UNIVERSITY OF CALIFORNIA
RIVERSIDE

Leveraging Mass Spectrometry for Structural Analysis: Peptide Isomers to Intact Proteins

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Brielle L. Van Orman

September 2024

Dissertation Committee:
Dr. Ryan R. Julian, Chairperson
Dr. Leonard J. Mueller
Dr. Yinsheng Wang

Copyright by
Brielle L. Van Orman
2024

The Dissertation of Brielle L. Van Orman is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

There are far too many people to list here, but I am forever grateful to anyone who has supported me throughout this journey. I could not have made it this far without the mentorship, support, and encouragement of many people. First and foremost, thank you to my parents for teaching me the virtues of hard work and never, ever giving up. I know the time apart has been difficult, and my trips home were too far and few between. Thank you for always being there when I needed it. To my siblings Forrest, Wilsey, Kori, and Kevin, I know we don't talk often enough, but I know you were there in spirit to support me, thank you for being you. To the Gagnon family, thank you for all your support and for giving me a second home and many unforgettable moments. Your encouragement carried me to the finish line. Troy, thank you for being such a supportive partner and getting kitties with me. Your encouragement and patience mean the world to me.

I have been privileged to have had some incredible mentors over the years. I owe a huge thank you to my high school science teachers, Dan Glasser, Rhonda Kuntz, and Jan Stromsborg, for fostering my love of science. Dr. Alena Kubatova was an amazing advisor during my time at UND and gave me my first research experience. Throughout my time at UCR, I have been fortunate to encounter many exceptional instructors and researchers. Thanks to Dr. Wenwan Zhong and Dr. Min Xue for teaching informative courses and providing constructive feedback while serving on my committee. Thank you to Dr. Yinsheng Wang for teaching a great mass spec course and joining my committee late in the game. Dr. Len Mueller, thank you for valuable advice and serving on my

committee. Dr. Joey Genereux has been a great ally throughout my PhD, from sharing equipment and talking over experiments, to providing invaluable advice and perspective on many topics.

I owe my best memories during graduate school to my peers. I owe a big thanks to Chris Williams and Dr. Drew Kellum for always sparing an ear to listen and commiserate. I owe any success to the members of the Julian lab. Thank you to Dr. Tyler Lambeth for leading by example and helping me navigate the beginning of my graduate career. A special thanks to Dr. Ting Ting Quanrud for being the best mentor and friend, I couldn't do it without you! I went through it all with Evan Hubbard, and I couldn't ask for a better colleague. To Dr. Jacob Silzel, Dr. Gaurav Pandey, Dr. Sandeep Kumar, Thomas Shoff, Sharon Chen, Lin He, and Toma Chowdhury, thank you for the camaraderie and collaboration over the years.

Lastly, thank you to my advisor Dr. Ryan Julian. The instruction and the opportunities you have given me have been invaluable. I know I certainly wasn't the most gracious person to mentor, but I will forever be thankful that you didn't give up on me even during contentious times. The institutions and fields you are a part of and students you mentor are made immeasurably better because of your dedication to upholding standards and following data rather than what is popular, and I am no exception. Thank you for everything.

The text in this dissertation, in part, is a reprint of the materials as they appear in the following publications:

Chapter 3:

Van Orman, B. L.; Wu, H.-T.; R. Julian, R. Differentiation of Peptide Isomers by Excited-State Photodissociation and Ion–Molecule Interactions. *Physical Chemistry Chemical Physics* **2020**, *22* (41), 23678–23685.

ABSTRACT OF THE DISSERTATION

Leveraging Mass Spectrometry for Structural Analysis: Peptide Isomers to Intact Proteins

by

Brielle L. Van Orman

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, September 2024
Dr. Ryan R. Julian, Chairperson

Protein structure is fundamental for biological functioning. Although there are many safeguards for maintaining protein structural fidelity, proteins are still subject to modifications with deleterious effects. Isomerization of amino acids is one such modification resulting in changes to protein structure and function. Analytical methods to interpret protein structure and locate modifications are required to expand our knowledge of protein structure and its role in biology and disease. Mass spectrometry (MS) based methods are uniquely suited to address these needs. Herein, MS methods are discussed for applications in biomolecular structure studies in several contexts.

The development of a method to detect isomerization in human brain tissue is described in Chapter 2. Sample preparation methods, instrument methods, and data acquisition and analysis are described. Using the methods proposed, we located isomerization and post-translational modification (PTM) sites in tau protein and superoxide dismutase, both associated with neurodegenerative disease, in human brain

tissue. The methods developed will be used for further investigation of the role of these modifications in Alzheimer's disease and other neurodegenerative disorders.

Next, ultraviolet photodissociation (UVPD) was used to interrogate structural differences induced by isomerization in peptide isomers. Differences in chromophore absorption and ion-molecule reactions were used to differentiate synthetic peptides containing aspartic acid (Asp), glutamic acid (Glu), and serine (Ser) isomers. Molecular modeling was used to propose potential structures in the gas phase.

We also performed structure studies in the gas phase on whole proteins. Radical directed dissociation (RDD) and UVPD were used to fragment iodine-modified ubiquitin and myoglobin. Results of these experiments reveal conformational differences between charge states, including high charge states. They also provide information about structural differences and similarities between apo- and holo-myoglobin.

Lastly, we analyzed 193 nm UVPD and HCD data to test the statistical confidence of assigning internal fragments. We found that assignments of internal fragments feature a high probability of being assigned when an incorrect sequence is used for assignment. We also report ETD data on an intact protein and peptide to observe differences in terminal ions at varying activation time and propose why complement ions pairs are often not observed in top-down analyses of proteins.

TABLE OF CONTENTS

Chapter 1 Introduction 1

Protein Structure 1

Mass Spectrometry for Protein Structure 4

Radicals in MS 6

Isomers 8

AD, Proteostasis, and Aging 10

Internal ions 14

Conclusions 15

Chapter 2 Locating Tau Isomerization In Vitro and in Tissue Samples 21

Introduction 21

Isomerization 21

Detection in Tissue 23

Experimental 24

In Vitro aging 24

Digest 25

LC-MS/MS Analysis 25

Isomer Detection in Tissue 26

Results and Discussion 29

In Vitro aging 29

Isomer Detection in Tissue 31
Conclusions 37
References 38
Chapter 3 Differentiation of Peptide Isomers by Excited-State Photodissociation and Ion-Molecule Interactions 42
Introduction 42
Methods 45
Peptide synthesis 45
UVPD experiments 45
Data 45
Simulated Annealing 46
Results and Discussion 46
Simulations 54
Ion-molecule reactions 57
Conclusions 59
Supplemental Information 60
References 61
Chapter 4 Radical initiated photodissociation and radical directed dissociation provide complementary gas phase structural information 65
Introduction 65

Experimental 68
Materials 68
Modification of Proteins 68
Mass Spectrometry 68
Results and Discussion 69
Conclusion 83
References 85
Chapter 5 False positives in internal fragment assignments 88
Introduction 88
Experimental 89
Materials 89
Data analysis 89
Experimental 90
Results and Discussion 90
Experimental 90
ClipsMS 99
Conclusion 107
Supporting Information 109
References 110
Chapter 6 Concluding Remarks 114

LIST OF FIGURES

Figure 1.1 The 26S proteasome, a 2.5 megaDalton complex responsible for protein degradation. Modeled from PBD 5GJQ. 3

Figure 1.2 Roepstorff notation in example peptide. 5

Figure 1.3 Radical precursors 4IB, iodo-tyrosine, and 4IPT from left to right. 8

Figure 1.4 When Asp is isomerized, active sites may not be able to form necessary interactions with substrates. 13

Figure 2.1 Progression of isomerization of Asp387 in recombinant tau. 30

Figure 2.2 Isomer distribution in hippocampal sample TDH peptide. 32

Figure 2.3 Asp387 in a frontal cortex sample. a) Asp387 isomer distribution in the sarkosyl pellet. b) Asp387 isomer distribution in the tau pellet. Tau isomerization is more advanced in the tau pellet. 34

Figure 2.4 All detected modifications mapped across tau sequence. Oxidation sites are highlighted in red, deamidation in green, phosphorylation in yellow, and isomerization in pink. 36

Figure 3.1 Representative mass spectra from isomers of 4IB-VKLDSG (underlined residue locates isomerized position) used to quantitate chromophore excitation at 213 nm a) L-Asp and b) D-Asp and at 266 nm c) L-Asp and d) D-Asp. 48

Figure 3.2 Relative chromophore excitation for a) [4IB-VKLDSG+H]+ b) [4IB-VKLDHG+H]+ and c) [4IB-VKLDHG+2H]2+ following excitation at 213 nm (shown in blue) and 266 nm (shown in orange). Error bars represent standard error of regression. 51

Figure 3.3 Relative chromophore excitation in peptides a) [4IB-AIPVSR+H]+, b) [4IB-VHLGGEGYK+H]+, and c) [4IB-VTIHEGGPWFK+H]+ at UVPD 213 nm (shown in blue) and 266 nm (shown in orange). Error bars represent standard error of regression. 53

Figure 3.4 Simulated annealing results for 4IB-VKLDHG protonated on lysine. Scatter plots on the left show the potential energy of structures within 50 kJ of the lowest energy structure versus the distance between the iodine atom on 4IB and the central carbon of the aspartic acid in the side chain. On the right, the number of structures within each distance bin are shown in a histogram. 56

Figure 3.5 Formation of +O (shown in grayblue) and +O2 (shown in greenorange) adducts in a) [4IB-AIPVSR+H]+ and b) [4IB-VKLDSG+H]+. These adducts were only observed in significant abundances in 213 nm experiments. Error bars represent standard error of regression. 58

Figure S3.1 Full range of structures generated by simulated annealing of 4IB-VKLDHG, where Fig. 6 shows structures with potential energy within the lowest 50 kJ. 60

Figure 4.1 PD provides structural information from alpha radical dissociation. a) PD spectrum of iodo-ubiquitin in the 7+ charge state. b) Locations of the radical fragments mapped on the protein sequence. c) Locations of the radical induced fragments on the crystal structure. d) Location of radical fragments and radical origin mapped on ubiquitin crystal structure with secondary structures shown. All fragments and Tyr59 are located on flexible regions. 71

Table 4.1 Distance of Tyr59 from alpha carbons of residues where radical induced fragments are located in the 7+ PD spectrum, from crystal structure**.** 73

Figure 4.2 PD spectrum of iodo-ubiquitin in the 5+ charge state. No backbone fragmentation is present. 74

Figure 4.3 RDD provides structural information at charge states where PD does not. a) PD spectrum of iodo-ubiquitin in the 5+ charge state. No backbone fragmentation is present. b) Location of radical fragments and radical origin mapped on ubiquitin sequence. c) Location of radical fragments and radical origin mapped on ubiquitin crystal structure. 76

Figure 4.4 Specific charge states result in specific radical dissociations. a) The 10+ charge state of iodo-ubiquitin results in a very intense -106Y fragment. b) When the -106Y fragment is reisolated and subject to further photoactivation, it fragments with very high specificity into an a+1/y-2 pair, indicating an alpha radical. c) Radical fragments plotted by charge state. -106Y is highly charge state dependent, suggesting a conformation in which the tyrosine sidechain is constrained from the rest of the backbone. 78

Figure 4.5 RDD of apo and holo myoglobin produce similar but distinct fragmentation. Radical fragments of 8+ holo myoglobin plotted on a) the crystal structure and b) protein sequence. Radical fragments of 7+ apo myoglobin plotted on c) the crystal structure and d) protein sequence. 80

Figure 4.6 The apo and holo versions of the protein share surprisingly similar features. Here the locations of RDD fragments are plotted on PDB structure 1BJE. a) The remote fragments in the holo protein are Phe138 and His119. The heme group is displayed in Kelly gprotein are apo protein features remote fragments at His82 and His116, denoted with arrows. c) The secondary structures of myoglobin shown. His119 and His116 (bottom circle) are close in sequence, and will always have to located near each other. However, Phe138 and His82 (top circle) are located on different secondary structures. It is surprising that the contact between the two structures is preserved in both apo and holo forms of the protein. 83

Figure 5.1 Fragmentation observed for ETD of 18+ histone protein 3 with a) 1 ms activation time, b) 2 ms activation time, and c) 10 ms activation time. 93

Figure 5.2 Proportion of assigned ions with complement pairs and proportion of ions that cannot be assigned as terminal ions for charge states a) 18+, b) 22+, and c) 24+. 94

Figure 5.3 ETD Fragmentation of the 6+ charge state of peptide from histone protein 3 (1-20). Activation times of a) 1ms, b) 5 ms, and c) 10 ms. 98

Figure 5.4 Percentages of non-terminal ions assigned with the correct sequence (blue bar) and incorrect sequences (orange bars). Error tolerances were 2, 5, and 10 ppm; results are from HCD data from a-c) myoglobin, d-f) carbonic anhydrase, g-k) aldolase, and j-l) enolase. 100

Figure 5.5 For HCD data, there are trends based on protein size. The proportion of ions assigned using the shuffled sequences divided by assignments using the correct sequence plotted against protein size and the standard deviation of the shuffled sequences plotted against protein size for a, b) 2 ppm error tolerance; c, d) 5 ppm, and e, f) 10 ppm. 102

Figure 5.6 Percentages of non-terminal ions assigned with the correct sequence (blue bar) and incorrect sequences (orange bars). Error tolerances were 2, 5, and 10 ppm; results are from 193 nm PD data from a-c) myoglobin, d-f) carbonic anhydrase, g-k) aldolase, and j-l) enolase. 104

Figure 5.7 For UVPD data, trends based on protein size are less apparent. The proportion of ions assigned using the shuffled sequences divided by assignments using the correct sequence plotted against protein size and the standard deviation of the shuffled sequences plotted against protein size for a, b) 2 ppm error tolerance: c, d) 5 ppm, and e, f) 10 ppm. 106

LIST OF SCHEMES

Scheme 1.1 Aspartic acid isomerization mechanism 8

Scheme 2.1 Flowchart of tissue sample fractionation. 27

Scheme 3.1 Different solvation environments created by alternate arrangements of charged and polar groups could potentially influence the photochemistry of an active chromophore. 43

Scheme 4.1 The experimental workflow for generating radicals on proteins for PD dissociation or RDD..... 67

Scheme 4.2 Proposed mechanism for loss of tyrosine sidechain. 77

LIST OF TABLES

Table 2.1 Concentrations of fractions from initial fractionation protocol. For the rest of the samples analyzed, fractionation was stopped after separation into two fractions, sarkosyl soluble and insoluble. 33

Table 3.1 Numerical values for PD yields 54

Table 3.2 Numerical values for ion-molecule oxygen adducts..... 59

Table 4.1 Distance of Tyr59 from alpha carbons of residues where radical induced fragments are located in the 7+ PD spectrum, from crystal structure**.**..... 73

Chapter 1
Introduction

Protein Structure

Proteins come in many shapes and sizes. From the humble ubiquitin protein to megadalton protein complexes, proteins perform innumerable critical biological functions. Proteins account for more than half of a cell's dry mass and are indisputably important players in biological function.[1] Their constituent molecules, amino acids, vary widely in chemical properties. With twenty canonical amino acids available, the possible structure of a protein is extremely diverse. The wide-ranging structural possibilities allow proteins to be the jacks of all trades of biomolecules. Similarly to archaeologists studying ancient tools, we are interested in determining what tasks specific proteins perform based on their structure. Similarly to tools, with time and exposure to non-ideal conditions, proteins may be altered or damaged. Studying these changes may be of utmost importance to understanding, and someday preventing and treating, the multitude of diseases and disorders that arise from alterations in protein structure.

Protein structure has several orders. Primary structure is determined by the order of amino acids. The amino acids are linked together through the amide group of one amino acid to the carboxyl group of the adjacent amino acid to form the protein backbone. The backbone is comprised of these stable bonds starting with an N-terminus of the amide group of the first amino acid and ending with a C-terminus of the carboxyl group of the last amino acid.

Secondary structure arises from hydrogen bonds between atoms in the protein backbone. These interactions lead to formation of regions of stability such as alpha-helices and beta sheets. Conversely, regions can lack these interactions, leaving disordered or flexible sections of the protein backbone.

The tertiary structure of proteins arises from interactions between sidechains of amino acids, in which secondary structures can be brought in proximity. Salt bridges can also be formed between amino acid sidechains of opposite charge. Proteins with multiple cysteine residues can also form disulfide bonds. These interactions work to give a protein its overall three-dimensional shape and contribute to protein structure stability and maintaining its fidelity. Lastly, quaternary structure arises from protein subunits or whole proteins interacting with one another to form protein complexes.

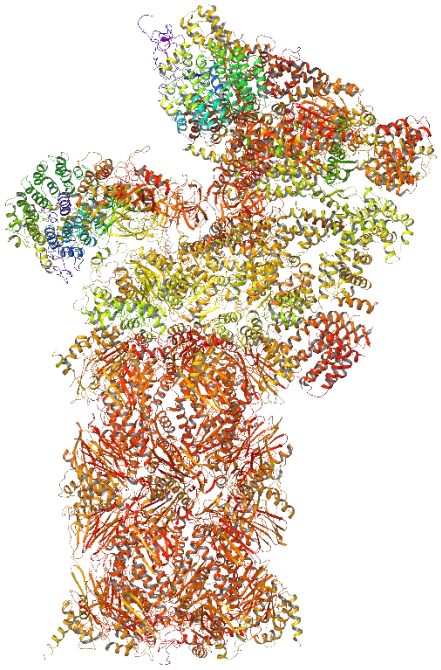


Figure 1.1 The 26S proteasome, a 2.5 megaDalton complex responsible for protein degradation. Modeled from PBD 5GJQ.

There are several analytical techniques that can capture protein structure with very high resolution, including x-ray crystallography, solid-state nuclear magnetic resonance (NMR), and cryo-electron microscopy.[2-4] These methods have provided invaluable information for many proteins, and doubtlessly new methods will continue to emerge and expand our structural databases. Although these are highly accurate methods to obtain protein structures, they do have limitations. Proteins are dynamic structures; the methods listed above rely on crystallization, which will not capture dynamic structures. Additionally, some proteins are not amenable to crystallization, or it may be prohibitively difficult to obtain enough material to form the crystal. Fortunately, other avenues of

analytical characterization may be ideal for elucidation of protein structure; particularly of interest is mass spectrometry.

Mass Spectrometry for Protein Structure

Mass spectrometry (MS) is an extremely versatile and widespread analytical technique. To measure mass to charge, samples must be ionized. This enables analysis of molecules in the gas phase. Although initially restricted to small molecule analysis, visionary researchers such as Klaus Biemann recognized the potential of MS as a protein characterization technique long before the required ionization technology was available.[5] The advent of soft ionization techniques matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) enabled MS to analyze much larger molecules.[6,7] Indeed, in the intervening years, MS has been used to characterize nucleic acids, glycans, proteins, and other biomolecules.[8-10]

There are several mechanisms for fragmenting biomolecules in a mass spectrometer. The most common ones use collisions with a neutral gas, usually helium or nitrogen. In the case of peptides and proteins, collisions with the gas will cause the molecules to heat up, and the heating will cause the protein or peptide to be fragmented at the peptide bond, located between the amide and carboxyl groups of adjacent amino acids. The N-terminal and C-terminal fragments are referred to as b- and y-ions respectively. These fragmentation methods are collision-induced dissociation (CID) or higher energy collision dissociation (HCD).

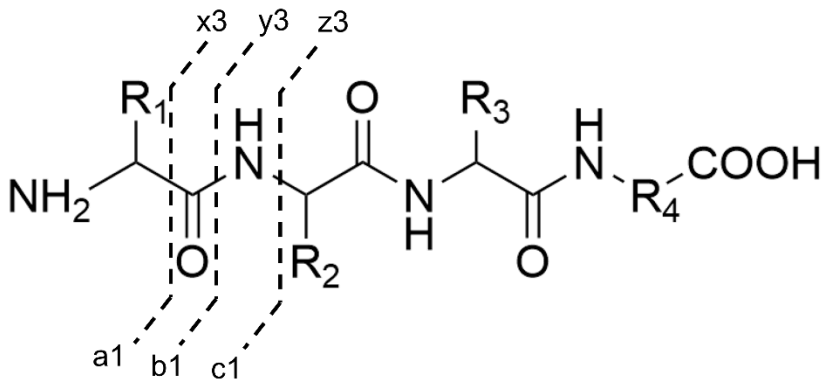


Figure 1.2 Roepstorff notation in example peptide.

Electrons can also be used to induce fragmentation in polypeptides. Electron capture dissociation (ECD) utilizes electrons that are captured by multiply charged cations, followed by subsequent neutralization of the electron, forming an excited radical which then cleaves the N-alpha carbon bond with high specificity. These fragments are referred to as c- and z-ions in the N- and C-terminal directions respectively. Disulfide bonds are also cleaved with high specificity. Electron transfer dissociation (ETD) works very similarly, but with a radical anion used to transfer the electron to the analyte.

There are several methods for acquiring structural information from MS data. Electron methods such as electron capture or transfer dissociation (EC/TD), are also excellent at protein sequencing, however for larger molecules, fragmentation is often limited to terminal regions.

Photons can also be used to fragment biomolecules for structural information. Photons can access pathways to generate fragments that neither collisional activation nor electron-based fragmentation can.[11] Infrared multi-photon dissociation (IRMPD) utilizes photons with a wavelength of 10.6 μm, typically produced continuously by a CO_2 laser. Most organic molecules will absorb at this wavelength, and as they absorb multiple photons, step-wise ion dissociation will occur with the weakest bonds dissociating first. (Brodbelt review). Photons in the ultraviolet range can also be utilized for accessing higher electronic states resulting in dissociation directly or resulting in internal conversion. The Julian lab has demonstrated the utility of 213 and 266 nm photons for selective cleavage of C-I[12] and C-S bonds[13-15], which will be discussed further subsequently. Ultraviolet photodissociation (UVPD) can be used to obtain structural information about various biomolecules including lipids, carbohydrates, and proteins by various approaches.[16]

Radicals in MS

One strategy that can yield structural information is the use of radicals with mass spectrometry. One example of this type of method is fast photochemical oxidation of proteins (FPOP). FPOP utilizes a laser to hydrolyze hydrogen peroxide in solution to generate hydroxyl radicals. The generated hydroxyl radicals then irreversibly label the protein, with rates dependent on reactivity with sidechains and solvent accessibility. FPOP is very useful for protein structure studies because it does not disrupt structure and its timescale is short relative to protein dynamics.[17]

Unlike FPOP and other methods that use radicals to label proteins, radicals can also be used to elucidate protein structure information via unique fragmentation patterns. Radical directed dissociation (RDD) relies on a radical precursor being selectively introduced on the peptide or protein of interest. Radical precursors include iodo-tyrosine, 4-iodobenzoic acid, and 4-iodophenylthiourea.[18,19,20] Once the precursor is introduced, photoactivation with a 213 or a 266 nm laser can generate a radical via homolytic cleavage of the precursor. Once there is a radical generated, additional collisional energy can be input to the protein, which causes migration of the radical. The radical will migrate to sites nearby, causing unique fragmentation. RDD fragments include c, z-1, a, x, x-H_2O, and sidechain losses. In addition to the unique fragments generated by RDD, the mechanisms that generate these fragments predispose them to occur at specific residues based on beta H-bond dissociation energy.[21] Because the radical will travel through 3D space, residues where these fragments occur must be located near the radical precursor, thus providing structural information in the gas phase. It has been demonstrated that RDD is sensitive even to subtle changes in structure, with the ability to differentiate between isomers, something that is impressive considering that there is no mass difference between isomeric forms.[22,23]

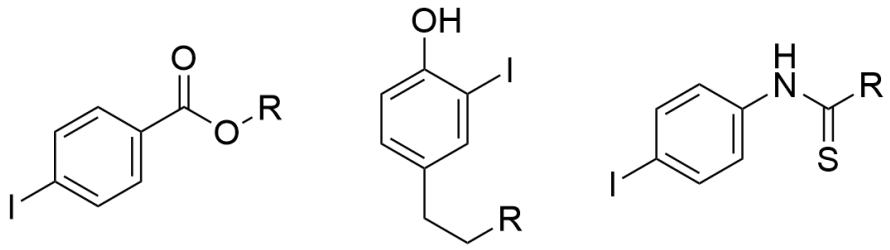
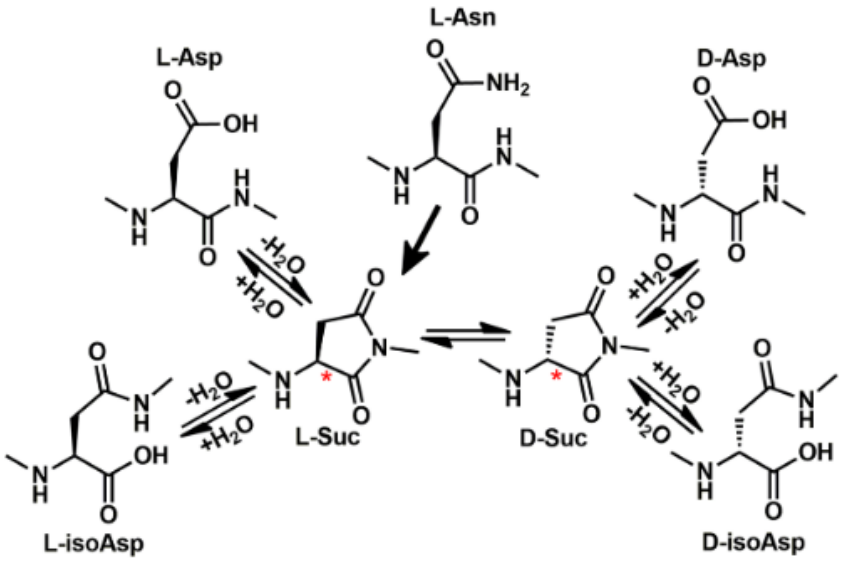


Figure 1.3 Radical precursors 4IB, iodo-tyrosine, and 4IPT from left to right.

Isomers

MS is actually well-suited for detecting PTMs. The mass shifts associated with PTMs such as phosphorylation, ubiquitination, acetylation, methylation, and others combined with fragmentation information allows PTM mapping with standard MS methods. However, biomolecules in living organisms are subject to other modifications. One class of these modifications is referred to as spontaneous chemical modifications (SCMs). These modifications include deamidation and isomerization.



Scheme 1.1 Aspartic acid isomerization mechanism

The modification occurs when the amino acid loses NH_2 or water from the side chain and forms a succinimide ring intermediate through attacking the backbone. Deamidation can occur at asparagine (Asn) and to lesser extent glutamine (Gln). Isomerization occurs most rapidly at aspartic acid (Asp) followed distantly by serine (Ser), and other amino acids at even slower rates. Once the intermediate is formed, the ring can be reopened through hydrolysis of the carbonyl group. However, the intermediate can epimerize, leaving the reopened ring in the opposite chirality. Additionally, the ring can be hydrolyzed at the wrong carbonyl group, resulting in an extra carbon in the amino acid sidechain and missing a carbon atom in the backbone. This is a total of four possible conformations resulting from isomerization. While there is mass shift of 1 Da from deamidation due to the NH_2 group being replaced by OH_2, it can still be difficult to detect with MS due to resulting in the same mass as the C^{13} isotope of the unmodified peptide. However, isomerization does not result in a mass shift, making it even more challenging to detect by MS. Although it may seem counterintuitive, MS is disputably the best method by which to detect isomer, and the Julian lab has developed multiple methodologies for doing so, which serves as a testament to the utility of MS as a tool for studying biomolecular structure.[24,25]

SCMs are differentiated from PTMs because they are not enzymatically regulated. As the name suggests, these modifications arise from chemistry in the surrounding environment. Although they are spontaneous, in most cases a significant amount of time is required for SCMs to arise. This time is longer than the lifetime of most proteins in the

human body ranging from less than an hour to 22 hours, leaving a small minority vulnerable to SCMs.[26] Long-lived proteins (LLPs) are a class of proteins loosely defined by having half-lives longer than 48 hours.[27] They are spread throughout the body, and many are associated with disease.

Even though these modifications are subtle in comparison to the structure of a whole protein, SCMs can have profound physiological consequences. Many biological processes depend on structural recognition of proteins by enzymes, chaperone proteins, and other biomolecules and interactions that are highly dependent on atom placement in space. When an amino acid is isomerized, the structure may be altered such that these interactions are weakened, or even no longer take place at all. The disruption of these interactions can have profound consequences for biological functions; however isomerization is often ignored compared to other PTMs due to the difficulties involved in detecting isomerization.

