	# C-Term Frags	Sequence Coverage
ADH	Yes	24	18	12%
Aldolase	Yes	19	16	8%
GSTA1	Yes	25	20	21%
Enolase	Yes	27	18	8%
Creatine Kinase	No	NA	NA	NA
GND1	No	NA	NA	NA
AqpZ	Yes	63	60	35%
Hemoglobin	Yes	α=9 β=6	α=10 β=6	α=13% β=8%
Hemoglobin	Refer to row above	Refer to row above	Refer to row above	Refer to row above
Transthyretin	Yes	3	41	36%

REFERNCES

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.

CHAPTER 4

Native Top-Down Mass Spectrometry with Electron Capture Dissociation on an Orbitrap Mass Spectrometer Yields Sequence and Higher-Order Structure Information for Proteins and Protein Complexes

ABSTRACT: Electron capture dissociation (ECD) is a nonergodic top-down fragmentation technique where electrons directly interact with proteins and protein complexes to produce covalent fragments. Conventionally, ECD has been performed on Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers because the magnet used to radially trap ions can also be utilized to focus an electron beam. However, ECD has recently been adapted to work on other types of instruments including orbitrap mass spectrometers. With this new development it is possible to obtain much of the same information that can be collected with FT-ICR instruments. This study reveals that ECD on an orbitrap-based instrument can provide sequence information for the NIST monoclonal antibody. The sequence coverage for the heavy chain was 31.60% and the sequence for the light chain was 10.94%. Fragments from ECD fragmentation can reveal the location of modifications along the protein sequence. Fragmentation of phosphorylated tau protein reveal the locations of the phosphorylation sites to be at the boundary of the microtubule binding domain and the proline rich domain and the C-terminal domain. ECD of large protein complexes can reveal regions that are solvent exposed regions that form the interface. The research here indicates that ECD of the aldolase homotetramer produces fragments from the solvent exposed C-terminus and are absent from the interface forming N-terminus.

Keywords: Native Mass Spectrometry, Top-Down Mass Spectrometry (TDMS), Electron Capture Dissociation (ECD), Electron Ionization Dissociation (EID), Higher-Order Structure

MAIN TEXT:

Top-down mass spectrometry, where an intact protein is dissociated into fragments and analyzed by mass spectrometry, has become a growing field for the analysis of proteins. Top-down mass spectrometry can be performed with collision-, electron-, or photon-based techniques and can be utilized to

sequence proteins,[1] locate modifications,[2] and return structure information.[3, 4] It has been found that electron-based techniques are uniquely suited to obtain structure information as the nonergodic nature of the technique dissociates peptide bonds without disturbing the structure of the intact protein. Electron capture dissociation (ECD) has conventionally been relegated to Fourier transform-ion cyclotron resonance instruments because the magnet can be used to focus the electron beam while simultaneously trapping protein ions.[5] However, ECD has recently been adapted to work on an orbitrap mass spectrometer. In this work, ECD on a UHMR instrument was performed to obtain sequence, modification, and structure information on a variety of proteins and proteins complexes.

To demonstrate the utility of ECD on an orbitrap-based mass spectrometer, ECD was performed on an intact native antibody. Antibodies have traditionally been sequenced with a bottom-up approach or a middle-down approach due to the size of the complex;[6] however, we have found that ECD with an orbitrap based instrument can reveal extensive sequence coverage of these molecules. ECD of a native mAb reveals numerous *c*/*z* fragments from the light chain and the heavy chain that can be mapped back to their respective sequence as well as charge reduced precursors. (Fig. 1A) The sequence coverage for the light chain is 32% (Fig. 1B) and the sequence coverage for the heavy chain is 11%. (Fig. 1C) The light chain has many more N-terminal fragments than C-terminal fragments since there is an interchain disulfide bond that requires breaking before a C-terminal fragment is released. The heavy chain contains modifications that and be mapped back to the sequence including an N-terminal pyroglutamic acid and a G1F glycan modification. Performing ECD on native antibodies could be beneficial for increasing the signal to noise of fragment peaks. Most antibody analysis is performed under denaturing conditions. However, the accumulation of charge under denaturing conditions leads to spread of the charge state distribution and loss of signal to noise. Fragmentation of native antibodies allows for the consolidation of charge, higher signal to noise for fragment peaks, and potentially higher sequence coverage for the complex.

ECD fragmentation with subsequent high energy C-trap dissociation (HCD) of the ECD products can also provide sequence information on protein complexes. To determine the types of products that result from this workflow a three-step experiment was carried out on the native tetramer of alcohol

dehydrogenase. First, in-source CAD was utilized to eject a monomer from the tetramer complex. The monomer was then subjected to ECD producing *c* and *z* fragments with reduced precoursors. Lastly the product ions were activated with HCD to obtain *b* and *y* fragments in addition to the *c* and *z* fragments. The spectrum resulting from this workflow revealed abundant *c*-ions from ECD as well as abundant *b*- and *y*-ions from subsequent HCD. (Fig. S1A) Analysis of the spectrum revealed many types of terminal fragment ions are produced from this workflow as well as *bz*, *cy*, and *cz* internal fragment ions adding up to a sequence coverage of 40%. (Fig. S1B) The fragmentation also revealed the presence of N-terminal acetylation, the location of the V58T mutation, and the location of a Zn^{2+} ion. ECD in combination with HCD reveals high sequence coverage for protein complex monomers and reveals the locations of multiple modifications along the protein sequence.

Electron capture dissociation with an orbitrap based instrument can be utilized to locate phosphorylation sites on the tau protein. Tau is a 40kDa intrinsically disordered amyloid protein that aggregates in Alzheimer's disease.[7] Tau is commonly phosphorylated, and it is thought that phosphorylation could trigger aggregation of the protein.[8] Tau phosphorylated with a CDK5 kinase was obtained and an MS1 spectrum revealed that 3 sites of the protein were dominantly phosphorylated. (Fig. S2) To determine the location of those phosphorylation sites, the 14+ charge state fragmented with ECD. The spectrum revealed charge reduced precursors as well as abundant c- and z- fragments. (Fig. 2A) It was found that the phosphorylation sites occur at S345 as well as a site between amino acid 184 and 227. (Fig. 2B) This data indicates that ECD fragmentation of large, phosphorylated proteins can reveal the location of phosphorylation sites which could contribute to increased aggregation of tau and the onset of neurodegenerative diseases such as Alzheimer's disease.

It has been reported that ECD of protein complexes can reveal structural information on those complexes. For example, ECD of yeast alcohol dehydrogenase reveals primarily *c*-fragments because the N-terminus is solvent exposed.[9] It has also been reported that other electron-based fragmentation techniques such as electron transfer dissociation (ETD) release fragments that are sensitive to protein structure.[10] To determine if ECD on an orbitrap instrument can also release structure sensitive fragments, a complex-down

and native experiment were performed on the aldolase homotetramer. Complex-down, where a monomer of aldolase is ejected with in-source CAD before ECD fragmentation (Scheme S1), reveals numerous abundant N and C-terminal fragments in the ECD spectrum. (Fig. 3A) Full analysis of the spectrum reveals that 46 N-terminal *c*-fragments and 43 C-terminal *z*-fragments are present in the spectrum. (Fig. 3B) In contrast, ECD of the intact native tetramer revealed abundant C-terminal *z*-fragments. (Fig. 3C) Complete analysis of this spectrum revealed that N-terminal *c*-fragments were not found in the spectrum but 54 C-terminal *z*-fragments were found. (Fig. 3D) The discrepancy between the complex-down and native workflows could indicate that there is a structural effect of fragments resulting from the native fragmentation workflow. Mapping the fragments resulting from native ECD of aldolase to the crystal structure indicated that the fragments resulted from the solvent exposed region of Aldolase. (Fig. 3D) This data indicates that fragments resulting from ECD on an orbitrap-based instrument can also reveal structure characteristics for those complexes.

It was also found that electron ionization dissociation (EID) can reveal structure information on protein complexes. EID dissociates peptide bonds with higher energy electrons than ECD and it has been found that EID of denatured proteins can result in higher sequence coverage compared to ECD;[11] however, it is not known if EID can release fragments in structure dependent manner. To investigate this, ECD and EID were performed on the native ADH homotetramer. ECD of ADH revealed charged reduced species in addition to covalent fragments released from the intact tetramer. (Fig. S3A) The fragmentation pattern reveals numerous N-terminal *c*-fragments since the N-terminal region of ADH is solvent exposed. (Fig. S3B) For comparison EID was also performed on the ADH homotetramer. The spectrum revealed charge reduced species, fragment ions, and charge oxidized species which is characteristic of EID fragmentation. (Fig. S3C) The fragment ions present in the spectrum are exclusively N-terminal *c*-fragments as is shown in the fragment location map. (Fig. S3B) Mapping the fragments from EID on the crystal structure of ADH reveals the fragments result from the solvent exposed region of the N-terminus much like ECD. This example indicates that EID of protein complexes releases fragments in a conformationally sensitive manner and reveals structural characteristics of those protein complexes.

This research indicates that ECD on an orbitrap mass spectrometer can reveal sequence, modification, and structure information on large proteins and protein complexes. ECD on a native antibody reveals sequence coverage of 32% for the light chain and 11% for the heavy chain as well as the location of various modifications on the heavy chain. ECD with subsequent HCD of ejected ADH monomers reveals intense fragments from both techniques as well as internal fragments which add up to a sequence coverage of 40%. ECD of phosphorylated tau revealed the location of the phosphorylation sites to be on the C-terminal domain and the interface between the proline rich domain and the microtubule binding domain. Native TD-MS of the native aldolase homotetramer with ECD reveals fragments that correspond to the solvent exposed C-terminus which contrasts with fragmentation of the monomer with ECD. Lastly, EID of the ADH homotetramer reveals the release of N-terminal *c*-fragments and is comparable to ECD of the same complex. This is consistent with the fact that the N-terminus of ADH is solvent exposed. ECD of proteins with an orbitrap-based mass spectrometer reveals results that are similar to ECD with other mass spectrometers and provide information relevant to the sequence and structure of the proteins.

FIGURES

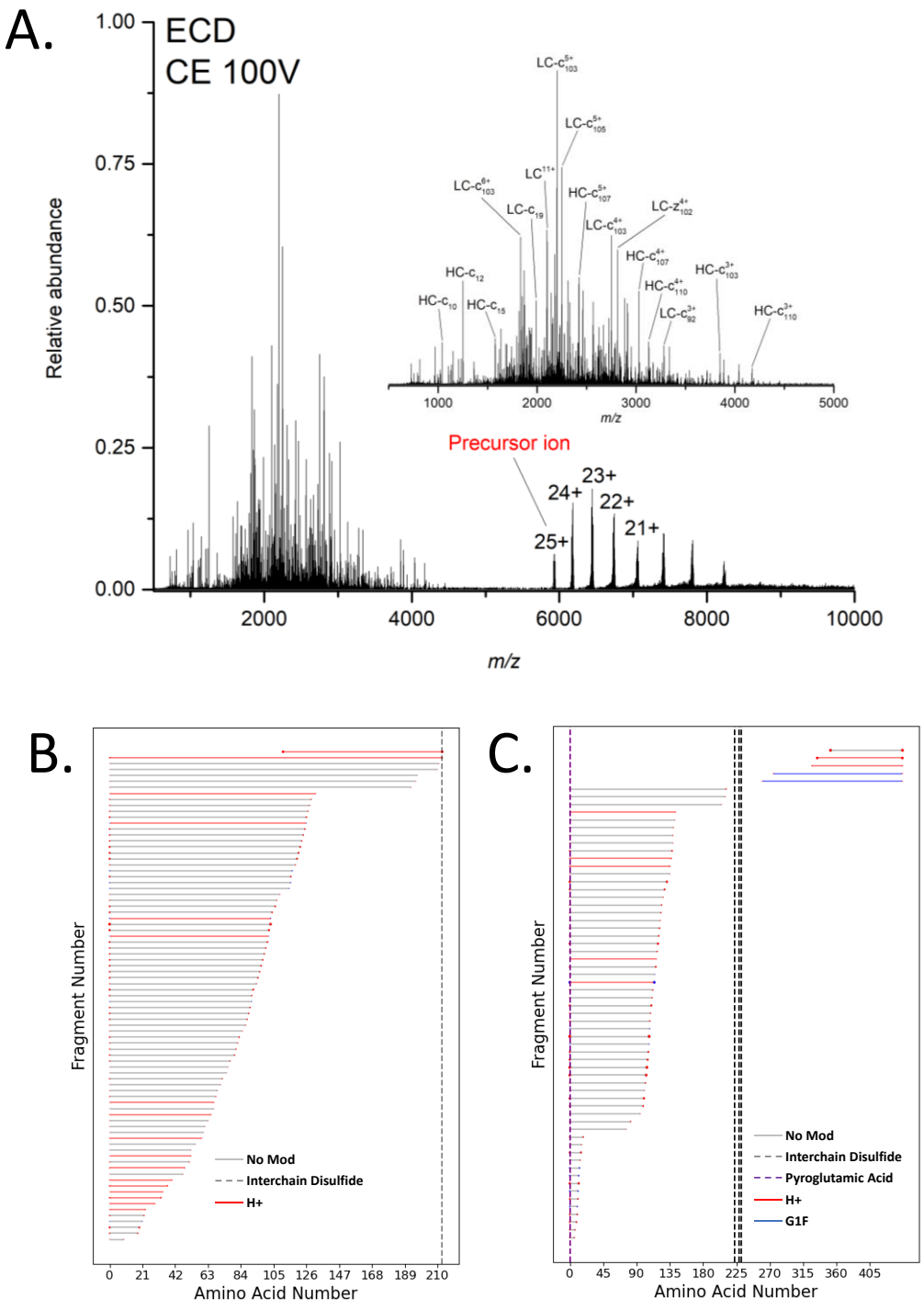


Figure 1: An ECD spectrum of the 25+ charge state of the intact NIST antibody along with the corresponding fragmentation location maps for the B.) light chain and C.) the heavy chain. Notice how fragments come from both chains and that modifications on the heavy chain can be located.

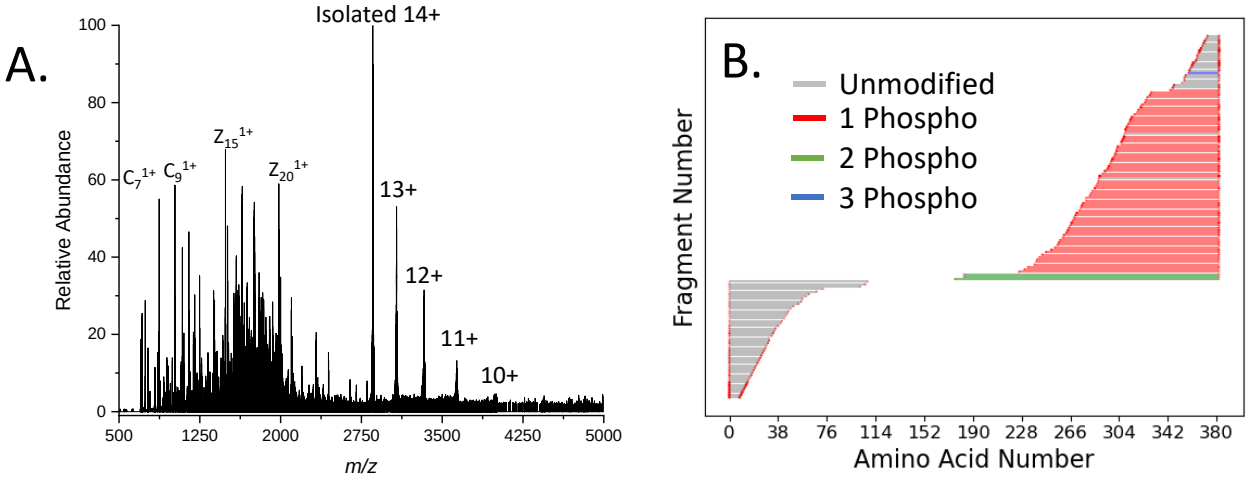


Figure 2: A.) An ECD spectrum of the 14+ charge state of tau phosphorylated for 20 mins with a CDK5 kinase and B.) the corresponding fragment location map. The data indicates that tau is phosphorylated on the boundary of the proline rich domain and the microtubule binding domain and the C-terminal domain of tau.

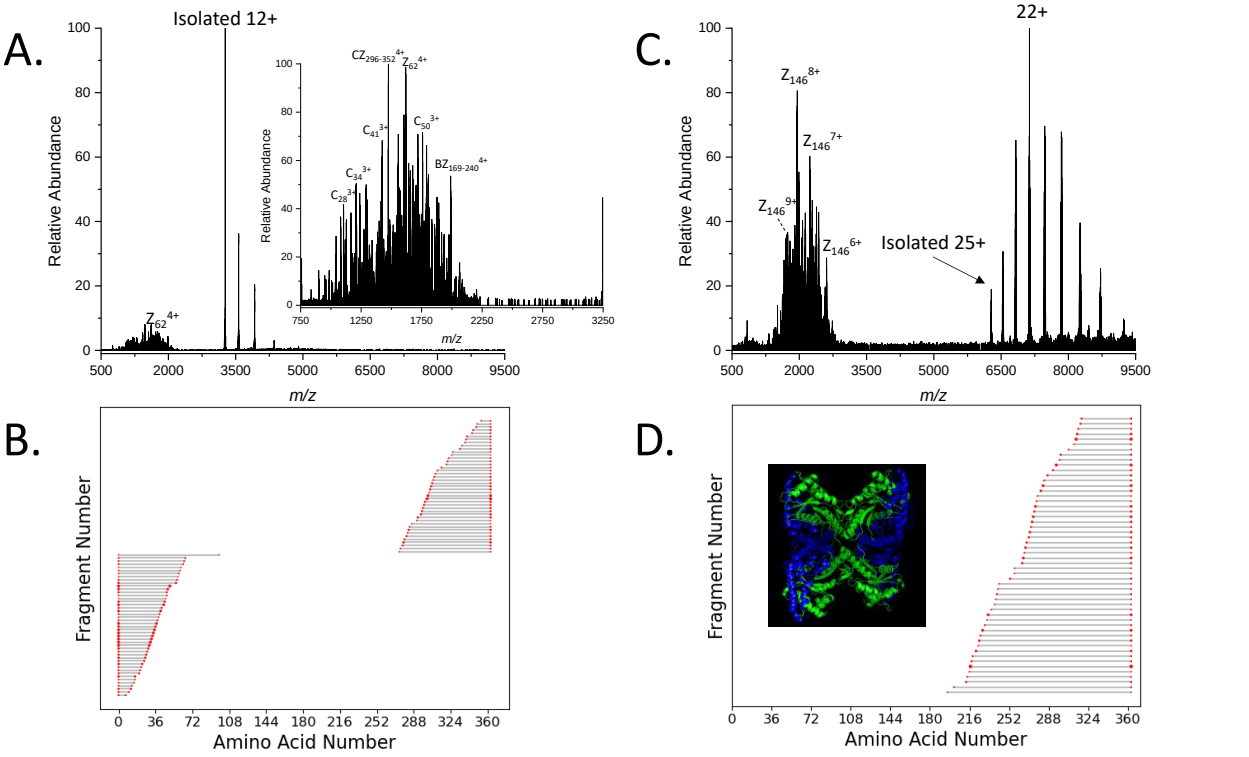


Figure 3: A.) A complex-down ECD spectrum of the aldolase homotetramer with B.) the corresponding location map and C.) a native ECD spectrum the aldolase homotetramer with D.) the corresponding fragment location map as well as the structure for the native tetramer (Blue=C-terminal coverage). Notice that exclusively C-terminal *z*-fragments are observed in the native TD-MS spectrum, indicating the C-terminus is solvent exposed. Mapping the fragments on the crystal structure supports this notion.

REFERENCES

1. 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.

2. 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.

3. 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.

4. 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.

5. Tsybin, Y. O.; Quinn, J. P.; Tsybin, O. Y.; Hendrickson, C. L.; Marshall, A. G., Electron capture dissociation implementation progress in Fourier transform ion cyclotron resonance mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2008,** *19* (6), 762-771.

6. Ayoub, D.; Jabs, W.; Resemann, A.; Evers, W.; Evans, C.; Main, L.; Baessmann, C.; Wagner-Rousset, E.; Suckau, D.; Beck, A. In *Correct primary structure assessment and extensive glyco-profiling of cetuximab by a combination of intact, middle-up, middle-down and bottom-up ESI and MALDI mass spectrometry techniques*, MAbs, Taylor & Francis: 2013; pp 699-710.

7. Johnson, V. E.; Stewart, W.; Smith, D. H., Widespread tau and amyloid-beta pathology many years after a single traumatic brain injury in humans. *Brain pathology* **2012,** *22* (2), 142-149.

8. Morris, M.; Knudsen, G. M.; Maeda, S.; Trinidad, J. C.; Ioanoviciu, A.; Burlingame, A. L.; Mucke, L., Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. *Nature neuroscience* **2015,** *18* (8), 1183-1189.

9. 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.

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. 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. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (9), 1896-1902.

SUPPORTING INFORMATION

MATERIALS AND METHODS

ADH, aldolase, and the NIST mAb were dissolved in 200mM ammonium acetate. These samples were desalted with 10kDa filters or 6k Bio-Rad filters and diluted to 10μM for analysis. 1N3R tau was expressed in *E. coli* and phosphorylated with a CDK5 kinase for 20min. The solution was then buffer exchanged into 20mM ammonium acetate. All samples were diluted to 10μM. The solutions were electrosprayed on a Thermo UHMR (Thermo Fisher Scientific, San Jose, CA) with metal-coated borosilicate capillaries (Au/Pd-coated, 1 μm i.d.) on a nanospray ionization source. Complex-down experiments were performed by using in-source dissociation to eject a monomer, guiding the monomer through the ECD cell, and activating the monomer with 50V of HCD energy. nTD-MS was performed by guiding intact protein complexes though the ECD cell and activating the complex with 50-110V of HCD energy.

UHMR TD-MS data was deconvoluted with Biopharma Finder 3.2. Deconvoluted peaks were assigned to *c*, *z*, and internal fragments by ClipsMS 2.0.0 with an error tolerance of 3-5ppm. The addition of a hydrogen atom was searched as an unlocalized modification. For ADH, modifications including an N-terminal acetylation, a V58T mutation, and a Zn^{2+} ion were searched. Phosphorylation modifications were added to tau for localization. Fragmentation of ADH (PDB: 4W6Z) and aldolase (PDB: 1ADO) was mapped onto crystal structures with Pymol 2.4.1.

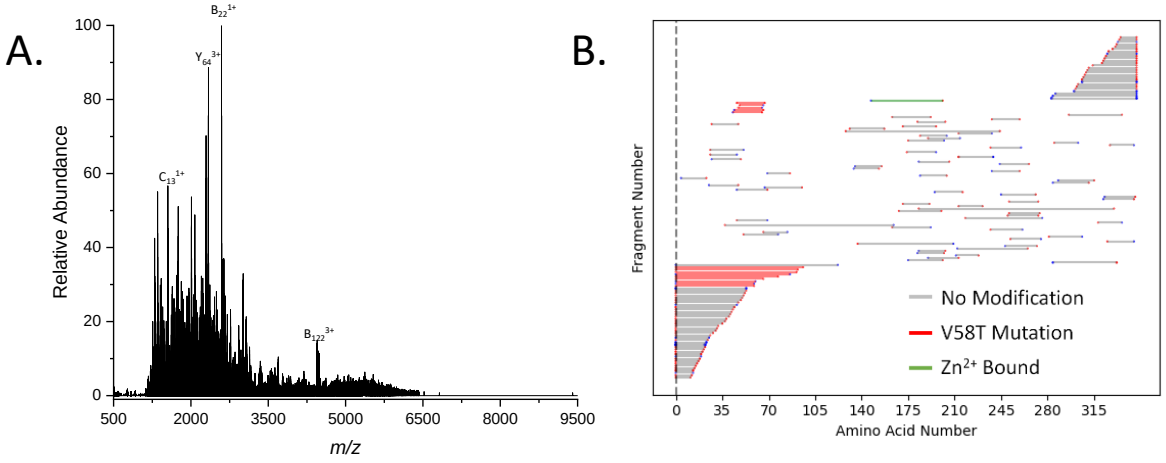


Figure S1: A.) ECD of ejected ADH monomers with subsequent HCD of the ECD products with B.) the corresponding fragmentation location map. The resulting *b*-, *c*-, *y*-, *z*-, and internal fragments add up to a sequence coverage of 40%.

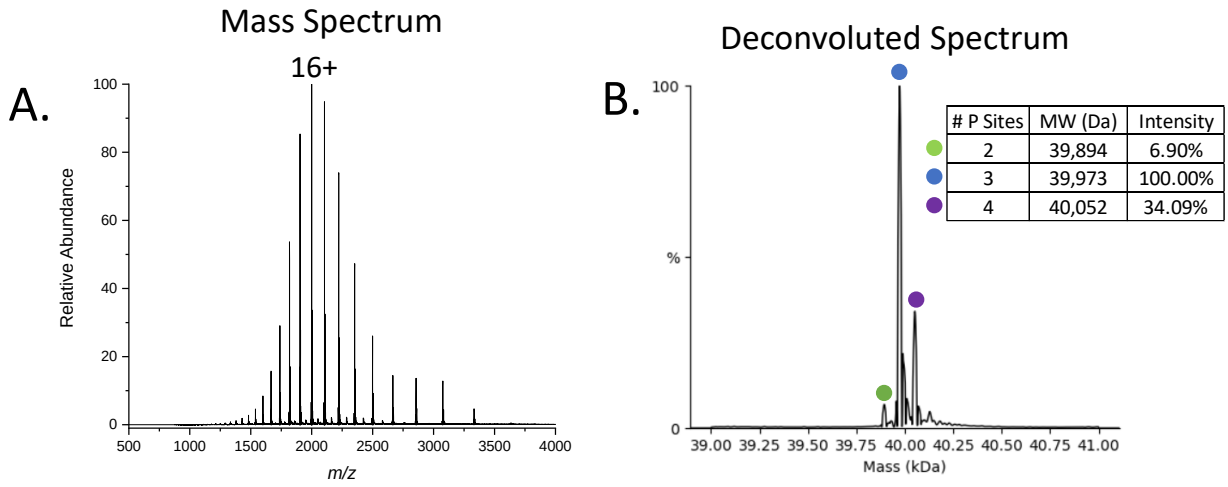


Figure S2: A.) A mass spectrum of phosphorylated tau and B.) the deconvoluted spectrum indicating that tau predominantly has 3 phosphorylation sites and up to 4 phosphorylation sites.

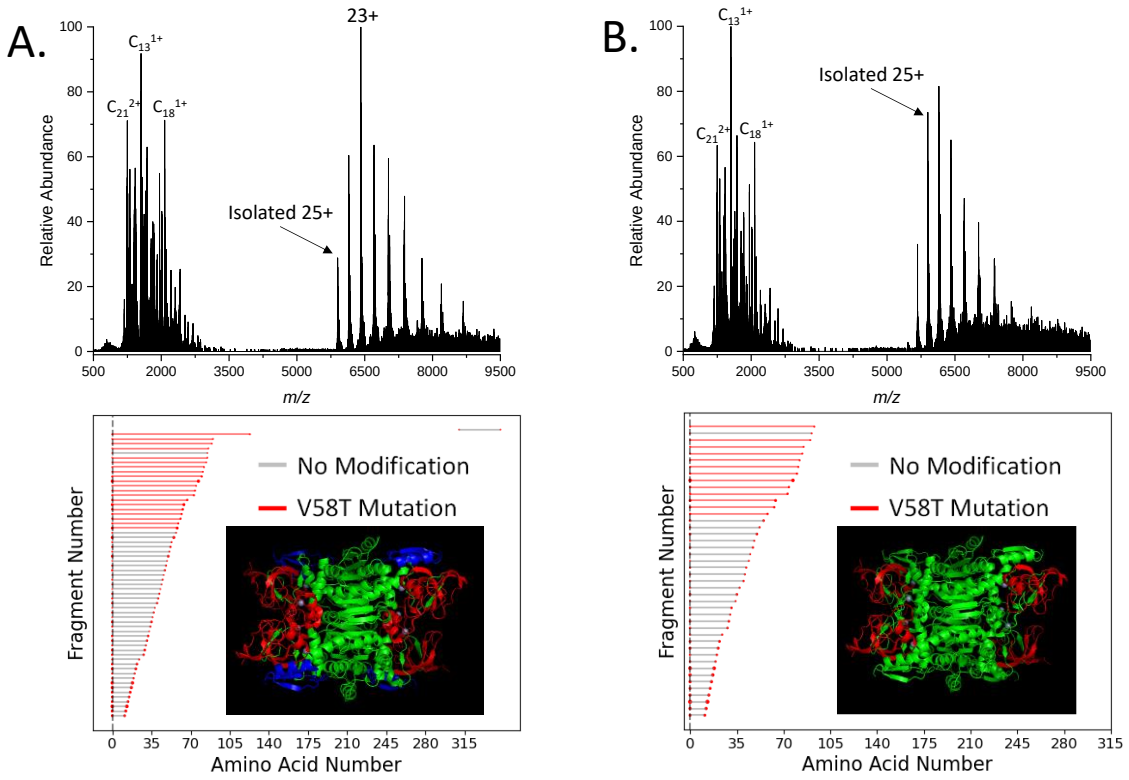
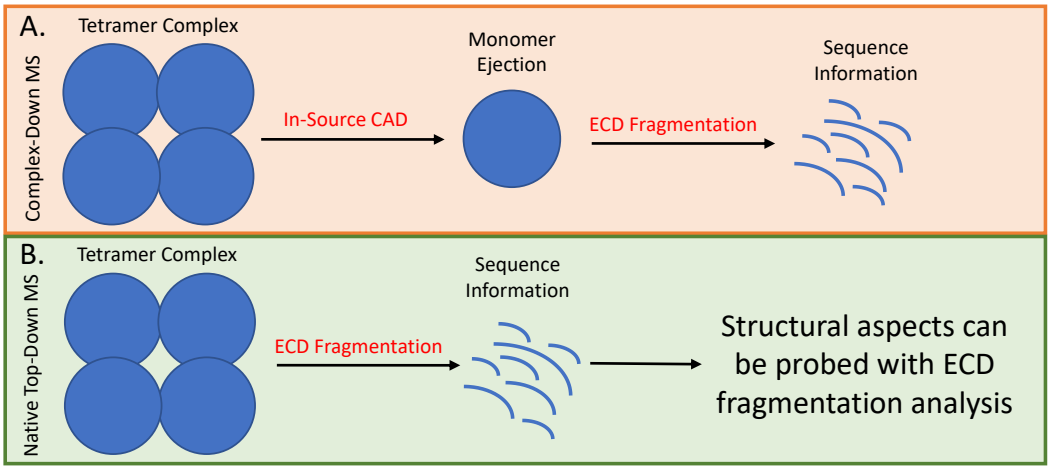


Figure S3: A.) ECD and B.) EID spectrum of the alcohol dehydrogenase (ADH) 147 kDa homotetramer with the corresponding fragment location maps and ADH structure (Red=N-terminal coverage and Blue=C-terminal coverage). The red lines correspond to a V58T mutation, and the vertical dotted line indicates the presence of an N-terminal acetylation. Notice that charge oxidation is not observed in the ECD spectrum but is observed in the EID spectrum. Mainly N-terminal *c*-fragments are observed in both spectra indicating that N-terminus of ADH is solvent exposed.



