UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mass Spectrometry-Based Techniques for Studying Isomerization and Protein Structure

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

Doctor of Philosophy

in

Chemistry

by

Jacob W. Silzel

September 2023

Dissertation Committee:
Dr. Ryan Julian, Chairperson
Dr. Joseph Genereux
Dr. Linlin Zhao

The Dissertation of Jacob W. Silzel is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

There are a lot of people I owe thanks to for supporting my journey as a chemist. First, I would like to thank my parents, John and Lisa, who not only provided a fundamental foundation of math and science knowledge that prepared me well for college and grad school, but have also supported and encouraged me every step of the way. I'd like to thank my sisters Emily, Faith, and Grace as well as my grandparents Bing, Barbara, Wayne, and Mary for all the support and encouragement I've received from them as well. I also owe a lot to my professors from Biola University, in particular Drs. Jonah Chang, Dana Johnson, Jessica Lu, and Xidong Chen. Without mentorship, guidance, and encouragement from these individuals I would not be the chemist I am today.

I would like to thank my wife Courtney for all the support and love she has shown me. Even when we were just dating and engaged, she cared for me and supported me. I couldn't ask for a better spouse than her, and every day I am thankful I get to be married to her. I would also like to thank my Lord and savior Christ Jesus, by whom all things were created, for loving me before I loved Him, and for the salvation He has provided me. Completing a PhD in chemistry is one of many blessings He has given me.

I would like to thank all the amazing scientists I met in the Julian lab, who all contributed to my studies and helped and encouraged me in many ways: James Bonner, Lance Talbert, Yana Lyon, Dylan Riggs, Tyler Lambeth, Ting Ting Wu, Brielle Van Orman, Evan Hubbard, Gaurav Pandey, Sandeep Kumar, Thomas Shoff, Isabelle Cordill, Sharon Chen, and Lin He. A special thanks goes to Julian lab alums Tyler Lambeth, Dylan Riggs, Yana Lyon, and Lance Talbert for fielding a deluge of questions from me

when I started in the lab and bringing me up to speed on many of the techniques and knowledge necessary. I would also like to thank my labmate Evan Hubbard for his friendship, grad school wouldn't have been the same without him. Among others at UCR, I would like to thank my friends Dr. Drew Kellum and Jin Tang. A special thanks goes to my committee members Dr. Joey Genereux and Dr. Linlin Zhao as well as outside members Dr. Vince Lavalo and Dr. Sean O'Leary for guidance and recommendations during my SYRE, orals, and my defense.

Finally, I would like to thank my PI Dr. Ryan Julian. Ryan pushed me to become a better scientist and provided many opportunities to grow and develop as a chemist. I am truly thankful that Ryan brought me into his lab and allowed me to work on such interesting and engaging projects. There are few people in my life outside of family who have had such a large impact on who I am, and Ryan is one of them. I would not be the chemist, mass spectrometrist, writer, presenter, or scientist in general that I am today without Ryan, and for that I will be thankful for the rest of my life.

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

Chapter 2: Silzel, J. W.; Murphree, T. A.; Paranji, R. K.; Guttman, M. M.; Julian, R. R. Probing the Stability of Proline Cis/Trans Isomers in the Gas Phase with Ultraviolet Photodissociation. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (9), 1974–1980.

Chapter 3: Silzel, J. W.; Lambeth, T. R.; Julian, R. R. PIMT-Mediated Labeling of L-Isoaspartic Acid with Tris Facilitates Identification of Isomerization Sites in Long-Lived Proteins. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (3), 548–556.

Chapter 4: Silzel, J. W.; Ben-Nissan, G.; Tang, J.; Sharon, M.; Julian, R. R. Influence of Asp Isomerization on Trypsin and Trypsin-like Proteolysis. *Anal. Chem.* **2022**, *94* (44), 15288–15296.

Chapter 5: Silzel, J. W.; Julian, R. R. RDD-HCD Provides Variable Fragmentation Routes Dictated by Radical Stability. *J. Am. Soc. Mass Spectrom.* **2023**, *34* (3), 452–458.

The co-author Ryan R. Julian listed in these publications directed and supervised the research which forms the basis for this dissertation. All other authors provided technical expertise and/or performed specific experiments.

ABSTRACT OF THE DISSERTATION

Mass Spectrometry-Based Techniques for Studying Isomerization and Protein Structure

by

Jacob W. Silzel

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, September 2023
Dr. Ryan Julian, Chairperson

The proper structure and function of proteins in living organisms is essential to life. Even small changes to the amino acid residues that make up proteins can drastically alter both structure and function. As proteins age, they accumulate post-translational modifications and spontaneous chemical modifications such as isomerization, which can drastically affect both structure and function. In long-lived proteins (LLPs) where little to no turnover occurs, significant amounts of isomerization can build up over time. Long-lived proteins have been implicated in a growing number of human health conditions. Thus, understanding the relationship between these LLPs and isomerization is growing in importance. Additionally, the connection between isomerization and protein structure/function is complex and requires further unraveling to understand completely. In this work, we detail new mass spectrometry methods for identifying isomerization, and explore the consequences of isomerization on enzymatic digestion. First, we start by

examining cis/trans proline isomerization and establish that mass spectrometry can distinguish these isomers if proper care is taken to select the appropriate ESI conditions along with charge solvation to preserve solution-state structure into the gas phase. Next, we detail a L-isoAsp-specific covalent labeling technique and validate this method on a 72-year-old eye lens sample. This technique allows major isomerization sites to be quickly and easily identified through introduction of a mass shift which can be searched for just like a regular post-translational modification. Following that, the consequences of Asp isomerization in substrates of trypsin-like proteases are studied, revealing that even robust enzymes such as trypsin are unable to recognize isomerized residues in the active site. In addition, the 20S proteasome is found to be unable to digest isomerized substrate, carrying important implications in the ultimate fate of isomerized LLPs. Finally, we examine the effects of using higher-energy collisional dissociation (HCD) to excite the hydrogen-deficient radical formed during radical-directed dissociation (RDD). RDD-HCD was found to create different fragments depending on the amount of HCD energy used and revealed insight into radical-based fragmentation. Overall, this work highlights mass spectrometry as a powerful tool to study the structural and functional consequences of isomerization in proteins.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Introduction

Protein Structure and Function

For thousands of years the domestic horse provided critical functions within human civilization. From agriculture to transportation to military warfare, the roles horses played are as important as they are varied, and horses were carefully bred for specific traits that made them more desirable in their intended role.[1] In that sense, it is no overstatement to say that proteins are the workhorses of the cell, and they have an even greater level of importance in living organisms than the horse did for human civilization. Proteins are a very diverse group of biomolecules, possessing a very wide range of different roles to play in the cell. These roles are tied very closely to the individual protein's structure, which is inseparable from its' function. Proteins are large polypeptide chains made up of amino acid subunits, of which there are 20 canonical amino acids. Each amino acid side chain offers different chemical options, such as basicity, acidity, hydrophobicity, or hydrophilicity. With 20 available subunits to choose from with such wide-ranging differences in chemical properties, the number of different combinations available is vast. These basic building blocks determine a lot about what the overall protein structure will look like and how the protein will function. However, the sequence of a protein alone is only part of the picture when it comes to protein structure.

Protein structure is more complicated than simply the amino acid sequence, or primary structure, and there are several different levels of protein structure beyond primary. When amino acids are combined into a polypeptide chain, the interactions between the

backbone carbonyl oxygens and backbone hydrogens result in H-bonding that causes structural rearrangement. The most common shapes are alpha-helices and beta-sheets, which are known as secondary structure. As sections of the protein backbone form these shapes, amino acid side chains also interact with each other to influence the overall structure of the protein. Hydrophobic residues cluster together to avoid solvent exposure while hydrophilic residues seek water. Positively charged residues form salt bridges with negatively charged ones. Cysteine residues form disulfide bonds, linking entire sections of the protein together in a large cycle. These interactions between side chains greatly influence the overall folding of the protein and are known as tertiary structure. While many proteins' structures stop at tertiary, a protein with multiple subunits possesses even higher order quaternary structure, which is the overall 3D conformation adopted by the complex as the individual subunits interact noncovalently with each other.

For a protein, possessing the proper structure is essential to functioning correctly, and if a protein becomes misfolded, then the original function is often either diminished or lost. Not only do misfolded proteins lose their intended functions, they can also be toxic and lead to disease. For example, mutations in human lysozyme can lead to greater structural instability, leading to unfolding, which not only results in loss of function but also causes toxic aggregate species to form.[2] Amyloid beta and tau, both intrinsically disordered, are known to aggregate and form plaques and neurofibrillary tangles, and these proteins are both implicated at the center of neurodegenerative diseases such as Alzheimer's disease.[3,4] It is clear that when proteins lack structure, they can pose a threat to living

organisms, highlighting the importance of maintaining proper protein structure as well as the importance of understanding modifications that can lead to a loss of structure.

Post-Translational Modifications

During their residence time in a living organism, proteins can undergo many post-translational modifications (PTMs). Post-translational modifications can be defined as chemical modification of specific residues, and there are more than 400 different types of PTMs that proteins can experience.[5] Some of the most common PTMs are acetylation, phosphorylation, methylation, and ubiquitination. Most PTMs are enzymatically driven and act as mechanisms through which protein function can be tightly regulated, or new functionality imparted to the protein. For example, acetylation is carried out by N-terminal and lysine acetyltransferases and has a wide range of consequences, such as allowing interactions with the Golgi apparatus, prevention of migration from the cytosol to the endoplasmic reticulum, and modulating protein-protein interactions.[6] N-terminal acetylation is also important for maintaining protein structure by stabilizing flexible N-termini and preventing misfolding and aggregation.[7] In the case of phosphorylation, enzymes called kinases recognize and phosphorylate specific sites on proteins, and phosphorylation acts as a mechanism through which protein structure and function can be regulated.[8] Phosphorylation is reversible, and can act as a molecular switch for changing enzyme functions in response to metabolism changes.[9] Phosphorylation also results in a change in structure in the protein itself, which causes the change in function. For example, phosphorylation of cardiac myosin-binding protein C causes eight specific

residues normally buried in an alpha-helix to become more exposed, which may be critical for binding during normal cardiac function.[10] Phosphorylation can also serve as an important disease biomarker, such as in the case of Tau hyperphosphorylation at certain Ser/Thr residues, which is indicative of Alzheimer's disease.[11] Abnormal phosphorylation also affects tau structure, causing aggregation and oligomers to form, as well as altering function, as phosphorylated tau interacts with microtubules less than normal tau.[12] Methylation is achieved through methyltransferases and is important for regulating many different cellular functions.[13] Although methylation is a very small PTM, its' size belies its' complexity. For example, there are so many different possibilities for histone methylation alone that it has been hypothesized that there is a histone code read by enzymes in the cell that controls and modulates many different biological functions depending on the methylation state.[14] Ubiquitination on the other hand is a much larger modification in which a 76-residue ubiquitin protein is attached to another protein via a ubiquitin ligase enzyme.[15] Ubiquitination is also quite complex, as different sites can be ubiquitinated, and each site of modification can be ubiquitinated by chains of ubiquitin linked together (called polyubiquitination), each eliciting a different cellular response. Some of the consequences of ubiquitination is recognition by the proteasome for degradation, and adding new capabilities for protein-protein interactions through ubiquitin recognition domains once ubiquitin is attached to a protein. It is clear that post-translational modifications further expand the functionality of proteins beyond the 20 canonical amino acids, significantly affect protein structure, and can be used for context-sensitive modulation of protein function.

Spontaneous Chemical Modifications

PTMs often occur for the purpose of intentional modulation of protein structure for a gain of function, but proteins can undergo modifications that are not as tightly regulated and often result in the disruption of protein structure and function. Spontaneous chemical modifications (SCMs) are similar to PTMs but are not controlled enzymatically and are often the result of protein aging and/or damage. One such example of a SCM is protein oxidation. Protein oxidation is most common at aromatic and thiol-containing residues such as Cys, Met, Trp, His, Tyr, and Phe.[16] Oxidation is often caused by exposure to UV light, and involves either direct or indirect reaction of oxygen with these amino acids in proteins. Proteins can undergo many changes after oxidation, such as cross-linking, increased likelihood of unfolding, changes in conformation, as well as changes in binding to cofactors and metal ions.[17] For human growth hormone, Met oxidation leads to thermal destabilization of protein folding,[18] and although no change in conformation was found, oxidation led to a higher propensity for aggregation in the recombinant protein.[19] Furthermore, oxidation has been found to be present in many diseases, as either a consequence of other protein damage occurring or as a causative factor.[20] Oxidation is perhaps the most common SCM, but there are others to consider in the study of proteins as well.

Racemization or isomerization of residues from L- to D-amino acids is another SCM that can occur in proteins over time. In nature, amino acids exist predominantly in the L-form, and cellular machinery is almost exclusively set up to accommodate only L-amino

acids. This inherent homochirality of biological systems is fundamental to life as we know it on planet Earth.[21] Although naturally-occurring D-amino acids are found in toxins and neuropeptides from organisms such as platypus, frogs, mussels, octopi, and snails, these occurrences are rare compared to the vast majority of amino acids found in nature and their role in these organisms is tightly regulated.[22] D-amino acids can also arise spontaneously from L-amino acids isomerizing to D-amino acids, which can drastically alter protein structure and function. Since the machinery of the cell is finely tuned for L-amino acids, isomerization can cause loss of function in many processes.

Of the canonical amino acids, Aspartic acid is most susceptible to isomerization and undergoes this process most readily. This is due in part to Aspartic acid's unique structure which facilitates isomerization through a 5-membered succinimide ring intermediate.[23] Because of this intermediate, Aspartic acid can form additional isomers besides simple L- to D-inversion of the alpha-carbon stereocenter. Nucleophilic attack from the nitrogen in the backbone to the carbonyl carbon in the Asp side chain creates a five-membered L-succinimide ring, which is capable of racemization via an enol intermediate to the D-succinimide form. Hydrolysis at either of the two carbonyls in L-/D-succinimide results in the ring reopening and forming either L-/D-Asp or L-/D-isoAsp depending on which carbonyl is hydrolyzed.

Isomerization can have a significant effect on protein structure and function. For example, Asp isomerization in a monoclonal antibody was found to significantly decrease binding affinity to IgE, which drastically alters the efficacy of the antibody.[24] Isomerization in substrates can also disrupt the ability of the antibody to recognize the

substrate, such as in the case of Aβ-specific antibodies which have been found to have decreased binding affinities for isomerized Aβ.[25] Not only are antibody-substrate activities affected by isomerization in either the substrate or antibody, enzymatic activity is affected by isomerization as well. In Ribonuclease A, it was found that changing the Asp at position 121 from L-Asp to either L-iso, D-, or D-isoAsp resulted in a complete loss in catalytic activity for RNA cleavage, since Asp121 forms a hydrogen bond with His119 that is necessary for enzymatic activity.[26] The ability for kinases to phosphorylate target residues is affected by isomerization as well. Phosphorylation of eye lens cyrstallin αB at Ser59, an important phosphorylation site for regulating oligomerization, was found to be inhibited by Ser59 epimerization and nearby Asp isomerization.[27]

Although Asp is the amino acid which is most likely to isomerize, there are other amino acids which can undergo epimerization or isomerization. Similarly to Aspartic acid, Glutamic acid isomerizes through a 6-membered glutarimide intermediate.[28] However, Glu isomerization is much slower than Asp isomerization. Serine is also known to epimerize and is the second-fastest residue to epimerize behind Aspartic acid, however the mechanism of Serine epimerization is not well understood.[29]

Deamidation of Asparagine and Glutamine is another common SCM that occurs in proteins. Asn deamidation is closely related to Asp isomerization due to proceeding through the succinimide intermediate. Asn deamidation has been extensively studied and specific rates of deamidation calculated for model peptides and proteins.[30,31] Similar to Asp isomerization, Asn deamidation can cause structural changes in proteins as well as loss in function.[32,33]

Structural Elements Affecting Protein Vulnerability to SCMs

Although SCMs can cause structural changes in proteins, the initial protein structure can result in greater susceptibility to SCMs. For example, protein structure is one of the decisive factors in determining the extent to which a protein is oxidized.[34] Structural factors such as increased accessibility, electrostatics, and misfolding all increase the likelihood of oxidation in proteins.[35] For isomerization and deamidation, there are specific structural factors that are known to increase or decrease the rate. One such factor is local sequence effects. Aspartic acid residues with N-terminal residues that have smaller side chains reduce the amount of steric hindrance and increase the propensity for succinimide formation. Asp-Gly, Asp-Ala, or Asp-Ser are the three motifs that undergo isomerization the quickest.[36] Similarly, Asn-Gly, Asn-Ala, and Asn-Ser were found to be the motifs which deamidate the fastest.[30] In addition to primary sequence, secondary and tertiary protein structure can significantly affect isomerization rates as well. For example, two different monoclonal antibodies possessing similar Asp-Gly motifs but one antibody possessing greater flexibility and solvent accessibility near the Asp-Gly motif were both found to isomerize, but the antibody with greater solvent accessibility and flexibility at the Asp-Gly motif isomerized quicker.[37] Both solvent accessibility and flexibility are affected by protein secondary and tertiary structure. Regions where there is greater order in protein structure results in less deamidation and isomerization, while disordered regions in proteins results in greater deamidation and isomerization due to the lack of constraint from structural effects.[38,39,40,41] Thus, Asp residues with a C-terminal Gly, Ala,

or Ser will isomerize quickest, and the rate of isomerization will be increased or decreased by the protein structure or lack thereof at the site in question.

Understanding structure and the factors that influence it is vital to understanding the role and function of proteins in the cell. A full picture of structure and function starts but does not end with the primary structure. Grasping the entirety of a protein's structure requires grasping the covalent and noncovalent interactions that occur between the side chains and shining light on how these interactions build into the overall 3D picture, as well as understanding the PTMs and SCMs that a protein undergoes and how these influence structure. Although many analytical tools are helpful in studying these complex and multilayered interactions, no one tool is as powerful for the study of proteins as mass spectrometry.

Mass Spectrometry

In the early 1950s, mass spectrometry was utilized mainly by petroleum companies to analyze hydrocarbons and saw limited use by chemists in measurements of isotopic ratios of gases. MS started to emerge as the premier analytical technique that it is today when Klaus Biemann was struck with the inspiration to apply mass spectrometry to sequencing peptides.[42] However, instrumentation limitations held mass spectrometry back from its potential, as ionization of large biomolecules remained cumbersome due to the inherent harshness of the ionization methods in use at the time. It wasn't until the advent of electrospray ionization (ESI), which enabled peptides and proteins to be gently lifted into the gas phase with ease, that mass spectrometry would explode in relevance for the study

of proteins.[43] Since then, a plethora of MS-based techniques for studying proteins and protein structure have been developed.

One of the most important aspects of mass spectrometry that many of these techniques are based on is the ability to fragment molecules in the gas phase. Without this ability, the usefulness of mass spectrometry as an analytical technique would be severely limited, as there is only so much information that can be gleaned from the intact mass of a molecule, particularly for large biomolecules such as peptides and proteins. Some of the methods of fragmentation available are collision-induced dissociation (CID), higher-energy collisional dissociation (HCD), electron transfer dissociation or electron capture dissociation (ETD/ECD), ultraviolet photodissociation (UVDP), and radical-directed dissociation (RDD).

CID is the most common MS/MS fragmentation technique employed. CID involves slow heating of molecules over many low-energy collisions with gas atoms, usually helium, until the energy threshold required to break bonds is reached. For peptides, this usually results in cleavage of the peptide bond, and this produces predominantly b- and y-fragments. HCD is similar to CID but differs in that instead of many low-energy collisions, HCD is instead performed by accelerating ions into a chamber filled with nitrogen gas. Fewer collisions occur, but each collision is much higher in energy relative to the collisions that occur in CID.[44] Similar to CID, HCD produces mostly b- and y-fragments, but there are notable differences between CID and HCD. For example, HCD tends to produce smaller fragments at higher HCD energies, since larger fragments tend to undergo multiple collisions and experience secondary fragmentation events.[45]

In ETD/ECD, a reagent is introduced to the molecules in the mass spectrometer and reacts with the analytes creating a hydrogen-rich radical. This can result in backbone c- and z-ions for ETD/ECD on peptides and proteins, fragments which are not normally observed in CID or HCD. Important to Asp isomerization, ETD/ECD have been found to create diagnostic c+57 and z-57 ions for isoAsp, which have been used to identify Asp isomerization.[46]

UVPD is a technique in which a laser is used to irradiate molecules in the mass spectrometer, which causes specific fragmentation based on the wavelength of the laser. 213nm is the most common wavelength used for UVPD, but other wavelengths such as 193nm have been used as well.[47,48] UVPD can create a wide range of different fragments, including a-, b-, c-, x-, y-, and z-ions, and +/-1Da variants of each of these ions are also possible.[49] More selective dissociation also frequently occurs with UVPD, such as dissociation of C-S and S-S bonds, as these bonds absorb specifically at the wavelengths used in UVPD, allowing disulfides to be easily elucidated.[50,51]

RDD is a novel technique in which a hydrogen-deficient radical is created during 266nm MS^2 photodissociation (PD) of a chromophore-tagged analyte, followed by MS^3 isolation and collisional activation of the radical to induce migration and radical-driven fragmentation.[52] In order to create the radical during MS^2 PD, it is important to select a chromophore that has high PD yield at 266nm. Typical chromophores for RDD include 4-iodobenzoic acid, 3-iodo-tyrosine, and 4IB-18-crown-6,[53] although more novel chromophores have also been employed, such as isothiocyanates,[54] quinones,[55] and naphthalene.[56] Similar to UVPD, RDD can create a-, b-, c-, x-, y-, and z-ions, as well as

neutral side chain losses unique to each amino acid. Due to the migration of the radical through the available space of the peptide/protein, RDD is sensitive to structural differences and has been used to distinguish peptide epimers.[57,58] RDD is a powerful, structurally sensitive fragmentation technique for the study of protein structure as well as SCMs such as isomerization.

Besides being able to fragment molecules in the gas phase, coupling of separation techniques to mass spectrometry is perhaps one of the greatest advancements for the study of proteins. These separation techniques frequently involve bottom-up approaches in which proteins are digested by an enzyme such as trypsin prior to separation, where the individual peptide sequences as well as PTMs and/or SCMs can be mapped by fragmentation.[59] In this manner, identifying peptides on the order of thousands from very complex samples becomes possible through techniques such data-independent acquisition (DIA).[60] Additionally, separation of intact proteins is a growing field in proteomics and overcomes the difficulty of translating from peptides to proteins.[61] There are a few different modes of separation that have been applied to the study of proteins, the most common of which is reversed phase liquid chromatography (RP-LC). RP-LC separation of peptides relies on differences in interactions between the peptide and the hydrophobic stationary phase. If the correct mobile phases, gradient, and stationary phase are selected, separation even of small chemical differences caused by SCMs such as isomerization can be achieved, as even inversion of a stereocenter can cause significant differences in how peptide epimers interact with the stationary phase.[62] In the quest for better separation, novel stationary phases such as chiral phases have been employed in the separation of

peptide epimers and show promise for rapid separation of isomers.[63] Besides liquid chromatography, there are other MS-compatible separation techniques that have been used to study proteins. Capillary electrophoresis (CE) is frequently coupled to mass spectrometry and rather than interactions with a stationary phase, CE relies on separation based on charge and mass. CE has successfully been used to separate deamidated peptides based on charge differences caused by deamidation, as well as separate Asp and isoAsp isomers which is attributed to small differences in the pKas for Asp and isoAsp.[64,65] Gas-phase separation of proteins has also been developed through ion mobility mass spectrometry (IM-MS). IM-MS separates ions based on their collision-cross sections which are closely related to their mobility in a buffer gas. Ion mobility has also been used to separate Asp isomers[66,67] and can even be employed to pinpoint the location of isomerization within peptides by performing MS^n ion mobility separation on MS/MS fragments.[68,69] The ease with which mass spectrometry can be coupled with separation techniques such as these, as well as the many fragmentation techniques available make mass spectrometry a powerful tool for the analysis of protein structure.

Conclusion

The proper structure and function of proteins is vitally important for their roles in the cell, and PTMs and SCMs can drastically upset both structure and function. Although isomerization is usually considered an invisible SCM to mass spectrometry, techniques such as chromatographic separation and RDD enable isomerization to be identified and studied. The work described in this dissertation highlights new advances in mass

spectrometry for studying the consequences of isomerization on protein structure. Beginning in chapter 2, we take a detailed look at proline isomerization and establish that 266nm PD can distinguish between cis and trans proline isomers. Because proline isomers are prone to rearrangement, this study also provides insight into methods for preservation of solution-state structures into the gas phase for analysis. In chapter 3, we detail a covalent labeling technique specific for L-isoAsp which allows major sites of isomerization and long-lived proteins to be identified with ease. Our experiments characterize the covalent modification features such as diagnostic fragmentation and demonstrate the utility of this technique on a 72-year-old human eye lens sample. In chapter 4, the consequences of isomerization on trypsin and other trypsin-like proteases such as chymotrypsin and the 20S proteasome will be examined. Our experiments reveal that these proteases are not able to cleave isomerized substrates as effectively as the all-L substrates and have important implications for the cellular degradation of isomerized proteins. Finally, in chapter 5 we explore RDD-HCD as a new MS^n technique for studying protein structure. RDD-HCD provides tunable fragmentation based on the amount of HCD energy used and provides insights into the mechanism of HCD and radical fragmentation in the gas phase.

References

[1] Klecel, W.; Martyniuk, E. From the Eurasian Steppes to the Roman Circuses: A Review of Early Development of Horse Breeding and Management. *Animals* **2021**, *11* (7), 1859. https://doi.org/10.3390/ani11071859.

[2] Booth, D. R.; Sunde, M.; Bellotti, V.; Robinson, C. V.; Hutchinson, W. L.; Fraser, P. E.; Hawkins, P. N.; Dobson, C. M.; Radford, S. E.; Blake, C. C. F.; et al. Instability, Unfolding and Aggregation of Human Lysozyme Variants Underlying Amyloid Fibrillogenesis. *Nature* **1997**, *385* (6619), 787–793. https://doi.org/10.1038/385787a0.

[3] Armstrong, R. A. The Molecular Biology of Senile Plaques and Neurofibrillary Tangles in Alzheimer's Disease. *Folia Neuropathol.* **2009**, *47* (4), 289–299.

[4] Winklhofer, K. F.; Tatzelt, J.; Haass, C. The Two Faces of Protein Misfolding: Gain- and Loss-of-Function in Neurodegenerative Diseases. *EMBO J.* **2008**, *27* (2), 336–349. https://doi.org/10.1038/sj.emboj.7601930.

[5] Ramazi, S.; Zahiri, J. Post-Translational Modifications in Proteins: Resources, Tools and Prediction Methods. *Database* **2021**, *2021* (7), 1–20. https://doi.org/10.1093/database/baab012.

[6] Drazic, A.; Myklebust, L. M.; Ree, R.; Arnesen, T. The World of Protein Acetylation. *Biochim. Biophys. Acta - Proteins Proteomics* **2016**, *1864* (10), 1372–1401. https://doi.org/10.1016/j.bbapap.2016.06.007.

[7] Holmes, W. M.; Mannakee, B. K.; Gutenkunst, R. N.; Serio, T. R. Loss of Amino-Terminal Acetylation Suppresses a Prion Phenotype by Modulating Global Protein Folding. *Nat. Commun.* **2014**, *5* (1), 4383. https://doi.org/10.1038/ncomms5383.

[8] Johnson. The effects of phosphorylation on the structure and function of proteins. Annual review of biophysics and biomolecular structure. 1993;22(1). doi:10.1146/annurev.bb.22.060193.001215

[9] Humphrey, S. J.; James, D. E.; Mann, M. Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. *Trends Endocrinol. Metab.* **2015**, *26* (12), 676–687. https://doi.org/10.1016/j.tem.2015.09.013.

[10] Colson, B. A.; Thompson, A. R.; Espinoza-Fonseca, L. M.; Thomas, D. D. Site-Directed Spectroscopy of Cardiac Myosin-Binding Protein C Reveals Effects of Phosphorylation on Protein Structural Dynamics. *Proc. Natl. Acad. Sci.* **2016**, *113* (12), 3233–3238. https://doi.org/10.1073/pnas.1521281113.

[11] Suárez-Calvet, M.; Karikari, T. K.; Ashton, N. J.; Lantero Rodríguez, J.; Milà-Alomà, M.; Gispert, J. D.; Salvadó, G.; Minguillon, C.; Fauria, K.; Shekari, M.; et al. Novel Tau Biomarkers Phosphorylated at T181, T217 or T231 Rise in the Initial Stages of the Preclinical Alzheimer's Continuum When Only Subtle Changes in Aβ Pathology Are Detected. *EMBO Mol. Med.* **2020**, *12* (12). https://doi.org/10.15252/emmm.202012921.

[12] Noble, W.; Hanger, D. P.; Miller, C. C. J.; Lovestone, S. The Importance of Tau Phosphorylation for Neurodegenerative Diseases. *Front. Neurol.* **2013**, *4 JUL* (July), 1–11. https://doi.org/10.3389/fneur.2013.00083.

[13] Polevoda, B.; Sherman, F. Methylation of Proteins Involved in Translation. *Mol. Microbiol.* **2007**, *65* (3), 590–606. https://doi.org/10.1111/j.1365-2958.2007.05831.x.

[14] Chatterjee, J.; Rechenmacher, F.; Kessler, H. N -Methylation of Peptides and Proteins: An Important Element for Modulating Biological Functions. *Angew. Chemie Int. Ed.* **2013**, *52* (1), 254–269. https://doi.org/10.1002/anie.201205674.

[15] Komander, D. The Emerging Complexity of Protein Ubiquitination. *Biochem. Soc. Trans.* **2009**, *37* (5), 937–953. https://doi.org/10.1042/BST0370937.

[16] Grassi, L.; Cabrele, C. Susceptibility of Protein Therapeutics to Spontaneous Chemical Modifications by Oxidation, Cyclization, and Elimination Reactions. *Amino Acids* **2019**, *51* (10–12), 1409–1431. https://doi.org/10.1007/s00726-019-02787-2.

[17] Davies, M. J.; Truscott, R. J. W. Photo-Oxidation of Proteins and Its Role in Cataractogenesis. *J. Photochem. Photobiol. B Biol.* **2001**, *63* (1–3), 114–125. https://doi.org/10.1016/S1011-1344(01)00208-1.

[18] Mulinacci, F.; Capelle, M. A. H.; Gurny, R.; Drake, A. F.; Arvinte, T. Stability of Human Growth Hormone: Influence of Methionine Oxidation on Thermal Folding. *J. Pharm. Sci.* **2011**, *100* (2), 451–463. https://doi.org/10.1002/jps.22293.

[19] Mulinacci, F.; Poirier, E.; Capelle, M. A. H.; Gurny, R.; Arvinte, T. Influence of Methionine Oxidation on the Aggregation of Recombinant Human Growth Hormone. *Eur. J. Pharm. Biopharm.* **2013**, *85* (1), 42–52. https://doi.org/10.1016/j.ejpb.2013.03.015.

[20] Hawkins, C. L.; Davies, M. J. Detection, Identification, and Quantification of Oxidative Protein Modifications. *J. Biol. Chem.* **2019**, *294* (51), 19683–19708. https://doi.org/10.1074/jbc.REV119.006217.

[21] Blackmond, D. G. The Origin of Biological Homochirality. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (5), a002147–a002147. https://doi.org/10.1101/cshperspect.a002147.

[22] Mast, D. H.; Checco, J. W.; Sweedler, J. V. Advancing D-Amino Acid-Containing Peptide Discovery in the Metazoan. *Biochim. Biophys. Acta - Proteins Proteomics* **2021**, *1869* (1), 140553. https://doi.org/10.1016/j.bbapap.2020.140553.

[23] 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. https://doi.org/10.1016/s0021-9258(19)75855-4.

[24] Wakankar, A. A.; Borchardt, R. T.; Eigenbrot, C.; Shia, S.; Wang, Y. J.; Shire, S. J.; Liu, J. L. Aspartate Isomerization in the Complementarity-Determining Regions of Two Closely Related Monoclonal Antibodies. *Biochemistry* **2007**, *46* (6), 1534–1544.

[25] Readel, E. R.; Patel, A.; Putman, J. I.; Du, S.; Armstrong, D. W. Antibody Binding of Amyloid Beta Peptide Epimers/Isomers and Ramifications for Immunotherapies and Drug Development. *Sci. Rep.* **2023**, *13* (1), 12387. https://doi.org/10.1038/s41598-023-38788-1.

[26] Sakaue, H.; Kinouchi, T.; Fujii, N.; Takata, T.; Fujii, N. Isomeric Replacement of a Single Aspartic Acid Induces a Marked Change in Protein Function: The Example of Ribonuclease A. *ACS Omega* **2017**, *2* (1), 260–267. https://doi.org/10.1021/acsomega.6b00346.

[27] Lyon, Y. A.; Collier, M. P.; Riggs, D. L.; Degiacomi, M. T.; Benesch, J. L. P.; 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.

[28] Riggs, D. L.; Silzel, J. W.; Lyon, Y. A.; Kang, A. S.; Julian, R. R. Analysis of Glutamine Deamidation: Products, Pathways, and Kinetics. *Anal. Chem.* **2019**, *91* (20), 13032–13038. https://doi.org/10.1021/acs.analchem.9b03127.

[29] Hooi, M. Y. S.; Raftery, M. J.; Truscott, R. J. W. Age-Dependent Racemization of Serine Residues in a Human Chaperone Protein. *Protein Sci.* **2013**, *22* (1), 93–100. https://doi.org/10.1002/pro.2191.

[30] Robinson, N. E.; Robinson, A. B. Molecular Clocks. *Proc. Natl. Acad. Sci.* **2001**, *98* (3), 944–949. https://doi.org/10.1073/pnas.98.3.944.

[31] Robinson, N. E.; Robinson, A. B. Deamidation of Human Proteins. *Proc. Natl. Acad. Sci.* **2001**, *98* (22), 12409–12413. https://doi.org/10.1073/pnas.221463198.

[32] Chaves, J. M.; Srivastava, K.; Gupta, R.; Srivastava, O. P. Structural and Functional Roles of Deamidation and/or Truncation of N- or C-Termini in Human AA-Crystallin. *Biochemistry* **2008**, *47* (38), 10069–10083. https://doi.org/10.1021/bi8001902.

[33] Lampi, K. J.; Wilmarth, P. A.; Murray, M. R.; David, L. L. Lens β-Crystallins: The Role of Deamidation and Related Modifications in Aging and Cataract. *Prog. Biophys. Mol. Biol.* **2014**, *115* (1), 21–31. https://doi.org/10.1016/j.pbiomolbio.2014.02.004.

[34] Höhn, A.; König, J.; Grune, T. Protein Oxidation in Aging and the Removal of Oxidized Proteins. *J. Proteomics* **2013**, *92*, 132–159. https://doi.org/10.1016/j.jprot.2013.01.004.

[35] Chang, R. L.; Stanley, J. A.; Robinson, M. C.; Sher, J. W.; Li, Z.; Chan, Y. A.; Omdahl, A. R.; Wattiez, R.; Godzik, A.; Matallana-Surget, S. Protein Structure, Amino Acid Composition and Sequence Determine Proteome Vulnerability to Oxidation-induced Damage. *EMBO J.* **2020**, *39* (23), 1–21. https://doi.org/10.15252/embj.2020104523.

[36] Clarke, S. Propensity for Spontaneous Succinimide Formation from Aspartyl and Asparaginyl Residues in Cellular Proteins. *Int. J. Pept. Protein Res.* **2009**, *30* (6), 808–821. https://doi.org/10.1111/j.1399-3011.1987.tb03390.x.
[37] Wakankar, A. A.; Borchardt, R. T.; Eigenbrot, C.; Shia, S.; Wang, Y. J.; Shire, S. J.; Liu, J. L. Aspartate Isomerization in the Complementarity-Determining Regions of Two Closely Related Monoclonal Antibodies. *Biochemistry* **2007**, *46* (6), 1534–1544.
[38] 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.
[39] Lyon, Y. A.; Sabbah, G. M.; Julian, R. R. Identification of Sequence Similarities among Isomerization Hotspots in Crystallin Proteins. *J. Proteome Res.* **2017**, *16* (4), 1797–1805. https://doi.org/10.1021/acs.jproteome.7b00073.
[40] Xie, M.; Schowen, R. L. Secondary Structure and Protein Deamidation. *J. Pharm. Sci.* **1999**, *88* (1), 8–13. https://doi.org/10.1021/js9802493.
[41] STEVENSON, C. L.; FRIEDMAN, A. R.; KUBIAK, T. M.; DONLAN, M. E.; BORCHARDT, R. T. Effect of Secondary Structure on the Rate of Deamidation of Several Growth Hormone Releasing Factor Analogs. *Int. J. Pept. Protein Res.* **2009**, *42* (6), 497–503. https://doi.org/10.1111/j.1399-3011.1993.tb00356.x.
[42] Biemann, K. Laying the groundwork for proteomics Mass spectrometry from 1958 to 1988. Int. J. Mass Spectrom. 2007, 259, 1-7.
[43] Fenn, J.B.; Mann, M.; Meng, C.K.; Wong, S.F.; Whitehouse, C.M. Electrospray ionization for the mass spectrometry of large biomolecules. Science 1989, 246(4926), 64- 71
[44] Xia, Y.; Liang, X.; McLuckey, S. A. Ion Trap versus Low-Energy Beam-Type Collision-Induced Dissociation of Protonated Ubiquitin Ions. Anal. Chem. 2006, 78 (4), 1218– 1227
[45] Samgina, T. Y.; Vorontsov, E. A.; Gorshkov, V. A.; Artemenko, K. A.; Zubarev, R. A.; Lebedev, A. T. Mass Spectrometric de Novo Sequencing of Natural Non-Tryptic Peptides: Comparing Peculiarities of Collision-Induced Dissociation (CID) and High Energy Collision Dissociation (HCD). Rapid Commun. Mass Spectrom. 2014, 28 (23), 2595– 2604
[46] Sargaeva, N. P.; Lin, C.; O'Connor, P. B. Identification of Aspartic and Isoaspartic Acid Residues in Amyloid β Peptides, Including Aβ1–42, Using Electron-Ion Reactions. Anal. Chem. 2009, 81 (23), 9778–9786
[47] Fornelli, L.; Srzentić, K.; Toby, T. K.; Doubleday, P. F.; Huguet, R.; Mullen, C.; Melani, R. D.; dos Santos Seckler, H.; DeHart, C. J.; Weisbrod, C. R.; et al. Thorough Performance Evaluation of 213 Nm Ultraviolet Photodissociation for Top-down Proteomics. *Mol. Cell. Proteomics* **2020**, *19* (2), 405–420. https://doi.org/10.1074/mcp.TIR119.001638.
[48] Lanzillotti, M.; Brodbelt, J. S. Comparison of Top-Down Protein Fragmentation Induced by 213 and 193 Nm UVPD. *J. Am. Soc. Mass Spectrom.* **2023**, *34* (2), 279–285. https://doi.org/10.1021/jasms.2c00288.
[49] Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651. https://doi.org/10.1021/ja4029654.
[50] Bonner, J.; Talbert, L. E.; Akkawi, N.; Julian, R. 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.
[51] Quick, M. M.; Crittenden, C. M.; Rosenberg, J. A.; Brodbelt, J. S. Characterization of Disulfide Linkages in Proteins by 193 Nm Ultraviolet Photodissociation (UVPD) Mass Spectrometry. *Anal. Chem.* **2018**, *90* (14), 8523–8530. https://doi.org/10.1021/acs.analchem.8b01556.
[52] 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.

[53] Hubbard. Elucidating Biomolecular Structure Through Bond-Selective Radical Fragmentation. Vol 9. Royal Society of Chemistry, 2020.
[54] 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.
[55] 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.
[56] Diedrich, J. K.; Julian, R. R. Site-Specific Radical Directed Dissociation of Peptides at Phosphorylated Residues. *J. Am. Chem. Soc.* **2008**, *130* (37), 12212–12213. https://doi.org/10.1021/ja8023719.
[57] Tao, Y.; Quebbemann, N. R.; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820. https://doi.org/10.1021/ac3013434.
[58] 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.
[59] Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394. https://doi.org/10.1021/cr3003533.
[60] Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; et al. OpenSWATH Enables Automated, Targeted Analysis of Data-Independent Acquisition MS Data. *Nat. Biotechnol.* **2014**, *32* (3), 219–223. https://doi.org/10.1038/nbt.2841.
[61] Melby, J. A.; Roberts, D. S.; Larson, E. J.; Brown, K. A.; Bayne, E. F.; Jin, S.; Ge, Y. Novel Strategies to Address the Challenges in Top-Down Proteomics. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (6), 1278–1294. https://doi.org/10.1021/jasms.1c00099.
[62] Winter, D.; Pipkorn, R.; Lehmann, W. D. Separation of Peptide Isomers and Conformers by Ultra Performance Liquid Chromatography. *J. Sep. Sci.* **2009**, *32* (8), 1111–1119. https://doi.org/10.1002/jssc.200800691.
[63] Du, S.; Readel, E. R.; Wey, M.; Armstrong, D. W. Complete Identification of All 20 Relevant Epimeric Peptides in β-Amyloid: A New HPLC-MS Based Analytical Strategy for Alzheimer's Research. *Chem. Commun.* **2020**, *56* (10), 1537–1540. https://doi.org/10.1039/C9CC09080K.
[64] Faserl, K.; Sarg, B.; Maurer, V.; Lindner, H. H. Exploiting Charge Differences for the Analysis of Challenging Post-Translational Modifications by Capillary Electrophoresis-Mass Spectrometry. *J. Chromatogr. A* **2017**, *1498*, 215–223. https://doi.org/10.1016/j.chroma.2017.01.086.
[65] Boni, S. De; Neusüß, C.; Pelzing, M.; Scriba, G. K. E. Identification of Degradation Products of Aspartyl Tripeptides by Capillary Electrophoresis-Tandem Mass Spectrometry. *Electrophoresis* **2003**, *24* (5), 874–882. https://doi.org/10.1002/elps.200390110.
[66] G. Nagy, K. Kedia, I. K. Attah, S. V. B. Garimella, Y. M. Ibrahim, V. A. Petyuk and R. D. Smith, Separation of β-Amyloid Tryptic Peptide Species with Isomerized and Racemized L-Aspartic Residues with Ion Mobility in Structures for Lossless Ion Manipulations, Anal. Chem., 2019, 91(7), 4374–4380
[67] X. Zheng, L. Deng, E. S. Baker, Y. M. Ibrahim, V. A. Petyuk and R. D. Smith, Distinguishing D- and L-Aspartic and Isoaspartic Acids in Amyloid β Peptides with Ultrahigh Resolution Ion Mobility Spectrometry, Chem. Commun., 2017, 53(56), 7913–7916
[68] C. Jia, C. B. Lietz, Q. Yu and L. Li, Site-Specific Characterization of d-Amino Acid Containing Peptide Epimers by Ion Mobility Spectrometry, Anal. Chem., 2014, 86(6), 2972–2981
[69] Wu, H. T.; Julian, R. R. Two-Dimensional Identification and Localization of Isomers in Crystallin Peptides Using TWIM-MS. *Analyst* **2020**, *145* (15), 5232–5241. https://doi.org/10.1039/d0an01036g.