AD, Proteostasis, and Aging

Alzheimer's disease (AD) is by far the leading cause of dementia and cognitive impairment around the world, and is the fifth leading cause of death for adults over age 65. Its estimated total cost in healthcare alone in 2020 was $305 billion, and as the global population continues aging, it is project to reach over $1 trillion by midcentury.[28] Not only is AD a very economically devasting disease, it is a harrowing condition for both those who suffer as well as caregivers and loved ones. Both informal caregivers such as family members and formal caregivers such as healthcare providers experience depression and anxiety at rates higher than 30% and 40% respectively. Additionally, they

experience these mental health problems at higher levels than the general population and caregivers of patients with other disorders.[29] The chronic stress of caregiving can also manifest as physical problems such as cardiovascular disease.[30]

Despite the more than $40 billion spent in searches for AD treatment, only eight AD drugs are currently approved by the Food and Drug Administration (FDA), and only one of these is to treat the disease progression, while the rest are approved to treat symptoms only.[31] Most treatments have been focused on clearing aggregated protein. Aggregates of beta amyloid and tau protein are well known hallmarks of AD pathology. However, treatments attempting to prevent or reduce protein aggregates have shown, at best, small improvements in disease prognosis, and at worst, cause declines in patient health.[32]

Although protein aggregation is a well-known aspect of AD pathology, there are other aspects of the disease that are underexplored. Also observed in AD pathology are bodies similar to those observed in another group of disorders known as lysosomal storage disorders (LSDs).[33] These disorders are due to defects that affect lysosomal function. Lysosomes are organelles that feature a membrane, low pH, and vesicles containing enzymes. Lysosomes also contain transport systems and are able to secrete their contents.[34,35] They play important roles in autophagy, a pathway for molecular degradation, in which hydrolases contained in the lysosome degrade proteins, cytoplasmic organelles, and other macromolecules.[36] However, the pathway of autophagy is a complex biological process, and lysosomes interact with many molecules and contain a variety of constituents. This makes lysosomes subject to disorders caused by a variety

of defects. These include defects in proteins that break down glycans, such as alpha and beta mannosidosis, and aspartylglucosaminuria, known as glycoproteinoses, defects in enzymes that degrade lipids such as Niemann-Pick disease, Gaucher's disease, and Tay-Sachs disease, known as sphingolipdoses, and defects in lysosomal transport and other lysosomal proteins.[37]

In contrast to AD, most LSDs cause severe symptoms early in life. However, as discussed, LSDs are due to defects in various enzymes and proteins in the lysosome or autophagy pathway, typically causing severe symptoms early in life.[38] AD occurs later in life, with more than 90% of patients over age 65.[39] It has been documented that there is lysosomal dysfunction in AD brains.[40,41] One possible explanation for the difference in onset between AD and LSDs is that in LSDs, the enzymes and other parts of the autophagy pathway are defective or modified, but in AD, the substrates are modified over time by SCMs, appearing only later in life when metabolism and proteostasis naturally decline enough for SCMs to begin accumulating. Previous work has demonstrated that Asp isomerization inhibits proteolysis by cathepsins, a group of enzymes that constitute a major part of lysosomal proteases.[42,43]

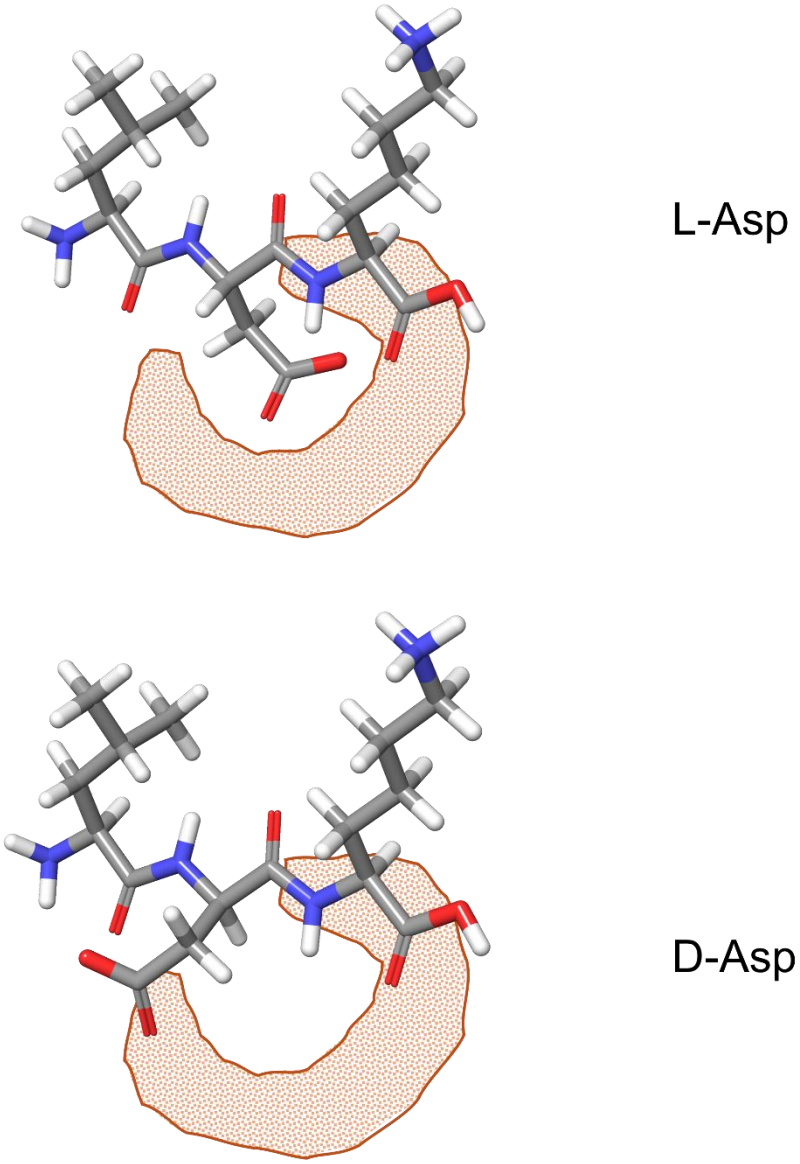


Figure 1.4 When Asp is isomerized, active sites may not be able to form necessary interactions with substrates.

Asp isomers have also been identified in relation to AD.[44] However most focus has been on amyloid-β, a protein well-known for its pathology in AD. However, tau protein is also present in irregular forms in AD. It is known to be hyperphosphorylated and aggregated in AD brains.[45] An under-investigated possibility is isomerization of tau and its possible role in lysosomal dysfunction in AD.

Internal ions

As spectrometer sensitivity and resolving power continues improving, top-down applications become more advanced. Oftentimes, top-down spectra are heavily populated, and many of the fragments cannot be assigned as terminal ions, especially in high energy dissociation methods such as UVPD and HCD. In recent years, these fragments have been attributed to internal ions, or terminal ions that have been further fragmented. Since these would two cleavages, they can be assigned as ax, ay, az, and so on, and there are several softwares available that have capabilities to assign these types of fragments.[46,47]

However, there are several issues with assigning these types of fragments. When multiple cleavages occur, there can be multiple possible fragments with the same exact masses. In these cases, even if the mass accuracy of the instrument used is very high, there is not currently a method of differentiating these types of isomers without further fragmentation. Further fragmentation has been demonstrated to definitively assign internal ions; however, this is not always practical, especially for large proteins where there may be hundreds of fragments.[48] Additionally, if two cleavages are taking place to generate an internal ion, there will be two terminal ions generated for each internal ion, thus making terminal ions more abundant than internal ions. This alone makes it

inherently more unlikely for internal fragments to be observed than terminal fragments. Another pitfall in assigning internal ions is that there is little experimental evidence to confirm that internal ions are created through the same pathways as terminal ions, yet it has been taken as a foregone conclusion by some that these are the ion types of internal fragments. In reality, the functional groups present after cleavages resulting in terminal ions differ from those present in intact peptides and proteins and may undergo different pathways upon further activation. In Chapter 5 we discuss the validity of fragments assigned as internal ions and propose possible alternatives to these assignments.

Conclusions

Proteins are vital biomolecules, and their function, or dysfunction, is dictated by their structure. Due to the diverse nature of proteins, studying their structure in various forms requires a diverse suite of tools. We are interested in utilizing mass spectrometric methods to interrogate protein and peptide structure and properties. This dissertation describes different methods for obtaining structural information from mass spectrometry data, with the application of examining protein modifications pertinent to aging and neurodegeneration discussed in Chapter 2. Methods for discriminating peptide isomers in the gas phase by examining chromophore absorption and ion-molecule reactions will be discussed in Chapter 3. Chapter 4 will explore protein structural data in the gas phase using RDD. Lastly, Chapter 5 will discuss the validity of assigning non-terminal fragments generated from intact proteins as internal ions, and possible alternative identities of these fragments.

References

[1] Delgado, F. F.; Cermak, N.; Hecht, V. C.; Son, S.; Li, Y.; Knudsen, S. M.; Olcum, S.; Higgins, J. M.; Chen, J.; Grover, W. H.; Manalis, S. R. Intracellular Water Exchange for Measuring the Dry Mass, Water Mass and Changes in Chemical Composition of Living Cells. *PLOS ONE* **2013**, *8* (7), e67590. https://doi.org/10.1371/journal.pone.0067590.

[2] Ilari, A.; Savino, C. Protein Structure Determination by X-Ray Crystallography. In *Bioinformatics: Data, Sequence Analysis and Evolution*; Keith, J. M., Ed.; Humana Press: Totowa, NJ, 2008; pp 63–87. https://doi.org/10.1007/978-1-60327-159-2_3.

[3] Kay, L. E. NMR Studies of Protein Structure and Dynamics. *Journal of Magnetic Resonance* **2011**, *213* (2), 477–491. https://doi.org/10.1016/j.jmr.2011.09.009.

[4] Bonomi, M.; Vendruscolo, M. Determination of Protein Structural Ensembles Using Cryo-Electron Microscopy. *Current Opinion in Structural Biology* **2019**, *56*, 37–45. https://doi.org/10.1016/j.sbi.2018.10.006.

[5] Biemann, K. The Coming of Age of Mass Spectrometry in Peptide and Protein Chemistry. *Protein Science* **1995**, *4* (9), 1920–1927. https://doi.org/10.1002/pro.5560040926.

[6] Hillenkamp, F.; Karas, M. Matrix-Assisted Laser Desorption/Ionisation, an Experience.*International Journal of Mass Spectrometry* **2000**, *200* (1), 71–77. https://doi.org/10.1016/S1387-3806(00)00300-6.

[7] Whitehouse, C. M.; Dreyer, R. N.; Yamashita, Masamichi.; Fenn, J. B. Electrospray Interface for Liquid Chromatographs and Mass Spectrometers. *Anal. Chem.* **1985**, *57* (3), 675–679. https://doi.org/10.1021/ac00280a023.

[8] Banoub, J. H.; Newton, R. P.; Esmans, E.; Ewing, D. F.; Mackenzie, G. Recent Developments in Mass Spectrometry for the Characterization of Nucleosides, Nucleotides, Oligonucleotides, and Nucleic Acids. *Chem. Rev.* **2005**, *105* (5), 1869–1916. https://doi.org/10.1021/cr030040w.

[9] Ruhaak, L. R.; Xu, G.; Li, Q.; Goonatilleke, E.; Lebrilla, C. B. Mass Spectrometry Approaches to Glycomic and Glycoproteomic Analyses. *Chem. Rev.* **2018**, *118* (17), 7886–7930. https://doi.org/10.1021/acs.chemrev.7b00732.

[10] *Resolution of protein structure by mass spectrometry - Vandermarliere - 2016 - Mass Spectrometry Reviews - Wiley Online Library*. https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/full/10.1002/mas.21450 (accessed 2024-04-22).

[11] Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chemical Society Reviews* **2014**, *43* (8), 2757–2783. https://doi.org/10.1039/C3CS60444F.

[12] Ly, T.; Julian, R. R. Elucidating the Tertiary Structure of Protein Ions in Vacuo with Site Specific Photoinitiated Radical Reactions. *J. Am. Chem. Soc.* **2010**, *132* (25), 8602–8609. https://doi.org/10.1021/ja910665d.

[13] Diedrich, J. K.; Julian, R. R. Site-Selective Fragmentation of Peptides and Proteins at Quinone-Modified Cysteine Residues Investigated by ESI-MS. *Anal. Chem.* **2010**, *82* (10), 4006–4014. https://doi.org/10.1021/ac902786q.

[14] Talbert, L. E.; Julian, R. R. Methionine and Selenomethionine as Energy Transfer Acceptors for Biomolecular Structure Elucidation in the Gas Phase. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (9), 1601–1608. https://doi.org/10.1007/s13361-019-02262-y.

[15] Bonner, J.; E. Talbert, L.; Akkawi, N.; R. Julian, R. Simplified Identification of Disulfide, Trisulfide, and Thioether Pairs with 213 Nm UVPD. *Analyst* **2018**, *143* (21), 5176–5184. https://doi.org/10.1039/C8AN01582A.

[16] Brodbelt, J. S.; Morrison, L. J.; Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem. Rev.* **2020**, *120* (7), 3328–3380. https://doi.org/10.1021/acs.chemrev.9b00440.

[17] Li, K. S.; Shi, L.; Gross, M. L. Mass Spectrometry-Based Fast Photochemical Oxidation of Proteins (FPOP) for Higher Order Structure Characterization. *Acc. Chem. Res.* **2018**, *51* (3), 736–744. https://doi.org/10.1021/acs.accounts.7b00593.

[18] Ly, T.; Julian, R. R. Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130* (1), 351–358. https://doi.org/10.1021/ja076535a.

[19] Lyon, Y. A.; Sabbah, G. M.; Julian, R. R. Differences in α-Crystallin Isomerization Reveal the Activity of Protein Isoaspartyl Methyltransferase (PIMT) in the Nucleus and Cortex of Human Lenses. *Experimental Eye Research* **2018**, *171*, 131–141. https://doi.org/10.1016/j.exer.2018.03.018.

[20] Lambeth, T. R.; Julian, R. R. Efficient Isothiocyanate Modification of Peptides Facilitates Structural Analysis by Radical-Directed Dissociation. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (8), 1338–1345. https://doi.org/10.1021/jasms.1c00237.

[21] Sun, Q.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R. Side Chain Chemistry Mediates Backbone Fragmentation in Hydrogen Deficient Peptide Radicals. *J. Proteome Res.* **2009**, *8* (2), 958–966. https://doi.org/10.1021/pr800592t.

[22] Lyon, Y. A.; Sabbah, G. M.; Julian, R. R. Differences in α-Crystallin Isomerization Reveal the Activity of Protein Isoaspartyl Methyltransferase (PIMT) in the Nucleus and Cortex of Human Lenses. *Experimental Eye Research* **2018**, *171*, 131–141. https://doi.org/10.1016/j.exer.2018.03.018.

[23] Tao, Y.; Julian, R. R. Identification of Amino Acid Epimerization and Isomerization in Crystallin Proteins by Tandem LC-MS. *Anal. Chem.* **2014**, *86* (19), 9733–9741. https://doi.org/10.1021/ac502296c.

[24] Wu, H.-T.; Riggs, D. L.; Lyon, Y. A.; Julian, R. R. Statistical Framework for Identifying Differences in Similar Mass Spectra: Expanding Possibilities for Isomer Identification. *Anal. Chem.* **2023**, *95* (17), 6996–7005. https://doi.org/10.1021/acs.analchem.3c00495.

[25] Wu, H.-T.; Van Orman, B. L.; Julian, R. R. Localizing Isomerized Residue Sites in Peptides with Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2024**, *35* (4), 705–713. https://doi.org/10.1021/jasms.3c00373.

[26] Eden, E.; Geva-Zatorsky, N.; Issaeva, I.; Cohen, A.; Dekel, E.; Danon, T.; Cohen, L.; Mayo, A.; Alon, U. Proteome Half-Life Dynamics in Living Human Cells. *Science* **2011**, *331* (6018), 764–768. https://doi.org/10.1126/science.1199784.

[27] Truscott, R. J. W.; Schey, K. L.; Friedrich, M. G. Old Proteins in Man: A Field in Its Infancy. *Trends in Biochemical Sciences* **2016**, *41* (8), 654–664. https://doi.org/10.1016/j.tibs.2016.06.004.

[28] Wong, W. Economic Burden of Alzheimer Disease and Managed Care Considerations. *Am J Manag Care* **2020**, *26* (8 Suppl), S177–S183. https://doi.org/10.37765/ajmc.2020.88482.

[29] Sallim, A. B.; Sayampanathan, A. A.; Cuttilan, A.; Chun-Man Ho, R. Prevalence of Mental Health Disorders Among Caregivers of Patients With Alzheimer Disease. *Journal of the American Medical Directors Association* **2015**, *16* (12), 1034–1041. https://doi.org/10.1016/j.jamda.2015.09.007.

[30] Mausbach, B. T.; Patterson, T. L.; Rabinowitz, Y. G.; Grant, I.; Schulz, R. Depression and Distress Predict Time to Cardiovascular Disease in Dementia Caregivers. *Health Psychology* **2007**, *26* (5), 539–544. https://doi.org/10.1037/0278-6133.26.5.539.

[31] FDA-Approved Treatments for Alzheimer's. **2024**.

[32] Vaz, M.; Silvestre, S. Alzheimer's Disease: Recent Treatment Strategies. *European Journal of Pharmacology* **2020**, *887*, 173554. https://doi.org/10.1016/j.ejphar.2020.173554.

[33] Chung, K. M.; Hernández, N.; Sproul, A. A.; Yu, W. H. Alzheimer's Disease and the Autophagic-Lysosomal System. *Neuroscience Letters* **2019**, *697*, 49–58. https://doi.org/10.1016/j.neulet.2018.05.017.

[34] Arai, K.; Shimaya, A.; Hiratani, N.; Ohkuma, S. Purification and Characterization of Lysosomal H(+)-ATPase. An Anion-Sensitive v-Type H(+)-ATPase from Rat Liver Lysosomes. *Journal of Biological Chemistry* **1993**, *268* (8), 5649–5660. https://doi.org/10.1016/S0021-9258(18)53369-X.

[35] Luzio, J. P.; Rous, B. A.; Bright, N. A.; Pryor, P. R.; Mullock, B. M.; Piper, R. C. Lysosome-Endosome Fusion and Lysosome Biogenesis. *Journal of Cell Science* **2000**, *113* (9), 1515–1524. https://doi.org/10.1242/jcs.113.9.1515.

[36]Parzych, K. R.; Klionsky, D. J. An Overview of Autophagy: Morphology, Mechanism, and Regulation. *Antioxid Redox Signal* **2014**, *20* (3), 460–473. https://doi.org/10.1089/ars.2013.5371.

[37] Vellodi, A. Lysosomal Storage Disorders. *British Journal of Haematology* **2005**, *128* (4), 413–431. https://doi.org/10.1111/j.1365-2141.2004.05293.x.

[38] Staretz-Chacham, O.; Choi, J. H.; Wakabayashi, K.; Lopez, G.; Sidransky, E. Psychiatric and Behavioral Manifestations of Lysosomal Storage Disorders. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **2010**, *153B* (7), 1253–1265. https://doi.org/10.1002/ajmg.b.31097.

[39] Association, A. 2015 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia* **2015**, *11* (3), 332–384. https://doi.org/10.1016/j.jalz.2015.02.003.

[40] Whyte, L. S.; Lau, A. A.; Hemsley, K. M.; Hopwood, J. J.; Sargeant, T. J. Endo-Lysosomal and Autophagic Dysfunction: A Driving Factor in Alzheimer's Disease? *Journal of Neurochemistry* **2017**, *140* (5), 703–717. https://doi.org/10.1111/jnc.13935.

[41] Cataldo, A. M.; Petanceska, S.; Terio, N. B.; Peterhoff, C. M.; Durham, R.; Mercken, M.; Mehta, P. D.; Buxbaum, J.; Haroutunian, V.; Nixon, R. A. Aβ Localization in Abnormal Endosomes: Association with Earliest Aβ Elevations in AD and Down Syndrome. *Neurobiology of Aging* **2004**, *25* (10), 1263–1272. https://doi.org/10.1016/j.neurobiolaging.2004.02.027.

[42] Pandey, G.; Julian, R. R. LC–MS Reveals Isomeric Inhibition of Proteolysis by Lysosomal Cathepsins. *Analysis & Sensing* **2022**, *2* (4), e202200017. https://doi.org/10.1002/anse.202200017.

[43] Lambeth, T. R.; Riggs, D. L.; Talbert, L. E.; Tang, J.; Coburn, E.; Kang, A. S.; Noll, J.; Augello, C.; Ford, B. D.; Julian, R. R. Spontaneous Isomerization of Long-Lived Proteins Provides a Molecular Mechanism for the Lysosomal Failure Observed in Alzheimer's Disease. *ACS Cent. Sci.* **2019**, *5* (8), 1387–1395. https://doi.org/10.1021/acscentsci.9b00369.

[44] Mukherjee, S.; Perez, K. A.; Lago, L. C.; Klatt, S.; McLean, C. A.; Birchall, I. E.; Barnham, K. J.; Masters, C. L.; Roberts, B. R. Quantification of N-Terminal Amyloid-β Isoforms Reveals Isomers Are the Most Abundant Form of the Amyloid-β Peptide in Sporadic Alzheimer's Disease. *Brain Communications* **2021**, *3* (2), fcab028. https://doi.org/10.1093/braincomms/fcab028.

[45] Alonso, A. del C.; Grundke-Iqbal, I.; Iqbal, K. Alzheimer's Disease Hyperphosphorylated Tau Sequesters Normal Tau into Tangles of Filaments and Disassembles Microtubules. *Nat Med* **1996**, *2* (7), 783–787. https://doi.org/10.1038/nm0796-783.

[46] 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. https://doi.org/10.1021/acs.jproteome.0c00952.

[47] Zhu, Y.; Liu, Z.; Liu, J.; Zhao, H.; Feng, R.; Shu, K.; Wang, F.; Chang, C. Panda-UV Unlocks Deeper Protein Characterization with Internal Fragments in Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem*. **2024**, *96* (21), 8474–8483. https://doi.org/10.1021/acs.analchem.4c00253.

[48] Schmitt, N. D.; Berger, J. M.; Conway, J. B.; Agar, J. N. Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem*. **2021**, *93* (16), 6355–6362. https://doi.org/10.1021/acs.analchem.0c04670.

Chapter 2

Locating Tau Isomerization *In Vitro* and in Tissue Samples

Introduction

Microtubule associated protein tau is known to be associated with neurodegenerative disorders, including Alzheimer's disease (AD). In AD pathology, aggregated tau is observed as neurofibrillary tangles (NFTs).[1-3] Tau is an intrinsically disordered protein (IDP), meaning that it lacks defined secondary structures.[4,5] This lack of structure leaves tau protein more vulnerable to misfolding and aggregation.[6] While associated with AD, it is unclear what role NFTs play in contributing to the disease.

Isomerization

The attributes that leave tau susceptible to aggregation also leave it prone to other modifications. Spontaneous chemical modifications (SCMs) are modifications that are not enzymatically regulated; they occur when chemical conditions are sufficient. These include deamidation and isomerization.[7] A flexible protein backbone, such as the ones disordered regions feature, also facilitates SCMs. One of the most important requisites for SCMs is time.[8] The vast majority of the proteome is turned over at a rate that prevents these modifications from occurring.[9] Even in a protein with a structural predisposition for SCMs it can take days, or even weeks, for modifications to accumulate to a detectable amount.

A study by our group recently found that isomerization of Asp387 in tau tracked cognitive decline more accurately than aggregation in AD brain samples.[10] Strikingly, it

differentiated control, sporadic, and autosomal dominant cohorts, even in control high groups which featured aggregation on par with AD pathology. This suggests that isomerization may be a better indicator of AD disease state and may even be contributing to AD progression. The required amount of time for the amount of isomerization observed in the study above suggest a decline in autophagic flux; proteins are not turned over, allowing SCMs to take place. Additionally, our group has demonstrated isomerization can inhibit proteolysis; the cathepsins, a group of enzymes present in the lysosome responsible for protein degradation are unable to cut amino acids at or surrounding an isomerized residue.[11,12] Due to these findings, we hypothesize that once modifications begin occur due to slowing autophagic flux, proteolysis is inhibited and even more modifications occur as the isomerized protein is unable to be degraded.

While protein sequence, structure, and half-life can all provide some predictive power, with molecules as large and complex as whole proteins, it is difficult to determine a rate of modification without experimental data. Additionally, isomerization rates across sites in a single protein can vary greatly. To further complicate rate determination, there is some isomer repair present *in vivo*; protein L-isoaspartyl methyltransferase (PIMT) is an enzyme that repairs the L-isoAsp form of Asp by transferring a methyl group to the succinimide intermediate, resulting in converting some of the isomerized residue to the canonical L-Asp form.[13] While the repair pathway appears inefficient and perhaps insignificant, mice lacking PIMT suffer devastating neurological damage and experience fatal epilepsy within 60 days.[14,15] To further complicate matters, the efficiency of PIMT varies due to genetics.[16] All of these factors make it difficult to determine the rate of tau

isomerization from data from tissue samples. Because of these complications, we set out to determine the rate of tau isomerization *in vitro*. Two conditions were tested using recombinant protein isoform tau-F or 2N4R, the longest tau isoform, to ensure modifications would not be missed in a shorter isoform. Aggregated and non-aggregated were tested; due to its presence in AD, the consequences of aggregation, if any, on isomerization are of importance.

Detection in Tissue

Isomerization has no detectable mass shift, making it difficult to detect using MS alone, although there are several ways to differentiate spectra from different isomers with MS.[17,18] Fortunately, liquid chromatography can often resolve peptide isomers with identical masses, and isomers can be detected in a routine proteomics workflow without significant modification. In fact, the previous study where the AD isomers were found, was not conducted with these modifications in mind.

The previous study utilized data independent acquisition (DIA) coupled with LC. DIA is well suited for global isomer searches; continuous simultaneous fragmentation of a range of m/z's greatly expands the number of data points for a peptide across a chromatographic peak compared to data dependent acquisition methods, where only one scan may be acquired per peak in a complex mixture, or even missed entirely. However, DIA comes with its own set of challenges and limitations. Because there are no specified precursors selected, it is virtually impossible to know which peptides fragment ions are generated from. To make use of DIA data, a lot of processing and additional software, time investment, and data expertise is necessary.

In contrast to global isomer searches, if a known isomerization target is of interest, DDA or targeted methods are very useful. If the abundance of the target is sufficient, a DDA method with a mass list and appropriate dynamic exclusion parameters can be used for isomer search and quantification. Parallel reaction monitoring (PRM) methods can also be very useful. These methods may also be useful for locating modifications such as phosphorylation, which is abundant and widely varied in neurodegenerative diseases, and difficult to map using DIA. In this study, we used multiple acquisition methods to map tau isomerization in human brain tissue and determine a method to detect isomers in tissue samples.

Experimental

In Vitro aging

Full-length recombinant 2N4R tau protein was purchased from R & D Biosystems. The non-aggregated samples were aliquoted in 1X PBS buffer, pH 7.4. Each aliquot was 50 uL with a concentration of 4.35 uM. Aggregated samples were prepared using heparin sodium salt. 10 uM protein in 1X PBS, pH 7.4 was combined with 1 mM DTT, 2 mM MgCl2, and approximately 10 uM heparin sodium salt. Protein samples were shaken for 48 hours at 37°C to induce aggregation.[19]

Samples were placed in 500 uL Eppendorf tubes and placed inside glass vials with 1 mL of water to prevent solvent evaporation. Samples were incubated in a sand bath at 37°C. Timepoints were taken at 0, 3, 6, 9, 12, and 16 weeks. After collection of the timepoints, they were incubated with 50 uL of 70% formic acid at room temperature for two hours in order to monomerize any aggregated protein. All solvent was then removed

by lyophilization, and samples were resuspended in 10 uL of 50 mM tris-HCl pH 8.0 with 6 M urea.

Digest

Cysteines were reduced and alkylated by adding 3 uL of 10 mM dithiothreitol (Sigma) and incubated 20 minutes at room temperature followed by adding 15 uL of 20 mM iodoacetamide (Sigma) and incubated at room temperature in the dark for 1 hour. Following cysteine alkylation, 100 uL of 50 mM tris, pH 7.6 was added (urea concentration ≤0.6 M). The pH was checked, and if necessary adjusted with 1 M tris-HCl, pH 9. Trypsin (Promega) was added in a 1:20 enzyme:protein ratio, and incubated for 18 hours at 37°C. Samples were then filtered and 100 uL of optima water with 0.1% formic acid was added. Aggregated samples were also run through a 3kDa molecular weight cutoff filter (Millipore Sigma) to retain any large irreversible aggregates.

LC-MS/MS Analysis

Samples were analyzed using an Ultimate 3000 (Thermo) coupled to an Orbitrap Fusion Lumos (Thermo). LC was done using a 25 cm 100 um column packed in house with 1.9 um C18-AQ resin (Dr. Maisch). Mobile phase A was Optima water with 0.1% formic acid, and mobile phase B was 80% acetonitrile, 20% Optima water, and 0.1% formic acid. The gradient was 1-7% B in 7 minutes, 7-15% B in 20 minutes, 15-50% B in 40 minutes, and 50-98% B in 5 minutes.

A parallel reaction monitoring (PRM) method was used with 60 targets. The MS1 resolution was 120,000 with a scan range of 350-1500 m/z. MS2 used HCD with NCE 33, an isolation window of 1.6, resolution of 15,000, and scan range of 350-2000 m/z.

Isomer Detection in Tissue

Homogenization

Tissue homogenization was performed using a Dounce homogenizer alone, a Dounce homogenizer followed by sonication with a probe sonicator, and Biomashers (Omni International). Tissue was suspended in 10x by mass of lysis buffer, consisting of 10 mM Tris-HCl, pH 7.4, 800 mM NaCl, 1 mM EDTA, 2mM DTT, and 10% sucrose. When the Dounce homogenizer was used, 0.5 to 1.0 grams of tissue was used for the extraction. When the Biomashers were used 100 to 200 mg of tissue were used for the initial homogenization. The three homogenization methods yielded roughly equal results, therefore all methods are sufficient and should be selected by the desired amount of starting material.

