UNIVERSITY OF CALIFORNIA
RIVERSIDE

Utilizing Mass Spectrometry to Reveal the Behavior of Lysosomal Enzymes and the Consequences of Protein Structure

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

Doctor of Philosophy

in

Chemistry

by

Tyler R. Lambeth

September 2021

Dissertation Committee:
Dr. Ryan Julian, Chairperson
Dr. Joseph Genereux
Dr. Haofei Zhang

The Dissertation of Tyler R. Lambeth is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would like to thank the individuals who helped me develop along my path and guided me towards my doctorate. Thank you to Tiffany Pawluck, who answered a nearly endless number of questions I had when first discovering my passion for chemistry. I thank Dr. David Saiki for always being a welcoming presence I could ask for help. I thank Dr. Andreas Gebauer for the opportunity to do research and get real hands-on experience working in a lab, and for his guidance in my transition to graduate school. Thank you to Dr. Haofei Zhang for being a part of my committee and providing insight during my exams. Thank you to Dr. Joseph Genereux for his continuing encouragement and advice from my first classes through my committee exams. And thank you to my parents, Michelle and Robert Lambeth, for always encouraging me to reach high and providing an unwavering foundation of support and stability.

Thank you to all of the friends I have made during my time at UC Riverside including Ryland Forsythe, Jan Phillip Scheifers, Nicole Perkins, Priyanka Sarkar, Ryan Stanton, Tabby Stanton, and Jessica Rodriguez among many others. I thank Dave Slocumb, who has been a trustworthy source of support and friendship throughout my journey. I thank the remarkable scientists I've had a chance to work with in my time in the Julian lab: Dylan Riggs, James Bonner, Lance Talbert, Yana Lyon, Guy Quanrud, Jacob Silzel, Hoi Ting Wu, Evan Hubbard, Brielle Van Orman, and Gaurav Pandey. The advice, support, and camaraderie I've shared with my friends and colleagues have been fundamental to my success during the last five years.

Finally, thank you to my advisor Dr. Ryan Julian for allowing me to work in his lab and for his mentorship in navigating me to success. The high standard upheld in the lab was demanding at times, but his sincere instruction, advice, and direction helped me develop into a determined critical thinker and problem solver. He has helped me grow not only as a scientist but as a speaker and a writer as well, and I am sincerely thankful for his guidance.

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

Chapter 2: 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.

Chapter 3: Lambeth, T.R.; Julian, R.R. Lysosomal proteolysis of amyloid beta is impeded by fibrils grown in both acidic and neutral pH environments. *bioRxiv* doi: 10.1101/2021.06.18.449015.

Chapter 4: Lambeth, T.R.; Julian, R.R. A two-trick pony: lysosomal protease cathepsin B possesses surprising ligase activity. *RCS Chem. Biol.* **2021**, *2(2)*, 606-611.

ABSTRACT OF THE DISSERTATION

Utilizing Mass Spectrometry to Reveal the Behavior of Lysosomal Enzymes and the Consequences of Protein Structure

by

Tyler R. Lambeth

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, September 2021
Dr. Ryan R. Julian, Chairperson

The biochemical processes responsible for proper cellular operation are wholly dependent on interactions between biomolecules constructed with canonical yet alterable structures. Proteins represent a unique class of biomolecules with multiple levels of structure that can be spontaneously or enzymatically modified, resulting in unintended consequences. Spontaneous deamidation of asparagine and isomerization of aspartic acid into four products (L-Asp, D-Asp, L-isoAsp, D-isoAsp), and epimerization of L-serine into D-serine, are prevalent modifications that alter primary structure. Additionally, modifications can occur in higher order structure as observed in the infamous fibrilization of amyloid beta, resulting in beta sheet structures which further associate into fibrils. One cellular aspect of particular interest is the lysosome-dependent autophagy system. As the protein degradation process that occurs in the lysosome requires binding between cathepsin proteases and protein substrates, this system is intimately affected by the underlying structures To understand the behavior of these enzymes and their relation to

substrate structure, we utilized mass spectrometry to identify products and characterize the effect of modifications on normal protease operation. Initial experiments revealed the inhibition of digestion by both isomerization and higher order modifications in amyloid beta fibrilization. Subsequent studies employing fluorescence microscopy affirmed the consequence of structural alterations on the lysosome in a quantitative manner when amino acid isomers were introduced into live cells. Importantly, digestion was found to be inhibited by a sequence containing D-isoAsp residues which prevented cathepsin binding. During these experiments, another enzymatic behavior was revealed in cathepsin B, where a previously unknown structural ligation reaction occurred between peptides. We performed further mass spectrometry experiments to characterize this reaction and determined that the ligation reaction occurs as a reversal of the proteolytic action by attaching a dipeptide to the C-terminus. Additionally, it was found that ligation could operate via an L-isoAsp interaction that could attach peptides of any length together. Finally, to expand our toolkit for examination of protein structure with mass spectrometry, we characterized the use of a commercially available iodophenyl molecule for facile attachment of a radical precursor using isothiocyanate chemistry. Photodissociation of the iodine bond generated a radical for structurally sensitive radical-directed dissociation experiments.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Introduction

Protein Structure

Among the family of biomolecules, no class is more diverse, and arguably more important, than proteins. Formed from fundamental amino acid building blocks that linearly combine, mature proteins reach a complexity of more than the sum of their parts through unique behaviors and properties that arise from their final structure. With just five residues a peptide can be formed in 3,200,000 ways, illustrating the high degree of variation that can be observed in just the primary structure of proteins that contain hundreds to thousands of residues. However, where proteins begin to develop their true character is in the formation of higher order structures. **Figure 1.1** shows the tiers of structure formed by chains of amino acids. The secondary structure is comprised of local hydrogen bonding interactions between backbone amide groups that form into orderly arrangements, most notably the alpha-helix and beta-sheet. These local structures provide one influence on the overall three-dimensional tertiary structure of a protein. The tertiary shape is further fashioned by interactions between side chains, where the amino acid residues have a large effect. Basic and acidic side chains holding charges can form salt bridges which shackle sections together, cysteine residues can pair with others in a disulfide bridge, and hydrophobic side chains can clump together to minimize their solvent exposure. Finally, the quaternary structure of a protein is determined by the oligomers formed from a collection of protein units. These structural elements determine the behavior and features of each protein, giving it a unique role in the biochemical machinery of the cell.

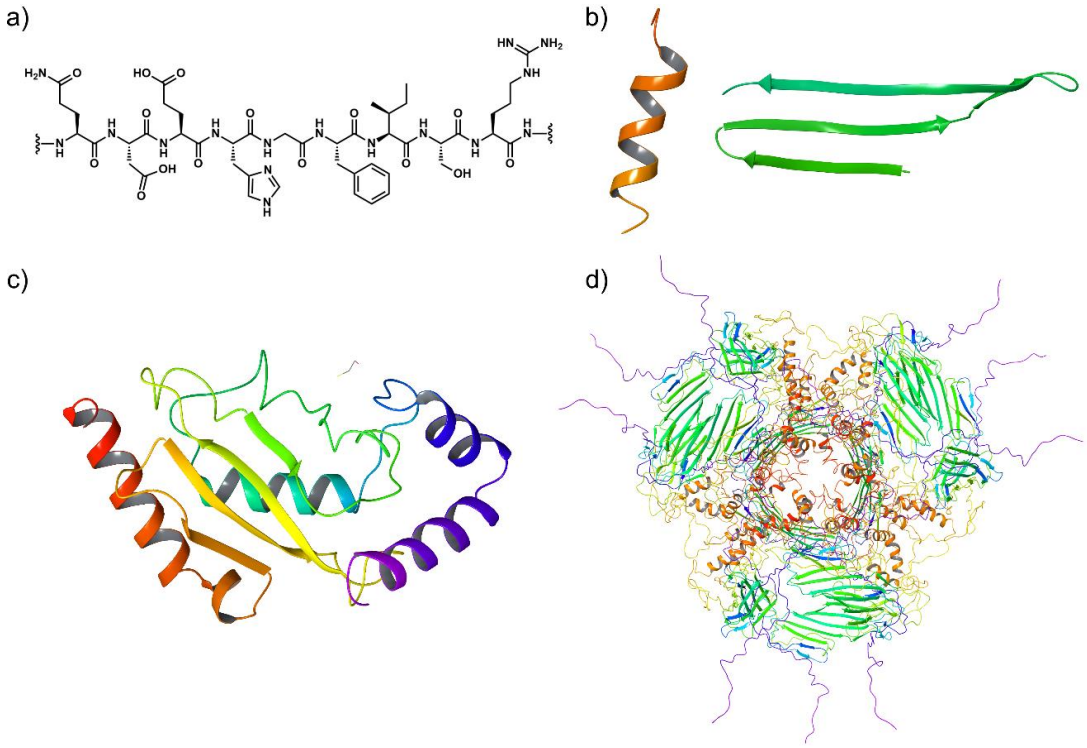


Figure 1.1. Structures formed by proteins modelled from PDB entry 3J07 of alpha B crystallin and 2BEP of ubiquitin conjugating enzyme E2-25K. **a)** The primary structure is constructed from amino acids linked via an amide bond. **b)** The secondary structure forms from hydrogen bonding between chains and typically forms into alpha-helices (left) or beta-sheets (right). **c)** The tertiary structure represents the overall protein monomer structure. **d)** The quaternary structure is formed from oligomerization of multiple monomer subunits.

Structure influences every aspect of protein behavior, and the proper function of a cell is wholly dependent on interactions between molecules. Proteins must be able to maintain structure as in the case of collagen, keratin, and elastin which serve to provide support for cells, tissues, hair, and nails.[1] Vitally important is the multi-faceted process of cell-signaling where proteins such as insulin bind to other proteins that serve as receptors,[2] immune system antibodies recognize targets through firm interactions,[3] and transport proteins such as hemoglobin move small molecules around by selectively binding or

releasing based on concentration[4]. A particularly prominent class of proteins are enzymes, which are necessary to perform cellular reactions in a highly specific and favorable manner. Cellular operations involving enzymes are diverse and numerous, notably including kinases and phosphatases involved in cell-signaling,[5] ligases which join biomolecules as in the ubiquitin system,[6] and the proteasome or autophagic proteases which degrade proteins[7] among many others. Each of these protein roles is fully dependent on maintaining the proper structural form of the protein with little room for error.

The complex structures formed are particularly sensitive to spontaneous alterations, which interfere with structural and functional roles in the cell.[8] One notorious source of protein damage is oxidation resulting from reactive oxygen species such as hydroxyls and peroxides.[9] These molecules can modify residues through covalent attachment, crosslinking, or fragmentation leading to myriad changes. Alterations to the primary structure subsequently affect the secondary and tertiary structure by modulating hydrophobic, backbone, and side chain interactions which can lead to aggregation or unfolding from what may only seem to be a small modification to a residue. While oxidation produces changes to atomic composition that make it straightforward to study, other sources of damage to protein structure can be difficult to trace. One of the most prevalent modifications in aged samples,[10,11] isomerization prevents a unique challenge to analysis by modifying the connectivity or orientation of atoms without an associated change in composition. In a similar vein, fibril formation in amyloids arises from secondary motifs spontaneously assembling into higher order structures, typically without

alteration of the primary residues. Like a lock and key dependent on shape, these alterations lead to consequences when interplay occurs between molecules.

Isomers

Isomerization is the rearrangement of a molecular structure that alters the spatial orientation of bonds or connectivity of the constituent atoms while retaining the same fundamental atomic composition. While this can take place in countless ways with small molecules, the most relevant isomerization in proteins occurs via side chain epimerization or most notoriously in the rearrangement of aspartic acid to isoaspartic acid. Epimerization, or the chiral inversion of one bond among multiple stereocenters, occurs primarily in the residues serine and aspartic acid when found in proteins. While the serine inversion mechanism is not well understood, the mechanism of isomerization in aspartic acid is well characterized. This process occurs through attack of the aspartic acid side chain by a backbone nitrogen, resulting in formation of a succinimide ring (**Figure 1.2**). The succinimide can be further epimerized to the D-succinimide before either form is hydrolyzed at a carbonyl to open as aspartic acid or isoaspartic acid. Asparagine can similarly cyclize to a succinimide ring by losing ammonia from the side chain; however, the ring is opened by water to produce aspartic acid isomers and does not return to asparagine. This represents a deamidation rather than isomerization of the side chain, and is accompanied by a mass shift.

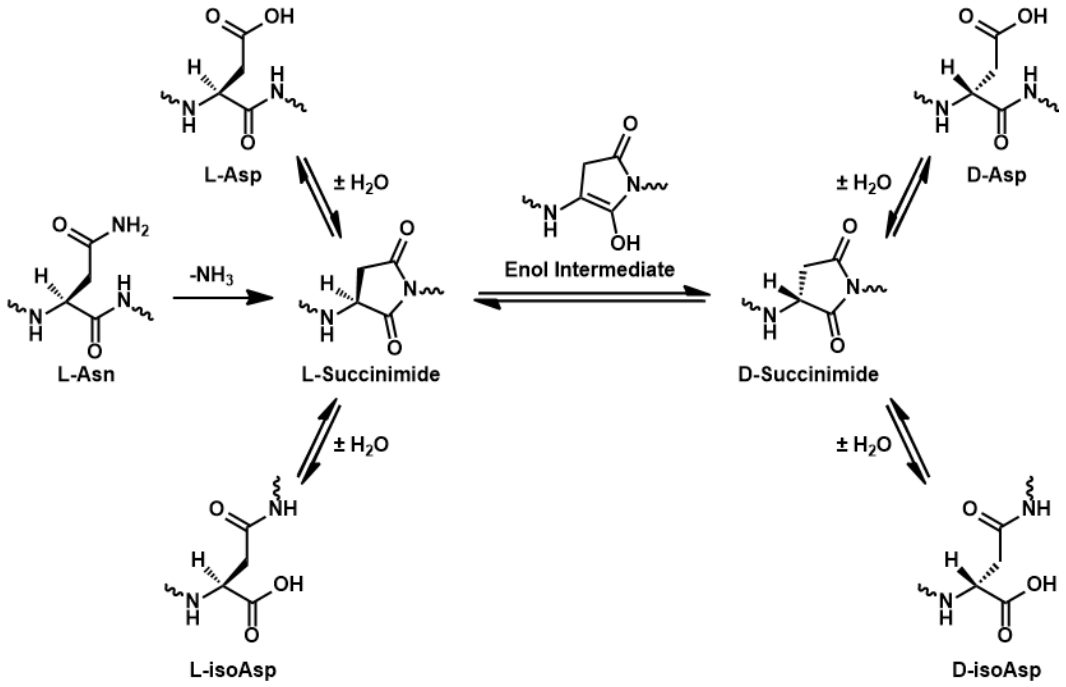


Figure 1.2. L-aspartic acid and L-asparagine proceed through a succinimide intermediate to produce isomers of aspartic acid. This ring can epimerize during the process, and the overall reaction yields four possible isomeric products (L-Asp, D-Asp, L-isoAsp, D-isoAsp).

While this may seem like a minor modification to a residue, deamidation and isomerization can provoke larger repercussions. For instance, it has long been known that aspartic acid isomerization can disrupt antibody function by distorting structure in the binding region, preventing the crucial interaction between the antibody and its target.[12,13] Additionally, deamidation of asparagine in RNase U2 induces a change in the secondary structure by partial unfolding of an alpha-helix, resulting in a lowered activity towards its substrate.[14] Transformation of higher order structure via aggregation is widely associated with aspartic acid isomers as well, which has been investigated in relation to Alzheimer's disease[15,16] and cataracts[17,18]. In protein interactions, aspartic acid isomers have been demonstrated to interfere with several proteases, causing them to be unable to bind and

cleave the isomerized peptide substrate.[19] Besides aspartic acid, serine epimerization is also observed in aged samples and has been found to interfere with kinase binding when present in its D-form.[20] Given the array of research demonstrating the complications arising from the isomerization of primary structure, we are interested in understanding the effect isomers have on biological systems by modulating enzymatic interactions.