CHAPTER 2: Probing the Stability of Proline Cis/Trans Isomers in the Gas Phase with Ultraviolet Photodissociation

Abstract

Although most peptide bonds in proteins exist in the trans configuration, when cis peptide bonds do occur, they can have major impact on protein structure and function. Rapid identification of cis-peptide bonds is therefore an important task. Peptide bonds containing proline are more likely to adopt the cis configuration because the ring connecting the sidechain and backbone in proline flattens the energetic landscape relative to amino acids with free side chains. Examples of cis proline isomers have been identified in both solution and in the gas phase by a variety of structure-probing methods. Mass spectrometry is an attractive potential method for identifying cis proline due to its speed and sensitivity, however the question remains whether cis/trans proline isomers originating in solution are preserved during ionization and manipulation within a mass spectrometer. Herein, we investigate the gas phase stability of isolated solution-phase cis and trans proline isomers using a synthetic peptide sequence with a Tyr-Pro-Pro motif. A variety of dissociation methods were explored to evaluate their potential to distinguish cis/trans configuration, including collision-induced dissociation, radical-directed dissociation, and photodissociation. Only photodissociation employed in conjunction with extremely gentle electrospray and charge solvation by 18-crown-6 ether was able to distinguish cis/trans isomers for our model peptide, suggesting that any thermal activation during transfer or while in the gas phase leads to isomer scrambling. Furthermore, the

necessity for 18-crown-6 suggests that intramolecular charge solvation taking place during electrospray ionization can override cis/trans isomer homogeneity. Overall, the results suggest that solution-phase cis/trans proline isomers are fragile and easily lost during electrospray, requiring careful selection of instrument parameters and consideration of charge solvation to prevent cis/trans scrambling.

Introduction

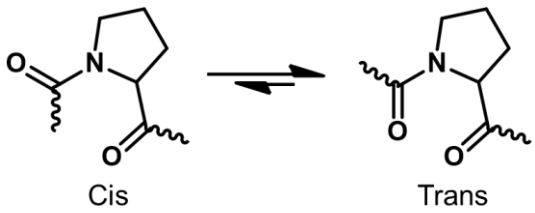
It has been hypothesized that one of the fundamental steps for life on Earth was the formation of the peptide bond, allowing the synthesis of proteins from amino acids.[1] Peptide bonds are rigid and planar, with local minima in either cis or trans configuration. The vast majority of peptide bonds in proteins adopt the trans configuration, which is energetically favorable relative to cis due to differences in sterics.[2] For non-proline peptide bonds the trans to cis rotational energy barrier is ~83.7 kJ/mol and the potential energy difference between the cis and trans isomers is ~10.5 kJ/mol.[3] The energetics are largely dictated by lower steric clash in the trans configuration. In contrast, the unique five-membered ring inherent to proline causes steric clash in both the cis and trans positions, lowering the transition barrier to ~54.4 kJ/mol and the potential energy difference to only ~2.1 kJ/mol (Scheme 2.1).[4] This results in Xxx-Pro peptide bonds being more likely to form cis isomers, although Xxx-Pro peptide bonds still favor trans over the cis configuration.[5]

There are specific sequence motifs that have been shown to increase the likelihood of cis isomer formation in proline-containing peptide bonds. A significant percentage of cis-

proline bonds are preceded by an aromatic residue, presumably due to stabilizing CH-π interactions between the aromatic residue and the proline.[5,6,7] Out of the three canonical aromatic amino acids, tyrosine precedes cis-proline most frequently in both model peptides and proteins.[2,6] In addition to cis-isomer stabilization from aromatic residues, sequential prolines also affect cis/trans isomerization propensity. In the case of a double-proline motif, four different isomers are possible due to cis/trans isomerization at the Xxx-Pro bond and at the Pro-Pro bond. Kinetic studies of model peptides containing Pro-Pro motifs suggest that sequential prolines significantly slow the rate of isomerization from cis to trans when compared to isomerization in model peptides with a single proline.[2] Combination of both CH-π and proline-proline interactions have been suggested to result in both a greater likelihood of cis-proline formation as well as a slower rate of isomerization, allowing cis-proline peptide bonds to be investigated with greater ease.[8]

When cis-proline is present, it significantly influences both structure and folding. The relatively slow timescale for proline cis/trans isomerization allows detailed study by NMR for characterization of protein folding events.[9] For example, heteronuclear 2D-NMR has been used to monitor folding of the proline-rich collagen triple helix, revealing that multiple cis/trans proline isomerization events occur during folding.[10] Differential scanning calorimetry coupled with NMR showed slow conversion of trans to cis proline at the Tyr28-Pro29 peptide bond in the homeodomain from the human transcription factor PBX, allowing direct detection of denatured state chemical shifts.[11] Proline cis/trans isomerization has also been shown to play roles in protein function. Peptidyl-prolyl isomerase-accelerated proline cis/trans isomerization in the Crk adaptor acts as a

switch between an autoinhibitory function (cis configuration) and an extended and uninhibited structure (trans configuration).[12] Cis-trans proline isomerization has also been established as a mechanism for the opening and closing of a neurotransmitter-gated ion channel.[13] Recently, the importance of cis-trans proline isomerization has been appreciated for modulating antigen-antibody interactions and the development of therapeutic antibodies including the anti-HIV antibody 10E8 that contains a Tyr-Pro-Pro (YPP) motif.[14-18] These examples highlight the biological relevance of proline cis/trans isomerization and establish the importance of utilizing existing techniques including NMR for discovery and characterization of proline cis/trans isomerization.



Scheme 2.1. Cis/trans isomerization of proline. The energy differences between the cis and trans configurations is ~2.1 kJ/mol.[4]

Gas phase studies of cis/trans proline isomerization have also revealed important insights. Ion mobility combined with mass spectrometry (IM-MS) can be used to separate conformations or derive structural information by comparing collision cross-sections with calculations. Multiple structure populations have been observed with IM-MS for proline-containing peptides, which have been attributed to cis/trans isomers of proline.[19,20,21] Similarly, the transition from polyproline I to polyproline II via the sequential cis-to-trans isomerization of each proline bond has been detailed.[22,23] Ion mobility has also been used

in conjunction with ultraviolet photodissociation (UVPD) to probe the differences in higher order gas-phase protein structure arising from proline cis/trans isomerization in ubiquitin.[24] In addition to ion mobility, IRMPD has been used to study gas-phase cis/trans isomers of prolyl bonds in polyproline.[25] IRMPD coupled with ion mobility identified a subpopulation within one of the major bradykinin structures resulting from isomerization to cis-proline at proline.[2,26] Despite the advances made in these studies, the full relationship between solution phase cis/trans isomers and gas phase cis/trans isomers remains an open question.

To examine known solution phase cis/trans proline isomers in the gas phase, the isomers must first be ionized and desolvated without causing structural rearrangement. Electrospray ionization (ESI) is the most commonly used method for attempting such transfers,[27] but the mechanism of ESI is not fully understood despite a great deal of investigation.[28,29,30] It is generally agreed that higher charge states are less likely to preserve solution-phase structure. Greater clarity has been provided by simulations in some cases, such as the chain ejection model proposed by Konermann.[31,32] However, proteins ionized by chain ejection bear little resemblance to their solution-phase structure, and in general, the clearest examples of gas-phase structure or structural transitions are those where the original native protein structure is lost. For example, highly charged ions that have essentially completely unfolded are clearly not native. Undisputed examples of structure retention remain more elusive.

While ESI is normally gentle enough to preserve covalent bonds during transfer into the gas phase, the removal of solvent creates a problem with respect to preservation of

solution-phase noncovalent interactions. These interactions are critical to stabilizing native protein structure and include details such as cis/trans configuration. In theory, mass spectrometric analysis of proline cis/trans isomerization should be readily achievable, but can cis/trans isomers present in solution be preserved during transfer into the gas phase? Although previous studies have indicated that transfer into the gas phase can lead to dramatic rearrangement of protein/peptide structure into non-native conformations, this question has not been fully explored in relation to proline cis/trans isomerization.[33]

In the course of this study, we utilize collisional activation, radical-directed dissociation (RDD), and photodissociation (PD) to assess cis/trans prolyl isomer stability during ESI. A model peptide (WSGYPPEE) derived from a sequence previously known to resolve into multiple liquid chromatography (LC) peaks corresponding to proline isomers is used. The results reveal that cis/trans prolyl isomers are fragile and subject to interconversion during ESI or if subjected to thermal activation. Preservation of these isomers from solution to gas phase for our model peptide requires both gentle ESI parameters and external charge solvation to minimize gas phase rearrangement. Even under these conditions, distinguishing cis/trans isomers is challenging, suggesting they are unlikely to interfere in experiments where not specifically targeted.

Experimental Methods

*Materials***.** Organic solvents and reagents were purchased from Fisher Scientific, Sigma-Aldrich, or Acros Organics and used without further purification. FMOC-protected amino acids and Wang resins were purchased from Anaspec, Inc or Chem-Impex International.

Peptide Synthesis. Peptides were synthesized manually following a FMOC-protected solid-phase peptide synthesis protocol.[34] FMOC-3-iodotyrosine replaced tyrosine during synthesis to provide a photocleavable radical precursor for RDD experiments. Following synthesis, peptides were stored frozen in 50/50 v/v acetonitrile/water.

Nuclear Magnetic Resonance. 6 mg of modified peptide WSG(I-Y)PPEE was dried fully and resuspended in in 0.104 mM 3-(Trimethylsilyl)-1-propanesulfonic acid (DSS), 50 mM sodium phosphate and pH corrected to pH 2.75 (final peptide concentration of 9.5 mM). The sample was then dried down by speedvac and resuspended in 99.99% D2O twice prior to NMR analysis. Samples were covered with aluminum foil whenever possible to prevent photodegradation.

All spectra were acquired on a Bruker AVANCE III 700 MHz solution NMR Spectrometer equipped with a proton optimized room temperature triple resonance TXI-1H{^{13}C,^{15}N} probehead. The sample was equilibrated for at least 10 minutes at a temperature of 283 K before spectra were recorded. 2D ^{1}H-^{1}H TOCSY (70 ms mixing time, 2048 points along F2 and 1024 points along F1 with 8 scans per increment), ^{1}H-^{1}H ROESY (200 ms mixing time, 2048 points along F2 and 1024 points along F1 with 8

scans per increment) and ^{1}H-^{13}C HSQC at 1% natural abundance of ^{13}C (2048 points along F2 and 512 points along F1 with 96 scans per increment) were recorded and spectra were processed and referenced to the DSS trimethyl silyl resonance using Bruker TopSpin software (version 4.0.8). ^{1}H resonances were assigned using the combination of crosspeaks from the HSQC and spin systems assigned from the TOCSY and ROESY using NMRFAM-SPARKY (version 1.470, powered by SPARKY version 3.190).

Assignment of all resolved peaks are shown in supporting table S1. All assigned chemical shifts were within one standard deviation from the average values reported in the Biological Magnetic Resonance Data Bank.[35] Cis/Trans proline assignments for each isomer were made using the relative intensity of the HA-HA or HD-HA crosspeak,[36] along with the carbon chemical shifts for the assigned prolines.[37] The relative abundance of each isomer was determined by the relative integration of a single resonance associated with each specific isomer. For example, the relative integration of the P5-α-trans-trans resonance was compared directly with P5-α-cis-trans and P5-α-trans-cis to determine the relative abundance of each population within the sample. The total relative populations were calculated from the P5 Ca, P6 Ca, and Y4 Cb cross-peak intensities.

Analysis. For LCMS, an Agilent 1100 binary pump was used with a Thermo BetaBasic-18 3 µm C18 150 mm x 2.1 mm column interfaced to a Thermo Fisher Scientific LTQ mass spectrometer with a standard ESI source. Samples were eluted using water with 0.1% formic acid as mobile phase A, and acetonitrile with 0.1% formic acid as mobile phase B. A Waters XBridge 3.5 µm C18 4.6 mm x 250 mm column interfaced to a Varian ProStar 320 UV/Vis detector and two ProStar 210 pumps were used to

fractionate cis/trans isomers from the iodo-tyrosine peptide variant prior to performing PD and RDD experiments. To minimize interconversion between cis and trans isomers during separation, the column was wrapped with ice packs and the solvents were kept in an ice bath. Separated fractions were prepared as ~20 μM samples in 49.5/49.5/1 (v/v/v) acetonitrile/water/formic acid with a ~1:1 molar ratio of peptide to 18C6 and infused into the LTQ linear ion trap using the standard electrospray ionization source. The LTQ was modified previously with a quartz window to allow 266 nm laser pulses from a Nd:YAG laser to excite the isolated precursor ions and form a radical on the peptide by photodissociation of the carbon—iodine bond. Following photodissociation, the radical product was further isolated and activated by collision-induced dissociation (CID) to promote radical-directed dissociation.

Instrument Parameters. Gentle ESI parameters were obtained by tuning the LTQ source voltages to maximize selective observation of serine octamer without generating any other serine clusters. This 'magic serine octamer' spectrum will only be observed if the ESI conditions are sufficiently gentle,[38] which does not correspond to high signal or analytical sensitivity. The instrument parameters for the gentle ESI tune are as follows: spray voltage 4.4kV, capillary temperature 120.0°C, capillary voltage 0.00V. The tube lens voltage was varied as follows: -35V, -25V, and +90V (corresponding to decreasingly gentle settings). To evaluate the effect of more standard ESI conditions, the instrument was tuned to the [M+H] + 18C6 adduct. The resulting parameters were spray voltage 4.4kV, capillary temperature 120.0°C, capillary voltage 23V, and tube lens +190.0V.

R_{isomer} Scores. We employ a method established by Tao *et al*.[39] to distinguish and quantify the differences in fragmentation patterns arising from two different isomers in PD and RDD spectra. R_{isomer} scores are determined by calculating a ratio of ratios R_1 / R_2, where R_1 and R_2 are ratios of two fragment ion intensities for isomers 1 and 2, respectively. Values of R_{isomer} close to one indicate fragmentation patterns that are nearly identical, while spectra with increasing differences in fragmentation give higher R_{isomer} scores. All possible pairs of fragment ions in two sets of spectra are compared, with the pair that gives the highest R_{isomer} score being selected as the pair that distinguishes the isomers the most. We define the minimum R_{isomer} scores that confirm the presence of isomeric forms to be those where differences in both reference peaks are outside two standard deviations.

Results and Discussion

CID and RDD Experiments. In this study, a model synthetic peptide WSGYPPEE (based on the 10E8 antibody) is used to study the retention of cis/trans proline during transfer into the gas phase by ESI. Isomerization of the Tyr-Pro and Pro-Pro bonds can yield up to four possible isomers (trans-trans, cis-trans, trans-cis, and cis-cis). When examined in electrospray-like solutions, three isomers are observed by NMR that correspond to WSG(iodo-Y)transPtransPGEE, WSG(iodo-Y)cisPtransPGEE, and WSG(iodo-Y)transPcisPGEE configurations, with relative intensities of 59.49 ± 5.6%, 33.54 ± 2.9% and 6.97 ± 2.8%, respectively (Figure S1). The lack of any signal observed

for the WSG(iodo-Y)cisPcisPGEE is consistent with the low propensity of peptides to adopt the higher energy cis-cis isomer.2

LCMS analysis of the same peptide (WSGYPPEE) produces a chromatogram with two prominent peaks that derive from the same m/z (Figure 2.1a). The relative abundances of the two isomers as calculated from peak areas are ~43% and ~57% respectively, indicating elution of the cis-trans isomer first followed by the trans-trans isomer. Collisional activation of the two isomers produces virtually identical fragmentation patterns (Figure 2.1b and 2.1c) that are incapable of distinguishing the two isomers. This failure may be in part due to the abundant $y4^+$ ion that dominates both spectra. It is produced by the proline effect that favors CID cleavage N-terminal to proline.[40] It is also likely that the thermal energy added for CID is sufficient to completely scramble cis/trans isomer configuration.

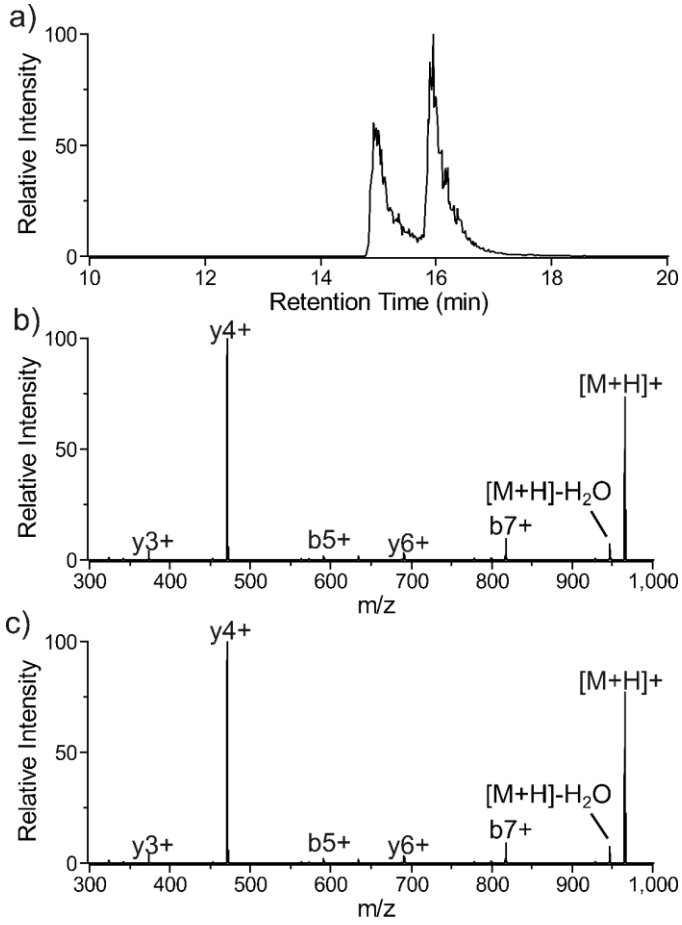


Figure 2.1. (a) C18 separation on ice of the synthetic peptide WSGYPPEE resulting in separation of the cis-trans and trans-trans isomers at retention time 14.4 minutes and 16 minutes, respectively. **(b)** CID spectrum of the cis-trans isomer eluting at 14.4 minutes and **(c)** CID spectrum for the trans-trans isomer eluting at 16 minutes. The CID spectra are indistinguishable.

Unfortunately, the proline effect may always be a potential problem for peptides that contain proline (and are therefore candidates for proline cis/trans isomerization). Furthermore, CID is not very sensitive to structure in general. However, we have demonstrated previously that radical-directed dissociation (RDD) can identify covalently constrained stereoisomers based on differences in fragmentation.[41,42] This approach can even successfully discriminate challenging isomers such as those derived from aspartic

acid and glutamic acid.[43,44] RDD is initiated by photodissociation of a carbon–iodine bond, allowing for selective creation of a radical following activation with ultraviolet light. 3-iodotyrosine was incorporated into our model peptide to facilitate RDD experiments. To establish whether modification of the tyrosine affected the propensity of the YP bond to form cis isomers, we attempted to separate isomers from the modified sequence with LC. Results show that WSG(iodo-Y)PPEE separates into two prominent peaks (Figure 2.2a), corresponding to cis-trans and trans-trans. To avoid the possibility that proline cis/trans isomers might be scrambled during transit into the gas phase, gentle ESI parameters were employed. Since ESI conditions are more difficult to precisely control when the instrument is operated in LCMS mode, experiments were conducted by direct infusion. The peptide was separated into two isomers via RP-HPLC with a C18 column while the solvents and column were chilled with an ice bath and cold packs, respectively. The two fractions corresponding to the separated isomer populations were then analyzed promptly using soft ESI tune conditions. The resulting RDD spectra are shown in Figures 2.2b and 2.2c. Despite the use of gentle ESI parameters, the data are still quite similar and yield an R_{isomer} score of 1.3. The exact identity of the C-terminal fragment used in the R_{isomer} score is unknown, but further CID on this ion results in an abundant proline effect y_4^+ fragment corresponding to PPEE from the original peptide. Graphing the intensities of the c-terminal and -129W fragments relative to $[M^{\bullet}\text{-NH3+H}]^+$ and representing the standard deviation of the mean with error bars further shows that these fragmentation patterns are not significantly different in intensity (Figure 2.2d).

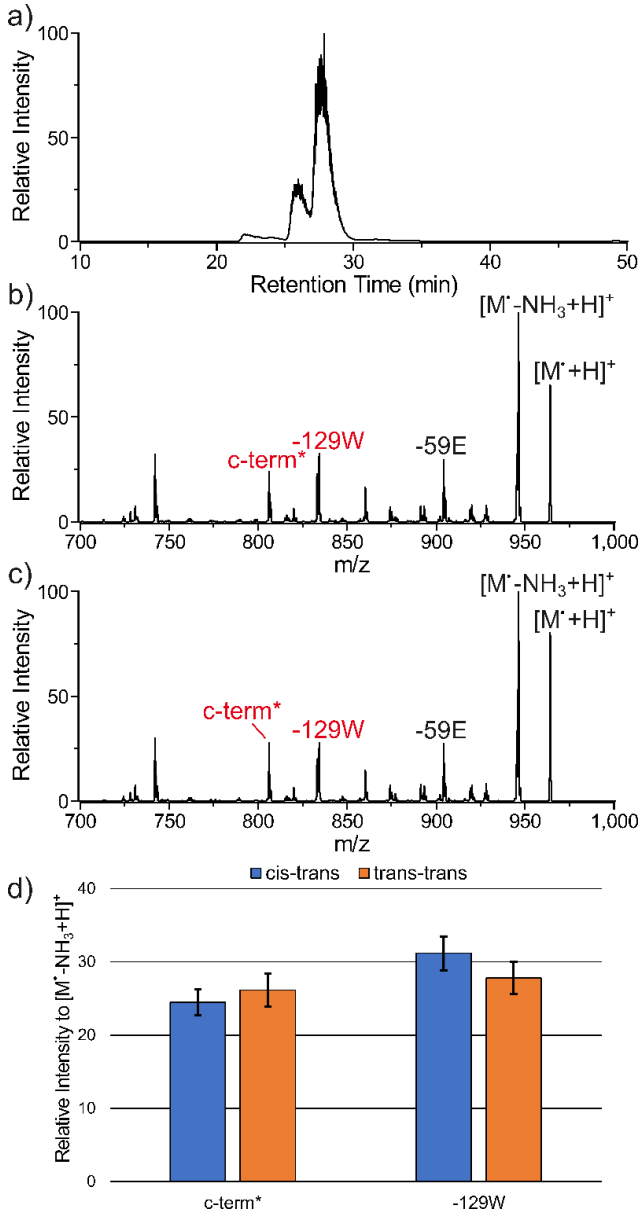


Figure 2.2. (a) C18 RP-HPLC separation on ice of the synthetic peptide WSG(3-iodo-Y)PPEE. Peaks corresponding to cis-trans and trans-trans isomers at ~25 minutes and ~27.5 minutes were collected on ice and promptly analyzed by RDD to yield the spectra in **(b)** and **(c)**. An R_{isomer} score of 1.3 was obtained using the -129W* and C-terminal fragments shown in red. For this data, a threshold R_{isomer} score of 1.9 would be required to confirm the presence of isomers. **(d)** Representation of the relative intensity of the two fragments with the standard deviation of the mean represented by error bars. Inclusion of error bars shows overlap between the -129W intensities and near overlap between the c-terminal RDD fragment. *Side chain losses here and throughout are represented by the shorthand notation corresponding to mass loss and amino acid.42

18C6 Photodissociation Experiments. Given that RDD is a structurally sensitive technique that is able to differentiate peptide diastereomers based on L/D isomerization, RDD should have been able to produce differences in fragmentation patterns if the proline cis/trans isomers were able to survive ESI without scrambling and radical migration occurred prior to scrambling once in the gas phase. To further dissect out these possibilities, experiments were conducted with the addition of 18-crown-6 (18C6), which has been shown to solvate charges and aid in the preservation of native-like structure during ESI.[45,46] Using the same source conditions as in the previous protocol with the addition of 18C6 to the isolated fractions prior to direct infusion yielded the PD spectra shown in Figures 2.3a and 2.3b. One of the primary losses corresponds to the loss of iodine, as expected. However, further collisional activation of the radical adduct $[M^{\bullet}+18C6+H]^+$ to obtain RDD data resulted in spectra that were once again very similar to each other (Figure 2.S2). It is possible that the collisional activation step in RDD enables cis/trans isomer scrambling prior to radical migration, reducing structural sensitivity.

However, two peaks $[M\text{-}NH_3+H]^+$ and $[M\text{-}CO_2+H]^+$ correspond to losses from the precursor peptide without loss of iodine. Comparison of these peaks yields a higher R_{isomer} score of 3.5, which is consistent with different isomeric forms. The loss of ammonia and CO_2 are likely due to direct dissociation,[47] meaning that these losses occur prior to any heating. The 18C6 adducts would then be lost afterwards following internal conversion of residual photon energy.[48] Collisional activation of the protonated peptide is dominated by the proline effect and yields a spectrum similar to that shown in Figure 1

for the non-iodinated peptide for which losses of CO_2 and ammonia are not notable, further suggesting these losses are uniquely attributable to UVPD.

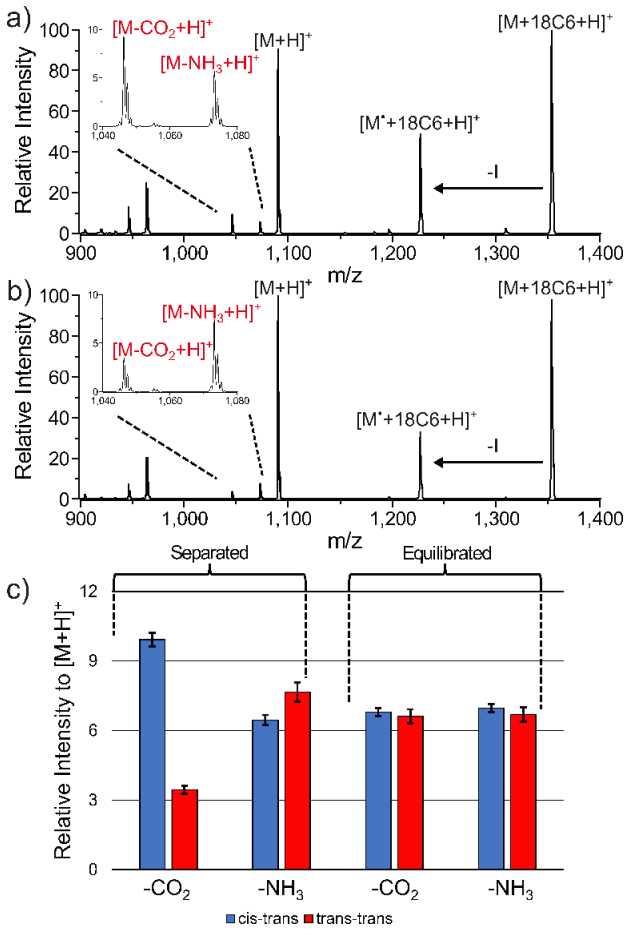


Figure 2.3. PD spectra resulting from photodissociation of the $[M+18C6+H]^+$ adduct for the cis-trans isomer **(a)** and the trans-trans isomer **(b)**. $[M-NH_3+H]^+$ and $[M-CO_2+H]^+$ are the two fragments that differ the most between the two spectra and are responsible for the R_{isomer} score of 3.5 (relative to minimum threshold of 1.3). **(c)** Bar plot of the intensity for each of these two fragments normalized to the $[M+H]^+$ ion intensity. Error bars represent the standard deviation of the mean, showing significant differences directly after separation, but virtually no difference after equilibration and repetition of the experiments.

To confirm that the differences in intensity noted in Figure 2.3a and 2.3b derive from cis or trans configuration, the two fractions were equilibrated at room temperature for 12

hours to allow for cis/trans interconversion. Repetition of the experiments with 18C6 and the same ESI parameters used prior resulted in PD spectra with a much lower R_{isomer} score of 1.3 (see Fig. 2.S3), confirming that the differences observed immediately following separation stem from the cis-trans/trans-trans configurations. The change in behavior is quantitatively illustrated in Figure 2.3c, where the average intensity and error are plotted for both experiments.

ESI Parameters and Preservation of Cis/Trans Configuration. Over the course of many past experiments, we have noted that the tube lens voltage has the largest influence on the gentleness of the ESI in our instrument. The results of variation of the tube lens voltage from -35, -25, +90 and +190V during collection of PD data with 18-crown-6 are summarized in Figure 4, which show that the losses of CO_2 and NH_3 are most significantly different at more negative tube lens voltages and progressively become less significant as the tube lens voltage increases. Calculating R_{isomer} scores corresponding to these spectra confirms the loss of distinct isomer populations, with the R_{isomer} scores decreasing from 3.5 to 1.6 as represented by the red line in Figure 2.4. Only the harshness of the ESI conditions is being changed in this series of results, which confirm that activation in the source region can lead to scrambling of cis/trans isomers.

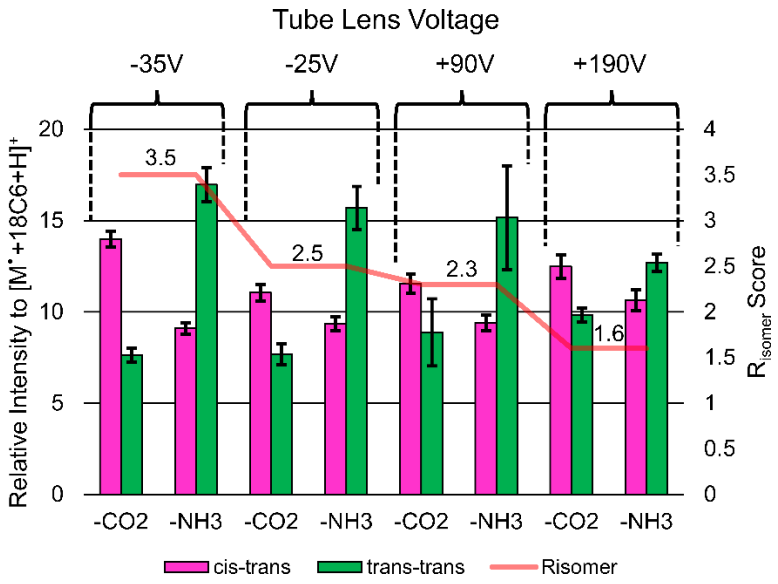


Figure 2.4. -CO_2 and -NH_3 intensities normalized to $[M^{\bullet}+18C6+H]^+$. Error bars represent the standard deviation of the mean. The cis-trans and trans-trans fractions are represented by the pink and green bars, respectively. The change in R_{isomer} score as determined from the -CO_2 and -NH_3 losses in each pair of fractions is represented by the transparent red line. The fragments gradually become more similar in intensity as tube lens voltage is increased, yielding a corresponding decrease in R_{isomer} score.

When the identical experiment is conducted without addition of 18C6, no significant difference is detected in the losses of CO_2 and NH_3 (see Fig 2.S4). Overall, our data reveal that a combination of charge solvation and gentle ESI conditions are necessary to preserve proline cis/trans isomers during transfer into the gas phase. Use of either of these measures alone is not enough to preserve the isomeric forms. These results suggest that the forces driving charge solvation in the gas phase are sufficient to cause rearrangement of peptide structure and overcome cis/trans isomerization barriers for proline. Consideration of the relevant energetics related to charge solvation is consistent with this idea. For example, estimates of the gas-phase binding energy of 18C6 to a protonated primary amine are in the range of 150-224 kJ/mol.[49,50] Although 18C6 is

ideal for solvating primary amines and solvation energies derived from more ad hoc intramolecular interactions within a peptide may be lower in magnitude, it is still easily conceivable that such solvation energies could exceed the 54.4 kJ/mol4 required for proline to overcome the barrier between trans/cis configuration. Given that the peptides studied herein are slow to isomerize, it is likely that other sequences would be equally or more difficult to preserve through the process of ESI. Indeed, the relevant energetics suggest that it may be possible for non-proline peptide bonds to be scrambled during ESI due to molecular rearrangement driven by charge solvation.

Conclusion

Our analysis of the WSGYPPEE synthetic peptide and an iodo-tyrosine variant highlight the difficulties associated with transfer of solution-phase cis/trans proline isomers into the gas phase for analysis. A combination of soft ESI conditions and charge solvation, for which we found 18C6 to be suitable, are required to enable transfer with high fidelity. It is also clear that collisional activation once in the gas phase can also facilitate cis/trans isomer scrambling. Although this suggests that future identification of cis/trans prolyl isomers by simple fragmentation-based experiments will be challenging, it also implies that other experiments examining stereoisomers or constitutional isomers under typical MS conditions are unlikely to be complicated by interference from cis/trans proline isomers. Furthermore, these results make it clear that intramolecular charge solvation in the gas phase is a powerful force that can easily overcome some structural preferences that may exist in solution. Due to their highly labile nature, the application of

gas-phase results to cis-trans prolyl isomers should probably be supplemented by experiments directly examining the same systems in solution.

Acknowledgements. The authors are grateful for funding from the NSF (1904577).

Supporting Information.

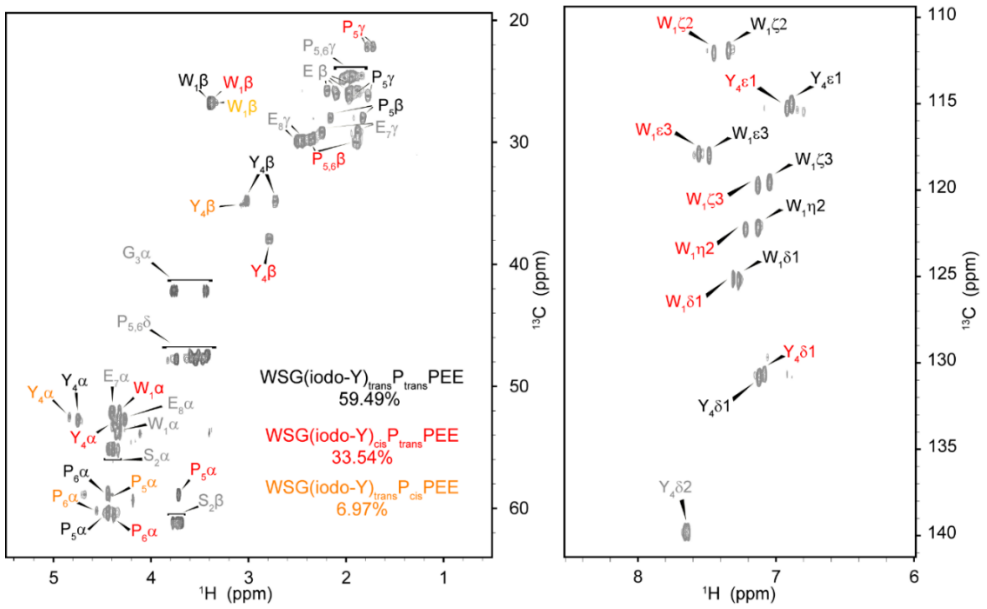


Figure 2.S1: 2D ^{1}H-^{13}C HSQC NMR spectra of peptide WSG(I-Y)PPEE at pH 2.8 with assigned cross peaks labeled. The aliphatic portion of the spectrum is shown on the left and the aromatic on the right. The peaks corresponding to peptide conformation $(iodo\text{-}Y)_{trans}P_{trans}P$, $(iodo\text{-}Y)_{cis}P_{trans}P$ and $(iodo\text{-}Y)_{trans}P_{cis}P$ are shown in black, red, and orange respectively. Crosspeaks that are not distinct for each isomer are labeled in gray. Due to low signal intensity, no peaks corresponding to the $(iodo\text{-}Y)_{trans}P_{cis}P$ conformation could be assigned in the aromatic region.

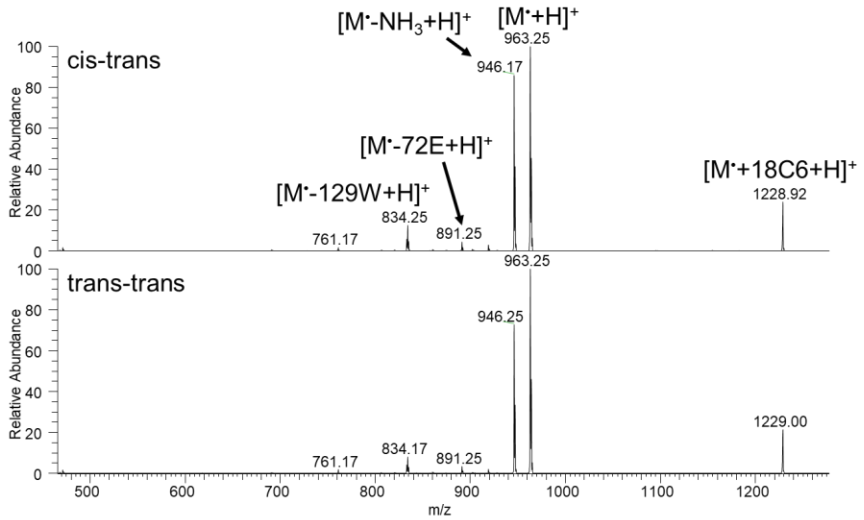


Figure 2.S2. RDD spectra for separated cis-trans and trans-trans isomers under gentle ESI conditions (tube lens -35V) with 18-crown-6. Photodissociation of $[M+18C6+H]^+$ is performed first, followed by isolation and excitation of $[M^{\bullet}+18C6+H]^+$.

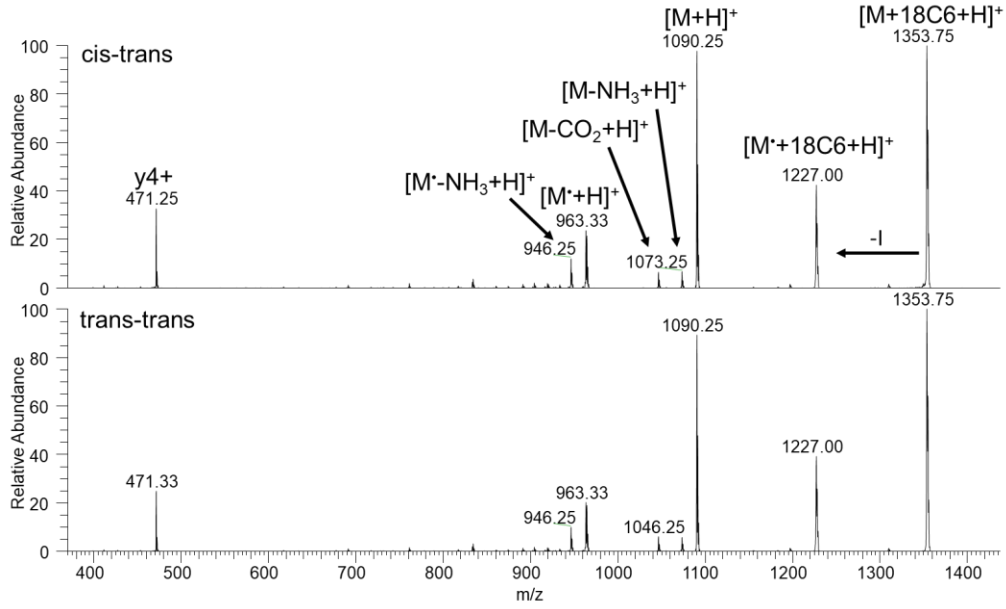


Figure 2.S3. Photodissociation of $[M+18C6+H]^+$ under gentle ESI conditions (tube lens -35V) and with 18-crown-6 after the separated cis-trans and trans-trans isomers are allowed to equilibrate at room temperature.