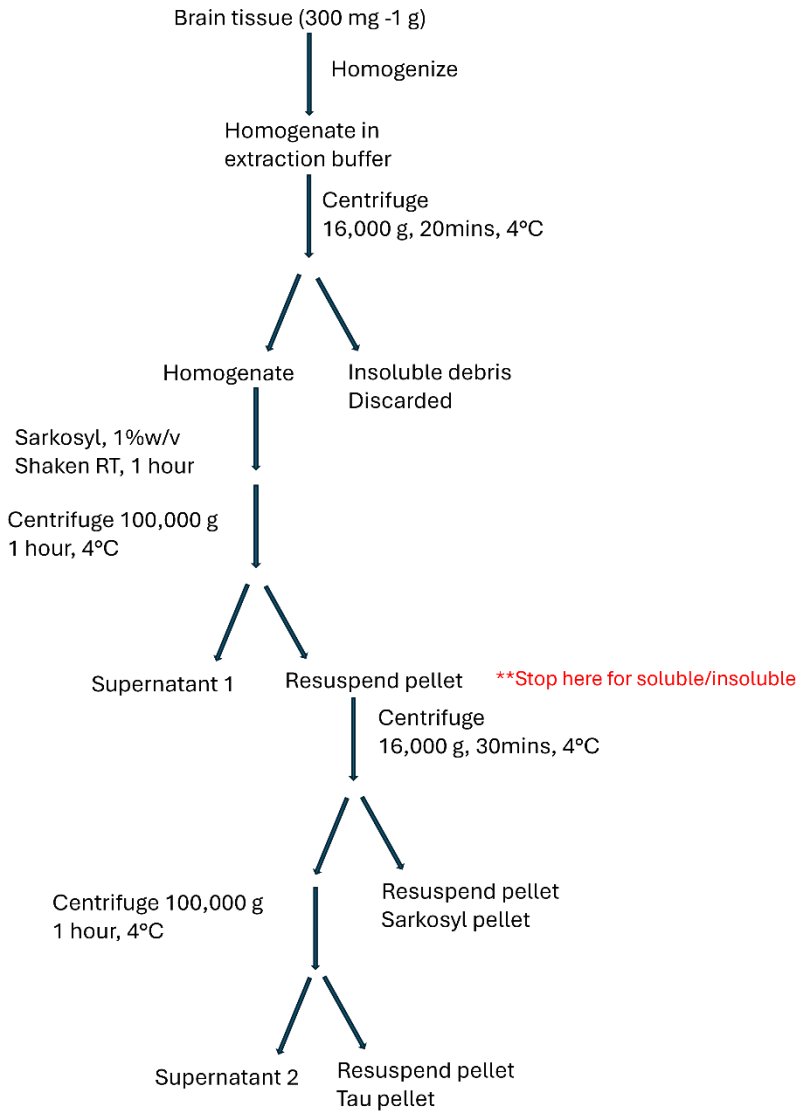
Fractionation

Following homogenization, the homogenate was centrifuged at 16,000 rpm at 4°C for 20 minutes to remove any large, insoluble debris. After the first homogenization, the supernatant was removed, and the pellet containing large debris was discarded.

Following initial fractionation, sarkosyl was added to 1% w/v to the supernatant and shaken at room temperature for one hour. Next, the supernatant was centrifuged at 100,000g and 4°C for 1 hour. The pellet was resuspended in 300 uL of tris buffer. If only soluble and insoluble fractions are desired, fractionation is complete at this step. If further fractionation is desired, the pellet can be resuspended and centrifuged at 16,000 g for 30 minutes at 4°C. The pellet can be resuspended, referred to as the sarkosyl pellet for this

manuscript. The supernatant can be centrifuged at 100,000 g at 4°C for another hour, resulting in two more fractions referred to as supernatant two and the tau pellet.

Scheme 2.1 Flowchart of tissue sample fractionation.



Data Acquisition

The appropriate data acquisition method is dependent on type of isomer search. For discovery isomer searches across the proteome, data independent acquisition (DIA) is useful. DIA eliminates the need to optimize dynamic exclusion times and minimizes isomer peaks that may be missed due to dynamic exclusion. The nature of DIA experiments enables a more comprehensive search through fragmenting everything in a defined m/z range, which allows the analysis of lower abundance peptides, and more confident assignments through continuous acquisition of fragment ions. This monitoring provides assignments of fragment ions throughout a chromatographic peak, which enables confident assignment of isomeric peaks.

Although DIA experiments are well suited for finding a large number of isomers across a proteome, they also intrinsically require more data processing, and care must be taken that the processing is appropriate. DIA also typically provides more information to filter through than other acquisition methods.[20] Lastly, DIA may not be suited for locating highly modified peptides. Peptides that are highly phosphorylated or deamidated are frequently missed or represented incorrectly by DIA or DIA processing software.

Once isomers have been identified, other acquisition methods may be suitable for isomer detection and quantification. Data dependent acquisition (DDA) can be used to detect and quantify isomers. Using a mass list is advised, and in some cases may be necessary to detect lower abundance targets. However, even with the use of a mass list, dynamic exclusion may lead to isomeric peaks being excluded from acquisition. Even in

cases where all isomeric peaks are identified, the MS2 scan may be triggered at any point across the chromatographic peak, making quantification more difficult.

An alternative to using DDA is using a targeted method such as parallel reaction monitoring (PRM). Relevant precursors should be identified by DIA, and the continuous monitoring for targets allows for detection and quantification even of low abundance targets. One drawback of PRM acquisition is a limited number of targets possible per experiment. Once isomers of interest were identified in DIA experiments, PRM was used to detect the same targets in other tissue samples.

Data analysis

Data was analyzed using Skyline. Variable modifications included in the search were deamidation of asparagine and glutamine, oxidation of methionine, and phosphorylation of serine, threonine, and tyrosine. Two missed cleavages were allowed. The search engine used was MS Amanda. The MS1 tolerance was 10 ppm, and the MS2 tolerance was 5 ppm. Two max variable modifications were allowed. For DIA spectra, the library was built in Skyline. Percent isomerization was calculated by extracting the peak areas from Skyline and dividing the total peak areas of the isomer peaks by the total peak area (including L-Asp).

Results and Discussion

In Vitro aging

The two conditions of recombinant protein featured different rates of isomerization of the TDHG peptide. Fig. 1 shows the percent isomerization at each timepoint. The soluble protein was modified at a faster rate, with half of the protein isomerized by 16 weeks.

The soluble protein also displayed a consistent trend, with the protein getting progressively more modified at each timepoint. The aggregated protein was modified at slower rate, with only approximately twenty-five percent modified after 16 weeks. Additionally, the aggregated protein displayed a less consistent trend than the soluble protein.

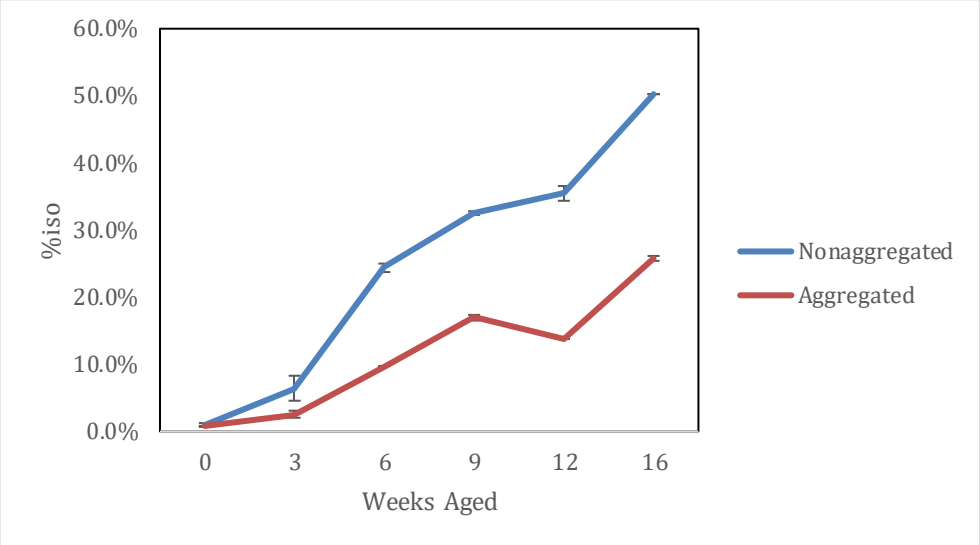


Figure 2.1 Progression of isomerization of Asp387 in recombinant tau.

These results are not surprising, due to the fact that aggregated protein does not form uniform structure, which could vary the amount of TDHGAEIVYK that is solvent exposed, slowing down isomerization in instances where it is not solvent accessible. It has been found that the 2N4R isoform of tau will form at least four different morphologies when aggregated in the presence of heparin and can interchange between them.[21] These results suggest aggregated tau observed in AD pathology may continue to

isomerize after aggregation, but highly isomerized protein more likely isomerized in a soluble form and then proceeded to aggregate.

Isomer Detection in Tissue

Initial experiments on unfractionated samples showed isomers across the tau sequence. Both DIA and DDA data was acquired and showed similar results. Isolation of tau using monoclonal antibodies was attempted, but it resulted in no protein being detected after the pulldown, which was confirmed via MS and Bradford assay. Initial results from the unfractionated sample that was not treated with antibodies showed isomerization in the TDH peptide, as well as sites of oxidation, deamidation, and phosphorylation throughout the sequence. The TDH isomerization present in this sample was 52%. The distribution of isomer peaks is shown in Fig. 2. The elution order of the isomers was previously determined using synthetic standards. As expected, the L-isoAsp form is the most prevalent, followed by D-isoAsp and D-Asp respectively.[22] Although the range of isomerization in the hippocampus previously reported (Evan) varied significantly, this value is consistent with an autosomal dominant hippocampus sample. The hippocampus is one of the earliest and most severely affected regions of the brain by AD.[23] If isomerization of Asp 387 is representative of AD pathology it would be expected that isomerization is higher in brain regions most affected by disease progression. These results provided preliminary results pointing us to sites of interest and led us to make modifications to the sample preparation.

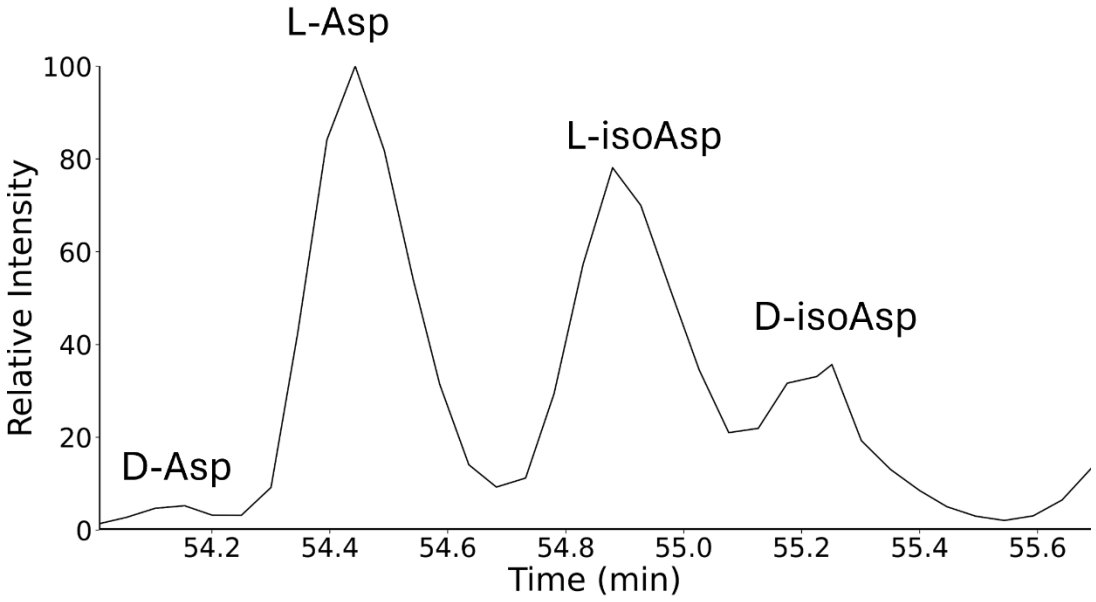


Figure 2.2 Isomer distribution in hippocampal sample TDH peptide.

Following the results with the pulldown, rather than try to isolate tau using antibodies, it was decided to concentrate tau using fractionation. Ultracentrifugation has been used previously to concentrate tau aggregates.[24] Two different fractionation methods were used, both utilizing sarkosyl and ultracentrifugation. The first resulted in four fractions to isolate tau from the rest of the sarkosyl insoluble proteins. However, it was determined that the concentration of tau was sufficient to detect with the rest of the insoluble proteins, and the concentration of the tau fraction was low and thus more difficult to detect.

Table 2.1 Concentrations of fractions from initial fractionation protocol. For the rest of the samples analyzed, fractionation was stopped after separation into two fractions, sarkosyl soluble and insoluble.

Fraction	Concentration
Supernatant 1	8.2 ug/uL
Supernatant 2	12 ug/uL
Sarkosyl pellet	1.6 ug/uL
Tau pellet	0.6 ug/uL

The next sample analyzed was an AD frontal cortex sample, split into four fractions. Both the sarkosyl and tau fractions featured Asp387 isomerization. Although the sarkosyl fraction had more total protein, the tau pellet featured more tau. The isomer distribution was also different between the two fractions. The sarkosyl fraction Asp387 was 50% isomerized, and the tau pellet was 61% isomerized. This difference could be attributed to isomerization leading to aggregation. The *in vitro* results suggest that once the protein is aggregated, isomerization is modulated, so it is possible the protein in the tau pellet was isomerized in the soluble form and later aggregated.

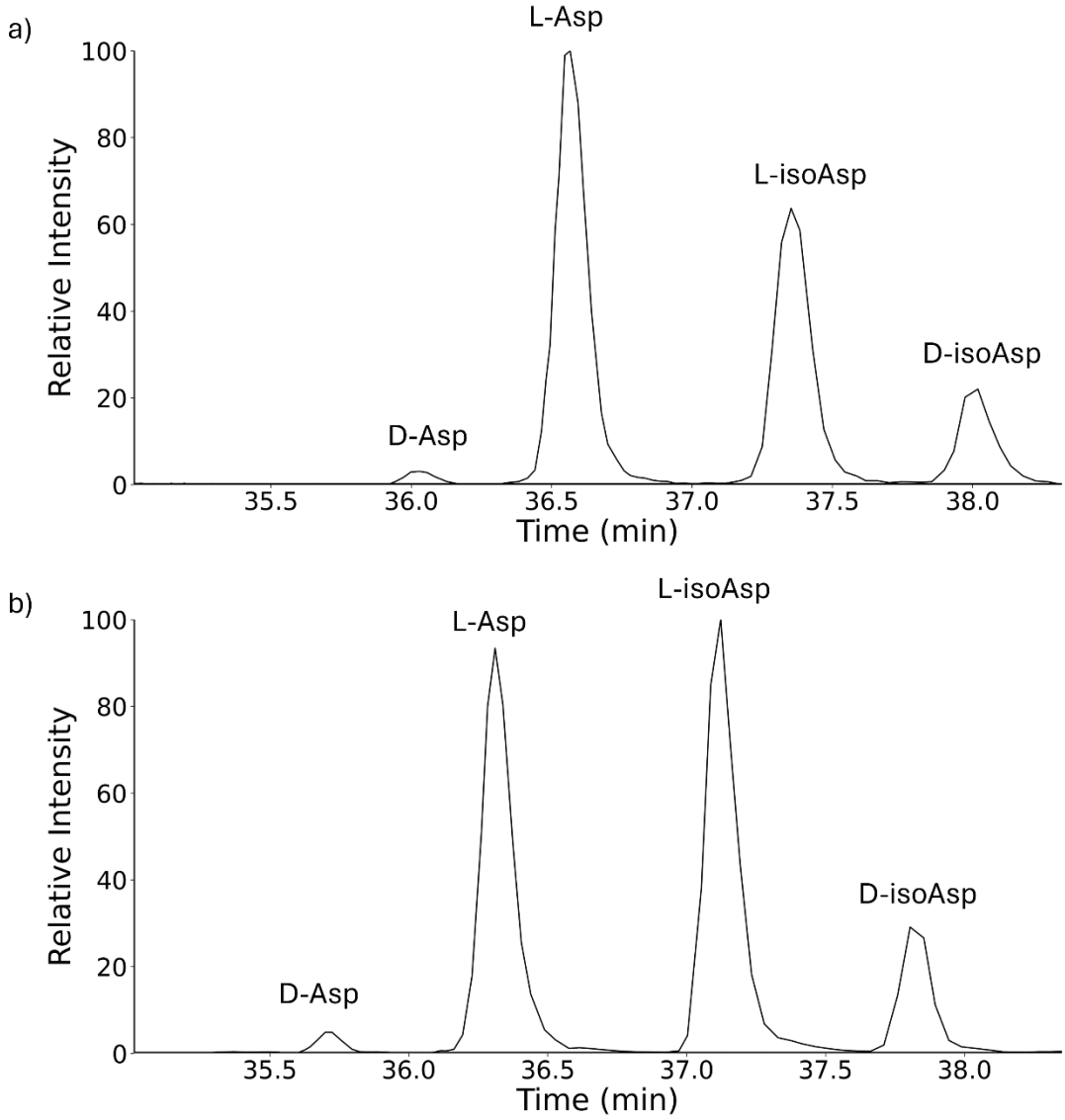


Figure 2.3 Asp387 in a frontal cortex sample. a) Asp387 isomer distribution in the sarkosyl pellet. b) Asp387 isomer distribution in the tau pellet. Tau isomerization is more advanced in the tau pellet.

Other proteins with modifications were identified as well, such as isomerization sites in superoxide dismutase (SOD1), a protein aggregated in ALS.[25] It has been found that SOD activity is reduced by over 50% in individuals with even mild cognitive impairment. There was also a strong correlation found between Mini Mental State Examination (MMSE) scores and reduced SOD activity.[26] The reduction of SOD activity leads to increased susceptibility to oxidative stress in the body. Isomerization of the protein could be a contributing factor to reduced activity. Another protein with isomerization sites was heat shock protein HSP-90. Some proteins involved with cell structure were also modified, such as ezrin, tropomyosin, and alpha-internexin.

Isomerization in frontal cortex has not been reported yet. The frontal cortex is another brain region that is affected in early AD stages.[27] It has also been found that the amount of synapse loss and other changes in the frontal cortex are correlated with cognitive decline. The high amounts of isomerization found in the frontal cortex sample examined suggest an advanced case of AD based on the results of the MacCoss study in the hippocampus and SMTG, but further investigation will be required to establish a correlation between amounts of isomerization in the frontal cortex and disease progression.

The results of the DDA and DIA experiments on the frontal cortex and hippocampus samples located modifications that were then used to compile a mass list for targeted data acquisition. The identified modifications across the tau protein sequence are shown in Fig. 4. A PRM mass list containing unmodified targets, and phosphorylation, oxidation, deamidation, and isomerization sites in tau protein, as well as isomerization sites in SOD,

and an amyloid beta peptide was compiled, for a mass list containing 116 targets. This number of targets when run on an Orbitrap Fusion Lumos should be sufficient to allow approximately 5 points per peptide across a 30 second chromatographic peak.[28] This number of data points should be sufficient for confident isomer identification, with the benefit of high reproducibility and quantification capabilities of PRM methods.[29]

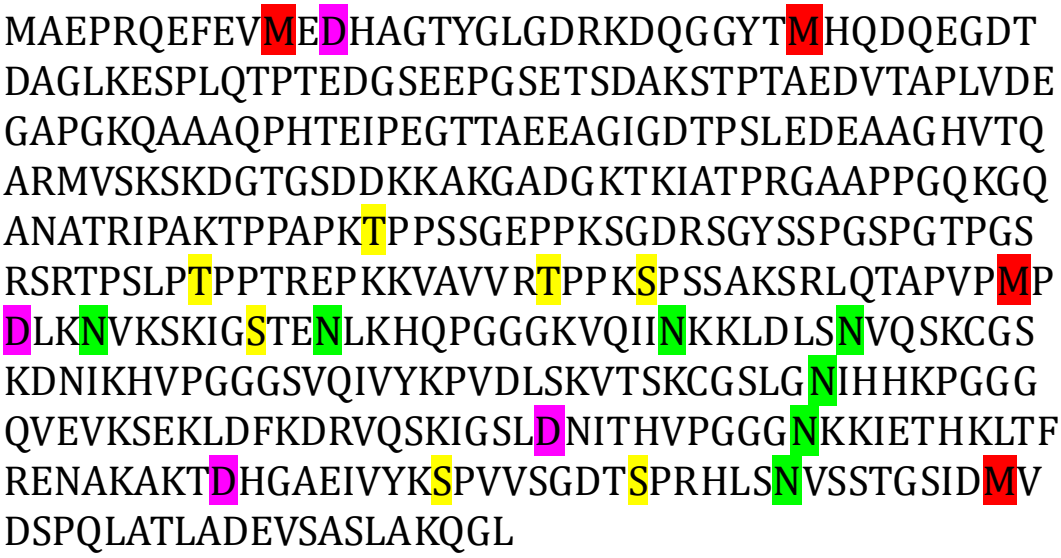


Figure 2.4 All detected modifications mapped across tau sequence. Oxidation sites are highlighted in red, deamidation in green, phosphorylation in yellow, and isomerization in pink.

The results discussed here have directed us towards best practices for isomer detection and describe a workflow starting from intact tissue to data analysis. While some conclusions can be drawn from the preliminary data presented here, more data collection in the future using the method developed here will be required to establish statistical correlations between isomerization and AD disease progression.

Conclusions

This chapter describes the development of a method for detecting isomerization *in vitro* and in tissue samples. Preliminary data shows the utility of the method but leaves some questions, such as is isomerization variable across brain regions, does isomerization occur in other tauopathies, does isomerization affect aggregation and others. Future work that can be done to address some of these questions include examining all brain regions from individuals diagnosed with AD and other disorders and collecting data from enough samples to generate more powerful statistical evidence of any correlation between isomerization and disease progression.

References

[1] Mandelkow, E.-M.; Mandelkow, E. Tau in Alzheimer's Disease. *Trends in Cell Biology* **1998**, *8* (11), 425–427. https://doi.org/10.1016/S0962-8924(98)01368-3.

[2] Meraz-Ríos, M. A.; Lira-De León, K. I.; Campos-Peña, V.; De Anda-Hernández, M. A.; Mena-López, R. Tau Oligomers and Aggregation in Alzheimer's Disease. *Journal of Neurochemistry* **2010**, *112* (6), 1353–1367. https://doi.org/10.1111/j.1471-4159.2009.06511.x.

[3] Avila, J. Tau Phosphorylation and Aggregation in Alzheimer's Disease Pathology. *FEBS Letters* **2006**, *580* (12), 2922–2927. https://doi.org/10.1016/j.febslet.2006.02.067.

[4] Battisti, A.; Ciasca, G.; Tenenbaum, A. Transient Tertiary Structures in Tau, an Intrinsically Disordered Protein. *Molecular Simulation* **2013**, *39* (13), 1084–1092. https://doi.org/10.1080/08927022.2013.794275.

[5] Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradović, Z. Intrinsic Disorder and Protein Function. *Biochemistry* **2002**, *41* (21), 6573–6582. https://doi.org/10.1021/bi012159+.

[6] Uversky, V. N. Intrinsically Disordered Proteins and Novel Strategies for Drug Discovery. *Expert Opinion on Drug Discovery* **2012**, *7* (6), 475–488. https://doi.org/10.1517/17460441.2012.686489.

[7] Truscott, R. J. W.; Schey, K. L.; Friedrich, M. G. Old Proteins in Man: A Field in Its Infancy. *Trends in Biochemical Sciences* **2016**, *41* (8), 654–664. https://doi.org/10.1016/j.tibs.2016.06.004.

[8] Toyama, B. H.; Hetzer, M. W. Protein Homeostasis: Live Long, Won't Prosper. *Nat Rev Mol Cell Biol* **2013**, *14* (1), 55–61. https://doi.org/10.1038/nrm3496.

[9] Mathieson, T.; Franken, H.; Kosinski, J.; Kurzawa, N.; Zinn, N.; Sweetman, G.; Poeckel, D.; Ratnu, V. S.; Schramm, M.; Becher, I.; Steidel, M.; Noh, K.-M.; Bergamini, G.; Beck, M.; Bantscheff, M.; Savitski, M. M. Systematic Analysis of Protein Turnover in Primary Cells. *Nat Commun* **2018**, *9* (1), 689. https://doi.org/10.1038/s41467-018-03106-1.

[10] Hubbard, E. E.; Heil, L. R.; Merrihew, G. E.; Chhatwal, J. P.; Farlow, M. R.; McLean, C. A.; Ghetti, B.; Newell, K. L.; Frosch, M. P.; Bateman, R. J.; Larson, E. B.; Keene, C. D.; Perrin, R. J.; Montine, T. J.; MacCoss, M. J.; Julian, R. R. Does Data-Independent Acquisition Data Contain Hidden Gems? A Case Study Related to Alzheimer's Disease. *J. Proteome Res.* **2022**, *21* (1), 118–131. https://doi.org/10.1021/acs.jproteome.1c00558.

[11] Lambeth, T. R.; Riggs, D. L.; Talbert, L. E.; Tang, J.; Coburn, E.; Kang, A. S.; Noll, J.; Augello, C.; Ford, B. D.; Julian, R. R. Spontaneous Isomerization of Long-Lived Proteins Provides a Molecular Mechanism for the Lysosomal Failure Observed in Alzheimer's Disease. ACS Cent. Sci. **2019**, 5 (8), 1387–1395. https://doi.org/10.1021/acscentsci.9b00369.

[12] Pandey, G.; Julian, R. R. LC–MS Reveals Isomeric Inhibition of Proteolysis by Lysosomal Cathepsins. *Analysis & Sensing* **2022**, *2* (4), e202200017. https://doi.org/10.1002/anse.202200017.

[13] Kagan, R. M.; McFadden, H. J.; McFadden, P. N.; O'Connor, C.; Clarke, S. Molecular Phylogenetics of a Protein Repair Methyltransferase. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **1997**, *117* (3), 379–385. https://doi.org/10.1016/S0305-0491(96)00333-1.

[14] Yamamoto, A.; Takagi, H.; Kitamura, D.; Tatsuoka, H.; Nakano, H.; Kawano, H.; Kuroyanagi, H.; Yahagi, Y.; Kobayashi, S.; Koizumi, K.; Sakai, T.; Saito, K.; Chiba, T.; Kawamura, K.; Suzuki, K.; Watanabe, T.; Mori, H.; Shirasawa, T. Deficiency in Protein L-Isoaspartyl Methyltransferase Results in a Fatal Progressive Epilepsy. *J. Neurosci.* **1998**, *18* (6), 2063–2074. https://doi.org/10.1523/JNEUROSCI.18-06-02063.1998.

[15] Qin, Z.; Dimitrijevic, A.; Aswad, D. W. Accelerated Protein Damage in Brains of PIMT+/− Mice; a Possible Model for the Variability of Cognitive Decline in Human Aging. *Neurobiology of Aging* **2015**, *36* (2), 1029–1036. https://doi.org/10.1016/j.neurobiolaging.2014.10.036.

[16] Kim, J.; Chen, B.; Bru, J.-L.; Huynh, E.; Momen, M.; Aswad, D. W. New Findings on SNP Variants of Human Protein L-Isoaspartyl Methyltransferase That Affect Catalytic Activity, Thermal Stability, and Aggregation. *PLOS ONE* **2018**, *13* (6), e0198266. https://doi.org/10.1371/journal.pone.0198266.

[17] Wu, H.-T.; Van Orman, B. L.; Julian, R. R. Localizing Isomerized Residue Sites in Peptides with Tandem Mass Spectrometry. J. Am. Soc. Mass Spectrom. **2024**, 35 (4), 705–713. https://doi.org/10.1021/jasms.3c00373.

[18] Wu, H.-T.; Riggs, D. L.; Lyon, Y. A.; Julian, R. R. Statistical Framework for Identifying Differences in Similar Mass Spectra: Expanding Possibilities for Isomer Identification. *Anal. Chem.* **2023**, *95* (17), 6996–7005. https://doi.org/10.1021/acs.analchem.3c00495.

[19] Montgomery, K. M.; Carroll, E. C.; Thwin, A. C.; Quddus, A. Y.; Hodges, P.; Southworth, D. R.; Gestwicki, J. E. Chemical Features of Polyanions Modulate Tau Aggregation and Conformational States. *J. Am. Chem. Soc.* **2023**, *145* (7), 3926–3936. https://doi.org/10.1021/jacs.2c08004.

[20] He, B.; Shi, J.; Wang, X.; Jiang, H.; Zhu, H.-J. Label-Free Absolute Protein Quantification with Data-Independent Acquisition. *Journal of Proteomics* **2019**, *200*, 51–59. https://doi.org/10.1016/j.jprot.2019.03.005.

[21] Zhang, W.; Falcon, B.; Murzin, A. G.; Fan, J.; Crowther, R. A.; Goedert, M.; Scheres, S. H. Heparin-Induced Tau Filaments Are Polymorphic and Differ from Those in Alzheimer's and Pick's Diseases. *eLife* **2019**, *8*, e43584. https://doi.org/10.7554/eLife.43584.