Scheme S1: The workflow of A.) complex-down TD-MS and B.) native TD-MS. nTD-MS can result in fragments that reveal structural information on protein complexes.

CHAPTER 5

Mass Spectrometry Structural Analysis of Intrinsically Disordered Phosphoproteins

Carter Lantz[1], Muhammad A. Zenaidee[1], Denise Tran[1], Karl Biggs[2], Gal Bitan[2], Rachel Loo[3], Joseph A. Loo[1]

[1]Department of Chemistry and Biochemistry, [2]Department of Neurology, David Geffen School of Medicine, [3]Department of Biological Chemistry University of California-Los Angeles, Los Angeles, CA 90095

ABSTRACT: Phosphorylation is a ubiquitous protein modification that is known to play important roles in many biological phenomena including cell signaling, the opening and closing of membrane protein channels, and even triggering of amyloid protein aggregation. Despite the important effect phosphorylation has on the function of proteins, the impact phosphorylation has on the structure of proteins is not well understood. To determine how phosphorylation affects the structure of proteins, top-down mass spectrometry (TD-MS) and ion mobility-mass spectrometry (IM-MS) were performed on various phosphorylated proteins and their dephosphorylated proteoforms. Top-down mass spectrometry (TD-MS) analysis with collision- and electron-based fragmentation techniques reveal the location of phosphorylation sites on the intrinsically disordered amyloid proteins β-casein and α-synuclein. It was also found that alkaline phosphatase dephosphorylates β-casein from the N-terminus to the C-terminus. Furthermore, IM-MS of commonly phosphorylated proteins such as β-casein, α-casein, ovalbumin, and phosvitin indicates that phosphorylation promotes compaction of protein structure in denaturing as well as native conditions. This observation is also shown for phosphorylation at serine 129 of the disease related amyloid protein α-synuclein. Compaction of proteins due to phosphorylation could suggest a mechanism that explains multiple biological phenomena seen in cellular mechanisms.

INTRODUCTION

Post-translational modifications (PTMs) (*e.g.*, acetylation, methylation, glycosolation, sumolation, ubiquitination, and phosphorylation) are covalent modifications added to proteins which can significantly alter the function of that protein. Of these modifications, phosphorylation has been shown to be one of the

most ubiquitous and consequential PTMs.[1] Phosphorylation of certain proteins has been shown to regulate important cellular processes such as translation[2] and mitosis,[3] open channels regulating calcium intake,[4] and activate p53, a cancer suppressing protein.[5] Furthermore, it has been reported that phosphorylation of amyloid proteins such as tau and α-synuclein can lead to aggregation and the onset of neurodegenerative diseases.[6-8] Phosphorylation has been shown to affect the function of proteins, although the effect phosphorylation has on the structure of a protein is not well understood.

Previous studies have attempted to provide insight into the structural effects of phosphorylation on peptides and proteins. Circular dichroism and ultracentrifugation of dephosphorylated and phosphorylated β-casein 1-25 shows that the structure of β-casein is more flexible when dephosphorylated.[9] Small-angle X-ray scattering has shown that phosphorylated peptides manifest a more compact tertiary structure than their corresponding dephosphorylated peptides.[10] Other research has suggested that the secondary structure of peptides[10] and proteins[11] can be altered by phosphorylation. Molecular simulations of peptides have shown that phosphorylation could compact peptides in the gas-phase by forming hydrogen bonds with other residues.[10, 12] Wong et al. also suggested that phosphorylated residues interact with the backbone of peptides in solution phase resulting in compaction of peptide structure.[13] The data from these studies have given us indications of how phosphorylation affects the structure of proteins; however, more information is needed to provide a more comprehensive view of how phosphorylation affects protein structure.

Bottom-up proteomics (BUP) has been utilized to localize sites of phosphorylation on various proteins including tau[14] histones,[15] β-casein,[16] and α-casein.[17] Although BUP has commonly been used to localize phosphorylation sites, any structural information on the protein is lost during the process of digestion and reverse phase liquid chromatography. To return relevant structural information, an intact protein must be ionized. With the development of electrospray ionization(ESI),[18] a soft ionization technique, proteins can be gently transferred into the gas-phase.[19, 20] This technique has been found to preserve weakly bound PTMs such as phosphorylation[21, 22] and even electrostatically bound compounds on native protein structures.[23-27] Once the intact protein is ionized into the gas-phase, modifications can be

localized with top-down mass spectrometry (TD-MS) and protein structure can be studied with ion-mobility mass spectrometry (IM-MS).

Top-down mass spectrometry (TD-MS) has become a rising field for the localization of PTMs including phosphorylation. TD-MS directly dissociates peptide bonds of intact proteins in the gas-phase using electron-, collision-, or photon-based fragmentation methods.[28] Fragments resulting from TD-MS can be mapped back to the protein sequence to provide relevant information on the sequence, [29] location of modifications,[30] and structural characteristics of proteins.[31] Measurement of dissociated products with Fourier transform-ion cyclotron resonance (FT-ICR) MS provides superior mass accuracy and resolution for accurate characterization and assignment of protein fragments.[32, 33] TD-MS in combination with FT-ICR has been utilized to localize phosphorylation sites on intact proteins including cardiac troponin 1,[34] β-casein,[35, 36] and mutant tau.[26] Top-down mass spectrometry has been found to be a useful tool when characterizing proteoforms of proteins and has previously been found to be an efficient way to localize phosphorylation sites.

Ion-mobility mass spectrometry (IM-MS) of intact proteins can be used to return relevant structural information on proteins and protein complexes. In IM-MS analytes are separated by collision cross section (CCS) so as to separate out proteins by their size.[37, 38] IM-MS data can be used to separate peptide mixtures[39, 40] as well as separate different conformations of proteins and protein complexes.[26, 41-43] IM-MS studies have performed on phosphorylated peptides showing that peptides have lesser drift time compared to their non phosphorylated counterparts indicating compaction of peptide structure.[12, 44-47] Eyers and coworkers recently showed that phosphorylation stabilizes cAMP-dependent kinase and shows an increase in early arriving confirmers in the ion mobiligram.[48] Our lab has shown that phosphorylation may promote more compact structures of a mutant tau protein.[26] IM-MS has proven to be a useful tool when characterizing protein structure and has been found to provide relevant information on changes to protein conformation due to modifications such as phosphorylation.

In this study, TD-MS and IM-MS are utilized to determine the effects phosphorylation has on the structure of proteins. Intact mass spectrometry and TD-MS efficiently characterizes the proteoforms in

solution and localizes where those sites are on the proteins. Phosphorylation seems to shift the structure of proteins by interacting with positively charged residues on the protein. This data indicates that phosphorylation compacts protein structure which may lead to alteration of protein function or the progression of diseases such as neurodegenerative diseases such as Parkinson's disease.

MATERIALS AND METHODS

β-casein, α-casein, and ovalbumin were bought from Milipore sigma already phosphorylated. Phosphorylated α-synuclein was obtained from Gal Bitan's lab. To dephosphorylated proteins, 1-3 units of alkaline phosphatase was inserted into solutions with 4.5-20ug of protein for 30mins on ice. Protein solutions sprayed in positive ion mode were dissolved in 50%ACN with 0.1% FA and protein solutions sprayed in negative ion mode were prepared in 50% ACN. The proteins were electrosprayed at a concentration of 10uM on a Synapt G2Si in Mobility-TOF mode. The proteins were sprayed with a capillary voltage of 1-2kV, a cone voltage of 20V, and a source offset of 40V. Upwards of 300 scans were collected for each experiment.

Proteins sprayed in native solutions were dissolved in 20mM ammonium acetate. Protein solutions were dephosphorylated by adding 1-2 units of alkaline phosphatase solution from Promega to 20ug of protein for 30 mins on ice. Alkaline phosphatase was allowed to dephosphorylated phosvitin for 24 hours in 4°C. Protein solutions were sprayed at a concentration from 10-30uM on a Synapt G2 in Mobility-TOF mode for both positive and negative ion mode. The capillary voltage was 0.9-1.5kV, a cone voltage was 20-40V, and a source offset of 40-80V. Upwards of 300 scans were collected for each experiment.

To obtain all phosphorylation states for β-casein, the protein was dissolved into 20mM ammonium acetate and diluted to 7.5M. 1.31 miliunits of alkaline phosphatase were inserted into 3.5ug of β-casein. Immediately after injection of the alkaline phosphatase, the solution of β-casein and alkaline phosphatase was inserted into a borosilicate nanoelectrospray tip and sprayed on a Waters Synapt G2 Si. 800 scans were acquired over 1 second intervals. The capillary voltage was 1.00kV, a cone voltage of 20V, and a source offset was 40V. The resulting spectrum was deconvoluted with UniDec to observe the summed intensities.[49]

β-casein was analyzed on a Bruker 15T FT-ICR SolariX instrument equipped with an infinity cell with metal-coated borosilicate capillaries (Au/Pd-coated, 1μM i.d.; Thermo Fisher Scientific) and sprayed at a flow rate of 10-40 nL/min with a nanospray ion source. Denatured β-casein was diluted to 10uM and sprayed in 50% ACN with 0.1% FA. The capillary voltage was set to 0.80kV, a skimmer voltage of 50V, a funnel voltage of 100V, and a time of flight of 0.8 seconds. Native β-casein was diluted to 10uM and sprayed in 20mM ammonium acetate with a capillary voltage of 1.20kV, a skimmer voltage of 25V, and a funnel voltage of 80V, and a time of flight of 1.0 second. 100 scans were averaged for each experiment. To perform top-down mass spectrometry analysis denatured β-casein was broadband fragmented with a CAD voltage of 3V and an EID bias of 25V. 400 scans were averaged for the experiment. Fragments were deconvoluted with Bruker software and analyzed with ClipsMS[50] with an error of 2ppm. Fragments were manually interpreted up to 5ppm.

To fragment all phosphorylation states of β-casein, protein was dissolved in 20mM ammonium acetate and diluted to 60uM. 1.31miliunits of solution was provided for 283ug of β-casein. At time points 0mins, 4 mins, 10mins, 15mins, and 30mins, 25uL of sample was placed in 25uL of ACN with 0.2% formic acid. The solutions were diluted to 20uM and sprayed with a capillary voltage of 0.80kV, a skimmer voltage of 50V, a funnel voltage of 100V, and a time of flight of 0.8 seconds. Each phosphorylation state from the 22+ charge state was fragmented with an m/z window of 3, a CAD voltage at 3 volts, and an EID bias of 30 volts. 100 scans were collected for each experiment. The data was deconvoluted with Bruker software and analyzed with ClipsMS with an error of 2ppm.

Native α-Synuclein was diluted to 10uM in sprayed in 20mM ammonium acetate. The solution was sprayed with a capillary voltage of 1.00kV, a skimmer voltage of 50V, a funnel voltage of 90V, and a time of flight of 0.80 seconds. 100 scans were collected for the experiment. To fragment native phosphorylated α-synuclein, the solution was sprayed with a capillary voltage of 1.00kV, a skimmer voltage of 50V, a funnel voltage of 90V, and a time of flight of 0.80 seconds, then the 15+ charge state was isolated with a m/z window of 20, and fragmented with an ECD bias of 1.0V. 400 scans were averaged for the experiment. The data was deconvoluted with Bruker software and analyzed with ClipsMS with an error of 2ppm.

RESULTS AND DISCUSSION

Characterization of Phosphorylated β-Casein

Intact mass spectrometry (MS1) and top-down mass spectrometry (TD-MS) were performed on β-casein to characterize the protein and locate the phosphorylation sites. MS1 analysis of denatured β-casein in positive ion mode revealed a bimodal distribution with the most abundant charge states for each maximum being 16+ and 22+. (Fig. 1A) The spectrum of native phosphorylated β-casein in positive ion mode showed a trimodal distribution with the most abundant charge states for each maximum at 11+, 16+, and 21+. (Fig. S1) The molecular weight corresponds to the mass of the protein containing 5 phosphorylation sites. The spectrum also reveals the presence of two variants known as the A1 and A2 variant which differ by one amino acid at position 67. (Fig. 1A)[51, 52] Top-down fragmentation with collisionally activated dissociation (CAD) and electron ionization dissociation (EID) was then performed on β-casein to localize the phosphorylation sites. The data revealed that phosphorylation sites were present on ser15, ser17, ser18, ser19, and ser35 which has been reported previously. (Fig. 1B) [26, 35] In addition, a *b*-fragment present at amino acid 74 contained a 40Da shift indicating the presence of the histidine variant at amino acid 67. (Fig. 1B) Alkaline phosphatase, an enzyme that has the capability to remove phosphorylation sites from a protein, was utilized to subsequently remove each phosphorylation site from β-casein and the states were isolated and fragmented with TD-MS. It was found that alkaline phosphatase removes phosphorylation sites from the N-terminus to the C-terminus. (Fig. S2) This indicates that the most N-terminal phosphorylation sites are more accessible to alkaline phosphatase and removal of the N-terminal sites provides access to other sites.

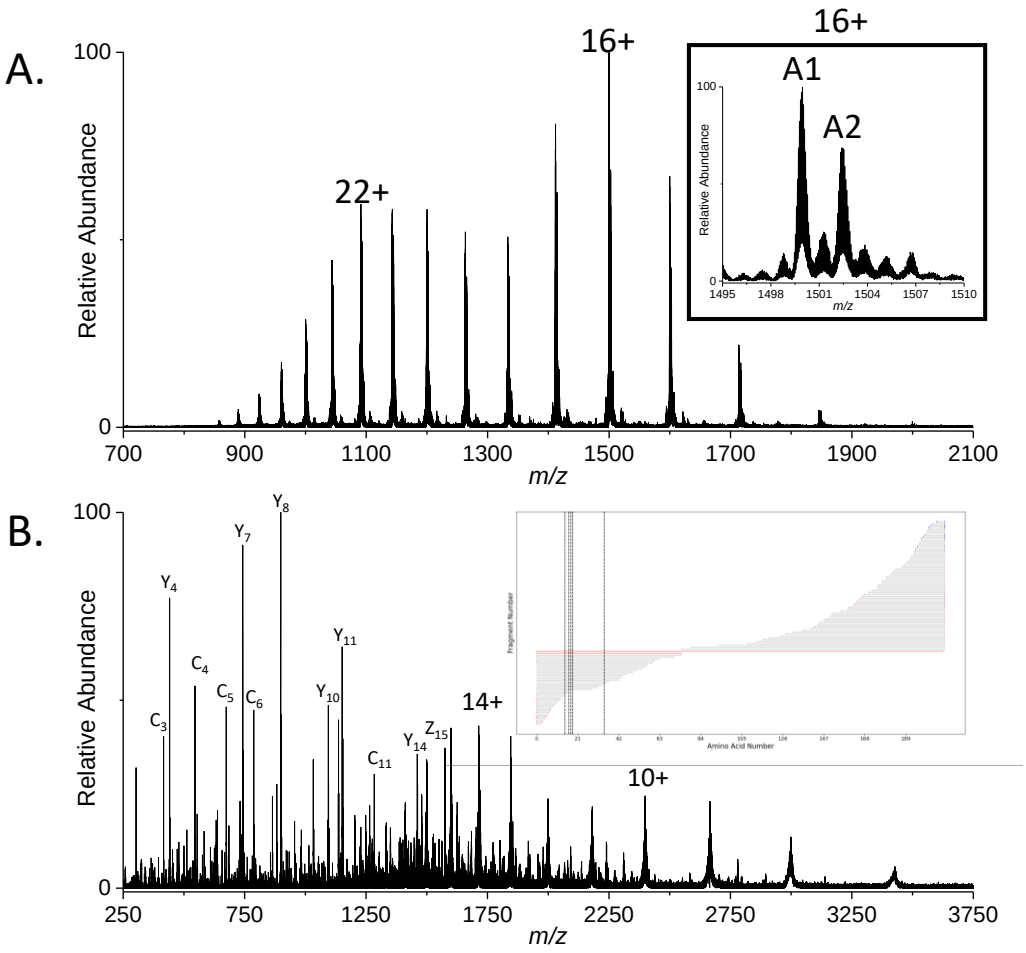


Figure 1: A.) A mass spectrum of denatured β-casein with an inset of the 16+ charge state showing the presence of the A1 and A2 variant, and B.) a spectrum of CAD/EID broadband fragmentation of phosphorylated β-casein with an inset of the fragment location map. The vertical dotted lines represent the location of the phosphorylation sites, and the red horizontal lines indicate existence of a point mutation from proline to histidine at position 67.

Structure Analysis of Denatured Phosphorylated β-Casein

To investigate the effect of phosphorylation on protein structure, dephosphorylated and phosphorylated β-casein were dissolved in denaturing solution and ion-mobility mass spectrometry (IM-MS) was performed on the solutions in positive-ion mode. The mass spectrum of dephosphorylated β-casein displayed a unimodal charge state distribution with 18+ as the most abundant charge state. (Fig. 2A) The mass spectrum of phosphorylated β-casein revealed a bimodal charge state distribution with the most abundant charge states for each maximum being 22+ and 15+. (Fig. 2B) A difference in charge state distribution often indicates a shift in protein conformation,[53-55] so to further investigate the structural effects

of phosphorylation on β-casein, the mobiligrams of the 16+ charge state of dephosphorylated and phosphorylated β-casein were extracted. The 16+ charge state showed two resolved peaks with times corresponding to 9.54ms and 11.75ms;(Fig. 2C) however, phosphorylated β-casein revealed early arriving conformers increase in abundance. (Fig. 2D) This trend is also present for the 15+ charge state which showed an increase in early arriving conformers when β-casein is phosphorylated. (Fig. S3A) The mobiligrams of the higher charge states such as the 22+ charge state showed a minimal shift in drift time. (Fig. S3B) It is possible that accumulation of charge on the protein abolishes any localized structures the protein may contain.[54, 55] Similarly, α-casein shows a similar trend where the charge state distribution shifts considerably when phosphorylated, and the mobiligrams show that early arriving conformers of α-casein increase in abundance when phosphorylated. (Fig. S4) The presence of early arriving conformers for many of the charge states suggests that phosphorylation compacts the structure of denatured proteins when electrosprayed in positive ion mode. Previous research has shown that phosphate groups interact with side chains of proteins through salt bridges resulting in intermolecular peptide interactions, intramolecular peptide interactions, and stable peptide helical structures.[10, 56-58] It is also known that salt bridges contribute significantly to protein structural stability in the gas-phase.[59-61] Therefore, it is possible that salt bridges form between phosphorylation sites and positively charged residues on the protein which results in a more compact structure of the protein.

A difference in electrospray polarity often shifts the charge state distribution of a protein.[62, 63] To determine whether the structural effect present in positive-ion mode is also present in negative-ion mode, dephosphorylated and phosphorylated β-casein were dissolved into denaturing solution and IM-MS was performed in negative ion mode. The mass spectrum of dephosphorylated β-casein displayed a mostly unimodal distribution with the most abundant charge state being 10-, although there was a less intense distribution of higher charge states with the most abundant charge state being 16-. (Fig. 2E) The mass spectrum of phosphorylated β-casein showed an increase in the abundance of the higher charge states revealing a distinct bimodal distribution with the most abundant charge states at each maximum being 20- and 12-. (Fig. 2F) This shift in charge state distribution suggests that a structural shift of β-casein due to

phosphorylation may be present negative-ion mode. The mobiligram of the 16- charge state of dephosphorylated β-casein reveals two dominate conformers at 10.09ms and 11.89ms. (Fig. 2G) The mobiligram of the 16- charge state for phosphorylated β-casein also showed the presence of those two conformers; however, the abundance of the early arriving conformers increases relative to the later arriving conformer. (Fig. 2H) The 10- (Fig. S5A) and 13- (Fig. S5B) charge states also showed an increase in early arriving conformers when β-casein is phosphorylated. Similarly, analysis of α-casein in negative-ion mode reveals a shift in charge state distribution due to phosphorylation, and the mobiligrams reveal the abundance of early arriving conformers increase when phosphorylated. (Fig. S6) The increase of early arriving conformers due to phosphorylation in negative-ion mode suggests the structure of proteins are compacted when phosphorylated. It is possible that salt bridges between negatively charged phosphates and positively charged residues are the reason for this compaction.

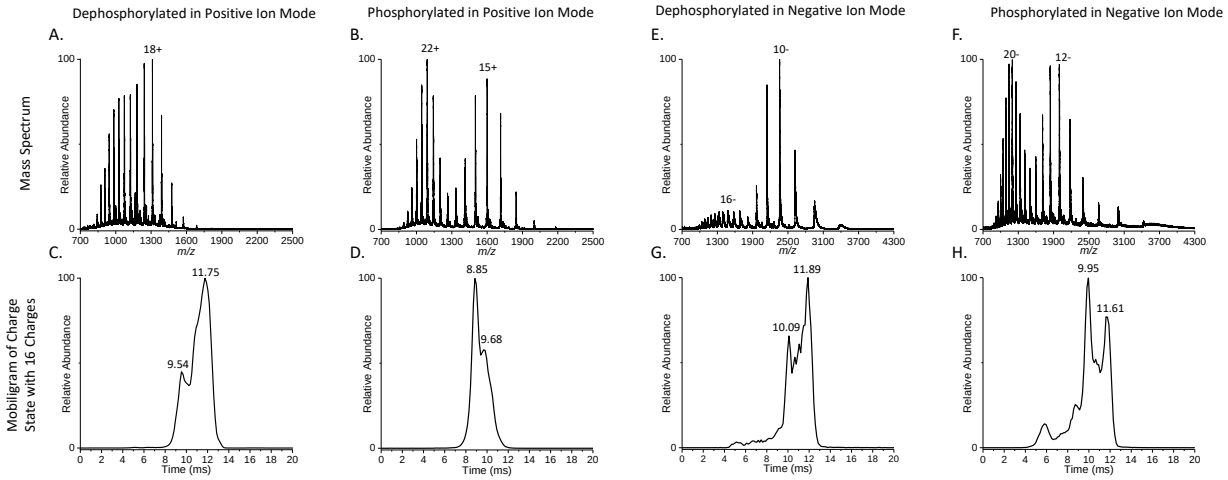


Figure 2: Spectra of denatured A.) dephosphorylated and B.) phosphorylated β-casein in positive ion mode with 16+ mobiligrams for C.) dephosphorylated and D.) phosphorylated β-casein, and spectra of E.) dephosphorylated and F.) phosphorylated β-casein in negative ion mode with 16- mobiligrams for G.) dephosphorylated and H.) phosphorylated β-casein.

Structure Analysis of Native Phosphorylated β-Casein

It is thought that native proteins tend to have more ordered arrangements of residues compared to their denatured counterparts.[64] Despite this, it is possible that phosphorylation could still have a structural effect on native proteins. To investigate this, β-casein was dissolved in 20mM ammonium acetate at pH 6.9

and IM-MS was performed in positive ion mode. The spectrum of dephosphorylated β-casein revealed a trimodal charge state distribution with the most abundant charge states at each maximum being 10+, 13+, and 21+. (Fig. 3A) The spectrum of phosphorylated β-casein showed a slight increase in the abundance of the lower charge states, although the spectrum still retained a trimodal distribution with the most abundant charge states for each maximum being 10+, 13+, and 22+. (Fig. 3B) The shift in charge state distribution was not as pronounced in native conditions compared to denatured conditions; however, it is still possible that a shift in structure is still present. The mobiligram of the 16+ charge state showed the presence of two conformers at 8.71ms and 10.25ms. (Fig. 3C) The mobiligram of the 16+ charge state of phosphorylated β-casein reveals conformers of similar size; however, the early arriving conformer increases in abundance relative to the late arriving conformer. (Fig. 3D) The mobiligrams for the lower charge states such as 10+ also reveal an increase in early arriving conformers when phosphorylated. (Fig. S7A) The mobiligram of the higher charge states such as 21+ may reveal a slight shift toward early arriving conformers, although the higher charges did not shift to the same degree as the lower charge states. (Fig. S7B) Furthermore, the spectra of native α-casein reveal only a minimal shift in charge state distribution when phosphorylation is present, although, early arriving conformers are shown to increase due to the presence of phosphorylation sites. (Fig. S8) The increase in early arriving conformers, especially for the lower charge states, indicates that phosphorylation compacts the structure of native proteins when electrosprayed in positive-ion mode. It is possible that salt bridges between phosphates and positively charge residues are also present in native conditions which lead to the compaction of the proteins.

To investigate how a difference in polarity would affect phosphorylated proteins in their native state, dephosphorylated and phosphorylated β-casein were electrosprayed under native conditions in negative-ion mode. The spectrum of dephosphorylated β-casein primarily displays a unimodal distribution with the maximum charge state being 10-, although there is a smaller distribution of higher charge states with the most abundant charge state being 16-. (Fig. 3E) The spectrum of phosphorylated β-casein revealed an increase in the abundance of the higher charge states and a distinct bimodal distribution with the most abundant charge states at each maximum being 11- and 22-. (Fig. 3F) The shift in charge state distribution

indicates that the structure of native β-casein also shifts due to phosphorylation in negative-ion mode. The mobiligram for the 16- charge state of dephosphorylated β-casein revealed two dominate conformers at 10.14ms and 11.91ms. (Fig. 3G) The mobiligram for the 16- charge state of phosphorylated β-casein reveals similar size conformers however the early arriving conformer increased in abundance relative to the late arriving conformer. (Fig. 3H) The 10- (Fig. S9A) and 13- (Fig. S9B) charge states also showed an increase in early arriving conformers when β-casein is phosphorylated. Furthermore, native α-casein shows a similar trend where the charge state distribution shifts considerably and the mobiligrams show that early arriving conformers increase in abundance when α-casein is phosphorylated. (Fig. S10) The increase in abundance of early arriving conformers indicates that the presence of salt bridges in native conditions could be compacting the structure of native β-casein when electrosprayed in negative-ion mode.

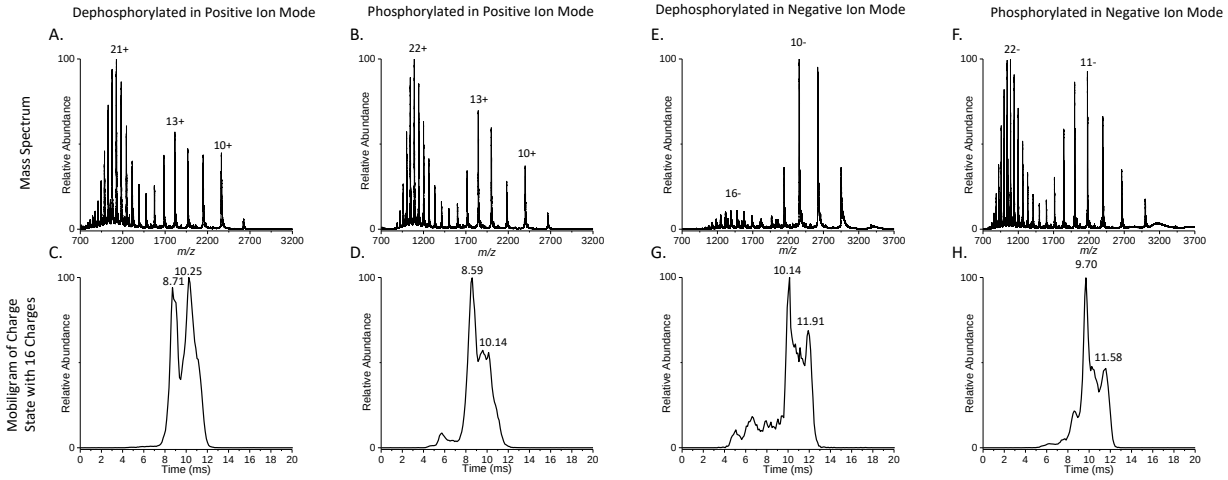


Figure 3: Spectra of native A.) dephosphorylated and B.) phosphorylated β-casein in positive ion mode with 16+ mobiligrams for C.) dephosphorylated and D.) phosphorylated β-casein, and spectra of native E.) dephosphorylated and F.) phosphorylated β-casein in negative ion mode with 16- mobiligrams for G.) dephosphorylated and H.) phosphorylated β-casein.

To further confirm the result that phosphorylation promotes compaction of protein structure, an experiment was also performed to determine if incremental compaction could be observed due to removal of each phosphorylation site on β-casein. β-casein was dissolved in native conditions with a minimal amount of alkaline phosphatase and electrosprayed in positive ion mode. In the mass spectrum, it was observed that each phosphorylation state of β-casein could be resolved, and a UniDec deconvolution confirms the presence of all phosphorylation states. (Fig. S11A) IM-MS of each phosphorylation state for

the 12+, 16+, and 21+ charge states showed that the addition of a phosphate on the protein incrementally gives rise to early arriving conformers. (Fig. S11B) This data further suggests that each phosphorylation site adds to the compaction of β-casein by forming salt bridges and further supports our hypothesis that phosphorylation compacts the structure of β-casein by forming salt bridges with positively charged residues.