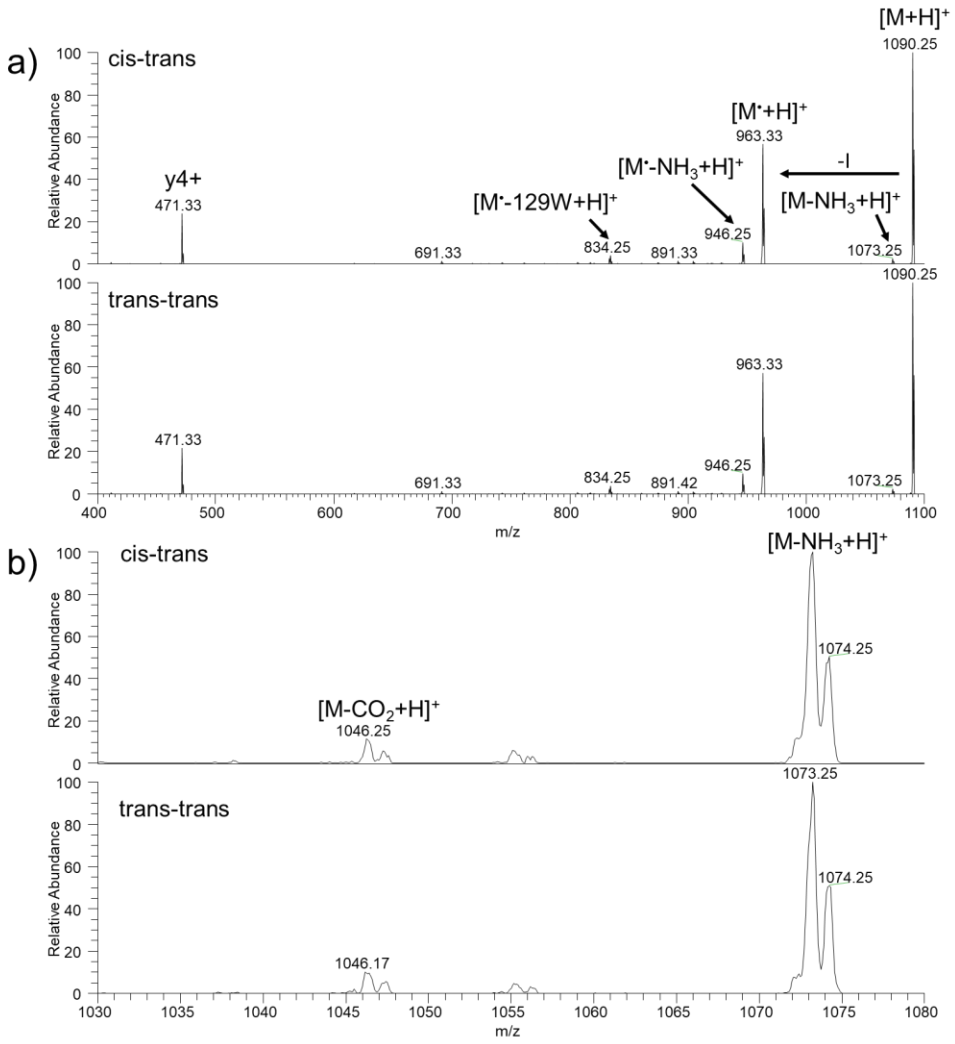


Figure 2.S4. (a) Full spectra of photodissociation of $[M+H]^+$ under gentle ESI conditions (tube lens -25V) after separation of cis-trans and trans-trans isomers. **(b)** Magnification of 1030-1080 m/z showing $[M-NH_3+H]^+$ and $[M-CO_2+H]^+$ neutral losses.

Table 2.S1: NMR Assignments

Conformer	residue	HA	HA2	HB	HB2	HG	HG2	HD	HD2	HE	HE2
1H (Trans-Trans)	W1	4.336		3.398				7.272			
1H (Cis-Trans)	W1	4.326		3.36				7.31			
1H (Trans-Cis)	W1			3.322							
1H	S2	4.437-4.337*		3.776-3.692*	3.776-3.692*						
1H	G3	3.763	3.441								
1H (Trans-Trans)	Iodo-Y4	4.749		3.027	2.733			7.12	7.634	6.921	
1H (Cis-Trans)	Iodo-Y4	4.372		2.769				7.081	7.658	6.886	
1H (Trans-Cis)	Iodo-Y4	4.836		3.077	3.051						
1H (Trans-Trans)	P5	4.444		2.16	1.832	1.974	1.773	3.739-3.431*			
1H (Cis-Trans)	P5	3.717		2.363	2.335	1.783	1.723	3.739-3.431*			
1H (Trans-Cis)	P5	4.4									
1H (Trans-Trans)	P6	4.442									
1H (Cis-Trans)	P6	4.379		1.911	1.882						
1H (Trans-Cis)	P6	4.555									
1H	E7	4.398		2.199-1.885*		2.247-1.879*					
1H	E8	4.266		2.199-1.885*		2.498-2.445*					
* several peaks were observed, but they could not be assigned to a specific conformation (cis/trans or trans/trans).											

…continued

Conformer	residue	HE3	HH2	HZ2	HZ3	CA	CB	CG	CD	CE	CD1	CE3	CZ2	CZ3	CH2	CD2
1H (Trans-Trans)	W1	7.483	7.13	7.341	7.045	53.385	26.745				125.228	117.992	111.906	119.505	122.15	
1H (Cis-Trans)	W1	7.557	7.22	7.448	7.132	51.932	26.748				125.132	117.892	112.064	119.729	122.27	
1H (Trans-Cis)	W1						26.776									
1H	S2					55.153	61.158									
1H	G3					42.129										
1H (Trans-Trans)	Iodo-Y4					52.723	34.785		130.833	115.011						139.795
1H (Cis-Trans)	Iodo-Y4					53.019	37.896		130.659	115.267						139.795
1H (Trans-Cis)	Iodo-Y4					52.497	35.117									
1H (Trans-Trans)	P5					60.289	27.99	26.188	47.781-47.495*							
1H (Cis-Trans)	P5					58.892	29.768	22.161	47.781-47.495*							
1H (Trans-Cis)	P5					58.889			47.781-47.495*							
1H (Trans-Trans)	P6					58.808										
1H (Cis-Trans)	P6					60.458	29.973									
1H (Trans-Cis)	P6					60.182										
1H	E7					52.087	26.014-25.	29.19								
1H	E8					52.668	26.014-25.	29.871								

References

[1] Rode, B. M. Peptides and the Origin of Life. *Peptides* **1999**, *20* (6), 773–786.
[2] Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G. Side Chain Effects On Peptidyl-Prolyl Cis/Trans Isomerization. *J. Mol. Biol.* **1998**, *279* (2), 449–460.
[3] Ramachandran, G. N.; Sasisekharan, V. Conformation of Polypeptides and Proteins. *Adv. Protein Chem.* **1968**, *23* (C), 283–437.
[4] Stewart, D. E.; Sarkar, A.; Wampler, J. E. Occurrence and Role of cis Peptide Bonds in Protein Structures. *J. Mol. Biol.* **1990**, *214* (1), 253–260.
[5] Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. Preferred Proline Puckerings in Cis and Trans Peptide Groups: Implications for Collagen Stability. *Protein Sci.* **2009**, *10* (12), 2627–2632.
[6] Wu, W. J.; Raleigh, D. P. Local Control of Peptide Conformation: Stabilization of Cis Proline Peptide Bonds by Aromatic Proline Interactions. *Biopolymers* **1998**, *45* (5), 381–394.
[7] Zondlo, N. J. Aromatic-Proline Interactions: Electronically Tunable CH/π Interactions. *Acc. Chem. Res.* **2013**, *46* (4), 1039–1049.
[8] Brown, A. M.; Zondlo, N. J. A Propensity Scale for Type II Polyproline Helices (PPII): Aromatic Amino Acids in Proline-Rich Sequences Strongly Disfavor PPII Due to Proline-Aromatic Interactions. *Biochemistry* **2012**, *51* (25), 5041–5051.
[9] Alderson, T. R.; Lee, J. H.; Charlier, C.; Ying, J.; Bax, A. Propensity for Cis -Proline Formation in Unfolded Proteins. *ChemBioChem* **2018**, *19* (1), 37–42.
[10] Buevich, A. V.; Dai, Q. H.; Liu, X.; Brodsky, B.; Baum, J. Site-Specific NMR Monitoring of Cis-Trans Isomerization in the Folding of the Proline-Rich Collagen Triple Helix. *Biochemistry* **2000**, *39* (15), 4299–4308.
[11] Farber, P.; Darmawan, H.; Sprules, T.; Mittermaier, A. Analyzing Protein Folding Cooperativity by Differential Scanning Calorimetry and NMR Spectroscopy. *J. Am. Chem. Soc.* **2010**, *132* (17), 6214–6222.
[12] Sarkar, P.; Reichman, C.; Saleh, T.; Birge, R. B.; Kalodimos, C. G. Proline Cis-Trans Isomerization Controls Autoinhibition of a Signaling Protein. *Mol. Cell* **2007**, *25* (3), 413–426.
[13] Lummis, S. C. R.; Beene, D. L.; Lee, L. W.; Lester, H. A.; Broadhurst, R. W.; Dougherty, D. A. Cis-Trans Isomerization at a Proline Opens the Pore of a Neurotransmitter-Gated Ion Channel. *Nature* **2005**, *438* (7065), 248–252.
[14] Shinoda, K.; Fujitani, H. Initiation of Prolyl Cis-Trans Isomerisation in the CDR-H3 Loop of an Antibody in Response to Antigen Binding. *Sci. Rep.* **2017**, *7* (1), 16964.
[15] Fassolari, M.; Chemes, L. B.; Gallo, M.; Smal, C.; Sánchez, I. E.; de Prat-Gay, G. Minute Time Scale Prolyl Isomerization Governs Antibody Recognition of an Intrinsically Disordered Immunodominant Epitope. *J. Biol. Chem.* **2013**, *288* (18), 13110–13123.
[16] Mitragotri, S.; Burke, P. A.; Langer, R. Overcoming the Challenges in Administering Biopharmaceuticals: Formulation and Delivery Strategies. *Nat. Rev. Drug Discov.* **2014**, *13* (9), 655–672.
[17] Huang, J.; Ofek, G.; Laub, L.; Louder, M. K.; Doria-rose, N. A.; Longo, N. S.; Imamichi, H.; Bailer, R. T.; Chakrabarti, B.; Sharma, S. K.; Alam, S. M.; Wang, T.; Yang, Y.; Zhang, B.; Migueles, S. A.; Wyatt, R.; Haynes, B. F.; Kwong, P. D.; Mascola, J. R.; Connors, M. Broad and Potent Neutralization of HIV-1 by a Gp41-Specific Human Antibody. *Nature* **2012**, *491* (V), 406–412.
[18] Masiero, A.; Nelly, L.; Marianne, G.; Christophe, S.; Florian, L.; Ronan, C.; Claire, B.; Cornelia, Z.; Grégoire, B.; Eric, L.; Ludovic, L.; Dominique, B.; Sylvie, A.; Marie G.; Francis, D.; Fabienne, S.; Cécile, C.; Isabelle, A.; Jacques, D.; Jérôme, D.; Bruno, G.; Katarina, R.; Jean-Michel, M.; Catherine, P. The Impact of Proline Isomerization on Antigen Binding and the Analytical Profile of a Trispecific Anti-HIV Antibody. *MAbs* **2020**, *12* (1).
[19] Counterman, A. E.; Clemmer, D. E. Cis-Trans Signatures of Proline-Containing Tryptic Peptides in the Gas Phase. *Anal. Chem.* **2002**, *74* (9), 1946–1951.

[20] Pierson, N. A.; Chen, L.; Russell, D. H.; Clemmer, D. E. Cis-Trans Isomerizations of Proline Residues Are Key to Bradykinin Conformations. *J. Am. Chem. Soc.* **2013**, *135* (8), 3186–3192.
[21] Voronina, L.; Masson, A.; Kamrath, M.; Schubert, F.; Clemmer, D.; Baldauf, C.; Rizzo, T. Conformations of Prolyl-Peptide Bonds in the Bradykinin 1-5 Fragment in Solution and in the Gas Phase. *J. Am. Chem. Soc.* **2016**, *138* (29), 9224–9233.
[22] Shi, L.; Holliday, A. E.; Shi, H.; Zhu, F.; Ewing, M. A.; Russell, D. H.; Clemmer, D. E. . Characterizing Intermediates along the Transition from Polyproline i to Polyproline II Using Ion Mobility Spectrometry-Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136* (36), 12702–12711.
[23] Shi, L.; Holliday, A. E.; Glover, M. S.; Ewing, M. A.; Russell, D. H.; Clemmer, D. E. Ion Mobility-Mass Spectrometry Reveals the Energetics of Intermediates That Guide Polyproline Folding. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (1), 22–30.
[24] Warnke, S.; Baldauf, C.; Bowers, M. T.; Pagel, K.; Von Helden, G Photodissociation of Conformer-Selected Ubiquitin Ions Reveals Site-Specific Cis / Trans Isomerization of Proline Peptide Bonds. *J. Am. Chem. Soc.* **2014**, *136* (29), 10308–10314.
[25] Martens, J. K.; Grzetic, J.; Berden, G.; Oomens, J. Gas-Phase Conformations of Small Polyprolines and Their Fragment Ions by IRMPD Spectroscopy. *Int. J. Mass Spectrom.* **2015**, *377* (1), 179–187.
[26] Morrison, L. J.; Wysocki, V. H. Low Energy CID and Action IRMPD Provide Insights into a Minor Subpopulation of the Gas-Phase Conformers of Triply Charged Bradykinin. *Int. J. Mass Spectrom.* **2015**, *391*, 2–10.
[27] Fenn, J. B. Electrospray Wings for Molecular Elephants (Nobel Lecture). *Angew. Chemie Int. Ed.* **2003**, *42* (33), 3871–3894. https://doi.org/10.1002/anie.200300605.
[28] Loscertales, I. G.; Fernández de la Mora, J. Experiments on the Kinetics of Field Evaporation of Small Ions from Droplets. *J. Chem. Phys.* **1995**, *103* (12), 5041–5060. https://doi.org/10.1063/1.470591.
[29] Kebarle, P.; Verkerk, U. H. Electrospray: From Ions in Solution to Ions in the Gas Phase, What We Know Now. *Mass Spectrom. Rev.* **2009**, *28* (6), 898–917. https://doi.org/10.1002/mas.20247.
[30] Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9.
[31] Metwally, H.; Duez, Q.; Konermann, L. Chain Ejection Model for Electrospray Ionization of Unfolded Proteins: Evidence from Atomistic Simulations and Ion Mobility Spectrometry. *Anal. Chem.* **2018**, *90* (16), 10069–10077.
[32] Konermann, L.; Metwally, H.; Duez, Q.; Peters, I. Charging and Supercharging of Proteins for Mass Spectrometry: Recent Insights into the Mechanisms of Electrospray Ionization. *Analyst* **2019**, *144* (21), 6157–6171.
[33] Breuker, K.; McLafferty, F. W. Stepwise Evolution of Protein Native Structure with Electrospray into the Gas Phase, 10^{-12} to 10^{2} s. *Proc. Natl. Acad. Sci.* **2008**, *105* (47), 18145–18152.
[34] Hood, C. A., Fuentes, G., Patel, H., Page, K., Menakuru, M., Park, J. H. Fast Conventional Fmoc Solid-Phase Peptide Synthesis With. *J. Pept. Sci.* **2008**, 14, 97–101.
[35] Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Wenger, R. K.; Yao, H.; Markley, J. L. BioMagResBank. *Nucleic Acids Res.* **2007**, *36* (Database), D402–D408.
[36] Grathwohl, C.; Wüthrich, K. Nmr Studies of the Rates of Proline Cis - Trans Isomerization in Oligopeptides. *Biopolymers* **1981**, *20* (12), 2623–2633.
[37] Shen, Y.; Bax, A. Prediction of Xaa-Pro Peptide Bond Conformation from Sequence and Chemical Shifts. *J. Biomol. NMR* **2010**, *46* (3), 199–204.
[38] Spencer, E. A. C.; Ly, T.; Julian, R. R. Formation of the Serine Octamer : Ion Evaporation or Charge Residue? *Int. J. Mass Spectrom.* **2008**, *270*, 166–172.
[39] Tao, W. A.; Zhang, D.; Nikolaev, E. N.; Cooks, R. G. Copper (II) -Assisted Enantiomeric Analysis of D , L -Amino Acids Using the Kinetic Method : Chiral Recognition and Quantification in the Gas Phase. *J. Am. Chem. Soc.* **2000**, 122, 43, 10598-10609.
[40] Grewal, R. N.; Aribi, H. El; Harrison, A. G.; Siu, K. W. M.; Hopkinson, A. C. Fragmentation of Protonated Tripeptides : The Proline Effect Revisited. *J. Phys. Chem. B* **2004**, 108, 15, 4899-4908.
[41] 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.

[42] Tao, Y.; Quebbemann, N. R.; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820.
[43] Riggs, D. L.; Silzel, J. W.; Lyon, Y. A.; Kang, A. S.; Julian, R. R. Analysis of Glutamine Deamidation: Products, Pathways, and Kinetics. *Anal. Chem.* **2019**, *91*, 13032–13038.
[44] Lyon, Y. A.; Sabbah, G. M.; Julian, R. R. Identification of Sequence Similarities among Isomerization Hotspots in Crystallin Proteins. *J. Proteome Res.* **2017**, *16* (4), 1797–1805.
[45] Ly, T.; Julian, R. R. Using ESI-MS to Probe Protein Structure by Site-Specific Noncovalent Attachment of 18-Crown-6. *J. Am. Soc. Mass Spectr.* **2006** 17, 1209–1215.
[46] Bonner, J. G.; Hendricks, N. G.; Julian, R. R. Structural Effects of Solvation by 18-Crown-6 on Gaseous Peptides and TrpCage after Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (10), 1661–1669.
[47] Pérot, M.; Lucas, B.; Barat, M.; Fayeton, J. A.; Jouvet, C. Mechanisms of UV Photodissociation of Small Protonated Peptides. *J. Phys. Chem. A* **2010**, *114* (9), 3147–3156.
[48] Kadhane, U.; Andersen, J. U.; Ehlerding, A.; Hvelplund, P.; Kirketerp, M.-B. S.; Lykkegaard, M. K.; Nielsen, S. B.; Panja, S.; Wyer, J. A.; Zettergren, H. Photodissociation of Protonated Tryptophan and Alteration of Dissociation Pathways by Complexation with Crown Ether. *J. Chem. Phys.* **2008**, *129* (18), 184304.
[49] Julian, R. R.; Beauchamp, J. L. Site Specific Sequestering and Stabilization of Charge in Peptides by Supramolecular Adduct Formation with 18-Crown-6 Ether by Way of Electrospray Ionization. *Int. J. Mass Spectrom.* **2001**, *210* (1–3), 613–623.
[50] Chen, Y.; Rodgers, M. T. Structural and Energetic Effects in the Molecular Recognition of Protonated Peptidomimetic Bases by 18-Crown-6. *J. Am. Chem. Soc.* **2012**, *134* (4), 2313–2324.

CHAPTER 3: PIMT-Mediated Labeling of L-isoAspartic Acid with Tris Facilitates Identification of Isomerization Sites in Long-Lived Proteins

Abstract

Isomerization of individual residues in long-lived proteins (LLPs) is a subject of growing interest in connection with many age-related human diseases. When isomerization occurs in LLPs, it can lead to deleterious changes in protein structure, function, and proteolytic degradation. Herein, we present a novel labeling technique for rapid identification of L-isoAsp using the enzyme Protein L-isoAspartyl Methyltransferase (PIMT) and tris. The succinimide intermediate formed during reaction of L-isoAsp-containing peptides with PIMT and S-Adenosyl Methionine (SAM) is reactive with tris base and results in a tris-modified aspartic acid residue with a mass shift of +103 Daltons. Tris-modified aspartic acid exhibits prominent and repeated neutral loss of water when subjected to collisional activation. In addition, another dissociation pathway regenerates the original peptide following loss of a characteristic mass shift. Furthermore, it is demonstrated that tris modification can be used to identify sites of isomerization in LLPs from biological samples such as the lens of the eye. This approach simplifies identification by labeling isomerization sites with a tag that causes a mass shift and provides characteristic loss during collisional activation.

Introduction

Although most of the tens-of-thousands of proteins found in the human body turn over rapidly, there are others with substantially longer half-lives. These long-lived proteins (LLPs) have half-lives ranging from 48 hours to roughly the equivalent to human lifetime.[1] Rapid recycling of short-lived proteins is accompanied by a substantial energetic cost, but one advantage is that the window for spontaneous chemical modifications to occur is minimized. In contrast, LLPs can accumulate many modifications during their residence time in an organism.[2] For example, isomerization of aspartic acid residues builds up gradually in LLPs, which can cause significant impact on structure and function. Such chemistry is observed in the crystallins of the human eye lens,[3] which represent extreme LLP's due to the absence of protein turnover in mature lens fiber cells.[4] The accumulation of isomerization in lens crystallins increases with age and has been connected with protein structure and solubility,[5] including changes that may be associated with cataract formation.[6] For example, recent work has shown that as αB-crystallin ages, isomerization of asp 109 causes disruption of a dimer-stabilizing salt bridge and facilitates greater aggregation of the protein.[7]

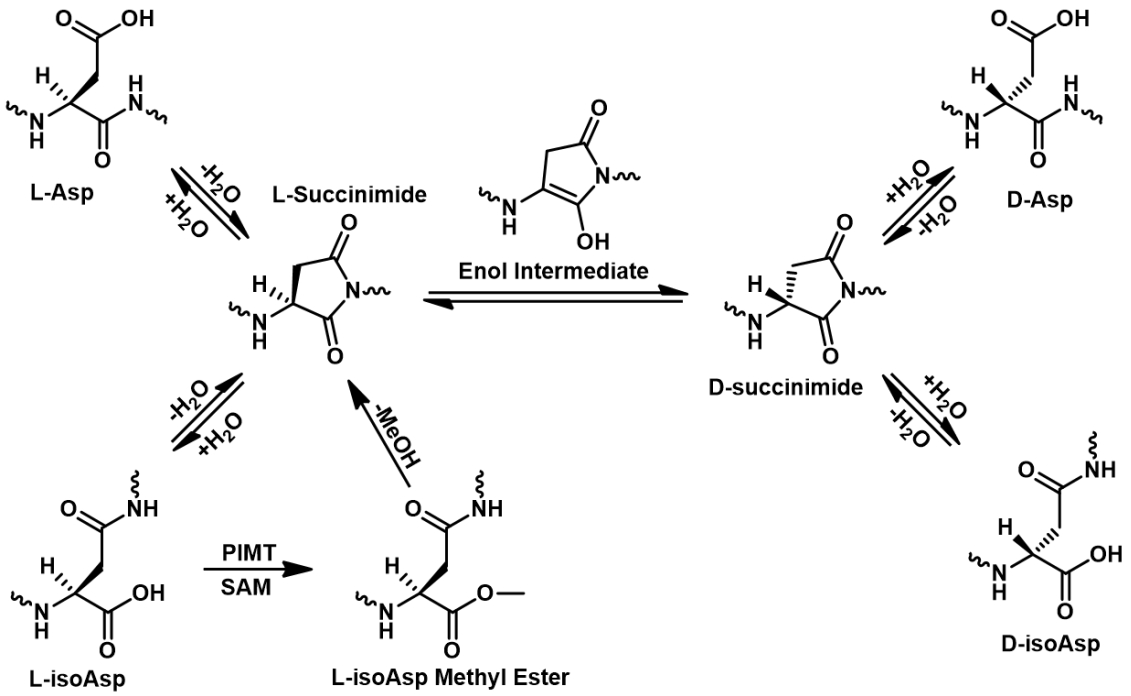
The detection of isomerization in LLPs poses a challenge to mass spectrometry because there is no accompanied mass shift, which has led to reliance on separation by liquid chromatography to aid in identification. Multiple peaks in a chromatogram for a peptide sequence containing aspartic acid or other residues capable of isomerization are often indicative of the presence of isomers. Reverse-phase liquid chromatography frequently enables separation of isomers and has been used to this end to examine isomerization in

LLPs in the eye lens,[8] as well as achieving separation of synthetic Aβ epimers.[9] More recently, ion mobility has been used for separations in conjunction with mass spectrometry for analysis of isomerization.[10,11,12] In fact, ion mobility can sometimes be used to identify actual sites of isomerization by separate analysis of the b/y ion fragments produced by collision-induced dissociation (CID).[13,14] In addition to separation-based methods, fragmentation techniques such as radical-directed dissociation (RDD) can be used to identify the presence of isomers.[15] ETD/ECD has also been shown to be capable of identification of L-isoAsp via diagnostic c+57 and z-57 ions,[16] although these unique ions are frequently quite low in abundance relative to other fragment ions.[17]

The structures of isomerized residues in proteins differ substantially from the canonical form, which can prevent enzymes from accommodating them into active sites for processing. For example, recent work showed that lysosomal cathepsins are unable to completely digest isomerized protein substrates.[18] Similarly impaired recognition and degradation of isomerized substrates has been used intentionally to enrich isomer populations with proteases such as aminopeptidase M (APM).[19] It is also possible to use enzymes capable of recognizing specific isomers. For example, with aspartic acid isomers the activities of Asp-N, Paenidase, and Protein L-isoAspartyl Methyltransferase (PIMT) can be used advantageously to target specific substrates.[20] Asp-N reacts with L-Asp only, cleaving peptide bonds on the N-terminal side of L-Asp. Paenidase, on the other hand, recognizes D-Asp exclusively and cleaves substrates on the C-terminal side of D-Asp.[21] PIMT is a protective enzyme that combats age-related isomerization in

proteins by selective recognition of L-isoAsp residues, which are methylated and partially repaired back to L-Asp.

PIMT is unusual in its ability to recognize L-isoAsp, but it is widespread throughout many different organisms including gram-negative bacteria, plants, fungi, invertebrates, and vertebrates and is thought to be essential to most life forms.[22] Studies of PIMT in mice and zebrafish confirm the essential nature of this enzyme, as mice lacking PIMT died of epileptic seizures within 4-8 weeks of birth,[23,24] and PIMT knockdown in zebrafish larvae resulted in decreased survivability and disruption of brain calcium homeostasis.[25] PIMT-mediated repair of L-isoAsp begins with methylation of the side chain by reaction with S-Adenosyl Methionine (SAM). The methylester intermediate converts readily to a five-membered succinimide ring, which then undergoes hydrolysis and reopens to form one of the four possible aspartic acid isomers as illustrated in Scheme 3.1. The primary product upon reopening of the succinimide ring is L-isoAsp,[26,27] which can be methylated again. Eventually in the presence of sufficient PIMT and SAM, L-isoAsp is funneled through this mechanism to form predominantly L-Asp with minor amounts of the D-Asp isomers. In addition to recognition of L-isoAsp, PIMT is also sensitive to the identities of the surrounding residues in the P2 to the P3' positions.[28] Although D-Asp is also recognized by PIMT, L-isoAsp is preferred ~1000-fold over D-Asp as a substrate, likely due to repositioning of the Asp side chain which may introduce steric clash at the PIMT active site.[29]



Scheme 3.1: Aspartic acid isomerization and PIMT-mediated repair of L-isoAsp

PIMT has been explored as a tool for the identification of L-isoAsp. The methylation of L-isoAsp with PIMT generates products with mass shifts of +14 Da (for the methyl ester) and -18 Da (for the succinimide ring). In principle, these mass shifts could be used to identify peptides containing L-isoAsp, but both products are short-lived intermediates with half-lives on the order of hours that are not stable in aqueous solutions.[27,30] Due to these factors, samples cannot be analyzed multiple times if locating these intermediates is the goal. However, the succinimide intermediate's innate reactivity presents an opportunity if a mass-shifting reagent that can react on a timescale competitive with water can be used. This can be achieved with Oxygen-18 water, yielding aspartic acid residues labeled with a mass shift of +2 Da after hydrolysis of the succinimide.[31]

Hydrazine can also react with the succinimide ring, resulting in a mass shift of +14 Da at the modified aspartic acid and allowing for affinity enrichment via aldehyde resins.[32,33] In this case, reaction of L-isoAsp with PIMT in the presence of 1M hydrazine was able to compete with the ongoing hydrolysis of the succinimide intermediate, resulting in some aminolysis product formation. The succinimide ring appears to be amine reactive in general, as it has also been found that tris can react in solution with succinimide rings formed during deamidation of IgG peptides.[34] While the covalent modification of peptides with tris during deamidation is potentially undesirable, in the context of L-isoAsp discovery, it presents an opportunity for a convenient labeling technique requiring nothing more than concentrated tris buffer and PIMT.

Herein, we present a simple PIMT-mediated method to label L-isoAsp with tris. PIMT is used to specifically convert L-isoAsp to a succinimide form, followed by aminolysis of the succinimide with tris base. This introduces an easily detectable mass shift of +103 Da at the modified aspartic acid and leads to retention-time shifts that allow separation from unmodified peptides. Collisional activation of the tris-modified peptides results in unusually abundant losses of multiple water molecules. The tris modification itself can also be lost, restoring the original, unmodified peptide. The method is demonstrated with L-isoAsp identification in an eye lens digest.

Experimental Methods

*Materials***.** Organic solvents and reagents were purchased from Fisher Scientific, Sigma-Aldrich, or Acros Organics and used without further purification. FMOC-

protected amino acids and Wang resins were purchased from Anaspec, Inc or Chem-Impex International. PIMT was purchased from Fitzgerald Industries (Cat. # 80R-1342), SAM was purchased from Sigma (a7007), and tris base from Fisher Bioreagents (BP152).

Peptide Synthesis. Peptides were synthesized manually following a FMOC-protected solid-phase peptide synthesis protocol.[35] Following synthesis, peptides were RP-HPLC purified using a Thermo UltiMate 3000 RS variable wavelength detector and pump connected to a Phenomenex Kinetex 5µm EVO C18 250 x 21.2 mm column (P/N 00G-4633-P0-AX) with Phenomenex SecurityGuard PREP Cartridge (P/N AJ0-9145). Samples were purified using water with 0.1% formic acid as mobile phase A, and acetonitrile with 0.1% formic acid as mobile phase B. Purified peptides were stored frozen in 50/50 v/v acetonitrile/water. Peptides were lyophilized and reconstituted in Fisher Optima water, prior to determining concentrations and reaction with PIMT in buffer.

PIMT Reactions on L-isoAsp Synthetic Peptides. Purified L-isoAsp synthetic peptides were combined together at a concentration of 40 µM for each sequence and reacted with PIMT at a molar ratio of 1:50 PIMT:Peptide with 400 µM SAM and 1.0M Tris-HCl pH 7.4 at 37C for 24 hours. All time points were quenched with TFA to pH <4 to stop further reaction with PIMT.

PIMT Reactions in Tryptic Digests. Tryptic digests of the crystallins from the eye lens were prepared according to a previously established method.[36] 10µg of desalted and lyophilized tryptic digests of 72-year-old crystallins from the water insoluble nucleus fraction in the eye lens were reconstituted to 0.2 µg/µL in 1.0M Tris-HCl pH 7.4, 400µM

SAM and reacted with 5 µg PIMT for 24h at 37C. Time points were quenched with TFA to pH <4 to halt reaction with PIMT.

Analysis. For LCMS analysis of synthetic peptides, an Agilent 1100 binary pump was used with a Thermo BetaBasic-18 3 µm C18 150 mm x 2.1 mm column interfaced to a Thermo Fisher Scientific LTQ mass spectrometer. Samples were eluted using water with 0.1% formic acid as mobile phase A, and acetonitrile with 0.1% formic acid as mobile phase B. A gradient of 5-65%B in 60 minutes was used to separate isomers. For sequential collisional activation of APSWFD^{+Tris}TGLSEMR, synthetic peptides were first desalted and then analyzed via nano-ESI with tips pulled from 1mm OD x 0.58mm ID x 100mm L borosilicate glass (Harvard Apparatus, P/N 30-0016). For LCMS analysis of tryptic digests of crystallins from human eye lens, a Thermo UltiMate 3000 RSLC was used with an emitter pulled from 100 µm ID silica (Polymicro technologies, P/N 1068150023) packed to a length of 24cm with Reprosil-Pur 120 C18-AQ 3um (Dr. Maisch GmbH) interfaced to a Thermo Fisher Scientific Orbitrap Velos Pro. Samples were eluted using 0.1% formic acid in water as mobile phase A, and 0.1% formic acid in 80% acetonitrile as mobile phase B. Samples were desalted by loading onto a PepMap100 5µm C18 1mm x 5mm trap cartridge (Thermo, P/N 160434) at 25 µL/min for 2 minutes at 1%B. Peptides were then separated with a gradient of 5-65%B over the course of 120 minutes. During data collection, the Orbitrap was operated in dynamic exclusion mode with a repeat count of 4, repeat duration of 15s, and an exclusion duration of 15s. Data was analyzed using MaxQuant (version 1.6.5.0) and searched against proteins previously established to be present in the eye lens.[37] For the search

parameters, trypsin with a maximum of 2 missed cleavages was specified, and carbamidomethylation on cysteines was set as a fixed modification. Met oxidation, Asn deamidation, pyro-Gln/Glu formation, and custom Asp+tris and deamidated Asn+tris modifications were set as variable modifications.

Results and Discussion

Labeling of Synthetic L-isoAsp Peptides with Tris. Reaction of synthetic the L-isoAsp-containing peptide APSWF**D**TGLSEMR (L-isoAsp in bold) with PIMT and SAM in tris buffer yields one baseline-resolved peak and two partially resolved peaks with +103 Da mass shifts, corresponding to addition of tris base to the aspartic acid side chain (Figure 3.1a). The presence of three peaks with the same mass shift suggests that the tris base attacked both L- and D-succinimide rings, forming a combination of L/D and L/D-iso tris-modified Asp. Fragmentation of the mass-shifted peaks reveals similar spectra for all three, a prominent neutral loss of water and production of a series of b- and y-ions (Figure 3.1b). Many of the b- and y-ions formed during collisional activation of the tris-modified peptide are accompanied by peaks corresponding to secondary loss of water from the fragment ions. Assignment of the b- and y-ions yields the largest number of matches when the tris modification is located on the aspartic acid. To evaluate if the unusual extent of water loss is connected to the tris modification, a CID spectrum for the unmodified peptide was obtained, see Figure 3.1c. The loss of water is significantly less

abundant for the unmodified peptide. Taken together, these results suggest that tris easily loses water, which may be useful for the identification of tris-modified peptides.

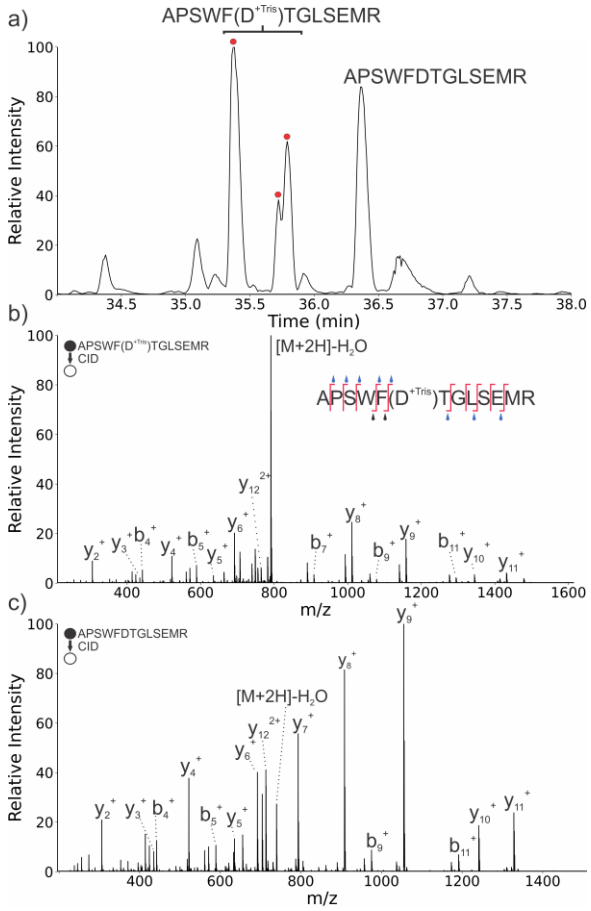
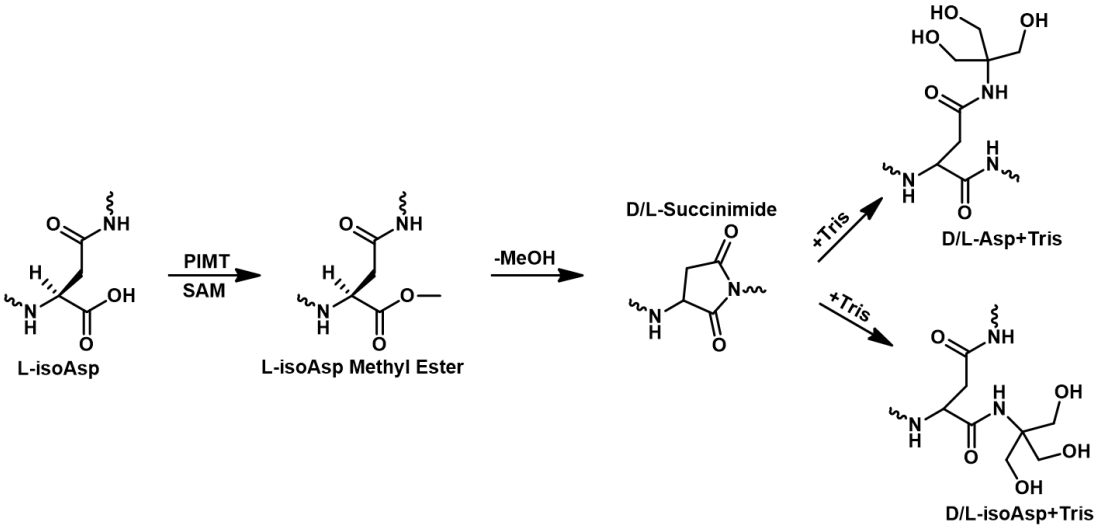


Figure 3.1: (a) Chromatogram of synthetic L-isoAsp peptide APSWF**D**TGLSEMR modified with tris. Tris-modified peaks are labeled with red dots. (b) CID of tris-modified APSWFDTGLSEMR at RT 35.4 min. Water droplets over b/y labels indicate observed b_n/y_n-H_2O fragments, blue droplets contain the tris modification while the black droplets do not. (c) Collisional dissociation of unmodified APSWFDTGLSEMR at RT 36.4 min.

Scheme 3.2 outlines a proposed mechanism for formation of tris-modified aspartic acid. Following reaction of L-isoAsp with PIMT and SAM, the methylated side chain readily converts to a succinimide intermediate. Normally this succinimide is hydrolyzed, forming one of four different isomers of aspartic acid. However, in the presence of high

concentrations, tris base competes with water for nucleophilic attack of the succinimide ring, yielding tris-modified isomers in addition to those created by standard hydrolysis. Although it should be noted that only three out of the four possible tris modification products were observed in Figure 3.1a, coelution of a fourth tris modification product in one of the three observed peaks is possible.



Scheme 3.2: PIMT-mediated tris modification of L-isoAsp residues.

To further explore the favorable water loss observed following collisional activation of APSWFD^{+Tris}TGLSEMR, additional experiments were conducted by nano-ESI. CID of the modified peptide is shown in Figure 3.2a, producing water loss as the most favorable fragment by far (25.8% of the total fragment abundance). Further re-isolation and fragmentation of the water-loss peak in an MS^3 experiment yields a second water loss as the base peak in addition to b- and y-fragment ions that often exhibit additional water losses. Subsequent isolation and fragmentation of the double-water-loss peak yielded an abundant third dehydration, although slightly less intense than the most abundant y-ions (Figure 3.2c). Astonishingly, a significant amount of a fourth water loss can be observed

in an MS^5 experiment, likely originating from another residue in the peptide (Figure 3.2d). In each MS^n experiment, all the b- and y-ions localize dehydration to the tris-modified aspartic acid, except for y_{10} ions (labeled in red) in the MS^4 and MS^5 steps which each contain one less neutral loss than expected. This indicates that although the first water loss likely occurs most favorably from the tris-modified aspartic acid, other residues in the peptide may contribute to the subsequent losses, of which Ser(3) is the most likely source.

Another feature observed in Figure 3.2 is a decrease in the fractional abundance of the neutral water loss relative to the b- and y-ions in the MS^n spectra as each neutral loss is re-isolated and fragmented. This may indicate that the first dehydration from the tris-modified peptide is most favorable while the subsequent dehydrations decrease in favorability, with the added collisional energy instead forming b- and y-ions. In cases where the b- and y-ion abundance and/or coverage is not sufficient to establish sequence identity and/or the location of the tris modification during the MS^2 step, re-isolation and fragmentation of the water loss in an MS^3 step could be used to produce more backbone fragmentation and sequence information.