[22] Clarke, S. Propensity for Spontaneous Succinimide Formation from Aspartyl and Asparaginyl Residues in Cellular Proteins. *International Journal of Peptide and Protein Research* **1987**, *30* (6), 808–821. https://doi.org/10.1111/j.1399-3011.1987.tb03390.x.

[23] Rao, Y. L.; Ganaraja, B.; Murlimanju, B. V.; Joy, T.; Krishnamurthy, A.; Agrawal, A. Hippocampus and Its Involvement in Alzheimer's Disease: A Review. *3 Biotech* **2022**, *12* (2), 55. https://doi.org/10.1007/s13205-022-03123-4.

[24] Mukherjee, S.; Dubois, C.; Perez, K.; Varghese, S.; Birchall, I. E.; Leckey, M.; Davydova, N.; McLean, C.; Nisbet, R. M.; Roberts, B. R.; Li, Q.-X.; Masters, C. L.; Streltsov, V. A. Quantitative Proteomics of Tau and Aβ in Detergent Fractions from Alzheimer's Disease Brains. *Journal of Neurochemistry* **2023**, *164* (4), 529–552. https://doi.org/10.1111/jnc.15713.

[25] Milani, P.; Gagliardi, S.; Cova, E.; Cereda, C. SOD1 Transcriptional and Posttranscriptional Regulation and Its Potential Implications in ALS. *Neurology Research International* **2011**, *2011* (1), 458427. https://doi.org/10.1155/2011/458427.

[26] Ansari, M. A.; Scheff, S. W. Oxidative Stress in the Progression of Alzheimer Disease in the Frontal Cortex. *Journal of Neuropathology & Experimental Neurology* **2010**, *69* (2), 155–167. https://doi.org/10.1097/NEN.0b013e3181cb5af4.

[27] DeKosky, S. T.; Scheff, S. W. Synapse Loss in Frontal Cortex Biopsies in Alzheimer's Disease: Correlation with Cognitive Severity. *Annals of Neurology* **1990**, *27* (5), 457–464. https://doi.org/10.1002/ana.410270502.

[28] Ronsein, G. E.; Pamir, N.; von Haller, P. D.; Kim, D. S.; Oda, M. N.; Jarvik, G. P.; Vaisar, T.; Heinecke, J. W. Parallel Reaction Monitoring (PRM) and Selected Reaction Monitoring (SRM) Exhibit Comparable Linearity, Dynamic Range and Precision for Targeted Quantitative HDL Proteomics. *J Proteomics* **2015**, *113*, 388–399. https://doi.org/10.1016/j.jprot.2014.10.017.

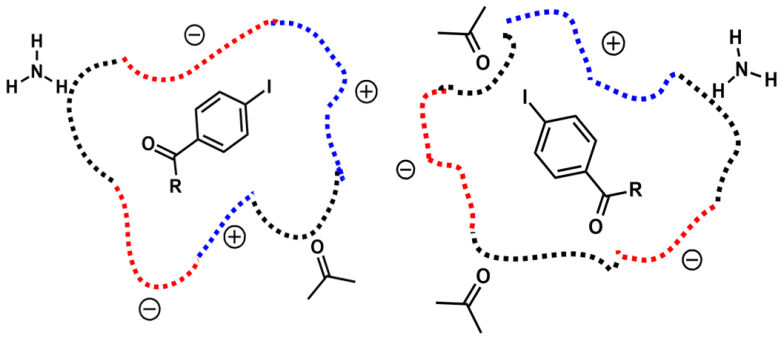
[29] Rauniyar, N. Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *Int J Mol Sci* **2015**, *16* (12), 28566–28581. https://doi.org/10.3390/ijms161226120.

Chapter 3

Differentiation of Peptide Isomers by Excited-State Photodissociation and Ion-Molecule Interactions

Introduction

Absorption of a particular wavelength of light in molecules is primarily dictated by the proper organization of electrons within chemical bonds to form a suitable chromophore. However, absorption can also be influenced by the environment surrounding a chromophore (as shown in Scheme 1). For example, the maxima at which the same chromophore can absorb or emit can shift if the solvent is changed. Such solvochromatic effects are complex,[1] but they fundamentally derive from differences in the polarity of solvents and how solvent molecules interact with the chromophore. In the context of larger molecules such as proteins, solvation of the chromophore can be achieved either partially or entirely by the molecule itself. This allows intrinsic chromophores such as tryptophan to be used as sensitive probes of protein structure.[2] Structural changes due to modulation of external factors such as pH or temperature can be detected if they influence the local hydrogen bonding, salt bridges, or hydrophobic pockets surrounding the tryptophan sidechain. This approach has been further extended by the development and use of extrinsic dyes that can be appended to specific sites and tailored to provide information about highly complex protein systems.[3] These examples make it clear that chromophoric absorption of light can be a powerful structural tool, one that is potentially sensitive to minor changes in environment.



Scheme 3.1 Different solvation environments created by alternate arrangements of charged and polar groups could potentially influence the photochemistry of an active chromophore.

Indeed, recent work has demonstrated that solvatochromism can be used to distinguish similar compounds, including isomers.[4] Isomers and their significance are often overlooked. Isomers are very important in biology, for example, spontaneous isomerization of amino acids within long-lived proteins is known to occur by several mechanisms.[5-7] Isomerization within these proteins can prevent their degradation, which may be important in several age-related diseases including Alzheimer's disease and cataracts.[8-10] Due to the nature of isomers, they are inherently difficult to differentiate. Isomers within non-biological materials can have effects on physical properties as well. For example, isomers of tetrathienoanthracene have been found to have different optical and electrochemical properties, as well as differing thermodynamic stabilities of radical cations.[11] In addition to materials, synthesis and differentiation of pharmaceutically relevant isomers is important. Inverted chirality in drugs can greatly reduce efficacy, and in some cases even cause harm, as demonstrated in studies on antitumor, antidepressant, and antipsychotic drugs.[12,13]

Fragmentation is another strategy for isomer identification and includes methods that utilize photochemistry to cleave bonds. Infrared multiphoton dissociation (IRMPD) has been used to successfully differentiate similar structures, including isomers. For example, glucose disaccharide linkage isomers can be identified with wavelength-tunable IRMPD-MS in combination with lithium tags because fragment ion abundance at specific wavelengths differs between isomers.[14] Small changes in structure have been differentiated by combining IR and ultraviolet (UV) excitation to analyze cold H_2O^+-benzo crown ether complexes.[15] In this system, IR-UV spectra in the region of C-H and O-H stretches are compared. Another study using UVPD to generate a radical, followed by addition of collisional energy to yield radical-directed dissociation (RDD) reported the ability to differentiate glycan isomers.[16] UVPD has also proven useful in the characterization of protein conformers related to cis/tran isomerization at proline.[17] Excited state chemistry involving photo-detachment of electrons has recently enabled circular dichroism-like characterization of DNA strands in the gas phase.[18] Structural information can also be obtained by action spectroscopy, which has revealed structural differences in sequence mutants of amyloid-beta.[19]

In the current study, peptides were tagged with *para*-iodobenzoic acid (4IB) to determine whether small structural changes such as isomerized residues would sufficiently influence chromophore excitation and subsequent fragmentation to allow isomer identification. Peptides containing aspartic acid (Asp), glutamic acid (Glu), and serine (Ser) isomers were subjected to UVPD at 213 nm and 266 nm. Photoexcitation of 4IB leads primarily to loss of iodine following excited state dissociation of the carbon–iodine bond for all

isomers, but the relative fragment-ion intensity was found to be sensitive to isomeric form for some peptides. Molecular dynamics calculations reveal that isomers adopt different structural ensembles in the gas phase, which alters solvation and absorption of UV photons by 4IB.

Methods

Peptide synthesis. The peptides AIPVSR (L-Ser and D-Ser forms), VKLDHG (L-Asp, D-Asp, L-isoAsp, D-isoAsp), VKLDSG (L-Asp, D-Asp, L-isoAsp, D-isoAsp), VHLGGEGYK (L-isoGlu, D-isoGlu), and VTIHEGGPWFK (L-Glu, D-isoGlu) were synthesized as previously described and tagged on bead with *para*-iodobenzoic acid (4IB) on the N-terminus in order to create a photocleavable bond for radical formation.[20]

UVPD experiments. UVPD experiments at 213 nm were conducted using an Orbitrap Velos Pro with an HCD (higher-energy collisional dissociation) cell modified with a quartz window to accept pulses from a 213 nm FQSS 213-Q4 laser. Photoactivation was carried out by inserting ions without collisional excitation and trapping them for 300 ms with the laser firing at 1000 Hz rep rate. UVPD experiments at 266 nm were performed in an LTQ ion trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA). The ion trap was modified to accommodate an Nd:YAG laser (Continuum, Santa Clara, CA) with a quartz window at the back of the ion trap.

Data. In order to evaluate chromophore excitation, data points were first sorted by intensity and the lowest ten intensities were eliminated to omit blank scans. The intensity of the loss of iodine was then plotted against the total ion count and fit with a linear regression. The slope and error of the fit were used to determine the relative abundance

and standard deviation. To evaluate ion-molecule interactions in the 213 nm experiments, intensities of oxygen adducts were treated identically to loss of iodine.

Simulated Annealing. To find structural differences between the peptide isomers of 4IB-VKLDHG, simulated annealing was performed using MacroModel (Schrödinger Inc., Portland, Oregon). Using the OPLS2005 force field, structures were heated to 1000 K and cooled to 200 K, followed by minimization. The outputs from 1000 rounds of annealing were collected, sorted by potential energy, and the structures within 50 kJ of the lowest energy structure were retained. Maestro (Schrödinger Inc., Portland, Oregon) was used to extract distances between pairs of atoms for structural comparisons as described below.

Results and Discussion

To evaluate whether chromophore excitation in the gas phase is sensitive to changes in molecular structure arising from isomerization, we conducted UVPD experiments on a series of peptide isomers. In lieu of monitoring absorption or emission directly, the magnitude of fragmentation caused by the absorption of 266 or 213 nm light was monitored with mass spectrometry. Chromophores containing carbon-iodine bonds that are particularly labile when excited at these wavelengths were covalently attached to each isomer in identical fashion. The primary fragment following photoactivation is loss of iodine as illustrated for L/D Asp isomers of the peptide 4IB-VKLDSG in Fig. 1. The preferential loss of iodine holds true for both excitation wavelengths, although it can be observed that 266 nm light is more selective in exciting this fragmentation pathway almost exclusively. At first glance, the fragmentation spectra obtained for each isomer

appear identical, but closer inspection reveals subtle differences in the relative intensities. For example, the abundance of the c_4 fragment varies by isomer. Furthermore, the loss of iodine relative to the precursor ion varies by isomer for excitation at 213 nm (Figs. 1a and 1b).

Subsequent to loss of iodine in Figs. 1a and 1b, ion-molecule reactions leading to addition of oxygen are noted (observed as +O and $+O_2$ adducts). These reactions occur due to the presence of trace oxygen impurities in the HCD cell that is filled with N_2 collision gas. The radical peptides formed by loss of iodine can easily react with molecular oxygen, which is also a radical, by way of recombination.[21,22] The radicals are retained in the HCD cell for 300 ms, allowing ample time for these adducts to form even at very low oxygen levels. Addition of both +16 Da and +32 Da is observed. The addition of +16 Da likely occurs sequentially when the initial addition of O_2 is followed by UVPD of the resulting peroxyl bond by a subsequent laser shot leading to loss of O.[23] In support of this possibility, the relative abundance of the +16 Da peak in comparison to the +32 Da peak increased with increasing activation time as shown in Fig. S2. The abundance of oxygen adducts also appears to be sensitive to isomeric form.

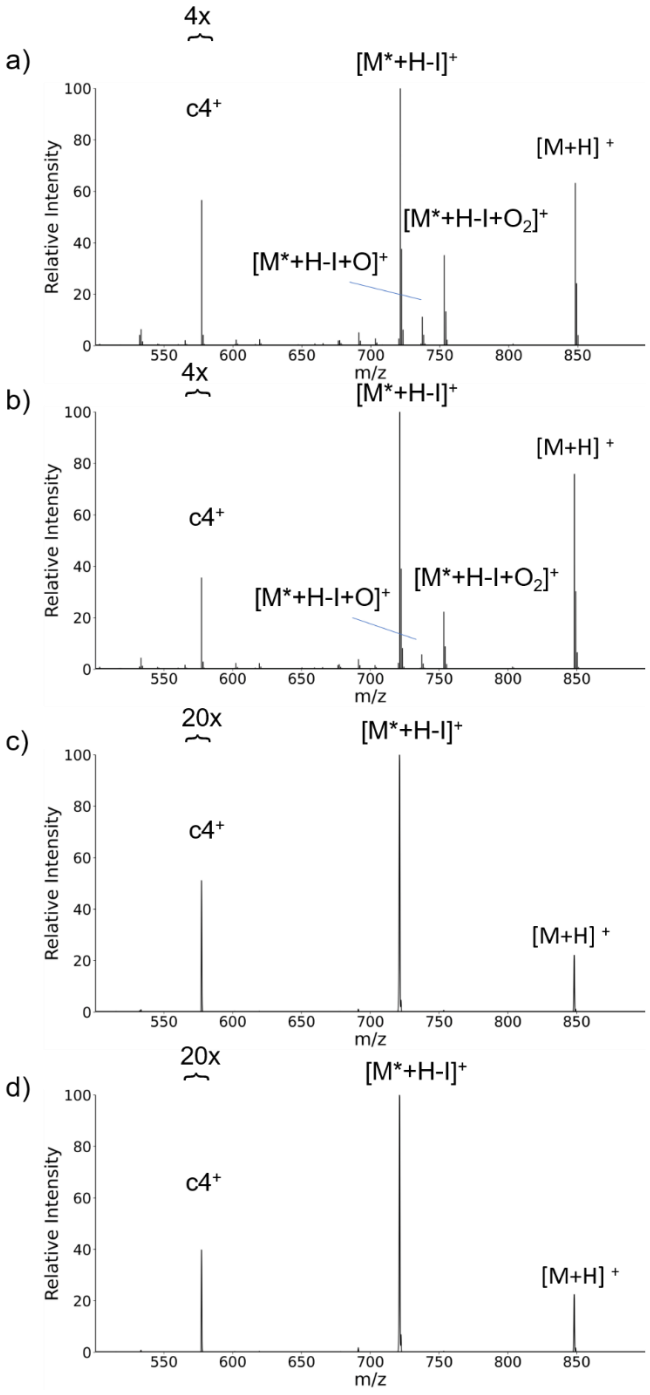


Figure 3.1 Representative mass spectra from isomers of 4IB-VKLDSG (underlined residue locates isomerized position) used to quantitate chromophore excitation at 213 nm a) L-Asp and b) D-Asp and at 266 nm c) L-Asp and d) D-Asp.

To facilitate quantitative comparisons, intensity data from mass spectra are presented in bar graph format as shown in Fig. 2 where the ratios of iodine loss to total ion count for peptides containing Asp isomers are shown. Numerical values for all UVPD yields are provided in Table 1. Asp isomers were selected for these experiments because they represent the most abundant site of isomerization in peptides and have been demonstrated to interfere with protein degradation in biology.[8,10,24] Excitation at both 266 nm (orange bars) and 213 nm (blue bars) was explored. The relative abundances are only relevant within each wavelength, i.e. only the orange bars should be compared to each other. Results for the singly protonated peptide [4IB-VKLDSG+H]$^+$ are shown in Fig. 2a, where excitation at 266 nm yields similar ratios for all isomers. In fact, for the 266 nm ratios, the maximum difference is 2% outside the standard deviation represented by the error bars. In contrast, 213 nm yields greater differences in loss of iodine for Asp isomers, with differences outside error ranging from 13% to 55%. Additional results for the peptide 4IB-VKLDHG are shown in Figs. 2b and 2c. Exchanging serine for histidine also allows the peptide to pick up another proton, enabling the influence of charge state to be examined. Interestingly, the L/D-Asp isomers in the +1 charge state can be differentiated from the L/D-isoAsp isomers following excitation at both 266 and 213 nm, although the magnitude of variation is not large (~3% and 52% max for 266 and 213 nm, respectively). The +2 charge state results in Fig. 2c are similar to those in Fig. 2a in that the 266 nm ratios are similar and vary less than those observed for 213 nm (although the ratios themselves are different for each peptide). Overall, the results in Fig. 2 suggest that 213 nm excitation is more sensitive to isomer structure. This may be a result of the

greater number of chromophores that 213 nm light can excite,[25] which include both the benzyl group and peptide bonds. In contrast, excitation at 266 nm is primarily limited to aromatic groups (including high absorption for 4IB).[26] The greater number of chromophores at 213 nm may enhance the likelihood for interaction via energy transfer to the labile carbon–iodine bond.

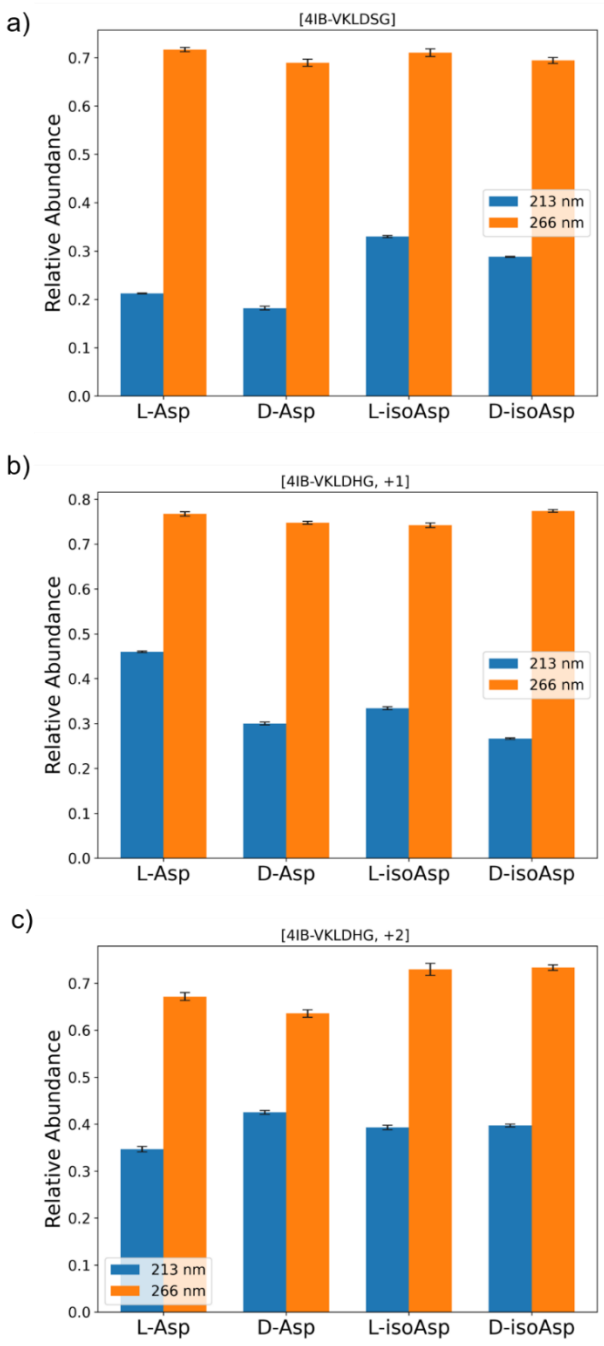


Figure 3.2 Relative chromophore excitation for a) [4IB-VKLDSG+H]+ b) [4IB-VKLDHG+H]+ and c) [4IB-VKLDHG+2H]2+ following excitation at 213 nm (shown in blue) and 266 nm (shown in orange). Error bars represent standard error of regression.

To explore how other sites of isomerization influence peptide structure and light absorption, we examined peptides isomerized at serine and glutamic acid (4IB-AIPVSR, 4IB-VHLGGEGYK, and 4IB-VTIHEGGPWFK), which are also modification sites of biological interest as serine is the second most abundant site of isomerization and glutamic acid can mimic aspartic acid isomerization pathways. [6,27] The peptide containing serine, 4IB-AIPVSR, is easily distinguished by 213 nm excitation (40% difference) but not by 266 nm light (error bars overlap). The serine side chain is small but is capable of forming hydrogen bonds. The results suggest that the serine side chain does not form a hydrogen bond near the 4IB group (due to absence of difference at 266 nm), but may form different hydrogen bonds with backbone groups that interact with 213 nm light. Peptides with glutamic acid isomers, 4IB-VHLGGEGYK and 4IB-VTIHEGGPWFK, contain additional native chromophores in the tyrosine and tryptophan side chains. Although there are measurable differences in the intensity of the loss of iodine for both peptides, the differences are not impressive, though statistically more relevant for the data acquired at 213 nm. These peptides are also longer (with 3 and 5 additional residues beyond those discussed above), which may lead to dilution of any differences in solvatochromatic effects because a lower percentage of the peptide is represented by the isomerized site.

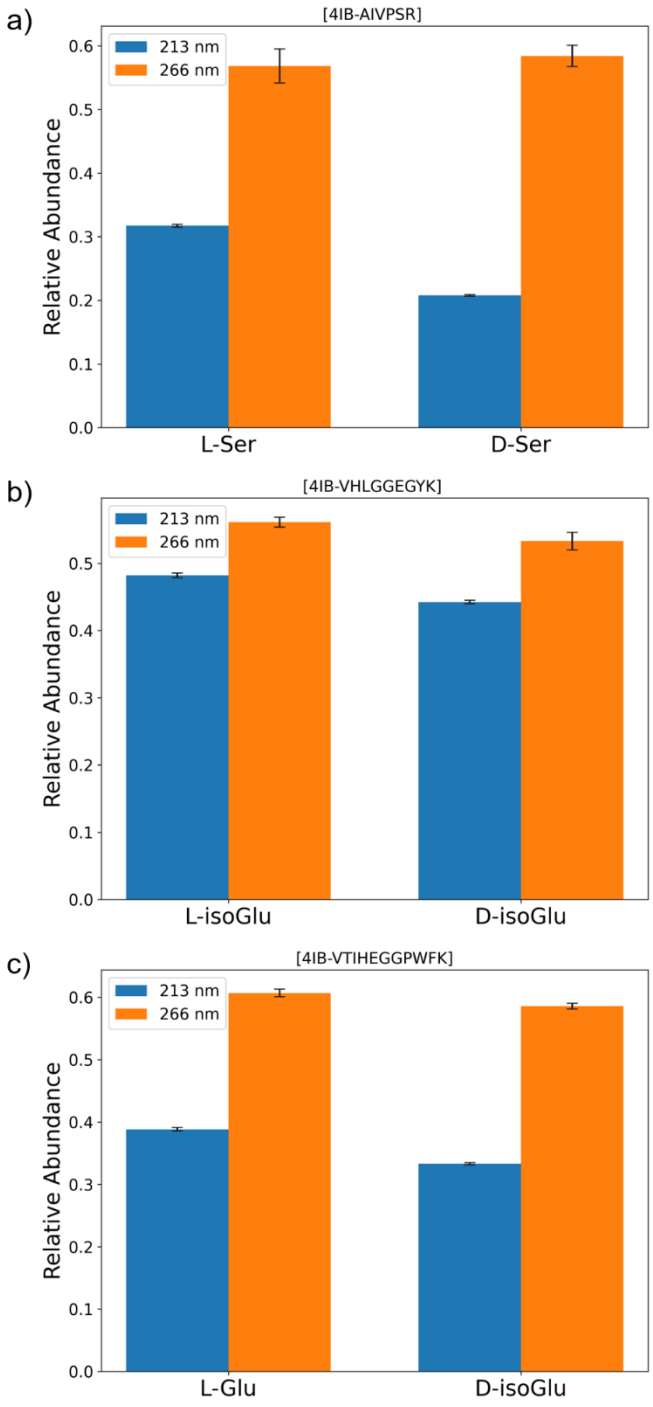


Figure 3.3 Relative chromophore excitation in peptides a) [4IB-AIPVSR+H]+, b) [4IB-VHLGGEGYK+H]+, and c) [4IB-VTIHEGGPWFK+H]+ at UVPD 213 nm (shown in blue) and 266 nm (shown in orange). Error bars represent standard error of regression.

Table 3.1 Numerical values for PD yields

Peptide	Isomer	Charge State	213 PD Yield	213 Error	266 PD Yield	266 Error
AIPVSR	L-Ser	+1	0.32	0.002	0.57	0.027
AIPVSR	D-Ser	+1	0.21	0.001	0.58	0.017
VKLDSG	L-Asp	+1	0.21	0.001	0.72	0.004
VKLDSG	D-Asp	+1	0.18	0.004	0.69	0.008
VKLDSG	L-isoAsp	+1	0.33	0.002	0.71	0.008
VKLDSG	D-isoAsp	+1	0.29	0.001	0.69	0.006
VKLDHG	L-Asp	+1	0.46	0.002	0.77	0.005
VKLDHG	D-Asp	+1	0.30	0.003	0.75	0.003
VKLDHG	L-isoAsp	+1	0.33	0.003	0.74	0.005
VKLDHG	D-isoAsp	+1	0.27	0.002	0.77	0.003
VKLDHG	L-Asp	+2	0.35	0.005	0.67	0.008
VKLDHG	D-Asp	+2	0.43	0.004	0.64	0.008
VKLDHG	L-isoAsp	+2	0.39	0.005	0.73	0.013
VKLDHG	D-isoAsp	+2	0.40	0.003	0.73	0.006
VHLGGEGYK	L-isoGlu	+1	0.48	0.004	0.56	0.007
VHLGGEGYK	D-isoGlu	+1	0.44	0.003	0.53	0.013
VTIHEGGPWFK	L-Glu	+1	0.39	0.003	0.61	0.006
VTIHEGGPWFK	D-isoGlu	+1	0.33	0.002	0.59	0.005

Simulations. Simulated annealing of each of the Asp isomers of 4IB-VKLDHG in the +1 charge state was carried out to explore low energy gas-phase structures and evaluate potential sources of the observed differences in photodissociation yields observed experimentally in Fig. 2b. Distances from the iodine atom of 4IB to the central carboxyl carbon of the Asp side chain are shown as a function of potential energy for all structures within 50 kJ/mol of the lowest energy conformation in Fig. 4. To the right of each scatter plot, a histogram of the distances is shown to allow quantitative comparison of the results. Both the scatter plots and histograms vary between isomers. The L-Asp peptide has isomers with distances spanning a wide range centered at ~11 Å. In contrast, D-Asp

appears to favor two structural populations with distances centered at 5-6 Å and ~11 Å. Interestingly, L-isoAsp appears to favor structures with shorter 5-6 Å distances, while D-isoAsp favors structures with a unique distance of ~7-8 Å. These simulations reveal that small changes, such as isomerization of a single residue can influence the distribution of structures and intramolecular solvation for a small chromophore such as 4IB. While it is not the intent of this project to identify the actual gas phase structure or structures for these isomeric peptides and additional simulations would be required on different charge state configurations to completely evaluate structural space, it is clear from the results in Fig. 4 that structural differences can occur as a result of single residue isomerization.

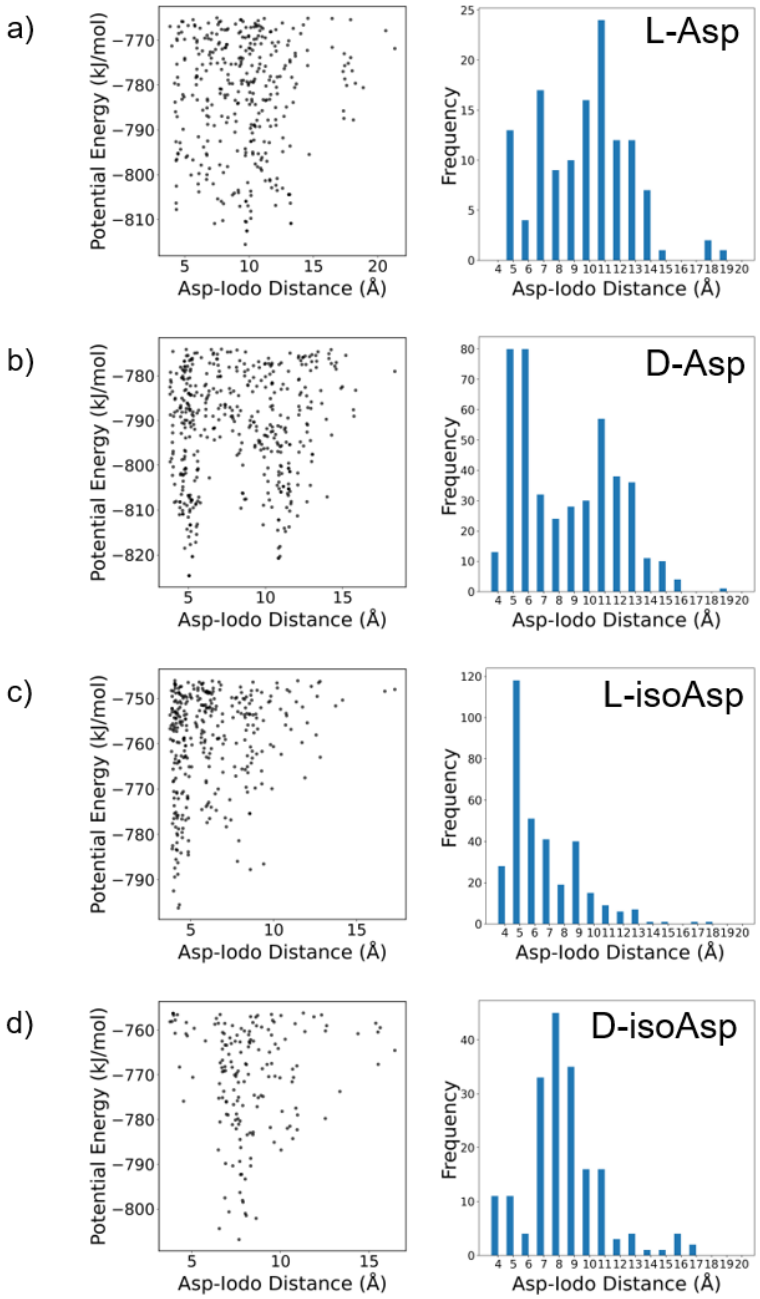


Figure 3.4 Simulated annealing results for 4IB-VKLDHG protonated on lysine. Scatter plots on the left show the potential energy of structures within 50 kJ of the lowest energy structure versus the distance between the iodine atom on 4IB and the central carbon of the aspartic acid in the side chain. On the right, the number of structures within each distance bin are shown in a histogram.