Structure Analysis of Other Phosphorylated Proteins

Other common phosphorylated proteins were analyzed to determine the effect of phosphorylation on protein structure. Phosvitin has been reported to be heavily phosphorylated and glycosylated. [65, 66] To investigate how phosphorylation affects phosvitin, the protein was dissolved in native solution and IM-MS in positive ion mode was performed. The spectrum of dephosphorylated phosvitin revealed a dominant unimodal distribution; however, the spectrum seems to be quite heterogeneous possibly due to all the glycosylation sites on the protein. (Fig. S12A) The spectrum of phosphorylated phosvitin revealed a heterogenous bimodal distribution indicating that various phosphorylation and glycosylation proteoforms are present in the spectrum. (Fig. S12B) IM-MS was performed on both spectra of phosvitin and a drift plot reveals that dephosphorylated phosvitin (Fig. S12C) favors extended conformers compared to phosphorylated phosvitin. (Fig. S12D) The presence of early arriving conformers indicates that phosphorylation does promote a more compact structure of phosvitin. This result supports data from a previous study that showed it is more difficult to digest phosvitin with cleavage enzymes when phosphorylated.[67] It is possible that a more compact structure of phosvitin could decreases the efficiency of the cleavage enzymes by restricting sites available to the enzymes.

Ovalbumin, a heterogeneous monomeric protein with multiple sites of phosphorylation and glycosylation,[68] was analyzed in denaturing and native conditions in positive-ion mode. The spectra of dephosphorylated and phosphorylated ovalbumin are quite similar, although it was shown that the lower charge states of ovalbumin slightly increase in intensity relative to the higher charge states when ovalbumin is phosphorylated. (Fig. S13A) IM-MS of the 29+ (Fig. S13B), 32+ (Fig. S13C), and 35+ (Fig. S13D) charge states of phosphorylated and dephosphorylated ovalbumin shows that the drift time for

phosphorylated ovalbumin decreases indicating compaction of ovalbumin structure. In native conditions, the most intense charge state of dephosphorylated ovalbumin is 13+; however, when phosphorylated, the charge state distribution shifts and the most intense charge state becomes 13+. (Fig. S14A) This shift in distribution indicates phosphorylation affects ovalbumin structure. IM-MS of the 12+ (Fig. S14B) and 11+ (Fig. S14C) charge states for dephosphorylated and phosphorylated ovalbumin indicates there is a slight increase in early arriving conformers due to phosphorylation. The 13+ charge state for dephosphorylated and phosphorylated ovalbumin did not differ much in mobility. (Fig. S14D) This data suggests that more ordered proteins such as ovalbumin are also compacted when phosphorylated.

Characterization of Phosphorylated α-Synuclein

α-Synuclein is an intrinsically disordered amyloid protein. Aggregation of this protein in the brain has been shown to precipitate the death of neurons and incur the onset of neurodegenerative diseases such as Parkinson's disease.[69] Phosphorylation of S129 on α-synuclein is thought to modulate the aggregation rate; however, the mechanism for aggregation modulation is not known. In addition, there is much debate on how phosphorylation of this residue affects aggregation rate with some studies suggesting that phosphorylation at S129 increases aggregation[7] and other studies suggesting S129 phosphorylation decreases the aggregation rate.[70] It is possible that structure analysis of α-synuclein could provide some insight into the effect phosphorylation at S129 has on the aggregation of this proteins.

To characterize phosphorylated α-synuclein, MS and TD-MS analysis was performed on a proteoform of phosphorylated α-synuclein. MS1 analysis of the native protein displayed a unimodal distribution with the most abundant charge state at 13+. (Fig. 4A) MS1 analysis of the denatured protein was a unimodal distribution with the most abundant charge state at 15+. (Fig S15) The mass of the protein was revealed to be the molecular weight of the α-synuclein sequence with the addition of 1 phosphorylation site. TD-MS with ECD of the 15+ charge state was performed, and analysis of the resulting TD-MS spectrum indicated that α-synuclein is phosphorylated at serine 129. (Fig. 4B) Structure analysis of this α-synuclein proteoform could provide insight into how phosphorylation at this site modulates aggregation.

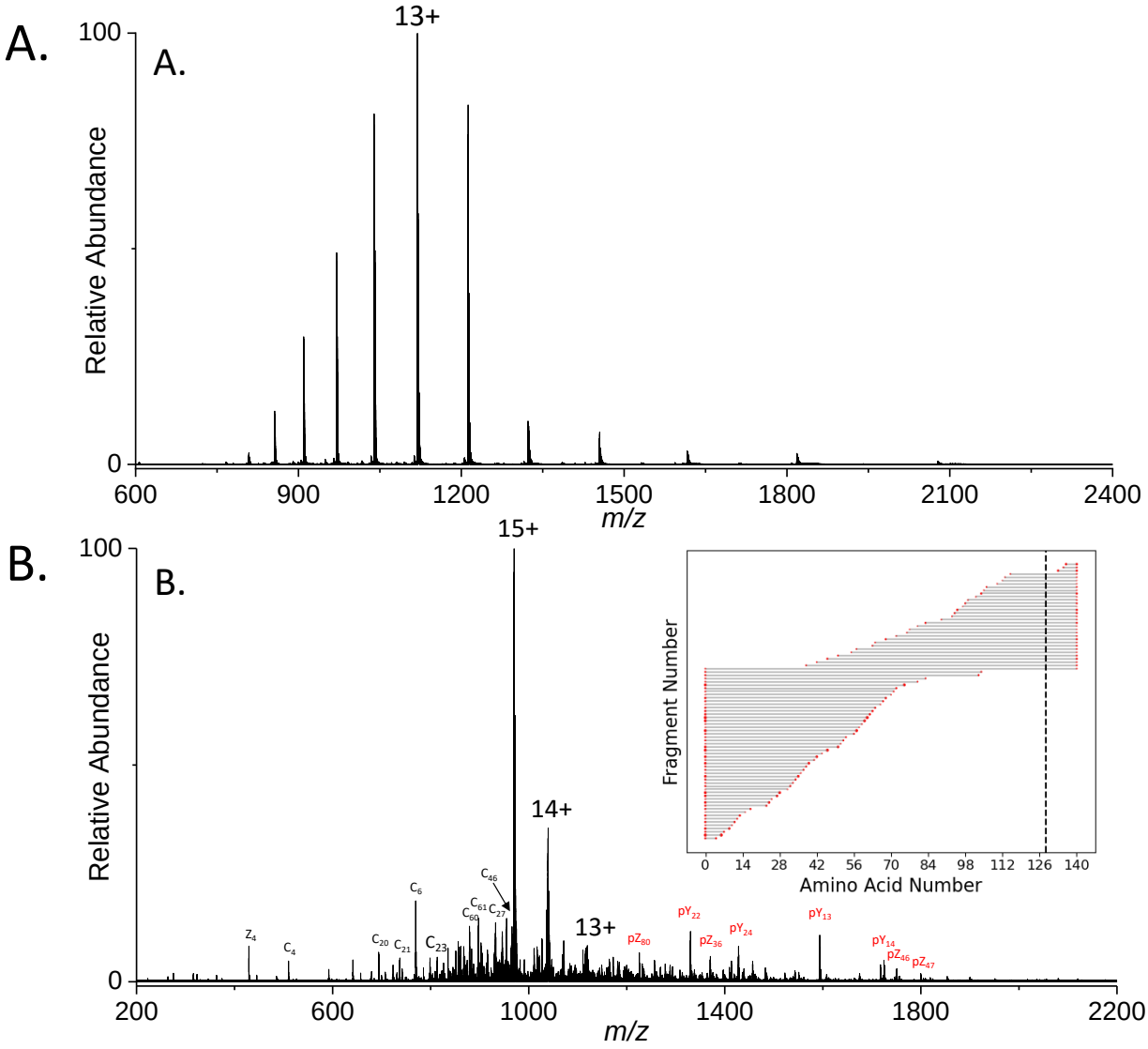


Figure 4: A.) A native mass spectrum of phosphorylated α-synuclein and B.) an ECD fragmentation spectrum of the 15+ charge state indicating that the location of phosphorylation is at serine 129 (dotted line).

Structure Analysis of Phosphorylated α-Synuclein

To determine the structural effect of S129 on native α-synuclein, dephosphorylated and phosphorylated α-synuclein were dissolved in native conditions and IM-MS was performed in positive-ion mode. The spectrum of dephosphorylated α-synuclein exhibited a unimodal distribution with the most intense charge state being 13+. (Fig. 5A) Phosphorylated α-synuclein showed a unimodal charge state distribution with the most abundant charges state being 14+. (Fig. 5B) The mobiligram for the 9+ charge

state of dephosphorylated α-synuclein reveals 4 conformers with drift times of 8.05ms, 9.26ms, 10.91ms, and 11.69ms. (Fig. 5C) The mobiligram for the 9+ charge state of phosphorylated α-synuclein reveals similar size conformers; however, the early arriving conformers increase in abundance relative to the late arriving conformer. (Fig. 5D) The mobiligram for the 10+ charge state also showed an increase in the abundance of early arriving conformers when α-synuclein was phosphorylated. (Fig. S16A) The higher charge states such as the 15+ did not change shape or drift time significantly due to phosphorylation (Fig. S16B) The spectrum of denatured phosphorylated α-synuclein reveals a similar distribution to dephosphorylated α-synuclein, and the mobiligrams do not show much of a difference due to phosphorylation of the protein. (Fig. S17) The increase in abundance of early arriving conformers in native conditions reveals phosphorylation compacts native α-synuclein in positive ion mode. It is possible that salt bridges between the phosphate on S129 and positively charged residues compact the structure of the protein.

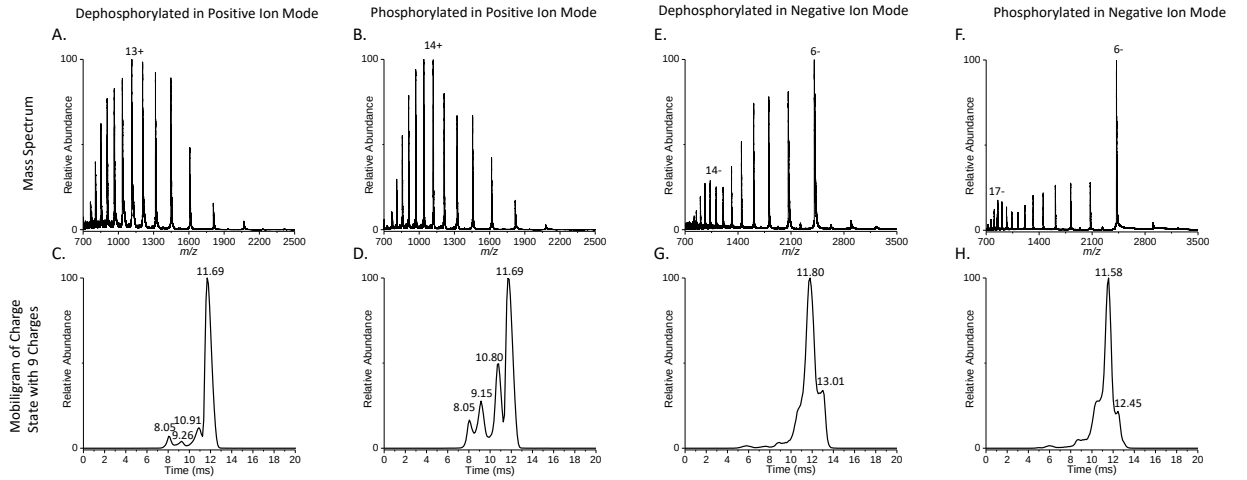


Figure 5: Spectra of native A.) dephosphorylated and B.) phosphorylated α-synuclein in positive ion mode with 9+ mobiligrams for C.) dephosphorylated and D.) phosphorylated α-synuclein, and spectra of native E.) dephosphorylated and F.) phosphorylated α-synuclein in negative ion mode with 9- mobiligrams for G.) dephosphorylated and H.) phosphorylated α-synuclein.

To further investigate how phosphorylation affects the structure of α-synuclein, dephosphorylated and phosphorylated α-synuclein were dissolved in native conditions and IM-MS was performed in negative-ion mode. The charge state distribution of dephosphorylated α-synuclein revealed a bimodal distribution with the most abundant charge states for each maximum being 6-, and 14-. (Fig. 5E) The charge state distribution of phosphorylated α-synuclein still revealed a bimodal distribution; however, the higher charge

states shifted so that the maxima were present at 17- and 6-. (Fig. 5F) The mobiligram for the 9- charge state of dephosphorylated α-synuclein reveals two conformers at drift time of 11.80ms and 13.01ms. (Fig. 5G) The mobiligram for the 9- charge state of phosphorylated α-synuclein revealed an increase in abundance of the early arriving conformer. (Fig. 5H) Ion mobility for the 10- charge state also revealed the presence of early arriving conformers when phosphorylated. (Fig. S18A) IM-MS of the 15- charge state for dephosphorylated and phosphorylated α-synuclein does not shift significantly in drift time or shape when phosphorylated. (Fig. S18B) The spectrum of denatured phosphorylated α-synuclein reveals a similar distribution to dephosphorylated α-synuclein; however, the mobiligrams show an increase in abundance of early arriving conformers when α-synuclein is phosphorylated. (Fig. S19) The increase in abundance of early arriving conformers due to phosphorylation indicates that phosphorylation compacts native α-synuclein in negative-ion mode due to the presence of salt bridges on the protein.

Phosphorylation seems to compact the structure of proteins when sprayed in native conditions. Structural changes have often been linked to an alteration in protein function.[71] It is possible that phosphorylation could induce a structural change in proteins which could alter their function. Many charge states of α-synuclein seem to indicate a compaction due to phosphorylation at serine 129. Compaction of α-synuclein due to phosphorylation at this site could reveal a mechanism for protein aggregation. Because of the location of S129 on the protein sequence, it is likely that phosphorylation at this residue compacts the C-terminal region of the protein which may expose the aggregation region of the protein leading to the creation of oligomers, fibrils, and larger aggregates.

CONCLUSION

This study indicates that mass spectrometry can efficiently characterize phosphorylated proteins and provide information on how those phosphorylation sites alters the structure of the protein. Top-down fragmentation techniques such as CAD, EID, and ECD can pinpoint all 5 phosphorylation sites on β-casein, and serine 129 on α-synuclein. In addition, top-down fragmentation can determine how phosphorylation sites are removed with alkaline phosphatase. IM-MS of phosphorylated proteins in denaturing as well as

native conditions shows the rise of early arriving conformers compared to their respective dephosphorylated proteoforms. This suggests that phosphorylation of proteins promotes compaction of those proteins presumably due to salt bridges that form between negatively charged phosphate groups and positively charged side chains. This trend is also observed when serine 129 is phosphorylated on the disease related amyloid protein α-synuclein. Compaction of α-synuclein due to phosphorylation of serine 129 could expose the aggregation region of α-synuclein and provide a mechanism for protein aggregation and the onset of neurodegenerative diseases. This data indicates that structural perturbations on protein structure due to phosphorylation could be a mechanism that cells use to alter protein function and how diseases such as Parkinson's disease can develop.

REFERENCES

1. Humphrey, S. J.; James, D. E.; Mann, M., Protein phosphorylation: a major switch mechanism for metabolic regulation. *Trends in Endocrinology & Metabolism* **2015,** *26* (12), 676-687.

2. Tomioka, M.; Shimobayashi, M.; Kitabatake, M.; Ohno, M.; Kozutsumi, Y.; Oka, S.; Takematsu, H., Ribosomal protein uS7/Rps5 serine-223 in protein kinase-mediated phosphorylation and ribosomal small subunit maturation. *Scientific reports* **2018,** *8* (1), 1-14.

3. Alexander, K. E.; Rizkallah, R., Aurora A phosphorylation of YY1 during mitosis inactivates its DNA binding activity. *Scientific reports* **2017,** *7* (1), 1-13.

4. Blesneac, I.; Chemin, J.; Bidaud, I.; Huc-Brandt, S.; Vandermoere, F.; Lory, P., Phosphorylation of the Cav3. 2 T-type calcium channel directly regulates its gating properties. *Proceedings of the National Academy of Sciences* **2015,** *112* (44), 13705-13710.

5. Chen, X.; Gohain, N.; Zhan, C.; Lu, W.; Pazgier, M.; Lu, W., Structural basis of how stress-induced MDMX phosphorylation activates p53. *Oncogene* **2016,** *35* (15), 1919-1925.

6. Carlomagno, Y.; Chung, D.-e. C.; Yue, M.; Castanedes-Casey, M.; Madden, B. J.; Dunmore, J.; Tong, J.; DeTure, M.; Dickson, D. W.; Petrucelli, L., An acetylation–phosphorylation switch that regulates tau aggregation propensity and function. *Journal of Biological Chemistry* **2017,** *292* (37), 15277-15286.

7. Sato, H.; Arawaka, S.; Hara, S.; Fukushima, S.; Koga, K.; Koyama, S.; Kato, T., Authentically phosphorylated α-synuclein at Ser129 accelerates neurodegeneration in a rat model of familial Parkinson's disease. *Journal of Neuroscience* **2011,** *31* (46), 16884-16894.

8. Grundke-Iqbal, I.; Iqbal, K.; Tung, Y.-C.; Quinlan, M.; Wisniewski, H. M.; Binder, L. I., Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proceedings of the National Academy of Sciences* **1986,** *83* (13), 4913-4917.

9. Farrell, H.; Qi, P.; Wickham, E.; Unruh, J., Secondary structural studies of bovine caseins: structure and temperature dependence of β-casein phosphopeptide (1-25) as analyzed by circular dichroism, FTIR spectroscopy, and analytical ultracentrifugation. *Journal of protein chemistry* **2002,** *21* (5), 307-321.

10. Rieloff, E.; Skepö, M., Phosphorylation of a Disordered Peptide—Structural Effects and Force Field Inconsistencies. *Journal of Chemical Theory and Computation* **2020,** *16* (3), 1924-1935.

11. Ettah, I.; Ashton, L., Determination of phosphorylation and deprotonation induced higher order structural transitions in αs-caseins. *Analytical chemistry* **2019,** *91* (21), 13940-13946.

12. Ruotolo, B. T.; Verbeck IV, G. F.; Thomson, L. M.; Woods, A. S.; Gillig, K. J.; Russell, D. H., Distinguishing between phosphorylated and nonphosphorylated peptides with ion mobility− mass spectrometry. *Journal of proteome research* **2002,** *1* (4), 303-306.

13. Wong, S. E.; Bernacki, K.; Jacobson, M., Competition between intramolecular hydrogen bonds and solvation in phosphorylated peptides: Simulations with explicit and implicit solvent. *The Journal of Physical Chemistry B* **2005,** *109* (11), 5249-5258.

14. Morris, M.; Knudsen, G. M.; Maeda, S.; Trinidad, J. C.; Ioanoviciu, A.; Burlingame, A. L.; Mucke, L., Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. *Nature neuroscience* **2015,** *18* (8), 1183-1189.

15. Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A., High throughput characterization of combinatorial histone codes. *Molecular & Cellular Proteomics* **2009,** *8* (10), 2266-2284.

16. Garcia, B. A.; Shabanowitz, J.; Hunt, D. F., Analysis of protein phosphorylation by mass spectrometry. *Methods* **2005,** *35* (3), 256-264.

17. Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jørgensen, T. J., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular & cellular proteomics* **2005,** *4* (7), 873-886.

18. 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.

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

20. Leney, A. C.; Heck, A. J., Native mass spectrometry: what is in the name? *Journal of the American Society for Mass Spectrometry* **2016,** *28* (1), 5-13.

21. Wu, Z.; Tiambeng, T. N.; Cai, W.; Chen, B.; Lin, Z.; Gregorich, Z. R.; Ge, Y., Impact of phosphorylation on the mass spectrometry quantification of intact phosphoproteins. *Analytical chemistry* **2018,** *90* (8), 4935-4939.

22. Siuti, N.; Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nature methods* **2007,** *4* (10), 817-821.

23. Vöpel, T.; Bravo-Rodriguez, K.; Mittal, S.; Vachharajani, S.; Gnutt, D.; Sharma, A.; Steinhof, A.; Fatoba, O.; Ellrichmann, G.; Nshanian, M., Inhibition of huntingtin exon-1 aggregation by the molecular tweezer CLR01. *Journal of the American Chemical Society* **2017,** *139* (16), 5640-5643.

24. Acharya, S.; Safaie, B. M.; Wongkongkathep, P.; Ivanova, M. I.; Attar, A.; Klärner, F.-G.; Schrader, T.; Loo, J. A.; Bitan, G.; Lapidus, L. J., Molecular basis for preventing α-synuclein aggregation by a molecular tweezer. *Journal of Biological Chemistry* **2014,** *289* (15), 10727-10737.

25. 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.

26. 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.

27. Sinha, S.; Lopes, D. H.; Du, Z.; Pang, E. S.; Shanmugam, A.; Lomakin, A.; Talbiersky, P.; Tennstaedt, A.; McDaniel, K.; Bakshi, R., Lysine-specific molecular tweezers are broad-spectrum

inhibitors of assembly and toxicity of amyloid proteins. *Journal of the American Chemical Society* **2011,** *133* (42), 16958-16969.

28. Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A., Top or middle? Up or down? Toward a standard lexicon for protein top-down and allied mass spectrometry approaches. *Journal of The American Society for Mass Spectrometry* **2019,** *30* (7), 1149-1157.

29. Loo, J. A.; Edmonds, C. G.; Smith, R. D., Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. *Science* **1990,** *248* (4952), 201-204.

30. 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.

31. 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.

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

33. Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B., Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nature methods* **2019,** *16* (7), 587-594.

34. Zhang, J.; Guy, M. J.; Norman, H. S.; Chen, Y.-C.; Xu, Q.; Dong, X.; Guner, H.; Wang, S.; Kohmoto, T.; Young, K. H., Top-down quantitative proteomics identified phosphorylation of cardiac troponin I as a candidate biomarker for chronic heart failure. *Journal of proteome research* **2011,** *10* (9), 4054-4065.

35. Shi, S. D.-H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W., Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Analytical Chemistry* **2001,** *73* (1), 19-22.

36. Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W., Plasma electron capture dissociation for the characterization of large proteins by top down mass spectrometry. *Analytical chemistry* **2003,** *75* (7), 1599-1603.

37. Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature chemistry* **2014,** *6* (4), 281-294.

38. May, J. C.; Jurneczko, E.; Stow, S. M.; Kratochvil, I.; Kalkhof, S.; McLean, J. A., Conformational landscapes of ubiquitin, cytochrome c, and myoglobin: Uniform field ion mobility measurements in helium and nitrogen drift gas. *International journal of mass spectrometry* **2018,** *427*, 79-90.

39. Bush, M. F.; Campuzano, I. D.; Robinson, C. V., Ion mobility mass spectrometry of peptide ions: effects of drift gas and calibration strategies. *Analytical chemistry* **2012,** *84* (16), 7124-7130.

40. Hoaglund-Hyzer, C. S.; Clemmer, D. E., Ion trap/ion mobility/quadrupole/time-of-flight mass spectrometry for peptide mixture analysis. *Analytical chemistry* **2001,** *73* (2), 177-184.

41. Lermyte, F.; Williams, J. P.; Brown, J. M.; Martin, E. M.; Sobott, F., Extensive charge reduction and dissociation of intact protein complexes following electron transfer on a quadrupole-ion mobility-time-of-flight MS. *Journal of The American Society for Mass Spectrometry* **2015,** *26* (7), 1068-1076.

42. Campuzano, I. D.; Giles, K., Historical, current and future developments of travelling wave ion mobility mass spectrometry: A personal perspective. *TrAC Trends in Analytical Chemistry* **2019,** *120*, 115620.

43. Konijnenberg, A.; Ranica, S.; Narkiewicz, J.; Legname, G.; Grandori, R.; Sobott, F.; Natalello, A., Opposite structural effects of epigallocatechin-3-gallate and dopamine binding to α-synuclein. *Analytical chemistry* **2016,** *88* (17), 8468-8475.

44. Ruotolo, B. T.; Gillig, K. J.; Woods, A. S.; Egan, T. F.; Ugarov, M. V.; Schultz, J. A.; Russell, D. H., Analysis of phosphorylated peptides by ion mobility-mass spectrometry. *Analytical chemistry* **2004,** *76* (22), 6727-6733.

45. Thalassinos, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H., Characterization of phosphorylated peptides using traveling wave-based and drift cell ion mobility mass spectrometry. *Analytical chemistry* **2009,** *81* (1), 248-254.

46. Kerr, T. J.; Gant-Branum, R. L.; McLean, J. A., Multiplexed analysis of peptide functionality using lanthanide-based structural shift reagents. *International journal of mass spectrometry* **2011,** *307* (1-3), 28-32.

47. Glover, M. S.; Dilger, J. M.; Acton, M. D.; Arnold, R. J.; Radivojac, P.; Clemmer, D. E., Examining the influence of phosphorylation on peptide ion structure by ion mobility spectrometry-mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2016,** *27* (5), 786-794.

48. Byrne, D. P.; Vonderach, M.; Ferries, S.; Brownridge, P. J.; Eyers, C. E.; Eyers, P. A., cAMP-dependent protein kinase (PKA) complexes probed by complementary differential scanning fluorimetry and ion mobility–mass spectrometry. *Biochemical Journal* **2016,** *473* (19), 3159-3175.

49. 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.

50. 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.

51. Kamiński, S.; Cieślińska, A.; Kostyra, E., Polymorphism of bovine beta-casein and its potential effect on human health. *Journal of applied genetics* **2007,** *48* (3), 189-198.

52. 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. *Journal of Mass Spectrometry* **2019,** *54* (6), 527-539.

53. Konermann, L., A minimalist model for exploring conformational effects on the electrospray charge state distribution of proteins. *The Journal of Physical Chemistry B* **2007,** *111* (23), 6534-6543.

54. Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Analytical chemistry* **2010,** *82* (22), 9557-9565.

55. Picache, J. A.; Rose, B. S.; Balinski, A.; Leaptrot, K. L.; Sherrod, S. D.; May, J. C.; McLean, J. A., Collision cross section compendium to annotate and predict multi-omic compound identities. *Chemical science* **2019,** *10* (4), 983-993.

56. Errington, N.; Doig, A. J., A phosphoserine− lysine salt bridge within an α-helical peptide, the strongest α-helix side-chain interaction measured to date. *Biochemistry* **2005,** *44* (20), 7553-7558.

57. Woods, A. S.; Ferré, S., Amazing stability of the arginine− phosphate electrostatic interaction. *Journal of proteome research* **2005,** *4* (4), 1397-1402.

58. Bailey, L. S.; Alves, M.; Galy, N.; Patrick, A. L.; Polfer, N. C., Mechanistic insights into intramolecular phosphate group transfer during collision induced dissociation of phosphopeptides. *Journal of Mass Spectrometry* **2019,** *54* (5), 449-458.

59. Zhang, Z.; Browne, S. J.; Vachet, R. W., Exploring salt bridge structures of gas-phase protein ions using multiple stages of electron transfer and collision induced dissociation. *Journal of The American Society for Mass Spectrometry* **2014,** *25* (4), 604-613.

60. Julian, R. R.; Hodyss, R.; Beauchamp, J., Salt bridge stabilization of charged zwitterionic arginine aggregates in the gas phase. *Journal of the American Chemical Society* **2001,** *123* (15), 3577-3583.

61. Hall, Z.; Hernández, H.; Marsh, J. A.; Teichmann, S. A.; Robinson, C. V., The role of salt bridges, charge density, and subunit flexibility in determining disassembly routes of protein complexes. *Structure* **2013,** *21* (8), 1325-1337.

62. Liko, I.; Hopper, J. T.; Allison, T. M.; Benesch, J. L.; Robinson, C. V., Negative ions enhance survival of membrane protein complexes. *Journal of the American Society for Mass Spectrometry* **2016,** *27* (6), 1099-1104.

63. Watt, S. J.; Oakley, A.; Sheil, M. M.; Beck, J. L., Comparison of negative and positive ion electrospray ionization mass spectra of calmodulin and its complex with trifluoperazine. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* **2005,** *19* (15), 2123-2130.

64. Beychok, S.; De Lozé, C.; Blout, E., Helix contents of solutions of native and acid-denatured ferrihemoglobin and ferrimyoglobin. *Journal of Molecular Biology* **1962,** *4* (6), 421-429.

65. Czernick, D.; Liu, J.; Serge, D.; Salih, E., Topographical distribution of phosphorylation sites of phosvitins by mass spectrometry. *Journal of proteomics* **2013,** *83*, 76-98.

66. Raikos, V.; Hansen, R.; Campbell, L.; Euston, S. R., Separation and identification of hen egg protein isoforms using SDS–PAGE and 2D gel electrophoresis with MALDI-TOF mass spectrometry. *Food chemistry* **2006,** *99* (4), 702-710.

67. Samaraweera, H.; Moon, S. H.; Lee, E. J.; Grant, J.; Fouks, J.; Choi, I.; Suh, J. W.; Ahn, D. U., Characterisation of phosvitin phosphopeptides using MALDI-TOF mass spectrometry. *Food chemistry* **2014,** *165*, 98-103.

68. Yang, Y.; Barendregt, A.; Kamerling, J. P.; Heck, A. J., Analyzing protein micro-heterogeneity in chicken ovalbumin by high-resolution native mass spectrometry exposes qualitatively and semi-quantitatively 59 proteoforms. *Analytical chemistry* **2013,** *85* (24), 12037-12045.

69. Marques, O.; Outeiro, T., Alpha-synuclein: from secretion to dysfunction and death. *Cell death & disease* **2012,** *3* (7), e350-e350.

70. Stephens, A. D.; Zacharopoulou, M.; Moons, R.; Fusco, G.; Seetaloo, N.; Chiki, A.; Woodhams, P. J.; Mela, I.; Lashuel, H. A.; Phillips, J. J., Extent of N-terminus exposure of monomeric alpha-synuclein determines its aggregation propensity. *Nature communications* **2020,** *11* (1), 1-15.

71. Krishnamurthy, H.; Gouaux, E., X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature* **2012,** *481* (7382), 469-474.

SUPPORTING INFORMATION

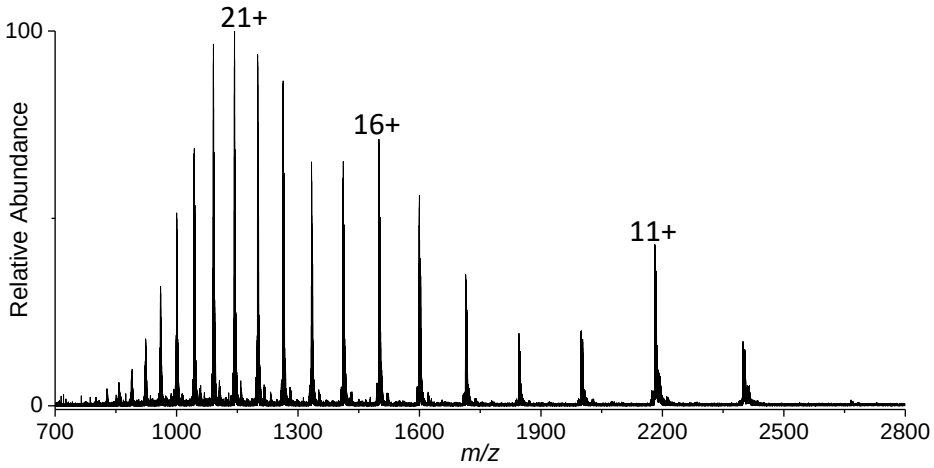


Figure S1: A FT-ICR spectrum of native β-casein containing all 5 phosphorylation sites. The wide charge state distribution indicates that this protein is intrinsically disordered and the trimodal distribution indicates that multiple conformations of this protein exist in native conditions.