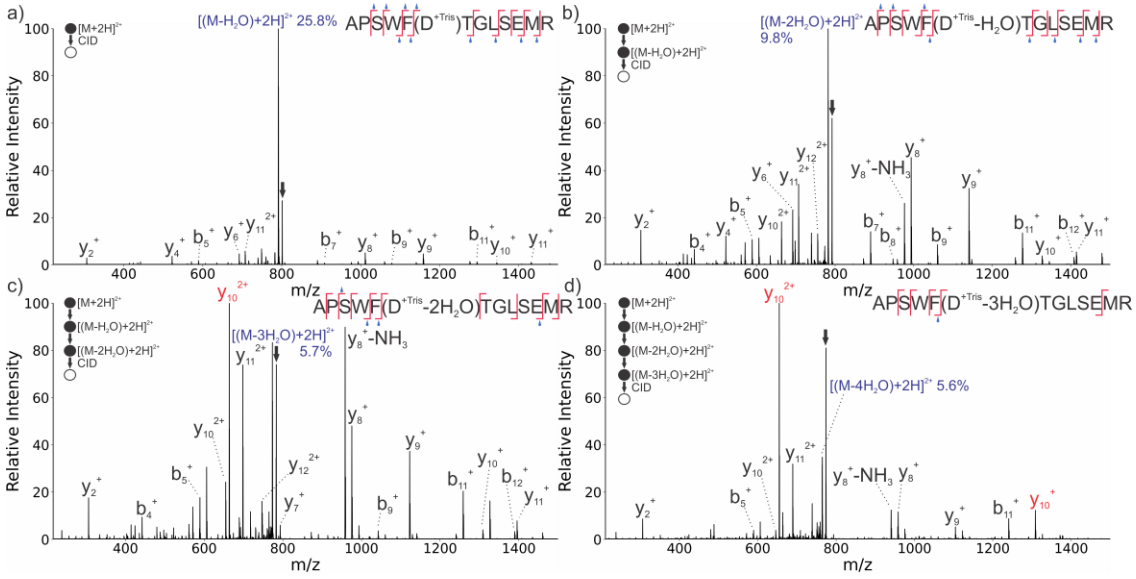


Figure 3.2: Sequential isolation and fragmentation of the neutral loss of water from APSWFD^{+Tris}TGLSEMR. Precursor ions are indicated with black arrows. a) MS^2, b) MS^3, c) MS^4, d) MS^5. y_{10} ions labeled in red correspond to a y_{10} ion with one water loss (c), and y_{10} ions with two water losses (d), with the missing water loss in both cases likely originating from Ser(3). Fractional abundance of the neutral water loss is indicated by percent.

To determine whether PIMT-mediated tris labeling of L-isoAsp residues works for other sequences, a mixture of seven synthetic crystallin- and Aβ-derived L-isoAsp peptides was incubated with PIMT and SAM in 1.0M Tris pH 7.4 for 24h (APSWF**D**TGLSEMR, TVL**D**SGISEVR, HFSPE**D**LTVK, VGS**D**KGAIIG, IQTGL**D**ATHAER, GYQYLLEPG**D**FR, and GISEVRS**D**R). All peptides were modified by tris in amounts ranging from 3% to 64%, calculated as a percent of the total peak area for the unmodified and modified sequence (Figure 3.S1). The average modification percentage for these seven sequences was 32%. Tris modification of GISEVRSDR (3%) was considerably lower than for the other peptides, with the second lowest modification percentage being GYQYLLEPGDFR (18%). PIMT is known to interact with the three residues C-terminal to L-isoAsp and is known to have decreased activity for substrates

near the C-terminus.[28] It is likely that the low yield of tris modification for GISEVRSDR (3%) results from poor binding and methylation by PIMT. The effect of pH was investigated by repeating the experiment at pH 6.6, 7.0, 7.8, and 8.2. Significant changes in the tris modification yield were not observed and increasing the pH to 9.0 and 10.0 was found to drastically decrease the yield, presumably due to reduced PIMT activity (Figure 3.S2, 3.S3).

Interestingly, collisional activation of all six peptides yielded neutral loss of water from the precursor as the base peak in the spectrum, further supporting the possibility that water loss may be a signature for tris-modified peptides. Neutral water losses from b- and y-ions were also present in all spectra (see Figure 3.3). In some cases, neutral losses of water from fragments were exclusively consistent with dehydration of the tris modification (TVLDSGISEVR, VGSDKGAIIG, and IQTGLDATHAER). For other sequences (APSWFDTGLSEMR, HFSPEDLTVK, and GYQYLLEPGDFR), most water losses could be attributed to the tris modification, but some fragments were observed that corresponded to water loss from other residues.

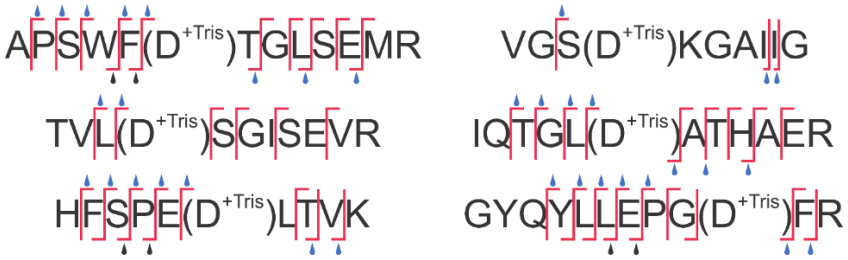


Figure 3.3: b- and y-fragments of tris-modified peptides. Water drops indicate b_n-H_2O or y_n-H_2O fragments, blue droplets contain the tris modification while black droplets do not.

Loss of Tris and Asp Regeneration During CID. In addition to loss of water, collisional activation of APSWFD^{+Tris}TGLSEMR also yields a -103 Da neutral loss with relative intensity of ~9% (Figure 3.4a). This mass loss is identical to the mass increase following addition of tris to the peptide. This same neutral loss is also observed for the other synthetic sequences modified with tris, in abundances ranging from 0.5% - 10% relative to the base peak. Re-isolation and activation of the -103 Da neutral loss yields a spectrum that very closely matches the CID data for the unmodified peptide (compare Figure 3.4b and Figure 3.1c), suggesting that the original peptide (or an isomeric form of it) is regenerated by partial loss of the tris modification. For this to occur, tris would need to leave behind an -OH group on the aspartic acid sidechain while the remainder of the modification was lost. One possible mechanism that could account for these observations is that a small population of tris rearranges in the gas phase to the ester form. This rearranged intermediate can then undergo a one-step fragmentation yielding loss of tris as an oxetane and regenerating original aspartic acid side chain (Figure 3.4c). It is also possible that a subpopulation of tris modified asp already exists as the ester in solution and this species simply degrades during CID as outlined in Figure 3.4c. Although the mechanistic details are not particularly important in terms of isomer identification, the characteristic loss could be utilized as further validation of tris modification in an analysis of complex samples.

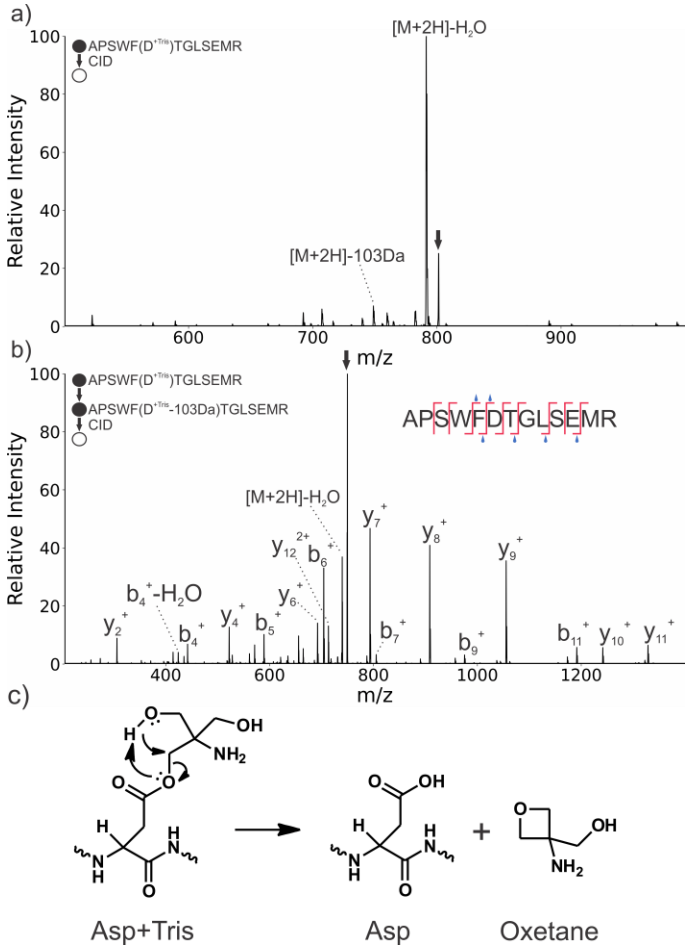


Figure 3.4: a) CID of APSWFD^{+Tris}TGLSEMR yields a peak corresponding to a neutral loss of 103 Daltons. b) Re-isolation and fragmentation of this loss confirms this peak to be the original sequence minus the tris modification. c) Proposed mechanism for the regeneration of the original unmodified peptide during fragmentation.

Although PIMT is capable of accepting a wide range of amino acid combinations N- and C-terminal to L-isoAsp,[28] some sequence motifs near L-isoAsp may slow or prevent PIMT repair and thus reduce the efficiency of tris modification. A sequence in which aspartic acid and serine have isomerized in proximity to each other could present complications, as inversion of the serine stereocenter may impede accommodation within the PIMT active site. Studies of very long-lived proteins such as the crystallins have

shown that serine is also susceptible to epimerization from L-serine to D-serine.[38,39] PIMT repair of the synthetic crystallin-derived peptide TVL**D**SGISEVR L-isoAsp shows near complete repair after 24h, with most of the product being the L-Asp epimer with some D-isomer products at retention time 25 minutes (Figure 3.5a). On the other hand, PIMT repair of the L-isoAsp/D-Ser(5) epimer of TVL**DS**GISEVR after 24h reveals the L-isoAsp to be mostly untouched by PIMT, and only a small amount of the L-Asp D-Ser(5) epimer is present (Figure 3.5b). Both the L-isoAsp and L-isoAsp/D-Ser(5) exhibit tris modification at the aspartic acid, although only 11% of the peptide was modified for the L-isoAsp/D-Ser(5) epimer, whereas 45% of the L-isoAsp epimer was modified. This is likely due to reduced PIMT activity caused by disruptive non-covalent interactions between D-serine and the active site. However, the fact that the PIMT/Tris reaction still modified 11% of the L-isoAsp/D-Ser(5) sequence indicates that even in cases where aspartic acid and serine have isomerized in proximity to each other in a sequence, tris modification of isomerized aspartic residues may not be hindered completely. Sequential aspartic acid repeats are another sequence motif that can hinder PIMT repair of L-isoAsp, and these repeat motifs can also be hotspots of Asp isomerization.[28,40] However, the results in Figure 3.5c show that 24h reaction of the crystallin-derived sequence KMEIV**D**DDVPSL L-isoAsp(6) results in 18% tris modification despite the presence of sequential Asp residues. In such acidic repeat motifs, double Asp isomerization is possible, resulting in two sequential L-isoAsp residues. In this case, it is expected that double L-isoAsp residues would result in low PIMT recognition and tris modification, since introduction of a second L-isoAsp would likely only exacerbate the problems

presented by acidic repeats to PIMT's active site. However, unless double isoAsp formation was dominant, some fraction of the single modification would persist and be available for tris labeling.

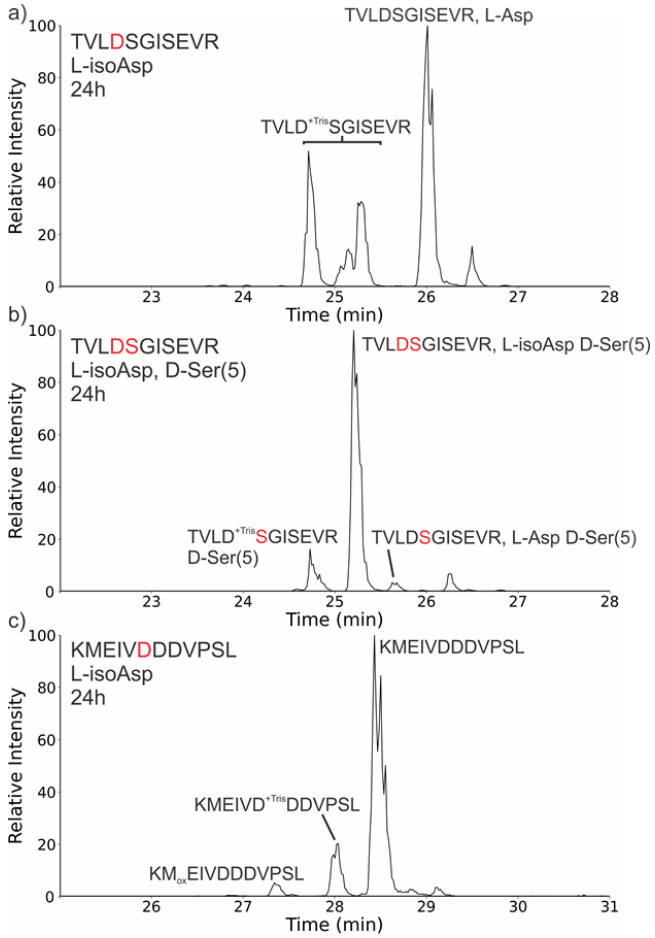


Figure 3.5: a) 24h reaction of TVL**D**SGISEVR L-isoAsp with PIMT and SAM, resulting in repair of L-isoAsp and three peaks corresponding to the tris modification of the aspartic acid. b) 24h PIMT repair of TVL**DS**GISEVR L-isoAsp/D-Ser(5), resulting in modest tris modification of the aspartic acid, and a large amount of L-isoAsp/D-Ser(5) precursor remaining. c) 24h PIMT repair of KMEIV**D**DDVPSL, resulting in tris modification.

Labeling of Aspartic Acid Isomerization Sites in Long-Lived Proteins. The introduction of a mass shift at L-isoAsp should allow identification within proteomics experiments by traditional PTM-based search methods. To explore the feasibility of this

idea, we reconstituted a tryptic digest of a 72-year-old human eye lens in 1M Tris at pH 7.4 and reacted it for 24h with PIMT in the presence of SAM. The extracted ion chromatogram before incubation with PIMT is shown in Figure 3.6a for a representative peptide, HFSPEDLTVK. A distinct L-isoAsp peak is notable. In Figure 3.6b, the extracted ion chromatogram for the same peptide is shown after incubation with PIMT and SAM in tris. The L-isoAsp peptide is essentially absent. In addition, new peaks corresponding to the peptide mass +103 Da are found as shown in Figure 3.6c. CID confirmed that the +103 Da peak corresponded to tris modification of HFSPEDLTVK. In addition, 21 other tris modified peptides were identified in the sample out of 36 total asp-containing sequences. The identified tris modified peptides included six from Crystallin αA, five peptides from Crystallin αB, and peptides from Crystallin βB1, γD, βA1, γS, βA4, and βB2 (Fig 3.6d). Modified sequences were identified via CID and/or by comparison of chromatographic data obtained before and after reaction with PIMT, by specifically looking for diminution of the L-isoAsp peak and appearance of a new mass shifted peak (Figures 3.S4-3.S6). The αA and αB crystallins were the most abundant species present in the sample, a fact unsurprising considering that crystallins make up almost 90% of the protein content in the lens.[41] αA and αB exhibited the highest number of modifications among the crystallins, with most aspartic acids from these proteins being modified. However, some aspartic acid residues were not modified during the reaction, likely due to a lack of sufficient isomerization/L-isoAsp at these locations. Two non-crystallin proteins with sequences containing aspartic acid were also identified, phakinin

and retinal dehydrogenase (BFSP2 and AL1A1 in Figure 3.6d). Tris-modified aspartic acids were not detected in sequences from these proteins.

Figure 3.6d shows the percent isomerization (blue) versus the percent tris modification found for each sequence (orange). The results suggest that >10% isomerization is required for tris modification to be detectable. Importantly, peptides from Crystallin βA1 and γS were modified by tris although no isomerization was detected prior. In these instances, it is likely that the L-isoAsp peptide did not chromatographically separate from the L-Asp version for the unmodified peptides. For cases where more than 10% isomerization was detected but no accompanying tris modification was found, some of these sequences were present in very low abundance, making detection of a tris modification difficult. Sequence effects may explain the remaining absences, such as for MEIIDDDVPSFHAHGYQEK from crystallin βB2, which is ~36% isomerized but was not tris modified. This peptide contains an acidic repeat, which is the most likely site of isomerization in this sequence but is also likely to hinder PIMT repair as discussed earlier. The varying preferences with which PIMT binds to different sequence motifs may also explain in part why the relationship between percent isomerization and percent tris modification differs significantly for different peptides. In any case, peptides with a higher percent isomerization are for the most part more likely to be modified, indicating that this method is most effective at labeling major sites of isomerization.

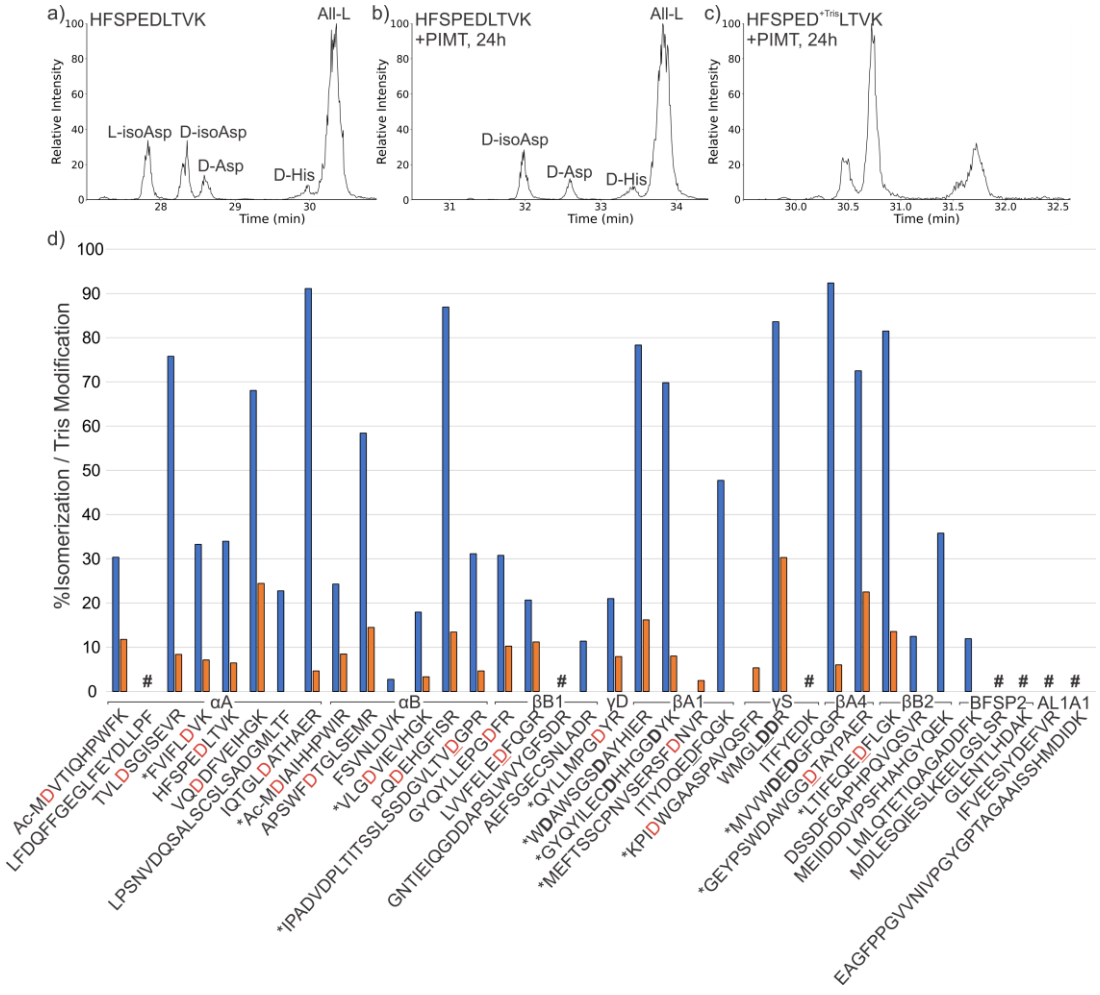


Figure 3.6: a) HFSPEDLTVK extracted ion chromatogram prior to reaction with PIMT and b) after 24h reaction with PIMT. c) Extracted ion chromatogram of HFSPED^{+Tris}LTVK showing three peaks corresponding to tris modification after reaction with PIMT. d) Asp-containing sequences identified in a tryptic digest of 72 year old crystallin. Bar plots in blue show percent isomerization prior to addition of PIMT and percent tris modification for each sequence after PIMT addition are shown by the orange bar plots. Percent isomerization was calculated as [sum of isomer peak areas / (sum of isomer peak areas + all-L peak area)] x 100. Red Asp residues indicate tris-modified sites, bold Asps indicate sequences where determining the exact site of modification was not possible, and underlined Asp residues indicate sites that underwent deamidation. Pound signs designate sequences for which neither isomerization nor tris modification was observed. Asterisks indicate tris-modified sequences identified by the +103 Da mass shift appearing after addition of PIMT, modified sequences without an asterisk were identified by both mass shift and accompanying CID spectra.

Conclusion

Spontaneous isomerization of Asp residues is the most abundant non-mass-shifting modification that occurs in LLPs and is potentially important for understanding the underlying causes of many age-related diseases. Herein, we have demonstrated that L-isoAsp can be effectively labeled by simply incubating samples in tris buffer with PIMT. The resulting tris addition yields mass-shifted peaks that can be treated as typical PTMs and identified using standard data analysis already implemented in proteomics experiments. Out of ten proteins identified with aspartic acid-containing sequences from the eye lens tryptic digest, eight were labeled at least once with tris. If a single L-isoAsp modification site can be identified, it reveals not only the location of one potential function-altering structural change, but also confirms the status of the entire protein as long-lived, which could then become the subject for a more targeted investigation. Although methods for directly identifying isomers with mass spectrometry continue to develop, the challenge becomes more difficult as sample complexity increases, making alternative approaches that simplify the problem by selectively introducing a mass-shifting reagent worthy of further pursuit.

Acknowledgements. The authors are grateful for funding from the National Institute on Aging, R01AG066626. The authors would also like to thank Dr. Yana A. Lyon for processing and tryptic digestion of the 72-year-old WI nucleus eye lens fraction used in this work.

Supporting Information.

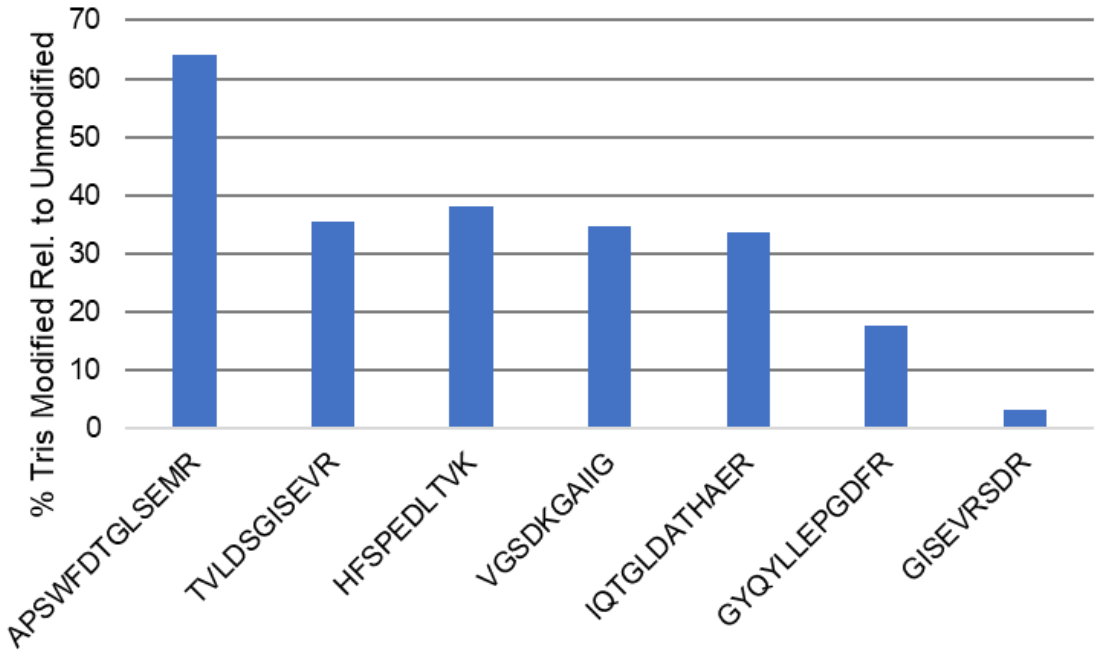


Figure 3.S1: Relative abundance of tris-modified L-isoAsp-containing synthetic sequences after 24h reaction with 1:50 (M/M) PIMT and 400uM SAM in 1.0M Tris pH 7.4. The relative abundance was calculated as a percent of the total peak area for the unmodified and modified sequence.

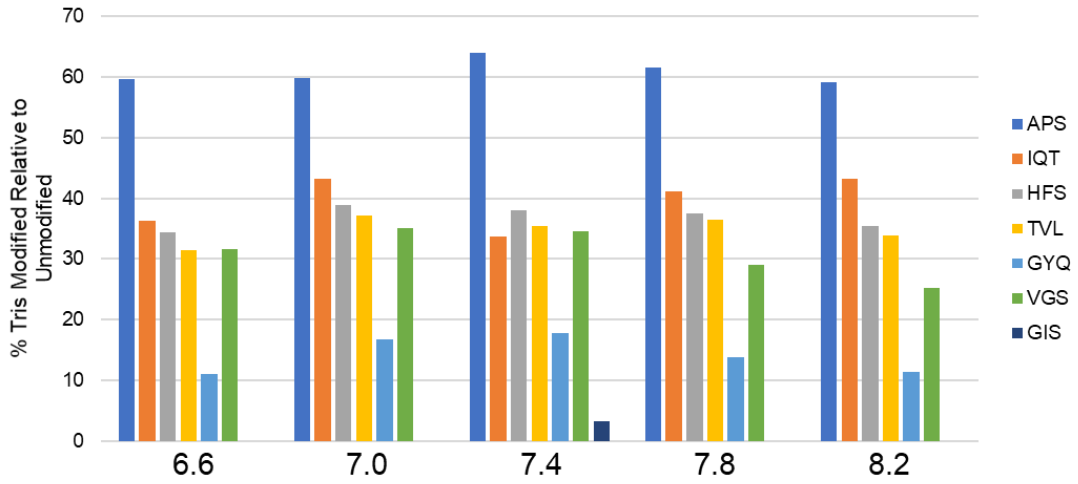


Figure 3.S2: Percent tris modification for each synthetic sequence (APSWFDTGLSEMR, IQTGLDATHAER, HFSPEDLTVK, TVLDSGISEVR, GYQYLLEPGDFR, and VGSDKGAIIG) at different pH (6.6, 7.0, 7.4, 7.8, 8.2). Tris modification of GISEVRSDR was only tested at pH 7.4.

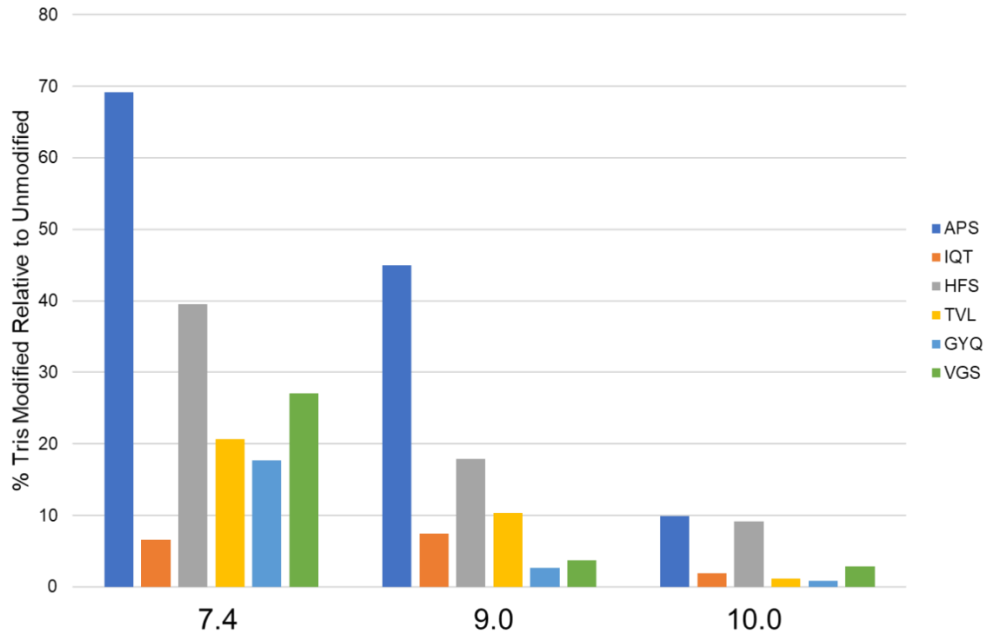
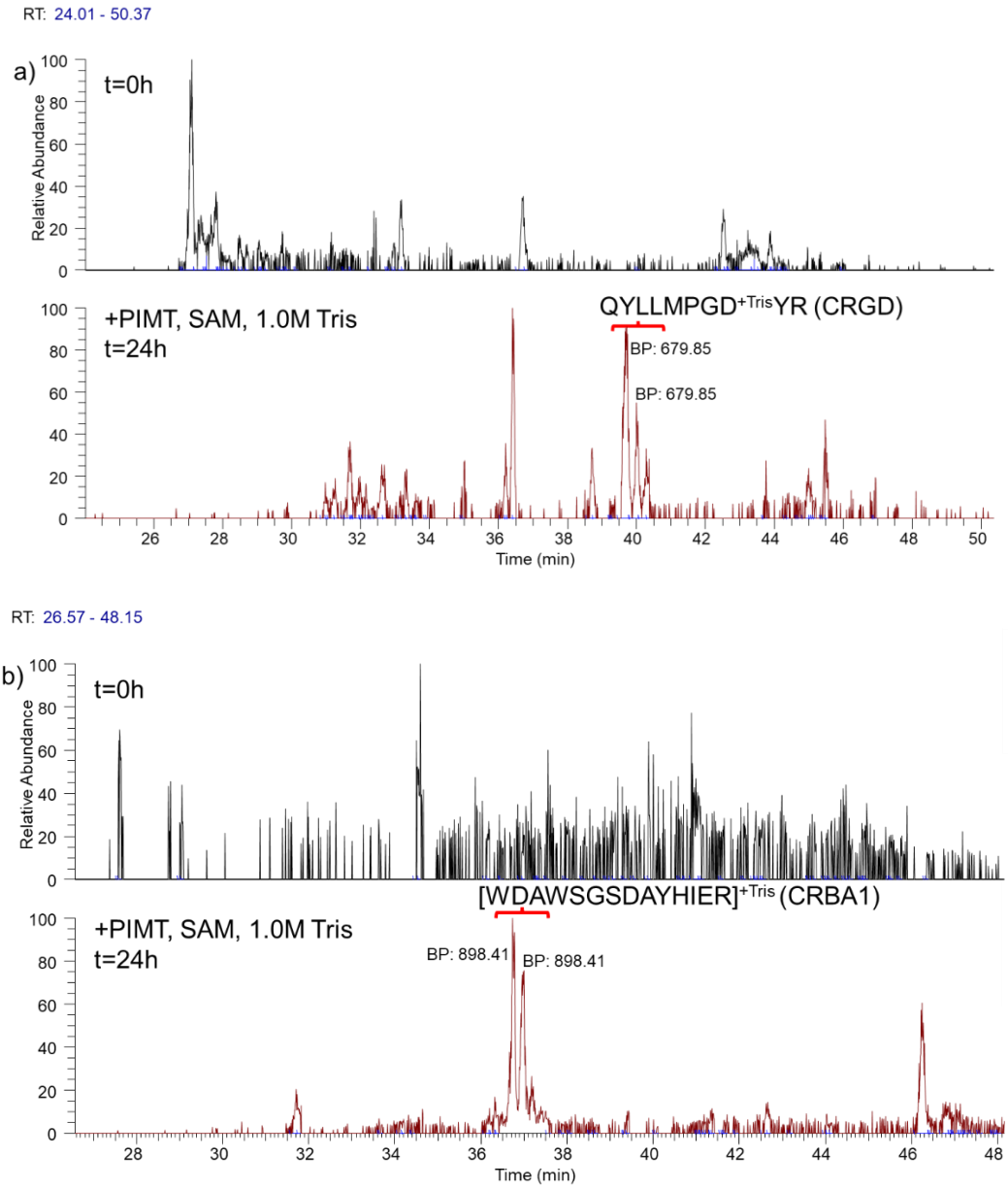
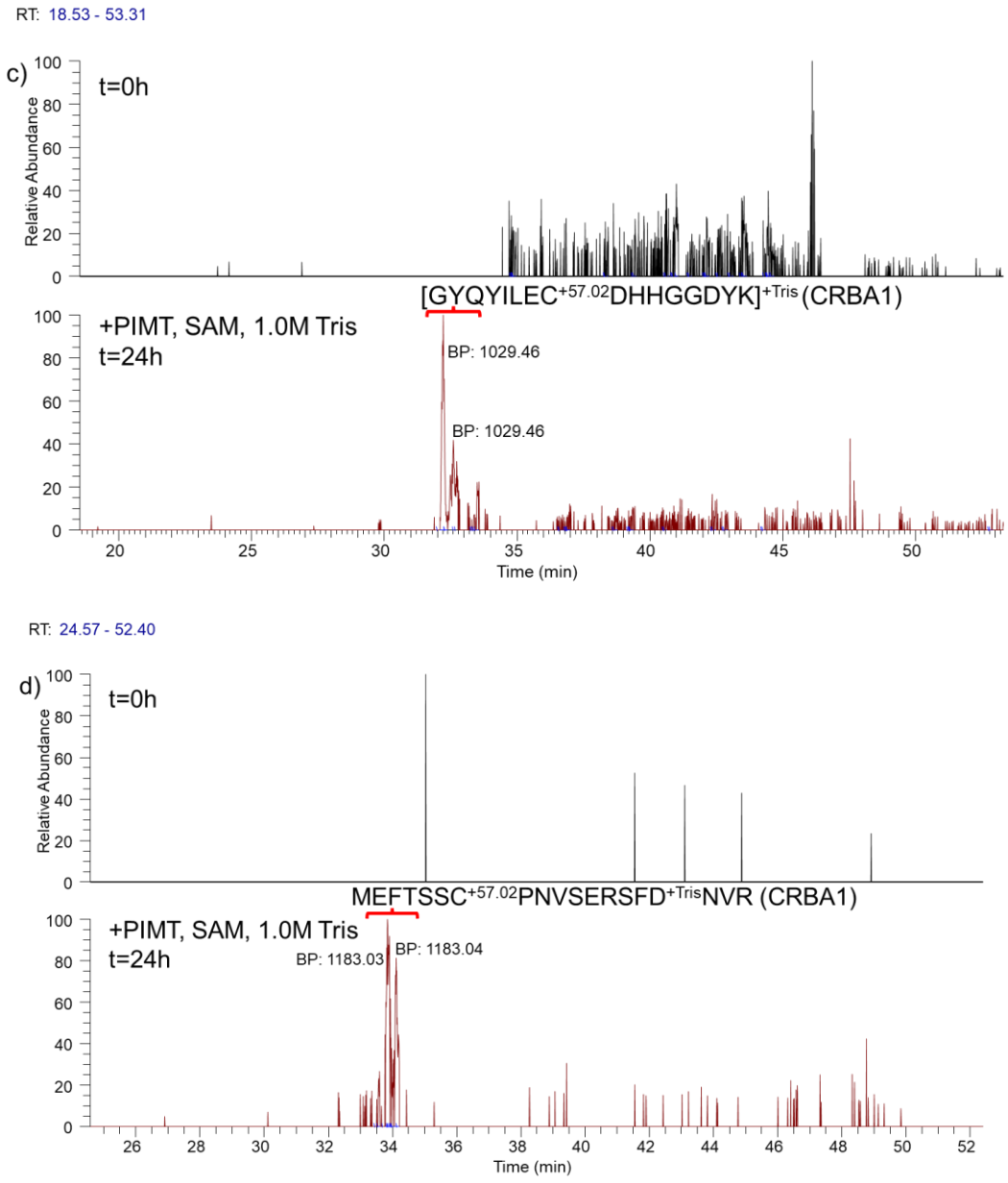
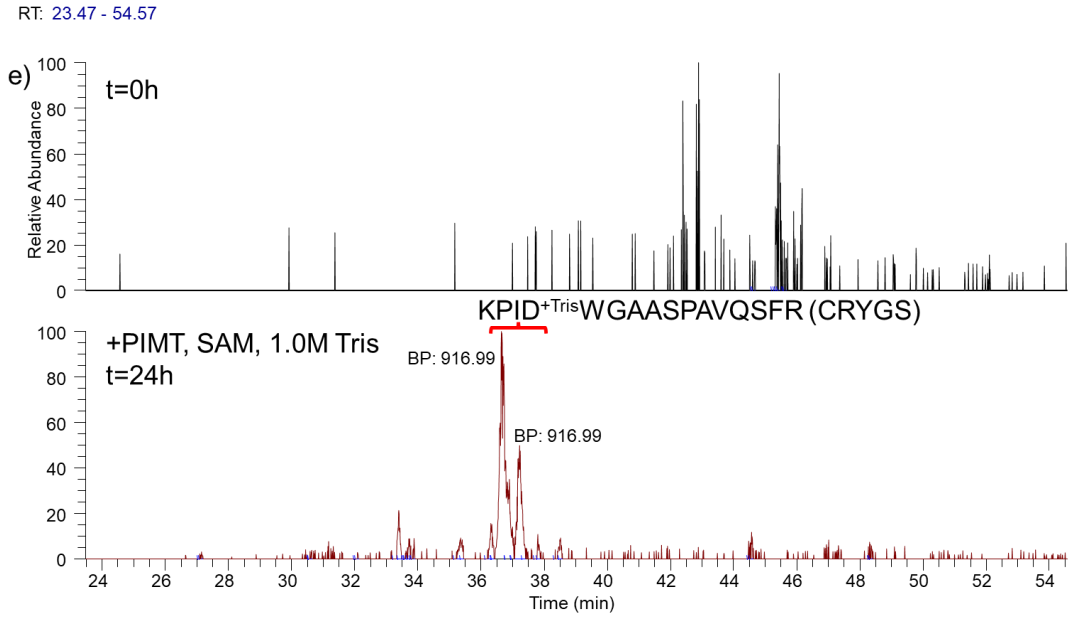
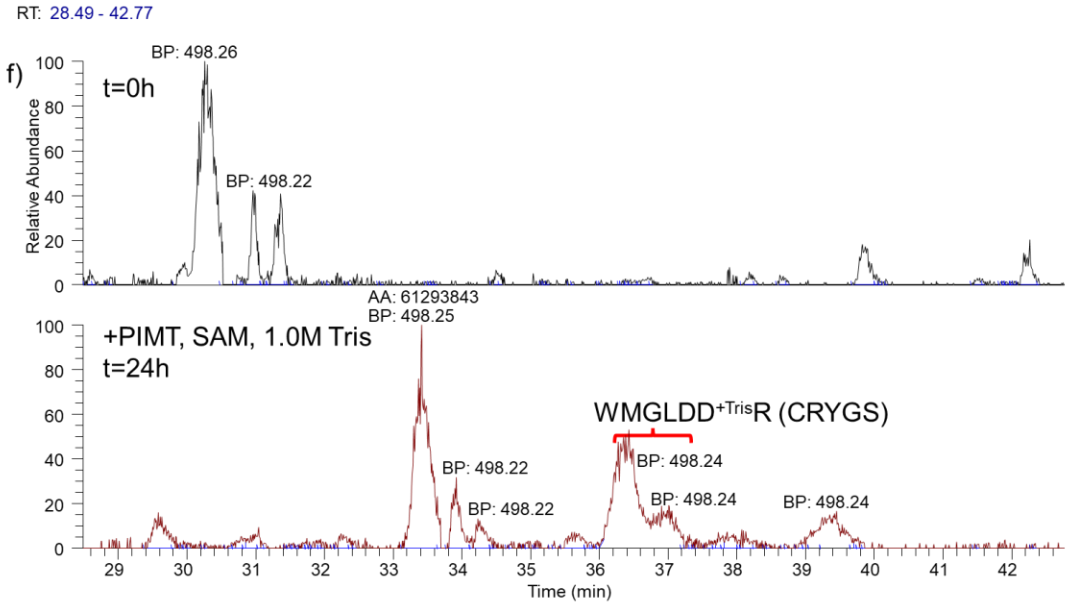


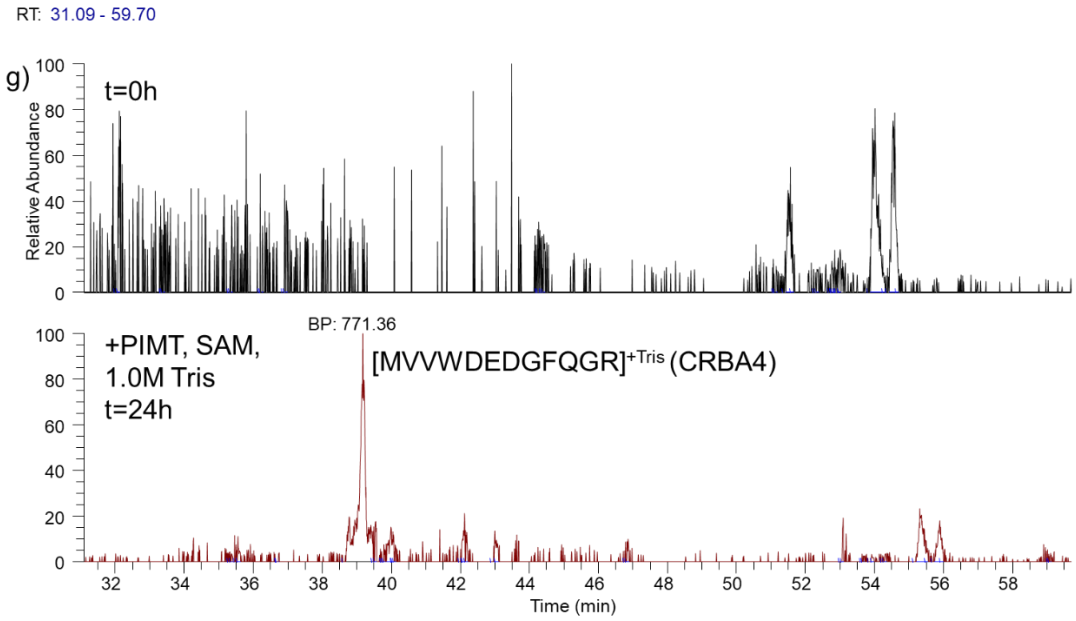
Figure 3.S3: Percent tris modification for synthetic sequences at pH 7.4, 9.0, and 10.0. Synthetic peptides were reacted in 1.0M Tris at the above pH's for 24h with 1:50 (M/M) PIMT, 400uM SAM.