Ion-molecule reactions. Ion-molecule reactions have been used for decades to examine peptide structure in the gas phase.[28-30] In particular, ion-molecule reactions between radical species and oxygen are convenient when carried out with adventitious/contaminant oxygen.[31-34] The same structural differences between isomeric peptides that influence chromophore excitation (such as solvation or absence of it) can also potentially affect the rates of ion-molecule reactions. As discussed above, trace levels of O_2 in the HCD cell lead to ion molecule reactions between the radical peptides and O_2 (see Fig. 1). The relative intensities of the +O and $+O_2$ adducts for 4IB-VKLDSG and 4IB-AIPVSR peptide isomers are shown in Fig. 5. Interestingly, for 4IB-AIPVSR the epimers can be distinguished by both the +O and $+O_2$ adducts. Furthermore, the ratio of the two adducts appear to be correlated with the +O adduct appearing at roughly half the intensity of the $+O_2$ adduct for both isomers. For 4IB-VKLDSG, all for isomers can be identified as well, but the intensities of the +O and $+O_2$ adducts are not correlated between isomers. The most intense and least intense adducts are not the same for the +O and $+O_2$ adducts. The results in Fig. 5 suggest that ion-molecule reactions are very sensitive to three-dimensional structure, enabling identification of isomeric form.

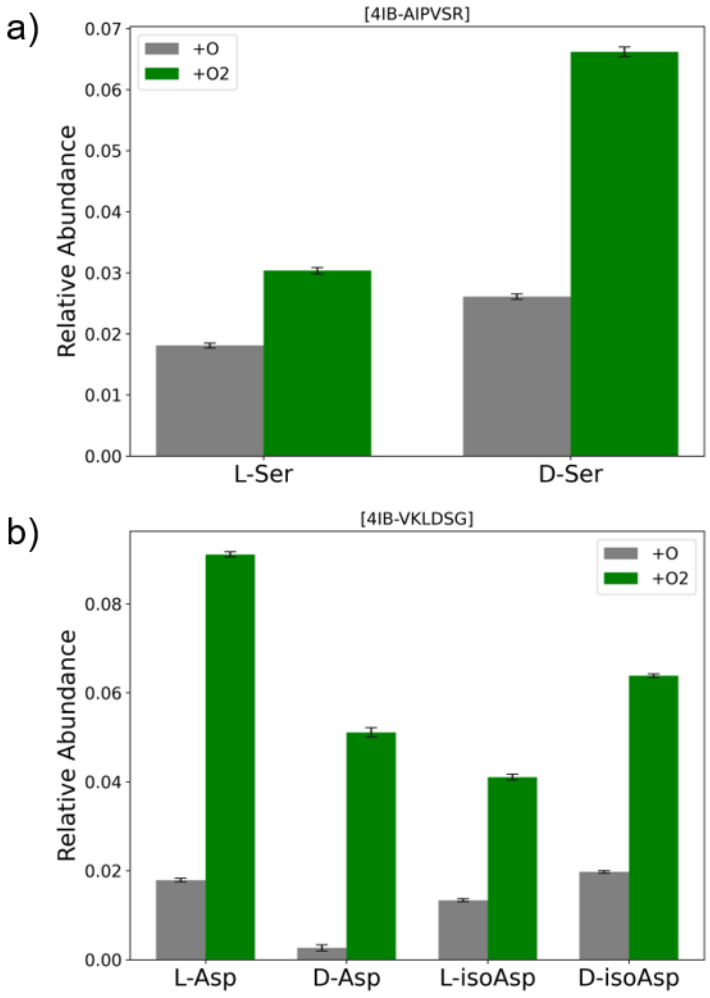


Figure 3.5 Formation of +O (shown in gray) and +O2 (shown in green) adducts in a) [4IB-AIPVSR+H]+ and b) [4IB-VKLDSG+H]+. These adducts were only observed in significant abundances in 213 nm experiments. Error bars represent standard error of regression.

Table 3.2 Numerical values for ion-molecule oxygen adducts

Peptide	Isomer	Charge State	+16 Yield	+16 Error	+32 Yield	+16 Error
VKLDHG	L-Asp	+2	0.018	0.0004	0.091	0.0006
VKLDHG	D-Asp	+2	0.003	0.0007	0.051	0.0011
VKLDHG	L-isoAsp	+2	0.013	0.0004	0.041	0.0007
VKLDHG	D-isoAsp	+2	0.020	0.0003	0.064	0.0004

Conclusions

We have demonstrated that solvatochromic perturbations due to differences in three-dimensional structure between isomeric peptides lead to differences in the absorption of UV light. These differences can be easily quantified by examining the extent of iodine loss in peptides with carbon–iodine containing chromophores. The excitation wavelength is important, with our results suggesting that better structural sensitivity is achieved with 213 nm light. This may be attributable to excitation of additional chromophores such as peptide bonds. The results suggest that sensitivity to isomeric form is reduced as peptide size increases, implying there may be an upper limit to the size of isomer that can be identified. Ion-molecule reactions between the radicals generated by UVPD and adventitious molecular oxygen proved to be quite sensitive to isomeric form. In systems where such reactions can be carried out without modification of instrumentation, this approach may be attractive.

Supplemental Information

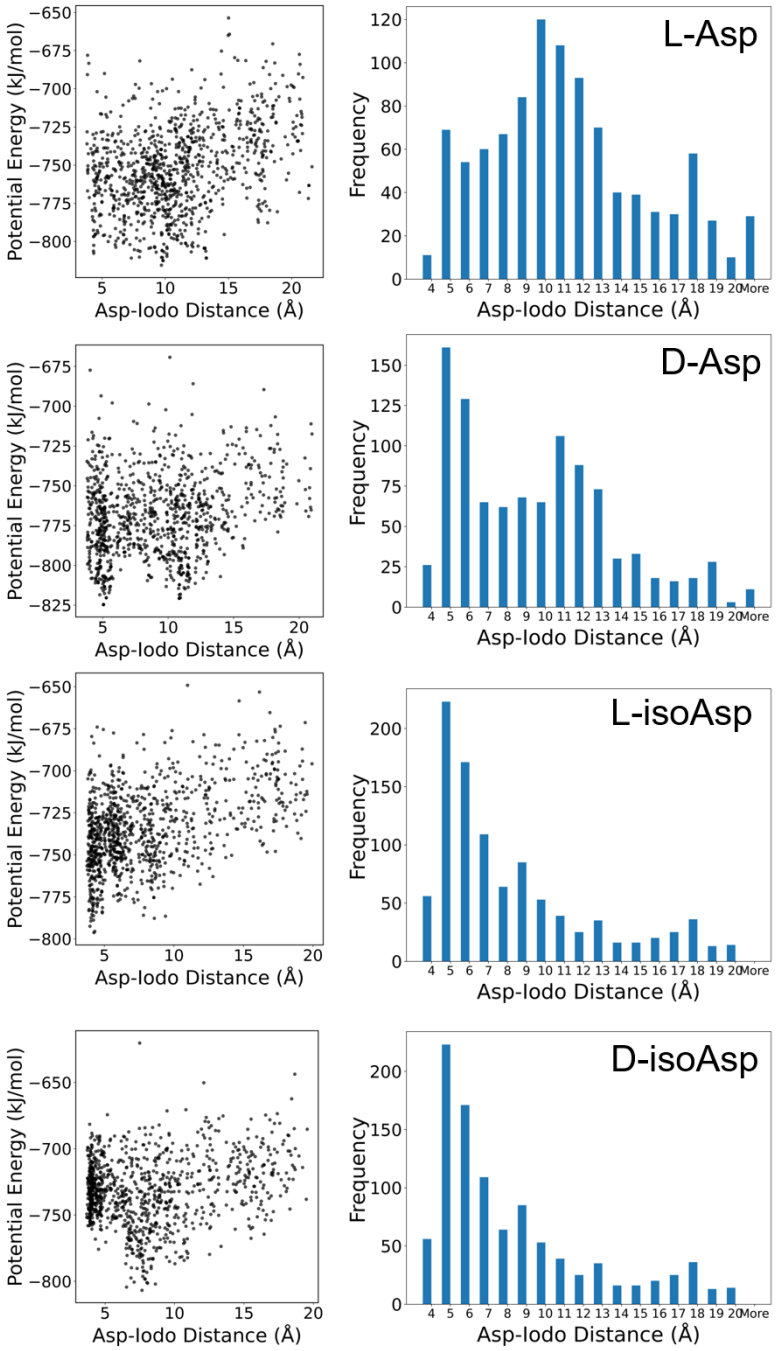


Figure S3.1 Full range of structures generated by simulated annealing of 4IB-VKLDHG, where Fig. 6 shows structures with potential energy within the lowest 50 kJ.

References

[1] Marini, A.; Muñoz-Losa, A.; Biancardi, A.; Mennucci, B. What Is Solvatochromism? *J. Phys. Chem. B.* **2010**, *114* (51), 17128–17135.

[2] Ghisaidoobe, A.; Chung, S. Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques. *Int. J. Mol. Sci.* **2014**, *15* (12), 22518–22538.

[3] Loving, G. S.; Sainlos, M.; Imperiali, B. Monitoring Protein Interactions and Dynamics with Solvatochromic Fluorophores. *Trends Biotechnol.* **2010**, *28* (2), 73–83.

[4] Madeira, P. P.; Loureiro, J. A.; Freire, M. G.; Coutinho, J. A. P. Solvatochromism as a New Tool to Distinguish Structurally Similar Compounds. *J. Mol. Liq.* **2019**, *274*, 740–745.

[5] Robinson, N. E.; Robinson, Z. W.; Robinson, B. R.; Robinson, A. L.; Robinson, J. A.; Robinson, M. L.; Robinson, A. B. Structure-Dependent Nonenzymatic Deamidation of Glutaminyl and Asparaginyl Pentapeptides. *J. Pept. Res.* **2004**, *63* (5), 426–436. https://doi.org/10.1111/j.1399-3011.2004.00151.x.

[6] Riggs, D. L.; Silzel, J. W.; Lyon, Y. A.; Kang, A. S.; Julian, R. Analysis of Glutamine Deamidation: Products, Pathways, and Kinetics. *Anal. Chem*. **2019**, *91*, 13032−13038. https://doi.org/10.26434/chemrxiv.8852246.v1.

[7] Radkiewicz, J. L.; Zipse, H.; Clarke, S.; Houk, K. N. Accelerated Racemization of Aspartic Acid and Asparagine Residues via Succinimide Intermediates: An Ab Initio Theoretical Exploration of Mechanism. *J. Am. Chem. Soc.* **1996**, *118* (38), 9148–9155. https://doi.org/10.1021/ja953505b.

[8] Lambeth, T. R.; Riggs, D. L.; Talbert, L. E.; Tang, J.; Coburn, E.; Kang, A. S.; Noll, J.; Augello, C.; Ford, B. D.; Julian, R. R. Spontaneous Isomerization of Long-Lived Proteins Provides a Molecular Mechanism for the Lysosomal Failure Observed in Alzheimer's Disease. *ACS Cent. Sci.* **2019**, *5* (8), 1387–1395. https://doi.org/10.1021/acscentsci.9b00369.

[9] Lyon, Y. A.; Collier, M. P.; Riggs, D. L.; Degiacomi, M. T.; P Benesch, J. L.; Julian, R. R. Structural and Functional Consequences of Age-Related Isomerization in α-Crystallins. *J. Biol. Chem.* **2019,** *294* (19), 7546–7555. https://doi.org/10.1074/jbc.RA118.007052.

[10] Geiger, T.; Clarke, S. Deamidation, Isomerization, and Racemization at Asparaginyl and Aspartyl Residues in Peptides: Succinimide-Linked Reactions That Contribute to Protein Degradation. *J. Biol. Chem.* **1987,** *262* (2), 785-794.

[11] Brusso, J. L.; Hirst, O. D.; Dadvand, A.; Ganesan, S.; Cicoira, F.; Robertson, C. M.; Oakley, R. T.; Rosei, F.; Perepichka, D. F. Two-Dimensional Structural Motif in Thienoacene Semiconductors: Synthesis, Structure, and Properties of Tetrathienoanthracene Isomers. *Chem. Mater.* **2008**, *20* (7), 2484–2494. https://doi.org/10.1021/cm7030653.

[12] Arnesano, F.; Pannunzio, A.; Coluccia, M.; Natile, G. Effect of Chirality in Platinum Drugs. *Coordination Chemistry Reviews.* **2015**, *284*, 286–297. https://doi.org/10.1016/j.ccr.2014.07.016.

[13] Coutts, R. T.; Baker, G. B. Implications of Chirality and Geometric Isomerism in Some Psychoactive Drugs and Their Metabolites. *Chirality* **1989**, *1* (2), 99–120. https://doi.org/10.1002/chir.530010204.

[14] Polfer, N. C.; Valle, J. J.; Moore, D. T.; Oomens, J.; Eyler, J. R.; Bendiak, B. Differentiation of Isomers by Wavelength-Tunable Infrared Multiple-Photon Dissociation-Mass Spectrometry: Application to Glucose-Containing Disaccharides. *Anal. Chem.* **2006**, *78* (3), 670–679. https://doi.org/10.1021/ac0519458.

[15] Inokuchi, Y.; Ebata, T.; Rizzo, T. R. UV and IR Spectroscopy of Cold H 2 O + −Benzo-Crown Ether Complexes. *J. Chem. Phys. A* **2015**, *119*, 11113-11118. https://doi.org/10.1021/acs.jpca.5b07033.

[16] Riggs, D. L.; Hofmann, J.; Hahm, H. S.; Seeberger, P. H.; Pagel, K.; Julian, R. R. Glycan Isomer Identification Using Ultraviolet Photodissociation Initiated Radical Chemistry. *Anal. Chem.* **2018**, *90* (19), 11581–11588. https://doi.org/10.1021/acs.analchem.8b02958.

[17] Warnke, S.; von Helden, G.; Pagel, K. Analyzing the Higher Order Structure of Proteins with Conformer-Selective Ultraviolet Photodissociation. *Proteomics* **2015**, *15* (16), 2804–2812.

[18] Daly, S.; Rosu, F.; Gabelica, V. Mass-Resolved Electronic Circular Dichroism Ion Spectroscopy. *Science* **2020**, *368* (6498), 1465–1468.

[19] Daly, S.; Kulesza, A.; Poussigue, F.; Simon, A.-L.; Choi, C. M.; Knight, G.; Chirot, F.; MacAleese, L.; Antoine, R.; Dugourd, P. Conformational Changes in Amyloid-Beta (12–28) Alloforms Studied Using Action-FRET, IMS and Molecular Dynamics Simulations. *Chem. Sci.* **2015**, *6* (8), 5040–5047.

[20] Chan, W. C.; White, P. D. *Fmoc Solid Phase Peptide Synthesis : A Practical Approach*; Oxford University Press, 2000.

[21] Moore, B. N.; Blanksby, S. J.; Julian, R. R. Ion-Molecule Reactions Reveal Facile Radical Migration in Peptides. *Chem. Commun* **2009**, 5015–5017. https://doi.org/10.1039/b907833a.

[22] Maccarone, A. T.; Kirk, B. B.; Hansen, C. S.; Griffiths, T. M.; Olsen, S.; Trevitt, A. J.; Blanksby, S. J. Direct Observation of Photodissociation Products from Phenylperoxyl Radicals Isolated in the Gas Phase. *J. Am. Chem. Soc.* **2013**, *135* (24), 9010–9014. https://doi.org/10.1021/ja402610s.

[23] Lightfoot, P. .; Cox, R. .; Crowley, J. .; Destriau, M.; Hayman, G. .; Jenkin, M. .; Moortgat, G. .; Zabel, F. Organic Peroxy Radicals: Kinetics, Spectroscopy and Tropospheric Chemistry. *Atmos. Environ. Part A. Gen. Top.* **1992**, *26* (10), 1805–1961. https://doi.org/10.1016/0960-1686(92)90423-I.

[24] Ritz-Timme, S.; Collins, M. J. Racemization of Aspartic Acid in Human Proteins. *Ageing Research Reviews*. **2002**, *1* (1), 43–59. https://doi.org/10.1016/S0047-6374(01)00363-3.

[25] Pereverzev, A. Y.; Kopysov, V. N.; Boyarkin, O. V. Peptide Bond Ultraviolet Absorption Enables Vibrational Cold-Ion Spectroscopy of Nonaromatic Peptides. *J. Phys. Chem. Lett.* **2018**, *9* (18), 5262–5266.

[26] Ly, T., Zhang, X., Sun, Q., Moore, B., Tao, Y., & Julian, R. R. (2011). Rapid, quantitative, and site specific synthesis of biomolecular radicals from a simple photocaged precursor. *Chem. Comm.*, *47*(10), 2835.

[27] Hooi, M. Y. S.; Truscott, R. J. W. Racemisation and Human Cataract. D-Ser, D-Asp/Asn and d-Thr Are Higher in the Lifelong Proteins of Cataract Lenses than in Age-Matched Normal Lenses. *Age.* **2011**, *33* (2), 131–141. https://doi.org/10.1007/s11357-010-9171-7.

[28] Lebrilla, C. B. Ion-Molecule Reactions as Probes of Gas-Phase Structures of Peptides and Proteins. *Mass Spectrom. Rev.* **1997**, *16* (2), 53–71.

[29] Karanji, A. K.; Khakinejad, M.; Kondalaji, S. G.; Majuta, S. N.; Attanayake, K.; Valentine, S. J. Comparison of Peptide Ion Conformers Arising from Non-Helical and Helical Peptides Using Ion Mobility Spectrometry and Gas-Phase Hydrogen/Deuterium Exchange. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (12), 2402–2412.

[30] Osburn, S.; Burgie, T.; Berden, G.; Oomens, J.; O'Hair, R. A. J.; Ryzhov, V. Structure and Reactivity of Homocysteine Radical Cation in the Gas Phase Studied by Ion–Molecule Reactions and Infrared Multiple Photon Dissociation. *J. Phys. Chem. A* **2013**, *117* (6), 1144–1150.

[31] Xia, Y.; Chrisman, P. A; Pitteri, S. J.; Erickson, D. E.; McLuckey, S. A. Ion/Molecule Reactions of Cation Radicals Formed from Protonated Polypeptides via Gas-Phase Ion/Ion Electron Transfer. *J. Am. Chem. Soc.* **2006**, *128* (36), 11792–11798.

[32] Moore, B. N.; Blanksby, S. J.; Julian, R. R. Ion–Molecule Reactions Reveal Facile Radical Migration in Peptides. *Chem. Commun.* **2009**, *33*, 5015-5017.

[33] Barlow, C. K.; Wright, A.; Easton, C. J.; O'Hair, R. A. J. Gas-Phase Ion-Molecule Reactions Using Regioselectively Generated Radical Cations to Model Oxidative Damage and Probe Radical Sites in Peptides. *Org. Biomol. Chem.* **2011**, *9* (10), 3733.

[34] Zhang, X.; Julian, R. R. Photoinitiated Intramolecular Diradical Cross-Linking of Polyproline Peptides in the Gas Phase. *Phys. Chem. Chem. Phys.* **2012**, *14* (47), 16243.

Chapter 4

Radical initiated photodissociation and radical directed dissociation provide complementary gas phase structural information

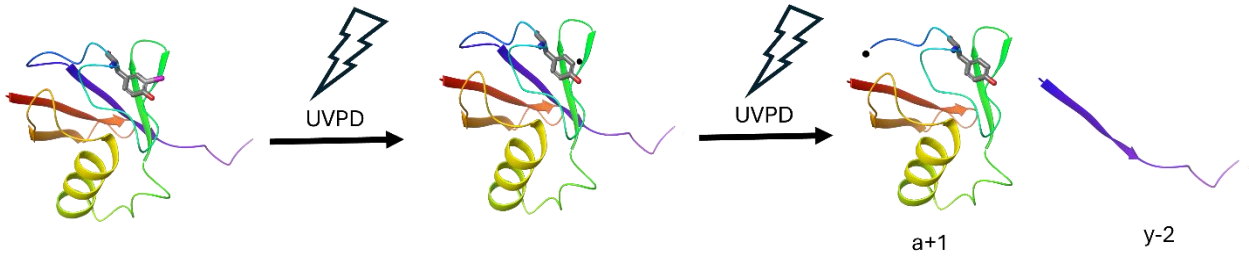
Introduction

Protein structure is one of the cardinal properties of biological function. Protein structure in many conditions is of interest to understanding biological function and other properties; however, the gas phase is a difficult set of conditions for studying structure, but it may be the best alternative for instances of samples that are not amenable to current techniques. There are several different existing mass spectrometry (MS) techniques for probing gas phase structure. Methods such as limited proteolysis or ion mobility can be combined with MS to provide structural information.[1] Ion-ion and ion-molecule reactions can also be utilized to provide gas phase structure information. [2-5] These techniques leverage chemical crosslinks or hydrogen/deuterium exchange. Traditional MS fragmentation techniques can also provide structural information, particularly collision and electron-based methods.[6,7] However, these techniques are limited in that they yield fragments by dissociation of bonds that are present at each residue in a protein, which is very useful for interrogating primary sequence, but may be less effective for determination of higher order structure than alternative techniques.[8,9]. Additionally, electron based techniques may miss significant regions of a protein, especially in cases where native structure is preserved, requiring a disruption to protein folding to generate fragments.[10] It has also been shown that fragment ion generation during ECD of low

charge states of ubiquitin is significantly inhibited due to noncovalent interactions preventing separation of the fragment ions.[11]

One method that has been especially useful for elucidating protein gas phase structure using specific fragmentation is radical directed dissociation (RDD).[12-14] This method requires a radical precursor in the form of a photocleavable bond. Proteins can be modified selectively at tyrosine residues using chloramine T iodination.[15,16] In addition to modifying only tyrosine, and histidine with a lower efficiency, the reaction can be performed at biological pH. Under these gentle conditions, the protein is able to retain some of its native structure, and only solvent exposed residues will be modified.

Once the protein is modified, a radical is generated by ultraviolet photoactivation with a 213 or 266 nm laser. This activation is performed without significant heating of the protein, and elements of the protein structure can be preserved. After the radical is formed, additional collisional energy can be added in an MS^3 step to cause the radical to migrate. The migration will be limited to residues that are in close spatial proximity, and the migration site is determined by thermodynamics and kinetics.[17] The radical will cause characteristic fragmentation such as a-, c-, and z-1 fragments that normally would not be observed using collisional activation methods, which informs structure about where the fragmentation occurs. The fragmentation is dependent on beta C-H bond dissociation energies.[18] The radical can also be further photoactivated to produce radical induced dissociation. The specificity of the modification, combined with the unique fragment ions generated by radical dissociation provide structural information that may not be available in other types of experiments such as electron based methods or 193 nm UVPD.



Scheme 4.1 The experimental workflow for generating radicals on proteins for PD dissociation or RDD.

RDD was previously performed on various charge states of iodo-ubiquitin to obtain structural information.[12] The results reported were from experiments performed on an LTQ classic ion trap mass spectrometer (Thermo Fisher, Waltham, MA). While an impressive amount of structural information was obtained from these preliminary results, the experiments pushed the resolution limits of the instruments, and much of the data was unassignable. In contrast, the tribrid instruments available at the time of this study offer a vast increase in resolution, sensitivity, and m/z range over previous instrumentation.[19,20] Thus, using an Orbitrap Fusion Lumos (Thermo Fisher, Waltham, MA), were able to obtain even more structural information from iodo-ubiquitin using RDD, and additional information in the MS^2 step using alpha radical activation. From the radical dissociation we are able to determine structural features in the gas phase of different charge states of ubiquitin and myoglobin.

Experimental

Materials

Ubiquitin from bovine erythrocytes, myoglobin from horse heart, and sodium metabisulfite were all purchased from Sigma Aldrich (St. Louis, MO). Chloramine T trihydrate and sodium iodide were purchased from Thermo Fisher Scientific (Waltham, MA).

Modification of Proteins

Solvent exposed tyrosine residues were iodinated as previously described.[21] Briefly, in 100-200 uL of water, protein was modified by adding 1 molar equivalent of sodium iodide followed immediately by the addition of 2 molar equivalents of chloramine t in 3-6 uL of water. The protein and iodinating reagents were incubated at room temperature for varying amounts time depending on protein. After incubation, the reaction was quenched with 4 molar equivalents of sodium metabisulfite in water and desalted using Pierce protein concentrators (Thermo Scientific).

Mass Spectrometry

Experiments were performed on an Orbitrap Fusion Lumos using nano electrospray ionization. Proteins were sprayed at a concentration of 10 uM in either Optima water with 0.1% formic acid (Fisher) or 200 mM ammonium acetate at pH 7. Resolution was 120K. For RDD experiments, the radical was generated using photodissociation from a 266 nm laser (Brand) in the MS2 step. The radical was then isolated, and additional CID energy was introduced in the MS3 step, with the exception of the native apo myoglobin data. For this experiment, additional CID energy was applied in MS2 prior to radical generation in

order to remove the heme group, thus requiring an MS4 step to perform RDD. Data was analyzed with MASH explorer.[22]

Results and Discussion

Experimental results for iodo-ubiquitin are shown in figure 1. The PD spectrum of iodo-ubiquitin in the 7+ charge state is shown in Fig. 1a. The C-I bond readily absorbs at 266 nm, resulting in homolytic cleavage which leaves a radical on the protein at the tyrosine sidechain, which is observed as the base peak in the spectrum. The UVPD activation time used in fig 1 was 200 ms. While the primary product of the photoactivation of the modified protein is loss of iodine, there are also substantial backbone fragments present in the spectrum. The other large peaks correspond to a+1/y-2 pairs at residues which are fragments generated from an alpha radical that is further photoactivated to generate the a/y-ion complement pairs. The PD activation time must be sufficient to both generate and photoactivate the resulting radical; the a/y pairs are not observed at lower activation PD times.