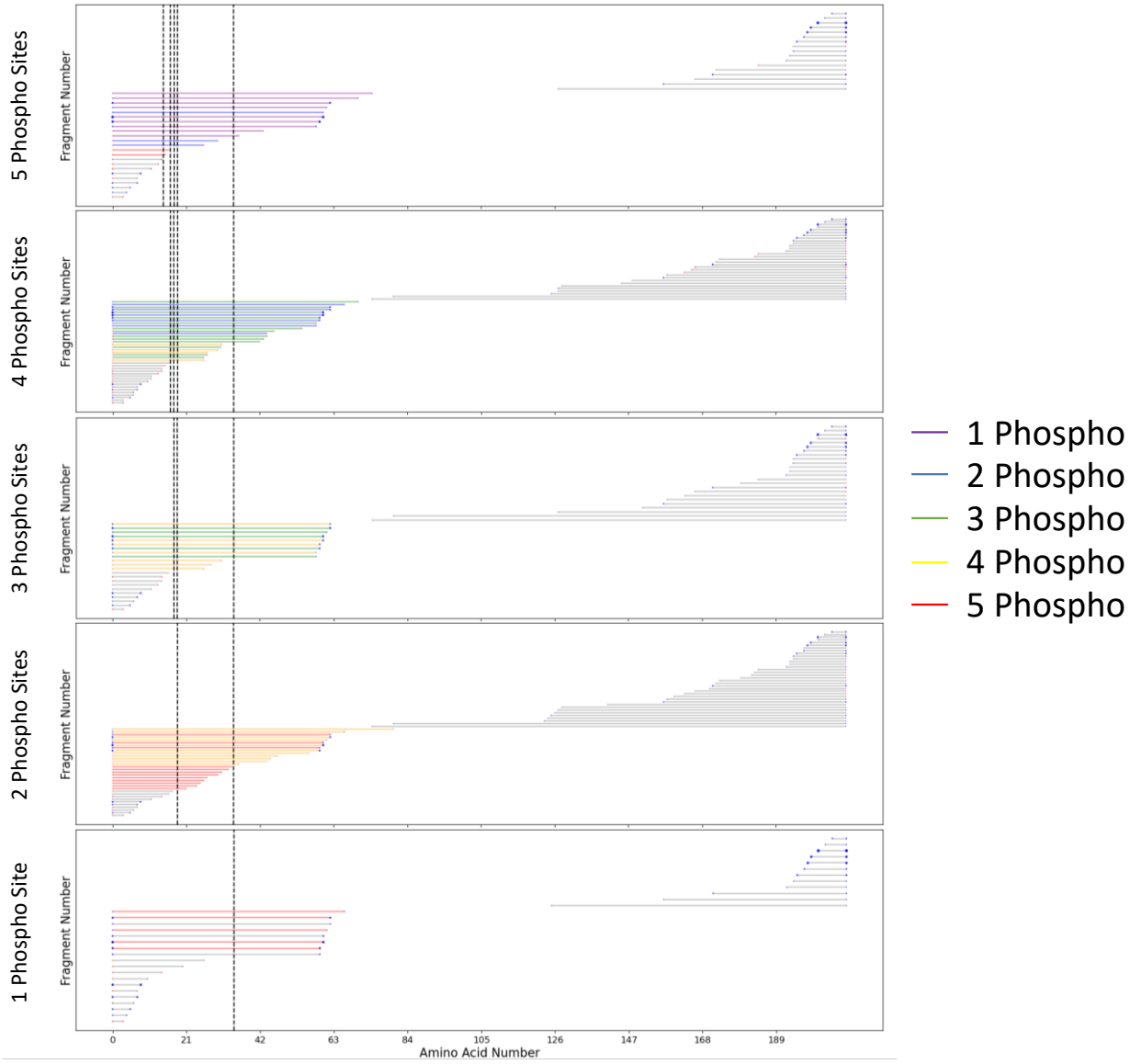


Figure S2: CAD/EID fragmentation of each phosphorylation state of β-Casein when dephosphorylated by alkaline phosphatase. The vertical dotted lines indicate the phosphorylation sites that are still present on the protein.

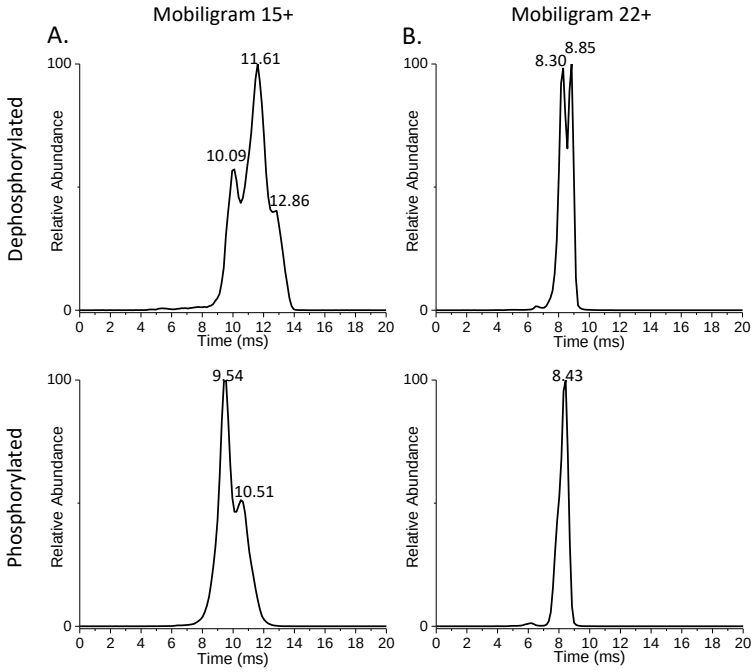


Figure S3: Mobiligrams of A.) the 15+ charge state and B.) the 22+ charge state for phosphorylated and dephosphorylated denatured β-casein. Notice how the mobiligrams reveal that early arriving conformers are favored when β-casein is phosphorylated.

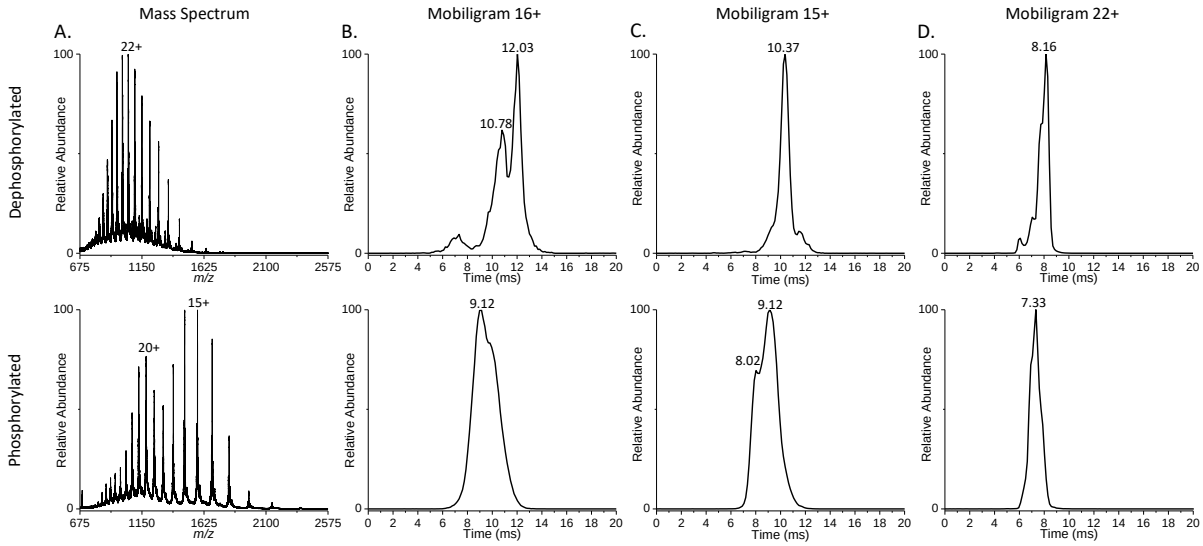


Figure S4: A.) Spectra of dephosphorylated and phosphorylated α-casein (23.0kDa) analyzed under denaturing conditions in positive-ion mode, and the corresponding B.) 16+, C.) 15+, and D.) 22+ mobiligrams for the two proteoforms. Notice that phosphorylated α-casein displays a different charge state distribution compared to dephosphorylated α-casein, and that mobiligrams indicate early arriving conformers are greater in abundance when α-casein is phosphorylated.

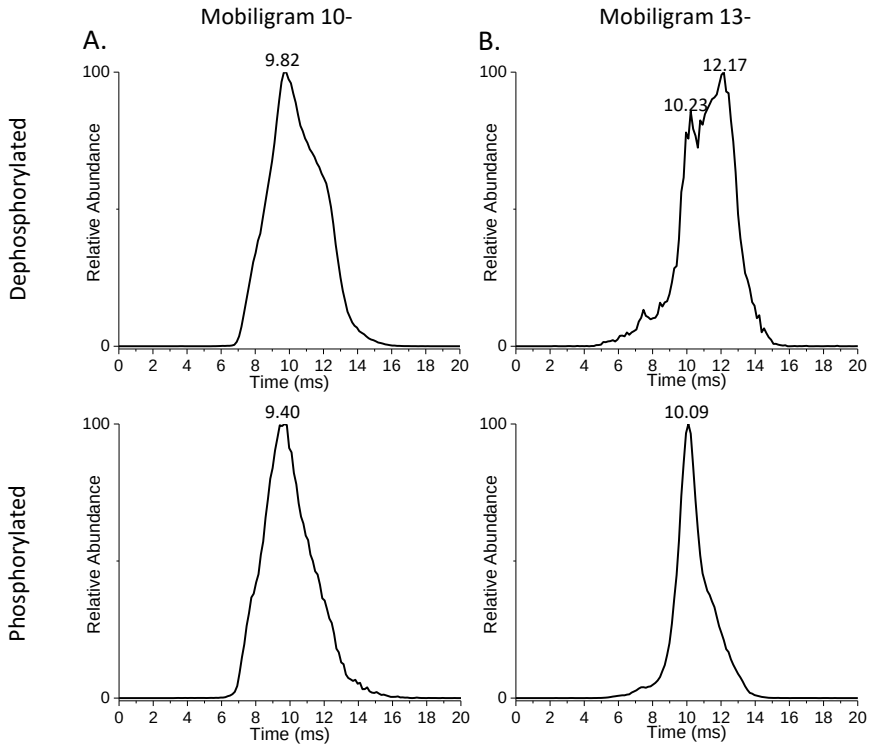


Figure S5: Mobiligrams of A.) the 10- charge state and B.) the 13- charge state for phosphorylated and dephosphorylated denatured β-casein. Notice how the mobiligrams reveal that early arriving conformers are favored when β-casein is phosphorylated.

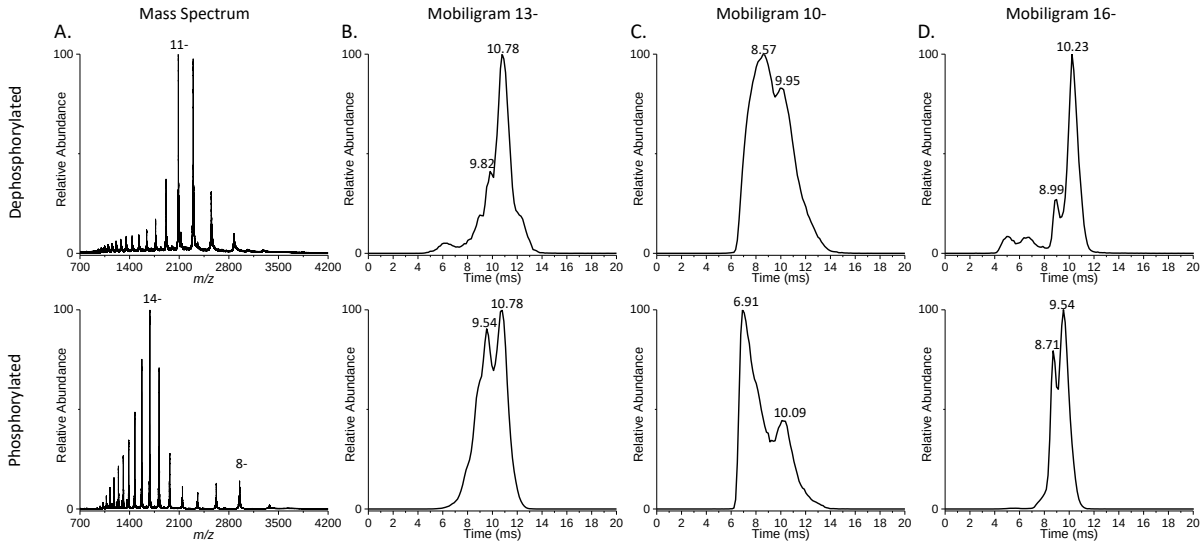


Figure S6: A.) Spectra of dephosphorylated and phosphorylated α-casein (23.0kDa) analyzed under denaturing conditions in negative-ion mode, and the corresponding B.) 13-, C.) 10-, and D.) 16- mobiligrams for the two proteoforms. Notice that phosphorylated α-casein displays a different charge state distribution compared to dephosphorylated α-casein, and that mobiligrams indicate early arriving conformers are greater in abundance when α-casein is phosphorylated.

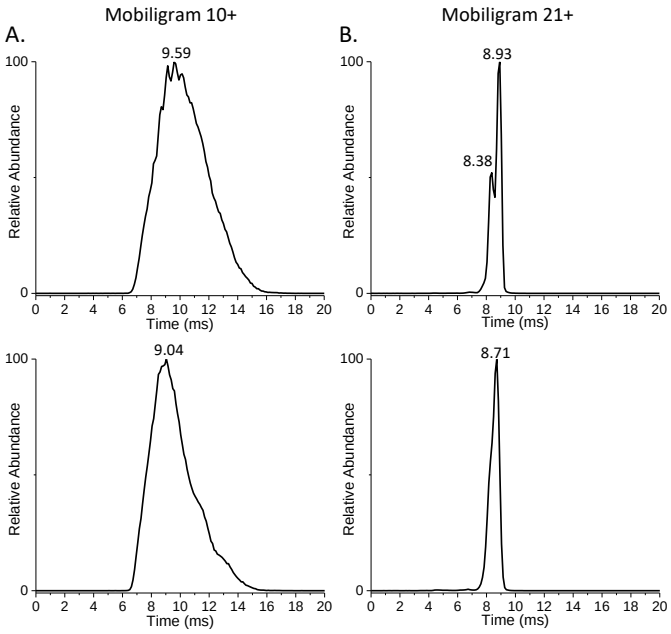


Figure S7: Mobiligrams of A.) the 10+ charge state and B.) the 21+ charge state for phosphorylated and dephosphorylated native β-casein. Notice how the 10+ mobiligrams reveal that early arriving conformers are favored when β-casein is phosphorylated. The 21+ does not shift considerably when phosphorylated possibly due to the accumulation of charge on the protein.

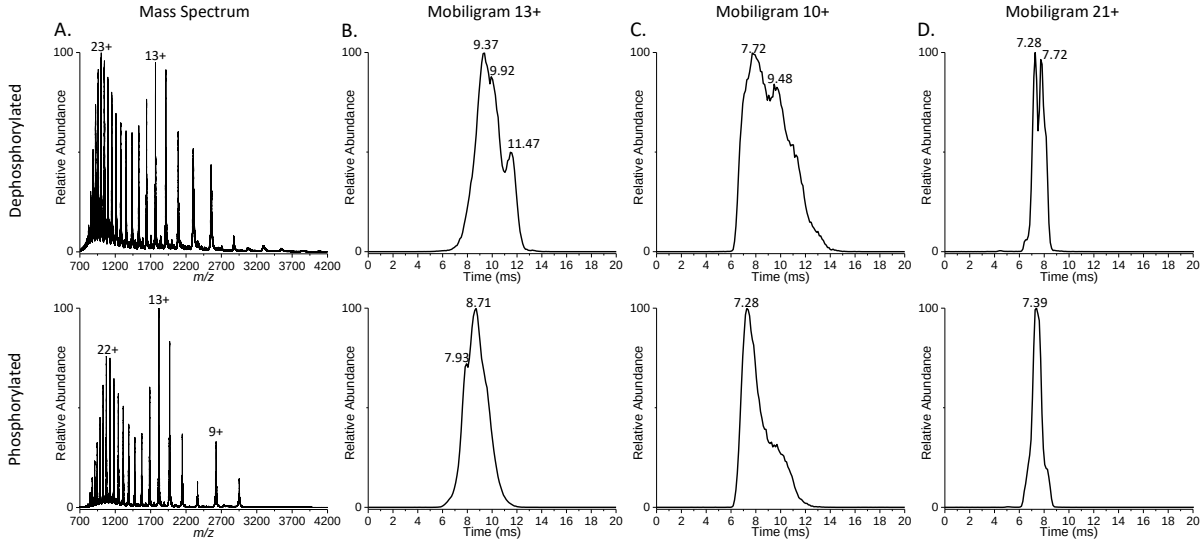


Figure S8: A.) Spectra of dephosphorylated and phosphorylated α-casein (23.0kDa) analyzed under native conditions in positive-ion mode, and the corresponding B.) 13+, C.) 10+, and D.) 21+ mobiligrams for the two proteoforms. Notice that phosphorylated α-casein displays a slightly different charge state distribution compared to dephosphorylated α-casein, and that mobiligrams indicate early arriving conformers are greater in abundance when α-casein is phosphorylated.

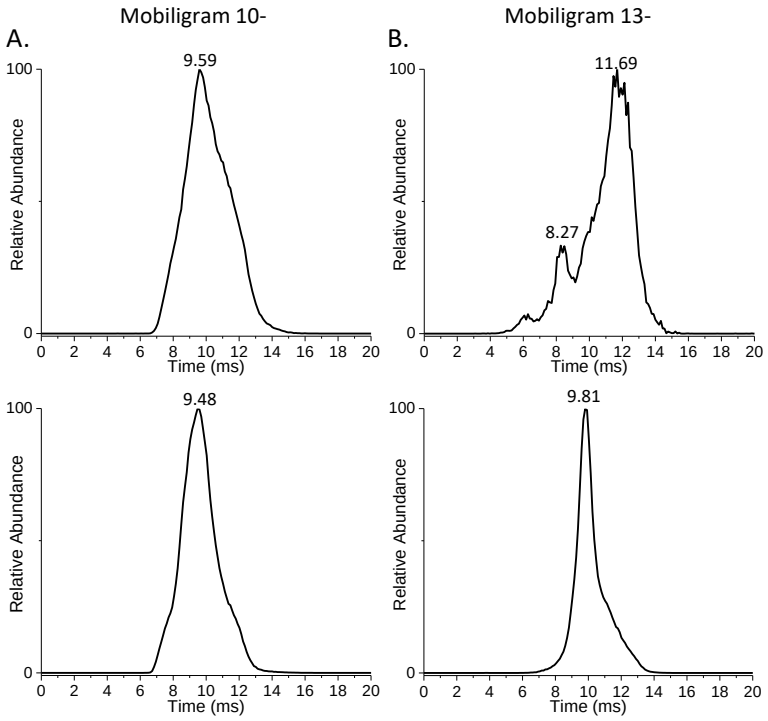


Figure S9: Mobiligrams of A.) the 10- charge state and B.) the 13- charge state for phosphorylated and dephosphorylated native β-casein. Notice how the mobiligrams reveal that early arriving conformers are favored when β-casein is phosphorylated.

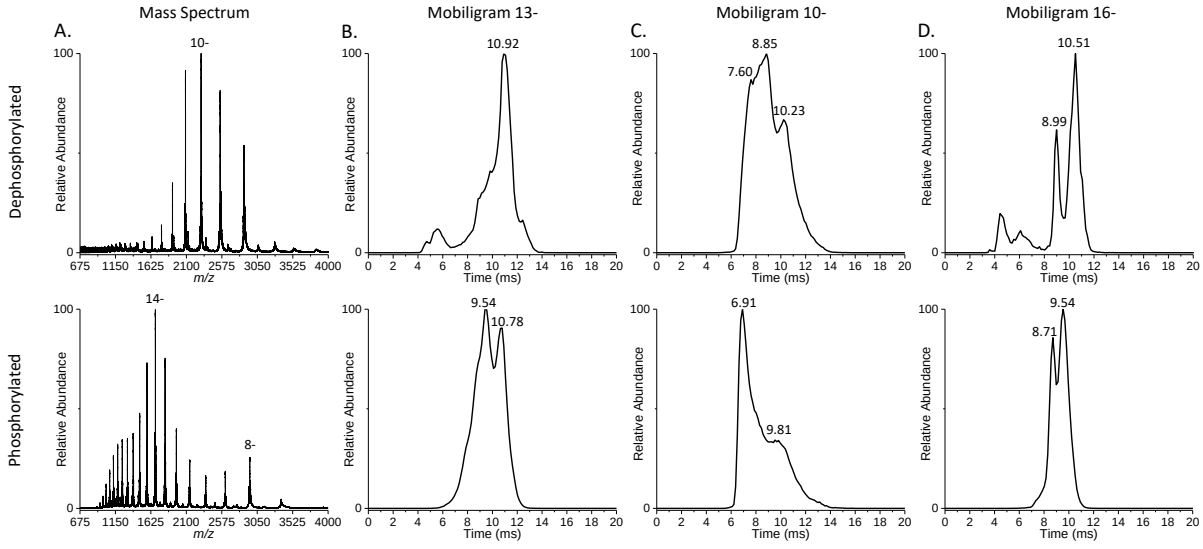


Figure S10: A.) Spectra of dephosphorylated and phosphorylated α-casein (23.0kDa) analyzed under native conditions in negative-ion mode, and the corresponding B.) 13-, C.) 10-, and D.) 16- mobiligrams for the two proteoforms. Notice that phosphorylated α-casein displays a different charge state distribution compared to dephosphorylated α-casein, and that mobiligrams indicate early arriving conformers are greater in abundance when α-casein is phosphorylated.

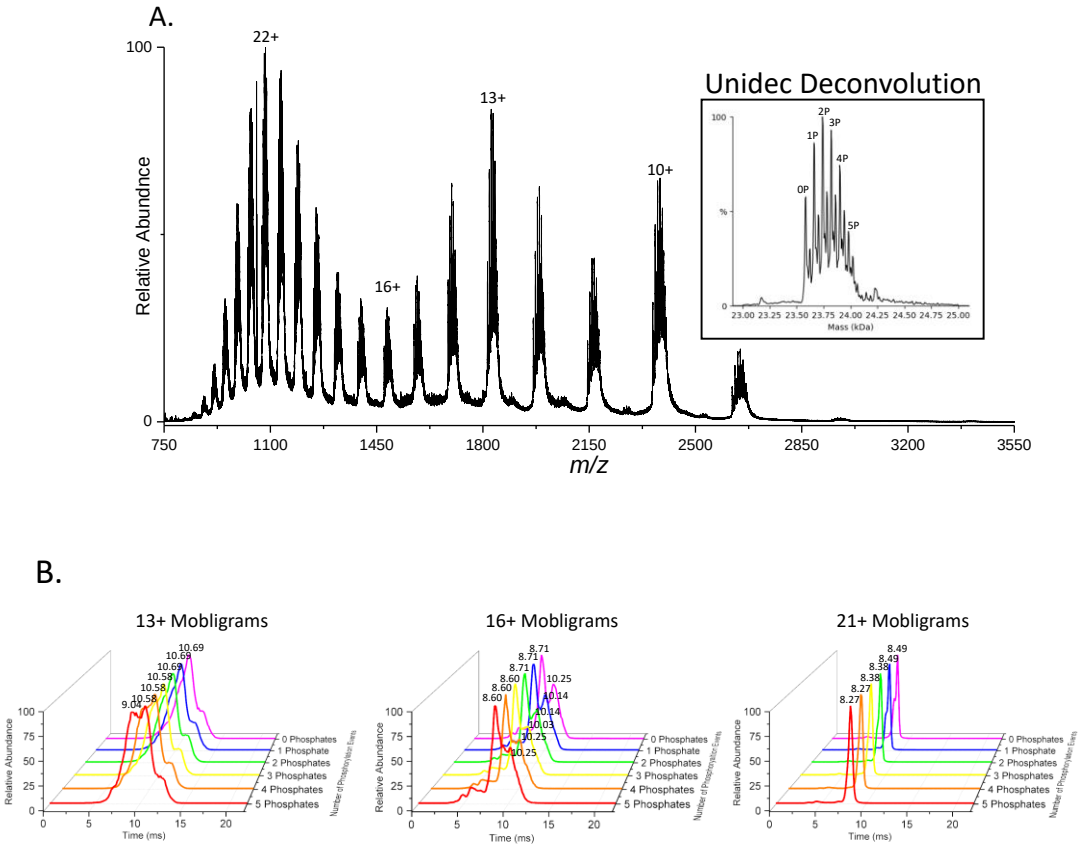


Figure S11: A.) A spectrum of native β-casein in positive-ion mode containing all 6 phosphorylation proteoforms and a UniDec deconvolution of the spectrum indicating all 5 phosphorylation sites can be observed. B.) Mobiligrams for all phosphorylation states of β-Casein for the 12+, 16+, and 21+ charge states. The mobiligrams for the 12+ and 16+ reveal incremental compaction when more phosphorylation sites are present on the protein. The 21+ mobiligram does not reveal much change; however, the drift time does decrease when more phosphates are present on the protein.

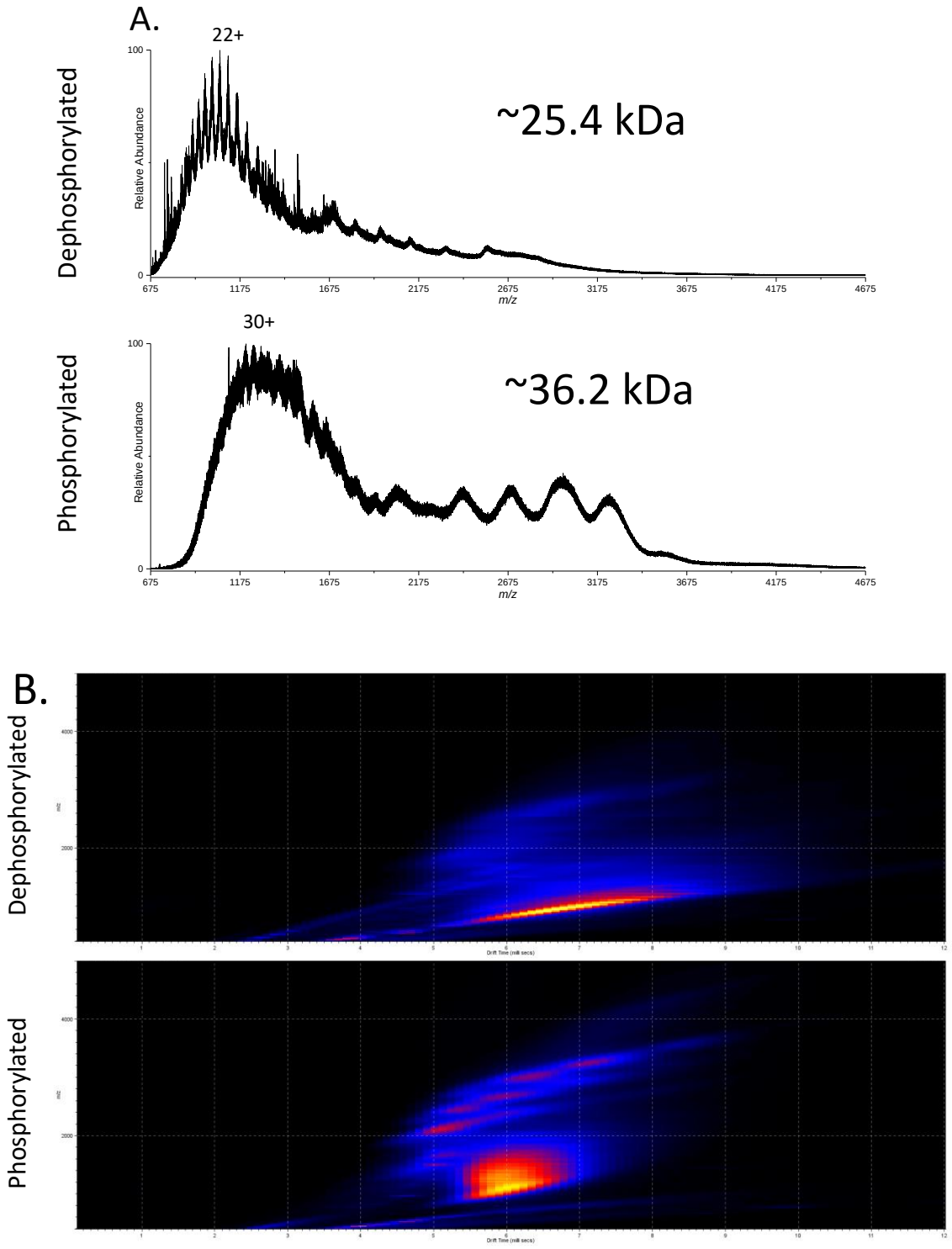


Figure S12: A.) Spectra of phosphorylated and dephosphorylated phosvitin and B.) drift plots of dephosphorylated (25.47kDa) and phosphorylated (36.19kDa) phosvitin, an intrinsically disordered phosphoprotein, in native solution positive-ion mode. The drift plots show that the structure of phosphorylated phosvitin is compacted in the gas-phase (narrower range of drift times).