Fibrils

Among protein structures, fibrils represent a notable class that is wide-spread among biological systems, operating in both advantageous and disadvantageous roles. Fibrils are formed as long and thin structures built from repeating subunits, which can adopt further structure by bending or coiling. In beneficial forms, fibrils function in a structural role by comprising architectures such as collagen, keratin, and cellulose. However, the primary focus with fibril research has been on the detrimental effects associated with amyloid aggregates. Amyloids form as aggregates of protein fibril structures, typically constructed from repeating β-sheets. Unlike isomers which modify the primary structure, fibrilization alters the higher order structures formed by the peptide while leaving the primary structure intact. These large, coalesced entities are associated with a multitude of diseases which are tied to a specific fibrilized species.[21] Prominently studied amyloids include α-Synuclein in Parkinson's disease,[22] prions in transmissible spongiform encephalopathies[23] and most infamously the amyloid beta peptide in Alzheimer's disease (AD).[24]

Amyloid beta (Aβ) is a peptide which can take multiple forms depending on its cleavage sites from the amyloid precursor protein (APP), forming possible sequences of 37-49 residues most commonly observed as 40 or 42 residues.24 After cleavage, the peptide backbone of Aβ folds over on itself, forming a beta-sheet motif stabilized by a salt bridge connection between residues Asp23 and Lys28.[25] These beta-sheets then combine together shown in **Figure 1.3**, forming a fibril which is elongated by the additional peptides. Further increasing the complexity of Aβ products, fibril structures solved by Nuclear magnetic resonance (NMR)[26,27] and Cryogenic electron microscopy (Cryo-EM)[28] demonstrate polymorphism with two major forms observed, a dimeric and trimeric form composed of two and three strands, respectively. Polymorphism affects not only the fibril form, but the size as well. While the primary structure remains untouched, the alterations to the higher order structure are known to inhibit proteolytic degradation by multiple proteases in the brain.[29] Because the formation of fibrils blocks access to the backbone and side chain residues used for protease-substrate binding, we are interested in studying the consequences of different fibril structural forms in relation to proteolytic clearance.

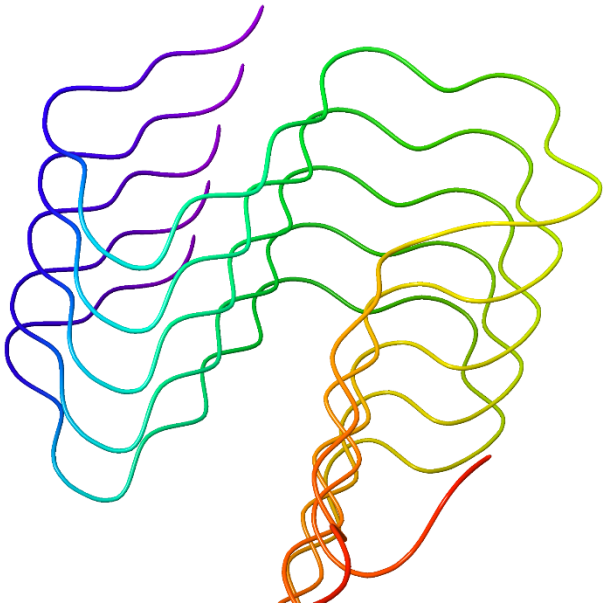


Figure 1.3. Stacked beta sheet structures constructed from the amyloid beta peptide modelled from PDB entry 5KK3.

Alzheimer's Disease

Since its discovery in diseased brains, Aβ research has been intrinsically linked with investigations into the pathology and cure to Alzheimer's disease. AD is characterized by the diagnostic hallmarks of amyloid plaques comprised of Aβ fibrils, and neurofibrillary tangles (NFT) consisting of hyper-phosphorylated tau protein.[30] The "Amyloid Cascade Hypothesis" has received the most attention as a possible pathological mechanism, centered around Aβ plaque and tau NFT deposition in the brain causing loss of memory and cognitive decline over time. This hypothesis draws connections to AD based on several pieces of evidence that indicate a role for Aβ. The gene for APP is located on chromosome 21, a chromosome most often associated with Down syndrome patients who are born with an extra copy and additionally exhibit AD-like pathology over the age of 40

years.[31] Genetic connections have been identified involving AD and presenilin which is involved in APP cleavage,[32,33] and apolipoprotein E which is thought to be involved in Aβ clearance,[34] further implying a role for Aβ.

Numerous drugs have been designed around this hypothesis, aimed at preventing Aβ production, plaque buildup, or clearing existing plaques. However, despite the implied role of Aβ and a multitude of human clinical trials, all drug trials except one have failed to arrest decline or result in cognitive improvement.[35] Additionally, some drugs such as AN-1792 caused additional brain atrophy even after successfully removing plaques. The only drug which has been approved by the FDA is the recently authorized Aducanumab, an antibody which targets amyloid plaques for clearance.[36] However, this decision was highly controversial due to a lack of strong evidence and questionable data analysis, leading 10 out of 11 members of the advisory committee to recommend not approving the drug and 3 members to resign following its approval. What the Aβ evidence and the failed drug trials may indicate is that Aβ plays a role in the disease, but that it is seemingly not the single causative agent in the pathology of AD.

A sometimes overlooked yet ubiquitous hallmark of AD is the accumulation and storage of lysosomes in the brain.[37] This feature of AD pathology results from failure of the autophagic pathways and actually precedes the formation of amyloid plaques.[38] Additionally, Aβ has been identified in autophagic organelles in AD such as endosomes,[39] multivesicular bodies,[40] and lysosomes[41]. A large and growing body of work has connected AD and autophagy, though a direct culprit in autophagic failure has not been identified. Due to the effect that protein structural modifications of

isomerization and fibrilization have on protease digestion, we are interested in studying how these structural modifications may affect lysosomal proteases as a potential factor in the autophagic failure observed in AD. Like AD, these processes have an inherent time factor to them, particularly isomerization, and a missing piece of the puzzle may be found in the damaging modifications.

The Lysosome

Proteostasis in the cell is an integral yet delicate process which balances the production and degradation of proteins within. The degradation portion exists as a clearance and recycling mechanism, performed by either the ubiquitin-proteasome apparatus or the system of lysosomal autophagy.[42] Multiple pathways operate to deliver undesirable biomolecules to the lysosome, either from outside the cell through endocytosis or from inside the cell through macro, micro, and chaperone-mediated autophagy.[43] The macroautophagy and endocytosis pathways transport products by encasing them in an acidic membrane, while the microautophagy and chaperone-mediated pathways deliver cytosolic components directly into the lysosome. Once inside the acidic lysosome, a battery of hydrolases lead to degradation of the biomolecules into their constituent building blocks. This system is both robust enough to handle a variety of biomolecules which may not even originate in the cell, and fragile enough that a defect in the machinery precipitates a class of diseases known as lysosomal storage disorders.[44] Many of these disorders are a result of a dysfunctional hydrolase which cannot degrade

its target substrate. When this occurs, the substrate accumulates in the lysosome leading to failure and storage of the organelle.[45]

For the hydrolases tasked with degrading proteins, this role is chiefly performed by the family of proteases known as cathepsins. There are 15 cathepsins with different catalytic residues including aspartic acid and serine, though 11 of 15 utilize cysteine as the cleavage residue.[46] These can be further subdivided into endopeptidases which cleave the peptide backbone internally, and exopeptidases which bind to the termini to cleave one or two residues at a time. The cathepsins have broad specificities tending to favor hydrophobic residues near the cleavage site46, however other selectivity has been observed such as cathepsin B favoring arginine,[47] cathepsin K preferring proline,[48] and cathepsin L favoring positively charged residues in some positions.[49] The combined efforts of the cathepsins fully degrade waste proteins, with the endopeptidases serving to break the protein into smaller peptide chunks and the exopeptidases finishing the job by binding and removing the constituent amino acids. As the lysosomal system can be disturbed by dysfunctional hydrolases, we are interested in whether perturbations to the protein structure can interfere with the cathepsins and likewise result in disruption of the autophagic process. Any flaw in the system can lead to lysosomal failure and storage, and Alzheimer's disease may represent a unique lysosomal storage disorder which results from substrate modification instead of systemic modification.

Mass Spectrometry

Since its invention in the early 1900's by J.J. Thomson, mass spectrometry has developed into one of the single most useful instruments employed by analytical chemists. Early uses of the technique allowed Thomson to discover the first isotopes of a stable element by separating a beam of neon particles, revealing two ions with different mass.[50] However, it wasn't until decades later that Klaus Biemann would establish the powerful capabilities of mass spectrometry by applying it to structural determination in organic chemistry. Initially, Biemann worked on the alkaloid sarpagine, and was able to match its proposed structure to a known structure by chemical modification and analysis of the fragmentation spectra.[51] Biemann next brought this capability to the field of biochemistry, aiming to develop a method for sequencing of proteins with mass spectrometry. The first protein sequence of Insulin had only been published five years before, determined through arduous chemical methods to break the sequence down stepwise until amino acids could be analyzed sequentially.[52] While this was able to sequence Insulin with its 51 residues, application to larger proteins would be exponentially more difficult. Biemann was able to ionize peptides for analysis and develop a method for direct fragmentation analysis rather than spectra matching.[53] However, this technique required chemical modification for volatility and a hard ionization method which produces extensive fragmentation in large biomolecules.

A major advance in protein analysis came with the development of electrospray ionization, a soft ionization method which could safely produce gas-phase ions of large biomolecules with very little fragmentation.[54] This technique carried an additional

advantage by ionizing liquid solvent directly, allowing it to be coupled with liquid-chromatography for facile separation of analytes followed by immediate analysis with no further steps. A host of techniques involving modification and digestion of proteins for liquid-chromatography mass spectrometry (LC-MS) developed for structural determination at a far more rapid pace than had been seen previously. Hydrogen-deuterium exchange (HDX) and fast photochemical oxidation of proteins (FPOP) are applied by labelling solvent-exposed protein surfaces to study structural form. HDX can be used to study conformational changes that occur from PTMs that result in a shift in higher order structure.[55] Similarly FPOP has been used to detect structural differences between a mutated antibodies and their original form.[56] Crosslinking is a technique which covalently connects residues before digestion, and is often employed in studying protein-protein interaction sites by linking interaction partners via nearby residues.[57]

Mass spectrometry experiments are largely dependent on the analysis of fragmentation, as intact mass analysis is limited by the vast overlap of molecular weights in chemical space. Gas-phase fragmentation of analytes can be achieved by many methods, but those commonly used are collision-induced dissociation (CID), electron-capture or electron-transfer dissociation (ECD/ETD), and radical-directed dissociation (RDD), across which the varying types of peptide fragments shown in **Figure 1.4** can be achieved. CID is the most widely employed technique which operates by heating a peptide through numerous gas-phase collisions until the lowest energy pathway results in cleavage of the amide bond, producing b- and y-ions.[58] ECD/ETD proceed by introducing a radical to generate a hydrogen-rich peptide (no loss of atoms) which leads to cleavage

of the N-C_α bond after heating, resulting in c- and z-ions which are not typically found in CID fragmentation spectra.[59] A distinct radical technique, RDD generates an odd-electron species by cleavage of a covalent bond to generate a hydrogen-deficient species.[60] In a hydrogen-deficient peptide an atom has been removed, leaving an unoccupied site where an atom from elsewhere in the structure can be abstracted for relocation. This in turn generates a radical at the site of abstraction, serving as a mechanism for radical migration throughout the structure before rearrangement leads to bond cleavage. These sites must be adjacent in space but may be distant in sequence, making RDD fragmentation inherently more sensitive to three-dimensional structure than other methods. Radicals in hydrogen-deficient peptides migrate to α-, β- and γ-carbons, leading to a variety of unique fragmentation when they rearrange neighboring bonds.[61] α- and γ-carbon radicals will rearrange the bonds of the side chain to produce partial and full side chain losses, respectively. Migration to the β-carbon will lead to cleavage of the C_α-carbonyl bond, producing a- and x-ions not found in other fragmentation methods. Additionally, x-ions are unstable and can break down to form z+1 ions.

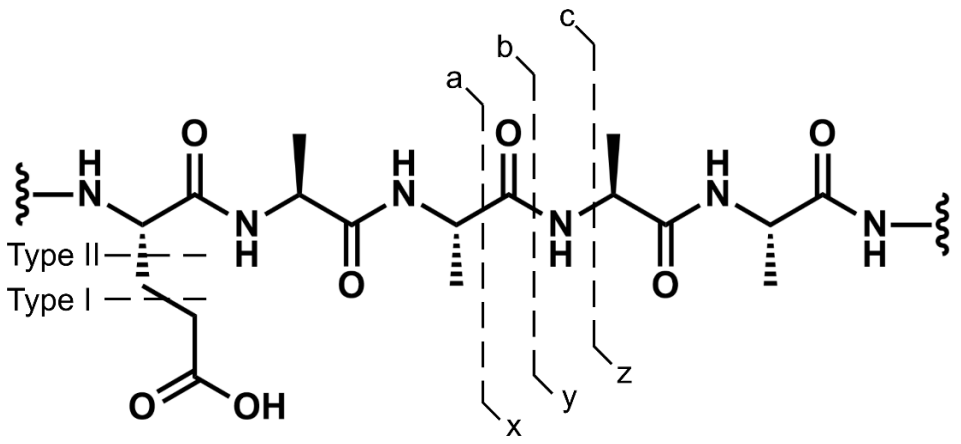


Figure 1.4. Mass spectrometry fragmentation nomenclature. a/x, b/y, and c/z ions represent backbone fragmentation while Type I/II fragments occur when part or all of a side chain is cleaved.

Fragmentation is useful for the host of structural information it contains upon analysis. Bottom-up proteomics in mass spectrometry is founded on this principle, allowing cellular proteins to be digested into peptides which are identified using their fragmentation and the sequences of the proteome.[62] These proteome sequences were in turn identified in large part by mass spectrometry through use of the peptide fragmentation to determine the primary structure.[63] Structural features can be discerned as well, for instance salt-bridges in zwitterions can be confirmed using c- and z-ions in a modified form of ETD,[64] and disulfide links can be found with UVPD[65]. While these techniques use the identity of the fragments for analysis, other information is also present in the fragment intensities. Importantly, fragment intensity can be used to differentiate amino acid isomers in proteins, which do not result in a mass shift but display different fragmentation intensities.[66,11]. Isomer fragmentation spectra can be quantitatively compared via an R_{isomer} score by using the ratios of fragments which change significantly between samples.[67] For these reasons we are interested in utilizing the ability of mass

spectrometry to reveal and discriminate structural differences in isomer and fibril modifications.

Proteins are vitally important to proper cellular function in life, and yet are susceptible to diverse alterations including some that are undetectable to many analytical techniques. Isomers and fibrils occur naturally in cellular systems and have established ties to disease pathologies, yet these modifications represent a challenge to study as they do not change the underlying composition of the protein. We are further interested in studying these structural changes when they lie at the intersection of autophagic failure and Alzheimer's disease, where they may play a causative role in lysosomal dysfunction. Accordingly, the work described in this dissertation aims to utilize mass spectrometry as a powerful tool to study these difficult to characterize modifications and their consequences in enzymatic interactions.