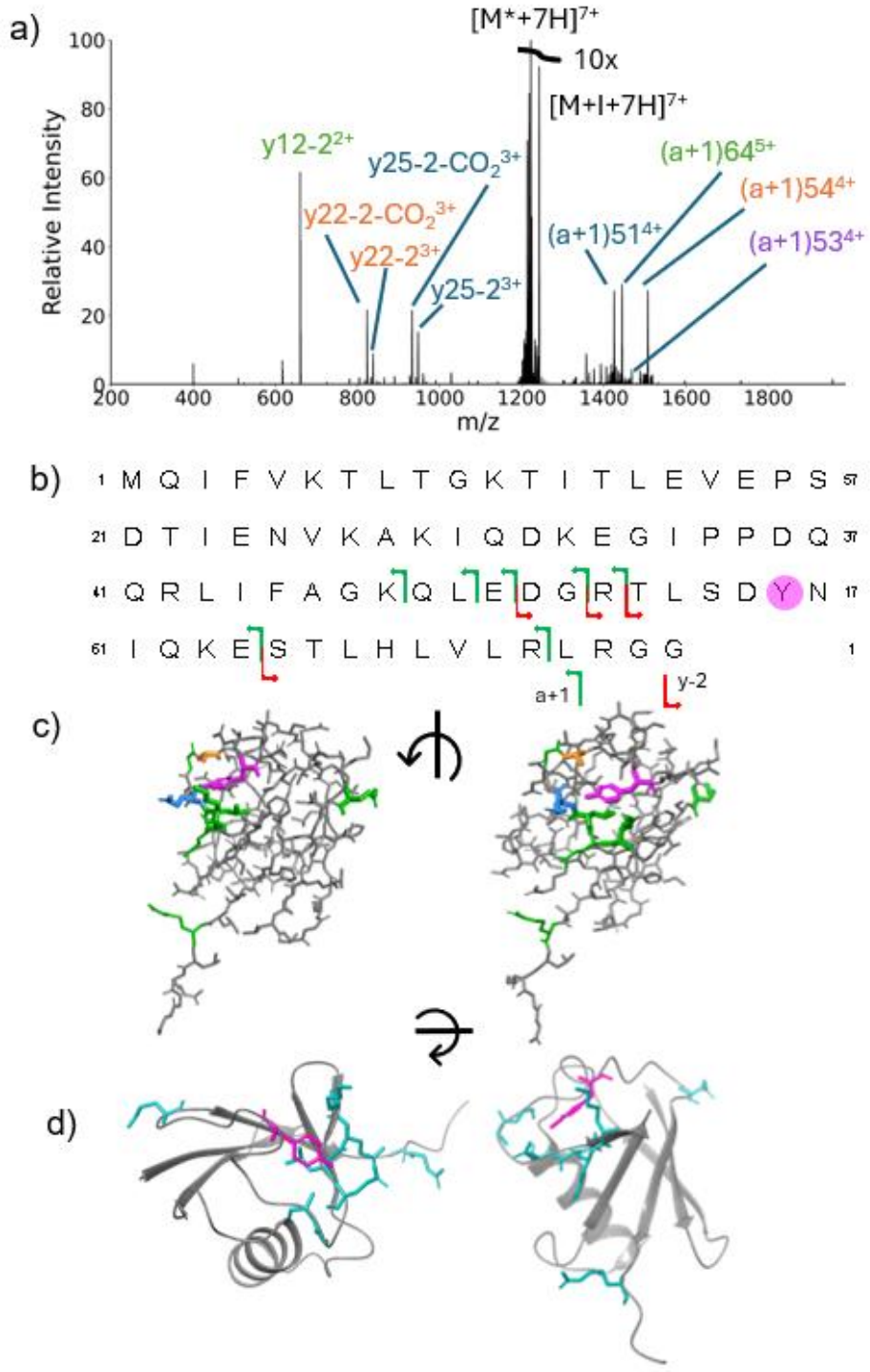

a)
[M*+7H]7+
10x
[M+I+7H]7+
y12-2 2+
y25-2-CO2 3+
y22-2-CO2 3+
y22-2 3+
y25-2 3+
(a+1)64 5+
(a+1)54 4+
(a+1)51 4+
(a+1)53 4+
Relative Intensity
m/z
b)
M Q I F V K T L T G K T I T L E V E P S
D T I E N V K A K I Q D K E G I P P D Q
Q R L I F A G K Q L E D G R T L S D Y N
I Q K E S T L H L V L R L R G G
a+1
y-2
c)
d)

Figure 4.1 PD provides structural information from alpha radical dissociation. a) PD spectrum of iodo-ubiquitin in the 7+ charge state. b) Locations of the radical fragments mapped on the protein sequence. c) Locations of the radical induced fragments on the crystal structure. d) Location of radical fragments and radical origin mapped on ubiquitin crystal structure with secondary structures shown. All fragments and Tyr59 are located on flexible regions.

Fig. 1b shows the a and y fragment ions mapped on the protein sequence. Ubiquitin features a single tyrosine residue at position 59. The fragment ions observed due to alpha radical activation are distributed beyond 10 residues of Tyr59. This distribution is indicative of protein folding preserved in the gas phase due to the ability of the radical to migrate to the residues located far away in sequence. The distances between the Tyr59 sidechain and the alpha carbons of the residues where fragment ions are located are shown in table 1. The distances are from crystal structure PDB ID 5TR4.[23] While the furthest distance in the crystal structure is 16 angstroms, this fragment is located at Arg72, which is located on the flexible C-terminal tail of the protein, shown in figure 1d. This makes it likely that the protein is able to adopt a conformation in the gas phase where Arg72 is in close enough proximity to Tyr59 for the radical to migrate to, enabling the generation of the a72 ion.

There is a cluster of radical induced fragments located from residues 48 to 55. There are two intense pairs of a+1/y-2 fragment ions located at residues 51 and 54, as well as a series of smaller a+1 ions. In the crystal structure, residues 51 and 54 are equidistant from Tyr59 at 5 angstroms. The relative intensity of both pairs is very similar, indicating that the radical has an equal chance of reaching the residues, perhaps pointing to this section of the structure being retained. Interestingly, both y-2 ions from these residues feature a

subsequent -CO2 peak. This is unusual, as the y-2 peak where these losses originate would not contain a radical, and must proceed through another pathway. The other most abundant backbone fragments are an a+1/y-2 pair at residue 64. Ser64 is located on the same loop as Tyr59 as well as Glu51 and Arg54, the sites of the other abundant a/y pairs. As shown in Fig 1d, Glu51 and Arg54 are located on either side of the Tyr59 sidechain in the crystal structure. Additionally, Cook et al. proposed a hydrogen bond between the Tyr sidechain OH and the amide of Glu51 from crystallographic data.[24] However, they noted that iodination of the Tyr would cause disruption of this local structure due to sterics. The MS data, in particular the abundance of the fragments at residue 64 suggest that there is a sizable amount of conformations where the Tyr sidechain is pointed in the opposite direction, bringing it in close proximity to generate the radical fragments observed there, but the abundance of the fragments at Glu51 and Arg54 suggest there is a sizable amount of conformations where the Tyr remains oriented near the residues despite the modification.

Table 4.1 Distance of Tyr59 from alpha carbons of residues where radical induced fragments are located in the 7+ PD spectrum, from crystal structure**.**

Residue	Distance (Å)
48	7.97
49	6.90
50	3.88
51	5.06
53	7.82
54	5.04
64	16.15

Fig. 2a shows the PD spectrum of the 5+ charge state of iodo-ubiquitin. Despite having the same PD activation time as the 7+ charge state of 200 ms, the spectrum features no backbone fragmentation. Additionally, the relative intensity of the loss of iodine peak is much lower than the 7+ charge state. Previously, it has been found that even minor alterations to structure can cause a detectable change in UV activation of a chromophore.[25] In this case, less activation of the lower charge state is likely attributable to H-bonding to the Tyr sidechain OH group.[26] A cluster of fragment ions is located within 30 m/z of the loss of iodine peak, but these are attributed to side chain and water losses. The sidechain losses are primarily type I losses, although there is a -74M peak that is a type II loss, which is a radical-initiated pathway. These losses point mostly toward a lack of fragmentation happening, however there could be some fragmentation occurring,

which is not observed due to the preservation of H-bonding. The lower amount of photoactivation observed at the lower charge could be explained by several possibilities. One possibility is some structural induced modulation of the homolytic cleavage of iodine, resulting in less radical being generated initially. Another possibility could be pathways available to the radical at this charge state that do not result in backbone cleavage. This lack of backbone fragmentation limits the amount of structural information that can be elucidated. However, the addition of collisional energy in an MS3 step can cause radical migration resulting in beta radical induced cleavages, known as radical directed dissociation (RDD).

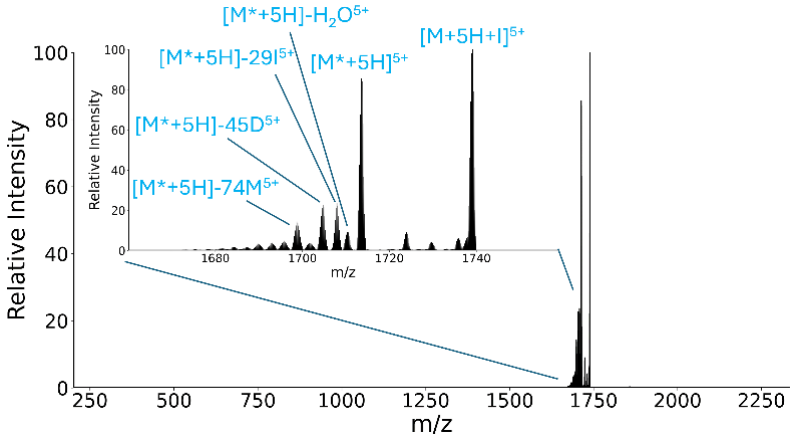


Figure 4.2 PD spectrum of iodo-ubiquitin in the 5+ charge state. No backbone fragmentation is present.

In contrast to the PD results, the RDD spectrum yields many backbone fragments. Some of these are b and y ions due to proton-initiated pathways, but most are from radical dissociation, including the most intense ions. The fragmentation is plotted on the sequence in figure 2a. Some ions observed include x-H2O, which are radical induced

fragments that can occur at serine and threonine. The 5+ charge state features four of them, located at threonines 12 and 55, and serines 57 and 65. Other radical induced fragments include a-, c-, and z-1 ions. The location of the fragment ions and Tyr59 are shown plotted on the crystal structure in fig 2b. The 5+ charge state was acquired by electrospray in ammonium acetate. Under these conditions, the protein may retain more of its native folded structure. However, the RDD data indicates that there are at least some conformations where the structure diverges significantly from the crystal structure. Radical induced fragments are located at Asp39 and Thr12, which are 14.94 angstroms and 20.61 angstroms away from Tyr59, respectively, in the crystal structure. It should be noted that these particular fragment ions have quite low relative abundances when compared to ions that are located closer in the crystal structure, which could indicate that the conformations that lead to fragmentation at Asp39 and Thr12 may be low in population compared to others.

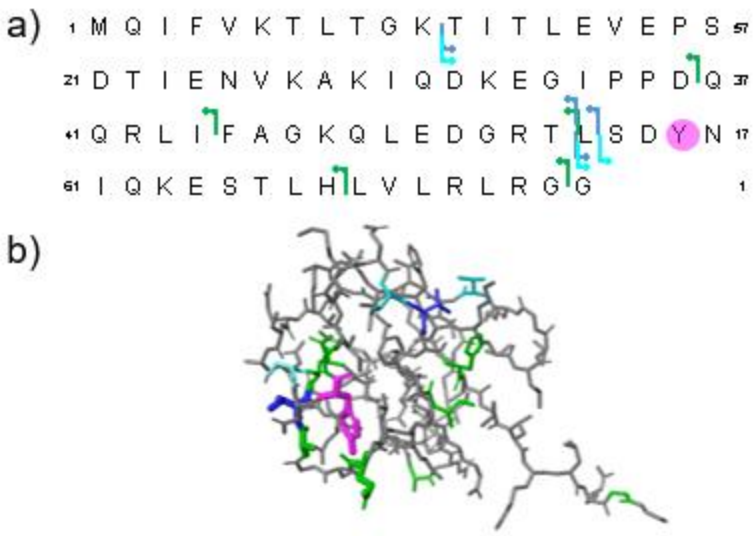
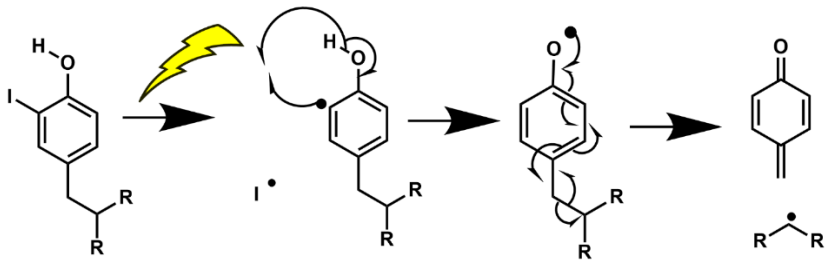


Figure 4.3 RDD provides structural information at charge states where PD does not. a) PD spectrum of iodo-ubiquitin in the 5+ charge state. No backbone fragmentation is present. b) Location of radical fragments and radical origin mapped on ubiquitin sequence. c) Location of radical fragments and radical origin mapped on ubiquitin crystal structure.

The RDD spectrum of the 10+ charge state is shown in figure 4a. There are backbone fragments spread throughout the spectrum, but the largest fragment ion by far is a loss of the tyrosine sidechain. The fragment does not appear in the PD spectrum. This fragment is also highly charge state-dependent; the fractional abundance of several RDD fragments are plotted against the charge state in figure 4b. The -106Y peak dominates at 10+ but is still present at 11+. This points to a conformation in which the sidechain is constrained away from the rest of the protein, making it favorable to abstract the hydrogen from the OH group of the tyrosine sidechain.



Scheme 4.2 Proposed mechanism for loss of tyrosine sidechain.

This could result in the radical being isolated on the tyrosine residue, which could lead to loss of the sidechain by the mechanism shown in scheme 2. If the -106Y peak is reisolated and additional photoactivation is applied in an MS4 step, the -106Y splits into an a+1/y-2 pair with very high specificity. This is shown in figure 4c. This fragmentation confirms that the sidechain loss leaves an alpha radial on the protein, which can be further activated to cause radical dissociation in a PD step. It has been demonstrated that tyrosine deprotonation facilitates electron detachment upon PD activation in negative mode.[27] This detachment yields selective alpha cleavage N-terminal to Tyr. This finding supports the proposal that the sidechain is being deprotonated via radical abstraction. Furthermore, when an eight residue section of the backbone containing Tyr is simulated, a conformational search using Maestro finds the lowest energy structure has the Tyr sidechain constrained in space.

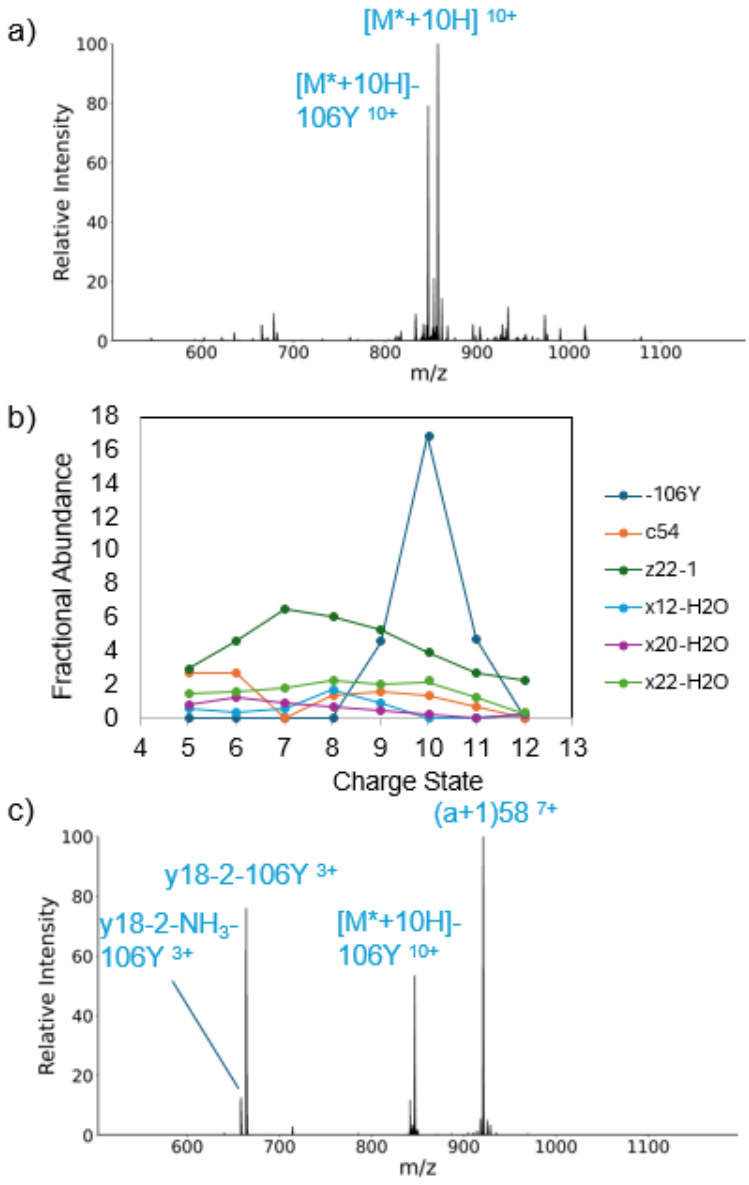


Figure 4.4 Specific charge states result in specific radical dissociations. a) The 10+ charge state of iodo-ubiquitin results in a very intense -106Y fragment. b) When the -106Y fragment is reisolated and subject to further photoactivation, it fragments with very high specificity into an a+1/y-2 pair, indicating an alpha radical. c) Radical fragments plotted by charge state. -106Y is highly charge state dependent, suggesting a conformation in which the tyrosine sidechain is constrained from the rest of the backbone.

Figure 5a displays the sites where RDD fragments are generated when starting from the 8+ charge state of iodinated myoglobin. Myoglobin contains two tyrosine residues, but Tyr103 is much more solvent exposed than Tyr146.[28] The primary product of the iodination reaction is a singly modified protein, indicating that most of the iodine is present at Tyr103. The largest peaks in the spectrum occur at 2422 and 2117 m/z, which correspond the loss of heme in the 7+ and 8+ charge states, respectively. Important features of this spectrum include a- and c-ions, as well as K/E sidechain losses. These fragments can be attributed to well-documented radical mechanisms, meaning they are most likely generated by radical migration rather than proton-initiated activation. Although typically sidechain losses can't be assigned to a particular residue, the residue directly n-terminal to the Tyr103 is a lysine, and therefore highly likely the origin of the sidechain loss. The heme group is also observed at 616 m/z.

Structural information can be better interpreted when fragments are plotted on the protein sequence. Figure 5b shows the RDD fragments plotted along the myoglobin sequence, with the iodinated tyrosine highlighted. Examination of the fragments along the sequence show that the RDD fragments are distributed widely throughout the sequence with respect to Tyr103, the origin of the radical. When these fragments are mapped on the native crystal structure[29] (Fig. 5b), it can be observed that most of the fragments generated are close in space to Tyr103 in the native structure. This proximity combined with the distance in sequence suggests that some elements of the native structure have been preserved in the gas phase. However, the distance from Tyr103 in

space of some fragmentation sites in the native structure, such Phe138 and His119 are indicative of the presence of some alternative structural conformations.

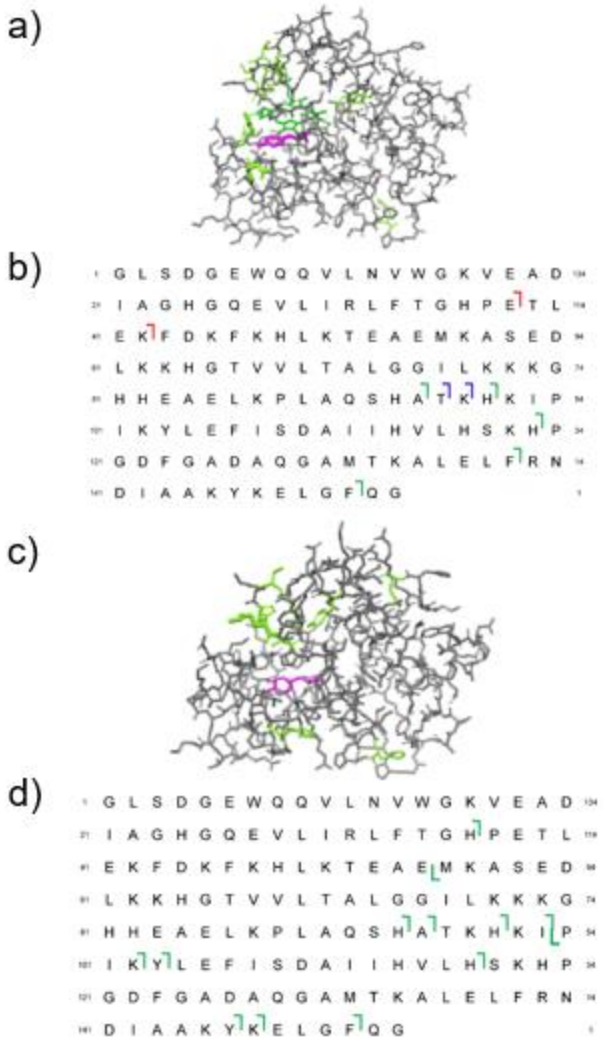


Figure 4.5 RDD of apo and holo myoglobin produce similar but distinct fragmentation. Radical fragments of 8+ holo myoglobin plotted on a) the crystal structure and b) protein sequence. Radical fragments of 7+ apo myoglobin plotted on c) the crystal structure and d) protein sequence.

When the apo protein, which has no reported crystal structure, is generated by applying CID energy before the photoactivation step, similar RDD fragmentation is produced (Fig 5c,d). The loss of the heme group results in a loss of charge, leaving predominantly the 7+ charge state. When the heme group is lost, expected backbone fragmentation that was missing in the holo protein data appears. In particular, a-ions at His36 and His82 are generated by RDD of the apo protein, but not the holo protein. These results indicate a quenching effect of the radical by the heme group. When the fragments are mapped on the native crystal structure, shown in Fig 5c, this effect becomes more apparent as His37 and His83 are 7.7 and 11.0 Å, respectively, away from the heme in the native crystal structure. Both of these are located on flexible regions of the protein, allowing the possibility for some of the native structure to be retained while His36 and 82 are close to the radical generated at Tyr103. The distant fragments observed in the apo protein data are in close proximity in sequence to the distant fragments observed in the holo data, however they are not located at the exact same residues. Specifically, fragments His82 and His116 are observed when the apo protein is fragmented, and Phe138 and His119 are observed when the holo protein is fragmented. These remote fragments are not exactly reproduced between the apo and holo forms, but their proximity in the protein sequence and native crystal structure suggest that in the alternate, non-native conformations some structural elements are being retained from the native structure.

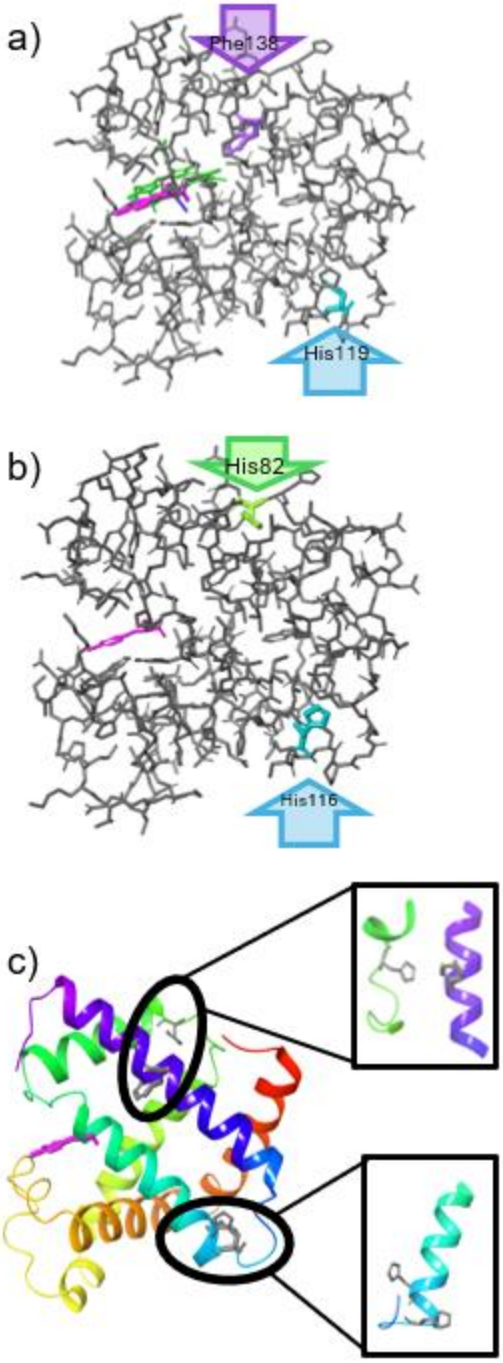

a)
Phe138
His119
b)
His82
His116
c)

Figure 4.6 The apo and holo versions of the protein share surprisingly similar features. Here the locations of RDD fragments are plotted on PDB structure 1BJE. a) The remote fragments in the holo protein are Phe138 and His119. The heme group is displayed in Kelly gprotein are apo protein features remote fragments at His82 and His116, denoted with arrows. c) The secondary structures of myoglobin shown. His119 and His116 (bottom circle) are close in sequence, and will always have to located near each other. However, Phe138 and His82 (top circle) are located on different secondary structures. It is surprising that the contact between the two structures is preserved in both apo and holo forms of the protein.

Additionally, when secondary structures are examined in the native crystal structure, His83 (a-ion observed in apo protein) and Phe138 (a-ion observed in holo protein) are located on different alpha helices. However, the helices are located directly adjacent to one another (shown Fig. 6a,b). The observation of RDD fragments from these regions during the fragmentation of the apo and holo forms of the protein suggest that the proximity of the alpha helices containing His82 and Phe138 is conserved in the apo alternate conformation and the holo alternate conformation as well as the native structure.

Conclusion

Structural information from gas phase molecules can be elucidated not only from RDD in the MS3 step, but from photoactivation alone when a radical precursor is present. At charge states where loss of iodine yield is reduced, RDD provides more backbone fragmentation than PD, providing more structural information. Both techniques provide useful complementary structural information. Radical induced dissociation was able to provide information about how non-covalent interactions affect fragmentation in the gas phase, including for conditions where a crystal structure is not available. Radical

precursors that could selectively modify other residues would further increase the utility of these methods.

References

[1] Vandermarliere, E.; Stes, E.; Gevaert, K.; Martens, L. Resolution of Protein Structure by Mass Spectrometry. *Mass Spectrometry Reviews* **2016**, *35* (6), 653–665. https://doi.org/10.1002/mas.21450.

[2] Cheung See Kit, M.; Carvalho, V. V.; Vilseck, J. Z.; Webb, I. K. Gas-Phase Ion/Ion Chemistry for Structurally Sensitive Probes of Gaseous Protein Ion Structure: Electrostatic and Electrostatic to Covalent Cross-Linking. *International Journal of Mass Spectrometry* **2021**, *463*, 116549. https://doi.org/10.1016/j.ijms.2021.116549.

[3] Kang, W.-Y.; Mondal, A.; Bonney, J. R.; Perez, A.; Prentice, B. M. Structural Elucidation of Ubiquitin via Gas-Phase Ion/Ion Cross-Linking Reactions Using Sodium-Cationized Reagents Coupled with Infrared Multiphoton Dissociation. *Anal. Chem*. **2024**, *96* (21), 8518–8527. https://doi.org/10.1021/acs.analchem.4c00442.

[4] Green, M. K.; Lebrilla, C. B. Ion-Molecule Reactions as Probes of Gas-Phase Structures of Peptides and Proteins. *Mass Spectrometry Reviews* **1997**, *16* (2), 53–71. https://doi.org/10.1002/(SICI)1098-2787(1997)16:2<53::AID-MAS1>3.0.CO;2-8.

[5] "Time-Resolved ElectroSpray Ionization Hydrogen-Deuterium Exchange Mass Spectrometry for Studying Protein Structure and Dynamics."

[6] Reid, G. E.; Wu, J.; Chrisman, P. A.; Wells, J. M.; McLuckey, S. A. Charge-State-Dependent Sequence Analysis of Protonated Ubiquitin Ions via Ion Trap Tandem Mass Spectrometry. *Anal. Chem*. **2001**, *73* (14), 3274–3281. https://doi.org/10.1021/ac0101095.

[7] Loo, J. A.; Goring, A. K.; Lantz, C.; Wei, B.; Zenaidee, M. A. Investigating 3D Structures of Native Proteins and Complexes through Electron-Based Dissociation. **2020**. https://doi.org/10.1039/9781839161056-00102.

[8] 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. *J Am Soc Mass Spectrom* **1990**, *1* (1), 53–65. https://doi.org/10.1016/1044-0305(90)80006-9.

[9] Zubarev, R. A. Electron-Capture Dissociation Tandem Mass Spectrometry. *Current Opinion in Biotechnology* **2004**, *15* (1), 12–16. https://doi.org/10.1016/j.copbio.2003.12.002.

[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. https://doi.org/10.1002/pmic.201400516.

[11] Ganisl, B.; Valovka, T.; Hartl, M.; Taucher, M.; Bister, K.; Breuker, K. Electron Detachment Dissociation for Top-Down Mass Spectrometry of Acidic Proteins. *Chemistry* **2011**, *17* (16), 4460–4469. https://doi.org/10.1002/chem.201003709.

[12] Ly, T.; Julian, R. R. Elucidating the Tertiary Structure of Protein Ions in Vacuo with Site Specific Photoinitiated Radical Reactions. *J. Am. Chem. Soc.* **2010**, *132* (25), 8602–8609. https://doi.org/10.1021/ja910665d.

[13] Ly, T.; Julian, R. R. Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130* (1), 351–358. https://doi.org/10.1021/ja076535a.

[14] Sun, Q.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R. Side Chain Chemistry Mediates Backbone Fragmentation in Hydrogen Deficient Peptide Radicals. *J. Proteome Res.* **2009**, *8* (2), 958–966. https://doi.org/10.1021/pr800592t.

[15] Miyashita, M.; Yamashita, S. Studies on Iodinated Compounds: VI. Separation Characteristics of Iodohistidines on Reversed-Phase High-Performance Liquid Chromatography. *Journal of Chromatography A* **1989**, *475* (2), 135–144. https://doi.org/10.1016/S0021-9673(01)89669-2.

[16] McGowan, E. B.; Stellwagen, E. Reactivity of Individual Tyrosyl Residues of Horse Heart Ferricytochrome c toward Iodination. *Biochemistry* **1970**, *9* (15), 3047–3053. https://doi.org/10.1021/bi00817a017.

[17] Tureček, F.; Julian, R. R. Peptide Radicals and Cation Radicals in the Gas Phase. *Chem. Rev.* **2013**, *113* (8), 6691–6733. https://doi.org/10.1021/cr400043s.

[18] Ly, T.; Julian, R. R. Tracking Radical Migration in Large Hydrogen Deficient Peptides with Covalent Labels: Facile Movement Does Not Equal Indiscriminate Fragmentation. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (6), 1148–1158. https://doi.org/10.1016/j.jasms.2009.02.009.