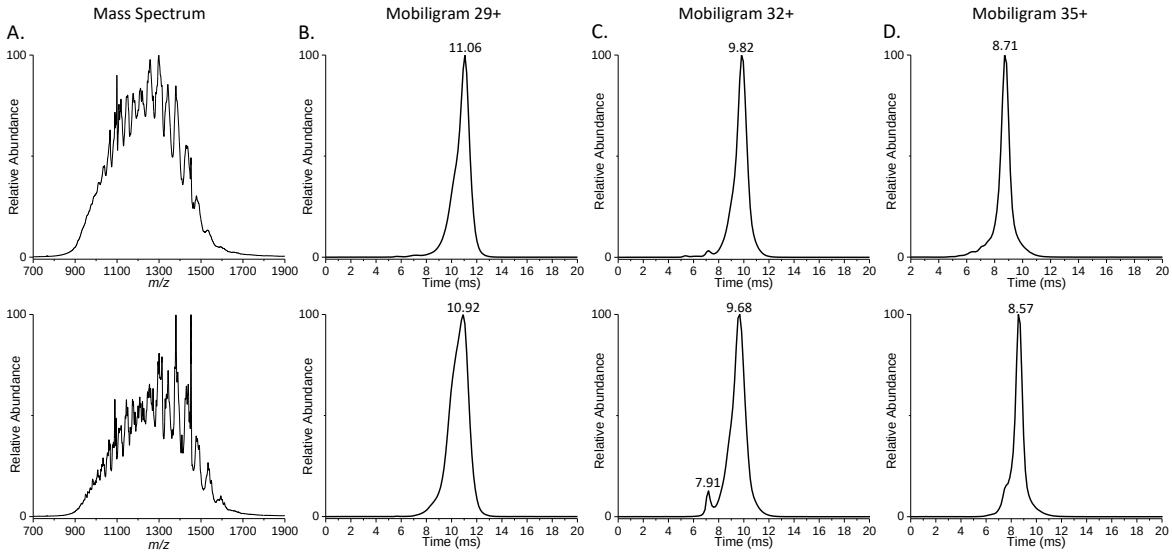


Figure S13: A.) Spectra of dephosphorylated and phosphorylated Ovalbumin (~45kDa) analyzed under denatured conditions in positive-ion mode, and the corresponding B.) 29+, C.) 32+, and D.) 35+ mobiligrams for the two proteoforms. Notice that phosphorylated ovalbumin displays a different charge state distribution compared to dephosphorylated α-casein, and that mobiligrams indicate a shift toward lower drift time when ovalbumin is phosphorylated.

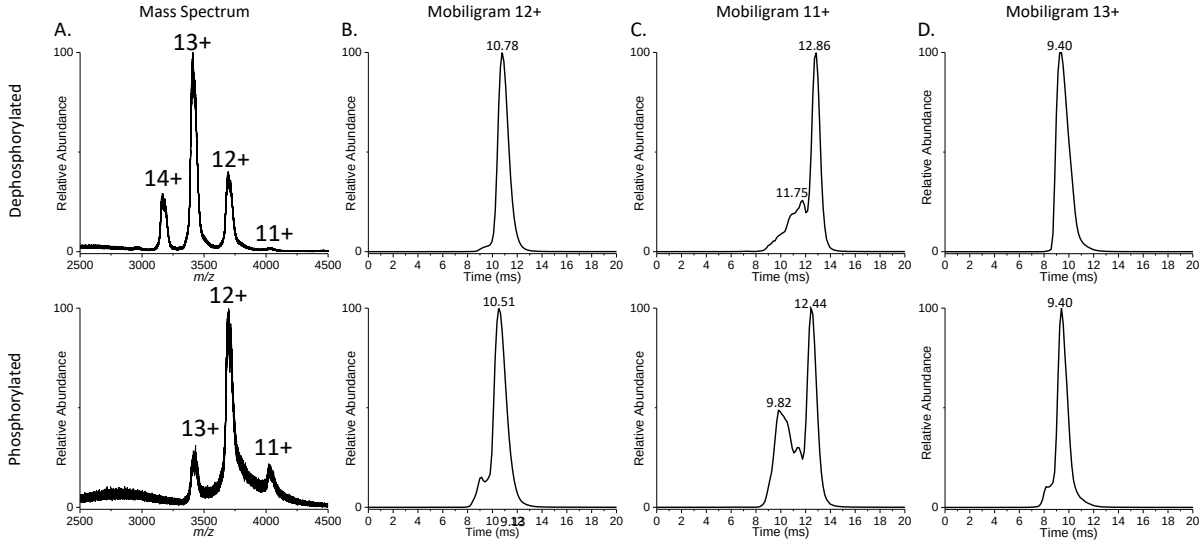


Figure S14: A.) Spectra of dephosphorylated and phosphorylated ovalbumin (~45kDa) analyzed under native conditions in positive-ion mode, and the corresponding B.) 12+, C.) 11+, and D.) 13+ mobiligrams for the two proteoforms. Notice that the charge state distribution for phosphorylated ovalbumin shifts toward lower charge states compared to dephosphorylated α-casein, and that mobiligrams reveal an increase in abundance of early arriving conformers when phosphorylated.

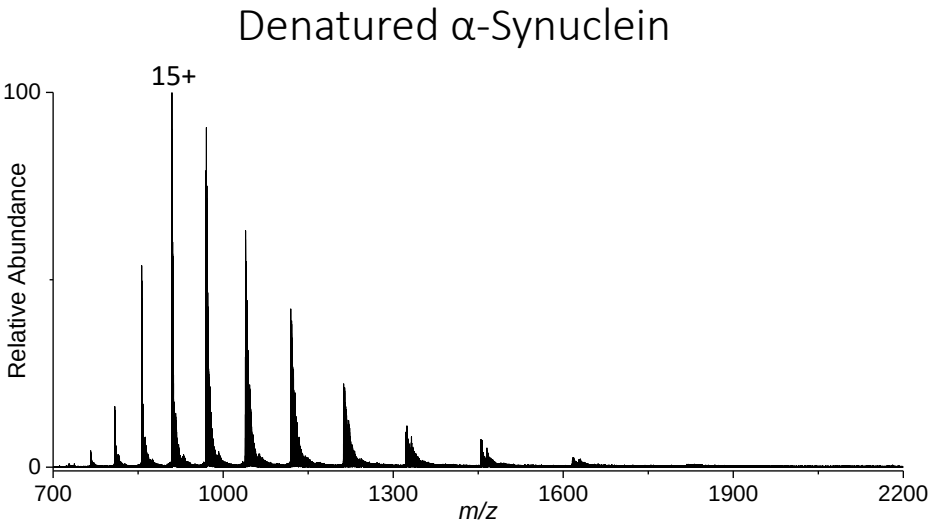


Figure S15: FT-ICR spectrum of denatured α-synuclein phosphorylated at serine 129.

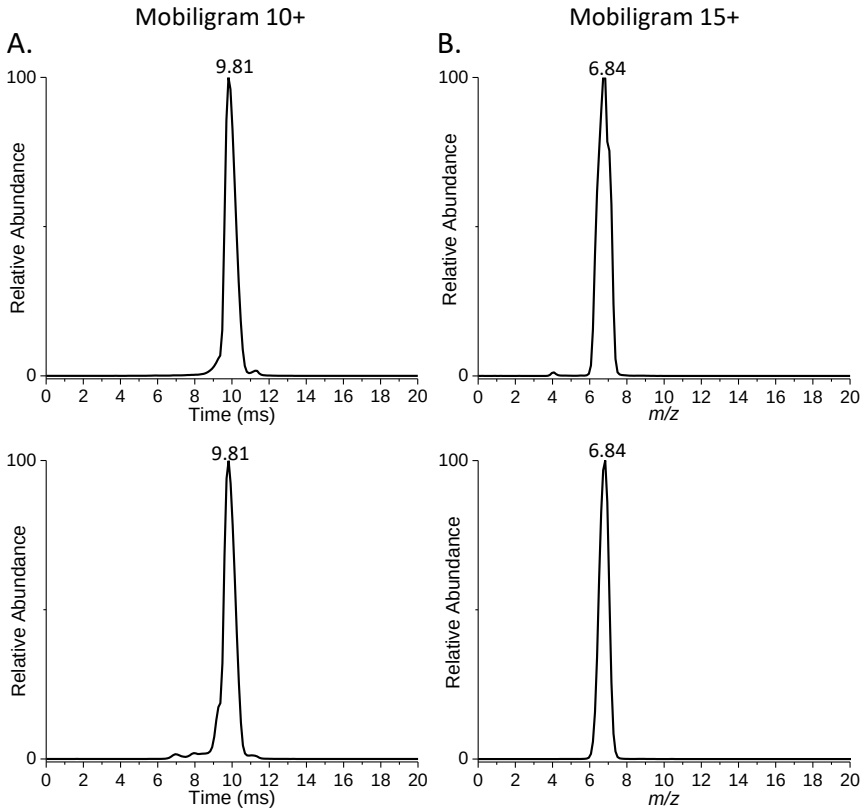


Figure S16: Mobiligrams of A.) the 10+ charge state and B.) the 15+ charge state for phosphorylated and dephosphorylated native α-synuclein. Notice how the higher charge states of α-synuclein do not shift considerably when phosphorylated.

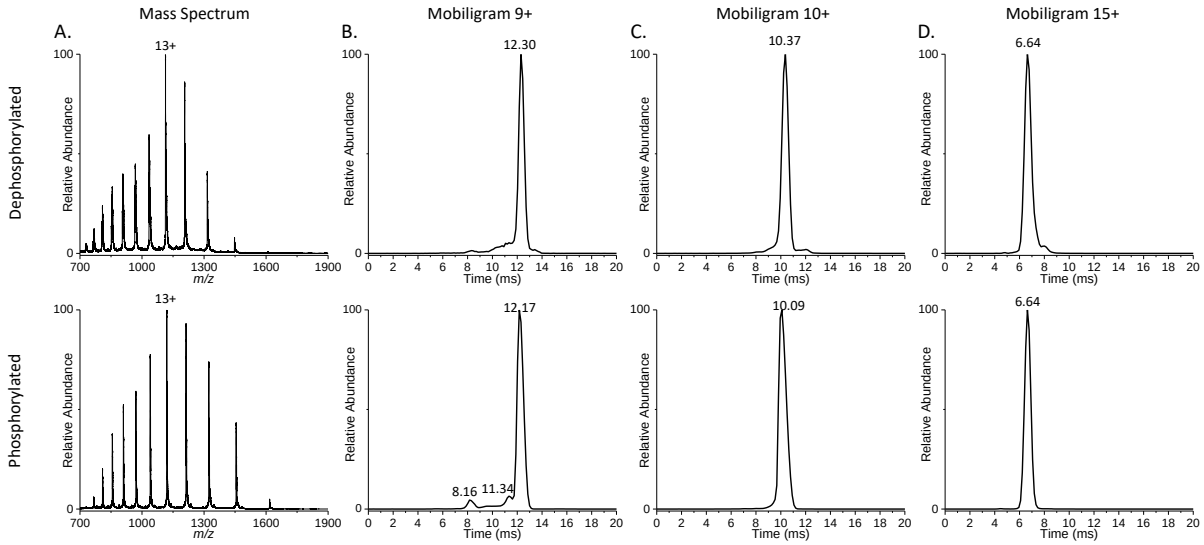


Figure S17: A.) Spectra of dephosphorylated and phosphorylated α-casein (14.5kDa) analyzed under denatured conditions in positive-ion mode, and the corresponding B.) 9+, C.) 10+, and D.) 15+ mobiligrams for the two proteoforms. Notice that the spectra and the mobiligrams do not shift considerably due to phosphorylation of α-synuclein.

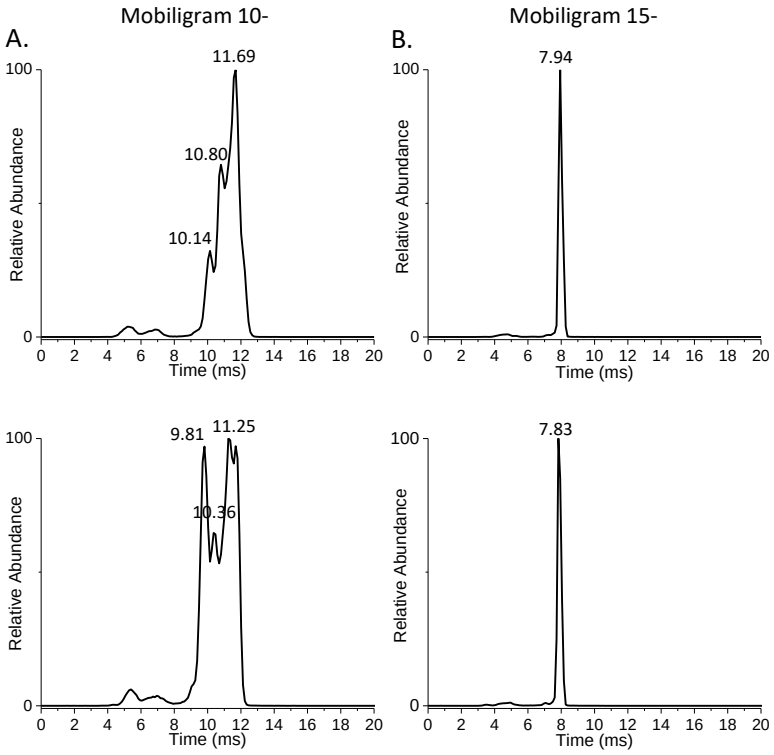


Figure S18: Mobiligrams of A.) the 10- charge state and B.) the 15- charge state for phosphorylated and dephosphorylated native α-synuclein. Notice how the mobiligrams reveal that early arriving conformers are favored when α-synuclein is phosphorylated.

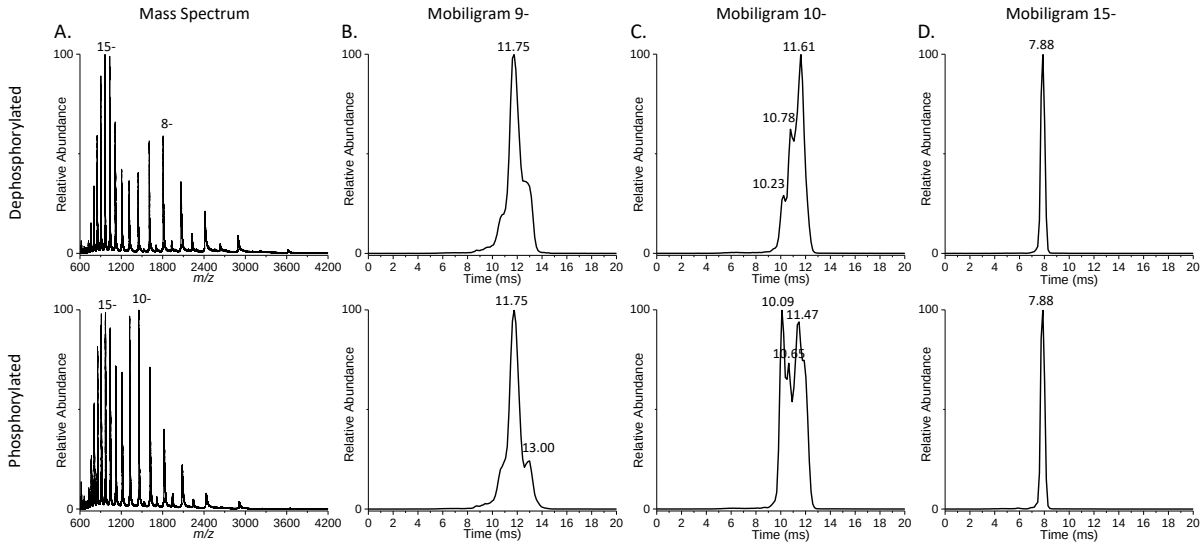


Figure S19: A.) Spectra of dephosphorylated and phosphorylated α-casein (14.5kDa) analyzed under denatured conditions in negative-ion mode, and the corresponding B.) 9-, C.) 10-, and D.) 15- mobiligrams for the two proteoforms. Notice how the 9- and the 10- mobiligrams reveal that early arriving conformers are favored when β-casein is phosphorylated. The 15- does not shift considerably when phosphorylated possibly due to the accumulation of charge on the protein.

CHAPTER 6

Characterization of CLR01 Binding on α-Synuclein with Native Top-Down Mass Spectrometry and Ion Mobility-Mass Spectrometry Reveals a Mechanism for Aggregation Inhibition

Carter Lantz[1], Jaybree Lopez[1], Andrew K. Goring[1], Muhammad A. Zenaidee[1], Karl Biggs[2], Julian Whitelegge[3], Gal Bitan[2], Rachel Loo[4], Joseph A. Loo[1,4]

[1]Department of Chemistry and Biochemistry, [2]Department of Neurology and Brain Research Institute, David Geffen School of Medicine, [3]The Pasarow Mass Spectrometry Laboratory, The Jane and Terry Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, [4]Department of Biological Chemistry, University of California-Los Angeles, Los Angeles, CA 90095

ABSTRACT: Parkinson's disease, a neurodegenerative disease that affects 15 million people worldwide, is characterized by aggregation of α-synuclein into Lewy Bodies in the neurons. Although this disease is prevalent in the population, a therapy or cure has yet to be found. Some small aromatic compounds have been reported to disrupt fibril formation. Among these compounds is a molecular tweezer compound known as CLR01 that binds lysine and arginine residues. This study aims to characterize how CLR01 interacts with various proteoforms of α-synuclein and how the structure of α-synuclein shifts due to that interaction. Native mass spectrometry (nMS) analysis of α-synuclein/CLR01 complexes reveals that multiple CLR01 molecules can bind α-synuclein with K_d values in the micromolar range, and that the binding of 1 CLR01 molecule shifts the ability for α-synuclein to bind other ligands. Electron capture dissociation (ECD) with Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometry (MS) of α-synuclein proteoforms bound with CLR01 can localize the sites of modification and indicates that CLR01 binds these proteoforms on the N-terminus. Ion mobility-mass spectrometry (IM-MS) reveals that CLR01 binding to α-synuclein compacts the structure of α-synuclein in the gas-phase. This data could indicate that when multiple CLR01 molecules bind the N-terminus of α-synuclein, the structure of the N-terminus of α-synuclein shifts toward a more compact state. This mechanism could explain why CLR01 binding prevents the formation of oligomers and fibrils and halts the progression of neurodegenerative diseases.

Keywords: Native Mass Spectrometry, Top-Down Mass Spectrometry, Ion-Mobility Mass Spectrometry, Proteoform, α-Synuclein

INTRODUCTION

Neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease are characterized by the formation of protein aggregates in the brain. Aggregates of amyloid proteins such as tau,[1] huntingtin,[2] and ataxin[3] have been correlated with the death of brain neurons and the onset of neurodegenerative disease symptoms. α-Synuclein (α-syn), a 14kDa intrinsically disordered amyloid protein, has been proposed to regulate synapse formation in neurons.[4, 5] Monomers of α-synuclein bind together in tight beta-sheet structures to form oligomers and fibrils[6, 7] and eventually form Lewy bodies in the brains of Parkinson's disease patients.[8] Although aggregation of α-syn has been studied extensively, the reason for this aggregation is largely unknown and a cure for this disease has not been found.

Although the reason for α-syn aggregation is not known, certain mutations and modifications have been found to modulate the rate of α-syn aggregation. For example, it has been reported that mutations such as A30P and A53T increase the rate of α-syn aggregation.[9] Post-translational modifications such as N-terminal acetylation have been reported to decrease the rate of α-syn aggregation[9-11] while phosphorylation at serine 129 has been correlated with increased α-synuclein aggregation.[12] In addition, the presence of metal ions such as copper[13] and manganese[14] have been linked to increased aggregation of α-syn. A shift in α-synuclein aggregation due to modifications is well documented; however, the reason for this alteration in aggregation potential is not well understood.

Although α-synuclein has been studied extensively, there is no therapy or cure for synucleinopathies. A multitude of small aromatic compounds have been found to have some inhibitory effect on fibril formation and disrupt the structures of existing fibrils.[15-19] Among these compounds is a small aromatic molecule known as CLR01. CLR01 is a molecular tweezer compound that binds lysine and arginine residues. We have previously found that CLR01 binds various amyloid proteins including α-synuclein.[20-24] Although some research has been done on CLR01 binding of α-synuclein, there is not much known about the binding interaction of CLR01 on different proteoforms of α-syn or the structural effect of CLR01 binding on the protein.

Over the past few decades, intact protein mass spectrometry (MS) has become a useful technique to characterize proteoforms of proteins and protein/ligand complexes. MS analysis of protein solutions has been utilized to determine the number of PTMs on a protein and the quantity of those PTMs on α-syn such as oxidation,[25] acetylation,[26] and phosphorylation.[27] In addition, native MS has been shown to readily determine binding of ligands on amyloid proteins including CLR01[21, 22] and metal ions.[13, 28, 29] MS1 analysis of proteins has found to provide valuable insight into the types of modifications present on proteins in solution and the extent of those modifications.

Electron capture dissociation (ECD) with Fourier transform-ion cyclotron resonance (FT-ICR) MS can be used to return relevant information on proteins including amyloid proteins. ECD is a nonergodic fragmentation technique that utilizes a beam of electrons to dissociate the peptide backbone without disrupting other intermolecular or intramolecular interactions.[30-32] Due to its nonergodic nature, ECD has been used for a variety of applications including protein sequencing, modification localization, and characterization of protein structure.[33] FT-ICR MS has been shown to provide superior mass accuracy and resolution for various compounds.[34] It has become a viable technique for collecting high resolution data of protein fragmentation from proteins[31] and protein complexes.[35] ECD in combination FT-ICR MS has previously been found to accurately pinpoint the binding of CLR01[20-24] and metal ions[13, 29] on various amyloid proteins including α-syn. In addition to locating the site of binding, it has been shown that TD-MS of can help simulate shifts in protein structure. Recent studies have shown that locating CLR01 binding with TD-MS can guide molecular dynamic simulations which revealed that CLR01 binding can shift the structure of N17-Huntingin and α-synuclein.[20, 22]

Ion mobility mass spectrometry (IM-MS) can be used to determine the overall 3D structure of intrinsically disordered proteins in the gas-phase. IM-MS can be utilized to determine the size of an ion in the gas-phase.[36] This measurement has been used to record conformational differences of proteins due to factors such as solution conditions,[37] collisional activation,[38, 39] and protein charging.[40] Binding of ligands has also been reported to shift the drift time of proteins.[41, 42] It has been found that small ligands and metal ions shift the conformational dynamics of amyloid proteins such as mutant tau[21] and α-syn.[13, 15, 29, 43, 44] IM-

MS has been found to reveal relevant structural information on proteins in the gas-phase and detect structural perturbations due to modifications.

In this study, the interaction between CLR01 and various proteoforms of α-syn is characterized. Native MS indicated that CLR01 binds a variety of α-synuclein proteoforms. Electron capture dissociation (ECD) can reveal the location of all modifications including the location of CLR01 binding to α-syn. In addition, IM-MS of α-synuclein/CLR01 complexes indicates that CLR01 compacts the structure of all α-syn proteoforms which may prevent aggregation of α-synuclein monomers. This research could shed light on how small molecule ligands interact with α-synuclein and how they inhibit aggregation of α-syn and other amyloid proteins.

MATERIALS AND METHODS

For native MS experiments, 10uM of protein was dissolved into 20mM ammonium acetate. CLR01 was added at various amounts to obtain concentrations of 10uM to 50uM. For analysis of acetylated and phosphorylated α-syn, WT α-syn, phosphorylated α-syn, and acetylated α-syn were added at 10uM for a total concentration of 30uM and 30uM of CLR01 was added to the solution. For metal ion experiments copper acetate was added at 10uM and manganese acetate was added at 50uM. CLR01 was added at 30uM for metal binding experiments.

For electron capture dissociation experiments 10 to 50uM CLR01 was added to 10uM solutions of α-synuclein in 20mM ammonium acetate. For solutions where metal was added 10uM copper acetate was added and 50uM manganese acetate was added to the solution with 30uM CLR01. Solutions were sprayed on a 15T Solarix FT-ICR instrument from Bruker. Peaks were isolated and fragmented with ECD with a bias of 1V and a pulse length of 10ms to 50ms. The spectra were deconvoluted with the DataAnalysis 4.0 or 5.0 and *c* and z fragments were matched to the deconvoluted list were assigned with ClipsMS[45] with an error of 5ppm. CLR01, manganese, and copper were added as unlocalized modifications and the A30P mutation, phosphorylation, and N-terminal acetylation were added as localized modifications.

For IM-MS experiments, IM was taken of the entire spectrum and data for each charge state was extracted over the $[M+nH]^{n+}$ peak. IM-MS of CLR01 binding to WT and A30P α-syn was taken with CLR01 inserted into the solution at a 5x molar concentration. IM-MS of acetylated and phosphorylated α-synuclein bound with CLR01 were analyzed at a 1:1 molar ratio. To analyze structural changes of CLR01 binding to copper bound α-syn, WT α-syn was added at a concentration of 10uM, copper acetate was added at a concentration of 10uM, and CLR01 was added at a concentration of 30uM. To analyze structural changes of CLR01 binding to manganese bound α-syn, WT α-syn was added at a concentration of 10uM, manganese acetate was added at a concentration of 50uM, and CLR01 was added at a concentration of 30uM.

RESLUTS AND DISCUSSION

Characterization of α-Synuclein CLR01 Binding

Previous research has found that CLR01 binds to wild type (WT) α-syn (α-syn) when combined at a 1:1 molar ratio.[20] However, it is known that greater inhibition of fibril formation can be achieved at higher concentrations of CLR01.[17] To investigate the interaction between CLR01 and α-syn at higher concentrations of CLR01, CLR01 was inserted into a native WT AS was inserted into a solution with a 5x molar concentration of CLR01. The resulting mass spectrum revealed a broad charge state distribution characteristic of intrinsically disordered proteins and multiple CLR01 binding events. (Fig. 1A) Deconvolution of the mass spectrum indicated that up to 5 binding events can be observed and 63% of all the CLR01 in the solution is bound to the protein. A solution of A30P α-syn, a proteoform that has been correlated with increased aggregation[9], and CLR01 were also sprayed in native conditions revealing that multiple CLR01 molecules can bind to A30P α-syn. (Fig. S1) Deconvolution of the spectrum indicated that 5 binding events can be observed on α-syn at this concentration and that 74% of CLR01 is bound. It is possible that the large number of CLR01 molecules bound to α-synuclein monomers at this concentration may prevent monomers from interacting with one another and aggregating into oligomers and fibrils.

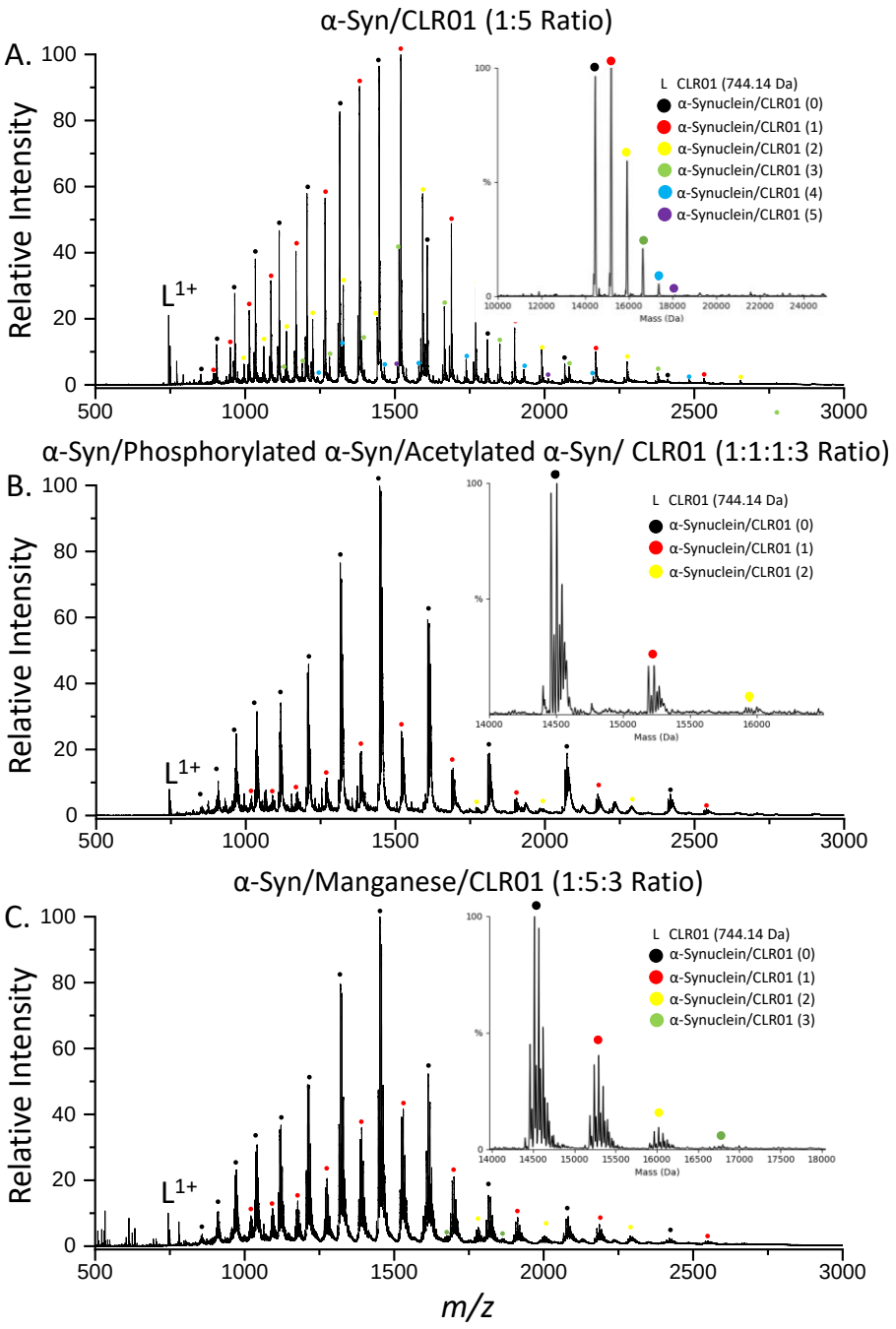


Figure 1: Mass spectra of solutions of A.) α-syn and CLR01 at a 1:5 molar ratio B.) α-syn, phosphorylated α-syn, acetylated α-syn, and CLR01 at a 1:1:1:3 molar ratio, and C.) α-syn, Mn^{2+}, and CLR01 at a 1:5:3 molar ratio with the corresponding deconvoluted mass spectra. The results reveal multiple CLR01 binding events and the increased binding of Mn^{2+} CLR01 molecules bind.