References

[1] Yeo, J.; Jung, G.; Tarakanova, A.; Martín-Martínez, F.J.; Qin, Z.; Cheng, Y.; Zhang, Y.-W.; Buehler, M.J. Multiscale modeling of keratin, collagen, elastin and related human diseases: Perspectives from atomistic to coarse-grained molecular dynamics simulations. *Extreme Mech. Lett.* **2018**, *20*, 112-124.

[2] Haeusler, R.A.; McGraw, T.E.; Accili, D. Biochemical and cellular properties of insulin receptor signalling. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 31-44.

[3] Lu, R.-M.; Hwang, Y.-C.; Liu, I-J.; Lee, C.-C.; Tsai, H.-Z.; Li, H.-J.; Wu, H.-C. Development of therapeutic antibodies for treatment of diseases. *J. Biomed. Sci.* **2020**, *27*, 1-30.

[4] Schechter, A.N. Hemoglbin research and the origin of molecular medicine. *Blood* **2008**, *112(10)*, 3927-3938.

[5] Cheng, H.-C.; Qi, R.Z.; Paudel, H.; Zhu, H.-J. Regulation and Function of Protein Kinases and Phosphatases. *Enzyme Res.* **2011**, *2011*, 1-4.

[6] Nakayama, K.I.; Nakayama, K. Ubiquitin ligases: cell-cycle control and cancer. *Nat. Rev. Cancer.* **2006**, *6*, 369-381.

[7] Lilienbaum, A. Relationship between the proteasomal system and autophagy. *Int. J. Biochem. Mol. Biol.* **2013**, *4(1)*, 1-26.

[8] Chondrogianni, N; Petropoulos, I.; Grimm, S.; Georgila, K.; Catalgol, B.; Friguet, B.; Grune, T.; Gonos, E.S. Protein damage, repair and proteolysis. *Mol. Aspects. Med.* **2014**, *35*, 1-71.

[9] Davies, M.J. Protein oxidation and peroxidation. *Biochem. J.* **2016**, *473*, 805-825.

[10] Lyon, Y.A.; Sabbah, G.M.; Julian, R.R. Identification of Sequence Similarities among Isomerization Hotspots in Crystallin Proteins. *J. Proteome Res.* **2017**, *16*, 1797-1805.

[11] Truscott, R.J.W.; Friedrich, M.G. The etiology of human age-related cataract. Proteins don't last forever. *Biochim. Biphys Acta.* **2016**, *1860*(1 Pt B), 192-198.

[12] Cacia, J.; Keck, R.; Presta, L.G.; Frenz, J. Isomerization of an Aspartic Acid Residue in the Complementarity-Determining Regions of a Recombinant Antibody to Human IgE: Identification and Effect on Binding Affinity. *Biochemistry* **1996**, *35*, 1897-1903.

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

[14] 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.

[15] Mitkevich, V.A.; Petrushanko, I.Y.; Yegorov, Y.E.; Simonenko, O.V.; Vishnyakova, K.S.; Kulikova, A.A.; Tsvetkov, P.O.; Makarov, A.A.; Kozin, S.A. Isomerization of Asp7 leads to increased toxic effect of amyloid-β42 on human neuronal cells. *Cell Death Dis.* **2013**, *4*, No. e939.

[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*, 9778-9786.

[17] 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.

[18] 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*, 131-141.

[19] Böhme, L.; Bär, J.W.; Hoffmann, T.; Manhart, S.; Ludwig, H.-H.; Rosche, F.; Demuth, H.-U. Isoaspartate residues dramatically influence substrate recognition and turnover by proteases. *Biol. Chem.* **2008**, *389*, 1043-1053.

[20] Lyon, Y.A.; Collier, M.P.; Riggs, D.L.; Degiacomi, M.T.; Benesch, J.L.P. Structural and functional consequences of age-related isomerization in α-crystallins. *J. Biol. Chem.* **2019**, *294(19)*, 7546-7555.

[21] Chiti, F.; Dobson, C.M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **2017**, *86*, 27-68.

[22] Baba, M.; Nakajo, S.; Tu, P.H.; Tomita, T.; Nakaya, K.; Lee, V.M.; Trojanowski, J.Q.; Iwatsubo, T. Aggregation of alpha-synuclein in Lew bodies of sporadic Parkinson's disease and dementia with Lewy bodies. *Am. J. Pathol.* **1998**, *152(4)*, 879-884.

[23] Cobb, N.J.; Surewicz, W.K. Prion Diseases and Their Biochemical Mechanisms. *Biochemistry* **2009**, *48(12)*, 2574-2585.

[24] Chen, G.-F.; Xu, T.-H.; Yan, Y.; Zhou, Y.-R.; Jiang, Y.; Melcher, K.; Xu, H.E. Amyloid beta: structure, biology, and structure-based therapeutic development. *Acta Pharmacol. Sin.* **2017**, *38*, 1205-1235.

[25] Lührs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Döbeli, H.; Schubert, D.; Riek, R. 3D structure of Alzheimer's amyloid-β(1-42) fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102(48)*, 17342-17347.

[26] Lu, J.-X. Qiang, W.; Yau, W.-M.; Schwieters, C.D.; Meredith, S.C.; Tycko, R. Molecular Structure of β-Amyloid Fibrils in Alzheimer's Disease Brain Tissue. *Cell* **2013**, *154(6)*, 1257-1268.

[27] Paravastu, A.K.; Leapman, R.D.; Yau, W.-M.; Tycko, R. Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105(47)*, 18349-18354.

[28] Gremer, L.; Schölzel, D.; Schenk, C.; Reinartz, E.; Labahn, J.; Ravelli, R.B.G.; Tusche, M.; Lopez-Iglesias, C.; Hoyer, W.; Heise, H.; Willbold, D.; Schröder, G.F. Fibril structure of amyloid-β(1-42) by cryo-electron microscopy. *Science* **2017**, *358*, 116-119.

[29] Chauhan, V.; Sheikh, A.M.; Chauhan, A.; Spivack, W.D.; Fenko, M.D.; Malik, M.N. Fibrillar amyloid beta-protein inhibits the activity of high molecular weight brain protease and trypsin. *J. Alzheimer's Dis.* **2005**, *7(1)*, 37-44.

[30] Ballard, C.; Gauthier, S.; Corbett, A.; Brayne, C.; Aarsland, D.; Jones, E. Alzheimer's disease. *Lancet*, **2011**, *377*, 1019-1031.

[31] Lott, I.T.; Head, E. Alzheimer's disease and Down syndrome: factors in pathogensis. *Neurobiol. Aging* **2005**, *26*, 383-389.

[32] Sherrington R.; Rogaev E.I.; Liang Y.; Rogaeva E.A.; Levesque G.; Ikeda M.; Chi H.; Lin C.; Li G.; Holman K.; Tsuda T.; Mar L.; Foncin J.F.; Bruni A.C.; Montesi M.P.; Sorbi S.; Rainero I.; Pinessi L.; Nee L.; Chumakov I.; Pollen D.; Brookes A.; Sanseau P.; Polinsky R.J.; Wasco W.; Da Silva H.A.; Haines J.L.; Perkicak-Vance M.A.; Tanzi R.E.; Roses A.D.; Fraser P.E.; Rommens J.M.; St George-Hyslop P.H.; Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **1995**, *375(6534)*,754-760.

[33] Rogaev EI.; Sherrington R.; Rogaeva E.A.; Levesque G.; Ikeda M.; Liang Y.; Chi H.; Lin C.; Holman K.; Tsuda T, Mar, L.; Sorbi, S.; Nacmias, B.; Piacentini, S.; Amaducci, L.; Chumakov, I.; Cohen, D.; Lannfelt, L.; Fraser, P.E.; Rommens, J.M.; St George-Hyslop, P.H. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* **1995**, *376(6543)*, 775-778.

[34] Corder, E.H.; Saunders, A.M.; Strittmatter, W.J.; Schmechel, D.E.; Gaskell, P.C.; Small, G.W.; Roses, A.D.; Haines, J.L.; Pericak-Vance, M.A. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **1993**, *261(5123)*, 921-923.

[35] Drachman, D.A. The amyloid hypothesis, time to move on: Amyloid is the downstream result, not cause, of Alzheimer's disease. *Alzheimer's Dementia* **2014**, *10*, 372-380.

[36] Crosson, F.J.; Covinsky, K.; Redberg, R.F.; Medicare and the Shocking US Food and Drug Administration Approval of Aducanumab: Crisis or Opportunity? *JAMA Intern Med.* [Online early access]. doi:10.1001/jamainternmed.2021.4610. Published online July 13, 2021. www.jamanetwork.com (accessed Jul 21, 2021).

[37] Nixon, R.A; Yang, D.-S. Autophagy failure in Alzheimer's disease-locating the primary defect. *Neurobiol. Dis.* **2011**, *43(1)*, 38-45.

[38] Cataldo, A.M.; Peterhoff, C.M.; Troncoso, J.C.; Gomez-Isla, T.; Hyman, B.T.; Nixon, R.A. Endocytotic pathway abnormalities precede amyloid beta deposition in sporadic Alzheimer's disease and Down syndrome: differential effects of APOE genotype and presenilin mutations. *Am. J. Pathol.* **2000**, *157(1)*, 277-286.

[39] Cataldo, A.M.; Petanceska, S.; Terio, N.B.; Peterhoff, C.M.; Durham, R.; Mercken, M.; Mehta, P.D.; Buxbaum, J.; Haroutunian, V.; Nixon, R.A. Abeta localization in abnormal endosomes: association with earliest Abeta elevations in AD and Down syndrome. *Neurobiol. Aging* **2004**, *25*, 1263-1272.

[40] Takahashi, R.H.; Milner, T.A.; Li, F.; Nam, E.E.; Edgar, M.A.; Yamaguchi, H.; Beal, M.F.; Xu, H.; Greengard, P.; Gouras, G.K. Interneuronal Alzheimer abeta42 accumulates in multivesicular bodies and is associated with synaptic pathology. *Am. J. Pathol.* **2002**, *161*, 1869-1879.

[41] Liu, R.Q.; Zhou, Q.H.; Ji, S.R.; Zhou, Q.; Feng, D.; Wu, Y.; Sui, S.F. Membrane localization of beta-amyloid 1-42 in lysosomes: a possible mechanism for lysosome labilization. *J. Biol. Chem.* **2010**, *285*, 19986-19996.

[42] Ciechanover, A. Intracellular protein degradation: from a vague idea thru the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. *Cell Death Differ.* **2005**, *12*, 1178-1190.

[43] Ballabio, A.; Bonifacino, J.S. Lysosome as dynamic regulators of cell and organismal homeostasis. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 101-118.

[44] Vellodi, A. Lysosomal storage disorders. *Br. J. Haematol.* **2004**, *128*, 413-431.

[45] Kiselyov, K.; Jennings Jr. J.J.; Rbaibi, Y.; Chu, C.T. Autophagy, Mitochondria and Cell Death in Lysosomal Storage Diseases. *Autophagy* **2007**, *3(3)*, 259-262.

[46] Turk, V.; Stoka, V.; Vasiljeva, O.; Renko, M.; Sun, T.; Turk, B.; Turk, D. Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochim. Biophys. Acta* **2012**, *1824*, 66-68.

[47] Cotrin, S.S.; Puzer, L.; Judice, W.A.D.; Juliano, L.; Carmona, A.K.; Juliano, M.A. Positional-scanning combinatorial libraries of fluorescence resonance energy transfer peptides to define substrate specificity of carboxypeptidases: assays with human cathepsin B. *Anal. Biochem.* **2004**, *335*, 244-252.

[48] Lecaille, F.; Weidauer, E.; Juliano, M.A.; Brömme, D.; Lalmanach, G. Probing cathepsin K activity with a selective substrate spanning its active site. *Biochem. J.* **2003**, *375*, 307-312.

[49] Puzer, L.; Cotrin, S.S.; Alves, M.F.M.; Egborge, T.; Araújo, M.S.; Juliano, M.A.; Juliano, L.; Brömme, D.; Carmona, A.K. Comparative substrate specificity analysis of recombinant human cathepsin V and cathepsin L. *Arch. Biochem. Biophys.* **2004**, *430*, 274-283.

[50] Thomson, J.J. Rays of Positive Electricity. *Proc. R. Soc. A.* **1913**, *89*, 1-20.

[51] Biemann, K. Four Decades of Structure Determination by Mass Spectrometry: From Alkaloids to Heparin. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 1254-1272.

[52] Sanger, F.; Thompson, E.O.P. The Amino-acid Sequence in the Glycyl Chain of Insulin. *Biochem. J.* **1953**, *53(3)*, 353-366.

[53] Biemann, K. Laying the groundwork for proteomics Mass spectrometry from 1958 to 1988. *Int. J. Mass Spectrom.* **2007**, *259*, 1-7.

[54] 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.

[55] Trabjerg, E.; Nazari, Z.E.; Rand, K.D. Conformational analysis of complex protein states by hydrogen/deuterium exchange mass spectrometry (HDX-MS): Challenges and emerging solutions. *TrAC, Trends Anal. Chem.* **2018**, *106*, 125-138.

[56] Cornwell, O.; Bond, N.J.; Radford, S.E.; Ashcroft, A.E. Long-Range Conformational Changes in Monoclonal Antibodies Revealed Using FPOP-LC-MS/MS. *Anal. Chem.* **2019**, *91(23)*, 15163-15170.

[57] Sinz, A. Investigation of protein-protein interactions in living cells by chemical crosslinking and mass spectrometry. *Anal. Bioanal. Chem.* **2010**, *397*, 3433-3440.

[58] Wells, J.M.; McLuckey, S.A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol.* **2005**, *402*, 148-185.

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

[60] Tureček, F.; Julian, R.R. Peptide Radicals and Cation Radicals in the Gas Phase. *Chem. Rev.* **2013**, *113*, 6691-6733.

[61] 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.

[62] Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198-207.

[63] Hunt, D.F.; Yates III, J.R.; Shabanowitz, J.; Winston, S.; Hauer, C.R. Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233-6237.

[64] Bonner, J.; Lyon, Y.A.; Nellessen, C.; Julian, R.R. Photoelectron Transfer Dissociation Reveals Surprising Favorability of Zwitterionic States in Large Gaseous Peptides and Proteins. *J. Am. Chem. Soc.* **2017**, *139(30)*, 10286-10293.

[65] 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*, 5176-5184.

[66] 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.

[67] 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*, 10598-10609.

CHAPTER 2: Spontaneous Isomerization of Long-Lived Proteins Provides a Molecular Mechanism for the Lysosomal Failure Observed in Alzheimer's Disease

Abstract

Proteinaceous aggregation is a well-known observable in Alzheimer's disease (AD), but failure and storage of lysosomal bodies within neurons is equally ubiquitous and actually precedes bulk accumulation of extracellular amyloid plaque. In fact, AD shares many similarities with certain lysosomal storage disorders though establishing a biochemical connection has proven difficult. Herein, we demonstrate that isomerization and epimerization, which are spontaneous chemical modifications that occur in long-lived proteins, prevent digestion by the proteases in the lysosome (namely the cathepsins). For example, isomerization of aspartic acid into L-isoAsp prevents digestion of the N-terminal portion of Aβ by cathepsin L, one of the most aggressive lysosomal proteases. Similar results were obtained after examination of various target peptides with a full series of cathepsins, including endo-, amino-, and carboxy-peptidases. In all cases peptide fragments too long for transporter recognition or release from the lysosome persisted after treatment, providing a mechanism for eventual lysosomal storage and bridging the gap between AD and lysosomal storage disorders. Additional experiments with microglial cells confirmed that isomerization disrupts proteolysis in active lysosomes. These results are easily rationalized in terms of protease active sites, which are engineered to precisely orient the peptide backbone and cannot accommodate the

backbone shift caused by isoaspartic acid or side chain dislocation resulting from epimerization. Although Aβ is known to be isomerized and epimerized in plaques present in AD brains, we further establish that the rates of modification for aspartic acid in positions 1 and 7 are fast and could accrue prior to plaque formation. Spontaneous chemistry can therefore provide modified substrates capable of inducing gradual lysosomal failure, which may play an important role in the cascade of events leading to the disrupted proteostasis, amyloid formation, and tauopathies associated with AD.