[19] Levy, M. J.; Washburn, M. P.; Florens, L. Probing the Sensitivity of the Orbitrap Lumos Mass Spectrometer Using a Standard Reference Protein in a Complex Background. *J. Proteome Res.* **2018**, *17* (10), 3586–3592. https://doi.org/10.1021/acs.jproteome.8b00269.

[20] Espadas, G.; Borràs, E.; Chiva, C.; Sabidó, E. Evaluation of Different Peptide Fragmentation Types and Mass Analyzers in Data-Dependent Methods Using an Orbitrap Fusion Lumos Tribrid Mass Spectrometer. *PROTEOMICS* **2017**, *17* (9), 1600416. https://doi.org/10.1002/pmic.201600416.

[21] Regoeczi, E. *Iodine Labeled Plasma Proteins*; CRC Press: Boca Raton, 2019. https://doi.org/10.1201/9780429288388.

[22] Guner, H.; Close, P. L.; Cai, W.; Zhang, H.; Peng, Y.; Gregorich, Z. R.; Ge, Y. MASH Suite: A User-Friendly and Versatile Software Interface for High-Resolution Mass Spectrometry Data Interpretation and Visualization. *J Am Soc Mass Spectrom* **2014**, *25* (3), 464–470. https://doi.org/10.1007/s13361-013-0789-4.

[23] PDB 5TR4

[24] Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8Åresolution. *Journal of Molecular Biology* **1987**, *194* (3), 531–544. https://doi.org/10.1016/0022-2836(87)90679-6.

[25] 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**. https://doi.org/10.1039/D0CP04111D.

[26] Kirk, B. B.; Trevitt, A. J.; Blanksby, S. J.; Tao, Y.; Moore, B. N.; Julian, R. R. Ultraviolet Action Spectroscopy of Iodine Labeled Peptides and Proteins in the Gas Phase. *J. Phys. Chem. A* **2013**, *117* (6), 1228–1232. https://doi.org/10.1021/jp305470j.

[27] Shaw, J. B.; Ledvina, A. R.; Zhang, X.; Julian, R. R.; Brodbelt, J. S. Tyrosine Deprotonation Yields Abundant and Selective Backbone Cleavage in Peptide Anions upon Negative Electron Transfer Dissociation and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2012**, *134* (38), 15624–15627. https://doi.org/10.1021/ja3032086.

[28] Davies, M. J. Identification of a Globin Free Radical in Equine Myoglobin Treated with Peroxides. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1991**, *1077* (1), 86–90. https://doi.org/10.1016/0167-4838(91)90529-9.

[29] Maurus, R.; Bogumil, R.; Nguyen, N. T.; Mauk, A. G.; Brayer, G. Structural and Spectroscopic Studies of Azide Complexes of Horse Heart Myoglobin and the His-64-->Thr Variant. *Biochem J* **1998**, *332 (Pt 1)* (Pt 1), 67–74. https://doi.org/10.1042/bj3320067.

Chapter 5

False positives in internal fragment assignments

Introduction

As mass spectrometers are developed with higher resolving power and mass accuracy, the more information can be extracted from acquired data. In particular, advances in MS technology has allowed top-down MS to expand to a much larger field than was possible in previous decades, enabling experiments such as top-down analysis of monoclonal antibodies, megadalton complexes, mapping of PTMs on intact proteins, and many more.[1-5] Multiple fragmentation techniques are available for top-down experiments, especially in tribrid instruments, such as HCD, ETD, and UVPD. These often yield very complex spectra, of which a large proportion may be unassignable after searching for expected terminal ions. Additionally, complement ion pairs are often not observed, even in highly charged analytes. If the cleavage to generate one terminal ion takes place, the other terminal ion should theoretically be observed, unless additional fragmentation is taking place. This suggests that some terminal ions are being further fragmented.

Recently, these unassignable fragments in top-down spectra of biomolecules have been attributed to internal fragments, or terminal fragments that are further fragmented.[6-9] When assigned, these fragments can greatly increase sequence coverage. Internal fragments are more likely to occur during high energy fragmentation methods, mainly HCD and UVPD. HCD primarily generates primarily a, b, y-type ions. When two cleavages occur, the number of possible ion types generated becomes larger, meaning that for every y-ion generated, there are possible ay and by ions generated. To further

complicate matters, there are possible isomers that have the same exact masses. While these are obstacles for assigning internal fragments in HCD experiments, UVPD data is even more difficult due to the variety of fragment ion types possible, particularly in 193 nm experiments.

MS^3 of known b- and y-ions has been shown to confidently assign internal ions.[10] However MS^3 may not always be practical, especially in larger proteins which may have many fragment ions and overlapping ions. There are currently multiple programs that have internal ion assignment features.[11-13] Utilizing ClipsMS, a program that does not require a data file but matches by mass, we investigate whether internal ions can statistically be confidently assigned. Here we examine 193 nm PD and HCD top-down data. We also perform ETD experiments to gain an understanding of what happens to terminal ions when activation is increased.

Experimental

Materials

Human histone protein 3.3 was generously provided by the Benjamin Garcia lab. Histone protein H3.3 (1-20) was purchased from Bachem (Torrance, CA, United States).

Data analysis

Thermo RAW files were downloaded from the Japan ProteOme Standard Reposistory (jPOSTrepo), accession number JPST001885.[14,15] The RAW files were deconvoluted using ExD Viewer (Agilent, Santa Clara, Ca, USA).[16] In ExD Viewer the fragmentation type was set to ExD, iterative matching set to multi-pass, and ion identification quality set to default. Following deconvolution, ions were exported to a csv file. Ions that were not

assigned as terminal ions by ExD Viewer were exported into a mass list and run in ClipsMS with the correct sequence and both terminal and internal ions on. Terminal and internal fragment errors were set to 2 ppm. The smallest internal fragment size was set to 5 amino acids. The masses assigned as internal ions when searched against the correct sequence were searched against three shuffled sequences with internal ions on and 2 ppm error tolerance. Protein sequences were shuffled using the shuffle protein feature on the Sequence Manipulation Suite.[17] ClipsMS features multiple possible assignments for internal ions. After each analysis, any duplicates were removed, and the number of assignments without duplicates was used for calculating the percent overlap.

Experimental

ETD experiments were performed using an Orbitrap Fusion Lumos (Thermo) using a resolution of 120,000. ETD spectra were collected at ETD reaction times of 0.5, 1, 2, 5, and 10 ms for 18+, 22+, and 24+ charge states. When 0.5, 1, or 2 ms ETD reaction times were used, data was collected with 50 microscans. When 5 or 10 ms reaction times were used, data was collected with 10 microscans. Data was then deconvoluted and assigned using ExD Viewer.

Results and Discussion

Experimental

ETD experiments were performed to identify complement ion pairs, or lack thereof, in a whole protein. ETD was chosen as the fragmentation method due to its predictable and reproducible fragment generation. Data previously reported shows all complement ions are observed where expected in ETD fragmentation of peptides.[18,19] The

complement pairs should also be observed in ETD of whole proteins unless the fragments are subject to further fragmentation. To test this, histone protein 3.3 was chosen as the analyte due to its distribution of charge, making it possible, in theory, to observe complementary ion pairs throughout the sequence. Very gentle activation was applied, starting with an ETD reaction time of 0.5 or 1 ms, with the priority of preserving complement pair fragments rather than the usual goal of maximum sequence coverage. The activation was then sequentially increased to see if ion complement pairs would be preserved or disappear with more activation.

Fragmentation at increasing activation time for the 18+ charge state is shown in Fig. 1. As shown, if there is not enough activation time, there are fewer fragments generated, and also fewer complements. At 1 ms activation time there are fewer fragments generated overall, with a sequence coverage of 61%. Of these fragments, only 27 c/z complement pairs are present.

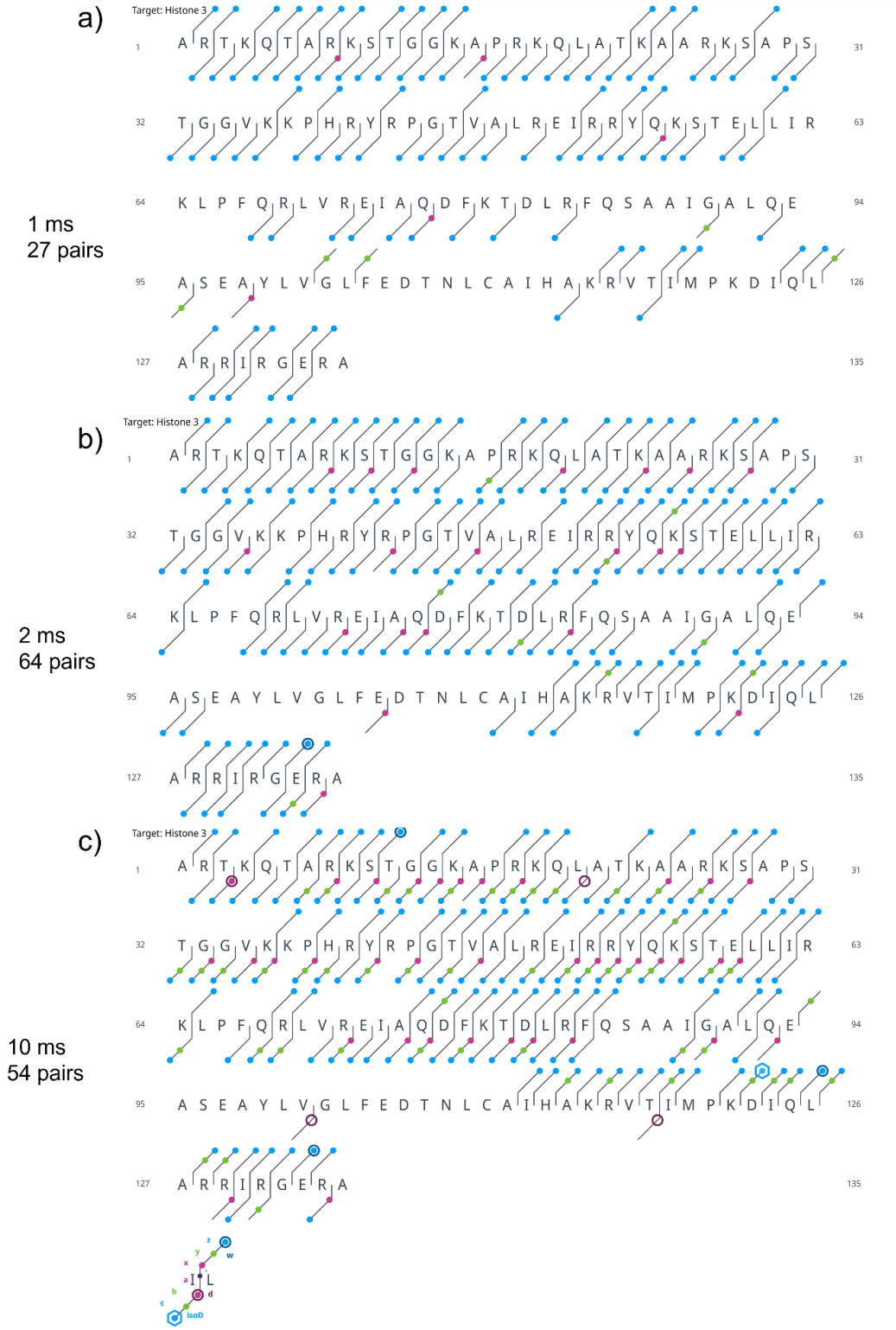

a)
Target: Histone 3
1 ms
27 pairs
b)
Target: Histone 3
2 ms
64 pairs
c)
Target: Histone 3
10 ms
54 pairs

Figure 5.1 Fragmentation observed for ETD of 18+ histone protein 3 with a) 1 ms activation time, b) 2 ms activation time, and c) 10 ms activation time.

When the activation time is increased to 2 ms, the number of both fragment ions overall and complement pairs increases substantially. Sequence coverage rises to 81%, and there are 64 complement pairs. If activation time is increased even further to 10 ms, the sequence coverage falls slightly to 80%, and the number of complement pairs drops to 54. When comparing the fragments assigned in the 2 ms and 10 ms data, the fragments that disappear with more activation appear to be large, with z-ions located near the n-terminus and c-ions located near the c-terminus disappearing. This is not surprising, given that large terminal ions are statistically less likely to be observed.[20] Additionally, the remaining half of the complement pair in these cases is typically a low intensity peak, meaning the complement ion is probably being generated, but subsequently fragmented resulting in an abundance that is not detectable. This trend occurs in multiple charge states of the histone protein. Fig. 2 shows the number of fragments observed in pairs as well as the number of fragments not assignable as terminal ions for three charge states of histone protein 3.

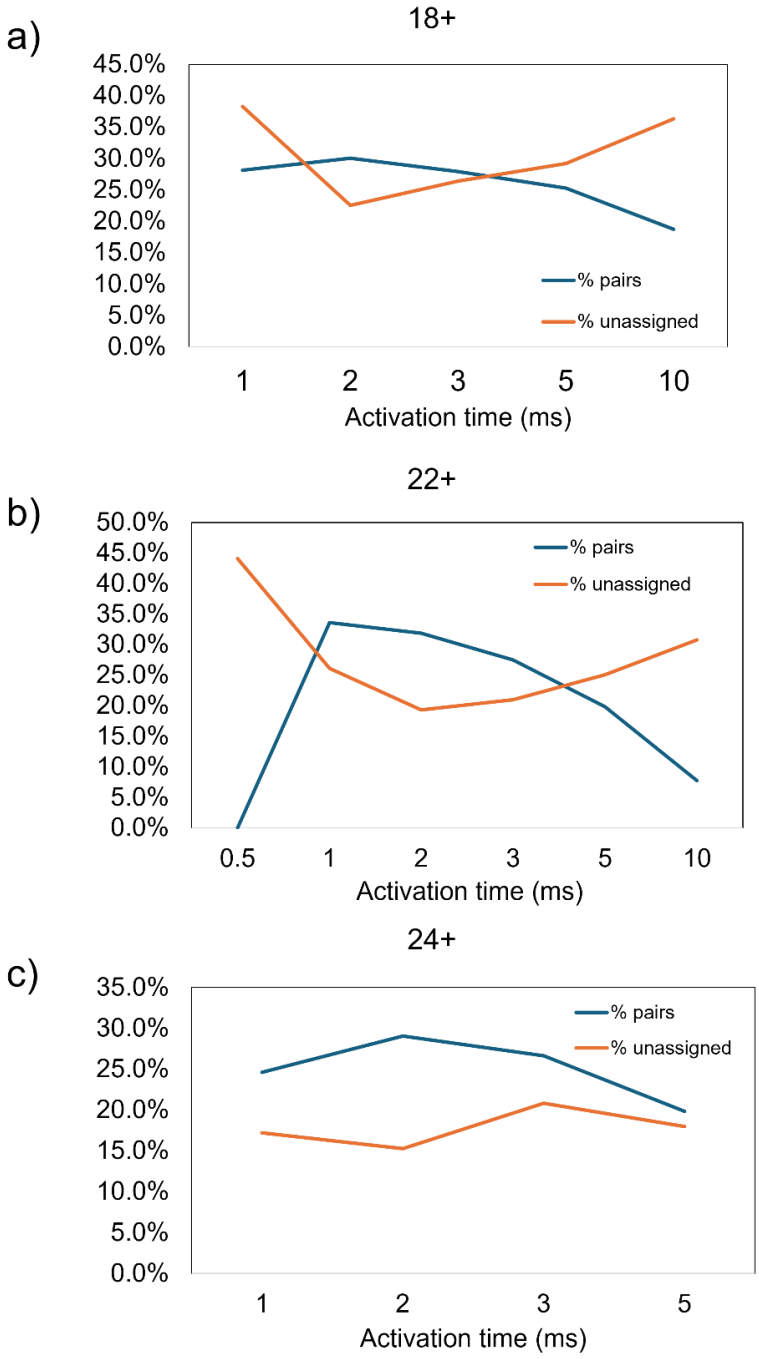


Figure 5.2 Proportion of assigned ions with complement pairs and proportion of ions that cannot be assigned as terminal ions for charge states a) 18+, b) 22+, and c) 24+.

The percentage of assigned fragments in complement pairs was calculated by dividing the number of fragments in pairs divided by total number of assigned fragments. The percentage of unassigned fragments was also calculated. When plotted by activation time, all examined charge states displayed a trend of having an inversely proportional relationship between percent of paired fragments and percent of unassigned fragments. The percentage of paired fragments appears to peak at 1 or 2 ms. These are also the points at which the fewest percent of unassigned fragments are present. As activation increases past 2 ms, the percentage of unassigned fragments increases, with the exception of the 24+ charge state, which still features a higher percentage of unpaired ions at 5 ms than 2 ms activation.

Internal ions have not been observed in electron-based activation of whole proteins.[21] In reality, it is likely that internal ions are generated in ETD when sufficient activation is present, however, the number of fragments that are possible make observation of them unlikely. Additionally, for every internal fragment generated, there are two terminal fragments generated, inherently making the intensity of the internal fragment lower than terminal fragments.

It is possible that there are fragment ions present that cannot be assigned as terminal ions that are not due to internal ions. During the ETD fragmentation method, the z-ion is left with the radical on the alpha position.[22] There is some evidence that the radical can migrate from this position based on the amino acid adjacent to the cleavage site.[23-26] Further activation of the radical-containing z-ions can lead to side-chain losses and

backbone cleavage.[27] It is possible that z-ions can undergo losses in ETD experiments without further activation, resulting in ion types that have not been identified yet.

To gain some understanding of what cleavages are generated by increased activation, ETD experiments on a truncated version of the histone protein (1-20) were performed. Although the number of complement c/z pairs only varied by 1 across activation times, the presence of other ion types varied by activation time. In particular, more a-, b-, and y-ions were present at longer activation times. The generation of these fragment types, which are usually seen in collisional activation spectra, are likely being generated by a process analogous to hot electron capture dissociation (HECD). Zubarev et al. report that if ECD is performed with high energy electrons, additional fragments from the expected c- and z-ions will be generated. These include, b- and y-ions as well as d- and w-ions, which are generated from side chain losses adjacent to the radical present in a· and z· ions.[28] These fragments are present in the intact protein and the peptide version. To determine more about the generation of these fragments, the peptide data was more closely examined.

By 5 ms of activation, the peptide features a-ions at almost every residue. There are also b- and y-ions at almost every residue. Additionally, there are several d- and w-ions. There are proportionally fewer of these ion types in the whole protein spectra. This is likely due to the large energy required to fragment a protein compared to a peptide. Internal fragments were also identified in the peptide. Due to the smaller size and less populated spectrum, internal ions can be assigned with more confidence than in large proteins. Most of the assigned internal ions are cz ions. The intensity of the internal ions

is quite low, and the cleavages that would result in the corresponding terminal ions, i.e. the c- and z-ions, had quite high intensities. This helps to explain why internal ions are not observed in intact proteins in ETD experiments. Even in a comparatively smaller molecule, the intensities of the internal ions observed are much lower than terminal ions, and these internal ions are only observed at residues that feature the most abundant terminal ions. When the size of the analyte is increased to the size of an intact protein, the ion count will be split between many channels, diluting signal. Additionally, spectra generated from fragmenting a whole protein will feature many more peaks than a peptide, making accurate deconvolution and subsequent peak assignment. These reasons may explain why internal fragments may be observed in peptides but not intact proteins.

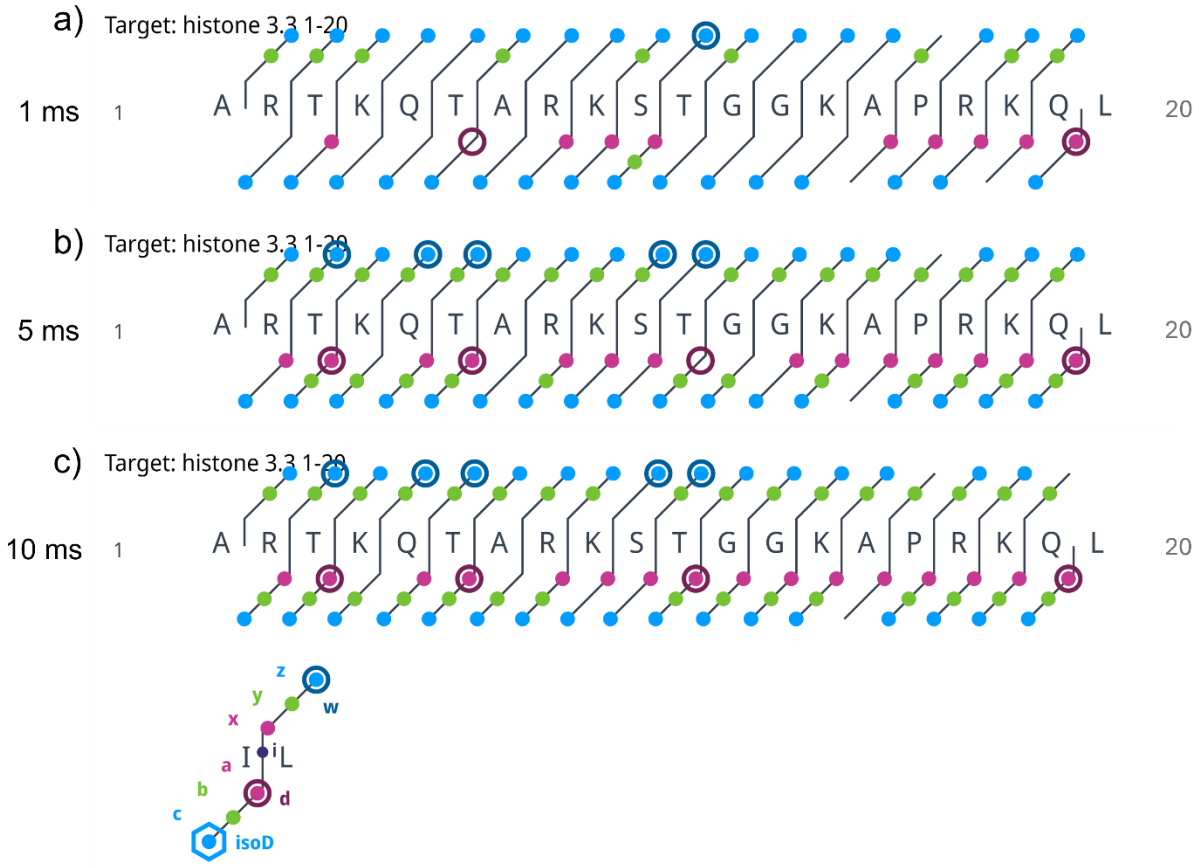


Figure 5.3 ETD Fragmentation of the 6+ charge state of peptide from histone protein 3 (1-20). Activation times of a) 1ms, b) 5 ms, and c) 10 ms.

Another possibility is that there are additional terminal fragments that are not assignable that are attributed to internal fragments. When a terminal fragment is generated, a functional group that differs from a protein terminus is left exposed. While these may behave similarly to the termini in an intact protein, they also may not. Further investigation of these possibilities should be pursued to determine whether fragments produced from multiple cleavage events behave differently than those produced from a single cleavage. To assign internal ions confidently, they should first be compared to

assigned terminal ions. Any terminal ions missing a complement ion should be identified and unassigned ions should be compared to the expected terminal ion to see if any losses would explain the discrepancy.

ClipsMS

Top-down HCD and UVPD data was analyzed to determine if internal fragments generated can be confidently assigned. Four proteins were examined: myoglobin, carbonic anhydrase, aldolase, and enolase. These proteins ranged in size from 16 to 46 kDa. Mass lists were generated by removing any assigned terminal ions. The mass lists were then searched against the correct sequence followed by three shuffled sequences. The results for the HCD data are shown in Fig. 3. The number of ions assigned as internal ions for the wrong sequence increased with protein size for HCD data, as well as the proportion of the mass list assigned.

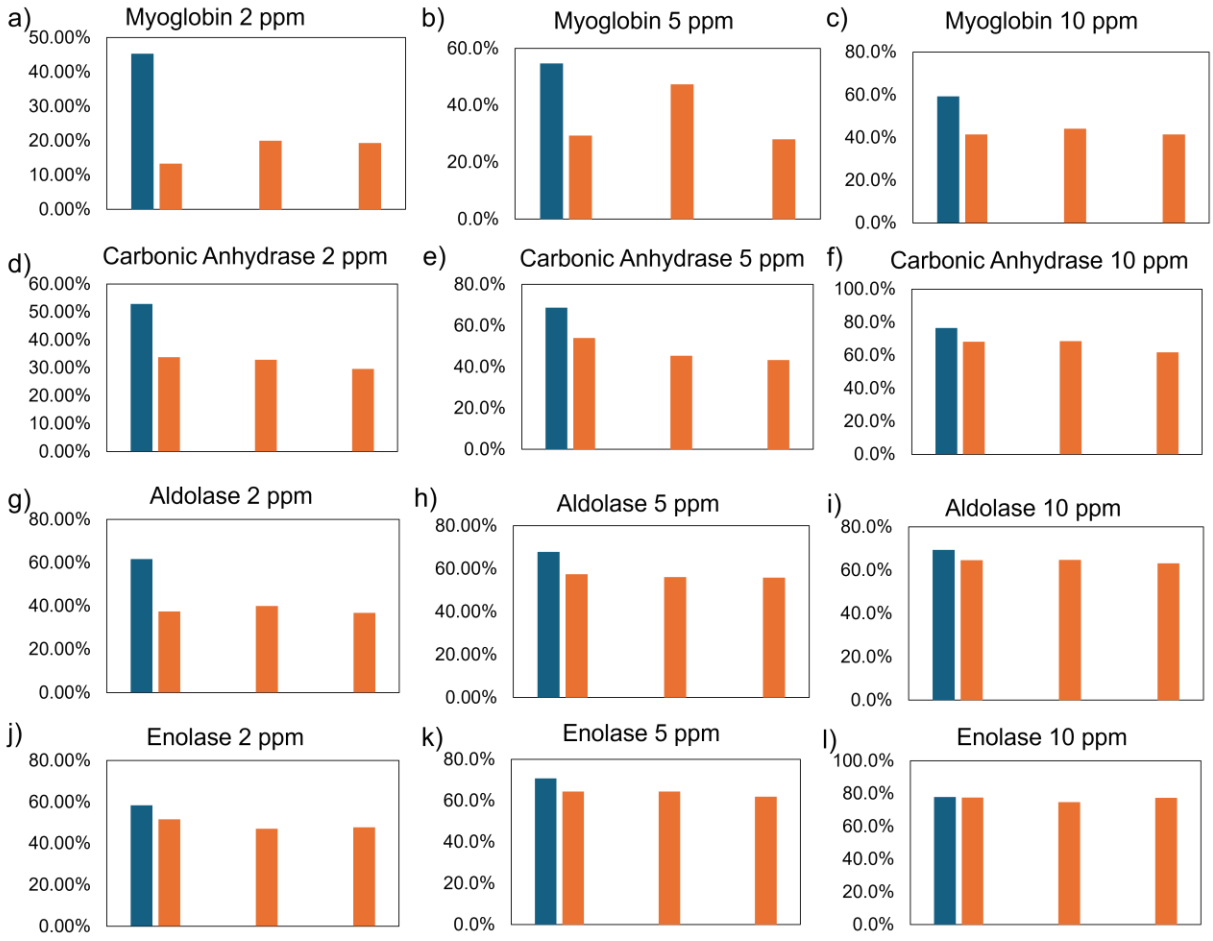


Figure 5.4 Percentages of non-terminal ions assigned with the correct sequence (blue bar) and incorrect sequences (orange bars). Error tolerances were 2, 5, and 10 ppm; results are from HCD data from a-c) myoglobin, d-f) carbonic anhydrase, g-k) aldolase, and j-l) enolase.

The lowest percentage of fragments assigned with the wrong sequence across all data was the myoglobin HCD data. When searched against the correct sequence with an error tolerance of 2 ppm, 45% of the mass list is assigned as internal ions, whereas the shuffled sequences only resulted in approximately 10-20% of the mass list being assigned. Myoglobin also had the most variation in results across the shuffled sequences

for HCD data. This result is likely due to smaller mass list and shorter sequence relative to the other proteins examined, although as error tolerance is increased to 10 ppm, variation decreases (Fig 4c). As protein size increases, the number of possible internal fragments increases, resulting in a larger proportion of the mass list being assigned. The HCD data for enolase featured the greatest similarity between the correct sequence and shuffled sequences. The correct sequence with an error tolerance of 2 ppm resulted in 58% of the mass list being assigned, while the shuffled sequences resulted in an average of 48% being assigned. The two proteins examined whose mass fell between, carbonic anhydrase and aldolase, featured percentages between myoglobin and enolase. The carbonic anhydrase analysis featured 53% of the mass list assigned with the correct sequence and an average of 32% assigned with the shuffled sequences, and aldolase resulted in 62% and 38% being assigned, respectively. The mass list also increased with protein size. In general, the number of assignments with both the correct and shuffled sequences increased with error tolerance. The ratio between the percent of the mass list assigned with shuffled sequences and the percent assigned with the correct sequence plotted against protein size is shown in Fig. 5. There is also a slight negative correlation between standard deviation of the shuffled assignments and protein size observed in the HCD data at error tolerances of 2 and 5 ppm, but by 10 ppm there is no trend and the differences between standard deviations becomes small. This is likely due to the increase in the mass list with protein size, and the number of assignments made with an increasing error tolerance. Previous work has shown there is not a strong correlation between number of observed internal ions and protein size with regard to HCD fragmentation.