Mass spectra were also collected at 1:0, 1:1, 1:2, 1:3, 1:4 molar ratios of WT α-synuclein and CLR01. These spectra were utilized to calculate K_d values for each binding event. (Fig. S2) It was found

that the first binding event had a K_d of 46μM. The K_d values for the second and third binding sites were 70.6μM and 89.2μM respectively, indicating enhanced binding at these sites. The K_d values for the fourth and 5th binding sites was 250.6 μM and 1M indicating these binding sites were nonspecific. It is interesting that CLR01 binding is enhanced at the 2nd and 3rd binding site, and it may be the case that cooperative binding occurs when more than one CLR01 molecule is bound.

It is known that some post-translational modifications (PTMs) can modulate the rate of amyloid protein aggregation including N-terminal acetylation and serine 129 phosphorylation. To investigate how these PTMs alter CLR01 binding, wild type α-synuclein, acetylated α-synuclein, and phosphorylated α-synuclein were added to the same native solution and sprayed with an equimolar ratio of CLR01. (Fig. 1B) The spectrum revealed baseline resolved peaks corresponding the acetylated and phosphorylated proteoforms of α-synuclein in addition to CLR01 bound proteoforms. Deconvolution of the spectrum indicates up to 2 CLR01 molecules can bind each proteoform. Comparison of the intensities for unbound and CLR01 bound proteoforms revealed acetylated and phosphorylated α-synuclein had a slightly lower affinity (96%-98%) for the CLR01 molecule than the WT proteoform possibly indicating that the presence of PTMs alter the binding characteristics of CLR01.

Mn^{2+} binding to α-syn has been shown to increase aggregation.[14] To determine the relationship between α-syn, Mn^{2+}, and CLR01, nMS was performed on a solution with a 1:5:3 molar ratio of α-syn to Mn^{2+} to CLR01. The resulting spectrum indicated multiple binding events of Mn^{2+} and CLR01. (Fig. 1C) Deconvolution of the spectrum indicates up to 5 Mn^{2+} binding events can be observed and 4 CLR01 molecules at this concentration. In addition, the spectrum indicates that binding of CLR01 enhances Mn^{2+} binding. When CLR01 is not bound to the protein, 86% of the α-synuclein is bound by Mn^{2+}. However, when 1 CLR01 molecule is bound, 89% of the α-synuclein is bound by CLR01, and when 2 CLR01 molecules are bound, only 92% of the α-synuclein is bound by Mn^{2+}. This data indicates that when CLR01 binds to α-syn, Mn^{2+} binding increases.

Cu^{2+} binding to α-syn has shown to increase the rate of α-syn aggregation.[13] To investigate the interaction between α-synuclein, Cu^{2+}, and CLR01, nMS was performed on a solution of α-syn/Cu/CLR01

at a 1:1:3 molar ratio. (Fig. S3) The spectrum reveals that most α-synuclein is bound with Cu^{2+} and CLR01 binding events can be observed. Deconvolution of the spectrum indicates up to 3 CLR01 molecules can bind α-syn. In addition, the spectrum reveals that multiple CLR01 binding events compete with Cu^{2+} for binding. In the deconvoluted spectrum, 65% of the α-synuclein is bound to Cu^{2+} bound when CLR01 is not bound. When 1 and 2 CLR01 molecules bind α-synuclein the amount of the Cu^{2+} proteoform does not change much at 64% and 68% respectively. However, when 3 CLR01 molecules are bound by α-synuclein 57% is bound by Cu^{2+}. This data suggests that when multiple CLR01 molecules bind to α-synuclein, they compete with Cu^{2+} for binding.

Top-Down Analysis of α-Synuclein/CLR01 Complexes

Our lab has previously fragmented the WT α-syn/CLR01 complex and it was determined that CLR01 binds predominantly to lysine 10 and lysine 12.[20] However, it is not known where other CLR01 molecules bind when multiple CLR01 molecules are bound to α-synuclein. To determine where the second CLR01 binding event occurs, the 13+ charge state of wildtype α-syn bound with 2 CLR01 molecules (m/z = 1223.85) was isolated and fragmented with ECD. (Fig. 2A) The spectrum revealed fragments that were not bound with CLR01, fragments with 1 CLR01 molecule bound, and fragments with 2 CLR01 molecules bound. Analysis of the unbound and bound fragments indicated that the secondary CLR01 binding location is at lysine 20 and 23. This data seems to suggest that CLR01 molecules coat the N-terminus of α-synuclein. ECD was also performed on A30P α-synuclein/CLR01 complex. ECD fragmentation of the A30P α-syn/CLR01 complex indicated the location of the A30P mutation and the location of the CLR01 binding event at on the N-terminus of this proteoform. (Fig. 2B) This data suggests that the binding site of CLR01 did not shift drastically from WT.

To further study the binding site of location of CLR01 on α-syn, ECD was performed on acetylated and phosphorylated α-syn bound with CLR01. Fragmentation of the acetylated α-syn/CLR01 complex showed fragments that contained the acetylation site and the CLR01 molecule. (Fig. S4A) CLR01 was found to bind the N-terminus of this proteoform; however, the interaction may be weaker due to the lesser number of fragments containing the CLR01 molecule. It is possible that the stabilization of the α-helix on

the N-terminus shifts the binding site of CLR01 to this location. Fragmentation of the phosphorylated α-syn/CLR01 complex indicated the location of the phosphorylation site at serine 129 and the location of CLR01 binding was similar to WT at the N-terminus of the protein. (Fig. S4B) This indicates that phosphorylation of α-syn on the C-terminus does not significantly shift the binding location of CLR01. This data suggests that the CLR01 binding does not shift significantly due to the presence of PTMs on α-syn.

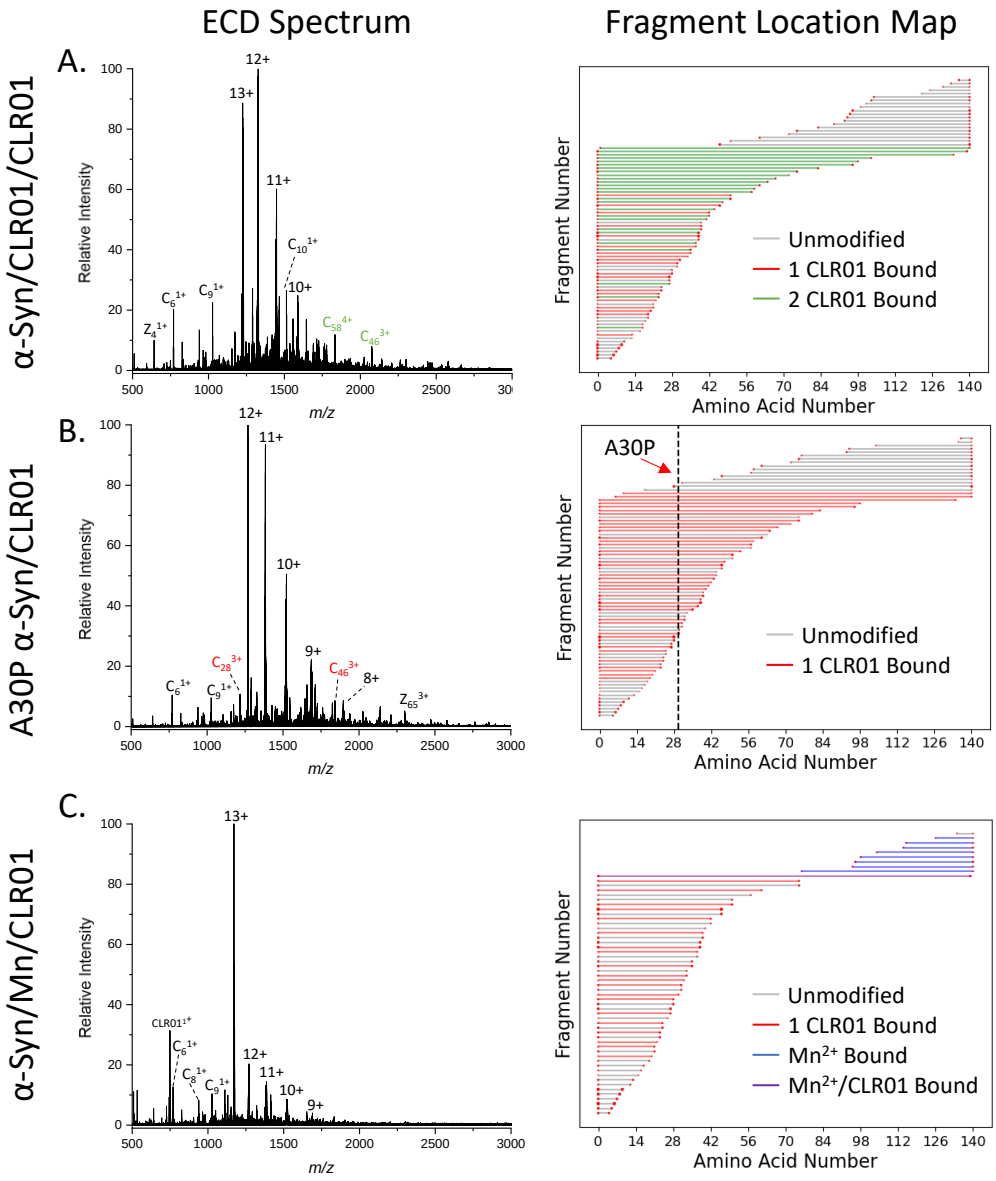


Figure 2: ECD fragmentation spectra for 13+ charge state of A.) the α-syn/CLR01/CLR01 complex, B.) the A30P α-syn/CLR01 complex, and E.) the α-syn/Mn^{2+}/CLR01 complexes with the corresponding fragment location maps. The maps reveal the location of the modifications and of the CLR01 binding location on the N-terminus of the protein.

Previous studies have utilized TD-MS to locate metal ions such as copper [13] and manganese [29] on α-syn. To characterize the interaction between α-syn, metal ions, and CLR01 molecules, ECD fragmentation was performed on α-syn/metal/CLR01 complexes. ECD fragmentation of the α-syn/ Mn^{2+}/CLR01 complex indicated that CLR01 binds to N-terminal region as expected and Mn^{2+} binds to the C-terminal region of α-Syn. (Fig. 2C) It is possible that CLR01 binding to the N-terminus could expose the Mn^{2+} binding region on the C-terminus which would explain the increased binding of Mn^{2+} when CLR01 molecules bind. ECD of the α-syn/ Cu^{2+}/CLR01 complex revealed peaks bound with Cu^{2+}, CLR01, and both Cu^{2+} and CLR01 which indicate that both CLR01 and Cu^{2+} bind primarily to the N-terminus of the protein. (Fig. S4C) The proximity of Cu^{2+} and CLR01 binding on α-syn may explain the decreased binding of Cu^{2+} when multiple CLR01 molecules bind the protein. Multiple CLR01 binding events on the N-terminus could shield the N-terminal region and prevent Cu^{2+} ions from binding. TD-MS with ECD of protein/ligand/metal complexes can accurately identify the location of the ligands and may explain why binding of metal ions is modulated by CLR01 binding.

Ion Mobility Analysis of α-Synuclein/CLR01 Complex

Ion-mobility mass spectrometry (IM-MS) has revealed that small molecules can shift the structure of amyloid proteins. Previous research has found that CLR01 compacts the structure of mutant tau upon binding.[21] In addition, it has also been found that EGCG compacts the structure of α-synuclein upon binding.[15] To determine if the structure of α-syn shifts upon CLR01 binding, IM-MS was performed on solutions of α-syn and CLR01. IM-MS of the 9+ charge state of the WT proteoform reveals the existence of 4 ensembles of ions. (Fig. 3A) IM-MS of the 9+ charge state with 1 CLR01 molecule bound reveals an increase in abundance of the early arriving ensembles (Fig. 3B). When additional CLR01 molecules bind, a greater increase in the abundance of early arriving ensembles is observed. (Fig. S5A) The 8+ charge state also shows an increase of early arriving ensembles much like that of the 9+ charge state. (Fig. S5B) The higher charge states such as 12+ do not show significant shifts in the mobiligram when CLR01 binds. (Fig. S5C) IM-MS of the A30P α-syn/CLR01 complexes reveals a similar pattern. The 9+ charge state of A30P α-syn reveals 4 ensembles, and when CLR01 binds, the early arriving ensembles increase in abundance.

(Fig. S6A) The 8+ charge state shows a similar trend. (Fig. S6B) The mobiligram of the 12+ charge state of A30P α-syn does not shift drastically indicating CLR01 does not affect the structure of the higher charge states of A30P. (Fig. S6C) The increase in abundance of early arriving ensembles for the 8+ and 9+ charge states due to CLR01 binding indicates that CLR01 compacts the structure of α-syn. It is possible this compaction is due to the presence of salt bridges formed by the negatively charged phosphates on CLR01 and positively charged residues of α-syn. The higher charge states do not shift considerably, possibly due to an accumulation of charge that extends the structure of the protein owing to charge repulsion between adjacent charge sites.[46, 47]

To further investigate the structural effect of CLR01 binding, IM-MS was performed on CLR01 bound forms of phosphorylated and acetylated α-syn. The 9+ IM profile of phosphorylated α-syn also revealed 4 ensembles of the proteins. (Fig. 3C) IM-MS of phosphorylated α-syn bound to CLR01 showed further increase in abundance of the early arriving ensembles. (Fig. 3D) The 8+ charge state (Fig. S7A) shows similar results to the 9+ charge state; however, the higher charge states such as 12+ did not shift significantly when bound with CLR01. (Fig. S7B) The 9+ IM profile of acetylated α-syn alsoreveals 4 ensembles of ions and the CLR01 bound state shows an increase in the abundance of early arriving ensembles indicating CLR01 compacts the structure of acetylated α-syn. (Fig. S8A) This observation is also observed for the 8+ charge state (Fig. S8B), although, the 12+ charge state does not shift significantly. (Fig. S8C) This data suggests that CLR01 compacts the structure of phosphorylated and acetylated proteoforms much like the WT proteoform of α-syn.

Previous research from our lab has found that Cu^{2+} binding increases early arriving ensembles of the protein.[13] To further study this proteoform of α-syn, IM-MS was performed on the unbound and CLR01 bound version of this proteoform. The 9+ charge state of Cu^{2+} bound α-syn revealed 4 ensembles of ions with early arriving ensembles significantly more abundant than WT. (Fig. 3E) When CLR01 binds the Cu^{2+} bound proteoform, the early arriving conformers further increase in abundance. (Fig. 3F) the 8+ charge state indicates similar trends. (Fig. S9A) The 12+ charge state as expected did not shift significantly when bound to either Cu^{2+} or CLR01. (Fig. S9B) It is possible that compaction of α-syn structure due to CLR01

binding could shield the region of Cu^{2+} binding. Mn^{2+} binding only slightly compacts α-syn structure; however, when CLR01 and Mn^{2+} both bind, the structure of α-syn is compounded. (Fig. S10A) This trend is also seen with the 8+ charge state. (Fig. S10B) The 12+ charge state of α-syn does not shift significantly as seen in the other spectra. (Fig. S10C) Compaction of protein structure may explain why Mn^{2+} binding is enhanced upon CLR01 binding.

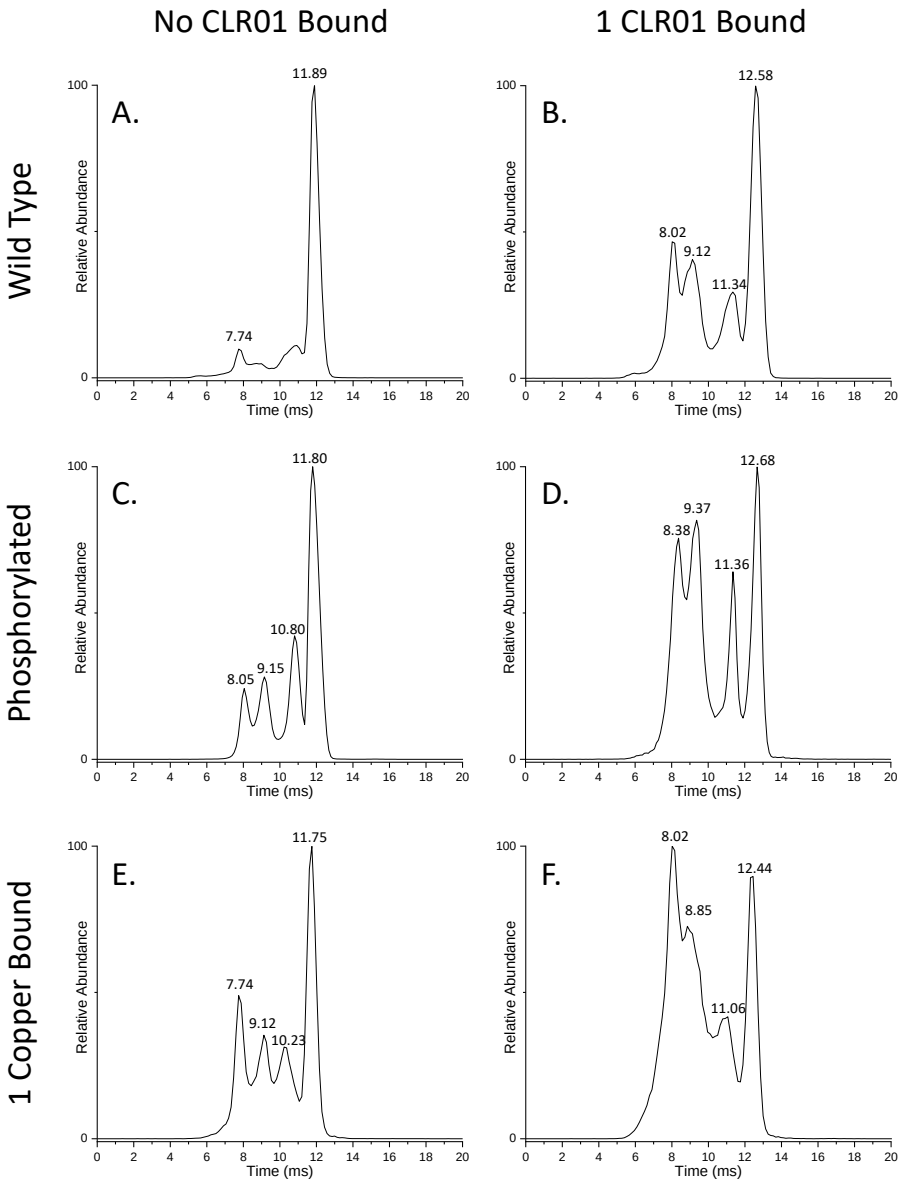


Figure 3: The mobiligrams for the 9+ charge state of A.) unbound and B.) CLR01 bound WT α-syn, C.) unbound and D.) CLR01 bound phosphorylated α-syn, and E.) unbound and F.) CLR01 bound Cu^{2+} bound α-syn. Notice for these proteoforms how CLR01 binding results in the increased abundance of early arriving ensembles of α-syn.

It is possible that compaction of α-syn due to CLR01 binding could be a possible mechanism CLR01 utilizes for preventing monomers from forming oligomer and fibrils. The disordered N-terminus of α-syn could expose its aggregation region and provide the opportunity for monomers to interact. From our data we found that multiple CLR01 molecules bind to the N-terminus of α-syn proteoforms and compacts its structure. This compaction of the N-terminus could shield the aggregation region of α-syn from other monomers resulting in a decreased probability for proteins to aggregate and form oligomers and fibrils. This could be the mechanism that CLR01 utilizes to prevent the aggregation of amyloid proteins.

CONCLUSION

In this study we analyzed how CLR01 interacts with various proteoforms of α-syn. It was found that CLR01 molecules can bind various proteoforms including mutated α-synuclein, PTM modified α-synuclein, and metal ion bound α-synuclein and that CLR01 binding modulates the binding of other ligands including other CLR01 molecules Mn^{2+} ions, and Cu^{2+} ions. Using ECD along with FT-ICR MS, it was possible to localize the modifications for all proteoforms and the CLR01 molecule along the protein sequence. It was found that for all proteoforms, CLR01 binds the N-terminal region of α-synuclein. IM-MS of various α-synuclein/CLR01 complexes indicates that CLR01 compacts the structure of the protein. It is possible that salt bridges formed by the negatively charged phosphates on CLR01 with positively charged residues of α-synuclein result in a more compact structure of the protein. This data could indicate the mechanism CLR01 utilizes to inhibit fibril formation. The compaction of the N-terminal region promoted by CLR01 could shield the region prone to aggregation from other α-synuclein monomers. This would prevent monomers from aggregating and forming oligomers and fibrils. This mechanism could potentially indicate how small molecules prevent fibril formation and provide information for the development of these small molecule inhibitors to halt the onset of neurodegenerative diseases.

REFERENCES

1. Johnson, V. E.; Stewart, W.; Smith, D. H., Widespread tau and amyloid-beta pathology many years after a single traumatic brain injury in humans. *Brain pathology* **2012,** *22* (2), 142-149.

2. Díaz-Hernández, M.; Moreno-Herrero, F.; Gómez-Ramos, P.; Morán, M. A.; Ferrer, I.; Baró, A. M.; Avila, J.; Hernández, F.; Lucas, J. J., Biochemical, ultrastructural, and reversibility studies on huntingtin filaments isolated from mouse and human brain. *Journal of Neuroscience* **2004,** *24* (42), 9361-9371.

3. Yamada, M.; Hayashi, S.; Tsuji, S.; Takahashi, H., Involvement of the cerebral cortex and autonomic ganglia in Machado-Joseph disease. *Acta neuropathologica* **2001,** *101* (2), 140-144.

4. Lautenschläger, J.; Stephens, A. D.; Fusco, G.; Ströhl, F.; Curry, N.; Zacharopoulou, M.; Michel, C. H.; Laine, R.; Nespovitaya, N.; Fantham, M., C-terminal calcium binding of α-synuclein modulates synaptic vesicle interaction. *Nature communications* **2018,** *9* (1), 1-13.

5. Bellani, S.; Sousa, V. L.; Ronzitti, G.; Valtorta, F.; Meldolesi, J.; Chieregatti, E., The regulation of synaptic function by α-synuclein. *Communicative & Integrative Biology* **2010,** *3* (2), 106-109.

6. Schweighauser, M.; Shi, Y.; Tarutani, A.; Kametani, F.; Murzin, A. G.; Ghetti, B.; Matsubara, T.; Tomita, T.; Ando, T.; Hasegawa, K., Structures of α-synuclein filaments from multiple system atrophy. *Nature* **2020**, 1-6.

7. Li, B.; Ge, P.; Murray, K. A.; Sheth, P.; Zhang, M.; Nair, G.; Sawaya, M. R.; Shin, W. S.; Boyer, D. R.; Ye, S., Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nature communications* **2018,** *9* (1), 1-10.

8. Baba, M.; Nakajo, S.; Tu, P.-H.; Tomita, T.; Nakaya, K.; Lee, V.; Trojanowski, J. Q.; Iwatsubo, T., Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies. *The American journal of pathology* **1998,** *152* (4), 879.

9. Stephens, A. D.; Zacharopoulou, M.; Moons, R.; Fusco, G.; Seetaloo, N.; Chiki, A.; Woodhams, P. J.; Mela, I.; Lashuel, H. A.; Phillips, J. J., Extent of N-terminus exposure of monomeric alpha-synuclein determines its aggregation propensity. *Nature Communications* **2020,** *11* (1), 1-15.

10. Kang, L.; Moriarty, G. M.; Woods, L. A.; Ashcroft, A. E.; Radford, S. E.; Baum, J., N-terminal acetylation of α-synuclein induces increased transient helical propensity and decreased aggregation rates in the intrinsically disordered monomer. *Protein Science* **2012,** *21* (7), 911-917.

11. Bu, B.; Tong, X.; Li, D.; Hu, Y.; He, W.; Zhao, C.; Hu, R.; Li, X.; Shao, Y.; Liu, C., N-terminal acetylation preserves α-synuclein from oligomerization by blocking intermolecular hydrogen bonds. *ACS Chemical Neuroscience* **2017,** *8* (10), 2145-2151.

12. Sato, H.; Arawaka, S.; Hara, S.; Fukushima, S.; Koga, K.; Koyama, S.; Kato, T., Authentically phosphorylated α-synuclein at Ser129 accelerates neurodegeneration in a rat model of familial Parkinson's disease. *Journal of Neuroscience* **2011,** *31* (46), 16884-16894.

13. Choi, T. S.; Lee, J.; Han, J. Y.; Jung, B. C.; Wongkongkathep, P.; Loo, J. A.; Lee, M. J.; Kim, H. I., Supramolecular modulation of structural polymorphism in pathogenic α-synuclein fibrils using copper (II) coordination. *Angewandte Chemie International Edition* **2018,** *57* (12), 3099-3103.

14. Uversky, V. N.; Li, J.; Fink, A. L., Metal-triggered structural transformations, aggregation, and fibrillation of human α-synuclein a possible molecular link between parkinson′ s disease and heavy metal exposure. *Journal of Biological Chemistry* **2001,** *276* (47), 44284-44296.

15. Konijnenberg, A.; Ranica, S.; Narkiewicz, J.; Legname, G.; Grandori, R.; Sobott, F.; Natalello, A., Opposite structural effects of epigallocatechin-3-gallate and dopamine binding to α-synuclein. *Analytical chemistry* **2016,** *88* (17), 8468-8475.

16. Hong, D.-P.; Fink, A. L.; Uversky, V. N., Structural characteristics of α-synuclein oligomers stabilized by the flavonoid baicalein. *Journal of molecular biology* **2008,** *383* (1), 214-223.

17. Prabhudesai, S.; Sinha, S.; Attar, A.; Kotagiri, A.; Fitzmaurice, A. G.; Lakshmanan, R.; Ivanova, M. I.; Loo, J. A.; Klärner, F.-G.; Schrader, T., A novel "molecular tweezer" inhibitor of α-synuclein neurotoxicity in vitro and in vivo. *Neurotherapeutics* **2012,** *9* (2), 464-476.

18. Bieschke, J.; Russ, J.; Friedrich, R. P.; Ehrnhoefer, D. E.; Wobst, H.; Neugebauer, K.; Wanker, E. E., EGCG remodels mature α-synuclein and amyloid-β fibrils and reduces cellular toxicity. *Proceedings of the National Academy of Sciences* **2010,** *107* (17), 7710-7715.

19. Xu, M.; Loa-Kum-Cheung, W.; Zhang, H.; Quinn, R. J.; Mellick, G. D., Identification of a new α-synuclein aggregation inhibitor via mass spectrometry based screening. *ACS chemical neuroscience* **2019,** *10* (6), 2683-2691.

20. Acharya, S.; Safaie, B. M.; Wongkongkathep, P.; Ivanova, M. I.; Attar, A.; Klärner, F.-G.; Schrader, T.; Loo, J. A.; Bitan, G.; Lapidus, L. J., Molecular basis for preventing α-synuclein aggregation by a molecular tweezer. *Journal of Biological Chemistry* **2014,** *289* (15), 10727-10737.

21. 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.

22. Vöpel, T.; Bravo-Rodriguez, K.; Mittal, S.; Vachharajani, S.; Gnutt, D.; Sharma, A.; Steinhof, A.; Fatoba, O.; Ellrichmann, G.; Nshanian, M., Inhibition of huntingtin exon-1 aggregation by the molecular tweezer CLR01. *Journal of the American Chemical Society* **2017,** *139* (16), 5640-5643.

23. Malik, R.; Meng, H.; Wongkongkathep, P.; Corrales, C. I.; Sepanj, N.; Atlasi, R. S.; Klärner, F.-G.; Schrader, T.; Spencer, M. J.; Loo, J. A., The molecular tweezer CLR01 inhibits aberrant superoxide dismutase 1 (SOD1) self-assembly in vitro and in the G93A-SOD1 mouse model of ALS. *Journal of Biological Chemistry* **2019,** *294* (10), 3501-3513.

24. Sinha, S.; Lopes, D. H.; Du, Z.; Pang, E. S.; Shanmugam, A.; Lomakin, A.; Talbiersky, P.; Tennstaedt, A.; McDaniel, K.; Bakshi, R., Lysine-specific molecular tweezers are broad-spectrum inhibitors of assembly and toxicity of amyloid proteins. *Journal of the American Chemical Society* **2011,** *133* (42), 16958-16969.

25. Leong, S. L.; Pham, C. L.; Galatis, D.; Fodero-Tavoletti, M. T.; Perez, K.; Hill, A. F.; Masters, C. L.; Ali, F. E.; Barnham, K. J.; Cappai, R., Formation of dopamine-mediated α-synuclein-soluble oligomers requires methionine oxidation. *Free Radical Biology and Medicine* **2009,** *46* (10), 1328-1337.

26. Sarafian, T. A.; Ryan, C. M.; Souda, P.; Masliah, E.; Kar, U. K.; Vinters, H. V.; Mathern, G. W.; Faull, K. F.; Whitelegge, J. P.; Watson, J. B., Impairment of mitochondria in adult mouse brain overexpressing predominantly full-length, N-terminally acetylated human α-synuclein. *PLoS One* **2013,** *8* (5), e63557.

27. Kellie, J. F.; Higgs, R. E.; Ryder, J. W.; Major, A.; Beach, T. G.; Adler, C. H.; Merchant, K.; Knierman, M. D., Quantitative measurement of intact alpha-synuclein proteoforms from post-mortem control and Parkinson's disease brain tissue by intact protein mass spectrometry. *Scientific reports* **2014,** *4*, 5797.

28. Ly, T.; Julian, R. R., Protein-metal interactions of calmodulin and α-synuclein monitored by selective noncovalent adduct protein probing mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2011,** *19* (11), 1663-1672.

29. Wongkongkathep, P.; Han, J. Y.; Choi, T. S.; Yin, S.; Kim, H. I.; Loo, J. A., Native top-down mass spectrometry and ion mobility MS for characterizing the cobalt and manganese metal binding of α-synuclein protein. *Journal of the American Society for Mass Spectrometry* **2018,** *29* (9), 1870-1880.

30. Haselmann, K. F.; Jørgensen, T. J.; Budnik, B. A.; Jensen, F.; Zubarev, R. A., Electron capture dissociation of weakly bound polypeptide polycationic complexes. *Rapid communications in mass spectrometry* **2002,** *16* (24), 2260-2265.

31. Shi, S. D.-H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W., Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Analytical Chemistry* **2001,** *73* (1), 19-22.

32. 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* (12), 5014-5030.