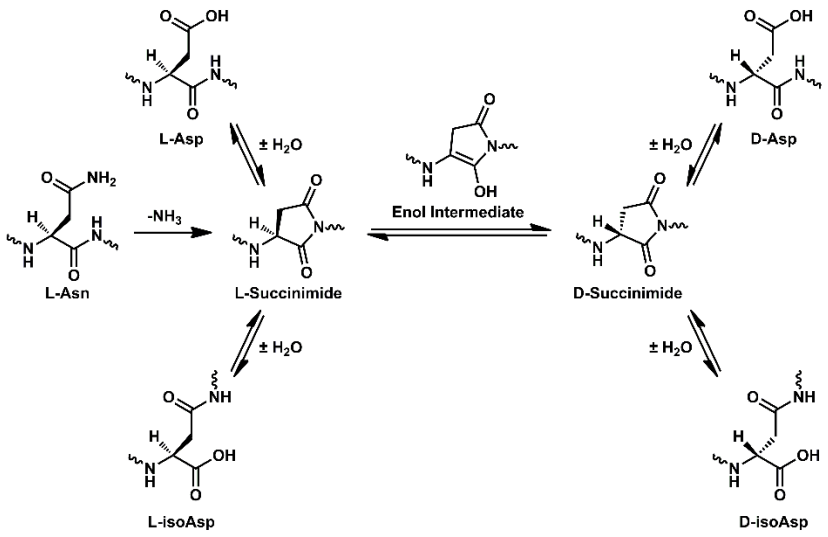
Introduction

The active balancing of protein synthesis and degradation, or proteostasis, is an ongoing and critical process in most cells.[1] Proteins must be created, carry out their requisite function, and then be recycled once they are no longer needed or have become nonfunctional. Several pathways are available for protein degradation, including the proteasome, macro-autophagy, micro-autophagy, and chaperone-mediated autophagy.[2,3] The autophagy-related pathways deliver proteins to lysosomes, which are acidic organelles containing a host of hydrolases, including many proteases.[4] Cargo taken into cells via endocytosis is also typically delivered to lysosomes for degradation. Regardless of the pathway, after cargo fuses with a lysosome, endopeptidases cleave proteins at internal sites, shortening proteins to peptides, which are then further digested from both termini by exopeptidases. After protein digestion has been completed, transporter proteins in the lysosomal membrane release (primarily) individual amino acids back into the cytosol for new protein synthesis or energy production.[5] Lysosomes are crucial for

maintaining cellular homeostasis, but they are also uniquely susceptible to problems when substrates cannot be hydrolyzed. For example, genetic modifications reducing the efficacy of a lysosomal hydrolase are the most common cause of lysosomal storage disorders. These devastating diseases involve 'storage' of failed lysosomal bodies within cells, which eventually leads to cell death and is particularly problematic for post-mitotic cells such as neurons.[6] Symptoms in lysosomal storage disorders usually emerge in infancy or childhood, are often associated with neurodegeneration, and are typically fatal.[7]

Long-lived proteins[8] are a primary target of the lysosome because they become modified and lose efficacy over time. A well-known example of this occurs with mitophagy,3 wherein old mitochondria are recycled in their entirety. Contributing factors that lead to long-lived protein deterioration include a variety of spontaneous chemical modifications, i.e. modifications not under enzymatic control.8 Some of these modifications are very subtle and difficult to detect, including isomerization and epimerization.[9] Isomerization occurs primarily at aspartic acid, when the side chain inserts into and elongates the peptide backbone (Scheme 2.1).[10] Identical products are also created during deamidation of asparagine, which further results in chemical transformation from one amino acid to another.[11] Epimerization occurs when an amino acid side chain inverts chirality from the L- to D- configuration. Peptide isomerization and epimerization do not have readily identifiable bioanalytical signatures, but both modulate structure in a subtle, yet significant way (see Fig. 2.1). Studies on the eye lens have shown that epimerization and isomerization are among the most abundant

modifications observed in extremely long-lived proteins.[12-14] However, knock-out experiments in mice have also revealed the importance of these modifications over much shorter timescales. For example, removal of the repair enzyme for L-isoAsp, protein-isoaspartyl methyl transferase (PIMT),[15] leads to lethal accumulation of isomerized protein in just 4-6 weeks.[16,17] This reveals that isomerization of aspartic acid is sufficiently dangerous that an enzyme has evolved to repair it.



Scheme 2.1. Pathways for isomerization of aspartic acid and deamidation of asparagine.

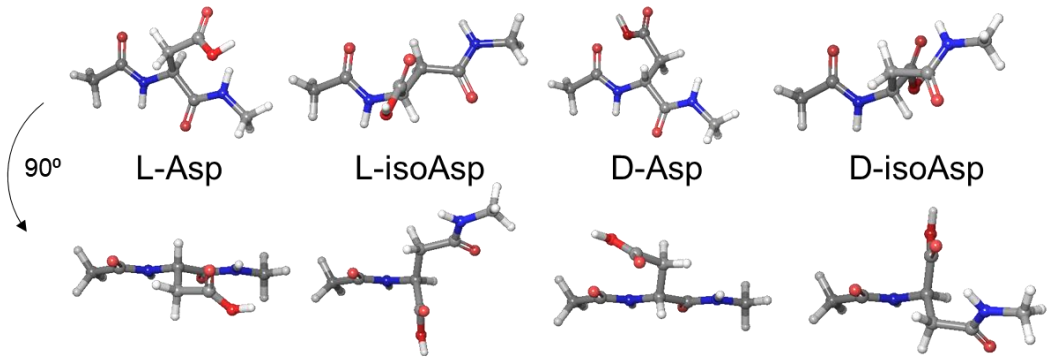


Fig. 2.1 Model structures of the aspartic acid isomers, where the iso-structure conformation closest to native backbone orientation is shown. Two views are illustrated for each isomer.

The importance of peptide isomers is further revealed in the uses nature has found for them. For example, single amino acid sites are intentionally epimerized in many venoms and in signaling neuropeptides in crustaceans.[18,19] The corresponding L-only peptides are not biologically active, confirming the importance of the chiral modifications. In addition, it is thought that epimerization is beneficial for these peptides because it allows them to escape, or prolong the time required for, proteolysis.[20] In fact, it is well-known that sites of epimerization and isomerization are both generally resistant to protease action, but the ramifications of such chemistry in the context of lysosome function have not been previously examined. Despite this absence, numerous studies have the established the importance of protein degradation in lysosomes. For example, knockout mice lacking cathepsin D grow normally for ~2 weeks but then die before the end of 4 weeks.[21] Examination of the neurons from these mice revealed an abundance of failed lysosomal bodies, similar to those observed in lysosomal storage disorders. Other research has shown that knockout mice lacking cathepsins B and L die within 2-4 weeks of birth. Again, accumulation of failed lysosomal bodies was observed in neurons of

these mice.[22] Although cathepsins can also be found outside the lysosome,[23] these results confirm significant, and likely fatal, impact on the lysosomal system when critical cathepsins are absent.

Amyloid aggregates or proteins that are otherwise insoluble are also targeted to lysosomes for degradation.[24] Amyloid aggregation has also captured the majority of attention as the potential cause of Alzheimer's disease (AD), but significant evidence also supports lysosomal storage as an underlying cause. For example, AD shares many pathological similarities with lysosomal storage disorders, including prolific storage of failed lysosomal bodies, accumulation of senile plaques, and formation of neurofibrillary tangles.[25,26] In fact, scanning-electron microscopy images of lysosomal storage (in neurons) are virtually indistinguishable between the two diseases. The lysosomal storage observed in AD precedes formation of amyloid deposits,[27] hinting that lysosomal malfunction may occur upstream of the events leading to extracellular amyloid aggregation. The parallels between the two diseases have also been offset by differences. For example, lysosomal storage disorders typically afflict youth and can progress rapidly, while AD typically occurs late in life over a longer timescale. Therefore, a mechanism accounting for the commonalities and differences between the diseases has been difficult to identify, but an intriguing possibility does exist.

The primary constituents of senile plaques, Aβ and tau, are both long-lived proteins that are subject to isomerization and epimerization.8 In fact, Aβ is significantly epimerized and isomerized in the brains of people with AD.[28] If isomerization and epimerization prevent lysosomal protein digestion, then a common link between

lysosomal storage disorders and AD would be established. In fact, AD would essentially represent a different type of lysosomal storage disorder, one that operates in reverse of the classical disease. Rather than failure of a *modified enzyme or modified transporter* to clear waste molecules, failure to digest or transport *modified waste* molecules would be operative and eventually lead to lysosomal storage. Close examination of another complex age-related disease, macular degeneration, reveals that there is precedence for substrate-induced lysosomal storage.[29]

Herein, we use mass spectrometry (MS) and liquid chromatography (LC) to demonstrate that isomerized or epimerized peptides resist degradation by cathepsins, including both endo- and exopeptidase activity. Important target peptides that are both long-lived and closely associated with AD were examined, including fragments of Aβ and tau. The results reveal that small peptide fragments comprised of residues surrounding isomerized or epimerized sites persist after digestion. Disrupted proteolysis was observed in both isolated reactions on model peptides in full-length Aβ, and in living cells, offering an explanation for the toxicity observed in previous experiments with cell and animal models employing isomerized Aβ (see discussion below). Additional experiments reveal that the rates of isomerization for the Asp residues in the N-terminal portion of Aβ are fast, providing a pathway for generation of these toxic species that could eventually lead to lysosomal failure and initiate other downstream consequences.

Results and Discussion

Defining limitations of cathepsin digestion

A series of isolated digestions of synthetic peptides in both canonical form and with isomerized or epimerized (iso/epi) sites were performed and the results are summarized in Fig. 2.2. Experiments were conducted with cathepsins D, L, B, and H. This collection includes all of the most abundant cathepsins and all modes of function, i.e. endo-, carboxy-, and aminopeptidases.[30,31] The peptide APSWFDTGLSEMR (αB^{57-69}), derived from αB-crystallin, was used as the initial test substrate. It contains both Ser and Asp residues known to be modified in the eye lens.[32] Furthermore, Ser59, Asp62, and Ser66 are each separated by other residues, allowing for semi-independent examination. Furthermore, the canonical sequence is a good substrate for proteolysis. Digestion of the native form with cathepsin D in acetate buffer at pH 4.5 yields the results shown in the upper part of Fig. 2.2a. The LC-MS derived ion chromatogram reveals many peptide fragments and almost complete consumption of the precursor. Clearly, the canonical all-L version of APSWFDTGLSEMR is easily digested. Substitution of L-Asp with L-isoAsp yields the lower chromatogram, where after 6 hours, the precursor remains basically untouched. A single modification therefore prevents cathepsin D from digesting an entire thirteen residue sequence, shutting down peptide hydrolysis at seven different sites. To more easily visualize the results in a condensed fashion, peptide fragments resulting from proteolysis are represented by color-coded lines below the peptide sequence as shown in Fig. 2.2b (full chromatograms are also provided in the supporting information). The data from Fig. 2.2a corresponds to the top two rows of the results shown in Fig. 2.2b. Data for

the other Asp isomers and both Ser epimers are shown in the remaining slots of Fig. 2.2b. All three non-native forms of aspartic acid essentially prevent digestion by cathepsin D. Furthermore, epimerization of the less bulky serine side chain also modulates cathepsin D action, preventing cleavage at one or more preferred sites even when the epimerized serine is located six residues away. Significant residual precursor is detected for all modifications, suggesting decreased affinity for the iso/epi modified peptides in general. Results for analogous experiments conducted in acetate buffer at pH 5.5 with cathepsin L are shown in Fig. 2.2c. The canonical peptide is digested into many peptide fragments, including small di- and tripeptides. Cathepsin L is one of the most aggressive lysosomal proteases and is able to cleave more sites in the iso/epi modified peptides relative to cathepsin D. Furthermore, precursor survival is not observed with cathepsin L. However, the sites where digestion occurs are all shifted well away from iso/epi modified residues in every instance, and the number of peptide fragments observed is still reduced relative to the canonical form. The results from cathepsin L and D reveal that digestion by endopeptidase action is significantly hampered by iso/epi modifications across wide regions of sequence.

The lysosomal task of reducing proteins and peptides into individual amino acids is never completed by endopeptidases, making examination of exopeptidases important. We used a palindromic peptide (RLHTIDITHLR) to systematically explore the limits of exopeptidase activity, and the results for experiments with cathepsins H and B are shown in Fig. 2.2d. For cathepsin B, the canonical sequence is rapidly degraded (CatB upper trace). None of the precursor remains and only a few fragments are detectable. This is

consistent with thorough digestion, producing amino acids or peptides too small to be retained on the column. In contrast, placement of an isomerized residue in the central position, D-isoAsp, halts digestion considerably (CatB lower trace). The most abundant product corresponds to a single cleavage, removal of the C-terminal LR dipeptide. When acting as an exopeptidase, cathepsin B preferentially removes dipeptides.30 Note, endopeptidase activity leads to the bond cleavages observed on the N-terminal side of the peptide. Similar results are obtained for cathepsin H, which behaves as an aminopeptidase, removing a single N-terminal amino acid at a time.30 The native peptide precursor is completely depleted (CatH upper trace), but a few larger peptide fragments remain relative to digestion by cathepsin B. This may relate to reduced affinity or slower progress due to removal of a single amino acid at a time. In any case, the isomerized peptide is digested noticeably less under identical conditions (CatH lower trace). Interestingly, cathepsin H is able to penetrate within one amino acid of the iso/epi residue compared with two for cathepsin B. This can be rationalized because cathepsin H does not need to accommodate two amino acids in the catalytic site. Some endopeptidase activity is also observed for cathepsin H. Similar results were obtained in experiments examining L-Ser versus D-Ser in the central position (see Fig. 2.S14). Taken together, these results illustrate significant disruption of proteolysis by iso/epi modifications for both the major endo- and exo- lysosomal peptidases.