However, the number of possible internal ions increases with protein size, making it more likely to incorrectly assign any observed non-terminal ion for larger proteins.

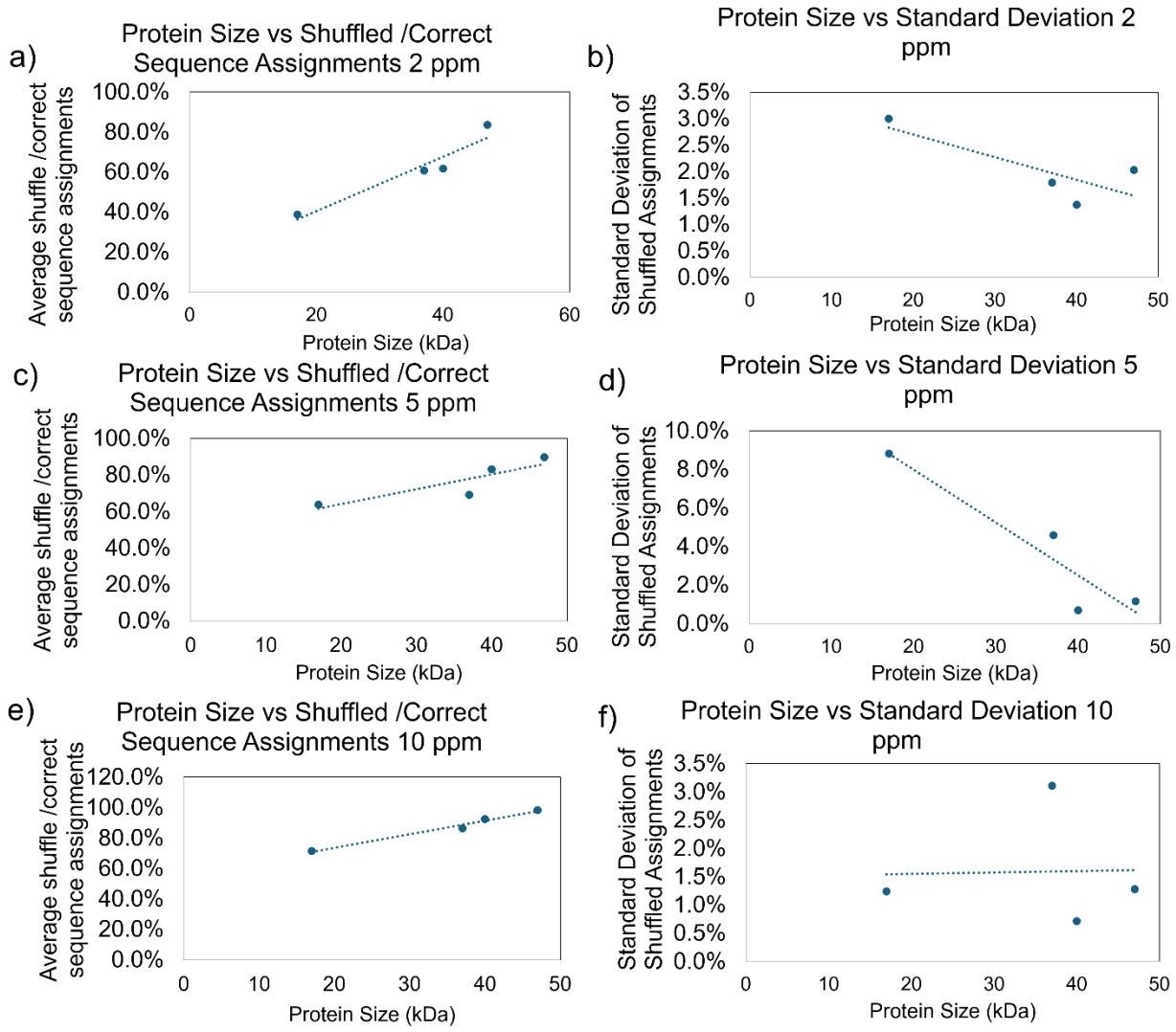


Figure 5.5 For HCD data, there are trends based on protein size. The proportion of ions assigned using the shuffled sequences divided by assignments using the correct sequence plotted against protein size and the standard deviation of the shuffled sequences plotted against protein size for a, b) 2 ppm error tolerance; c, d) 5 ppm, and e, f) 10 ppm.

The percentage of ions assigned followed the same trend for PD data, which is illustrated in Fig. 5. The lowest proportion of ions assigned with the shuffled sequence was again in the myoglobin data with an average of 15% of the ions assigned with a 2 ppm error tolerance, compared to 25% with the correct sequence. Aldolase and enolase data featured an average of 46% and 55% of ions assigned with shuffled sequences and 56% and 67% of ions assigned with the correct sequence, respectively. Carbonic anhydrase featured an average of 27% of ions assigned with shuffled sequences, and 23% assigned with the correct sequence. However, it should be noted that aldolase and enolase PD data was charge reduced using proton transfer, which helps to reduce the number of ambiguous peaks, while carbonic anhydrase and myoglobin were not. Once again, the mass list increased with protein size, and the proportion of shuffled assignments to assignments with the correct sequence loosely tracked with protein size (Fig. 7), although not as well as the HCD data. The percentages of masses assigned again increased for both the correct sequences and the shuffled sequences at a 2 ppm error tolerance, but any trend disappears when error tolerance is increased to 5 ppm. The standard deviation was again largest in the smallest protein, myoglobin, but did not feature a trend like the HCD data. There was not a trend in standard deviation at a 5 ppm error tolerance, but at 10 ppm, standard deviation had a loose negative correlation with size, although the differences between standard deviations was quite small.

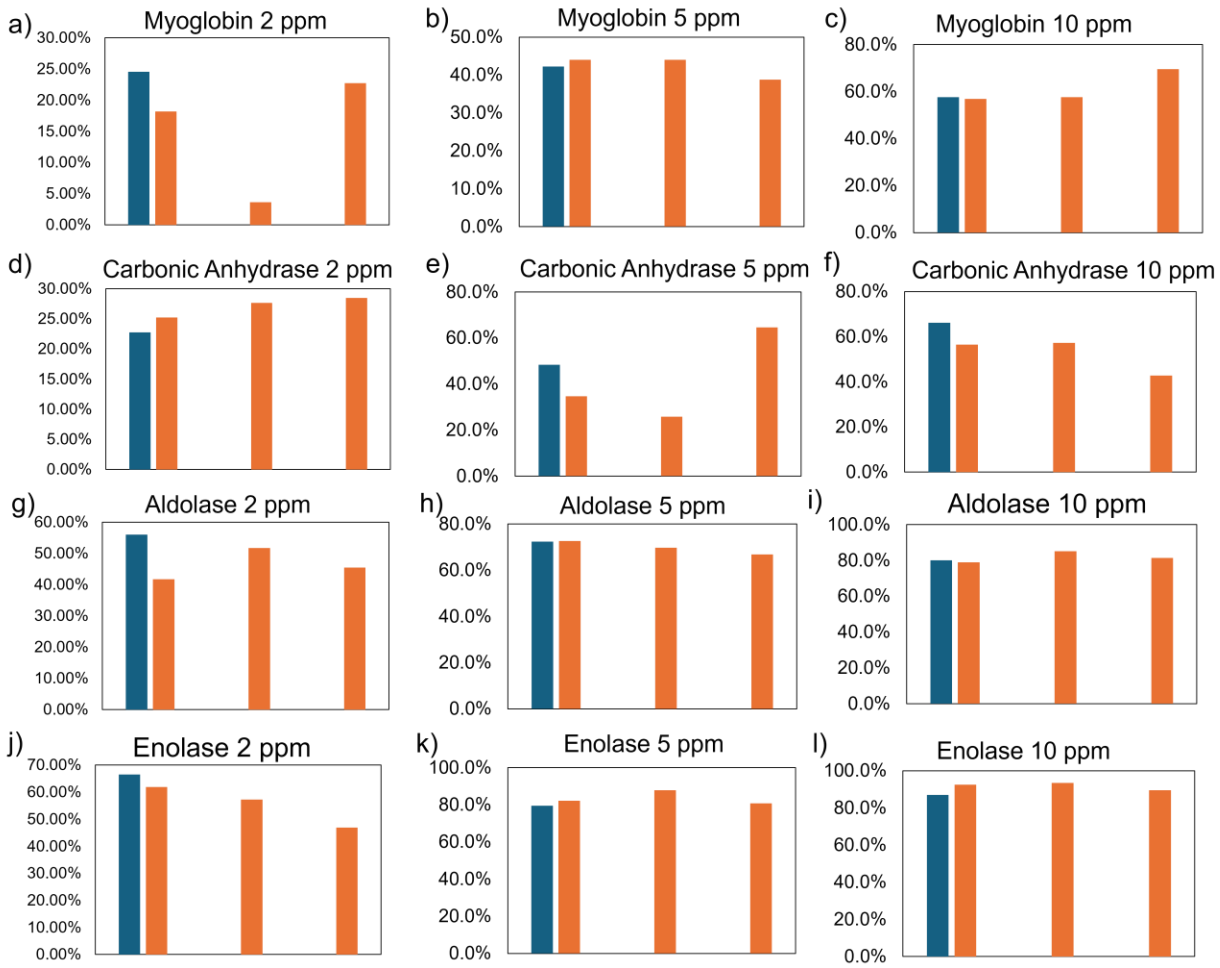


Figure 5.6 Percentages of non-terminal ions assigned with the correct sequence (blue bar) and incorrect sequences (orange bars). Error tolerances were 2, 5, and 10 ppm; results are from 193 nm PD data from a-c) myoglobin, d-f) carbonic anhydrase, g-k) aldolase, and j-l) enolase.

One possible reason for the amount of variation in the results relative to HCD data is the increased number of ion types in UVPD data. While internal ions in peptides observed from collisional activation have been identified as primarily ay and by ions, UVPD features c, x, and z ions, which could be further fragmented to form ax, az, bx, and other ion types.[29]

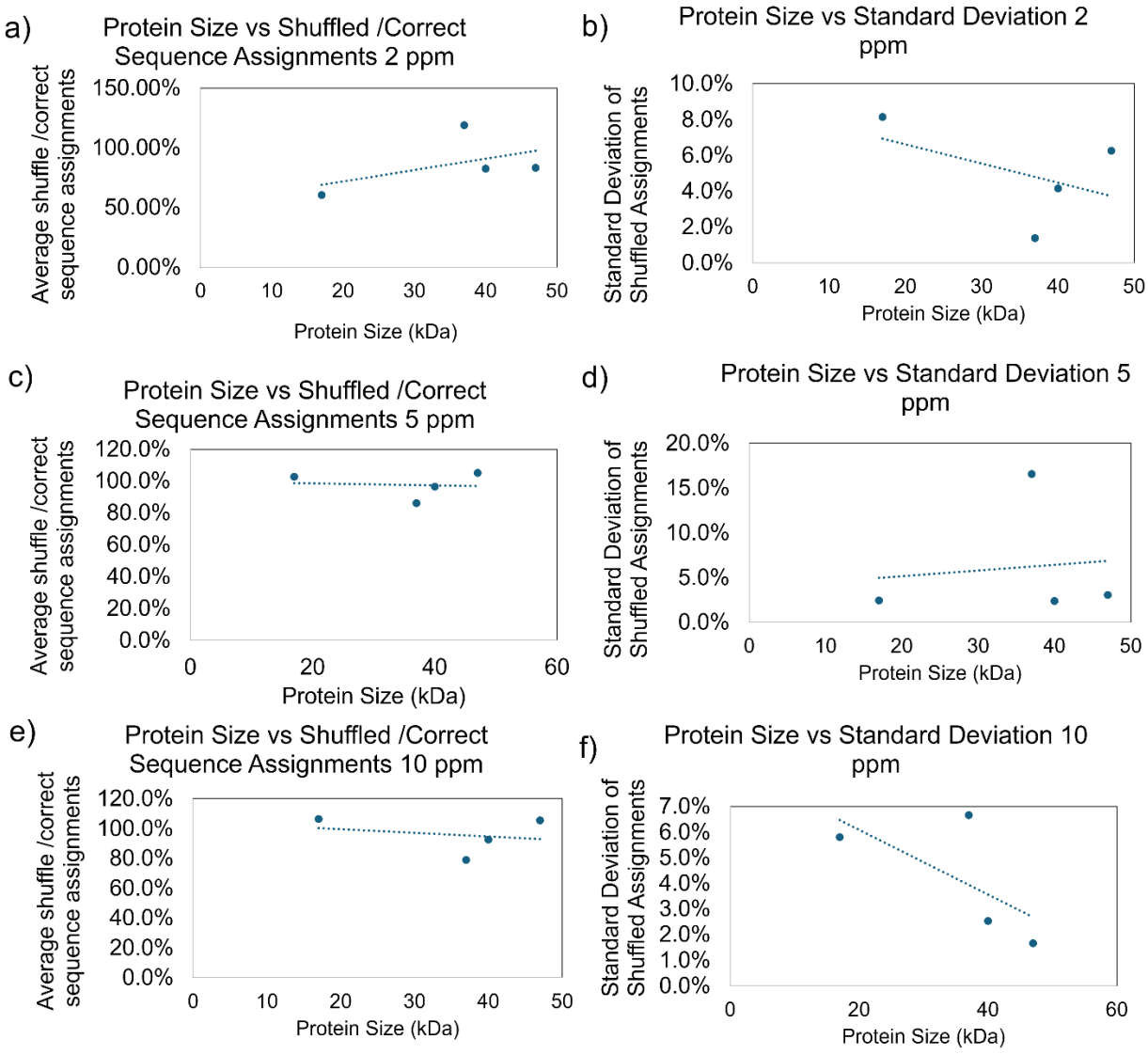


Figure 5.7 For UVPD data, trends based on protein size are less apparent. The proportion of ions assigned using the shuffled sequences divided by assignments using the correct sequence plotted against protein size and the standard deviation of the shuffled sequences plotted against protein size for a, b) 2 ppm error tolerance: c, d) 5 ppm, and e, f) 10 ppm.

UVPD fragmentation proceeds via two pathways. One of these is direct dissociation upon photon absorption. The other is internal conversion upon photon absorption, which causes ion heating and subsequent dissociation similar to collisional

activation. If the number of direct dissociations is too high relative to internal conversion, ions will be ripped into many pieces, and terminal ions will not be assignable. UVPD at 213 nm has some direct dissociation pathways, and 193 nm has many more due to peptide bond absorption.[30] This means there are likely some internal ions present in 193 nm data, however the statistical probability of correctly assigning is quite low and becomes lower with increasing protein size as shown above.

Part of the difficulty with correctly assigning internal ions, is due to the limited number of m/z values that can be resolved before there are overlapping, or isobaric, peaks. If a spectrum were shown with a peak perfectly resolved for every 5 ppm, there would be a finite number of peaks populating the spectrum. If more peaks were added, they would begin overlapping one another. This issue is further compounded in proteins, where masses are populated in clusters, with gaps between that have m/z values that are never observed.[31,32]

Conclusion

Ions that are not assigned as terminal ions are prone to error when assigned as internal ions using current techniques, even when stringent mass error tolerance is used. The likelihood of error is somewhat dependent on protein size and activation method. The proportion of ions that are not assignable as terminal ions increases with ETD activation time. The proportion of ions with a complement pair decreases proportionally as unassignable ions increases. The lack of z-ions compared to corresponding c-ions in ETD experiments suggests further fragmentation of z-ions. Further investigation

experimentally and computationally should be performed to ensure correct assignment of ions that cannot be assigned as regular terminal ions.

Supporting Information

Table S 5.1 Number of pairs of unassigned ions from ETD experiments on histone protein 3.

	18+					
activation time (ms)	Ion Pairs	Assigned fragments	% pairs	Total fragments	% assigned	% unassigned
1	27	192	28.1%	311	61.7%	38.3%
2	64	426	30.0%	550	77.5%	22.5%
3	54	387	27.9%	526	73.6%	26.4%
5	57	451	25.3%	637	70.8%	29.2%
10	54	577	18.7%	906	63.7%	36.3%
	24+					
activation time (ms)	Ion Pairs	Assigned fragments	% pairs	Total fragments	% assigned	% unassigned
1	32	260	24.6%	314	82.8%	17.2%
2	54	372	29.0%	439	84.7%	15.3%
3	41	308	26.6%	389	79.2%	20.8%
5	38	383	19.8%	467	82.0%	18.0%
	22+					
activation time (ms)	Ion Pairs	Assigned fragments	% pairs	Total fragments	% assigned	% unassigned
0.5	0	33	0	59	55.9%	44.1%
1	39	232	33.6%	314	73.9%	26.1%
2	62	389	31.9%	482	80.7%	19.3%
3	58	422	27.5%	534	79.0%	21.0%
5	38	383	19.8%	511	75.0%	25.0%
10	18	466	7.7%	673	69.2%	30.8%

References

[1] Fornelli, L.; Srzentić, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **2018**, *90* (14), 8421–8429. https://doi.org/10.1021/acs.analchem.8b00984.

[2] Greisch, J.-F.; Tamara, S.; A. Scheltema, R.; R. Maxwell, H. W.; D. Fagerlund, R.; C. Fineran, P.; Tetter, S.; Hilvert, D.; R. Heck, A. J. Expanding the Mass Range for UVPD-Based Native Top-down Mass Spectrometry. *Chemical Science* **2019**, *10* (30), 7163–7171. https://doi.org/10.1039/C9SC01857C.

[3] Zhang, H.; Ge, Y. Comprehensive Analysis of Protein Modifications by Top-Down Mass Spectrometry. *Circulation: Cardiovascular Genetics* **2011**, *4* (6), 711–711. https://doi.org/10.1161/CIRCGENETICS.110.957829.

[4] Liu, R.; Xia, S.; Li, H. Native Top-down Mass Spectrometry for Higher-Order Structural Characterization of Proteins and Complexes. *Mass Spectrometry Reviews* **2023**, *42* (5), 1876–1926. https://doi.org/10.1002/mas.21793.

[5] Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annual Review of Analytical Chemistry* **2016**, *9* (Volume 9, 2016), 499–519. https://doi.org/10.1146/annurev-anchem-071015-041550.

[6] Zenaidee, M. A.; Lantz, C.; Perkins, T.; Jung, W.; Loo, R. R. O.; Loo, J. A. Internal Fragments Generated by Electron Ionization Dissociation Enhance Protein Top-Down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (9), 1896–1902. https://doi.org/10.1021/jasms.0c00160.

[7] Beaumal, C.; Deslignière, E.; Diemer, H.; Carapito, C.; Cianférani, S.; Hernandez-Alba, O. Improved Characterization of Trastuzumab Deruxtecan with PTCR and Internal Fragments Implemented in Middle-down MS Workflows. *Anal Bioanal Chem* **2024**, *416* (2), 519–532. https://doi.org/10.1007/s00216-023-05059-x.

[8] Rahman, M.; Marzullo, B.; Holman, S. W.; Barrow, M.; Ray, A. D.; O'Connor, P. B. Advancing PROTAC Characterization: Structural Insights through Adducts and Multimodal Tandem-MS Strategies. *J. Am. Soc. Mass Spectrom.* **2024**, *35* (2), 285–299. https://doi.org/10.1021/jasms.3c00342.

[9] Kenderdine, T.; McIntyre, W.; Yassaghi, G.; Rollo, D.; Bunkowski, A.; Goerlach, L.; Suckau, D.; Tremintin, G.; Greig, M.; Bell, C.; Fabris, D. Integrating Internal Fragments in the Interpretation of Top-Down Sequencing Data of Larger Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **2023**, *34* (10), 2296–2307. https://doi.org/10.1021/jasms.3c00207.

[10] Schmitt, N. D.; Berger, J. M.; Conway, J. B.; Agar, J. N. Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem.* **2021**, *93* (16), 6355–6362. https://doi.org/10.1021/acs.analchem.0c04670.

[11] 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. https://doi.org/10.1021/acs.jproteome.0c00952.

[12] Zhu, Y.; Liu, Z.; Liu, J.; Zhao, H.; Feng, R.; Shu, K.; Wang, F.; Chang, C. Panda-UV Unlocks Deeper Protein Characterization with Internal Fragments in Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2024**. https://doi.org/10.1021/acs.analchem.4c00253.

[13] Larson, E. J.; Pergande, M. R.; Moss, M. E.; Rossler, K. J.; Wenger, R. K.; Krichel, B.; Josyer, H.; Melby, J. A.; Roberts, D. S.; Pike, K.; Shi, Z.; Chan, H.-J.; Knight, B.; Rogers, H. T.; Brown, K. A.; Ong, I. M.; Jeong, K.; Marty, M. T.; McIlwain, S. J.; Ge, Y. MASH Native: A Unified Solution for Native Top-down Proteomics Data Processing. *Bioinformatics* **2023**, *39* (6), btad359. https://doi.org/10.1093/bioinformatics/btad359.

[14] Okuda, S.; Watanabe, Y.; Moriya, Y.; Kawano, S.; Yamamoto, T.; Matsumoto, M.; Takami, T.; Kobayashi, D.; Araki, N.; Yoshizawa, A. C.; Tabata, T.; Sugiyama, N.; Goto, S.; Ishihama, Y. jPOSTrepo: An International Standard Data Repository for Proteomes. *Nucleic Acids Research* **2017**, *45* (D1), D1107–D1111. https://doi.org/10.1093/nar/gkw1080.

[15] Dunham, S. D.; Wei, B.; Lantz, C.; Loo, J. A.; Brodbelt, J. S. Impact of Internal Fragments on Top-Down Analysis of Intact Proteins by 193 Nm UVPD. *J. Proteome Res.* **2023**, *22* (1), 170–181. https://doi.org/10.1021/acs.jproteome.2c00583.

[16] Gunthals, A.; Sturgeon, D.; Gavrilenko, A.; Hakkila, B.; Sturgeon, S.; Gavrilenko, J.; Franklin, R.; Vasil'ev, Y.; Meeuwsen, J. ; Voinov, V. ; Beckman, J. Parameter-free Deconvolution and Visualization of Peptide and Protein Fragmentation Mass Spectra, 312344. Proceedings of the 71st ASMS Conference on Mass Spectrometry and Allied Topics, Houston, TX (June 2023).

[17] Stothard, P. The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *BioTechniques* **2000**, *28* (6), 1102–1104. https://doi.org/10.2144/00286ir01.

[18] Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2006**, *1764* (12), 1811–1822. https://doi.org/10.1016/j.bbapap.2006.10.003.

[19] Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors. *Anal. Chem.* **2007**, *79* (2), 477–485. https://doi.org/10.1021/ac061457f.

[20] 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. https://doi.org/10.1007/s13361-017-1823-8.

[21] Mikawy, N. N.; Rojas Ramírez, C.; DeFiglia, S. A.; Szot, C. W.; Le, J.; Lantz, C.; Wei, B.; Zenaidee, M. A.; Blakney, G. T.; Nesvizhskii, A. I.; Loo, J. A.; Ruotolo, B. T.; Shabanowitz, J.; Anderson, L. C.; Håkansson, K. Are Internal Fragments Observable in Electron Based Top-Down Mass Spectrometry? *Molecular & Cellular Proteomics* **2024**, 100814. https://doi.org/10.1016/j.mcpro.2024.100814.

[22] Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proceedings of the National Academy of Sciences* **2004**, *101* (26), 9528–9533. https://doi.org/10.1073/pnas.0402700101.

[23] J. M. Kempkes, L.; Martens, J.; Berden, G.; J. Houthuijs, K.; Oomens, J. Investigation of the Position of the Radical in z 3 -Ions Resulting from Electron Transfer Dissociation Using Infrared Ion Spectroscopy. *Faraday Discussions* **2019**, *217* (0), 434–452. https://doi.org/10.1039/C8FD00202A.

[24] Nguyen, H. T. H.; Shaffer, C. J.; Pepin, R.; Tureček, F. UV Action Spectroscopy of Gas-Phase Peptide Radicals. *J. Phys. Chem. Lett.* **2015**, *6* (23), 4722–4727. https://doi.org/10.1021/acs.jpclett.5b02354.

[25] Imaoka, N.; Houferak, C.; Murphy, M. P.; Nguyen, H. T. H.; Dang, A.; Tureček, F. Spontaneous Isomerization of Peptide Cation Radicals Following Electron Transfer Dissociation Revealed by UV-Vis Photodissociation Action Spectroscopy. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (9), 1768–1780. https://doi.org/10.1007/s13361-017-1871-0.

[26] Han, H.; Xia, Y.; McLuckey, S. A. Ion Trap Collisional Activation of c and Z• Ions Formed via Gas-Phase Ion/Ion Electron-Transfer Dissociation. *J. Proteome Res.* **2007**, *6* (8), 3062–3069. https://doi.org/10.1021/pr070177t.

[27] Chung, T. W.; Tureček, F. Backbone and Side-Chain Specific Dissociations of z Ions from Non-Tryptic Peptides. *J Am Soc Mass Spectrom* **2010**, *21* (8), 1279–1295. https://doi.org/10.1016/j.jasms.2010.02.018.

[28] Kjeldsen, F.; Haselmann, K. F.; Budnik, B. A.; Jensen, F.; Zubarev, R. A. Dissociative Capture of Hot (3–13 eV) Electrons by Polypeptide Polycations: An Efficient Process Accompanied by Secondary Fragmentation. *Chemical Physics Letters* **2002**, *356* (3), 201–206. https://doi.org/10.1016/S0009-2614(02)00149-5.

[29] 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. https://doi.org/10.1016/0168-1176(91)85054-P.

[30] R. Julian, R. The Mechanism Behind Top-Down UVPD Experiments: Making Sense of Apparent Contradictions. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (9), 1823–1826. https://doi.org/10.1007/s13361-017-1721-0.

[31] Wolski, W. E.; Farrow, M.; Emde, A.-K.; Lehrach, H.; Lalowski, M.; Reinert, K. Analytical Model of Peptide Mass Cluster Centres with Applications. *Proteome Sci* **2006**, *4* (1), 18. https://doi.org/10.1186/1477-5956-4-18.

[32] Zubarev, R.; Artemenko, K.; Zubarev, A.; Mayrhofer, C.; Yang, H.; Fung, Y. Early Life Relict Feature in Peptide Mass Distribution. *Open Life Sciences* **2010**, *5* (2), 190–196. https://doi.org/10.2478/s11535-009-0069-2.

Chapter 6

Concluding Remarks

Biomolecular structure analysis is an essential requirement for understanding the workings of biology, and how it changes in aging and disease. Through continued development of innovative analytical methods, we can further advance our knowledge and, from this perhaps develop effective treatment for neurodegenerative disorders. In this dissertation, mass spectrometry-based methods are described for a variety of applications related to understanding protein structure and properties.

In short, a mass-spectrometry method to detect modifications arising from spontaneous chemistry and post-translational modifications in recombinant protein and human brain tissue is described. *In vitro* aging of recombinant protein revealed differential rates of modification in soluble and aggregated protein. This method will also enable future studies to obtain data from samples collected from brains with neurodegenerative disease and controls to establish statistically significant differences between protein modifications the two groups. These results will help to reveal the potential role that spontaneous chemical modifications play in neurodegeneration and aging.

We also examined how isomers in peptides affected gas phase structure. By comparing chromophore absorption and ion-molecule reactions, peptides containing Asp, Ser, and Glu isomers were differentiated via simple UVPD experiments. Molecular modeling was also used to propose potential differing structures caused by the isomerization. These findings illustrate the structural impact that isomerization at one

amino acid residue can induce and show that mass spectrometry can reveal information about isomers even when no detectable mass shift is present.

Next, we used RDD and UVPD experiments to analyze the structure of intact proteins. A single tyrosine residue was covalently modified with iodine to introduce a radical precursor. These studies revealed structural differences, including between highly charged proteins, and apo- and holo-forms of protein. This type of analysis may provide structural information that cannot be obtained from other analytical techniques and can serve as validation for structural predictions. Additionally, it can further our understanding of protein behavior in the gas phase.

Lastly, we performed an analysis of 193 nm UVPD and HCD data to illustrate that current methods of assigning internal fragments may not provide results that are statistically confident. We showed that there is a large amount of overlap between assignments made with a search against the correct sequence, and assignments made with an incorrect sequence. Additionally, ETD experiments were performed on the intact histone protein 3 to explore why complement ion pairs are not frequently observed in top-down experiments. Ion complement pairs and the proportion of assignable terminal ions declined with more activation. This suggests that ions may not behave as expected when multiple cleavages occur, or terminal ions may be subject to further activation that has not been explored.

In summary, the work described in this dissertation illustrates the indisputable utility of mass spectrometry as an analytical tool for studying proteins and peptides. However, it still leaves unanswered questions that will be answered by the continuation

of the work presented here. One such question is the role of spontaneous chemical modifications in aging and neurodegeneration. By applying the method described in Chapter 2 to more samples, from a variety of conditions and brain regions, statistical power will reveal meaningful differences between healthy brains and those afflicted with neurodegeneration.

Other questions that have arisen during the scope of this work include are internal fragments really what we think they are? And if not, what could they be? Further experimental and theoretical investigations will expand our knowledge of ion behavior, and thus, further our ability to elucidate biomolecular structural information from mass spectrometry data. In conclusion, the work presented here and future work stemming from such questions will serve to expand our knowledge of structural biology and mass spectrometry capabilities.