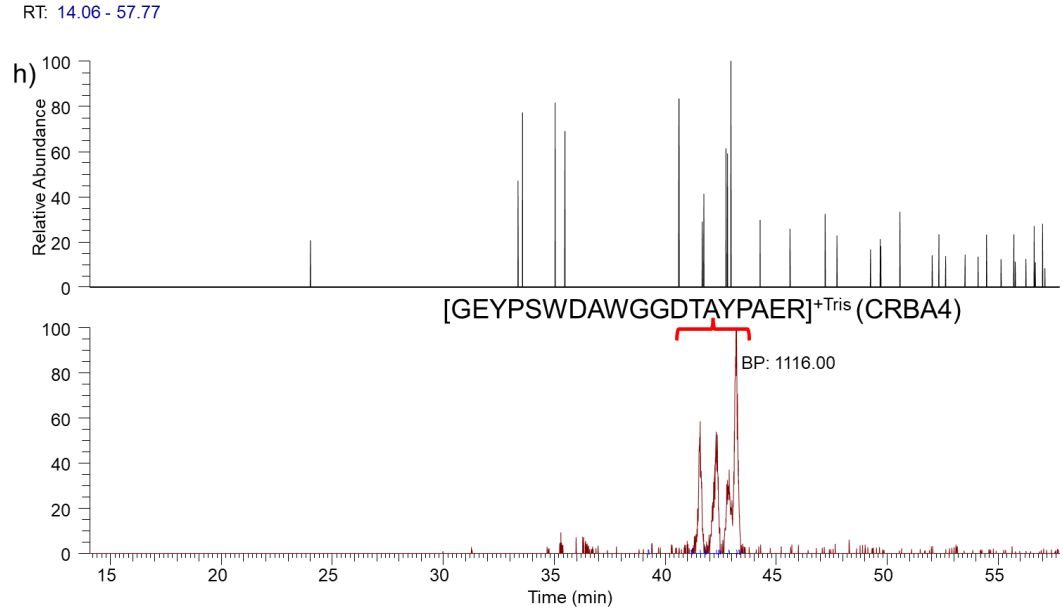

RT: 24.01 - 50.37
a)
t=0h
Relative Abundance
+PIMT, SAM, 1.0M Tris
t=24h
QYLLMPGD+TrisYR (CRGD)
BP: 679.85
BP: 679.85
Time (min)
RT: 26.57 - 48.15
b)
t=0h
Relative Abundance
[WDAWSGSDAYHIER]+Tris (CRBA1)
+PIMT, SAM, 1.0M Tris
t=24h
BP: 898.41
BP: 898.41
Time (min)

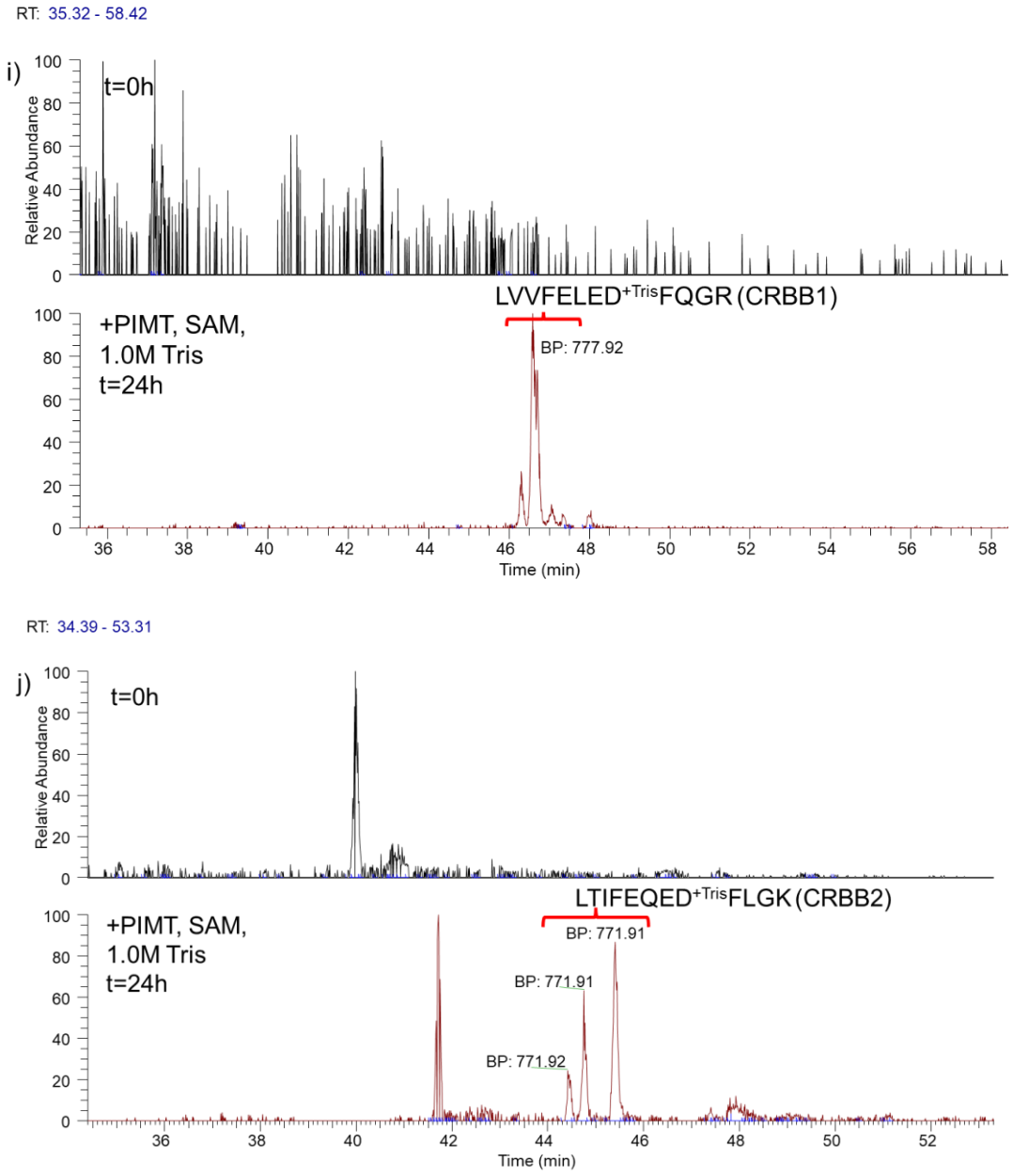

RT: 18.53 - 53.31
c)
t=0h
Relative Abundance
[GYQYILEC+57.02DHHGGDYK]+Tris (CRBA1)
+PIMT, SAM, 1.0M Tris
t=24h
BP: 1029.46
BP: 1029.46
Time (min)
RT: 24.57 - 52.40
d)
t=0h
Relative Abundance
MEFTSSC+57.02PNVSERSFD+TrisNVR (CRBA1)
+PIMT, SAM, 1.0M Tris
t=24h
BP: 1183.03
BP: 1183.04
Time (min)


RT: 23.47 - 54.57
e)
t=0h
Relative Abundance
KPID+TrisWGAASPAVQSFR (CRYGS)
+PIMT, SAM, 1.0M Tris
t=24h
BP: 916.99
BP: 916.99
Time (min)

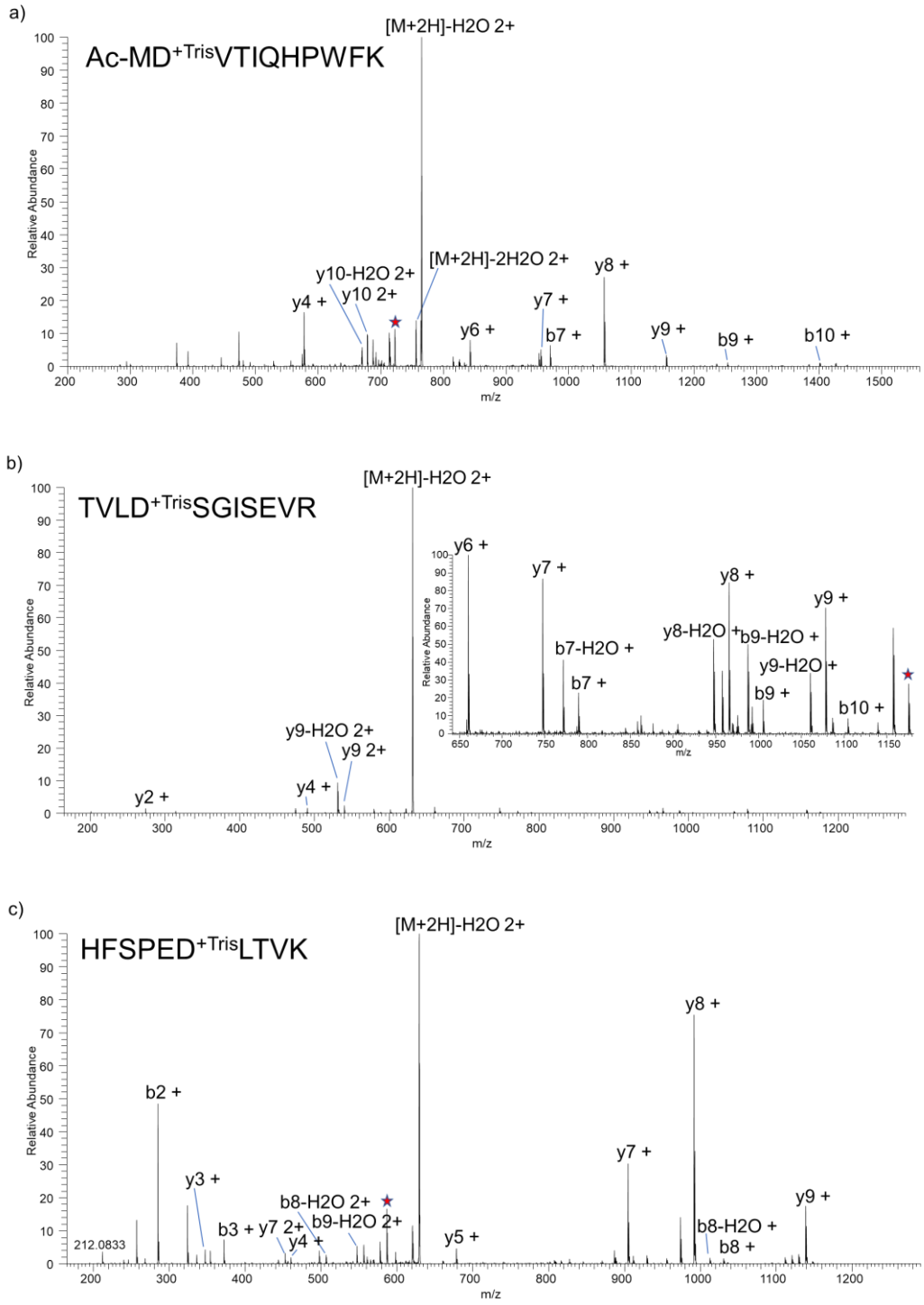

RT: 28.49 - 42.77
f)
t=0h
BP: 498.26
BP: 498.22
Relative Abundance
AA: 61293843
BP: 498.25
+PIMT, SAM, 1.0M Tris
t=24h
WMGLDD+TrisR (CRYGS)
BP: 498.22
BP: 498.22
BP: 498.24
BP: 498.24
BP: 498.24
Time (min)

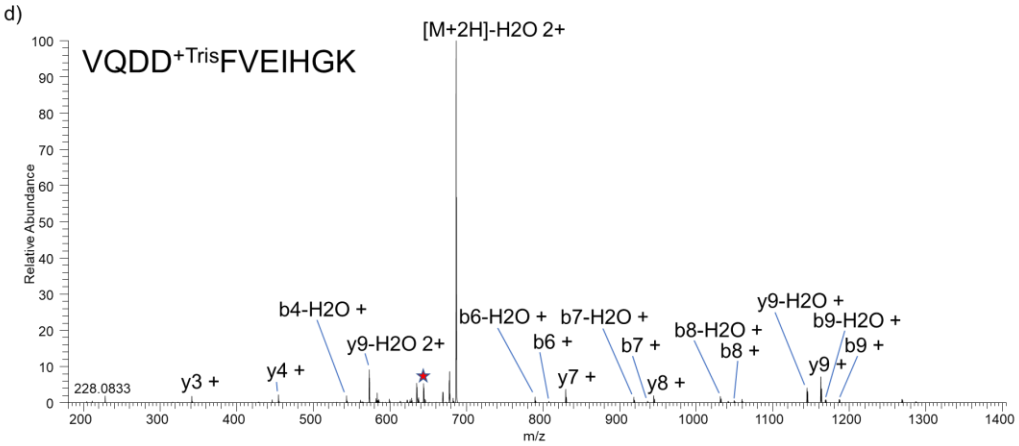

RT: 31.09 - 59.70
g)
t=0h
Relative Abundance
BP: 771.36
+PIMT, SAM,
1.0M Tris
t=24h
[MVVWDEDGFQGR]+Tris (CRBA4)
Time (min)

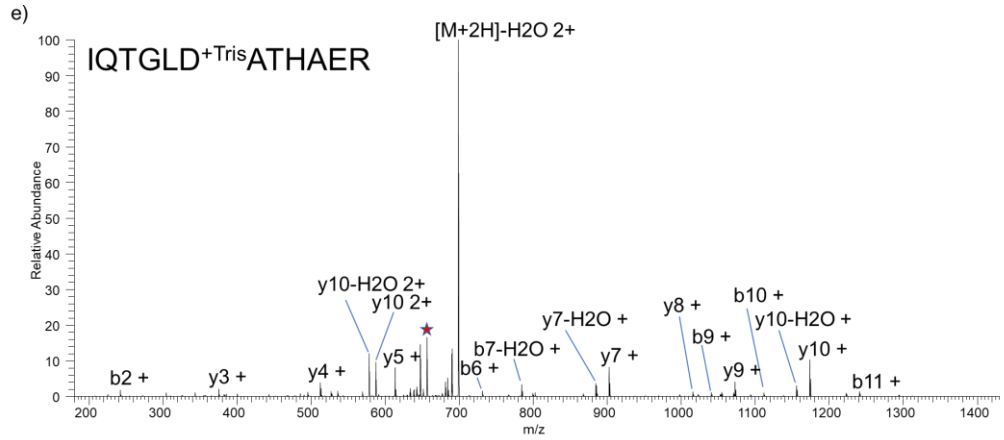

RT: 14.06 - 57.77
h)
Relative Abundance
[GEYPSWDAWGGDTAYPAER]+Tris (CRBA4)
BP: 1116.00
Time (min)

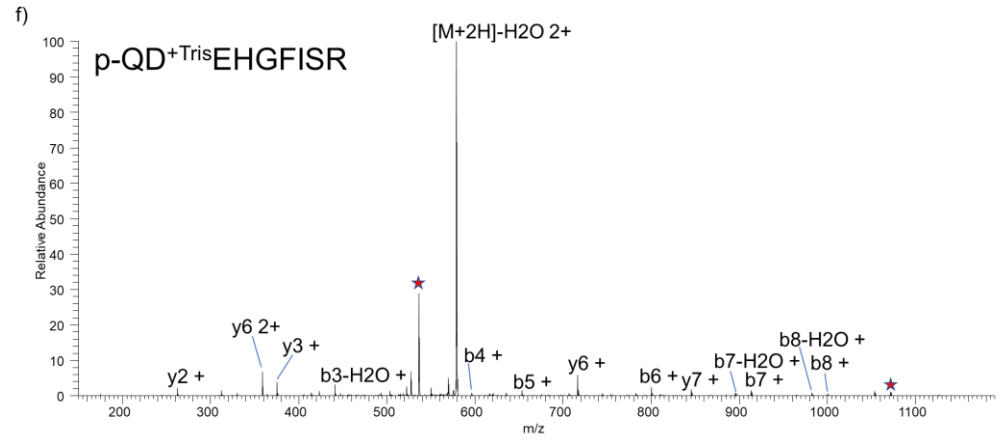


Figures 3.S4 a-j: Extracted ion chromatograms for tris modified peptides from crystallin γD (a), βA1 (b-d), γS (e-f), βA4 (g-h), βB1 (i), and βB2 (j) before PIMT reaction (t=0h) and after PIMT reaction (t=24h).


a)
Ac-MD+TrisVTIQHPWFK
[M+2H]-H2O 2+
[M+2H]-2H2O 2+
y10-H2O 2+
y10 2+
y4 +
y6 +
y7 +
b7 +
y8 +
y9 +
b9 +
b10 +
Relative Abundance
m/z
b)
TVLD+TrisSGISEVR
[M+2H]-H2O 2+
y6 +
y7 +
y8 +
y9 +
y8-H2O +
b9-H2O +
b7-H2O +
y9-H2O +
b7 +
b9 +
b10 +
y9-H2O 2+
y9 2+
y4 +
y2 +
Relative Abundance
m/z
c)
HFSPED+TrisLTVK
[M+2H]-H2O 2+
y8 +
b2 +
y7 +
y3 +
b8-H2O 2+
b3 + y7 2+
b9-H2O 2+
y4 +
y5 +
y9 +
b8-H2O +
b8 +
212.0833
Relative Abundance
m/z


d)
VQDD+TrisFVEIHGK
[M+2H]-H2O 2+
Relative Abundance
m/z
228.0833
y3 +
y4 +
b4-H2O +
y9-H2O 2+
b6-H2O +
b6 +
b7-H2O +
b7 +
y7 +
y8 +
b8-H2O +
b8 +
y9-H2O +
y9 +
b9-H2O +
b9 +


e)
IQTGLD+TrisATHAER
[M+2H]-H2O 2+
Relative Abundance
m/z
b2 +
y3 +
y4 +
y10-H2O 2+
y10 2+
y5 +
b6 +
b7-H2O +
y7-H2O +
y7 +
y8 +
b9 +
b10 +
y10-H2O +
y9 +
y10 +
b11 +


f)
p-QD+TrisEHGFISR
[M+2H]-H2O 2+
Relative Abundance
m/z
y2 +
y6 2+
y3 +
b3-H2O +
b4 +
b5 +
y6 +
b6 +
y7 +
b7-H2O +
b7 +
b8-H2O +
b8 +

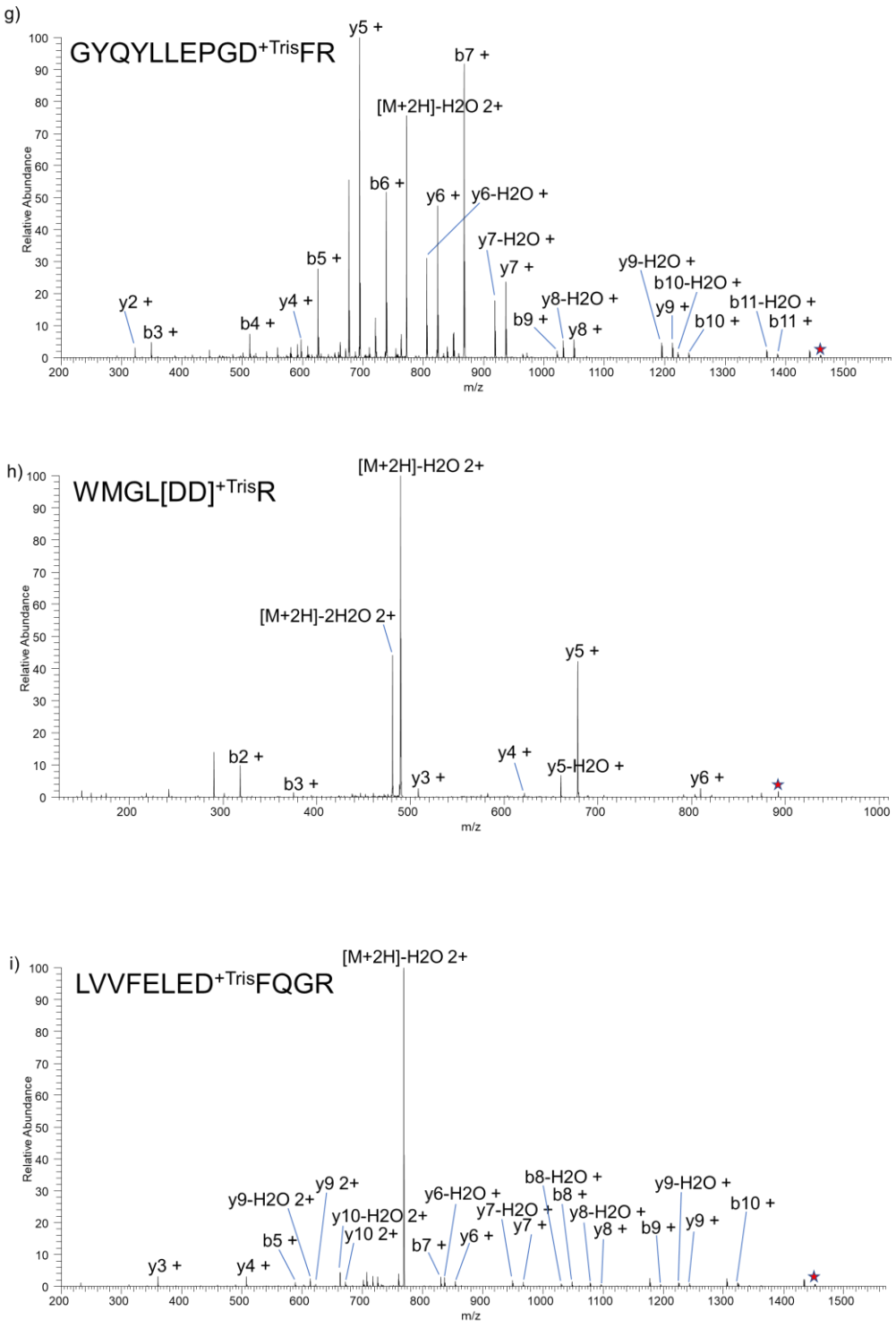
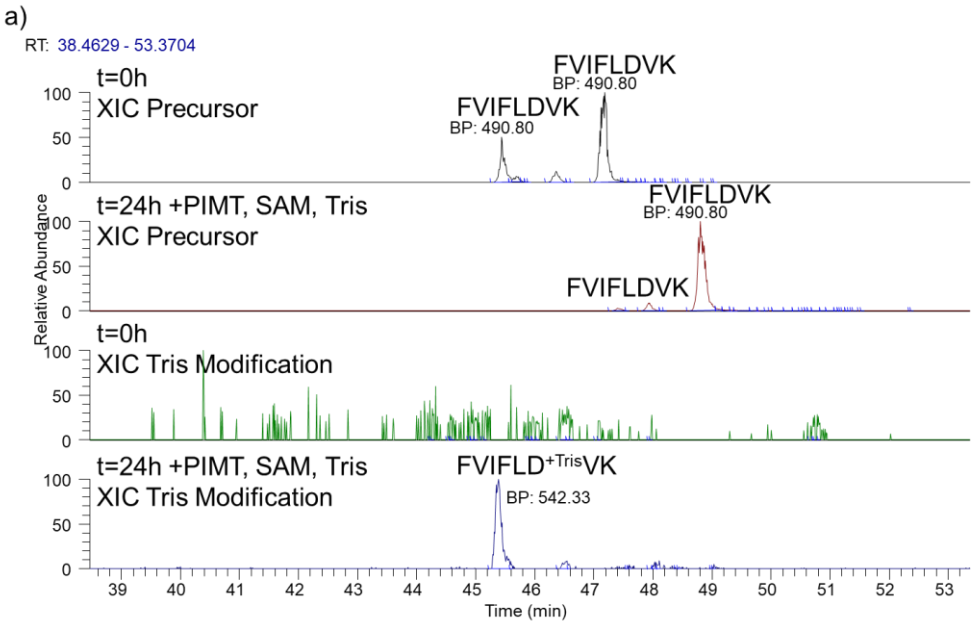
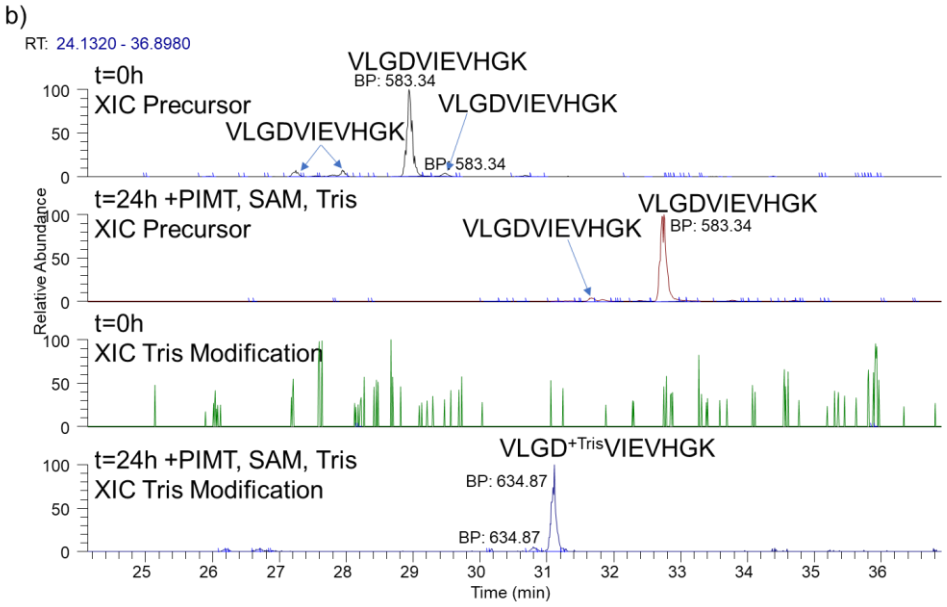


Figure 3.S5 a-i: CID spectra for tris-modified peptides from trypsin digest of 72-year-old crystallin samples reacted with PIMT and tris. Red star denotes the peak corresponding to concerted loss of tris and regeneration of asp side chain.


a)
RT: 38.4629 - 53.3704
t=0h
XIC Precursor
FVIFLDVK
BP: 490.80
FVIFLDVK
BP: 490.80
t=24h +PIMT, SAM, Tris
XIC Precursor
FVIFLDVK
BP: 490.80
FVIFLDVK
Relative Abundance
t=0h
XIC Tris Modification
t=24h +PIMT, SAM, Tris
XIC Tris Modification
FVIFLD+TrisVK
BP: 542.33
100
50
0
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
Time (min)


b)
RT: 24.1320 - 36.8980
t=0h
XIC Precursor
VLGDVIEVHGK
BP: 583.34
VLGDVIEVHGK
VLGDVIEVHGK
BP: 583.34
t=24h +PIMT, SAM, Tris
XIC Precursor
VLGDVIEVHGK
VLGDVIEVHGK
BP: 583.34
Relative Abundance
t=0h
XIC Tris Modification
t=24h +PIMT, SAM, Tris
XIC Tris Modification
VLGD+TrisVIEVHGK
BP: 634.87
BP: 634.87
100
50
0
25
26
27
28
29
30
31
32
33
34
35
36
Time (min)

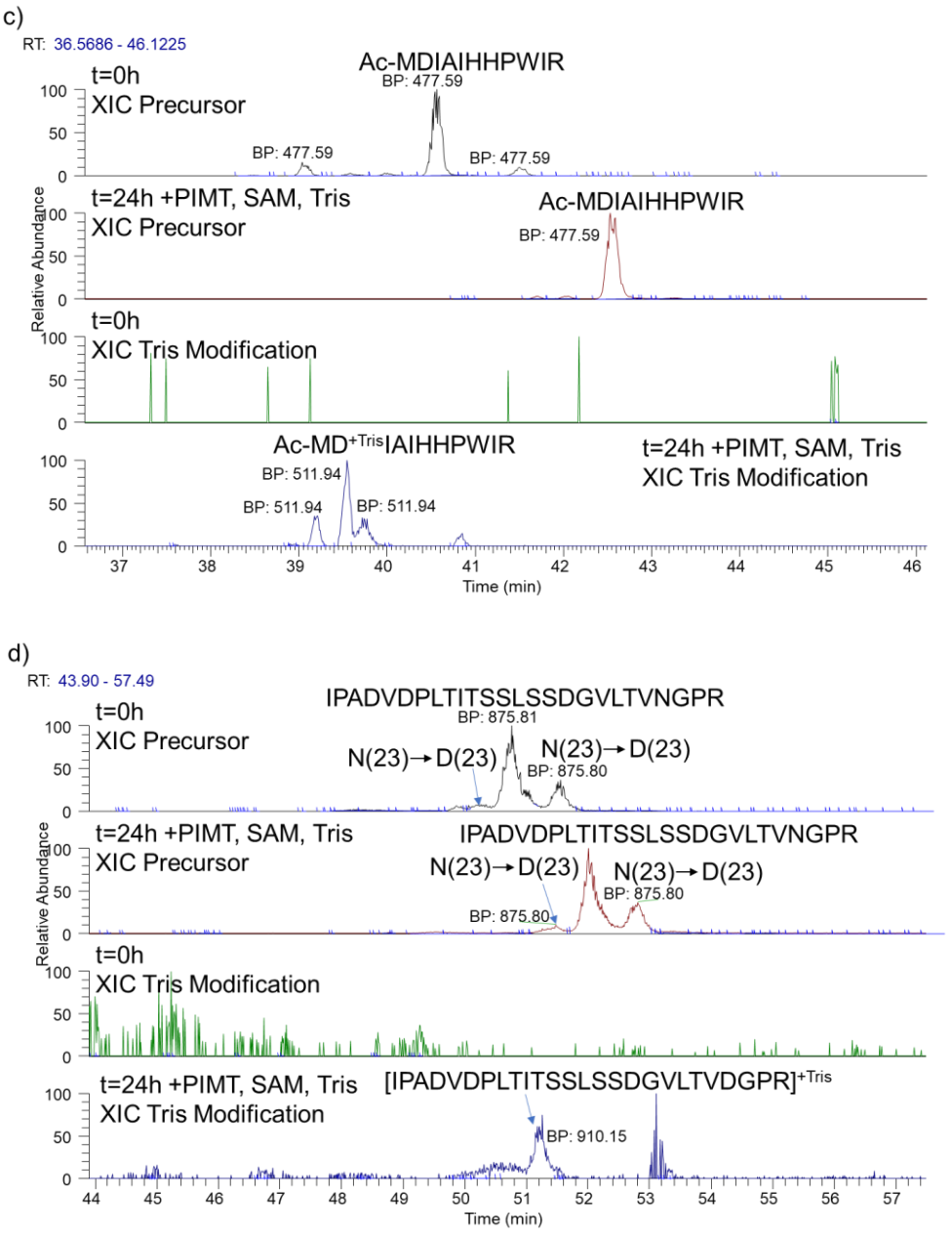


Figure 3.S6 a-d: Select extracted ion chromatograms (XIC) of the precursor and tris modification mass for tris-modified peptides. The initial time point (before addition of PIMT) is labeled as t=0h, t=24h is the endpoint taken for the reaction with PIMT and tris.

References

[1] Truscott, R. J. W.; Schey, K. L.; Friedrich, M. G. Old Proteins in Man: A Field in Its Infancy. *Trends Biochem. Sci.* **2016**, *41* (8), 654–664.
[2] Toyama, B. H.; Hetzer, M. W. Protein Homeostasis : Live Long, Won't Prosper. **2013**, *14* (January), 55–61.
[3] Fujii, N.; Ishibashi, Y.; Satoh, K.; Fujino, M.; Harada, K. Simultaneous Racemization and Isomerization at Specific Aspartic Acid Residues in AB-Crystallin from the Aged Human Lens. *Biochim. Biophys. Acta (BBA)/Protein Struct. Mol.* **1994**, *1204* (2), 157–163.
[4] Samuel Zigler, J.; Goosey, J. Aging of Protein Molecules: Lens Crystallins as a Model System. *Trends Biochem. Sci.* **1981**, *6* (1), 133–136.
[5] Lampi, K. J.; Wilmarth, P. A.; Murray, M. R.; David, L. L. Lens β-Crystallins: The Role of Deamidation and Related Modifications in Aging and Cataract. *Prog. Biophys. Mol. Biol.* **2014**, *115* (1), 21–31.
[6] Fujii, N.; Takata, T.; Fujii, N.; Aki, K. Isomerization of Aspartyl Residues in Crystallins and Its Influence upon Cataract. *Biochim. Biophys. Acta - Gen. Subj.* **2016**, *1860* (1), 183–191.
[7] Lyon, Y. A.; Collier, M. P.; Riggs, D. L.; Degiacomi, M. T.; Benesch, J. L. P.; Julian, R. R. Structural and Functional Consequences of Age-Related Isomerization in α-Crystallins. **2019**, *294* (14), 7546–7555.
[8] Fujii, N.; Sakaue, H.; Sasaki, H.; Fujii, N. A Rapid, Comprehensive Liquid Chromatography-Mass Spectrometry (LC-MS)-Based Survey of the Asp Isomers in Crystallins from Human Cataract Lenses*. *J. Biol. Chem.* **2012**, *287* (47), 39992–40002.
[9] Readel, E. R.; Wey, M.; Armstrong, D. W. Rapid and Selective Separation of Amyloid Beta from Its Stereoisomeric Point Mutations Implicated in Neurodegenerative Alzheimer's Disease. *Anal. Chim. Acta* **2021**, *1163*, 338506.
[10] Nagy, G.; Kedia, K.; Attah, I. K.; Garimella, S. V. B.; Ibrahim, Y. M.; Petyuk, V. A.; Smith, R. D. Separation of β-Amyloid Tryptic Peptide Species with Isomerized and Racemized L-Aspartic Residues with Ion Mobility in Structures for Lossless Ion Manipulations. *Anal. Chem.* **2019**, *91* (7), 4374–4380.
[11] Zheng, X.; Deng, L.; Baker, E. S.; Ibrahim, Y. M.; Petyuk, V. A.; Smith, R. D. Distinguishing D-and L-Aspartic and Isoaspartic Acids in Amyloid β Peptides with Ultrahigh Resolution Ion Mobility Spectrometry. *Chem. Commun.* **2017**, *53* (56), 7913–7916.
[12] 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 Commun.* **2021**, *3* (2), 1–17.
[13] Jia, C.; Lietz, C. B.; Yu, Q.; Li, L. Site-Specific Characterization of D -Amino Acid Containing Peptide Epimers by Ion Mobility Spectrometry. *Anal. Chem.* **2014**, *86* (6), 2972–2981.
[14] Wu, H. T.; Julian, R. R. Two-Dimensional Identification and Localization of Isomers in Crystallin Peptides Using TWIM-MS. *Analyst* **2020**, *145* (15), 5232–5241.
[15] Tao, Y.; Quebbemann, N. R.; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820.
[16] Sargaeva, N. P.; Lin, C.; O'Connor, P. B. Identification of Aspartic and Isoaspartic Acid Residues in Amyloid β Peptides, Including Aβ1−42, Using Electron−Ion Reactions. *Anal. Chem.* **2009**, *81* (23), 9778–9786.
[17] Ni, W.; Dai, S.; Karger, B. L.; Zhou, Z. S. Analysis of Isoaspartic Acid by Selective Proteolysis with Asp-N and Electron Transfer Dissociation Mass Spectrometry. *Anal. Chem.* **2010**, *82* (17), 7485–7491.
[18] 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.

[19] Livnat, I.; Tai, H. C.; Jansson, E. T.; Bai, L.; Romanova, E. V.; Chen, T. T.; Yu, K.; Chen, S. A.; Zhang, Y.; Wang, Z. Y.; Liu, D. D.; Weiss, K. R.; Jing, J.; Sweedler, J. V. A D-Amino Acid-Containing Neuropeptide Discovery Funnel. *Anal. Chem.* **2016**, *88* (23), 11868–11876.
[20] Fujii, N.; Maeda, H.; Takata, T.; Fujii, N.; Sakaue, H.; Nirasawa, S.; Takahashi, S.; Sasaki, H. Rapid Survey of Four Asp Isomers in Disease-Related Proteins by LC-MS Combined with Commercial Enzymes. *Anal. Chem.* **2015**, *87* (1), 561–568.
[21] Takahashi, S.; Ogasawara, H.; Hiwatashi, K.; Hori, K.; Hata, K.; Tachibana, T.; Itoh, Y.; Sugiyama, T. Paenidase, a Novel d-Aspartyl Endopeptidase from Paenibacillus Sp. B38: Purification and Substrate Specificity. *J. Biochem.* **2006**, *139* (2), 197–202.
[22] Kagan, R. M.; McFadden, H. J.; McFadden, P. N.; O'Connor, C.; Clarke, S. Molecular Phylogenetics of a Protein Repair Methyltransferase. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **1997**, *117* (3), 379–385.
[23] Yamamoto, A.; Takagi, H.; Kitamura, D.; Tatsuoka, H.; Nakano, H.; Kawano, H.; Kuroyanagi, H.; Yahagi, Y. I.; Kobayashi, S. I.; Koizumi, K. I.; Sakai, T.; Saito, K. I.; 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.
[24] 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. *Neurobiol. Aging* **2015**, *36* (2), 1029–1036.
[25] Soliman, R.; Cordero-Maldonado, M. L.; Martins, T. G.; Moein, M.; Conrotte, J.-F.; Warmack, R. A.; Skupin, A.; Crawford, A. D.; Clarke, S. G.; Linster, C. L. L-Isoaspartyl Methyltransferase Deficiency in Zebrafish Leads to Impaired Calcium Signaling in the Brain. *Front. Genet.* **2021**, *11*, 612343.
[26] Riggs, D. L.; Gomez, S. V; Julian, R. R. Sequence and Solution Effects on the Prevalence of D-Isomers Produced by Deamidation. *ACS Chem. Biol.* **2017**, *12* (11), 2875–2882.
[27] 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.
[28] Lowenson, J. D.; Clarke, S. Structural Elements Affecting the Recognition of L-Isoaspartyl Residues by the L-Isoaspartyl/D-Aspartyl Protein Methyltransferase. Implications for the Repair Hypothesis. *J. Biol. Chem.* **1991**, *266* (29), 19396–19406.
[29] Lowenson, J. D.; Clarke, S. Recognition of D-Aspartyl Residues in Polypeptides by the Erythrocyte L-Isoaspartyl/D-Aspartyl Protein Methyltransferase. Implications for the Repair Hypothesis. *J. Biol. Chem.* **1992**, *267* (9), 5985–5995.
[30] Murray, E. D.; Clarke, S. Metabolism of a Synthetic L-Isoaspartyl-Containing Hexapeptide in Erythrocyte Extracts. Enzymatic Methyl Esterification Is Followed by Nonenzymatic Succinimide Formation. *J. Biol. Chem.* **1986**, *261* (1), 306–312.
[31] Liu, M.; Cheetham, J.; Cauchon, N.; Ostovic, J.; Ni, W.; Ren, D.; Zhou, Z. S. Protein Isoaspartate Methyltransferase-Mediated 18 O-Labeling of Isoaspartic Acid for Mass Spectrometry Analysis. *Anal. Chem.* **2012**, *84* (2), 1056–1062.
[32] Alfaro, J. F.; Gillies, L. A.; Sun, H. G.; Dai, S.; Zang, T.; Klaene, J. J.; Byung, J. K.; Lowenson, J. D.; Clarke, S. G.; Karger, B. L.; Zhou, Z. S. Chemo-Enzymatic Detection of Protein Isoaspartate Using Protein Isoaspartate Methyltransferase and Hydrazine Trapping. *Anal. Chem.* **2008**, *80* (10), 3882–3889.
[33] Klaene, J. J.; Ni, W.; Alfaro, J. F.; Zhou, Z. S. Detection and Quantitation of Succinimide in Intact Protein via Hydrazine Trapping and Chemical Derivatization. *J. Pharm. Sci.* **2014**, *103* (10), 3033–3042.
[34] Kabadi, P. G.; Sankaran, P. K.; Palanivelu, D. V.; Adhikary, L.; Khedkar, A.; Chatterjee, A. Mass Spectrometry Based Mechanistic Insights into Formation of Tris Conjugates: Implications on Protein Biopharmaceutics. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (10), 1677–1685.
[35] Hood, C. A., Fuentes, G., Patel, H., Page, K., Menakuru, M., Park, J. H. Fast Conventional Fmoc Solid-Phase Peptide Synthesis With. *J. Pept. Sci.* **2008**, 14, 97–101.
[36] 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. *Exp. Eye Res.* **2018**, *171* (3), 131–141.

[37] Ahmad, M. T.; Zhang, P.; Dufresne, C.; Ferrucci, L.; Semba, R. D. The Human Eye Proteome Project: Updates on an Emerging Proteome. *Proteomics* **2018**, *18* (5–6), 1700394.
[38] Hooi, M. Y. S.; Raftery, M. J.; Truscott, R. J. W. Age-Dependent Racemization of Serine Residues in a Human Chaperone Protein. *Protein Sci.* **2013**, *22* (1), 93–100.
[39] 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 (Omaha).* **2011**, *33* (2), 131–141.
[40] Lyon, Y. A.; Sabbah, G. M.; Julian, R. R. Identification of Sequence Similarities among Isomerization Hotspots in Crystallin Proteins. *J. Proteome Res.* **2017**, *16* (4), 1797–1805.
[41] Sharma, K. K.; Santhoshkumar, P. Lens Aging: Effects of Crystallins. *Biochim. Biophys. Acta - Gen. Subj.* **2009**, *1790* (10), 1095–1108.