33. 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.

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

35. 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.

36. Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature chemistry* **2014,** *6* (4), 281-294.

37. Shi, H.; Atlasevich, N.; Merenbloom, S. I.; Clemmer, D. E., Solution dependence of the collisional activation of ubiquitin [M+ 7H] 7+ ions. *Journal of the American Society for Mass Spectrometry* **2014,** *25* (12), 2000-2008.

38. Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T., Collision induced unfolding of isolated proteins in the gas phase: past, present, and future. *Current opinion in chemical biology* **2018,** *42*, 93-100.

39. Bernstein, S. L.; Liu, D.; Wyttenbach, T.; Bowers, M. T.; Lee, J. C.; Gray, H. B.; Winkler, J. R., α-Synuclein: Stable compact and extended monomeric structures and pH dependence of dimer formation. *Journal of the American Society for Mass Spectrometry* **2004,** *15* (10), 1435-1443.

40. Jhingree, J. R.; Beveridge, R.; Dickinson, E. R.; Williams, J. P.; Brown, J. M.; Bellina, B.; Barran, P. E., Electron transfer with no dissociation ion mobility–mass spectrometry (ETnoD IM-MS). The effect of charge reduction on protein conformation. *International Journal of Mass Spectrometry* **2017,** *413*, 43-51.

41. Williams, J. P.; Brown, J. M.; Campuzano, I.; Sadler, P. J., Identifying drug metallation sites on peptides using electron transfer dissociation (ETD), collision induced dissociation (CID) and ion mobility-mass spectrometry (IM-MS). *Chemical communications* **2010,** *46* (30), 5458-5460.

42. Campuzano, I. D.; Lippens, J. L., Ion mobility in the pharmaceutical industry: an established biophysical technique or still niche? *Current Opinion in Chemical Biology* **2018,** *42*, 147-159.

43. Liu, Y.; Carver, J. A.; Calabrese, A. N.; Pukala, T. L., Gallic acid interacts with α-synuclein to prevent the structural collapse necessary for its aggregation. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2014,** *1844* (9), 1481-1485.

44. Moons, R.; van der Wekken-de Bruijne, R.; Maudsley, S.; Lemière, F.; Lambeir, A.-M.; Sobott, F., Effects of Detergent on α-Synuclein Structure. A Native MS-Ion Mobility Study. *International Journal of Molecular Sciences* **2020,** *21* (21), 7884.

45. 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.

46. Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Analytical chemistry* **2010,** *82* (22), 9557-9565.

47. Picache, J. A.; Rose, B. S.; Balinski, A.; Leaptrot, K. L.; Sherrod, S. D.; May, J. C.; McLean, J. A., Collision cross section compendium to annotate and predict multi-omic compound identities. *Chemical science* **2019,** *10* (4), 983-993.

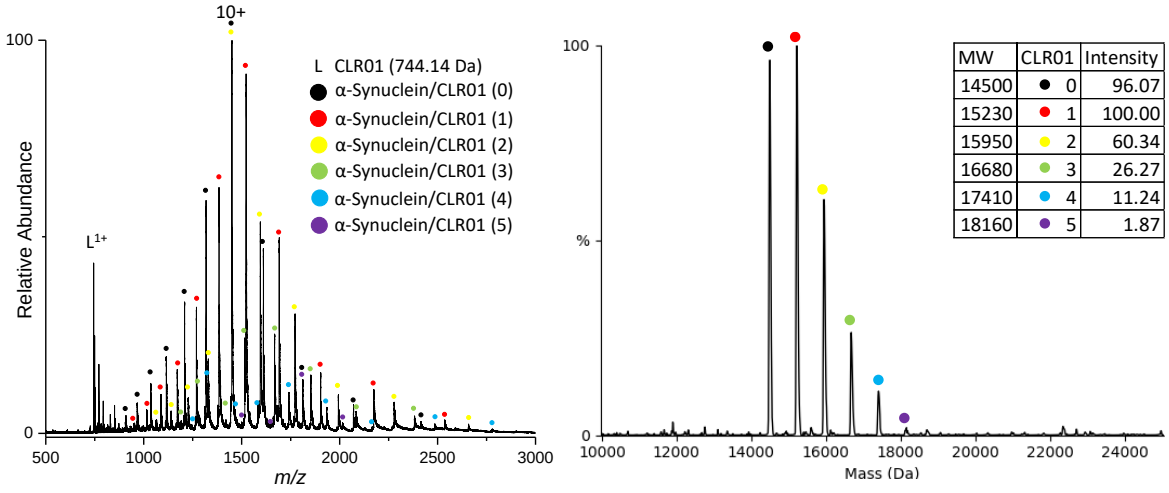


Figure S1: MS analysis of a solution with a 1:5 ratio of A30P α-syn to CLR01 and the UniDec deconvolution of the spectrum.

Kd's, uM				
K_i	K_{ii}	K_{iii}	K_{iv}	K_v
46.6	70.6	89.2	250.6	1000000

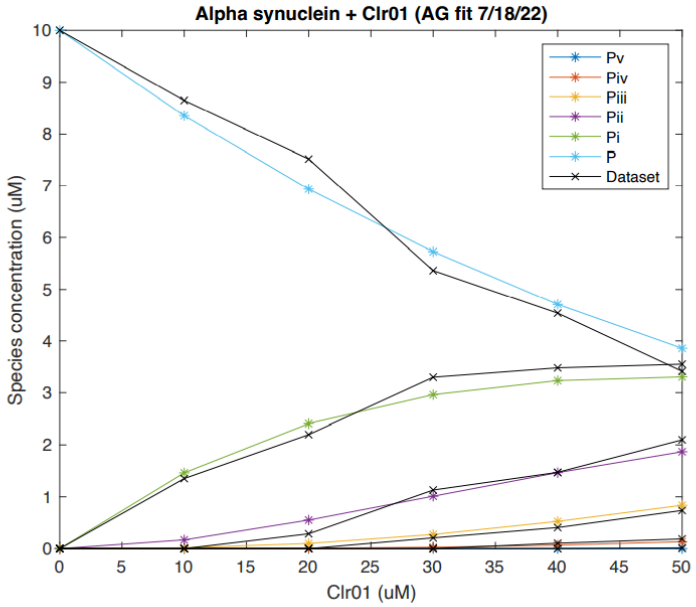


Figure S2: A graph revealing the K_d values for the 1 to 5 bound states of α-synuclein. The data indicates that CLR01 binds in the micromolar range.

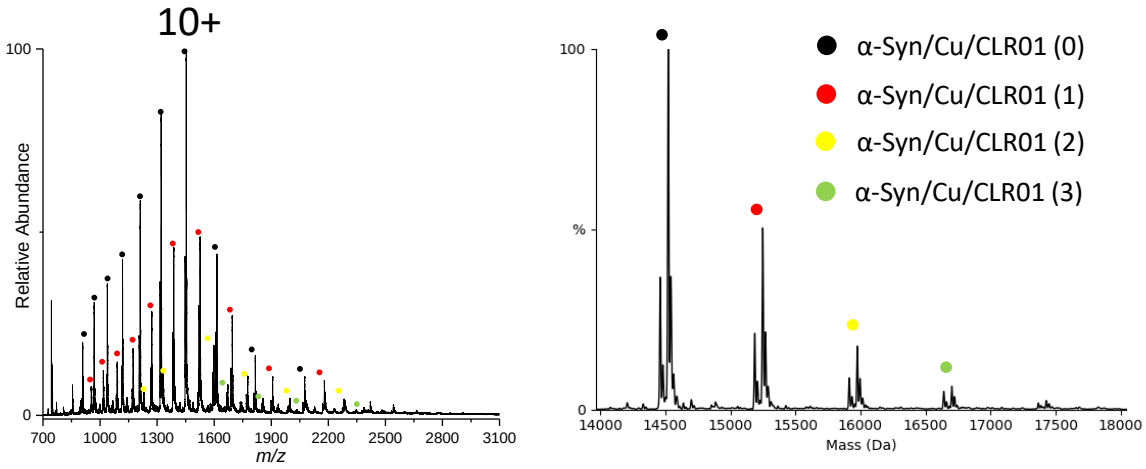


Figure S3: MS analysis of a solution with a 1:1:3 ratio of α-syn to Cu to CLR01 with the UniDec deconvolution of that spectrum.

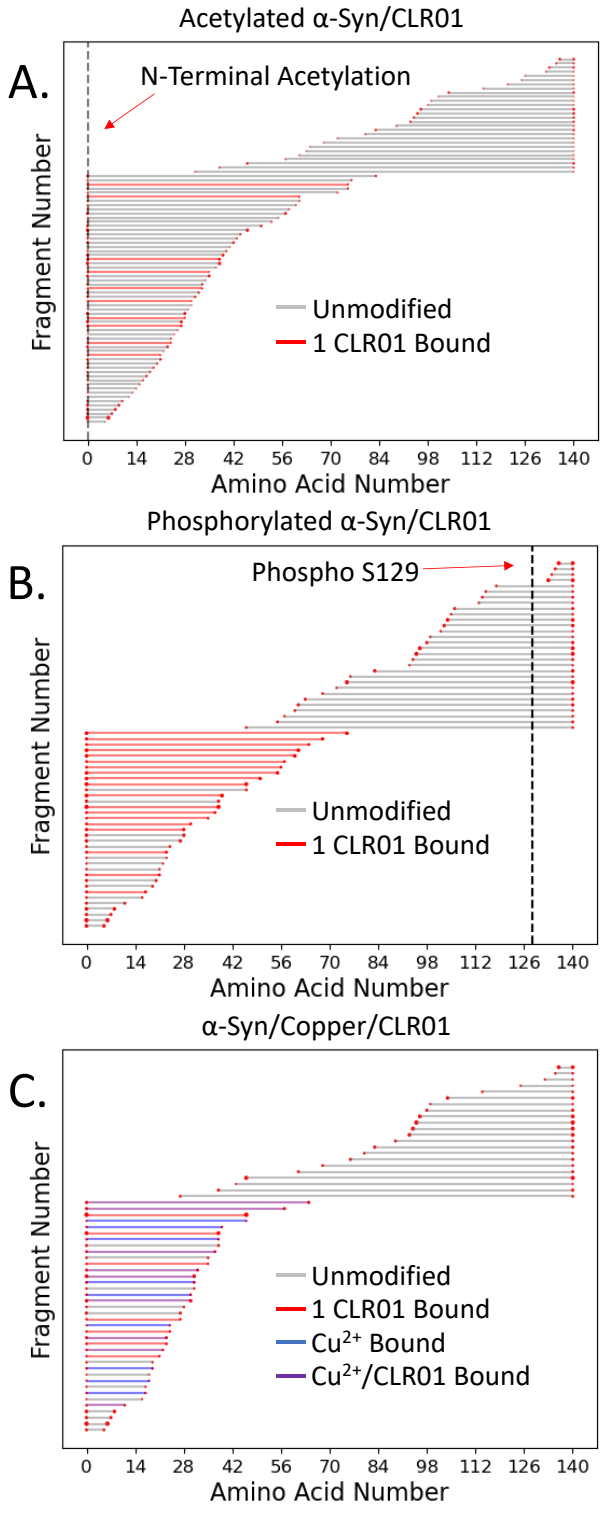


Figure S4: Fragment location maps for A.) the acetylated α-syn/CLR01 complex, B.) the phosphorylated α-syn/CLR01 complex, C.) and the α-syn/Cu/CLR01 complex. These maps indicate the location of the modification and that CLR01 binds to the N-terminus of the protein.

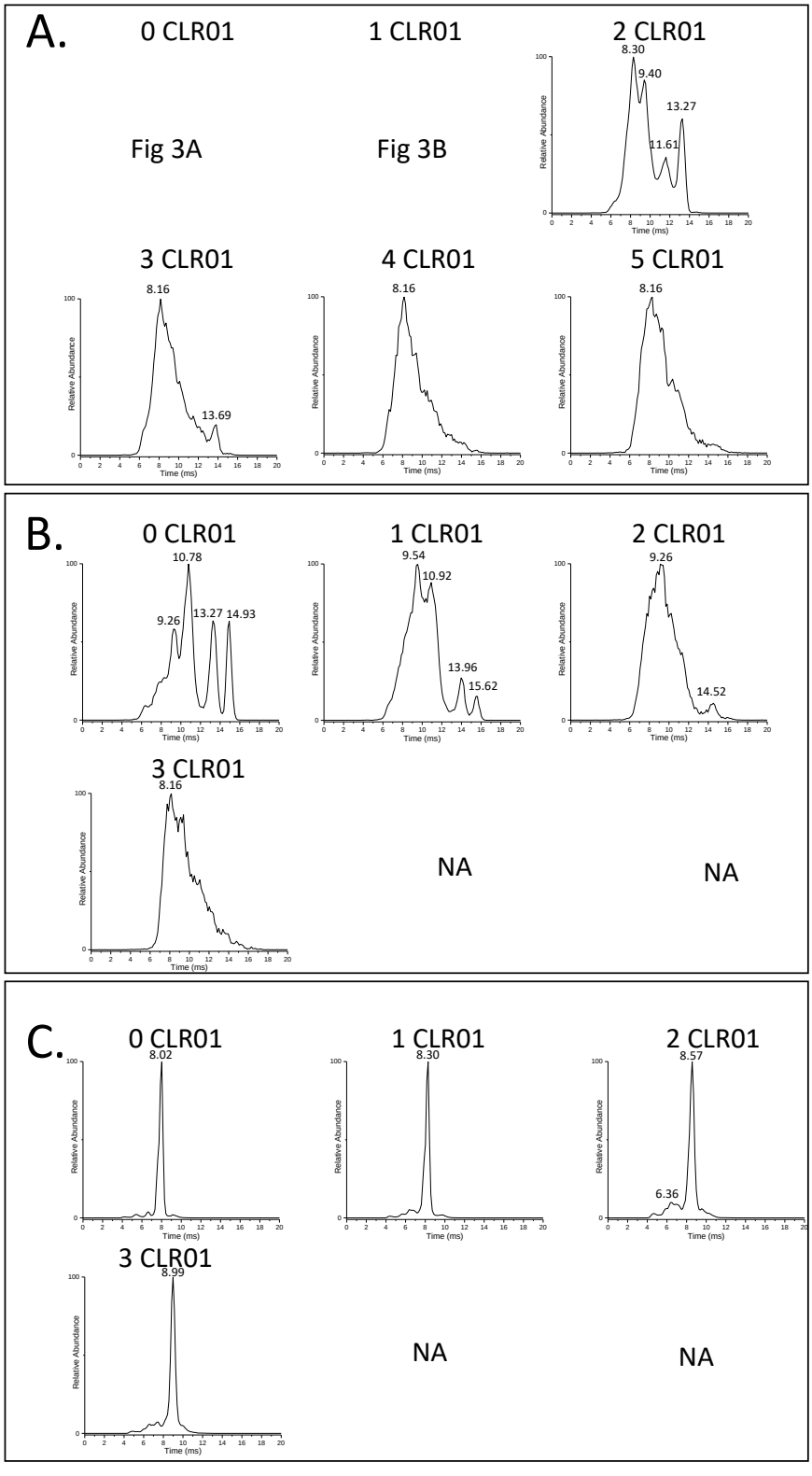


Figure S5: Ion mobility profiles for the CLR01 bound states of the A.) 9+, B.) 8+, and C.) 12+ charge states of α-syn.

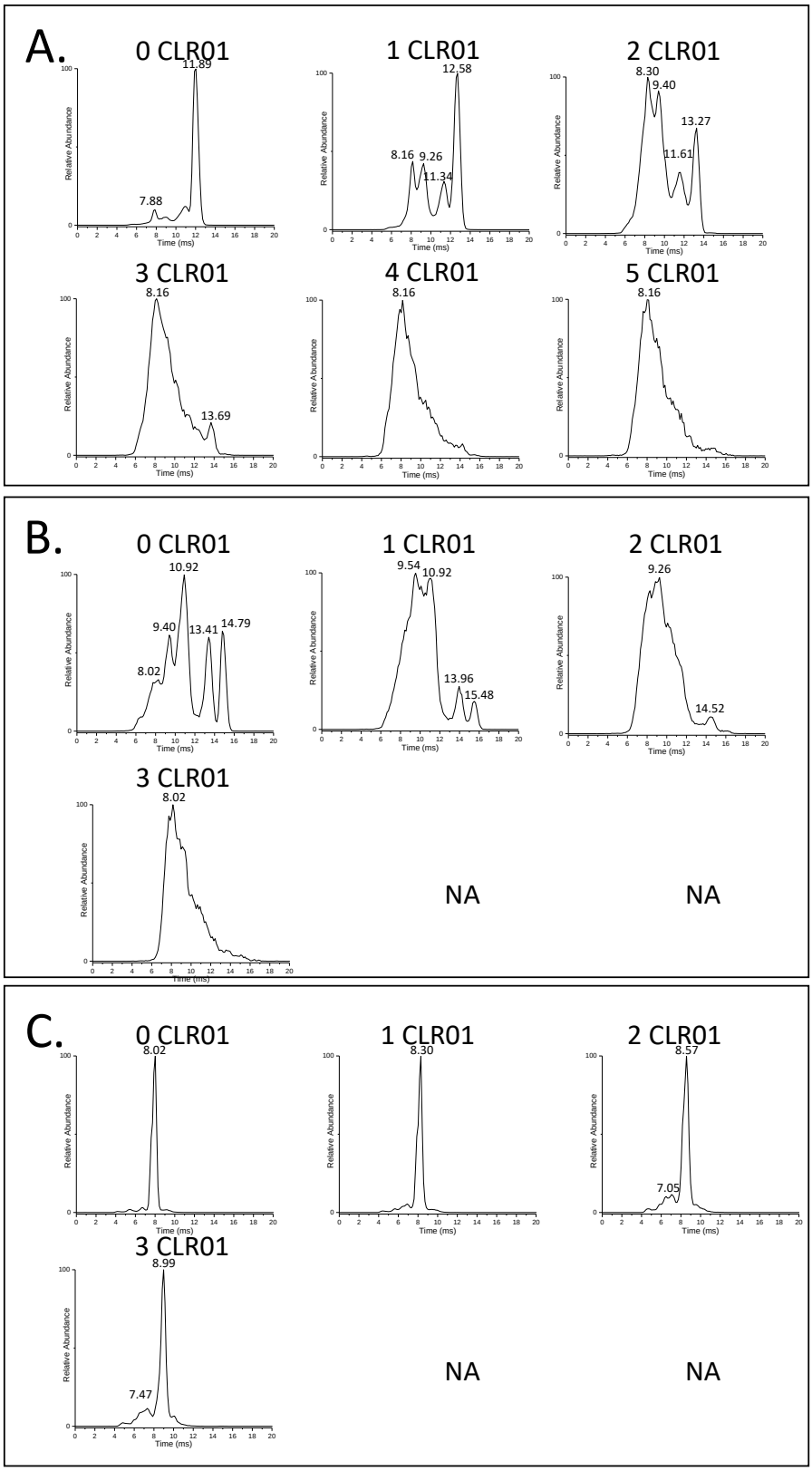


Figure S6: Ion mobility profiles for the CLR01 bound states of the A.) 9+, B.) 8+, and C.) 12+ charge states of A30P α-syn.

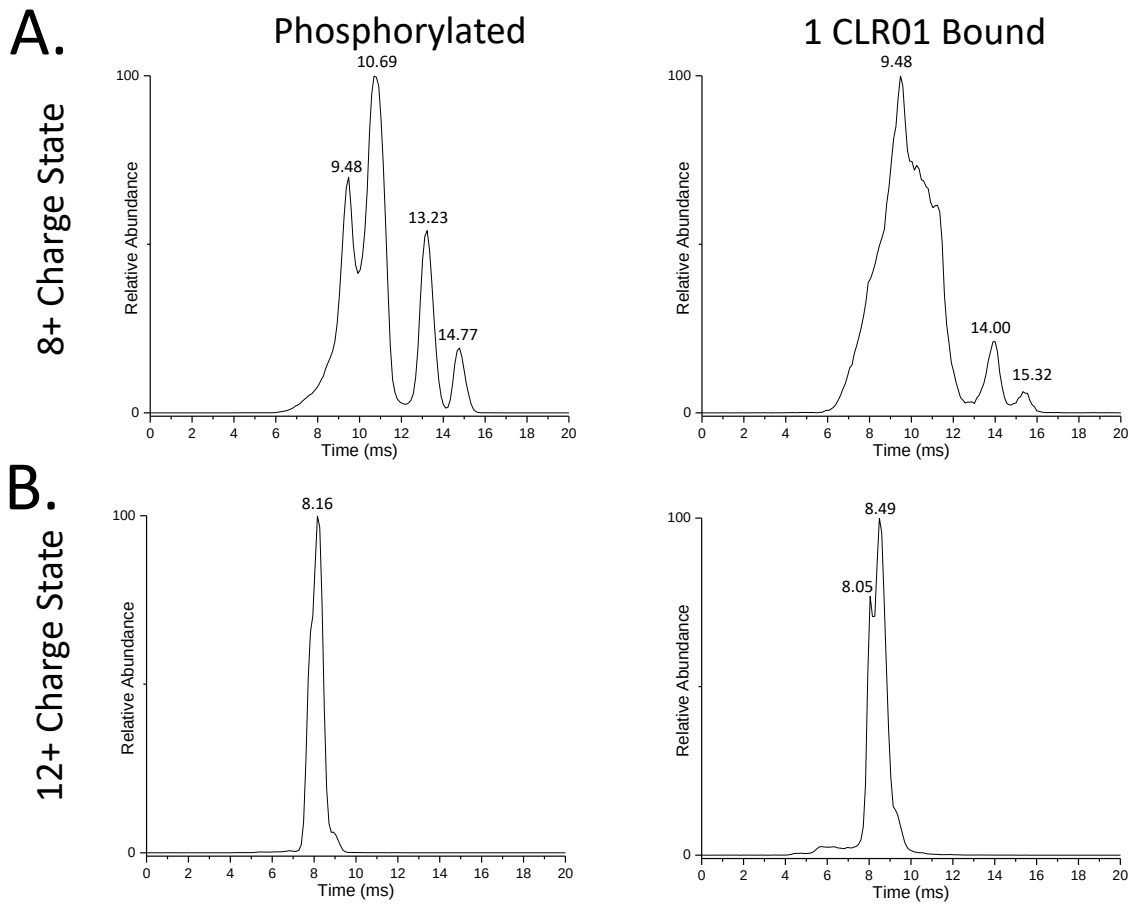


Figure S7: Ion mobility profiles for the unbound and CLR01 bound states of the 8+ and 12+ charge states of phosphorylated α-syn.

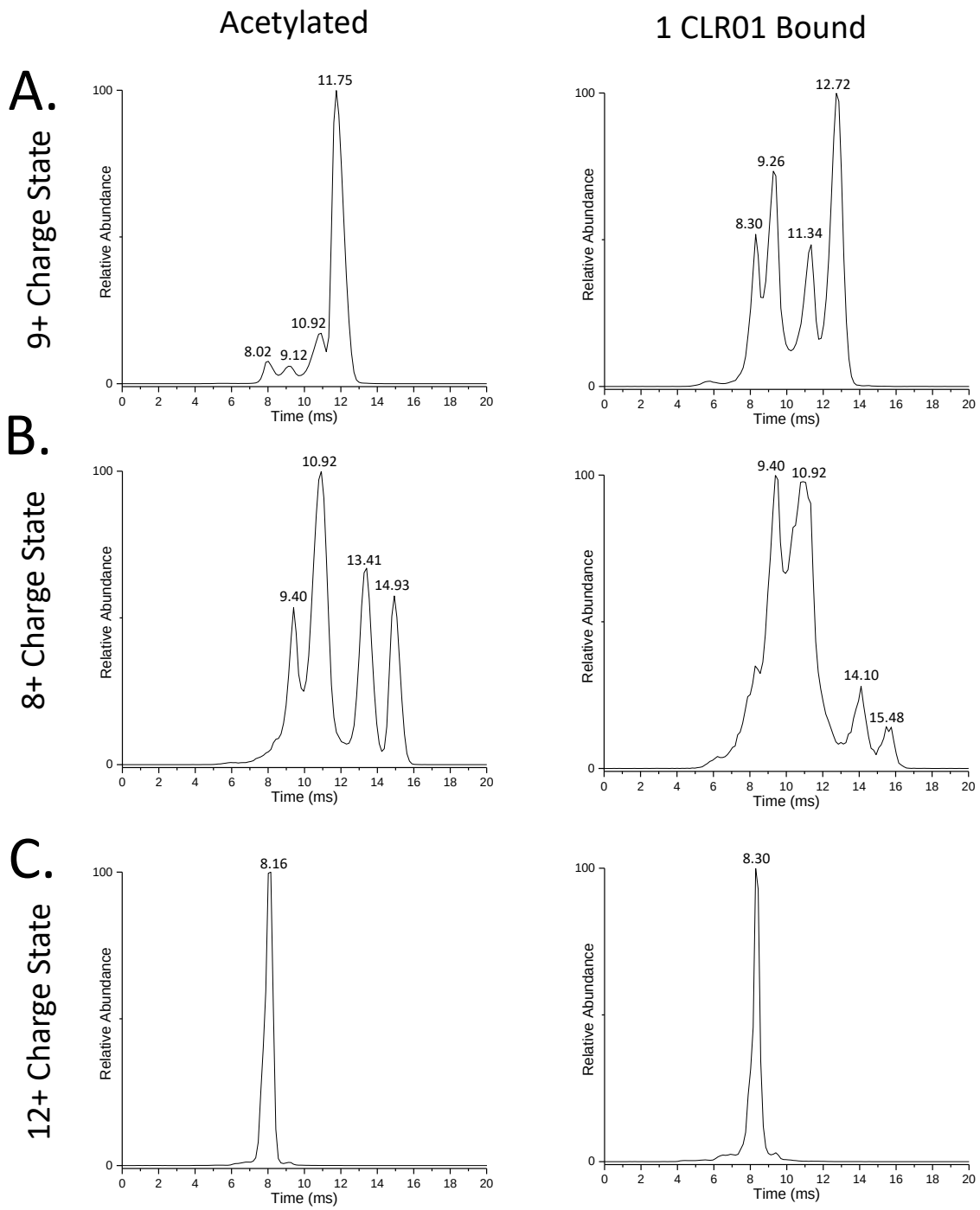


Figure S8: Ion mobility profiles for the unbound and CLR01 bound states of the A.) 9+, B.) 8+, and C.) 12+ charge states of acetylated α-syn.

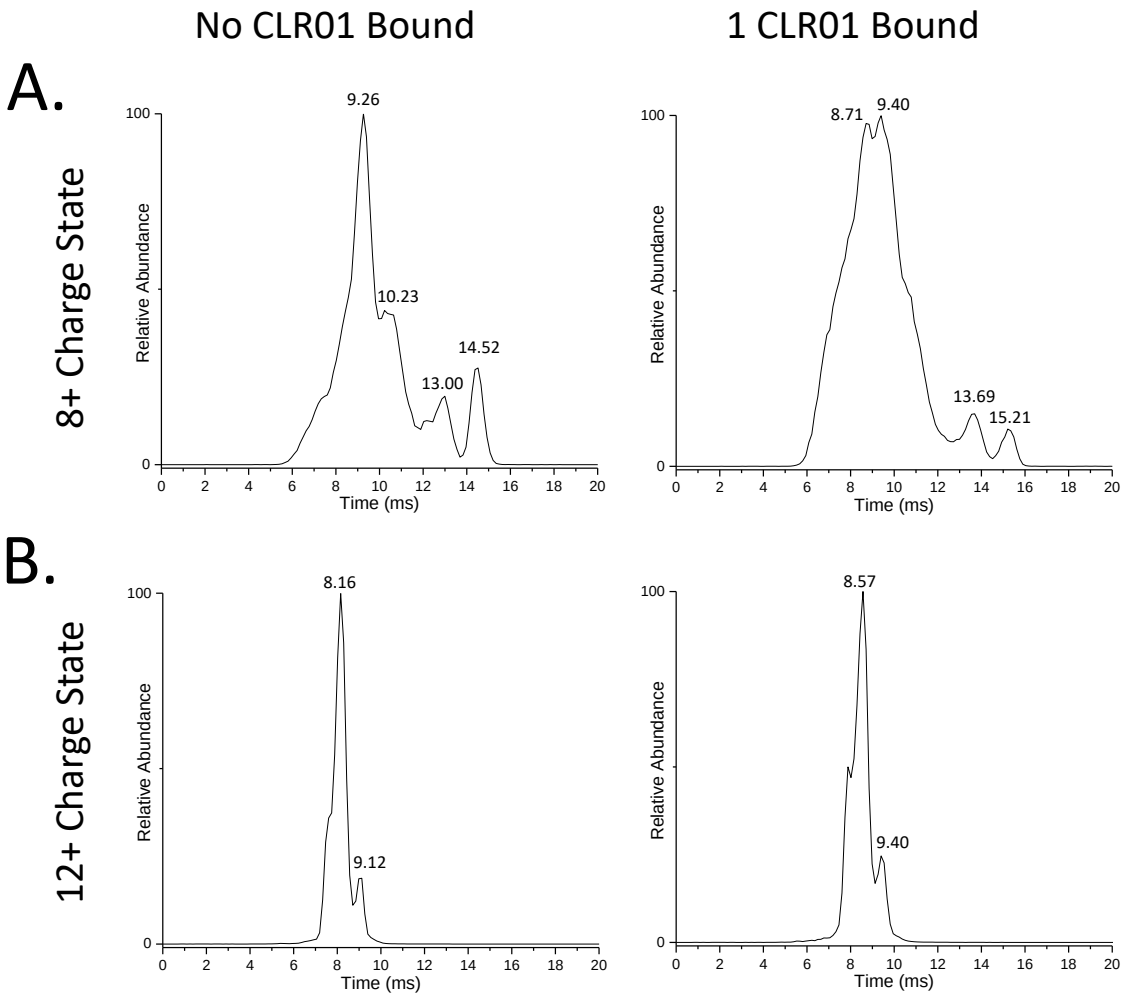


Figure S9: Ion mobility profiles for the unbound and CLR01 bound states of the A.) 8+ and B.) 12+ charge states of copper bound α-syn.