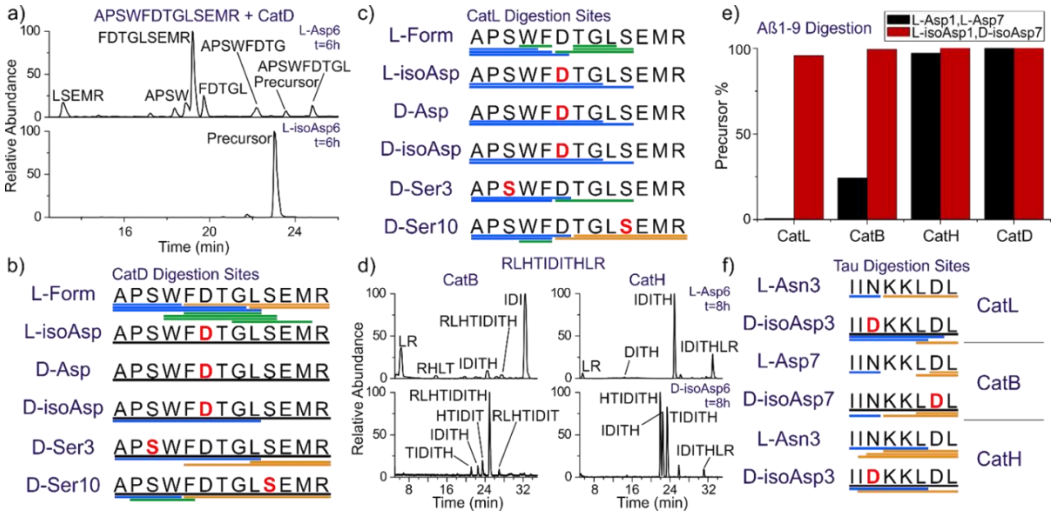


Figure 2.2. a) LC chromatogram for digestion of APSWFDTGLSEMR by cathepsin D. Summary of digestion by b) cathepsin D and c) cathepsin L. Each bar represents a fragment detected in the LC-MS chromatogram, color coded by N-terminal (blue), C-terminal (gold), and internal (green). Undigested precursor >50% relative intensity is represented by a black line. d) LC chromatograms for digestion of RLHTIDITHLR by exopeptidases cathepsins B and H for the native isomer (upper traces) and D-isoAsp isomer (lower traces). e) Summary of digestion of Aβ1-9 (L-Asp1, L-Asp7) vs (L-isoAsp1, D-isoAsp7) by major cathepsins. Only the canonical isomer is digested. f) Summary of digestion of ^{594}IINKKLDL^{601} from Tau using same color scheme.

The results for additional peptide targets relevant to AD are shown in Fig. 2.2e and 2f (Aβ1-9 and Tau ^{594}IINKKLDL^{601}). The two aspartic acids near the N-terminus of Aβ, Asp1 and Asp7, are highly isomerized in amyloid plaques,28 and represent an interesting target where multiple proximal iso/epi modifications can be found. Isomerization of Aβ is known to inhibit serum protease action, suggesting that cathepsins may likewise be stymied.[33] Experiments conducted on canonical Aβ1-9 and a double isomer (L-isoAsp1, D-isoAsp7) are summarized Fig. 2.2e, where the fraction of remaining precursor from each peptide is shown for each cathepsin. Cathepsin B and L easily deplete the precursor for the canonical peptide, but are unable to significantly reduce the amount of precursor for the double isomer. Interestingly, cathepsin D cleaves few sites[34] in Aβ and is unable

to cleave any portion of Aβ1-9 even in canonical form. Similarly, cathepsin H exhibits low affinity for the N-terminal residues in Aβ1-9, and digests the canonical peptide only marginally while leaving the isomerized form intact. The N-terminal portion of Aβ is therefore generally resistant to lysosomal protease action and home to multiple sites of modification that can further frustrate proteolysis, making the prospects for Aβ to contribute to lysosomal failure strong. Experiments on an aged sample of Aβ1-42 yielded similar results (Fig. 2.S13). The highly isomerized N-terminal region was not digested by Cathepsin L while digestion of the C-terminal portion not proximal to any isomerization was cleaved in comparable fashion for both native and aged Aβ1-42.

Tau-mediated pathology is also strongly associated with AD, making it an important target to consider.[35] Asn596 in Tau is known to deamidate,[36] which will yield conversion to Asp and iso/epi modifications according the pathway illustrated in Scheme 2.1. As a long-lived protein, Tau could also isomerize at Asp600. Isomerization at both sites is explored for the peptide fragment ^{594}IINKKLDL^{601} in Fig. 2.2f for cathepsins B, L, and H. The canonical peptide is rapidly consumed for all three cathepsins, but introduction of D-isoAsp at either position significantly perturbs the locations of proteolytic cleavage sites and leads to observation of abundant undigested precursor in all cases. These results reveal that inhibited proteolysis in the vicinity of iso/epi modified residues is likely a general feature for any peptide sequence, and long-lived proteins known to be modified in the brain will be difficult to for the lysosome to break down into amino acids.

Isomer digestion in living cells

To explore additional lysosomal proteases, experiments were conducted with fully active lysosomes in SIM-A9 mouse microglial cells, as shown in Fig. 2.3. For the peptide target, the N-terminal portion of Aβ was selected, and microglial cells were used because they are active participants in the clearance of Aβ within the brain.[37] Chimeric peptides (R_8-E_{edan}DAEFRHDK_{dab}G, where the Glu and Lys have been modified with edans and dabcyl, respectively) consisting of a cell-penetrating portion combined with an Aβ probe sequence were synthesized. Polyarginine was used for cell penetration, which is known to deliver cargo to the lysosome.[38] The probe portion of the peptide remains dark when intact as the edans fluorescence is efficiently quenched by dabcyl. Upon cleavage of the probe sequence, the quencher can separate and edans will emit broadly around 490nm. Results for Aβ1-7 (L-Asp1, L-Asp7) as the probe are shown in Fig. 2.3a, revealing that fluorescence is observed after 150 min as expected. In comparison, the D-isoAsp1/D-isoAsp7 probe yields lower intensity fluorescence in terms of quartile range, median, and number (including exceptionally bright cells), as shown in Fig. 2.3b. Statistical comparison of the results with the Mann-Whitney U test reveals that differences in digestion are significant for all time points. Higher resolution images confirmed that the fluorescence was punctate and overlapping with organelles stained by lysotracker, consistent with delivery to the endosomal/lysosomal system (Fig. 2.S8, 2.S9). Taken together, these results suggest that there is not an unknown protease in the lysosome engineered to digest iso/epi sites.

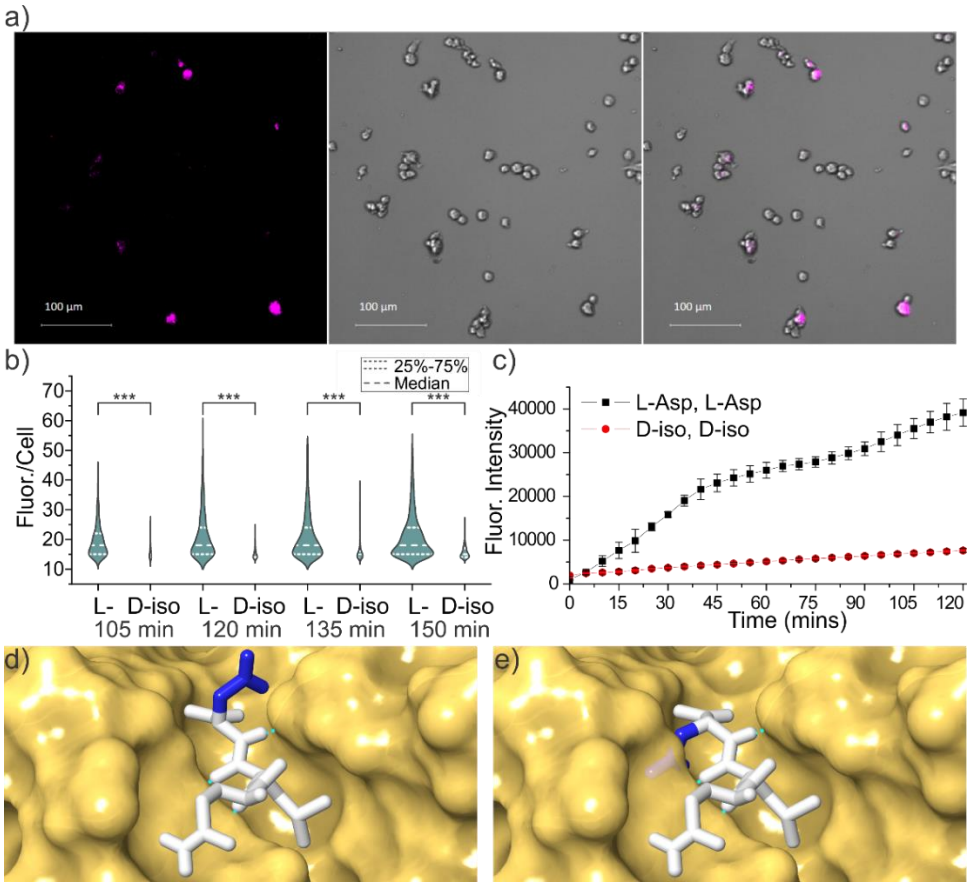


Figure 2.3. a) Sample images of SIM-A9 mouse microglial cells after 150 minute incubation with cleavable peptide target with all L-residues, fluorescence from 481-499 nm (left), brightfield (middle), and overlay (right). b) Violin plot showing quantitative comparison of fluorescence intensity per cell from Aβ1-7 cleavage for canonical and the D-isoAsp1/D-isoAsp7 isomers as a function of incubation time. *** $p < 0.001$ c) Fluorescence intensity as a function of time for incubation of same peptide with cathepsin L only. d) Active site of cathepsin L with native peptide substrate bound and e) mutated epimer with D-Asp sidechain highlighting inherent steric clash if backbone orientation is maintained. Structures derived from PDB ID: 3K24 with hydrogen bonds indicated by green dashed lines.

Interestingly, the microglial results can be largely recapitulated by examination of the same chimeric peptide incubated with only cathepsin L, as shown in Fig. 2.3c and Fig. 2.S1. Both the rates and magnitude of the differential closely match the results obtained in living cells. These findings are consistent with previous observations that cathepsin L is one of the most important lysosomal proteases and can account for ~40% of all protein digestion in the lysosome.30 The accurate reproduction confirms the validity of the LC-MS approach that yielded the results shown Fig. 2.2. Furthermore, the effects of iso/epi modifications are more accurately determined under controlled incubation where canonical peptides without additional modifications can be tested. For example, Aβ1-7 (D-isoAsp, D-isoAsp) itself is almost completely resistant to degradation, yet proteolysis with cathepsin L is increased by a factor of ~7 after decoration with hydrophobic chromophores needed for examination in cells (Fig. 2.S2). This suggests that the difference between digestions shown in Fig. 2.3b is significantly underestimated relative to the true inhibiting power of the D-isoAsp modifications.

The results in Fig. 2.2 and 2.3a,b can easily be rationalized by a molecular level inspection of the interaction between a protease and substrate peptide. In Fig. 2.3d, the X-ray crystal structure for binding of a peptide substrate to cathepsin L is shown.[39] The protease active site consists of a channel where several hydrogen bonds orient the peptide backbone of the substrate. Intimate contact and alignment of the substrate backbone is required to bring the cleavage site into proximity with the catalytic actors. Favorable or unfavorable interactions with side chains protruding above the groove determine the sequence selectivity, but introduction of a D-amino acid with the peptide backbone

remaining properly oriented would result in the side chain projecting directly into the wall of the binding groove (Fig. 2.3e). Similarly, isoAsp modifications disrupt both the backbone hydrogen bond partner spacing and relative orientation (Fig. 2.1), making for an even less tractable situation. These structural alterations make it impossible for iso/epi modified residues to fit properly into the catalytic binding site. Given the similarities inherent in the function and substrate for every protease, comparable complications are likely to exist for all proteases intended to cleave peptides comprised solely of canonical L-residues. Perhaps it is not surprising that poor proteolysis is observed for iso/epi modified peptides even in glial cells where a full complement of lysosomal proteases is available.

Timeframe for aspartic acid isomerization

Given that Aβ plays an important role in Alzheimer's disease (AD) and is highly isomerized in amyloid plaques,[28,40] we set out to determine the incubation times needed to yield such extensive modifications. Following incubation of Aβ 1-40, 1-42, and 1-9 in tris buffer at 37ºC, the degree of isomerization was measured and the results are shown as a function of time in Figs. 4a and 4b. To quantitate the isomerization of Asp1 and Asp7 independently, aged Aβ 1-40 and 1-42 were first digested with chymotrypsin, yielding ^{1}DAEF^{4} and ^{5}RHDSGY^{10} peptides, which were subsequently analyzed by LC-MS (see Fig. 2.S5, 2.S6). Isomerization occurs rapidly at both aspartic acids for both full length peptides, yielding roughly 14% combined isomerization within 30 days. This rate is comparable to previous examination[41] of Aβ1-16 and to isomerization of Asp151 in αA

crystallin (when determined for the peptide fragment ^{146}IQTGLDATHAER^{157}). [42] It is also consistent with other isomerization rates cited in the literature as shown in Fig. 2.4c,10,[43-45] where the only significantly faster rates involve Asp-Gly sequences. Detailed study of deamidation, which forms an identical succinimide ring intermediate preceding isomerization, revealed the fastest rates for analogous Asn-Gly sites.[46]

These experiments were conducted at μM concentrations, which is sufficient for the formation of amyloid fibrils. The presence of amyloid was examined by ThT assay after 7 days as shown in Fig. 2.4d. The assay reveals that Aβ1-42 had already formed fibrils within 7 days, while Aβ1-40 was just entering fibril formation, consistent with previous reports.[47] After digestion with chymotrypsin, the fluorescence diminishes substantially, suggesting that fibrils are broken up and should not significantly influence the analysis. Interestingly, amyloid formation appears to slightly increase the rate of isomerization for Asp1 but in general does not significantly influence the rates. This conclusion is further supported by the observation that the rates do not vary greatly from the results obtained for Aβ1-9, which does not form fibrils.41

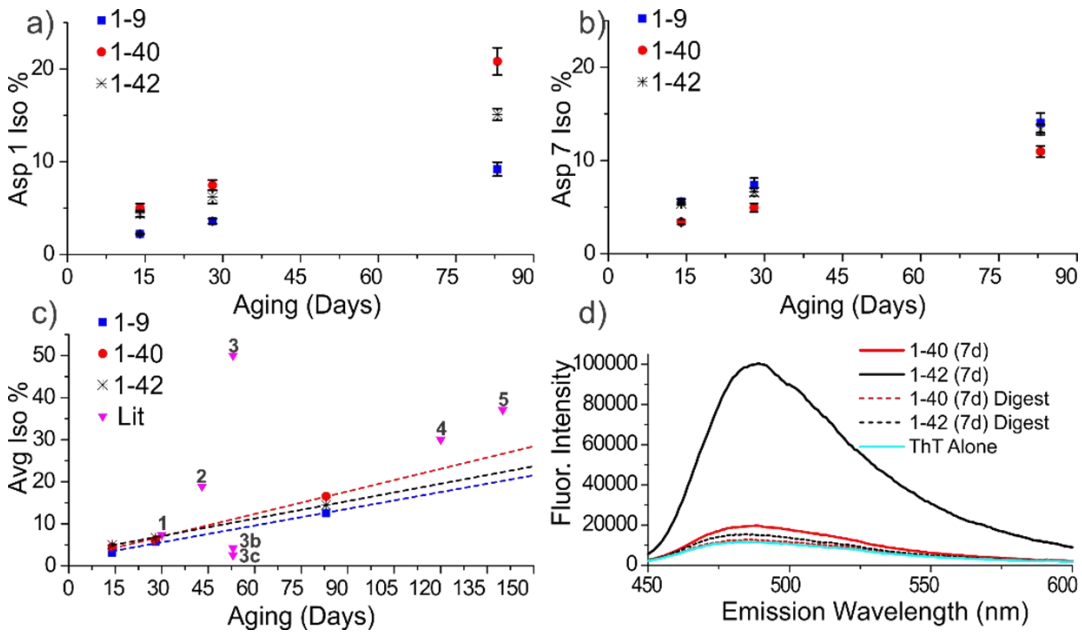


Figure 2.4. Isomerization % as function of time for a) Asp1 and b) Asp7. c) Average isomerization rate for Asp1 and Asp7 relative to rates from literature. d) ThT assay after 7 days confirming that any fibrils are largely digested during analysis. Data points: 1,43 2,42 3,44 3b,c (Estimated rate of the VYPDGA peptide from Literature point 3 modified to correspond to VYPDSA and VYPDAA based on known deamidation rates.46), 4,45 and 5.46