CHAPTER 4: The Influence of Asp Isomerization on Trypsin and Trypsin-like Proteolysis

Abstract

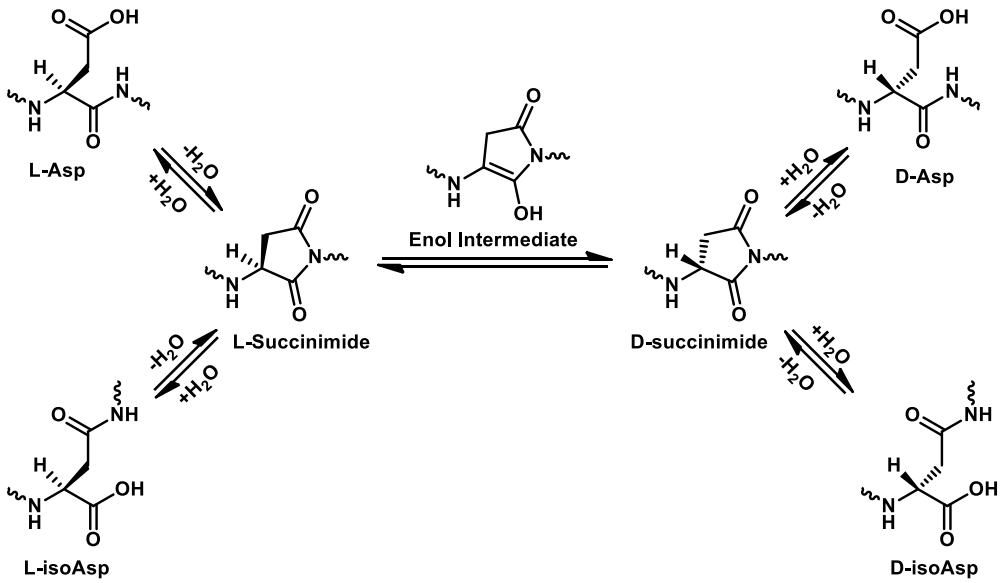
Long-lived proteins (LLPs), although less common than their short-lived counterparts, are increasingly recognized to play important roles in age-related diseases such as Alzheimer′s. In particular, spontaneous chemical modifications can accrue over time that serve as both indicators of and contributors to disrupted autophagy. For example, isomerization in LLPs is common and occurs in the absence of protein turnover while simultaneously interfering with protein turnover by impeding proteolysis. In addition to the biological implications this creates, isomerization may also interfere with its own analysis. To clarify, bottom-up proteomics experiments rely on protein digestion by proteases, most commonly trypsin, but the extent to which isomerization might interfere with trypsin digestion is unknown. Here, we use a combination of liquid chromatography and mass spectrometry to examine the effect of isomerization on proteolysis by trypsin and chymotrypsin. Isomerized aspartic acid and serine residues (which represent the most common sites of isomerization in LLPs) were placed at various locations relative to the preferred protease cleavage point to evaluate the influence on digestion efficiency. Trypsin was found to be relatively tolerant of isomerization, except when present at the residue immediately C-terminal to Arg/Lys. For chymotrypsin, the influence of isomerization on digestion was less predictable, resulting in long-range interference for some isomer/peptide

combinations. Given the trypsin- and chymotrypsin-like behavior of the 20S proteasome, and to further establish the biological relevance of isomerization in LLPs, substrates with isomerized sites were also tested against proteasomal degradation. Significant disruption of 20S proteolysis was observed, suggesting that if LLPs persist long enough to isomerize, it will be difficult for cells to digest them.

Introduction

Like a large movie with many supporting actors who appear only briefly, human biology is dominated by short-lived proteins that are synthesized and then degraded quickly. However, there also exist a number of long-lived proteins (LLPs) which, though their numbers may be few, can play important roles.[1] For example, mature fiber cells in the eye lens contain extreme LLPs, where protein lifetime is roughly equal to the age of the organism itself. Despite their lengthy lifetimes, the crystallin proteins in the eye lens are vital for maintaining sight.[2] An unavoidable consequence of increased lifespan is that LLPs have ample opportunity to accumulate post-translational modifications (PTMs), such as phosphorylation or glycosylation, or undergo spontaneous chemical modifications (SCMs), such as truncation, deamidation, oxidation, or isomerization.[3] Of these, isomerization is the most difficult to study because there is no change in atomic composition or mass, and the structural differences are subtle. At the same time, isomerization can be abundant, especially in the case of aspartic acid where transformation is facilitated by a five-membered succinimide intermediate produced by loss of water from the side chain (as illustrated in Scheme 4.1). This ring preferentially reopens by addition

of water to the opposite carbonyl, producing L-isoAsp where the natural backbone and side chain positions have inverted. In addition, the succinimide can switch chirality and reopen to produce D-Asp and D-isoAsp.



Scheme 4.1: Aspartic acid isomerization

The structural changes imparted by isomerization in LLPs can influence subsequent interactions with enzymes. For example, isomerization of serine (which occurs second in frequency after aspartic acid) strongly interferes with kinase activity and prevents phosphorylation.[4] It has also been recognized for some time that isomerization can inhibit protease action, which is relevant both for analytical analysis and biological function. In terms of analysis, modified protease behavior can be used to aid in the identification of peptides containing isomerized residues. For example, aminopeptidase M is unable to cleave D-amino acids from peptide N-termini, a shortcoming that has been used as an

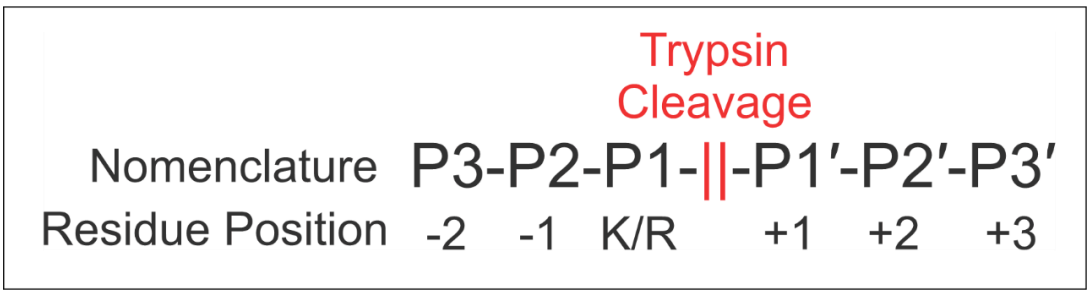
enrichment protocol to find D-amino acids in biological samples.[5] Asp-N cleaves at the N-terminal side of L-Asp residues, but it is unable to cleave isomerized Asp residues, allowing for missed cleavages to pinpoint isomerization in LLPs.[6] While examples are rare, certain enzymes recognize specific amino acid isomers and provide novel cleavage or modifications that can be used for identification of isomers. For example, protein isoaspartyl methyltransferase (PIMT) selectively binds L-isoAsp and methylates the side chain through S-adenosyl methionine, causing the succinimide ring to reform.[7] PIMT has been used in conjunction with succinimide-reactive reagents such as tris and hydrazine to introduce mass shifts at Asp isomerization sites, allowing easy identification of Asp isomers.[8,9] Paenidase is a novel protease which recognizes and cleaves at D-Aspartic acid and has been used to identify D-Asp in heavily isomerized proteins.[6,10] In addition, cathepsin B recognizes the L-isoAsp side chain as a C-terminus analog, resulting in endopeptidase cleavage at L-isoAsp as well as ligase activity.[11] For the majority of proteases however, isomerization causes structural changes that impede binding to the active site result and therefore reduce or prevent proteolysis.

Considering the biological perspective, substitution of D-amino acids in synthetic peptides used as antimicrobial drugs increases resistance to bacterial proteolysis thereby prolonging lifetime and improving effectiveness.[12] Similarly, peptide neurotransmitters and toxins have been found to contain D-amino acids which increase lifetimes by conferring resistance to proteolysis.[13,14] In terms of LLPs, isomerization can potentially disrupt proteostasis by preventing breakdown into constituent amino acids. For example, studies of the lysosomal cathepsins have shown that isomerized substrates cannot be fully

digested, which could eventually lead to lysosomal storage observed in Alzheimer′s disease (AD).[15,16]

Although the lysosome appears to be unable to degrade isomerized substrates, the ubiquitin-proteasome system is another possible pathway for protein degradation. Within the mammalian cell, the 26S proteasome is a major complex composed of a core 20S unit responsible for enzymatic digestion and 1-2 outer 19S regulatory units that act together for digestion and recycling of protein/peptide substrates.[17] Proteins tagged with ubiquitin are recognized by the 19S subunit and fed into the center of the complex for degradation. The core 20S proteasome unit can also operate independently without the attachment of 19S subunits and is an abundant form found in cells.[18] The number of 20S substrates is quite substantial, comprising around 20% of cellular proteins, which are recognized and degraded by 20S without ubiquitination.[19] Cleavage of substrates occurs through subunits β1, β2, and β5, which have caspase-, trypsin-, and chymotrypsin-like activity, respectively. In contrast to the 26S proteasome which recognizes ubiquitinated substrates via the 19S regulatory unit, the 20S proteasome recognizes and degrades damaged, unfolded, or otherwise disordered proteins.[20] Interestingly, disordered proteins or disordered regions in proteins are more likely to undergo significant amounts of Asn deamidation and Asp isomerization due to increased backbone flexibility and solvent accessibility.[21,22,23,24] In addition, isomerization itself can cause structural changes resulting in destabilization of ordered protein structure.[25,26] However, the question remains unanswered as to whether the proteasome can digest peptide/protein substrates that have been isomerized.

Although perturbed proteolysis has been recognized as a potential aid for analysis of isomerization, it may also present problems for bottom-up proteomic analyses,[27,28] which rely on digestion of whole proteins with proteases (most commonly trypsin), followed by tandem LC/MS analysis. Missed cleavages can result in significant problems for bottom-up proteomics such as fewer peptide identifications due to increased peptide size or inaccuracies in quantitative measurements due to inconsistent inclusion/exclusion of isomerized forms. Trypsin is the most abundant digestive enzyme present in the gut of all animals and is released to activate other digestive enzymes such as chymotrypsin as well as aid in the digestion of consumed protein.[29] Trypsin dominates bottom-up approaches because it yields low missed cleavages and strongly prefers cleavage at Lys/Arg residues, which have the correct abundance in proteins to generate peptides at roughly the right length for mass-spectrometric analysis. Despite being robust and very specific, trypsin activity can be affected by the residues surrounding Lys/Arg, and certain sequence motifs are known to impair digestion. For example, proline or basic residues at P1′ can impair digestion.[30] Acidic residues such as aspartic acid or glutamic acid in P2 or P1′ sites (see Scheme 4.2) are also known to cause moderate interference.[31] If canonical residues such as these affect trypsin, isomerized residues may also interfere with proteolysis.



Scheme 4.2. Standard proteolysis positioning

In this work we seek to answer two questions, (1) what is the effect of isomerization in proximity to trypsin and chymotrypsin cleavage sites on digestion efficiency, and (2) is the proteasome (which possesses trypsin and chymotrypsin-like activity) capable of degrading isomerized substrates? To this end, we test trypsin and chymotrypsin activity near isomerized residues in synthetic peptide substrates. Sequences were chosen with Asp residues near cleavage sites and tested by comparing trypsin/chymotrypsin cleavage of the all-L peptide against the sequences containing an isomerized Asp or epimerized Ser residue. Trypsin was found to primarily have difficulty processing isomerized substrates at the P1′ position, while for chymotrypsin, isomerization also caused substantial missed cleavages at sites further away. Similarly, the 20S proteasome was found to be unable to significantly degrade substrates with isomerized sites.

Experimental Methods

Materials. Organic solvents and reagents were purchased from Fisher Scientific, Sigma-Aldrich, or Acros Organics and used without further purification. FMOC-protected amino acids and Wang resins were purchased from Anaspec, Inc or Chem-Impex International. Trypsin was purchased from Sigma Aldrich (T0303). α-chymotrypsin was also purchased from Sigma Aldrich (C-4129).

Peptide Synthesis. Peptides were synthesized manually following a FMOC-protected solid-phase peptide synthesis protocol.[32] Following synthesis, peptides were RP-HPLC purified using a Thermo UltiMate 3000 RS variable wavelength detector and pump connected to a Phenomenex Kinetex 5µm EVO C18 250 x 21.2 mm column (P/N 00G-

4633-P0-AX) with Phenomenex SecurityGuard PREP Cartridge (P/N AJ0-9145). Samples were purified using water with 0.1% formic acid as mobile phase A, and acetonitrile with 0.1% formic acid as mobile phase B. Purified peptides were stored frozen in 50/50 v/v acetonitrile/water. Peptides were lyophilized and reconstituted in Fisher Optima water prior to determining concentrations and reaction with trypsin or chymotrypsin in buffer. Concentrations of peptides were obtained by measuring the absorbance at either 205nm or 280nm using an Agilent Cary 60 UV/Vis Spectrophotometer. Molar absorptivity constants were obtained from an online calculator made available by Marius Clore (https://spin.niddk.nih.gov/clore/).[33]

Digestion of Peptides with Trypsin/Chymotrypsin. All-L sequences and sequences containing isomers were digested separately with trypsin in equal ratios of enzyme to substrate (1:100 w/w). Chymotrypsin digestions were carried out at a ratio of 1:20 w/w enzyme to substrate. Trypsin digestions were carried out in 100mM Tris pH 7.8 at 37C. Chymotrypsin digestions were carried out in 100mM Tris pH 7.8, 10mM CaCl2 at 37C. All time points were quenched with 1% TFA to halt reaction progress prior to LCMS analysis.

Analysis. For LCMS analysis of synthetic peptides, an Agilent 1100 binary pump was used with a Thermo BetaBasic-18 3 µm C18 150 mm x 2.1 mm column interfaced to a Thermo Fisher Scientific LTQ mass spectrometer. Quenched reaction time points were diluted to a total concentration of ~5µM with water and 50µL of sample was loaded onto the sample loop for analysis of each time point. After sample loading but prior to analysis, peptides were desalted on-line by flushing the column at 1-5%B for 5 minutes with the

inject valve directing column output to waste. After 5 minutes the valve was switched to the inject position to direct samples to the MS for analysis. Samples were eluted using water with 0.1% formic acid as mobile phase A, and acetonitrile with 0.1% formic acid as mobile phase B. Gradients were optimized for separation and elution of digestion products for synthetic sequences used. For example, trypsin cleavage products of SEMRLEKDRFSVNL were separated over a gradient of 1-65%B in 60 minutes.

Proteasome Purification. Rat 20S proteasomes were purified from rat livers according to a previously-established protocol.[34] In brief, rat livers were homogenized on ice in lysis buffer (Tris-HCl 50 mM, pH 7.5, Sucrose 250 mM, EDTA 1 mM, DTT 1 mM, Benzonase endonuclease) and centrifuged at 4 °C for 15 min, at 1,000 *g*. The supernatant was mixed with NaCl to reach a final concentration of 0.5M and ultracentrifuged at 4 °C for 2 h, at 152,000 *g*. The resulting supernatant was ultracentrifuged again at 4 °C for 5 h, at 200,000 *g*. The resulting pellet was re-suspended in buffer containing Tris-HCl 25 mM, pH 7.5, EDTA 1 mM, DTT 1 mM and loaded onto a XK50/100 gel filtration column (•XK 50/100 column (GE Healthcare) packed with 1.8 L 4% Rapid Run™ Agarose Bead Standard 50-150 μm for gel filtration (GF) (Agarose Bead Technologies cat. # 4RRS-1000). Eluted fractions were tested by peptidase activity assay. Positive proteasome-containing fractions were pooled, loaded onto a Source 15Q strong anion exchange column (GE Healthcare cat. # 17-0947-01), and eluted over a gradient of 10-30% Buffer B (Tris-HCl 25 mM, pH 7.5, NaCl 1M, EDTA 1 mM, DTT 1 mM). Positive proteasome-containing fractions were pooled, loaded onto a HiTrap® DEAE Fast Flow weak anion exchange column (GE Healthcare cat. # 17-5055-01) and eluted over a step gradient of 10%, 30%, 40%, 50%,

60% and 70% buffer B. Positive proteasome-containing fractions were pooled, buffer exchanged to 10 mM phosphate buffer, pH 7.4 and loaded onto a CHT column •Tricorn 10/10 column packed with 8 ml of type I ceramic hydroxiapatite (CHT) (Bio-Rad cat. # 1582000). Proteins were eluted over a step gradient of 10%, 20%, 40%, 50% and 60% buffer (400 mM phosphate buffer, at pH 7.4). Positive proteasome-containing fractions were pooled and concentrated before storage.

Proteasome Digestion of Peptide Substrate. Peptides were dissolved in 150mM ammonium acetate to a concentration of 20µM, and frozen in small aliquots. On the day of the measurement, a peptide aliquot was thawed and diluted 10x with ammonium acetate and kept on ice. A sample of purified rat 20S proteasomes, at a concentration of 4µM was buffer exchanged into 150 mM ammonium acetate. For each measurement, 1µl of the 20S proteasome was mixed with 1µl of the peptide. The sample was immediately loaded into a nano-electrospray needle and measured continuously over a period of 20 minutes. In control reactions, each peptide was diluted to 1µM and measured without the 20S proteasome over the same period of time (20 minutes). Spectra were recorded on a modified Q Exactive Plus at a resolution of 70,000, at 250 millisecond injection time and with low averaging -5. Capillary voltage was set to 1.7 kV, at a temperature of 180°C. Fore-vacuum pressure was set to 1.5 mbar, and the trapping gas pressure to 1, corresponding to HV pressure of 3.4x10-5 mbar, and UHV pressure of 9.2x10-11 mbar. The source was operated at a constant energy of 2V in the flatapole bias and interflatapole lens. Bent flatapole DC bias and gradient were set to 2 and 10V, respectively, and the HCD cell was operated at an energy of 1V. In each experiment, 20 time points were taken

(averaged over minute, each). The results represent an average of three experiments, and error bars represent standard deviation. Peak intensities were extracted using the Qual Browser software, part of the Thermo Xcalibur 4.1.50 software package.

Results and Discussion

Isomerization near trypsin cleavage sites. To test the effect of different Asp isomers on trypsin cleavage in a substrate with multiple cleavage sites, the crystallin-derived synthetic sequence SEMRLEK**D**RFSVNL with either L-/L-iso/D-/D-isoAsp at Asp-8 was incubated with trypsin. After 24 hours, the digestions were halted with acid and the products analyzed via LCMS. The chromatogram for the all-L sequence (Fig. 4.1a) reveals cleavage products at each of the three trypsin cleavage sites in the peptide. The inset bar plot in Fig. 4.1a quantifies the fractional abundance of peptides resulting from hydrolysis at each site (blue) relative to the fractional abundance of peptides containing a missed cleavage at the same site (orange). For example, the blue bar for all-L Arg-4 (Fig. 4.1a) represents the sum of the fractional abundances for SEMR, LEKDRFSVNL, LEKDR, and LEK as these fragments are all products resulting from cleavage at Arg-4. The orange bar represents the fractional abundances for SEMRLEK and SEMRLEKDR, which both contain a missed cleavage at Arg-4.

The propensity for digestion as indicated by the size of the blue bars relative to the orange bars at each site reveals preferential digestion of canonical SEMRLEK**D**RFSVNL at Lys-7, followed by Arg-9 and Arg-4 (Fig. 1a). This result suggests that canonical Glu-6 and Asp-8 do not interfere with digestion at Lys-7 despite the known aversion of trypsin to

acidic residues in these positions. The propensities for digestion at each site change dramatically for the L-isoAsp peptide, particularly at Arg-4, which becomes the preferred site, and at Lys-7, which shifts from the most cleaved to least cleaved site (Fig. 4.1b). For the D-Asp peptide, Arg-4 is again preferred, but very little cleavage occurs at Lys-7 or Arg-9 (Fig. 4.1c). The D-isoAsp isomer shifts proteolysis from Lys-7 to either Arg-4 or Arg-9 (Fig. 4.1d). In summary, isomerization at Asp-8 largely impedes hydrolysis at Lys-7 regardless of the Asp isomer. Importantly isomers only reside in the P1′ position relative to Lys-7, while for Arg-4 and Arg-9, the isomers occupy sites P4′ and P2, respectively. Although Arg-4 is not a preferred cleavage site for the canonical peptide, it becomes the preferred cleavage site for all Asp isomers. This most likely occurs because Arg-4 is most distant from the site of isomerization. Although both Lys-7 and Arg-9 are immediately adjacent to the site of isomerization, significantly greater interference is observed for Lys-7. In fact, for D-isoAsp, cleavage at Arg-9 actually increases relative to the canonical peptide. This suggests that the P1′ position for trypsin allows the least flexibility in accommodating isomerized substrates, which may be analogous to the similar trend

observed previously for proline. Overall, the results clearly indicate that isomerization can interfere with trypsin digestion and increase the likelihood for missed cleavages to occur.

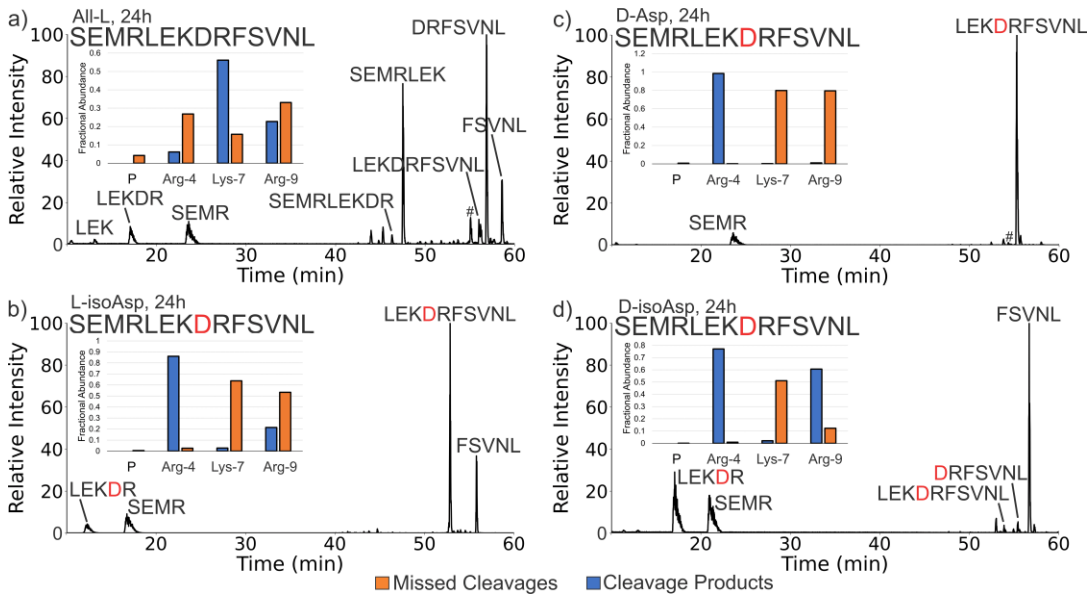


Figure 4.1: Chromatograms showing major products after 24 hours of trypsin digestion of SEMRLEKDRFSVNL for the (a) All-L, (b) L-isoAsp-8, (c) D-Asp-8, and (d) D-isoAsp-8 sequences. Bar plots are shown inside each chromatogram summarizing the propensity for cleavage at each of the three possible sites. Blue bars represent the sum of the fractional abundance of all peptides terminating at the cleavage site (i.e. fragments with C-terminal Lys/Arg or the corresponding complement fragments). Orange bars represent the sum of fractional abundance of peptides containing missed cleavages (e.g. any peptide containing RL for Arg-4). P denotes the fractional abundance of the precursor. # corresponds to precursor peptide

Isomerization-site proximity effect. Having established that trypsin is affected by isomerization near cleavage sites, we further examined a series of four synthetic sequences derived from highly isomerized LLPs. These peptides were chosen to test the effects of L-isoAsp in the P3, P2, P1′, and P2′ positions on trypsin digestion as shown in Fig. 4.2a. Each L-isoAsp-containing peptide and the corresponding all-L peptide were incubated separately with trypsin and the progress of cleavage compared at a series of time points

over the course of the reaction via LCMS. One example of raw data is shown in Fig. 4.2b, where it is clear that the canonical peptide is digested much more than the isomerized version. To quantitatively assess the relative extent of trypsin digestion for each isomerization site, rate constants for degradation of each peptide were calculated by fitting a first-order rate equation (Fig. 4.S1). The ratios of the all-L rate constants to the L-isoAsp rate constants are shown in Fig. 4.2c. Referencing the extent of digestion to the canonical peptide in each case allows examination of the influence of the L-isoAsp substitution independent of other sequence effects that may vary between peptides. Isomerization at P1′ (GDYKDSSDF) leads to a clear preference for digestion of the all-L form, as indicated by the large ratio shown in Fig. 4.2c. Digestion of the all-L peptide is also favored relative to L-isoAsp at the P3, P2, and P2′ positions, though to much lesser extent. The results again suggest that isomerization at P1′ is particularly problematic for trypsin recognition. Further investigation of this position revealed that L-isoAsp is not the only Asp isomer found to inhibit cleavage when present at P1′, D-Asp and D-isoAsp also inhibit trypsin cleavage, and D-Serine also yields similar effects (Fig. 4.S2). This suggests that while isomerization at other nearby locations causes a reduction in trypsin cleavage efficiency, isomerization at P1′ is particularly likely to slow proteolysis. These findings also agree with the data in Fig. 4.1 for SEMRLEKDRFSVNL, in which cleavage at Lys-7 was strongly perturbed by isomerization at Asp-8, which also occupies the P1′ position.

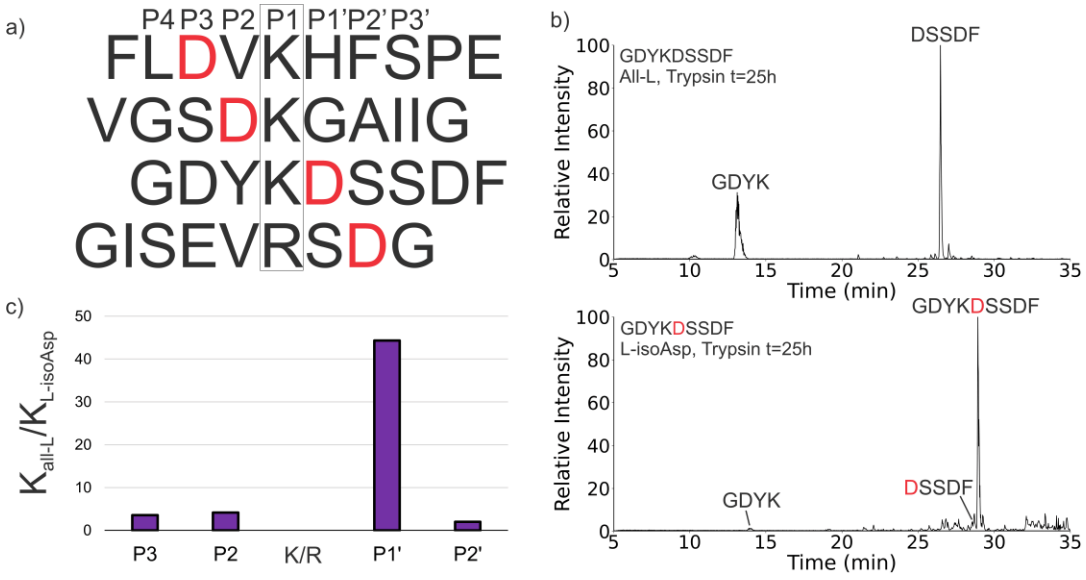


Figure 4.2: (a) Four sequences with an Asp residue at either P3, P2, P1′, or P2′ (labeled in red) were tested with trypsin. (b) Representative chromatogram showing trypsin cleavage of GDYKDSSDF All-L at t=25h and cleavage of GDYKDSSDF L-isoAsp-5 at t=25h. (c) Bar plot showing the comparison of the ratios of first-order rate constants for all-L precursor (Kall-L) and L-isoAsp precursor (KL-isoAsp). Larger bars indicate a higher preference for all-L sequence cleavage relative to the L-isoAsp containing sequence.

Chymotrypsin cleavage of isomerized substrates. Chymotrypsin shares many similarities with trypsin in terms of sequence, structure, and catalytic site while simultaneously exhibiting quite different substrate specificity.[35] Consequently, chymotrypsin is frequently used in bottom-up proteomics to produce digestion at aliphatic amino acids. To evaluate the effect of isomers on Chymotrypsin digestion, six variants of APSWFDTGLSEMR (all-L, L-isoAsp-6, D-Asp-6, D-isoAsp-6, D-Ser-3, D-Ser-10) were examined. After incubation with chymotrypsin for 21 hours, all digestion products were determined by LCMS and the results are shown in Fig. 4.3 as a function of cleavage propensity. The all-L sequence is preferentially cut at Phe-5 with minor activity at Trp-4 and Leu-9 (Fig. 4.3a). Digestion of the L-isoAsp-containing peptide reveals a shift in

behavior with the Leu-9 site being targeted, while no activity was observed at Trp-4 and only minimal cleavage at Phe-5 (Fig. 4.3b). This behavior can be most easily rationalized by the proximity of Trp-4 and Phe-5 to L-isoAsp. In contrast, the D-Asp variant yields virtually no chymotrypsin digestion at any of the three sites (Fig. 4.3c). Presumably the stereochemical inversion of the L-Asp side chain to D-Asp results in steric clash that largely prevents access to the chymotrypsin active site. Interestingly, D-isoAsp behaves similarly to L-isoAsp (Fig. 4.3d). Changing L-Ser-3 to D-Ser-3 does not completely impair digestion, but instead allows significant cleavage at Phe-5 while slightly increasing digestion at Leu-9 relative to the all-L sequence (Fig. 4.3e). In contrast, D-Ser-10 essentially only permits cleavage at Phe-5 while shutting down digestion at Leu-9 and Trp-4 (Fig. 4.3f). The reduced recognition at Leu-9 is easily explained by proximity, but the high digestion at Phe-5 coupled with reduced digestion at Trp-4, which is further away from the isomerized residue, suggests that longer range effects can sometimes influence behavior in unexpected ways. While D-Ser-3 and D-Ser-10 were both found to affect chymotrypsin cleavage, a direct comparison of Ser isomerization to Asp isomerization is difficult to make due to the positional differences relative to the chymotrypsin cleavage sites. To more directly compare Asp and Ser isomerization, we prepared the all-L serine analog APSWFSTGLSEMR as well as the D-Ser-6 epimer and digested with chymotrypsin (Fig. 4.3g and 4.3h). The all-L peptide yields results very similar to the original sequence (compare Fig. 4.3a/4.3g). Surprisingly, the D-Ser-6 epimer was almost untouched by chymotrypsin, yielding results similar to the D-Asp epimer for APSWFDTGLSEMR. This suggests that the sidechain stereochemistry may be more important than side chain identity

for impairing active site fit in some cases. In sum, these results show that chymotrypsin is sensitive to isomerization at a variety of positions relative to the cleavage site, which could make chymotrypsin a less predictable choice for the digestion of potentially isomerized LLPs.

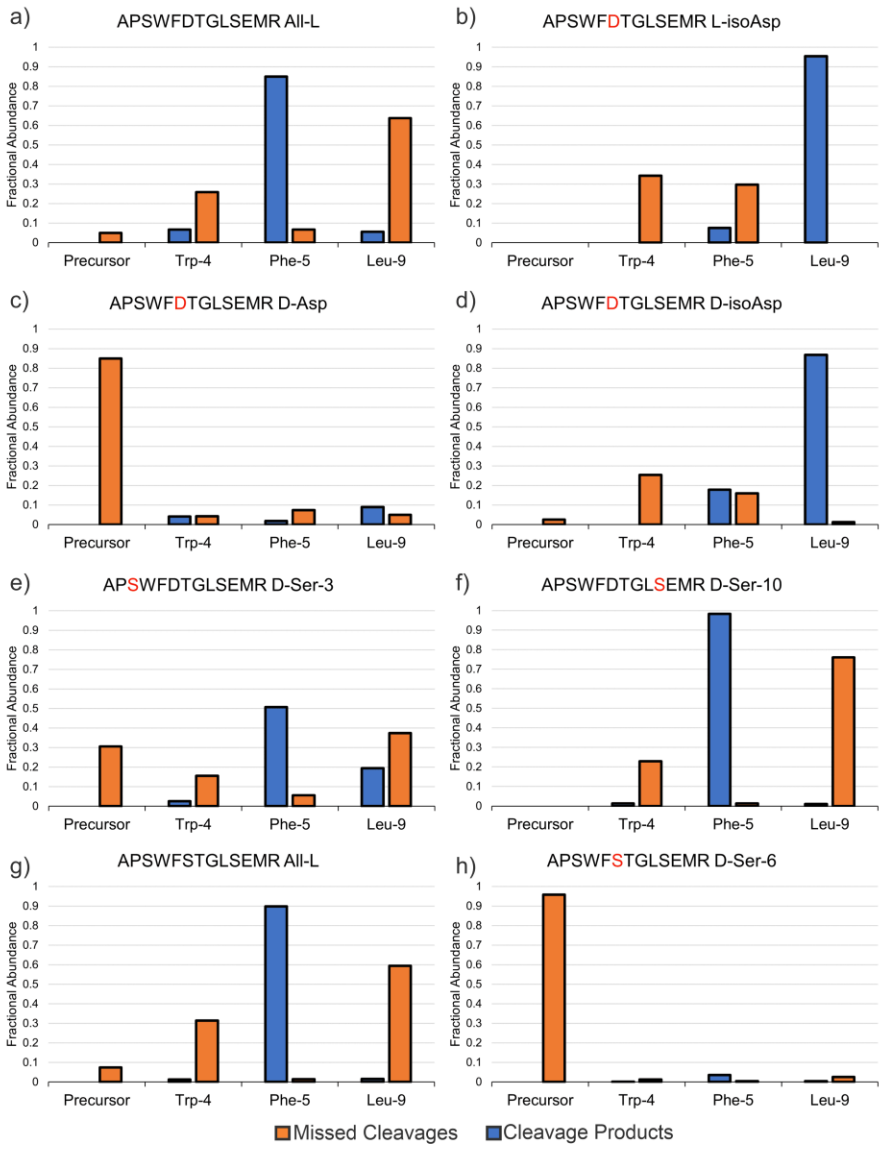


Figure 4.3: Chymotrypsin cleavage results at 21h of (a) APSWFDTGLSEMR All-L, (b) APSWF**D**TGLSEMR L-isoAsp, (c) APSWF**D**TGLSEMR D-Asp, (d) APSWF**D**TGLSEMR D-isoAsp, (e) AP**S**WFDTGLSEMR D-Ser-3, (f) APSWFDTGL**S**EMR D-Ser-10, (g) APSWFSTGLSEMR All-L, and (h) APSWF**S**TGLSEMR D-Ser-6. Blue bars represent the sum of fractional abundance of sequences produced from cleavage at a particular site (e.g. APSW and FDTGLSEMR for Trp-4 cleavage), while orange bars represent the fractional abundance of missed cleavage products at a specific site.

Proteasome digestion of isomerized substrates. Although not typically used for proteomics experiments, the proteasome houses both trypsin-like and chymotrypsin-like proteolytic sites as well as caspase-like activity. Furthermore, given that previous work has established that the lysosomal system cannot fully degrade isomerized peptide substrates, the potential digestion of isomerized peptides by the proteasome remains an important biological question. In order to test proteasome activity, larger peptide targets are required to make suitable substrates. We synthesized VKK**D**QLGKNE EGAPQEGILE **D**MPV**D**P**D**NEA YEMPSEEG, a sequence derived from alpha-synuclein, in both the all-L-Asp and all-L-isoAsp forms (L-isoAsp/Asp sites indicated in bold). The individual peptides were mixed with the 20S proteasome and monitored over 20 minutes via nano-electrospray. The signal for the intact all-L peptide decreased rapidly to zero over 10-15 minutes digestion time, while digestion product peaks appeared (Fig. 4.4a). In contrast, the all-(L-isoAsp) peptide persisted when incubated in the presence of the 20S proteasome, being virtually untouched by 20S digestion after 20 minutes and producing only minor cleavage products (Fig. 4.4b). Interestingly, the two main degradation products for the all-(L-Asp) sequence were identified as all-(L-Asp) 5-38, resulting from a single caspase-like cleavage at Asp-4, and all-(L-Asp) 5-25, resulting from sequential caspase cleavage at Asp-4 and Asp-25 (Fig. 4c). Changing Asp-4 and Asp-25 to L-isoAsp seems to prevent any caspase-like cleavage of the substrate, as these products are not observed in significant amounts for the all-(L-isoAsp) sequence. Furthermore, digestion is not shifted to other locations as was observed for trypsin/chymotrypsin. To quantify the 20S digestion results, the fractional abundance for each peptide precursor and observed cleavage products are

plotted versus reaction time in Figure 4.4d. The all-L substrate decreases in fractional abundance quickly (Fig. 4.4d, purple trace), while the all-(L-isoAsp) substrate changes very little in fractional abundance at any time point (Fig. 4.4d, yellow trace). Given that a full 16-residue portion of the peptide contains no isomerized sites, and multiple potential cleavage sites are located several residues from any isomers, it appears that the proteasome may have difficulty processing isomerized substrates. This may be related to reduced conformational flexibility available to substrates within the proteasome. Although this proximity may be efficient for promoting rapid, successive digestion events, it may also make it more difficult for modified substrates to be accommodated in the active sites.

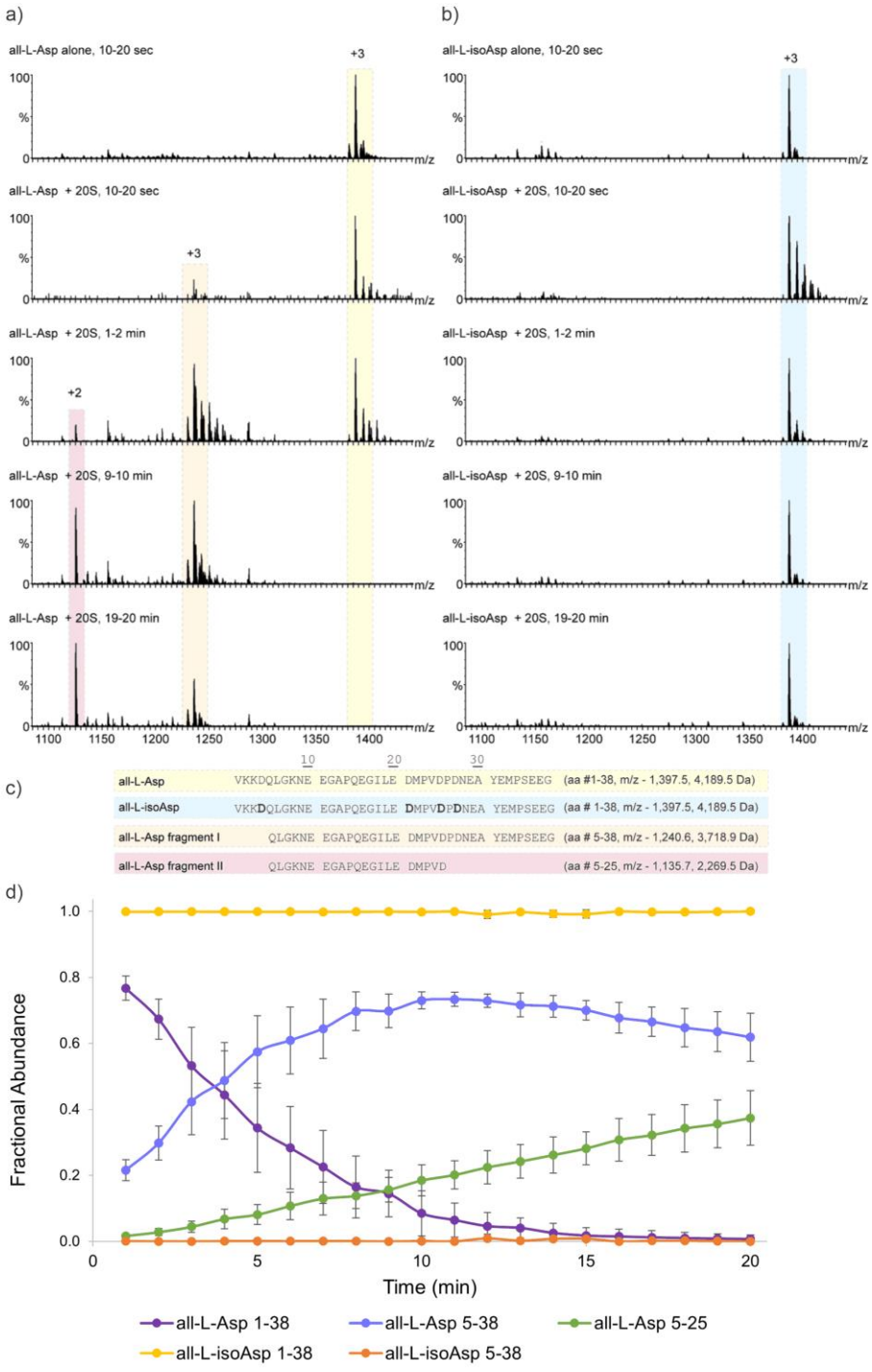

a)
b)
all-L-Asp alone, 10-20 sec
all-L-isoAsp alone, 10-20 sec
all-L-Asp + 20S, 10-20 sec
all-L-isoAsp + 20S, 10-20 sec
all-L-Asp + 20S, 1-2 min
all-L-isoAsp + 20S, 1-2 min
all-L-Asp + 20S, 9-10 min
all-L-isoAsp + 20S, 9-10 min
all-L-Asp + 20S, 19-20 min
all-L-isoAsp + 20S, 19-20 min
+3
+2
100
%
0
m/z
1100 1150 1200 1250 1300 1350 1400
c)
all-L-Asp VKKDQLGKNE EGAPQEGILE DMPVDPDNEA YEMPSEEG (aa #1-38, m/z - 1,397.5, 4,189.5 Da)
all-L-isoAsp VKKDQLGKNE EGAPQEGILE DMPVDPDNEA YEMPSEEG (aa # 1-38, m/z - 1,397.5, 4,189.5 Da)
all-L-Asp fragment I QLGKNE EGAPQEGILE DMPVDPDNEA YEMPSEEG (aa # 5-38, m/z - 1,240.6, 3,718.9 Da)
all-L-Asp fragment II QLGKNE EGAPQEGILE DMPVD (aa # 5-25, m/z - 1,135.7, 2,269.5 Da)
d)
Fractional Abundance
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
Time (min)
all-L-Asp 1-38
all-L-Asp 5-38
all-L-Asp 5-25
all-L-isoAsp 1-38
all-L-isoAsp 5-38

Figure 4.4: 20S Proteasome digestion of alpha-synuclein 1-38. (a) Averaged mass spectra for 20S digestion of all-(L-Asp) substrate at specific time ranges. Precursor is highlighted in yellow, major digestion products are highlighted in orange and red. (b) 20S digestion of all-(L-isoAsp) substrate. Precursor is highlighted in blue. (c) Sequences of precursor and cleavage products. Isomerized residues are indicated with bold letters. (d) Fractional abundance plot of precursors and cleavage products. Each point represents the average of three experiments, and error bars represent the standard deviation.