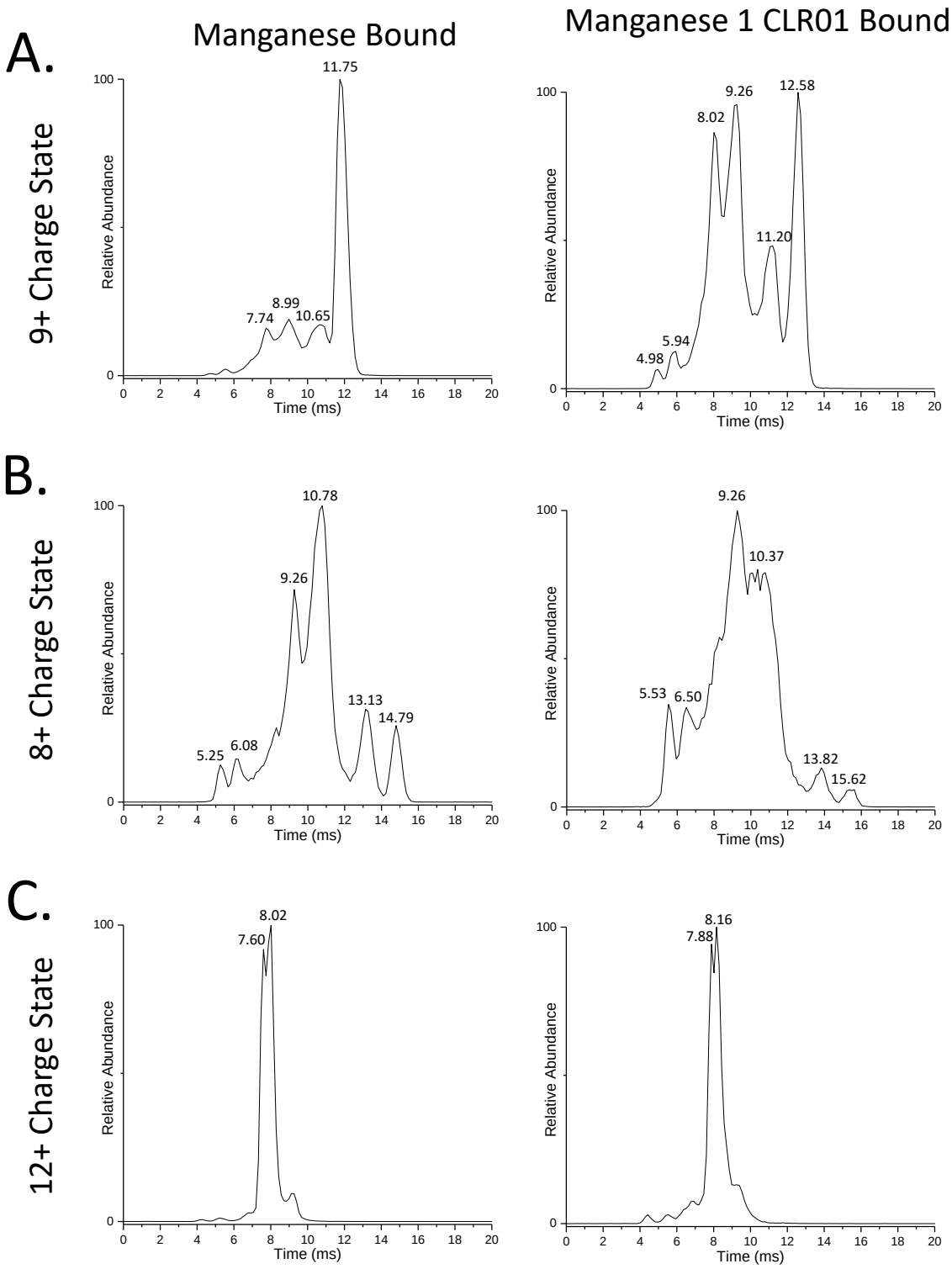


Figure S10: Ion mobility profiles for the unbound and CLR01 bound states of the A.) 9+, B.) 8+ and C.) 12+ charge states of manganese bound α-syn.

CHAPTER 7

Mass Spectrometry Analysis Reveals Size and Structure Information of Amyloid Protein Oligomers

ABSTRACT: Neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease are characterized by the aggregation of amyloid proteins in the brain. These aggregates can be large inclusions, fibrils, or small oligomers. It has been hypothesized that oligomers of amyloid proteins disrupt cellular processes in neurons causing their death; however, due to their small size and transient nature, analysis of these small aggregates have proven difficult. Here, we utilize native mass spectrometry (nMS) and top-down mass spectrometry (TD-MS) to provide information on the size of amyloid protein oligomers and information on the location of the aggregation interface along the sequence of the protein. We find that nMS analysis of amyloid-β oligomers reveals charge states while nMS of tau and α-synuclein oligomers only reveals oligomeric populations. In addition, we find that amyloid-β oligomers release monomers products under collision-based fragmentation while tau and α-synuclein oligomers release covalent fragments that reveal the location of the aggregation interface along the sequence of the protein. These techniques provide a promising way to characterize oligomers of amyloid proteins and may provide information on how these oligomers become toxic in neurodegenerative diseases.

INTRODUCTION

Neurodegenerative diseases such as Alzheimer's and Parkinson's disease are characterized by aggregated protein in the brain. In Alzheimer's disease, tau aggregates into neurofibrillary tangles,[1] in Parkinson's disease α-synuclein aggregates into Lewy bodies,[2] and in Machado Joseph's disease ataxin aggregates into nuclear inclusions.[3] Although these diseases are characterized by large aggregates of protein in the brain, smaller oligomers species of these proteins have also been reported to exist in the brains of neurodegenerative disease patients and have been hypothesized to be the toxic species of these diseases.[4] These aggregates are thought to disrupt cellular processes in the brain by inserting themselves into cell and organelle membranes causing destabilization of cellular processes.[5] Due to the contribution of these

oligomers to the progression of neurodegenerative diseases, it is important to characterize these oligomers and to understand how they form.

Amyloid protein aggregation is not fully understood, however there are factors which seem to affect the aggregation rate. It has been reported that post-translational modifications such as phosphorylation of tau[6] and α-synuclein[7] modulate the rate of aggregation and some sites have been found to be triggers for the aggregation process. Metal ion binding has also been linked to increased amyloid protein aggregation. For example, aluminum seems to increase aggregation of tau[8] and copper seem to increase the aggregation rate of α-synuclein.[9] Small molecules have been found to inhibit the progression of amyloid protein aggregation. For example, it has been reported that EGCG inhibits the formation of amyloid-β aggregates[10] and CLR01 inhibits α-synuclein aggregation.[11] Although aggregation has been studied extensively, little is known about amyloid proteins oligomers. Questions as fundamental as the size of the oligomers and the structure of the oligomer interface remain a mystery.

It has been hypothesized that oligomers are the toxic species in these neurodegenerative diseases. This hypothesis is based off the now controversial study reporting that oligomers of amyloid beta are toxic in mice.[4] This result is backed by other labs including the Mike Bower's lab where he reported the existence of dodecamer oligomers of amyloid-β.[12] A couple other studies have reported the location of amyloid protein aggregation interfaces[13] and the type of structure binding amyloid protein oligomers.[14] However, because of the transient nature of oligomers, it has been difficult to study them with conventional techniques. Mass spectrometry is uniquely qualified to study the size of these oligomers and the interface of these oligomers. In fact, there have been a few studies using native mass spectrometry,[13, 15] ion mobility mass spectrometry,[12, 15-17] and top-down mass spectrometry to obtain information on these oligomers.[13] In this study, we reveal how native mass spectrometry and top-down mass spectrometry can be utilized to return much needed information on amyloid protein oligomers formed in vitro.

MATERIALS AND METHODS

Oligomer solutions were created and buffer exchanged into 20mM ammonium acetate with 10k amicon filters from Sigma Aldrich. The solutions were directly sprayed on a Thermo UHMR (Thermo Fisher Scientific, San Jose, CA) with a voltage of 1.5-2.0kV. Amyloid-β oligomers were deconvoluted by simulating gaussian distributions from MATLAB and calculating the molecular weight from those values. The oligomers were activated with either 100V of in-source collisionally activated dissociation (CAD) or 200-300V of high-energy C-trap dissociation (HCD). The fragments were deconvoluted with Biopharma 3.2 and *b*- and *y*- fragments were searched with ClipsMS. Sodium ions were added as unlocalized modifications.

RESULTS

To obtain information on the size of amyloid protein oligomers, MS1 analysis was performed on incubated solutions of amyloid proteins. MS1 of amyloid-β oligomers reveals broad unresolved peaks that correspond to charge states. (Fig. 1A) The charge states range from 3+ to 9+ with the most abundant charge state being 6+. The average molecular weight of the oligomers is revealed to be about 55kDa which corresponds to oligomers comprised of 12 amyloid-β monomers. The unresolved nature of the charge states indicates there are multiple oligomer states. The distribution seems to indicate that oligomer states range from 11 monomers to 13 monomers. This analysis reinforces previous studies suggesting that amyloid-β oligomers aggregate into dodecamers.[4, 12] It is possible that this oligomer species is a toxic species and seeds larger aggregates of amyloid-β such as fibrils and plaques in the brain.

Other amyloid protein oligomers were analyzed to obtain size information. Tau K18+ oligomers revealed multiple peaks in the spectrum. These peaks did not correspond to charge states; however, they could correspond to oligomeric states. (Fig. 1B) This spectrum indicates that there are multiple populations of oligomers in the spectrum rather than one gaussian peak of oligomer states. More information is needed to determine the exact molecular weights of these oligomers or if certain sizes are more toxic than other oligomers; however, it is possible that these oligomers could be toxic and lead to the formation of larger

aggregates. α-Synuclein oligomers on the other hand did not reveal any size information. The peak detected revealed itself as one broad unresolved peak. (Fig. 1C) Presumably multiple oligomeric states, charge states, and other proteoforms are present within this one broad peak. Better desolvation and higher resolution are needed to determine the precise size information for these oligomers.

Top-down mass spectrometry has been performed on protein complexes including amyloid protein oligomers. TD-MS can provide the identity of the monomers in the complex and even provide information on structure characteristics.[13] HCD of amyloid-β oligomers revealed peaks corresponding to amyloid-β monomers as well as various fragments. (Fig. 2A) The presence of intense monomer peaks reveals that the strength of the oligomers is not as strong as other protein complexes such as ADH or aldolase. Analysis of the fragments present in the spectrum reveal fragments that contain the N-terminus and the C-terminus. The fragments reveal the binding of sodium ions which may contribute to the heterogeneity of the oligomers. HCD fragmentation of amyloid-β oligomers confirms the identity of the oligomers and reveals that different sodium bound proteoforms exist within the oligomers.

K18+ tau oligomers were also fragmented with high energy C-trap dissociation (HCD). Fragmentation of the oligomer peak did not reveal monomer peaks; however, it did reveal covalent fragments. (Fig. 2B) The fact there are no monomers in the spectrum revealed that the interaction between these monomers is stronger than the amyloid-β oligomers and could reveal structure information. It was found that the fragments that were present in the spectrum corresponded to C-terminal *y*-fragments. This indicates that the oligomer interface is on the N-terminus of the protein. This region contains the VQIVYK peptide which is a section of the tau sequence that is known to aggregate.[18] HCD fragmentation of tau oligomers reveals that the fragments contain structure information and suggests that the oligomer interface is on the N-terminus of the protein.

α-Synuclein oligomers were fragmented with in-source CAD to obtain modification and structure information on the oligomers. The fragmentation spectrum of α-synuclein oligomers did not reveal monomer peaks, although, the spectrum did reveal multiple covalent fragments that could be mapped to the protein sequence. (Fig. 2C) The spectrum revealed that exclusively *y*-fragments are released from the

oligomers and multiple sodium ions bind these oligomers. These sodium ions could contribute to the heterogeneity of the heterogeneity of the MS1 spectrum. Since no monomers are present in the spectrum, the interaction between these monomers is stronger than the amyloid-β oligomers, and the fragments could reveal structure information. The fragments in the spectrum only contain the C-terminus which indicates the aggregation interface is closer to the N-terminus. This lines up nicely with a previous fibril structure known as the rod polymorph,[19] so it is possible that the site of aggregation for these oligomers is close to the rod polymorph aggregation site.

CONCLUSION

Development of mass spectrometry techniques and instrumentation has provided the ability to analyze amyloid protein oligomers. We have found that oligomeric populations and in some cases charge states are able to be resolved when analyzing with MS. In addition, TD-MS with collision-based fragmentation techniques can reveal the identity of monomers present in the oligomers and in some cases information on the oligomer interface. In the future it may be possible to perform charge detection mass spectrometry (CD-MS) for direct mass analysis of the oligomers. This technique which measures the m/z and charge of an ion simultaneously can provide the mass of amyloid protein oligomers and the distribution of oligomeric units. In addition, ECD of amyloid protein oligomers could provide fragments that return more precise information on the location of the aggregation interface. These techniques should give us more information on amyloid protein oligomers and may explain how they form and become toxic species in neurodegenerative diseases.

FIGURES

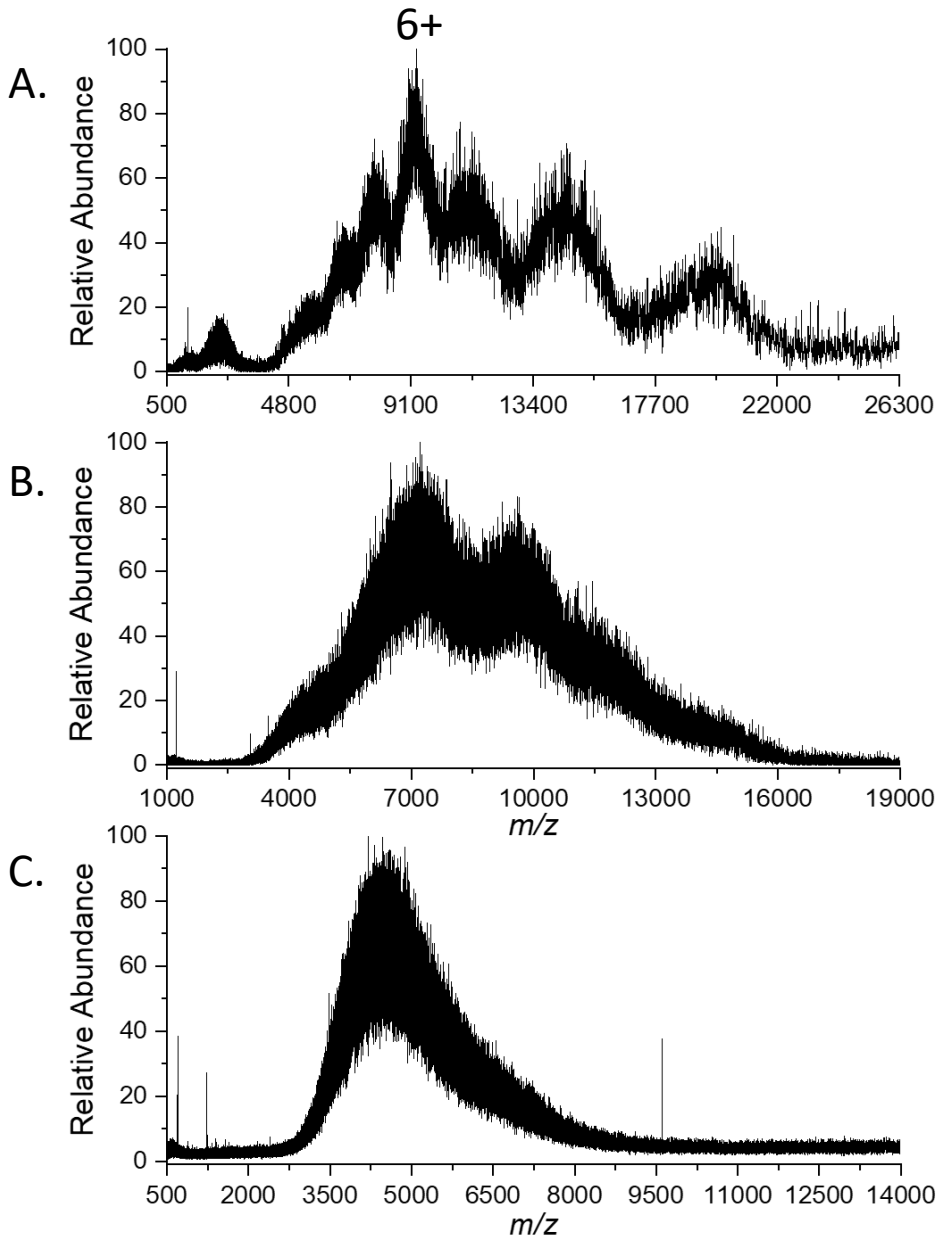


Figure 1: Spectra of A.) amyloid-β, B.) tau, and C.) α-synuclein oligomers. The amyloid-β spectrum reveals peaks corresponding to charge states, the tau spectrum reveals peaks corresponding to oligomeric states, and the α-synuclein spectrum does not reveal any size information.

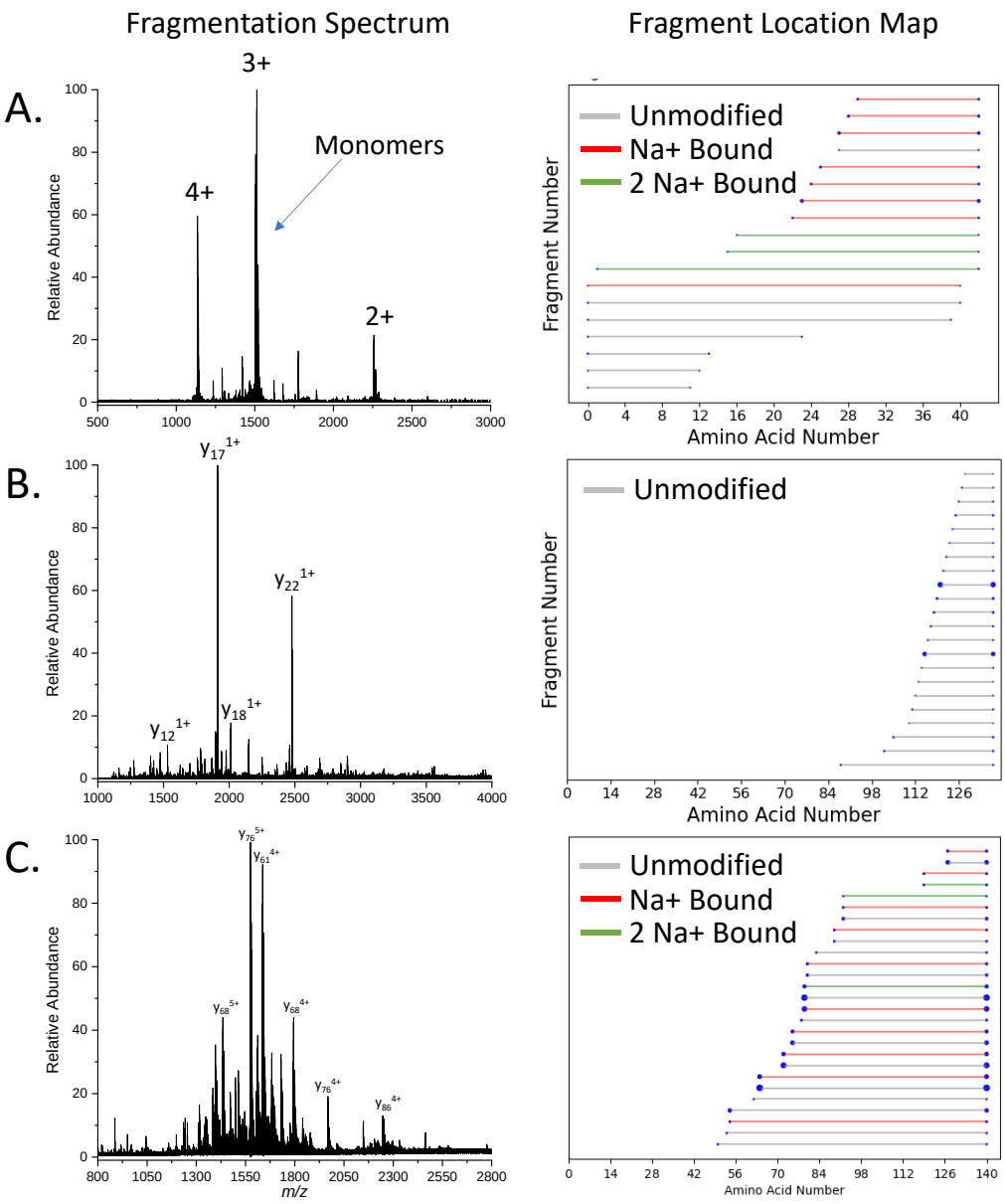


Figure 2: TD-MS spectra of A.) amyloid-β, B.) tau, and C.) α-synuclein oligomers with the corresponding fragment location maps. The amyloid-β spectrum reveals monomer peaks in addition to fragment peaks. The tau and α-synuclein spectra only reveal peaks corresponding to y-fragments which indicates that the aggregation interface is on the N-terminus of those oligomers.

REFERENCES

1. Johnson, V. E.; Stewart, W.; Smith, D. H., Widespread tau and amyloid-beta pathology many years after a single traumatic brain injury in humans. *Brain pathology* **2012,** *22* (2), 142-149.

2. Baba, M.; Nakajo, S.; Tu, P.-H.; Tomita, T.; Nakaya, K.; Lee, V.; Trojanowski, J. Q.; Iwatsubo, T., Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies. *The American journal of pathology* **1998,** *152* (4), 879.

3. Yamada, M.; Hayashi, S.; Tsuji, S.; Takahashi, H., Involvement of the cerebral cortex and autonomic ganglia in Machado-Joseph disease. *Acta neuropathologica* **2001,** *101* (2), 140-144.

4. Lesné, S.; Koh, M. T.; Kotilinek, L.; Kayed, R.; Glabe, C. G.; Yang, A.; Gallagher, M.; Ashe, K. H., A specific amyloid-β protein assembly in the brain impairs memory. *Nature* **2006,** *440* (7082), 352-357.

5. Ferreira, S. T.; Vieira, M. N.; De Felice, F. G., Soluble protein oligomers as emerging toxins in Alzheimer's and other amyloid diseases. *IUBMB life* **2007,** *59* (4-5), 332-345.

6. Martin, E. W.; Holehouse, A. S.; Grace, C. R.; Hughes, A.; Pappu, R. V.; Mittag, T., Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. *Journal of the American Chemical Society* **2016,** *138* (47), 15323-15335.

7. Fujiwara, H.; Hasegawa, M.; Dohmae, N.; Kawashima, A.; Masliah, E.; Goldberg, M. S.; Shen, J.; Takio, K.; Iwatsubo, T., α-Synuclein is phosphorylated in synucleinopathy lesions. *Nature cell biology* **2002,** *4* (2), 160-164.

8. Crapper, D.; Krishnan, S.; Dalton, A., Brain aluminum distribution in Alzheimer's disease and experimental neurofibrillary degeneration. *Science* **1973,** *180* (4085), 511-513.

9. Choi, T. S.; Lee, J.; Han, J. Y.; Jung, B. C.; Wongkongkathep, P.; Loo, J. A.; Lee, M. J.; Kim, H. I., Supramolecular modulation of structural polymorphism in pathogenic α-synuclein fibrils using copper (II) coordination. *Angewandte Chemie International Edition* **2018,** *57* (12), 3099-3103.

10. Ehrnhoefer, D. E.; Bieschke, J.; Boeddrich, A.; Herbst, M.; Masino, L.; Lurz, R.; Engemann, S.; Pastore, A.; Wanker, E. E., EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nature structural & molecular biology* **2008,** *15* (6), 558-566.

11. Prabhudesai, S.; Sinha, S.; Attar, A.; Kotagiri, A.; Fitzmaurice, A. G.; Lakshmanan, R.; Ivanova, M. I.; Loo, J. A.; Klärner, F.-G.; Schrader, T., A novel "molecular tweezer" inhibitor of α-synuclein neurotoxicity in vitro and in vivo. *Neurotherapeutics* **2012,** *9* (2), 464-476.

12. Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V., Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nature chemistry* **2009,** *1* (4), 326-331.

13. Lam, Y. P.; Wootton, C. A.; Hands-Portman, I.; Wei, J.; Chiu, C. K.; Romero-Canelon, I.; Lermyte, F.; Barrow, M. P.; O'Connor, P. B., Determination of the aggregate binding site of amyloid protofibrils using electron capture dissociation tandem mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (2), 267-276.

14. Stroud, J. C.; Liu, C.; Teng, P. K.; Eisenberg, D., Toxic fibrillar oligomers of amyloid-β have cross-β structure. *Proceedings of the National Academy of Sciences* **2012,** *109* (20), 7717-7722.

15. Bernstein, S. L.; Liu, D.; Wyttenbach, T.; Bowers, M. T.; Lee, J. C.; Gray, H. B.; Winkler, J. R., α-Synuclein: stable compact and extended monomeric structures and pH dependence of dimer formation. *Journal of the American Society for Mass Spectrometry* **2004,** *15* (10), 1435-1443.

16. Bleiholder, C.; Do, T. D.; Wu, C.; Economou, N. J.; Bernstein, S. S.; Buratto, S. K.; Shea, J.-E.; Bowers, M. T., Ion mobility spectrometry reveals the mechanism of amyloid formation of Aβ (25–35) and its modulation by inhibitors at the molecular level: epigallocatechin gallate and scyllo-inositol. *Journal of the American Chemical Society* **2013,** *135* (45), 16926-16937.

17. Murray, M. M.; Bernstein, S. L.; Nyugen, V.; Condron, M. M.; Teplow, D. B.; Bowers, M. T., Amyloid β protein: Aβ40 inhibits Aβ42 oligomerization. *Journal of the American Chemical Society* **2009,** *131* (18), 6316-6317.

18. Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J.; McFarlane, H. T., Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **2007,** *447* (7143), 453-457.

19. Li, B.; Ge, P.; Murray, K. A.; Sheth, P.; Zhang, M.; Nair, G.; Sawaya, M. R.; Shin, W. S.; Boyer, D. R.; Ye, S., Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nature communications* **2018,** *9* (1), 1-10.

CONCLUSION

This work reveals that mass spectrometry techniques can readily return relevant information on proteins and protein complexes. The development of algorithms to assign internal as well as terminal fragments generated by top-down mass spectrometry provides the ability to extract more information out of a mass spectrum than assignment of terminal fragments alone. Because internal fragments make up a significant number of observed fragments and those fragments cover the center of the protein sequence, they promise an increase in efficiency of characterizing protein sequences, localizing modifications, and revealing structure information. However, internal fragment duplicate assignments provide a significant hurdle to overcome. Fragments with the same chemical composition or fragments that are not distinguishable within the allotted mass error make assignment of some internal fragments difficult. Further evaluation of these fragments with more accurate mass analyzers, ion-mobility analysis, and MS^3 fragmentation of selected fragments may provide a greater understanding of internal fragment generation. Nevertheless, the assignment of internal fragments in a top-down mass spectrum shows promise of increased information gleaned from this data for the analysis of large proteins and protein complexes.

Recent developments in top-down mass spectrometry instrumentation have provided the ability to obtain relevant information on intact proteins and protein complexes. In this work it is described how in some circumstances the collisionally activated technique high energy C-trap dissociation (HCD) can reveal quaternary structure information on protein complexes. HCD is available on many Thermo instruments and is easy to implement in experimental workflows which should provide the ability for other labs to obtain structure information on their protein complexes relatively quickly and efficiently. It is possible that future research will provide insight into which proteins release covalent fragments that are sensitive to structural characteristics and why this phenomenon occurs. Moreover, we find that electron-based techniques (ExD) have their own advantages for fragmenting proteins. Electron-capture dissociation (ECD) of proteins with weakly bound PTMs such as phosphorylation can provide information on locations of those sites without disturbing the phosphate on the protein. Both ECD and electron ionization dissociation (EID) of protein complexes provide accurate structure information on solvent exposed regions of protein complexes. These

capabilities provide great promise for the study of these proteins; however, ECD tuning can be difficult and time consuming making it harder to implement in experimental workflows than collisionally activated techniques such as HCD. We hope with further development of ExD techniques and more widespread installation of these devices that electron-based fragmentation can become a routine fragmentation technique for the study of proteins.

Amyloid protein aggregation has been correlated to the death of neurons and the onset of neurodegenerative diseases. Characterization of amyloid protein monomers is important to determine the reason these proteins aggregate and for the discovery of potential therapies and cures. This work reveals that some sites of phosphorylation compact amyloid protein structure. This shift in protein dynamics could expose the aggregation region and allow monomers of amyloid proteins to interact and form oligomers and fibrils. However, this explanation is still speculative at this point. More research is needed to determine if the structural shift observed in the gas-phase is also present in solution or if another mechanism can explain the increase in aggregation potential of amyloid proteins. Furthermore, small molecule binding to amyloid proteins has been shown to decrease the aggregation rate. This work indicates that CLR01 binds the N-terminus of multiple proteoforms of α-synuclein with a K_d in the low micromolar range and that the molecule compacts the structure of the protein. The compaction of amyloid protein structure from small molecules could shield the aggregation region of amyloid proteins and prevent the protein from forming oligomers, fibrils, and larger aggregates. Small molecule inhibitors like CLR01 provide a promising way to inhibit aggregation, but challenges such as transporting them to brain neurons and indiscriminate binding to other proteins pose a problem. It is possible that modifying these compounds so they pass through to the brain and performing in vivo experiments to make sure these compounds do not significantly affect other cellular processes will yield a compound with the ability to inhibit fibril formation and to disrupt already existing aggregates in the brain.

Oligomers of amyloid proteins are thought to be the toxic species in neurodegenerative diseases. These small aggregates are thought to disrupt cellular processes which can cause havoc in brain neurons. Characterization of these oligomers is important, but their transient nature and small size make them

difficult to analyze. This work shows that mass spectrometry can provide relevant information on amyloid protein oligomers including their size and their aggregation interface. This data can be used to provide information about the formation and the toxicity of these oligomers. Although the mass spectrometry techniques utilized to obtain this information show promise for characterizing oligomers, the heterogeneous nature of the oligomers still pose a significant problem. Native mass spectrometry reveals that some oligomer solutions are too heterogeneous to obtain molecular weight information. Other recently developed techniques such as charge detection mass spectrometry (CD-MS) show promise for determining the exact molecular weights of these species. Future directions also include utilizing electron-based fragmentation techniques on an orbitrap-based instrument to reveal the location of the oligomer interface. It is possible that in time these mass spectrometry techniques will be used to quickly and efficiently characterize these small aggregates which could provide information on the toxicity of these species.