A framework connecting lysosomal failure and AD

Long-lived proteins are subject to many spontaneous chemical modifications, including subtle changes such as iso/epi modifications that may seem harmless and are easily overlooked. Nevertheless, heavy isotope pulse-chase experiments in mice have shown that long-lived proteins in the brain are more commonplace than previously realized and can persist for timespans exceeding one year.[48] These long-lived proteins are part of the overall equation that must be balanced to maintain proteostasis and will therefore be targeted for degradation at some point. Our results reveal that isomerized and epimerized sites in long-lived proteins resist digestion by the primary cathepsins present in lysosomes. Both epimerization (Ser and Asp) and isomerization (Asp) effectively

prevent proteolysis at the site of modification and nearby residues for both endo- and exopeptidases. Long-lived proteins targeted to the lysosome are therefore expected to produce residual peptide fragments that are too long to be recognized by the transporters responsible for releasing digested amino acids back to the cytosol. Additionally, the residual peptides will contain an unnatural amino acid that would be expected to further frustrate transporter recognition. Accumulation of these byproducts within the lysosomal machinery is therefore possible. In fact, interference with lysosomal function has already been documented in similar circumstances with pyroglutamate modified Aβ, where the influence on proteolysis is significantly less pronounced.[49]

We have demonstrated that iso/epi modifications significantly inhibit lysosomal digestion in glial cells, but prior work has additionally shown that such modifications are toxic. Makarov and coworkers have examined isomerization of the N-terminal portion of Aβ in relation to the idea that such modifications enhance amyloid formation in the presence of zinc ions. They found that isomerized Aβ1-42 was more toxic than the canonical form when incubated with several different cell lines (NSC-hTERT, SK-N-SH, and SH-SY5Y).[50] Furthermore, cell death by apoptosis rather than necrosis was more prevalent in the case of isomerized Aβ, indicating an alternate and more specific mechanistic pathway. Importantly, related experiments have demonstrated that Aβ localizes into the lysosome when incubated with SH-SY5Y cells,[51] suggesting that the toxicity could be reasonably attributed to lysosomal pathology instead. Toxic effects have also been found in animal studies.[52] Perhaps most strikingly, injection of isomerized Aβ1-16 leads to significantly increased amyloid plaque accumulation in 5XFAD

transgenic mice whereas canonical Aβ1-16 does not.[53] Importantly, Aβ1-16 does not contain the amyloid forming portion of the peptide.[54] Although these data could be interpreted to support to the zinc-mediated amyloid aggregation hypothesis, our findings suggest that disruption of the lysosomal system could also explain the results. Introduction of isomerized Aβ1-16 could lead to lysosomal failure, followed by disrupted proteostasis and the observed increase in amyloid plaque formation.

We have established that isomerization of Aβ is relatively fast. The residence time of Aβ in the human brain is difficult to determine due to the multiple destinations and pathways that can be taken, but studies have shown that the fraction of Aβ escaping into cerebrospinal fluid persists beyond 30 hours in a healthy individual.[55] Similar studies have shown that clearance rates for Aβ are mismatched relative to production in AD individuals,[56] which suggests that some fraction evades degradation and may persist for longer times. The rates in Fig. 2.4 allow for a small degree of isomerization (~0.2%) even within a 30 hour timeframe. Furthermore, any fraction of Aβ residing in the brain for a week or more would be expected to isomerize significantly. The N-terminal region of Aβ is disordered in amyloid structures determined by NMR,[54,57] which may allow free access to the required succinimide intermediate while providing some catalytic interactions that favor isomerization. Aβ is therefore a likely source of isomerized residues in the brain, but a few reports have shown that tau can also be isomerized due to deamidation at positions 596 and 698, or isomerization of Asp at positions 510 and 704.[58,59] The size and largely unstructured nature of tau[60] make it almost certain that other sites of isomerization also exist. There is ample evidence that the proteins most strongly

associated with AD pathology are subject to iso/epi modifications that could lead to lysosomal failure.

Conclusions

Iso/epi modifications are clearly generated on a relatively short timescale and prevent cathepsin digestion of nearby peptide bonds. Although other proteolytic pathways exist within cells that may also encounter difficulties with iso/epi modifications, lysosomes are uniquely vulnerable because undigested byproducts cannot escape the lysosomal membrane and can eventually cause failure and storage of the entire organelle. When this sequence of events is triggered in lysosomal storage disorders, the consequences are dramatic and often fatal. Malfunction of the lysosome is also strongly associated with the pathology of AD, as are misfolding and aggregation of both Aβ and Tau. Lysosomal failure caused by the iso/epi modifications documented to exist in both Aβ and Tau offers a direct connection between these observations and a potential new pathway to explore for the underlying cause and treatment of AD.

Acknowledgements: The authors are grateful for funding from the NIH (R01GM107099 to RRJ, and R01NS091616, R21NS106949, R25GM119975 to BDF). Min Xue is kindly thanked for allowing us to use his fluorescent plate reader. Hill Harman, Gal Bitan, Pablo Martinez, and Joe Loo are acknowledged for helpful discussions.

Supporting Information

Materials and Methods

Peptide Synthesis All peptides, including those with appropriate iso/epi modifications, were synthesized following an accelerated FMOC-protected solid-phase peptide synthesis protocol. Isomerized peptides were synthesized by incorporating the appropriate isomeric form of the amino acid residue during synthesis. Fluorophores Edans and Dabcyl were purchased as FMOC-protected amino acid conjugates and attached to peptides during synthesis. Following synthesis, peptides were purified on a Varian ProStar HPLC using a Phenomenex Jupiter Proteo C12 4 μm 90 Å 250 mm × 4.6 mm column and then validated by LC/MS. Purified peptides were stored frozen in 50/50 acetonitrile/water (v/v) and vacuum-dried before reconstitution in appropriate buffers.

In Vitro Aging

Lyophilized amyloid beta powder was disentangled via ammonium hydroxide treatment. Following disentanglement, aliquots containing 100 μg amyloid beta were lyophilized before reconstitution in 1 mL 50 mM Tris 7.4 pH at 37 °C. Samples were aged in microcentrifuge tubes in a sand-bath style incubator. Aliquots were removed at given timepoints and frozen until analysis. Aged samples were digested at room temperature for 18 h with chymotrypsin in 200 mM Tris buffer with 4 mM $CaCl_2$ and an enzyme:substrate weight ratio of 1:2 to ensure complete digestion of fibrils. Digested samples were frozen until analysis.

LC-MS Analysis

Samples were analyzed on an Agilent 1100 HPLC or a Thermo Fisher Easy nLC II interfaced with a Thermo Fisher LTQ or Thermo Fisher Velos Pro Orbitrap using electrospray ionization sources. For the Agilent 1100, peptides were separated on a Thermo Scientific BetaBasic C-18 3 μm, 150 x 2.1 mm column using water with 0.1% formic acid for mobile phase A, and methanol with 0.1% formic acid for mobile phase B. The Easy nLC II was operated with a capillary column packed in-house to a length of ~10 cm with Waters Acuity CSH C18 1.7 μm resin using a Shotgun Proteomics Inc high pressure vessel. For the Easy nLC II, mobile phase A was water with 0.1% formic acid, and mobile phase B was acetonitrile with 0.1% formic acid.

In Vitro Fluorimetry

Aliquots containing 75 μg of purified peptide were added to a 96-well plate in 50 mM acetate buffer pH 5.5 with 1 mM EDTA and 1 mM DTT to prevent active site oxidation. Cathepsin L was added at an enzyme:substrate ratio of 1:150. Digestion was then monitored using a Biotek Synergy H1 fluorescent plate reader with an excitation wavelength of 336 nm and emission wavelength of 490 nm. Fluorescence is plotted as the average of 3 runs and error bars represent the standard deviation.

The presence of amyloid fibrils was examined by ThT assay. Peptides were diluted to 2 μM in Tris buffer before addition of 30 μM ThT. Emission scans were performed on a QuantaMaster-400 fluorimeter, using an excitation wavelength of 440 nm.

In Vitro Digestion

Aliquots containing 75 μg of purified peptide were dissolved in 150 μL of buffer. Digestions utilizing cysteine cathepsins (CatL, CatB, CatH) were performed in 50 mM acetate buffer pH 5.5 with 1 mM EDTA and 1 mM DTT to prevent active site oxidation. Lower pH did not improve digestion (Fig. S11). Digestions with CatD were performed in 50 mM acetate buffer pH 4.5 to ensure protonation of the active site aspartic acid. Enzyme ratios were adjusted based on activity; CatL and CatB were used at an enzyme:substrate ratio of 1:150, CatD 1:50, and CatH 1:25. Increasing relative enzyme concentration and digestion time did not enable digestion of refractory sites (Fig. S12). Timepoint aliquots were taken at 0.5, 2, 4, 6, 8, 24, and 48 hours, and diluted to a final concentration of 5 μM with 0.2% TFA solution to pH quench the enzymatic reaction. Digested samples were frozen until analysis.

Microglial Cell Culture and Imaging

SIM-A9 microglial cells were seeded into 8-well chamber slides (Thermo Scientific, 154941) at a density of 2x 10^5 cells/cm^2 in complete media (DMEM:F12 + 10% FBS + 5% Horse serum + 1% penn/strep) and allowed to attach for 12 hours under standard conditions (37°C, 5% CO_2). At the start of experiments, 10 μM of either L- or D-isoAsp peptide was added. Digestion of the cell-penetrating peptide isomers was monitored in live cells on a Zeiss 880 laser scanning confocal microscope. The sample chamber was maintained at 37 °C for the duration of the live cell experiments. Images were collected using a Zeiss Plan-Apochromat 10x/0.45 M27 objective. Eight 850 μm x 850 μm images were collected and tiled as two sets of 2x2 images. Fluorescence quantitation is derived

from the full image. The data presented in Fig 2.3b encompasses 936 cells for the L-Asp peptide incubation (average = 234 per image, stdev = 36) and 1282 cells for the D-isoAsp peptide incubation (average = 320 per image, stdev = 51). Data from a second location Fig. 2.S1c which was also collected as 2x2 tiled images, and encompasses 986 cells for the L-Asp peptide incubation (average = 247 per image, stdev = 26) and 695 cells for the D-isoAsp peptide incubation (average = 174 per image, stdev = 20). Fluorescent signal was plotted as fluorescence per cell after establishing a signal threshold based on the negative control. Bright-field images were collected for cell-counting, and digestion was monitored by recording Edans derived fluorescent emission from 481-499 nm. Following imaging, cells were counted in FIJI, using a threshold image approach that was validated manually. Raw fluorescence data was recorded for both peptide isomers in addition to a negative control (i.e. cells without peptide). To eliminate background noise, scattering, and autofluorescence, only fluorescent signal that is above the negative control was used for quantitation. For cells with lysosomal labels, Invitrogen Lysotracker Red DND-99 was used in accordance with the manufacturer's instructions. Lysotracker was excited at 561 nm and emission was recorded from 580-735 nm.

Cell Fixing and Immunohistochemistry (IHC)

At fixation timepoints, cells were briefly washed with cold PBS. Cells were then fixed in 4% paraformaldehyde for 30 minutes at room temperature. After fixation, paraformaldehyde was removed and replaced with PBS. Fixed microglia were stained for ED-1 (CD86) with a secondary antibody conjugated to AlexaFluor 488 to aid in visualization. Alexafluor was excited at 488 nm and emission was recorded from 499-562

nm. Slides were kept at 4°C until viewing using the Zeiss 880 equipped with a Plan-Apochromat 63x/1.4 Oil DIC M27 objective.

Molecular Modeling

The active site of cathepsin L was modeled in Maestro 11 (Schrödinger, LLC). The PDB deposit, 3K24, was processed to include H-bonds and displayed using the surface representation. The QLA peptide bound in the crystal structure was modified to QLD and evaluated for potential H-bonds as well as steric clashes. The chirality of the alpha-carbon of the Asp residue was inverted to generate the D-Asp isomer.

Safety Statement

No unexpected or unusually high safety hazards were encountered.

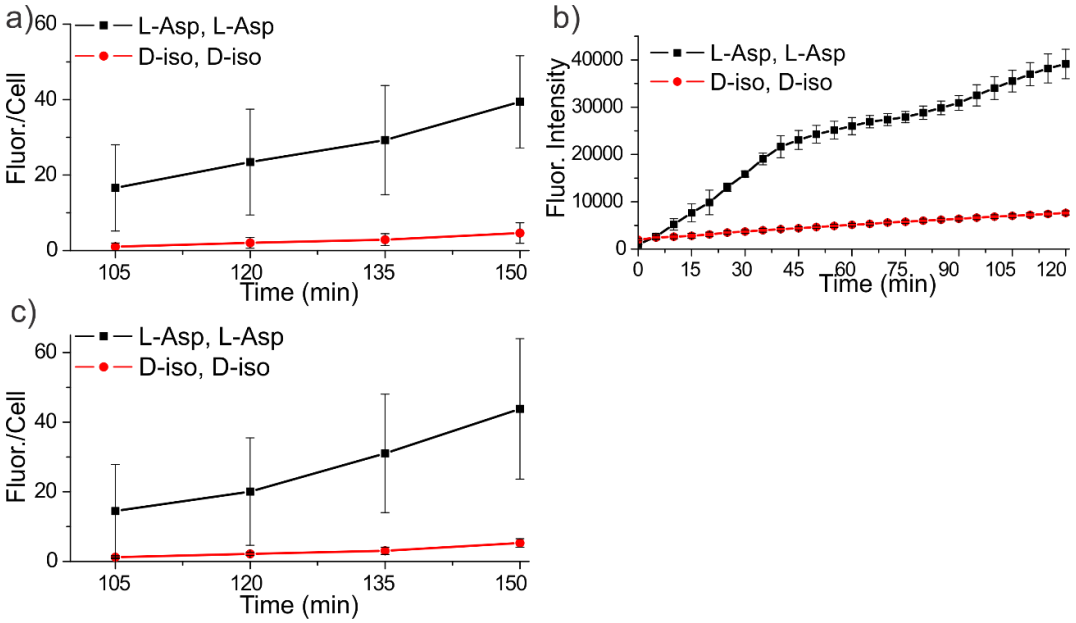


Figure 2.S1. Comparison of the microglial digestion rate data from Fig 2.3b replotted in the same format as the data shown in Fig 2.3c. a) replotted data, b) original Fig 2.3c for easy comparison. Values represent averages with error bars showing standard deviations. c) Similar results obtained from a second, independent location.