Conclusion

The inherent longevity of LLPs enables the possibility for Asp isomerization, which presents nearly universal problems for proteolysis. In the context of LLP analysis, trypsin exhibits the least sensitivity to isomerization and therefore remains the preferred protease for proteomic experiments. However, missed cleavages should be enabled during data analysis to identify those sites where isomerization may have blocked digestion, and special care will need to be taken in quantitative experiments if dealing with isomerized LLPs. In particular, for LCMS experiments where isomers elute with partial or complete separation, the quantitation should include all isomers, or (ideally) each isomer should be quantified individually. If isomerization is less pervasive, proteases other than trypsin will likely be suitable, although the same caveats for missed cleavages would still apply. In terms of cell biology, we demonstrated that the 20S proteasome, although containing trypsin-like sites, is not able to efficiently digest isomerized peptides. When considered in light of previous work showing that the lysosome is also incapable of digesting isomerized LLPs, it is not clear how cells (particularly post-mitotic cells) would be able to clear isomerized LLPs. Recent work has shown that Asp isomerization in tau is higher in AD versus control samples and is well-correlated with markers indicative of impaired autophagic flux.[36] Given that the burden of LLP isomerization is also unlikely to be reduced by the proteasome, the presence of isomerization should be an excellent indicator of disrupted proteostasis.

Acknowledgements. The authors are grateful for funding from the NIH (R01 AG066626). MS is grateful for the support of the Sagol Institute for Longevity Research grant and a Moross Proof-of-Concept grant. MS is the incumbent of the Aharon and Ephraim Katzir Memorial Professorial Chair.

Supporting Information

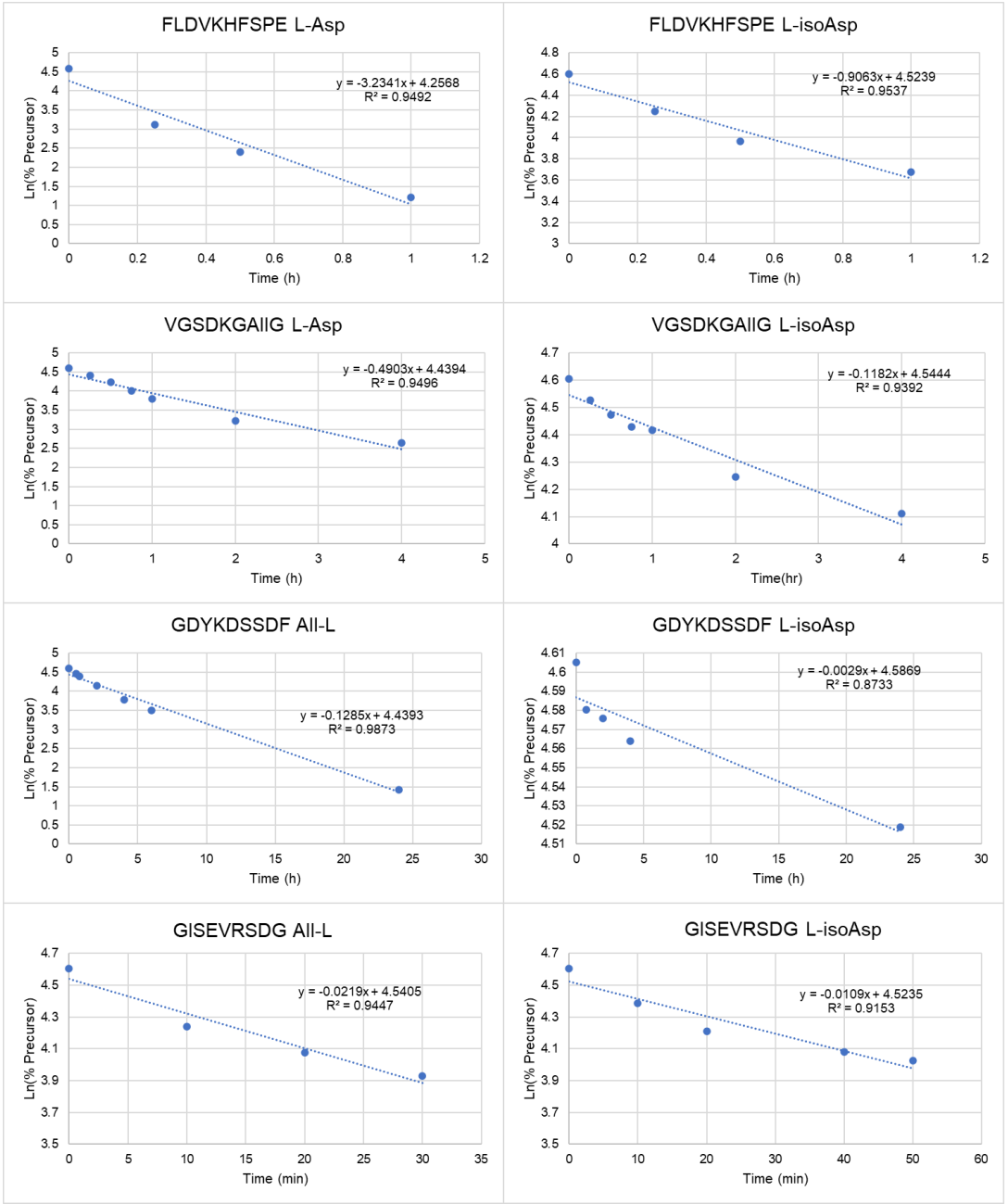


Figure 4.S1: Ln(% Precursor) vs time plots with linear fit for all-L and L-isoAsp peptides tested with trypsin. The rate constants for trypsin degradation of the precursor are given by the slopes of the best-fit line

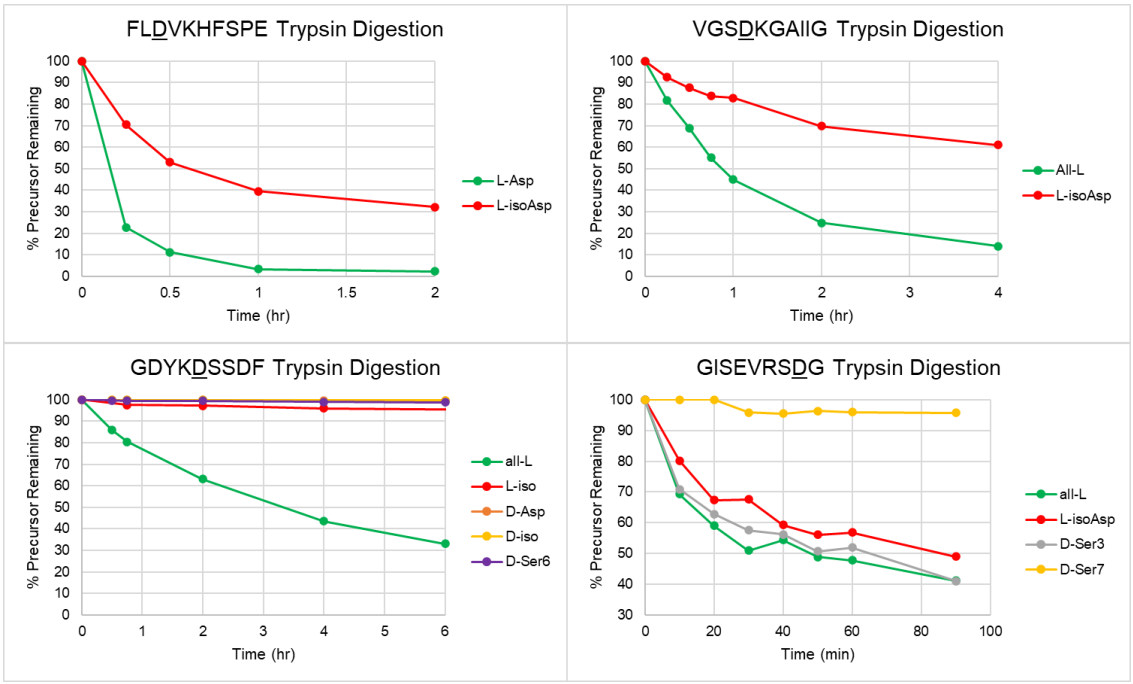


Figure 4.S2: % all-L, L-isoAsp, D-Asp, D-isoAsp, and D-Ser-X precursor remaining versus time for synthetic sequences tested with trypsin.

References

[1] Truscott, R. J. W.; Schey, K. L.; Friedrich, M. G. Old Proteins in Man: A Field in Its Infancy. *Trends Biochem. Sci.* **2016**, *41* -8, 654–664..

[2] Samuel Zigler, J.; Goosey, J. Aging of Protein Molecules: Lens Crystallins as a Model System. *Trends Biochem. Sci.* **1981**, *6* (1), 133–136.

[3] Toyama, B. H.; Hetzer, M. W. Protein Homeostasis : Live Long, Won't Prosper. **2013**, *14* (January), 55–61.

[4] Lyon, Y. A.; Collier, M. P.; Riggs, D. L.; Degiacomi, M. T.; Benesch, J. L. P.; Julian, R. R. Structural and Functional Consequences of Age-Related Isomerization in α-Crystallins. *J. Biol. Chem.* **2019**, *294* (19), 7546–7555.

[5] Livnat, I.; Tai, H. C.; Jansson, E. T.; Bai, L.; Romanova, E. V.; Chen, T. T.; Yu, K.; Chen, S. A.; Zhang, Y.; Wang, Z. Y.; et al. A D-Amino Acid-Containing Neuropeptide Discovery Funnel. *Anal. Chem.* **2016**, *88* (23), 11868–11876.

[6] Fujii, N.; Maeda, H.; Takata, T.; Fujii, N.; Sakaue, H.; Nirasawa, S.; Takahashi, S.; Sasaki, H. Rapid Survey of Four Asp Isomers in Disease-Related Proteins by LC-MS Combined with Commercial Enzymes. *Anal. Chem.* **2015**, *87* (1), 561–568.

[7] Lowenson, J. D.; Clarke, S. Structural Elements Affecting the Recognition of L-Isoaspartyl Residues by the L-Isoaspartyl/D-Aspartyl Protein Methyltransferase. Implications for the Repair Hypothesis. *J. Biol. Chem.* **1991**, *266* (29), 19396–19406.

[8] Silzel, J. W.; Lambeth, T. R.; Julian, R. R. PIMT-Mediated Labeling of L-Isoaspartic Acid with Tris Facilitates Identification of Isomerization Sites in Long-Lived Proteins. *J. Am. Soc. Mass Spectrom.* **2022**, *33* -3, 548–556.

[9] Alfaro, J. F.; Gillies, L. A.; Sun, H. G.; Dai, S.; Zang, T.; Klaene, J. J.; Byung, J. K.; Lowenson, J. D.; Clarke, S. G.; Karger, B. L.; et al. Chemo-Enzymatic Detection of Protein Isoaspartate Using Protein Isoaspartate Methyltransferase and Hydrazine Trapping. *Anal. Chem.* **2008**, *80* -10, 3882–3889.

[10] Takahashi, S.; Ogasawara, H.; Hiwatashi, K.; Hori, K.; Hata, K.; Tachibana, T.; Itoh, Y.; Sugiyama, T. Paenidase, a Novel d-Aspartyl Endopeptidase from Paenibacillus Sp. B38: Purification and Substrate Specificity. *J. Biochem.* **2006**, *139* (2), 197–202.

[11] Lambeth, T. R.; Dai, Z.; Zhang, Y.; Julian, R. R. A Two-Trick Pony: Lysosomal Protease Cathepsin B Possesses Surprising Ligase Activity. *RSC Chem. Biol.* **2021**, *2* (2), 606–611.

[12] Hamamoto, K.; Kida, Y.; Zhang, Y.; Shimizu, T.; Kuwano, K. Antimicrobial Activity and Stability to Proteolysis of Small Linear Cationic Peptides with D-Amino Acid Substitutions. *Microbiol. Immunol.* **2002**, *46* (11), 741–749.

[13] Checco, J. W.; Zhang, G.; Yuan, W.; Yu, K.; Yin, S.; Roberts-Galbraith, R. H.; Yau, P. M.; Romanova, E. V.; Jing, J.; Sweedler, J. V. Molecular and Physiological Characterization of a Receptor for D -Amino Acid-Containing Neuropeptides. *ACS Chem. Biol.* **2018,** *13*, 1343–1352.

[14] Genchi, G. An Overview on D-Amino Acids. *Amino Acids* **2017**, *49* (9), 1521–1533.

[15] 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.

[16] Pandey, G.; Julian, R. R. LC-MS Reveals Isomeric Inhibition of Proteolysis by Lysosomal Cathepsins. *Anal. Sens.* **2022**.

[17] Glickman, M. H.; Ciechanover, A. The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. *Physiol. Rev.* **2002**, *82* (2), 373–428.

[18] Kumar Deshmukh, F.; Yaffe, D.; Olshina, M.; Ben-Nissan, G.; Sharon, M. The Contribution of the 20S Proteasome to Proteostasis. *Biomolecules* **2019**, *9* -5, 190.

[19] Baugh, J. M.; Viktorova, E. G.; Pilipenko, E. V. Proteasomes Can Degrade a Significant Proportion of Cellular Proteins Independent of Ubiquitination. *J. Mol. Biol.* **2009**, *386* -3, 814–827.

[20] Pickering, A. M.; Davies, K. J. A. Degradation of Damaged Proteins. *Prog. Mol. Biol. Transl. Sci*. **2012**, *109*, 227–248.
[21] Xie, M.; Schowen, R. L. Secondary Structure and Protein Deamidation. *J. Pharm. Sci.* **1999**, *88* (1), 8–13.
[22] Stevenson, C. L.; Friedman, A. R.; Kubiak, T. M.; Donlan, M. E.; Borchardt, R. T. Effect of Secondary Structure on the Rate of Deamidation of Several Growth Hormone Releasing Factor Analogs. *Int. J. Pept. Protein Res.* **2009**, *42* -6, 497–503.
[23] Lyon, Y. A.; Sabbah, G. M.; Julian, R. R. Identification of Sequence Similarities among Isomerization Hotspots in Crystallin Proteins. *J. Proteome Res.* **2017**, *16* -4, 1797–1805.
[24] Wakankar, A. A.; Borchardt, R. T.; Eigenbrot, C.; Shia, S.; Wang, Y. J.; Shire, S. J.; Liu, J. L. Aspartate Isomerization in the Complementarity-Determining Regions of Two Closely Related Monoclonal Antibodies. *Biochemistry* **2007**, *46* -6, 1534–1544.
[25] Noguchi, S.; Miyawaki, K.; Satow, Y. Succinimide and Isoaspartate Residues in the Crystal Structures of Hen Egg-White Lysozyme Complexed with Tri-N-Acetylchitotriose. *J. Mol. Biol.* **1998**, *278* (1), 231–238.
[26] Noguchi, S. Structural Changes Induced by the Deamidation and Isomerization of Asparagine Revealed by the Crystal Structure of Ustilago Sphaerogena Ribonuclease U2B. *Biopolymers* **2010**, *93* (11), 1003–1010.
[27] Fujii, N.; Sakaue, H.; Sasaki, H.; Fujii, N. A Rapid, Comprehensive Liquid Chromatography-Mass Spectrometry (LC-MS)-Based Survey of the Asp Isomers in Crystallins from Human Cataract Lenses. *J. Biol. Chem.* **2012**, *287* (47), 39992–40002.
[28] 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 Commun.* **2021**, *3* (2), 1–17.
[29] Whitcomb, D. C.; Lowe, M. E. Human Pancreatic Digestive Enzymes. *Dig. Dis. Sci.* **2007**, *52* (1), 1–17.
[30] Siepen, J. A.; Keevil, E.-J.; Knight, D.; Hubbard, S. J. Prediction of Missed Cleavage Sites in Tryptic Peptides Aids Protein Identification in Proteomics. *J. Proteome Res.* **2007**, *6* (1), 399–408.
[31] Šlechtová, T.; Gilar, M.; Kalíková, K.; Tesařová, E. Insight into Trypsin Miscleavage: Comparison of Kinetic Constants of Problematic Peptide Sequences. *Anal. Chem.* **2015**, *87* (15), 7636–7643.
[32]Hood, C. A., Fuentes, G., Patel, H., Page, K., Menakuru, M., Park, J. H. Fast Conventional Fmoc Solid-Phase Peptide Synthesis With HCTU. *J. Pept. Sci.* **2008**, 14, 97–101.
[33] Anthis, N.J. & Clore, G.M. Sequence-Specific Determination of Protein and Peptide Concentrations By Absorbance at 205 nm. *Protein Sci*. **2013**, *22*, 851-858.
[34] Ben-Nissan G, Vimer S, Tarnavsky M, Sharon M. Structural Mass Spectrometry Approaches To Study The 20S Proteasome. *Methods Enzymol*. **2019**, *619*, 179-223.
[35] Ma, W.; Tang, C.; Lai, L. Specificity of Trypsin and Chymotrypsin: Loop-Motion-Controlled Dynamic Correlation as a Determinant *Biophys. J.* **2005,** *89*, 1183–1193.
[36] 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.; et al. Does Data-Independent Acquisition Data Contain Hidden Gems? A Case Study Related to Alzheimer's Disease. *J. Proteome Res.* **2022**, *21* (1), 118–131.

CHAPTER 5: RDD-HCD Provides Variable Fragmentation Routes Dictated by Radical Stability

Abstract

Radical-directed dissociation (RDD) is a fragmentation technique in which a radical created by selective 213/266nm photodissociation of a carbon-iodine bond is re-isolated and collisionally activated. In previous RDD experiments, collisional activation was effected by ion-trap collision-induced dissociation (CID). Higher-energy collisional dissociation (HCD) differs from CID both in terms of how ions are excited and in the number, type, or abundance of fragments that are observed. In this paper we explore the use of HCD for activation in RDD experiments. While RDD-CID favors fragments produced from radical-directed pathways such as a/z-ions and side chain losses regardless of the activation energy employed, RDD-HCD spectra vary considerably as a function of activation energy, with lower energies favoring RDD while higher energies favor products resulting from cleavage directed by mobile protons (b/y-ions). RDD-HCD therefore affords more tunable fragmentation based on the HCD energy provided. Importantly, the abundance of radical products decreases as a function of increasing HCD energy, confirming that RDD generally proceeds via lower energy barriers relative to mobile-proton-driven dissociation. The dominance of b/y ions at higher energies for RDD-HCD can therefore be explained by the higher survivability of fragments not containing the radical after the initial or subsequent dissociation events. Furthermore,

these results confirm previous suspicions that HCD spectra differ from CID spectra due to multiple dissociation events.

Introduction

Radical-directed dissociation (RDD) is a tandem mass spectrometry technique in which a radical is created site-specifically on a biomolecule and is subsequently activated by collisions to induce fragmentation. Prior to dissociation, the radical typically migrates to nearby sites by hydrogen atom abstraction, which lends a high degree of structural sensitivity to the method.[1] Migration sites are determined by a combination of structural constraints and the relative bond-dissociation energies of the initial and final sites.[2] Radicals can be created by addition of a chromophore such as 4-iodobenzoic acid, followed by highly specific photodissociation of the carbon-iodine bond, or by addition of a functional group labile to collisional activation, such as Tempo.[3] Radicals can then (in either case) be re-isolated and subjected to additional collisional activation in an MS^3 step. Differences in RDD fragmentation have allowed disambiguation of many classes of similar molecules, including lipid isomers varying only by the position of double bonds in the fatty acid chains.[4,5,6,7,8] RDD can also distinguish glycan oligomers differing in composition, configuration, and connectivity.[9,10] In particular, RDD is noted for sensitivity to stereochemistry, enabling distinction of glycosphingolipid epimers which only varied by the orientation of a single OH group in either the axial or equatorial position.[11] In similar applications, RDD has been used to distinguish isomeric peptides where a single side chain was inverted from the L to D configuration.[1]

RDD tends to produce different fragment ion types than those observed with traditional CID experiments on protonated analytes. For peptides, CID spectra are dominated by b/y fragments while RDD generates primarily a/x and c/z backbone fragments in addition to partial and complete side chain losses.2 It is interesting to consider why RDD fragmentation is different from traditional CID, given that both methods involve the use of collisional activation. One hint can be derived from a very small minority of peptides that behave as antioxidants and are able to sequester radicals and yield RDD spectra dominated by b/y ions.[12] This suggests that in the absence of radical sequestration, RDD is favored over proton-initiated fragmentation, presumably by facilitating lower-energy dissociation thresholds.[2] However, ion-trap CID is not well suited for varying energy deposition because ions are slowly heated by thousands of collisions and products are cooled immediately after creation. In contrast, higher-energy collisional dissociation (HCD) potentially allows for more control over the amount of energy deposited into ions.

Although similar methods, HCD and CID differ significantly in the mechanism by which energy is delivered to ions. In ion-trap CID, ions are resonantly excited to cause energetic collisions with helium gas. During the entire excitation period, ions are accelerated from one collision to another, eventually acquiring enough energy in small increments to fragment. In HCD, a beam of precursor ions is accelerated into a collision gas cell, allowing for a smaller number of collisions but at higher energy per collision relative to CID. Additionally, the timescales of energy transfer are quite different, with CID activation occurring over milliseconds while HCD takes only microseconds.[13] In

principle, these differences should allow precursors in HCD to be excited to higher energies than those attainable in ion trap CID.[14] In practice, HCD and CID spectra tend to contain many similar fragment ions but they are not identical, and the use of higher HCD energies leads to larger differences between the observed spectra.[15]

These fundamental mechanistic differences lead to changes in the relative abundances of fragments made as well as some differences in types of fragments made, and the differences between HCD and CID have been the subject of significant discussion. For example, CID spectra of peptides tend to contain mostly a, b, or y fragments and neutral losses, while HCD spectra contain these same fragments as well as more internal fragments and immonium ions.[16] In addition, higher HCD energies tend to reduce the abundance of larger fragments, presumably due to increased sequential fragmentation events that yield smaller fragments.[17] With large molecules such as peptides or proteins, more degrees of freedom necessitate higher energies for dissociation, and fragments often have sufficient internal energy for further degradation.[18] Since HCD provides more energy per collision than CID, the likelihood for sequential fragmentation is increased. Smaller fragment ions and singly-charged fragment ions make up a larger percentage of peptide HCD spectra and can represent the dominant fragments in some cases.[19] Large scale comparison of HCD and CID libraries revealed results consistent with previous observations.[20]

In this paper we explore the differences between RDD-CID and RDD-HCD, as well as track the changes in fragmentation observed as RDD-HCD energy is increased. Our results reveal that radical-based fragments are more favorably produced during RDD-CID

and low-energy RDD-HCD, while b/y fragments and smaller ions dominate higher-energy RDD-HCD. In addition, a stark contrast between the abundance of radical species and non-radical species is observed as HCD energy is increased, providing insight into the relative energy thresholds of RDD versus mobile proton-based dissociation.

Experimental

Materials. Organic solvents and reagents were purchased from Fisher Scientific, Sigma-Aldrich, or Acros Organics and used without further purification. FMOC-protected amino acids and Wang resins were purchased from Anaspec, Inc or Chem-Impex International. β-endorphin (YGGFMTSEKS QTPLVTLFKN AIIKNAYKKG E) was purchased from AnaSpec Inc. (Cat # 24319) and RRLIEDNEYTARG was purchased from Enzo (Cat # BML-P307-0001). AKAKTDHGAEIVYK was synthesized according to a modified solid-phase peptide synthesis protocol.[21]

Iodination and 4IB Modifications. β-endorphin was iodinated via reaction with NaI, Chloramine-T and Sodium metabisulfite in a manner to prevent excess iodination. Briefly, NaI and Chloramine T were combined in a 1:2 molar ratio prior to addition to β-endorphin. Following this, 1/3 molar equivalents of NaI:Chloramine-T were added to 20μL of 1mM β-endorphin in water and allowed to react for three minutes before addition of the next equivalent for a total of 1 molar equivalent of NaI at 9 minutes. The reaction was then quenched with 4x molar equivalents of sodium metabisulfite. 20μL of 1mM RRLIEDNEYTARG was iodinated by reaction of peptide, NaI, and Chloramine T at a 1:1:2 molar ratio for 10 minutes. At 10 minutes the reaction was quenched by

addition of 4x molar equivalents of sodium metabisulfite. AKAKTDHGAEIVYK was covalently modified with 4-iodobenzoic acid (4IB) via reaction with 4IB-N hydroxy succinimide (4IB-NHS). Briefly, 4IB-NHS was synthesized by reaction of 1:1:1 4IB:DCC:NHS (0.5mmol ea.) in 15mL dioxane for 12h under N_2. After 12h, the reaction precipitate was removed via filtration and dioxane was gently evaporated with N_2. Following this, covalent attachment of 4IB was achieved by reaction of 50µg AKAKTDHGAEIVYK in 25µL 100mM borate buffer (pH 8.5) with 25µL of 6.5mM 4IB-NHS (10-fold molar excess) in dioxanes for 1h. Iodo-RRLIEDNEYTARG and 4IB-AKAKTDHGAEIVYK were desalted on a MICHROM Bioresources peptide MicroTrap (P/N TR1/25109/02) directly following iodination to remove salts and reaction byproducts prior to MS analysis. Following iodination, iodo-β-endorphin was desalted on a MICHROM Bioresources protein MicroTrap (P/N TR1/25109/03).

Radical-Directed Dissociation Experiments. All experiments were performed on a Thermo Orbitrap Fusion Lumos. Peptides were introduced into the instrument via direct infusion using either a HESI source or a modified nano flex source from Thermo Scientific. The nano flex source was modified with a platinum wire to allow use of tips pulled from borosilicate glass (Harvard Apparatus GC100T-10). Peptide sprayed with the HESI source was diluted to 1µM in 50:50:0.1 H_2O:ACN:FA (v/v/v), while peptide sprayed with the nanoflex source was diluted to 1µM in water with 0.1% FA. Peptides were isolated using the quadrupole prior to either 213 nm or 266 nm photodissociation in the low pressure ion trap, after which the radical was re-isolated for either CID or HCD fragmentation and analysis in the Orbitrap mass analyzer. For RDD-CID, a single

normalized-collision-energy (referred to hereafter as CID energy) was selected at which the precursor was no longer observed to be the base peak with the exception of the results in Figure 1 where details are given. For RDD-HCD, normalized-collision-energies (HCD energy) were varied to produce changes in fragmentation.

Data Analysis. Following acquisition of data, deconvolution was performed in FreeStyle (v1.7) with Xtract with the analyzer type set to "OT," isotope table set to "protein," and the relative abundance threshold set to 1%. Fragment ions were assigned with a 0.01 Da tolerance. Following this, the fractional abundance of the deconvoluted data was calculated, where fractional abundance = intensity of fragment ion / sum of all fragment ion intensities. To aid in tracking the changes in fractional abundance as HCD energy is increased, % change was calculated using the first RDD-HCD energy as the initial point. % change was calculated as $[(FRAB_n – FRAB_0) / FRAB_0] * 100$. Sequence coverage plots were created using MASH (v2.2.0.33927). Side chain losses are common in RDD and are labeled by residue and approximate mass lost, i.e. -106Y indicates a 106 Da loss from the side chain of tyrosine.

Results and Discussion

RDD-CID vs RDD-HCD. To explore the effect of activation parameters on the relative abundances and fragment types observed in RDD-CID, experiments were performed on the 4+ charge state of iodo β-endorphin at various CID energies (Figure 5.1a-5.1c). The spectra are all quite similar, but a few minor differences can be noted. For example, the ions nearest to the precursor m/z are somewhat depleted at higher energies (i.e. -

NH_3/H_2O and -29I). This is likely due to leakage of the excitation waveform into these nearby m/z's. Other minor differences include a complementary pair of b_{12}/y_{19}-2 fragments, which are significantly more abundant at higher RDD-CID energies. To further quantify any changes in abundance as a function of CID energy, ratios of the relative intensities for a representative set of fragment pairs were calculated as shown in Fig. 5.1d. Constant ratios indicate similar relative ion abundances at all CID energies, which was the trend observed for most ions. Overall, the results illustrate that increasing the activation energy in ion-trap CID does not appreciably change the resulting fragmentation. This can be explained by considering that after dissociation, ions will no longer be resonantly excited but will instead undergo cooling collisions. Since the input of energy is slow and takes place over many small steps, it is not possible to raise the precursor ion energy significantly over the RDD thresholds. Although the collision energy step can be changed somewhat by altering the activation Q, changing this parameter also had little effect on the results (Fig. 5.S1).

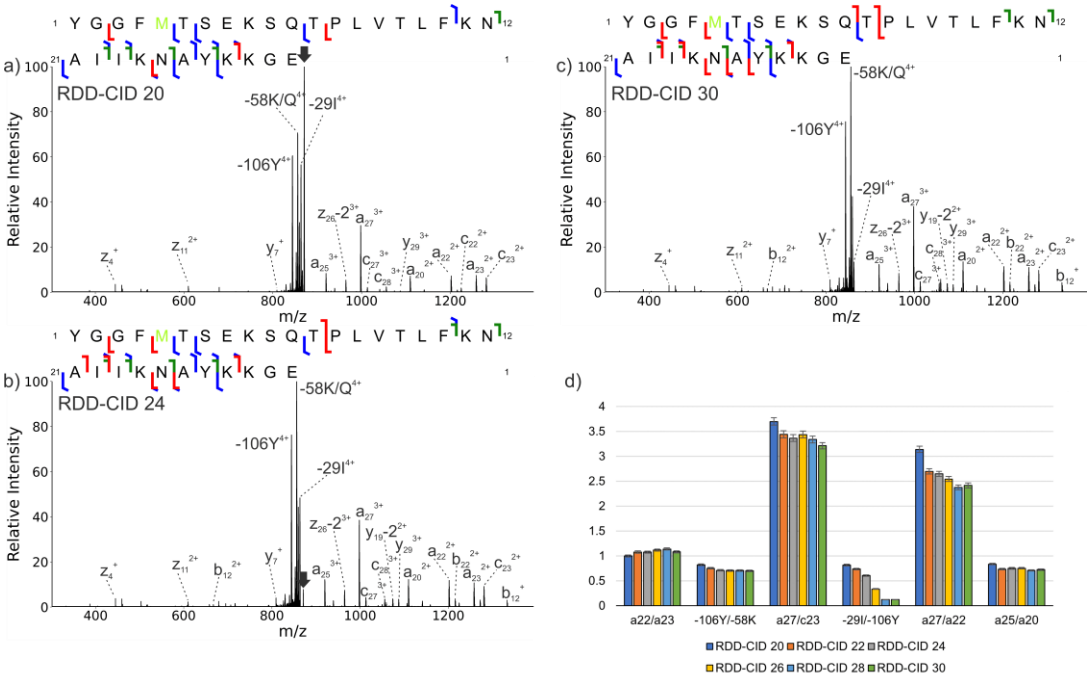


Figure 5.1: RDD-CID spectra and sequence ladders for β-endorphin at varying CID energies. (a-c) RDD-CID spectra for CID energies of 20, 24, and 30. d) ratios of intensities for selected fragment ion pairs a_{22}/a_{23}, -106Y/-58K, a_{27}/c_{23}, -29I/-106Y, a27/a22, and a25/a20. Error bars represent the standard deviation of the mean. The -29I/-106Y ratio exhibits the biggest change due to leakage of the excitation waveform fragmenting the -29I peak. In the sequence ladders, b/y fragments are shown in red, c/z in blue, and a fragments in green.

Results for an analogous series of RDD-HCD experiments are shown in Fig. 5.2. Even by casual observation, it is clear that the spectra in Fig. 2 change significantly as energy is increased. At the lowest HCD energy, the number, type, and abundances of fragment ions are similar (though not identical) to those obtained by RDD-CID (compare Fig. 5.1a and Fig. 5.2a). At higher energies, the number of fragment ions appears to increase, particularly in the lower m/z range while side chain losses are reduced in fractional abundance, particularly for the -106Y side chain loss (Fig. 5.2b-e). In contrast to RDD-CID, most of the fragment ion ratios tend to change as a function of energy in RDD-HCD as illustrated for several representative pairs in Fig. 5.2f. Since precursor ions are largely

accelerated prior to collisions in HCD, it is possible to access higher activation energies that facilitate alternative and/or sequential fragmentation pathways. Having noted some general trends, we now examine the effects of HCD energy on RDD product ions in greater detail.

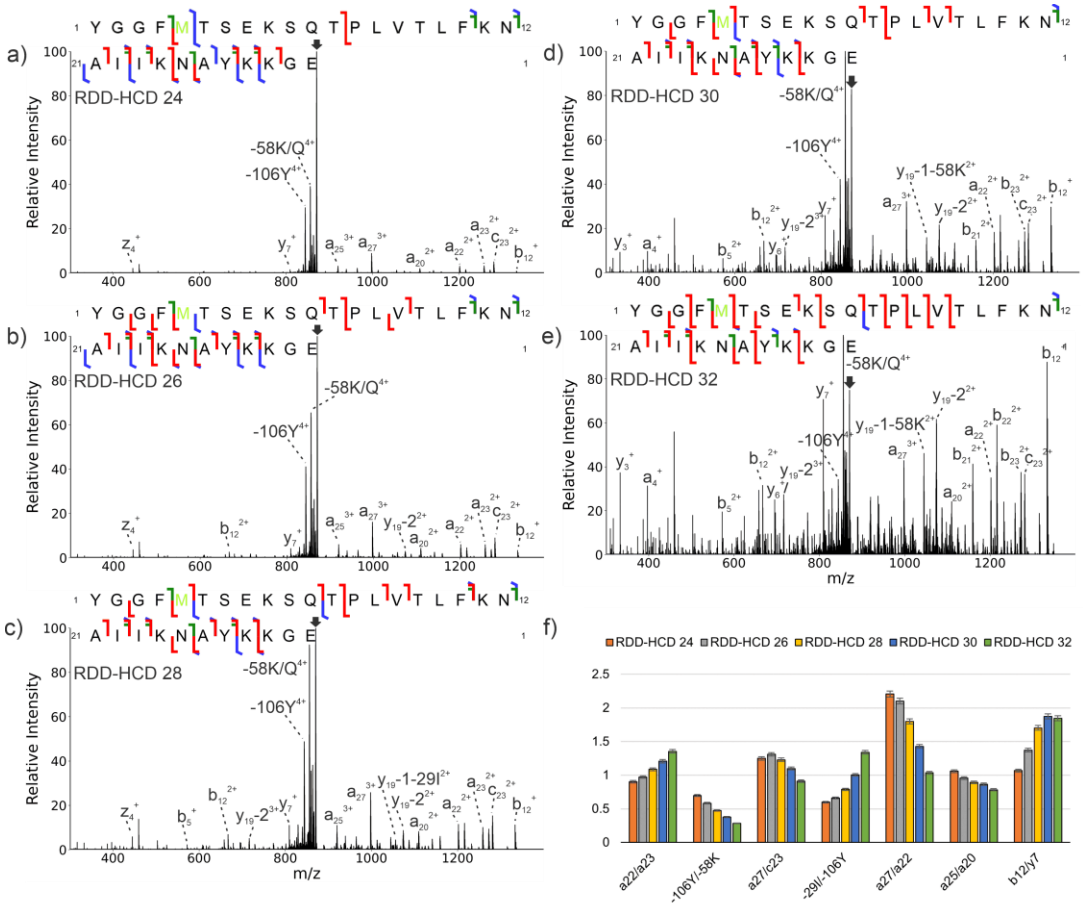


Figure 5.2: RDD-HCD spectra and sequence ladders for 4+ β-endorphin. a) RDD-HCD with a normalized collision energy (NCE) of 24, b) RDD-HCD, NCE 26, c) RDD-HCD, NCE 28, d) RDD-HCD, NCE 30, e) RDD-HCD, NCE 32, and f) ratio plots for a_{22}/a_{23}, -106Y/-58K, a_{27}/c_{23}, -29I/-106Y, a_{27}/a_{22}, a_{25}/a_{20}, b_{12}/y_7. Arrows indicate unfragmented precursor ion. In the sequence ladders, b/y fragments are shown in red, c/z in blue, and a fragments in green.

RDD-CID/HCD Neutral Loss Behavior with Increasing HCD Energy. Previous work has shown that for most peptides in the tryptic size regime, RDD produces side chain losses from the original precursor ion by three general mechanisms (called type I, II, and III).[2] To illustrate the behavior of these side chain losses as a function of activation energy, we plot the percent change in fractional abundance (FRAB) for several peptides in Fig. 5.3. The percent change is referenced to the lowest RDD-HCD energy (the second data point), with RDD-CID represented by the first data point. Only one RDD-CID data point is shown since RDD-CID spectra do not change appreciably with activation energy as shown previously in Fig. 5.1. Experiments on the 3+ charge state of 4IB-AKAKTDHGAEIVYK reveal a decrease in most side chain losses upon transition from RDD-CID to RDD-HCD (Fig. 5.3a). Subsequently increasing the HCD energy decreases the FRAB for all side chain losses, although the -106Y loss appears to be most labile. It is worth noting that the mechanism producing the -106Y loss leaves the radical species on the peptide while the -29I, -59E, and -45D losses leave behind even-electron peptides. These results suggest that radical species may more easily undergo subsequent fragmentation by HCD. Results for the 4+ charge state of β-endorphin are shown in Fig. 5.3b. Interestingly for this peptide, all side chain losses are more abundant in low energy HCD versus CID. At higher HCD energies, all of these products decrease in abundance, but again those losses that produce radical peptides (-71K, -106Y, and $-NH_3$) decrease more quickly. Very similar results are obtained for the 5+ charge state of β-endorphin (Fig. 5.S2), which may suggest that charge state does not play a significant role. The peptide fragments remaining behind after small molecule side chain losses are nearly the

same size as the initial peptide precursors, which should increase the probability for energetic secondary collisions. Such large species will also have nearly the same number of low energy dissociation pathways still available as the original precursor ion, therefore it is not surprising that further decay is observed at higher HCD energies. Similarly to the previous two examples, RDD-HCD experiments on RRLIEDNEYTARG result in decreases in FRAB for nearly all side chain loss products, with type III side chain -106Y exhibiting the quickest decrease in FRAB relative to the other side chain losses (Fig. 5.3c). Interestingly, the -87R loss was found to increase substantially with HCD energy and only begins to decrease in FRAB at much higher HCD energies. This behavior is quite opposite to the trend observed for other side chain losses and may be related to the fact that -87R is not a neutral loss since it contains a proton on the Arg side chain and leads to a reduction of charge state.

To further explore the stability of radical versus non-radical product ions, the peptides produced by side chain losses were re-isolated and subjected to identical CID activation in MS^4 experiments. -106Y and -71K exhibited a higher degree of fragmentation than the non-radical -58K loss (Fig. 5.S3). The FRAB of the intact -106Y, -71K, and -58K precursors were 0.18, 0.27, and 0.38, respectively. The higher residual precursor for the non-radical -58K loss is consistent with higher energy dissociation barriers for this fragment.

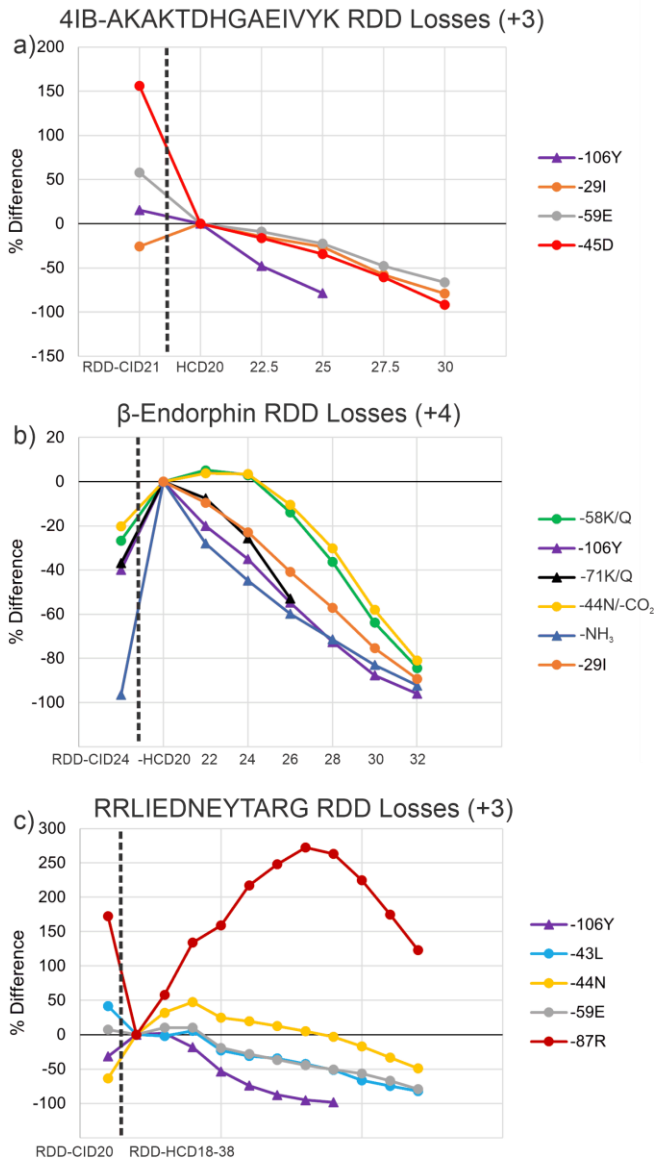


Figure 5.3: RDD-CID/HCD % change plots for the neutral loss fractional abundances for a) the 3+ charge state of 4IB-AKAKTDHGAEIVYK, b) the 4+ charge state of iodo- β-endorphin, and c) the 3+ charge state of RRLIEDNEYTARG. The first point on the plots is the initial RDD-CID point, while the subsequent points are RDD-HCD fractional abundances with the HCD energy increased at regular intervals. The dashed line separates the RDD-CID point from the RDD-HCD points. Radical species are represented with triangles at each data point, non-radical species are represented with circular points.

Stability of a-ions. RDD favors the production of a-ions at aromatic residues and Ser/Thr due to facile migration to the beta position.[2] The stability of a-ion products as a function of activation energy is shown in Fig. 5.4. For 4IB-AKAKTDHGAEIVYK in the 3+ charge state, a range of behaviors is observed, including ions that increase or decrease in abundance as a function of HCD energy or between HCD and CID (Fig. 5.4a). Closer inspection reveals that FRAB correlates well with ion length where longer a-ions lose FRAB at higher energies. In contrast, smaller ions tend to increase with higher energy, suggesting that they derive from secondary fragmentation of larger fragments. Notably, the a_4+1 fragment behaves in completely the opposite fashion, where this shortest fragment decreases in FRAB at higher energies. Importantly, the mechanism that generates this a_n+1 is known and is specific to sites N-terminal to Ser/Thr residues.[2] Although typical a-ions are non-radical species, the a+1 is a radical (see Scheme 5.S1c,d). Despite being the shortest fragment (and therefore less prone to additional collisions or secondary fragmentation), the radical nature of the a_4+1 ion must account for its fragility and reduced FRAB as HCD energy is increased. This interpretation is further confirmed upon consideration of the a-ions generated for 4+ β-endorphin, which includes a typical a_4 ion (Fig. 5.4b). For β-endorphin, all longer a-ions decrease in FRAB as HCD energy rises while the a_4 ion, which does not exist at lower energies, rises dramatically at higher HCD energy. RDD experiments on RRLIEDNEYTARG also confirm these results (Fig. 5.4c). A radical a_9+1 fragment formed at Thr-10 was observed during RDD-CID and low energy RDD-HCD but decayed in FRAB faster than any of the other a-ions as HCD energy was increased, including the corresponding non-radical a_9 fragment.

Interestingly, a_7 and a_8 both increase steadily in fractional abundance with increasing HCD energy. The behavior of these longer fragments with increasing HCD energy could be explained by degradation of the a_9 and a_9+1 fragments which may contribute to production of the a_7 and a_8 fragments. A similar pattern is observed for a_4, a_5, and a_6. The a_6 ion decreases slowly in FRAB as HCD energy is increased, while a_4 and a_5 increase moderately in FRAB. This increase in FRAB for a_4 and a_5 may be due to degradation of a_6 to a_4 and a_5, in a similar fashion to a_9.

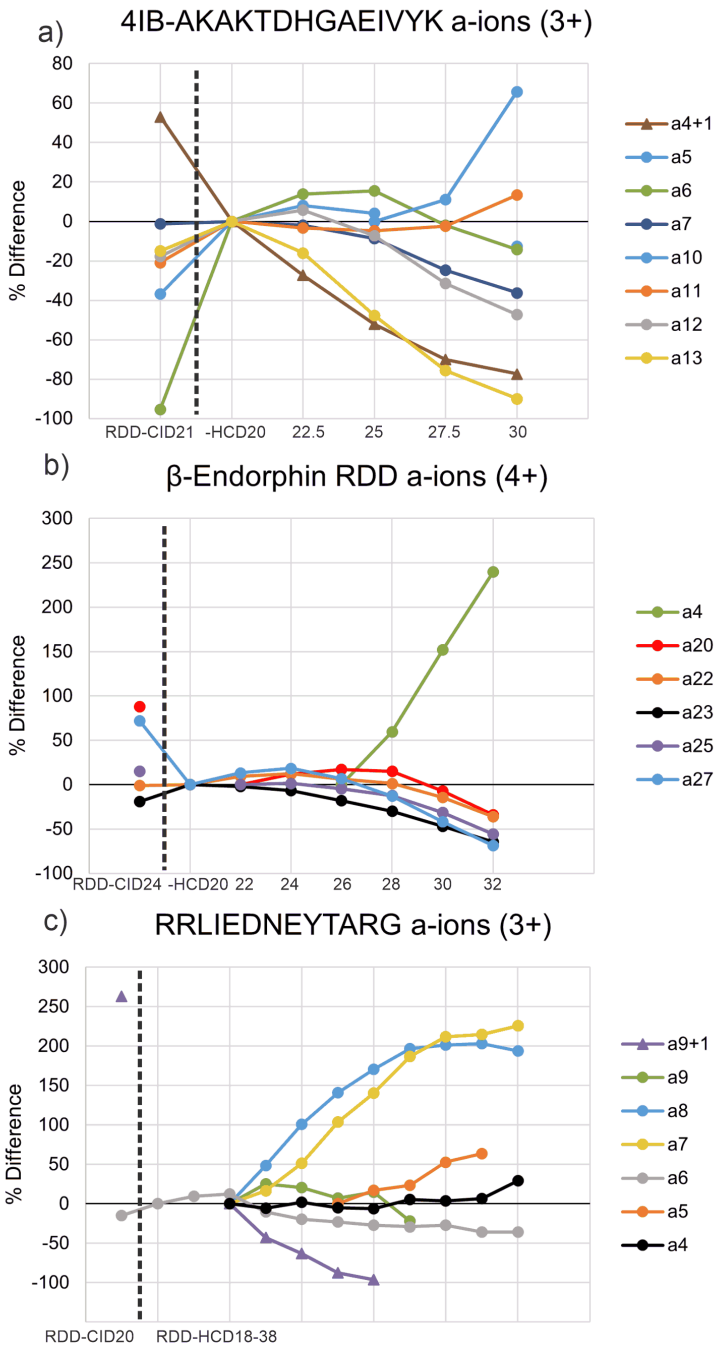


Figure 5.4: RDD-CID/HCD % change plots for a-ions. a) 3+ charge state of 4IB-AKAKTDHGAEIVYK, b) 4+ charge state of β-endorphin, c) 3+ charge state of RRLIEDNEYTARG. Radical a+1-species are represented with triangles at each data point, non-radical a-species are represented with circular points.

Further consideration of fragment size. The decay of larger fragments observed from the neutral losses in Figure 1 shows that there is a potential inverse relationship between fragment size and HCD energy. In order to explore this further, fragment length versus the sum of fractional abundances for all fragments of the same length are plotted for 4IB-AKAKTDHGAEIVYK in Figure 5.5a at three different HCD energies. At the lowest HCD energy, longer fragments that are 13 or 14 amino acids in length are present in higher FRAB, while these same fragments decrease in FRAB as HCD energy increases. Smaller fragments such as those 2, 3, 5, 6, or 7 amino acids in length tend to increase in FRAB with increasing HCD energy. This illustrates that higher HCD energy tends to produce smaller ions at the cost of less abundant larger ions. Exceptions to this trend, such as for fragments 4 amino acids in length, can be rationalized by radical fragility (i.e. the a_4+1 radical is a major contributor to this datapoint). For β-endorphin, fragments of length 2-21 are produced in higher FRAB at higher HCD energies while fragments of length 22-31 are lower in FRAB at higher HCD energies (Fig. 5.5b). It's clear that as HCD energy is increased, smaller fragments begin to dominate the spectrum while larger ones decrease. Examining the types of fragments observed under each experiment also reveals more about what is happening as HCD energy is increased. In Figure 5.5c, low energy RDD-HCD produces a few c/z and a fragments, and most of these fragments are long. With additional activation, more b/y, c/z, and a-ions are made, in addition to shorter fragments. Finally, at even higher energy RDD-HCD, b/y and shorter fragments begin to dominate. These transitions are consistent with diminution of radical fragments and sequential truncation of longer fragments.

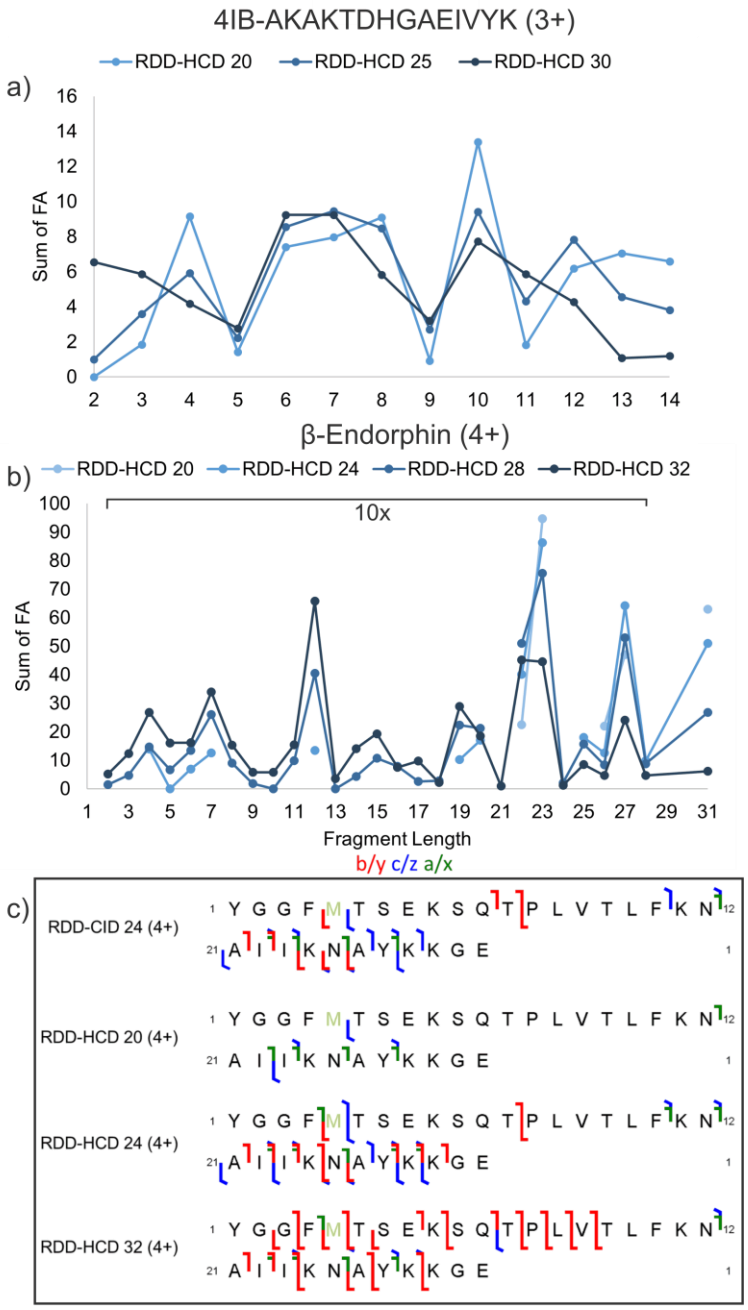


Figure 5.5: a) fractional abundance vs fragment length for 4IB-AKAKTDHGAEIVYK. b) Fractional abundance vs fragment length for β-endorphin (4+), with the FRAB for fragments 2-28 amino acids in length magnified 10x. c) Fragment type and sequence coverage for β-endorphin (4+) for RDD-CID24, RDD-HCD20, RDD-HCD24, and RDD-HCD32. b/y fragmentation is shown in red, c/z in blue, and a/x in green.

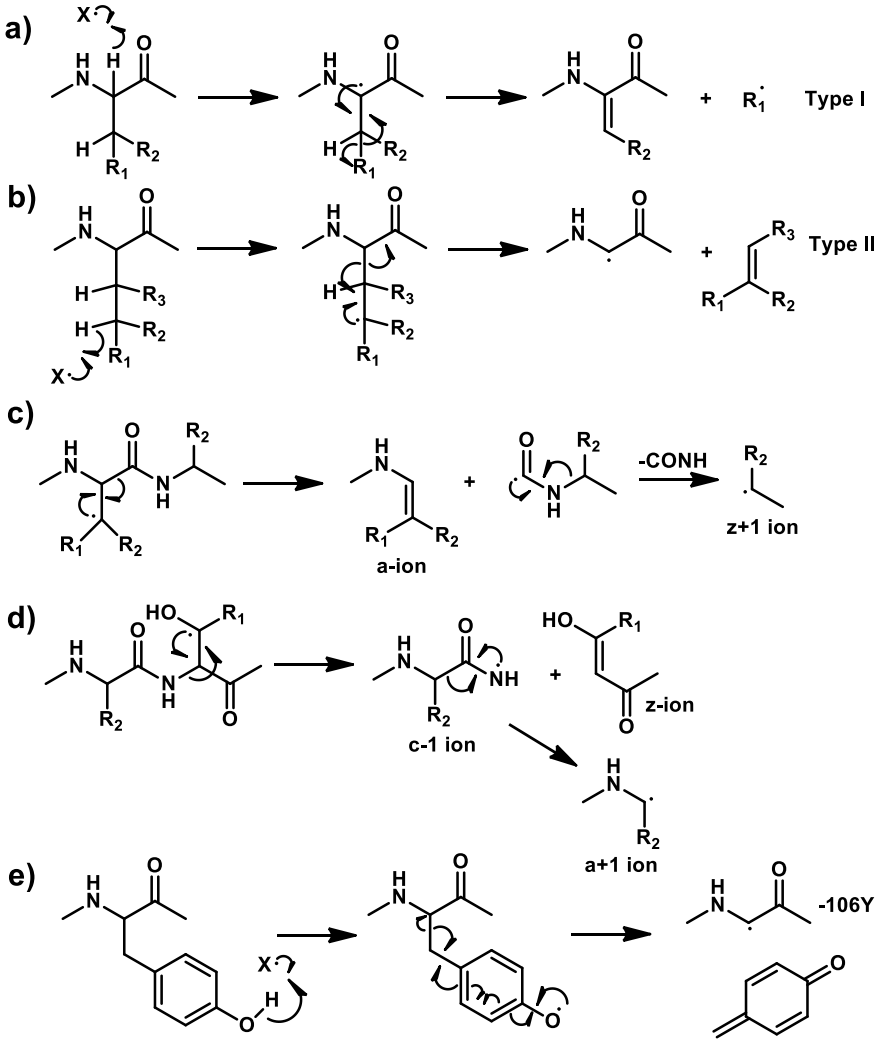
Conclusion

Our results illustrate differences in both the stability of radical versus non-radical fragment ions and differences in the mechanisms leading to fragmentation in CID versus HCD. The resonant excitation used in ion-trap CID does not afford easy access to activation energies exceeding the lowest energy dissociation pathways, meaning that RDD-CID experiments do not vary as a function of activation once the dissociation threshold is achieved. However, the radical peptide fragments produced by RDD are significantly less stable relative to the canonically protonated species (that are also products of RDD) and will easily undergo secondary fragmentation with high energy activation in HCD experiments. The dominant b/y ions observed at higher HCD energies therefore represent 'survivor' ions that have undergone secondary dissociation and have few low-energy dissociation pathways remaining available. Ironically, HCD also appears to enable lower energy activation than is possible with CID, as is most apparent in the second line of Fig. 5.5c. This allows tuning of fragmentation based on HCD energy to obtain spectra with either more or less of a specific fragment type, ranging from radical-dominated products to radical-less products. RDD-HCD therefore appears to be a versatile option to be included in the collection of MS^n methods.

Acknowledgements

The authors would like to thank John Syka, Chris Mullen, and Josh Hinkle from Thermo Fisher Scientific for valuable discussions and assistance with instrument modifications, as well as the NSF for funding (CHE-1904577), and Gaurav Pandey for peptide synthesis.

Supporting Information



Scheme 5.S1: Selected RDD mechanisms for (a) type I side chain losses, (b) type II side chain losses, (c) a and z+1 ion formation during RDD, (d) a+1 and z ion formation at Ser/Thr during RDD, R_1=H/CH_3, and (e) side chain loss from Tyrosine.

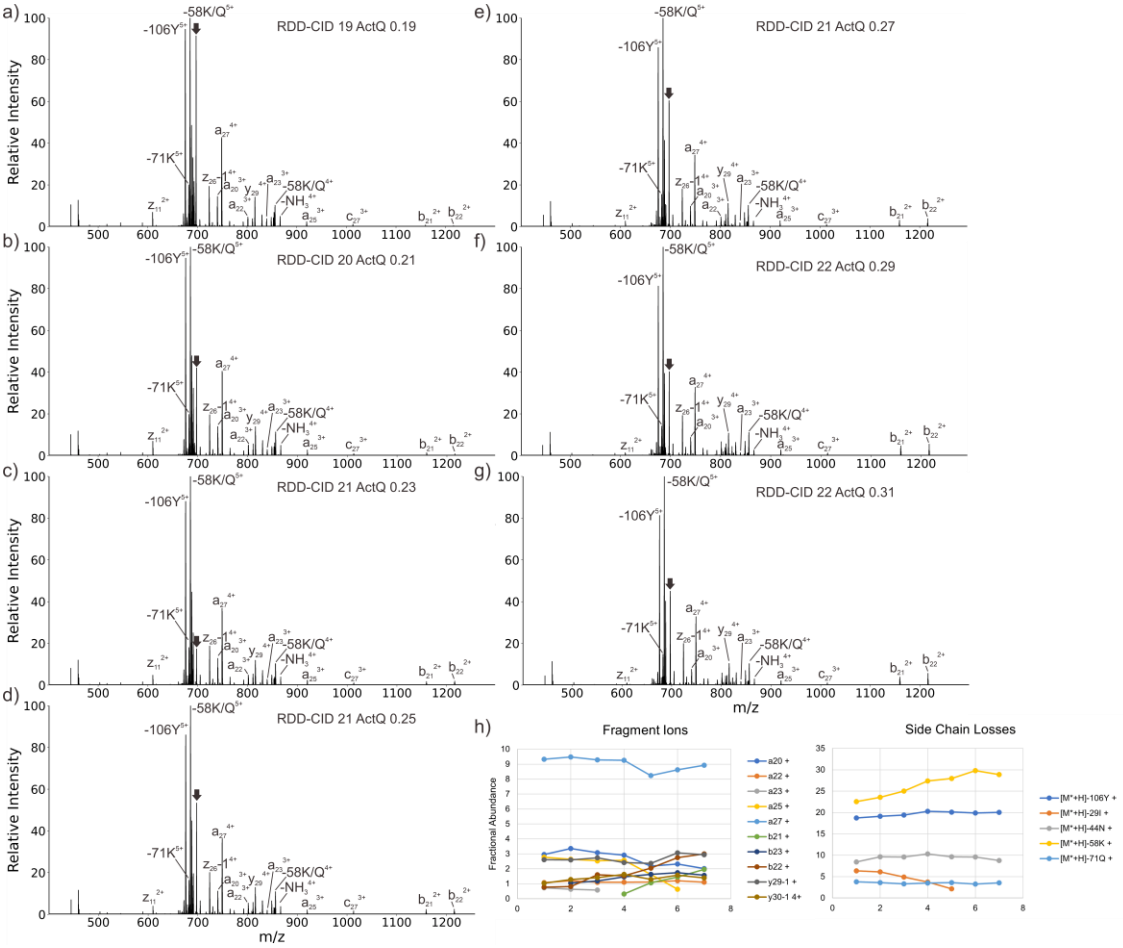


Figure 5.S1: Results of varying RDD-CID energy and Act Q values. Major peaks are labeled and precursor is indicated by the black arrow. a) RDD-CID 19 Act Q 0.19, b) RDD-CID 20 Act Q 0.21, c) RDD-CID 21 Act Q 0.23, d) RDD-CID 21 Act Q 0.25, e) RDD-CID 21 Act Q 0.27, f) RDD-CID 22 Act Q 0.29, g) RDD-CID 22 Act Q 0.31, h) fractional abundance plots versus Act Q values.

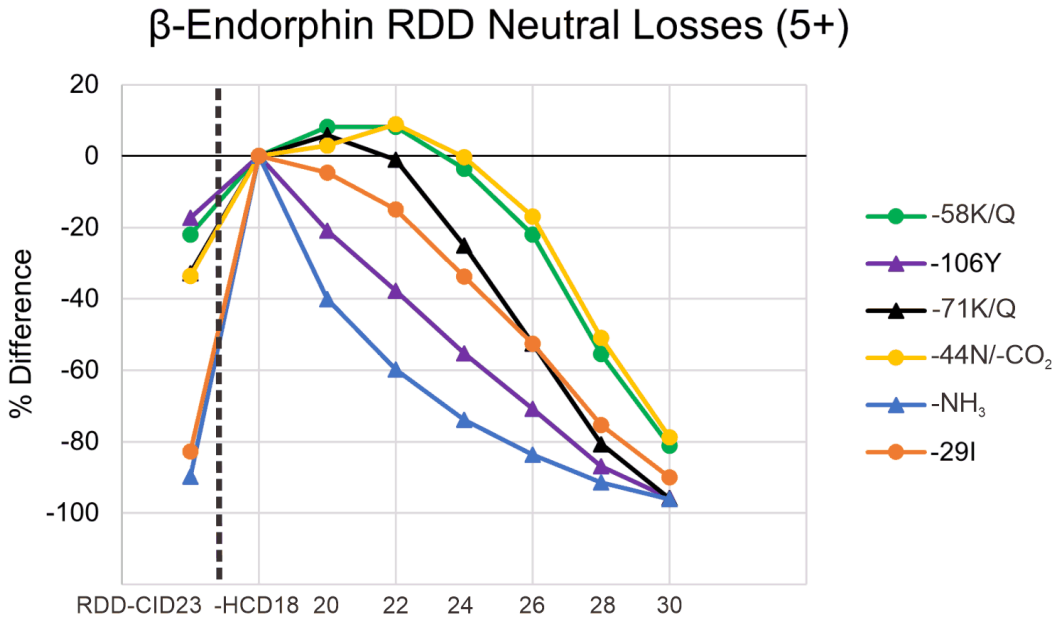


Figure 5.S2: % Difference plot for neutral losses from β-endorphin (5+). Losses in which the radical remains behind on the peptide are denoted by the triangle points, while losses in which the radical has left the peptide with the neutral fragment are denoted with circular points.

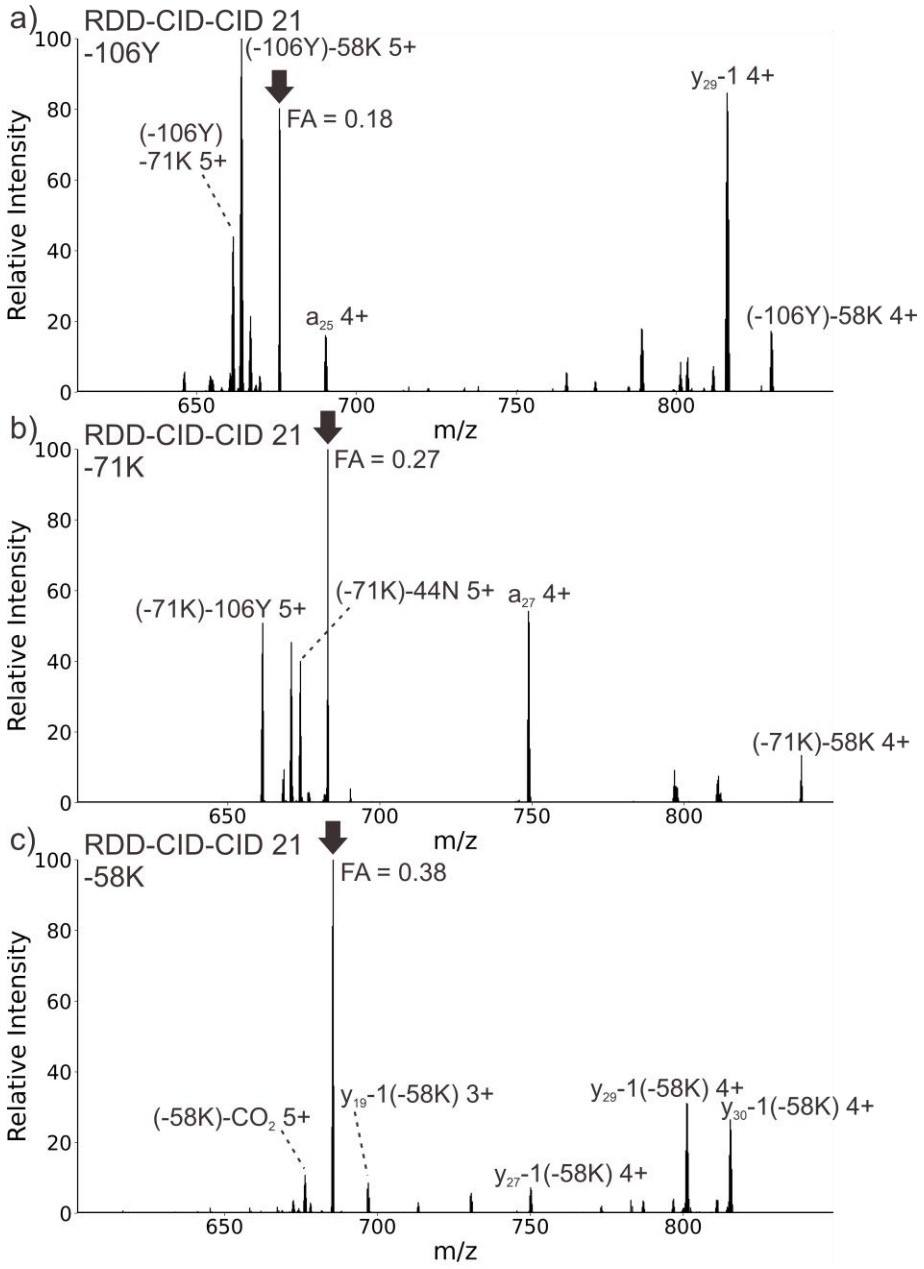


Figure 5.S3: PD-CID-CID (MS4) on a) the RDD-CID -106Y side chain loss from β-endorphin, b) the -71K side chain loss from β-endorphin, and c) the RDD-CID -58K side chain loss from β-endorphin. Fractional abundance (FA) of the remaining precursor is indicated.

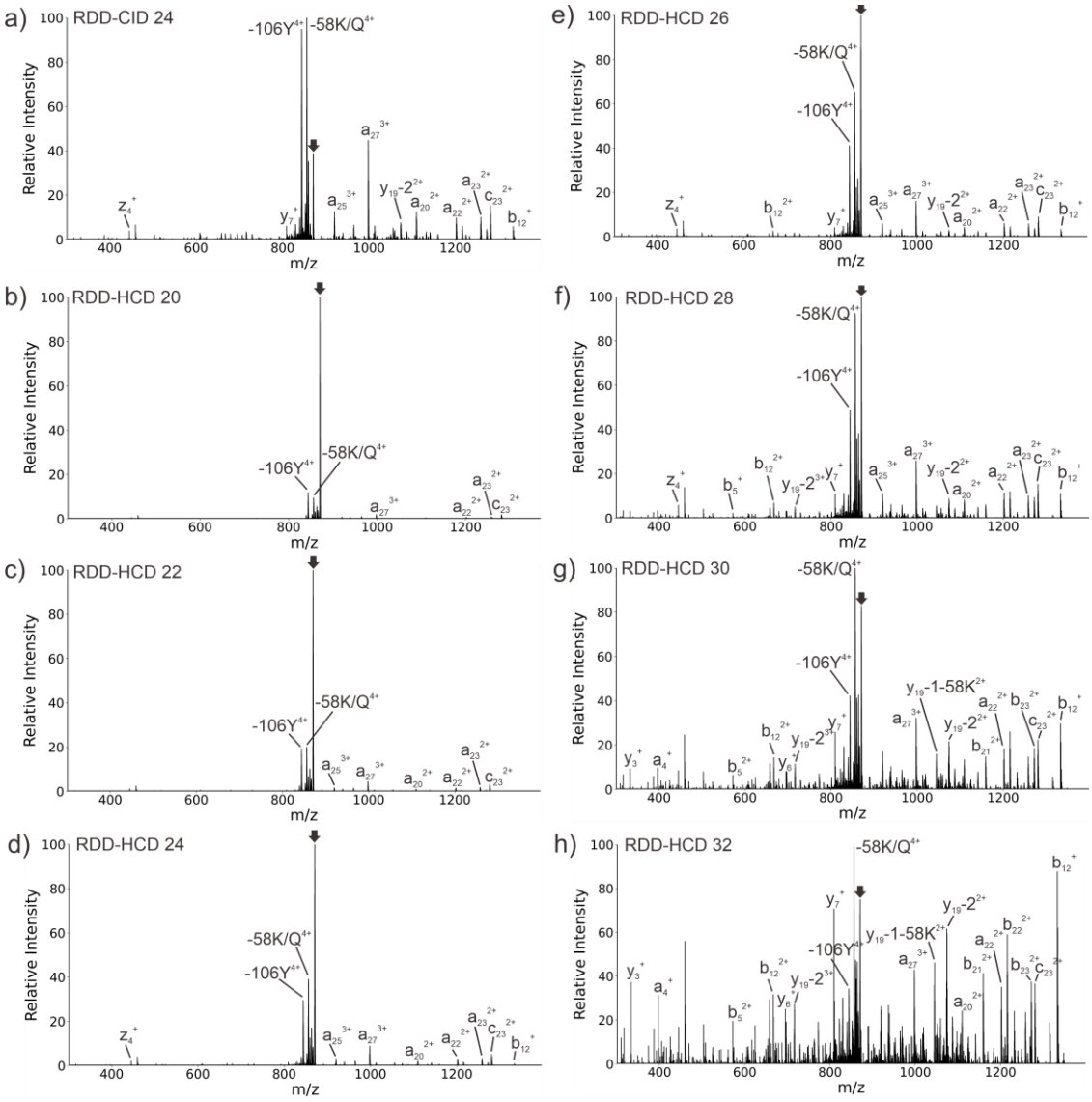


Figure 5.S4: β-endorphin mass spectra for RDD-CID and RDD-HCD at all selected HCD NCE. Major peaks are labeled, black arrow denotes precursor.

References

[1] Tao, Y.; Quebbemann, N. R.; Julian, R. R. Discriminating D-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (15), 6814–6820.
[2] 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.
[3] Lee, M.; Kang, M.; Moon, B.; Oh, H. Bin. Gas-Phase Peptide Sequencing by TEMPO-Mediated Radical Generation. *Analyst* **2009**, *134* (8), 1706.
[4] Pham, H. T.; Ly, T.; Trevitt, A. J.; Mitchell, T. W.; Blanksby, S. J. Differentiation of Complex Lipid Isomers by Radical-Directed Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (17), 7525–7532.
[5] Narreddula, V. R.; Boase, N. R.; Ailuri, R.; Marshall, D. L.; Poad, B. L. J.; Kelso, M. J.; Trevitt, A. J.; Mitchell, T. W.; Blanksby, S. J. Introduction of a Fixed-Charge, Photolabile Derivative for Enhanced Structural Elucidation of Fatty Acids. *Anal. Chem.* **2019**, *91* (15), 9901–9909.
[6] Pham, H. T.; Julian, R. R. Radical Delivery and Fragmentation for Structural Analysis of Glycerophospholipids. *Int. J. Mass Spectrom.* **2014**, *370*, 58–65.
[7] Zhao, X.; Xia, Y. Characterization of Fatty Acyl Modifications in Phosphatidylcholines and Lysophosphatidylcholines via Radical-Directed Dissociation. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (2), 560–568.
[8] Zhao, X.; Zhang, W.; Zhang, D.; Liu, X.; Cao, W.; Chen, Q.; Ouyang, Z.; Xia, Y. A Lipidomic Workflow Capable of Resolving Sn - and C-C Location Isomers of Phosphatidylcholines. *Chem. Sci.* **2019**, *10* (46), 10740–10748.
[9] 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.
[10] Fabijanczuk, K.; Gaspar, K.; Desai, N.; Lee, J.; Thomas, D. A.; Beauchamp, J. L.; Gao, J. Resin and Magnetic Nanoparticle-Based Free Radical Probes for Glycan Capture, Isolation, and Structural Characterization. *Anal. Chem.* **2019**, *91* (24), 15387–15396.
[11] Pham, H. T.; Julian, R. R. Characterization of Glycosphingolipid Epimers by Radical-Directed Dissociation Mass Spectrometry. *Analyst* **2016**, *141* (4), 1273–1278.
[12] Hamdy, O. M.; Alizadeh, A.; Julian, R. R. The Innate Capacity of Proteins to Protect against Reactive Radical Species. *Analyst* **2015**, *140* (15), 5023–5028.
[13] de Graaf, E. L.; Altelaar, A. F. M.; van Breukelen, B.; Mohammed, S.; Heck, A. J. R. Improving SRM Assay Development: A Global Comparison between Triple Quadrupole, Ion Trap, and Higher Energy CID Peptide Fragmentation Spectra. *J. Proteome Res.* **2011**, *10* (9), 4334–4341.
[14] Xia, Y.; Liang, X.; McLuckey, S. A. Ion Trap versus Low-Energy Beam-Type Collision-Induced Dissociation of Protonated Ubiquitin Ions. *Anal. Chem.* **2006**, *78* (4), 1218–1227.
[15] Diedrich, J. K.; Pinto, A. F. M.; Yates, J. R. Energy Dependence of HCD on Peptide Fragmentation: Stepped Collisional Energy Finds the Sweet Spot. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (11), 1690–1699.
[16] Michalski, A.; Neuhauser, N.; Cox, J.; Mann, M. A Systematic Investigation into the Nature of Tryptic HCD Spectra. *J. Proteome Res.* **2012**, *11* (11), 5479–5491.
[17] Samgina, T. Y.; Vorontsov, E. A.; Gorshkov, V. A.; Artemenko, K. A.; Zubarev, R. A.; Lebedev, A. T. Mass Spectrometric de Novo Sequencing of Natural Non-Tryptic Peptides: Comparing Peculiarities of Collision-Induced Dissociation (CID) and High Energy Collision Dissociation (HCD). *Rapid Commun. Mass Spectrom.* **2014**, *28* (23), 2595–2604.
[18] Marzluff E. M. and Beauchamp J. L.: Collisional Activation Studies of Large Molecules. In: Baer, T., Ng, C. Y., and Powis, I. (eds.) Large Ions: Their Vaporization, Detection and Structural Analysis. pp. 115-141. John Wiley & Sons, Chichester, UK (1996).
[19] Shao, C.; Zhang, Y.; Sun, W. Statistical Characterization of HCD Fragmentation Patterns of Tryptic Peptides on an LTQ Orbitrap Velos Mass Spectrometer. *J. Proteomics* **2014**, *109*, 26–37.
[20] Wilburn, D. B.; Richards, A. L.; Swaney, D. L.; Searle, B. C. CIDer: A Statistical Framework for Interpreting Differences in CID and HCD Fragmentation. *J. Proteome Res.* **2021**, *20* (4), 1951–1965.

[21] Hood, C. A.; Fuentes, G.; Patel, H.; Page, K.; Menakuru, M.; Park, J. H. Fast Conventional Fmoc Solid-Phase Peptide Synthesis With HCTU. *J. Pept. Sci.* **2008**, *14*, 97– 101.

CHAPTER 6: Concluding Remarks

Protein structure and function is easily perturbed by seemingly small changes in structure such as isomerization and can lead to drastic overall effects such as loss of function or aggregation. The study of isomerization is becoming increasingly important as the relevance of isomerization-prone LLPs in the context of human diseases is clearly established, and discovery of the underlying causes of complex human diseases such as AD requires that no stone be left unturned. In an effort to completely overturn the stone of isomerization in LLPs and its' potentially complex link to aging and human disease, it is important to develop the proper analytical techniques to study these modifications so that the intricate relationship between age-related isomerization, protein structure, and disease pathology can be further elucidated.

We started by establishing that mass spectrometry is capable of distinguishing cis/trans isomers in proline, structural motifs that are very fragile and prone to spontaneous. It was established that under the specific parameters of gentle ESI conditions and solvation by 18C6, differences in 266nm PD fragmentation patterns could be observed that differentiate cis/trans proline isomers. Overall, the results showed that mass spectrometry is capable of distinguishing not only cis/trans structural differences via 266nm PD, but also that through careful selection of ESI conditions and consideration of charge solvation, fragile conformations can be preserved from solution into the gas

phase for analysis. This highlights the strength of mass spectrometry for the study of protein structure and sets the stage for the study of isomerization.

Having established that differences in MS^n fragmentation patterns allows the differentiation of fragile structural differences, we turned to developing an L-isoAsp-specific covalent modification technique that could be used to rapidly screen large numbers of peptides for isomerization. Exploitation of the reactive 5-membered succinimide ring present during PIMT repair of L-isoAsp allowed a mass shift to be introduced that greatly reduced the difficulty of isomer discovery. This method was then validated on a 72-year old crystalline lens digest, where 8 out of 10 known LLPs were labeled with tris. In complex samples with hundreds of proteins and thousands of peptides, introducing a mass shift that can quickly and effectively identify isomerization and thus identify LLPs greatly reduces the amount of time needed to discover new LLPs and isomerization sites.

Next, we examined the consequences of isomerization on proteolysis, specifically trypsin, chymotrypsin, and the 20S proteasome. Comparison of sequences containing Asp isomers in proximity to trypsin cleavage sites revealed that proteolysis was impaired in all cases, but proteolysis was nearly completely inhibited when isomerization was present at the P1' position in trypsin substrates. Furthermore, isomerization was found to affect chymotrypsin even more than trypsin. In a substrate with multiple chymotrypsin cleavage sites, chymotrypsin would shift to cleavage sites further away from the point of introduced isomerization, leaving cleavage sites close to the isomerized residue untouched. For the 20S proteasome, very little cleavage of isomerized substrate was

observed versus cleavage of the all-L substrate. Since the lysosome is also known to be unable to handle isomerized substrates, the question remains what happens to isomerized LLPs and the undigested fragments that inevitably get left behind by the proteasome or lysosome.

Finally, a new MS^n method was explored for the study of isomerization and protein structure. In RDD, CID is typically used to cause radical migration and fragmentation, but in chapter 5 we study in detail the effects of using HCD at varying energies to excite the radical. It was discovered that lower energy HCD favored radical-based fragmentation, while higher HCD energies decreased the amount of radical-based fragmentation and more mobile proton-based fragmentation occurred, presumably due to the higher energy barrier required for mobile proton fragmentation being overcome at higher HCD energy. In addition, the type and extent of fragmentation changed significantly with HCD energy, effectively allowing the ions observed to be tuned by adjusting the HCD energy. RDD-HCD provides another powerful tool for analysis of peptide and protein structure.

Mass spectrometry is the preeminent method for studying proteins, yet one application that has remained on the fringes is the study of isomerization. Isomerization has gradually been growing as a sub-field of protein study, but there are significant gaps in knowledge. For example, if the lysosome and proteasome are both unable to degrade isomerized substrates, then what happens to these isomerized proteins and what is the final biological result? Furthermore, although we have a good grasp on isolated consequences of isomerization on protein structure and function in specific proteins, what is missing is an

overall understanding of exactly how isomerization fits into the bigger picture of human aging. To what extent is isomerization a driving force in human age-related pathology? Is it more a symptom of pathology, or a cause? All of these questions require a more thorough examination of isomerization through the lens of high-throughput proteomics. At this point in the field, MS-based techniques for the study of isomerization are emerging that should lower the difficulty barrier for wide-scale study of isomerization. Yet, it seems that it will take time for isomerization to enter the field's consciousness and be studied as thoroughly as modifications such as phosphorylation, as isomerization is still ignored in many large-scale proteomic studies. This effectively creates a blind spot in the literature, but with such a blind spot the field is ripe for new discoveries. Mass spectrometry is one of the best tools available to study isomerization and how it fits into the big picture of human disease.

Overall, the body of work in this thesis shows significant advances in techniques to identify isomerization, as well as advances in the understanding of the complex interactions between isomerization and proteolysis. As isomerization becomes more and more relevant in the study of human aging and disease, new mass spectrometry-based techniques will be needed for the discovery and study of isomers. This work provides a foundation of new techniques to further unravel the consequences of isomerization in proteins.