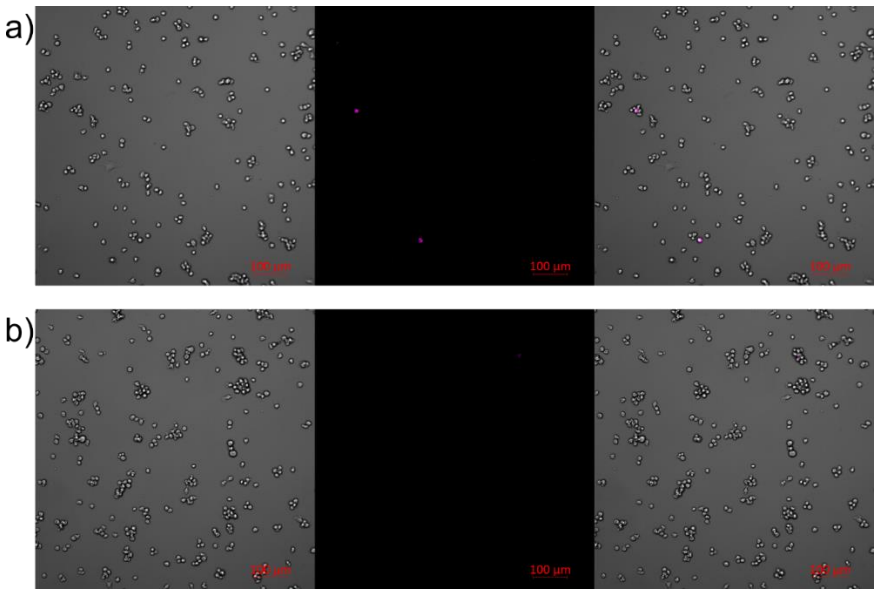


Figure 2.S2. Representative microscopy images of the brightfield, fluorescence emission, and merge (left, center, and right) depicting microglia incubated with the chimeric FRET peptide (R_8-E_{edan}DAEFRHDK_{dab}G in the L-Asp form (a) and the D-isoAsp form (b).

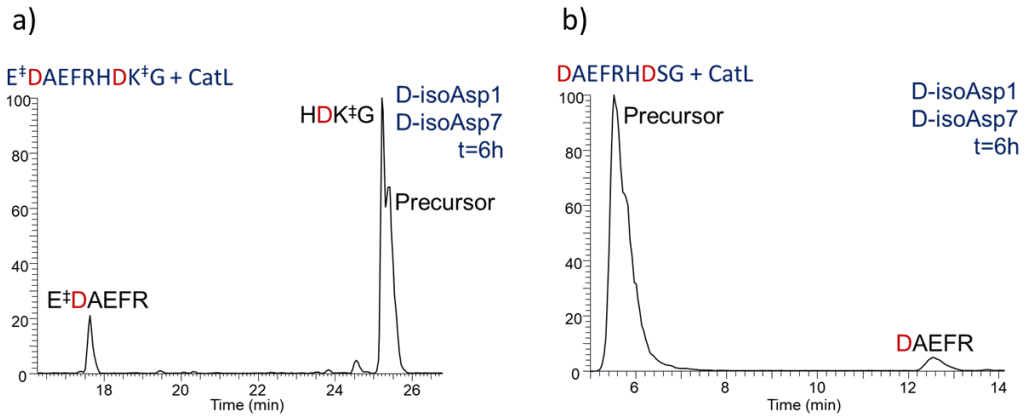


Figure 2.S3. Comparison of digestion of a) chromophoric and b) native-state peptides by LC-MS. Digestion proceeds more rapidly after addition of chromophoric probes for cell imaging.

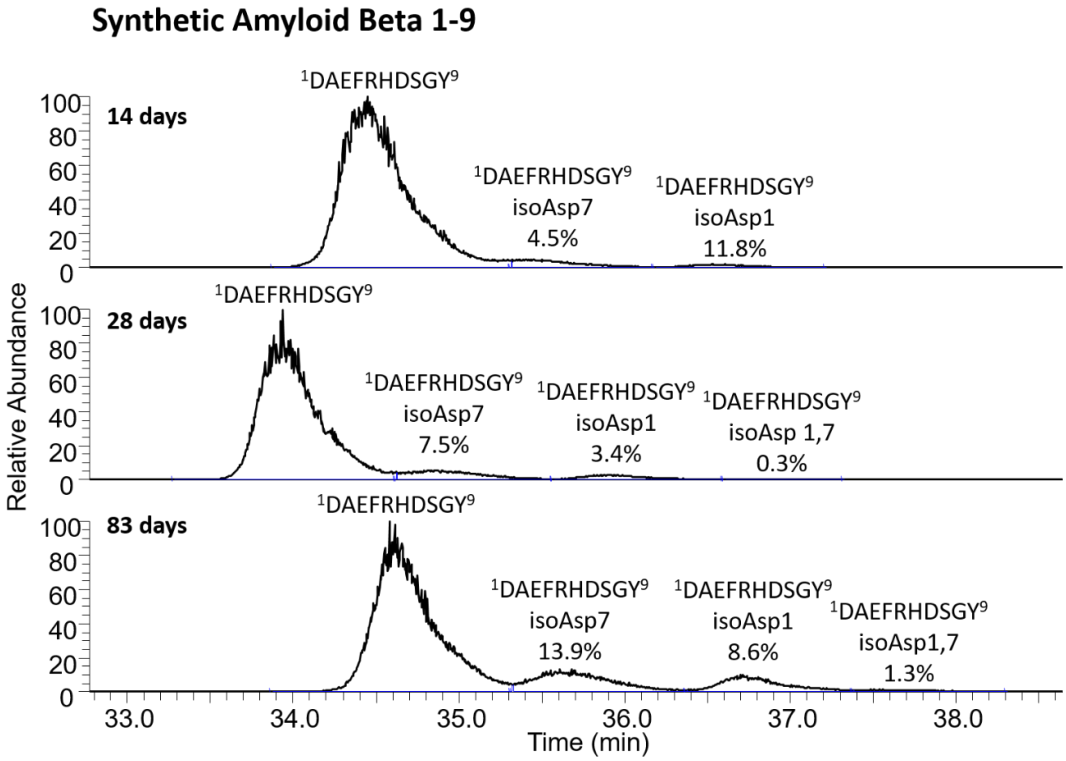


Figure 2.S4. LC-MS chromatograms corresponding to the aging data for synthetic amyloid beta 1-9 as presented in Figure 2.2.4. The isomerization rates are calculated based on LC-MS peak area.

Note: the synthetic amyloid beta 1-9 peptide yields sufficient chromatographic separation to quantitate the isomerization rate at both Asp 1 and Asp 7 simultaneously (as shown in Fig. 2.S4). Conversely, the full-length amyloid beta 1-40 and 1-42 proteins require digestion prior to isomer analysis via LC-MS. To quantitate the isomerization rate at Asp 1 and Asp 7, the full-length amyloid protein was aged intact, and then digested with chymotrypsin immediately before analysis. Quantitation is achieved by analyzing the relative peak area of the chymotryptic peptide products, ^{1}DAEF^{4} and ^{5}RHDSGY^{10} as depicted in Fig. 2.S5 and 2.S6 respectively.

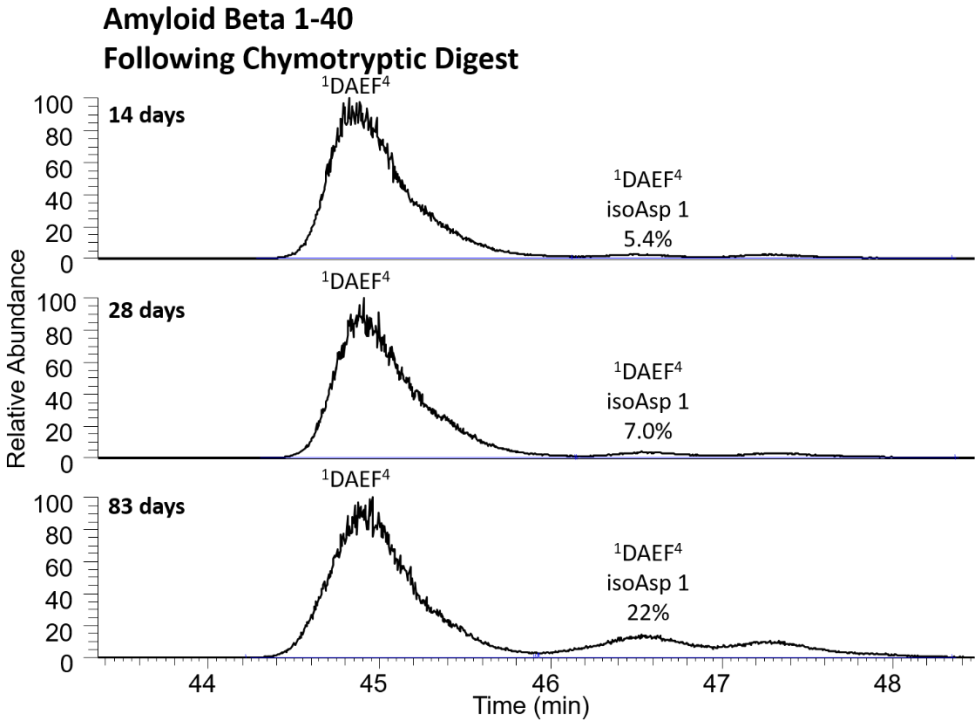


Figure 2.S5. LC-MS chromatograms corresponding to Aβ 1-40 isomerization rate data at Asp1 presented in Fig. 2.4. Following intact protein aging, Aβ was digested with chymotrypsin, and the isomerization rate was determined by LC-MS peak area for the digested peptide isomers.

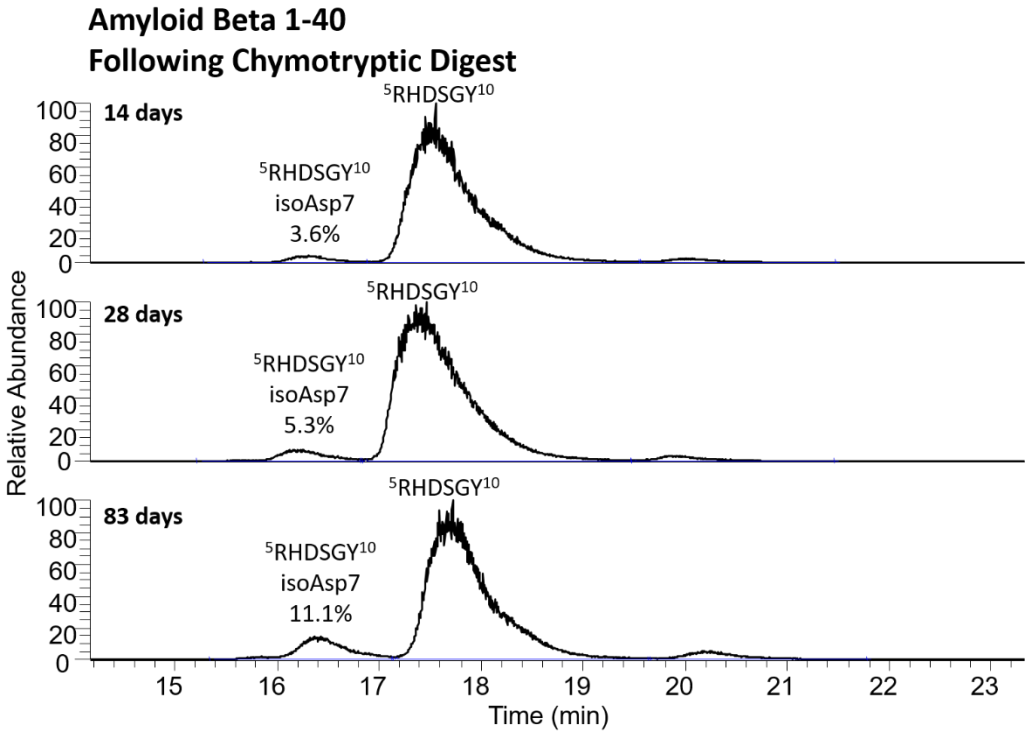


Figure 2.S6. LC-MS chromatograms corresponding to Aβ 1-40 isomerization rate data at Asp7 presented in Fig. 2.4. Following intact protein aging, Aβ 1-40 was digested with chymotrypsin, and the isomerization rate was determined by LC-MS peak area for the digested peptide isomers.

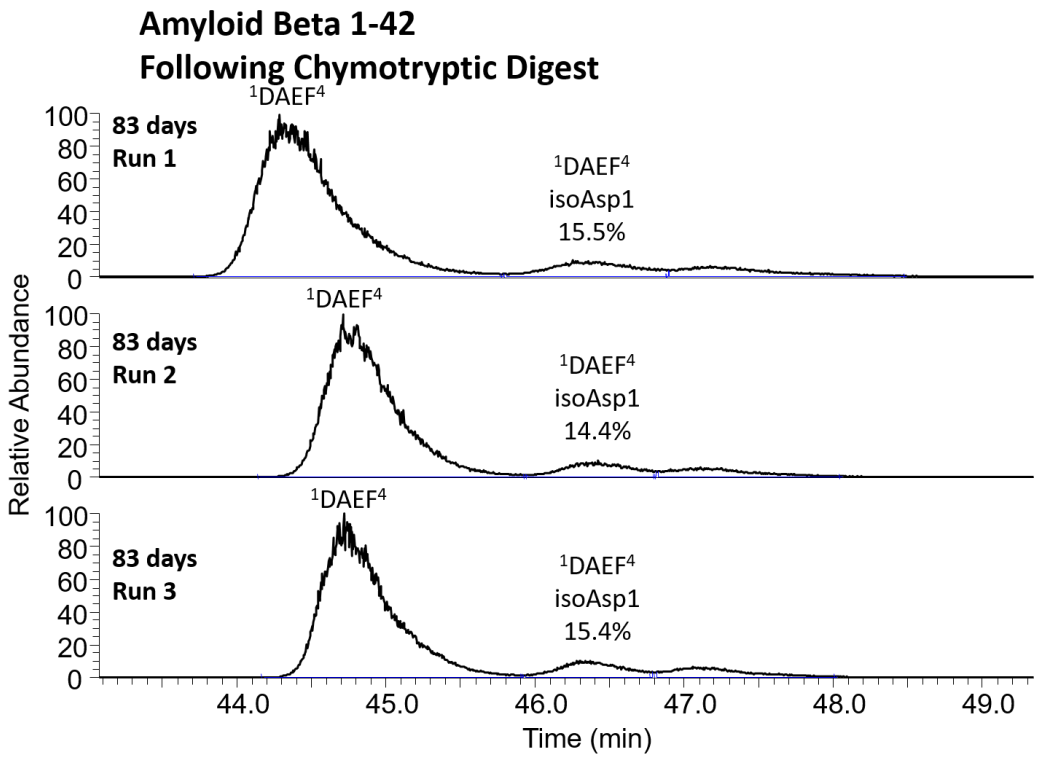


Figure 2.S7. Representative triplicate LC-MS chromatograms corresponding to Aβ 1-42 isomerization rate data at Asp1 presented in Figure 2.2.4. Following intact protein aging for 83 days, Aβ was digested with chymotrypsin, and the isomerization rate was determined by LC-MS peak area for the digested peptide isomers. These triplicate runs yield an average of 15.1% isomerization with a standard deviation of 0.64%.

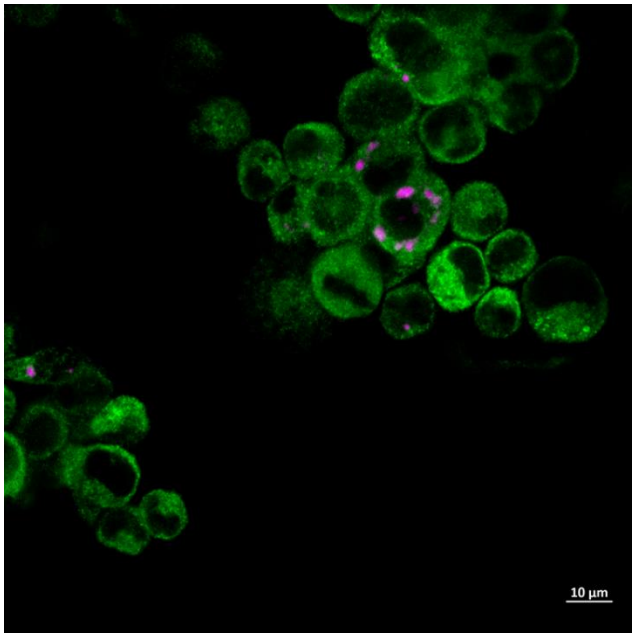


Figure 2.S8. Image of microglial cells under 63x magnification after 3 hour incubation with R_8-*E*$_{edan}$DAEFRHD*K*$_{dab}$G. Fluorescence (pink) is clearly observed in puncta, consistent with localization in lysosomes.

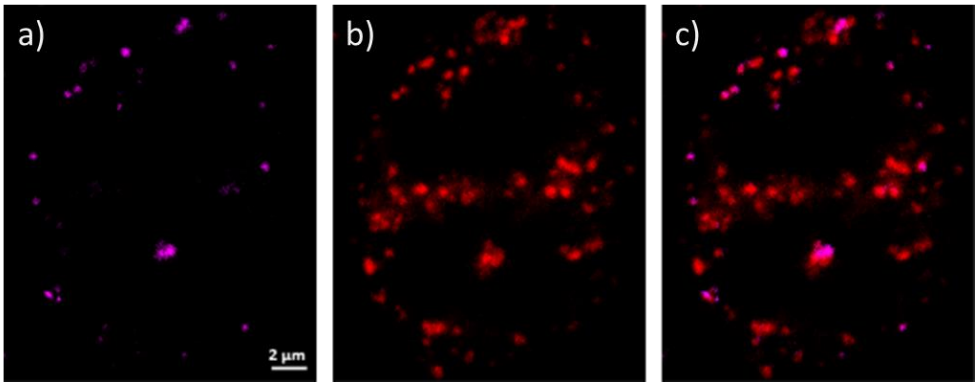


Figure 2.S9. Images of microglial cells under 63x magnification after 6 hour incubation with R_8-*E*$_{edan}$DAEFRHD*K*$_{dab}$G and lysotracker. a) Fluorescence of cleaved peptides. b) Fluorescence from lysotracker. c) Merged image illustrating co-localization. Cleaved peptides originate in puncta that are positive for staining with lysotracker.

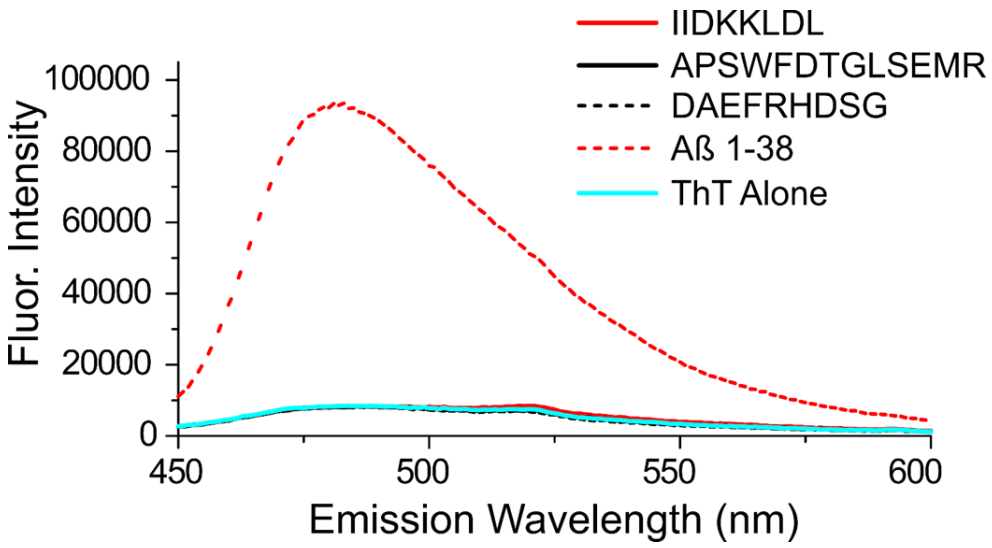


Figure 2.S10. Thioflavin-T (ThT) assay of model peptides to test for aggregation in acetate buffer after 30 minute incubation. Aβ 1-38 was used as a positive control. Lack of fluorescence for model peptides above the ThT background indicates that no aggregation is occurring prior to digestion.

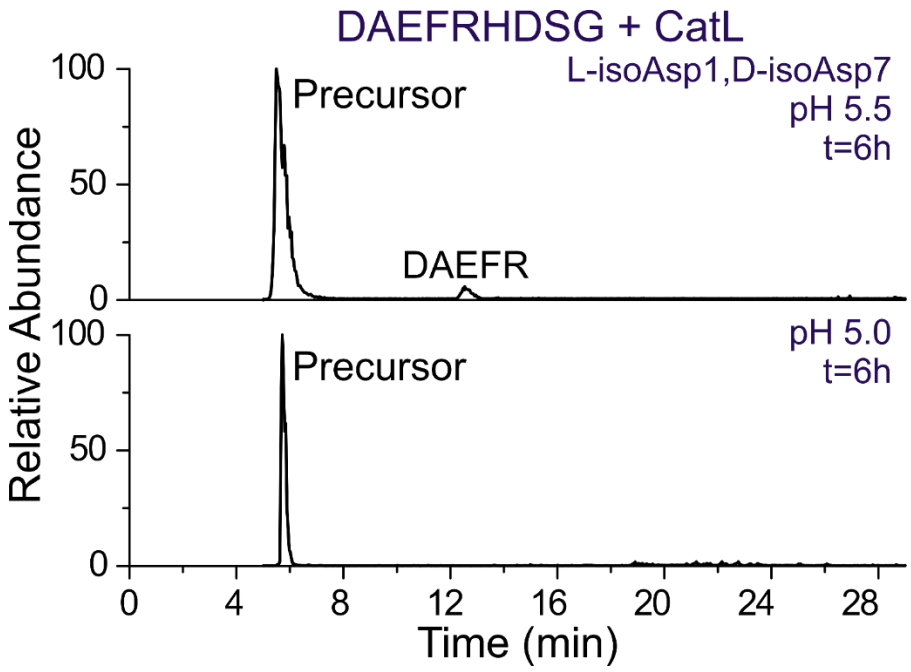


Figure 2.S11. Comparison of digestion efficiency at varying pH. Digestion of Aβ 1-9 by cathepsin L at pH 5.5 (upper trace) and pH 5.0 (lower trace). Less proteolytic cleavage is observed at the lower pH.

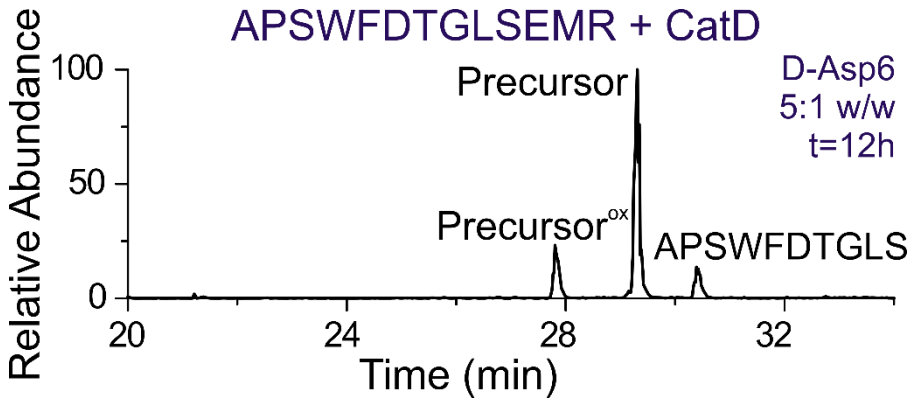


Figure 2.S12. Digestion of APSWFDTGLSEMR by cathepsin D at a lower substrate:enzyme ratio than used for Figure 2.2.2. Undigested precursor indicates that high concentrations of protease are unable to completely digest iso/epi modified peptides.

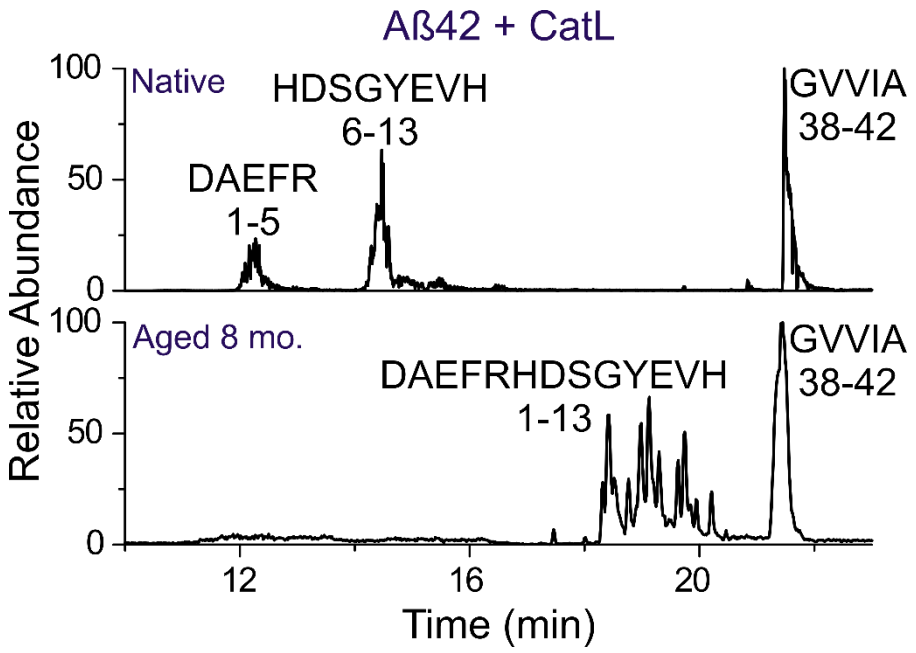


Figure 2.S13. Extracted ion chromatograms are shown for digestion by cathepsin L of: native all-L Aβ1-42 (upper trace), and Aβ1-42 aged for 8 months in Tris buffer at 37°C (lower trace). Ion intensity is extracted for: ^{1}DAEFR^{5}, ^{6}HDSGYEVH^{13}, ^{1}DAEFRHDSGYEVH^{13}, and ^{38}GVVIA^{42}. Abundant cleavage near Asp7 is observed in the native stock (upper trace), producing ^{1}DAEFR^{5} and ^{6}HDSGYEVH^{13} that are much less abundant in the aged sample (lower trace). The inverse relationship holds true for the abundance of ^{1}DAEFRHDSGYEVH^{13}, which is detected in numerous peaks due to isomerization being possible at Asp1 and/or Asp7. Isomerization of Asp7 prevents cleavage between residues 5 and 6. In contrast, the proteolytic fragment ^{38}GVVIA^{42} is located at the C-terminus in a region unaffected by isomerization. ^{38}GVVIA^{42} is detected in similar abundance in both spectra, indicating that only cleavage near isomerized residues is hindered.

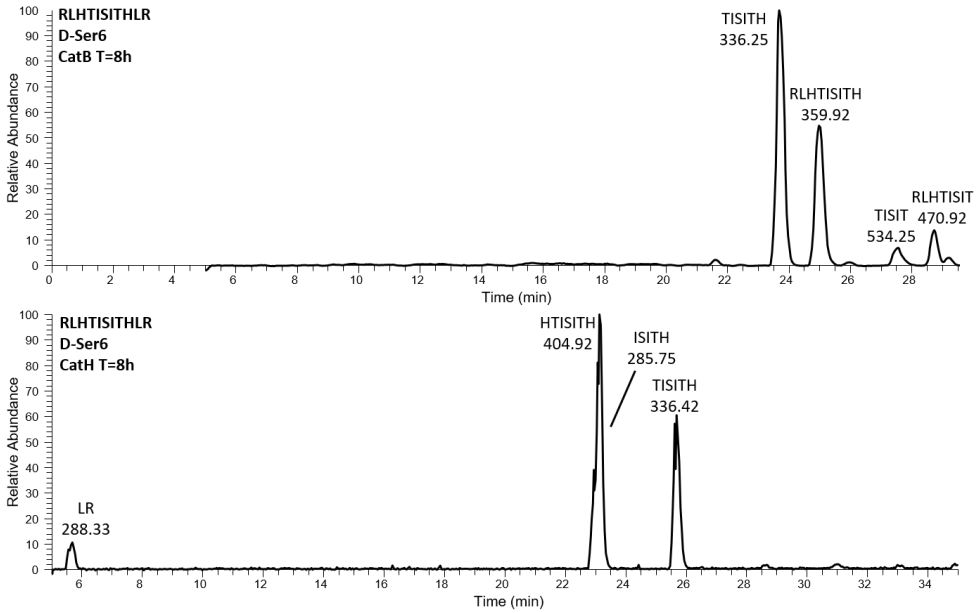


Figure 2.S14. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

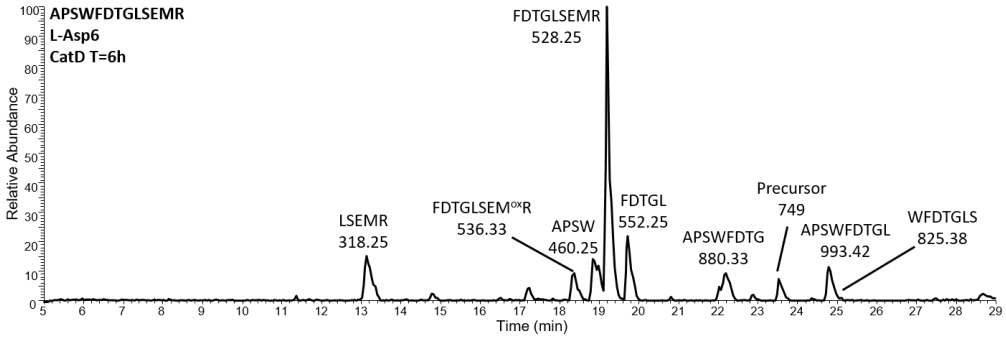


Figure 2.S15. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

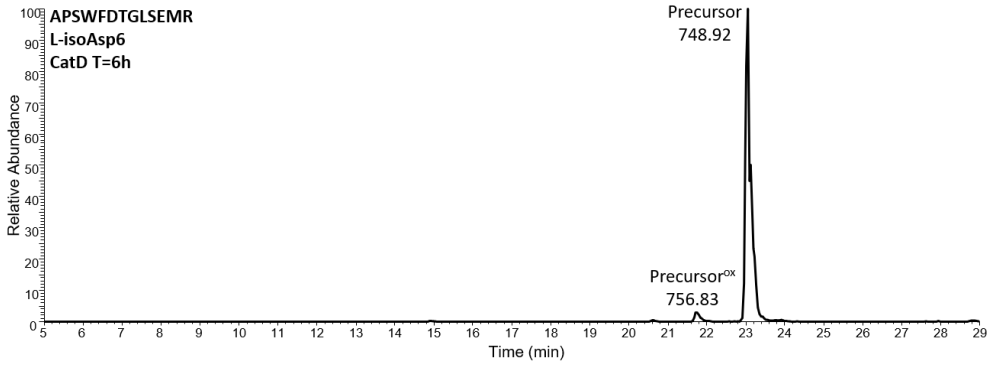


Figure 2.S16. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

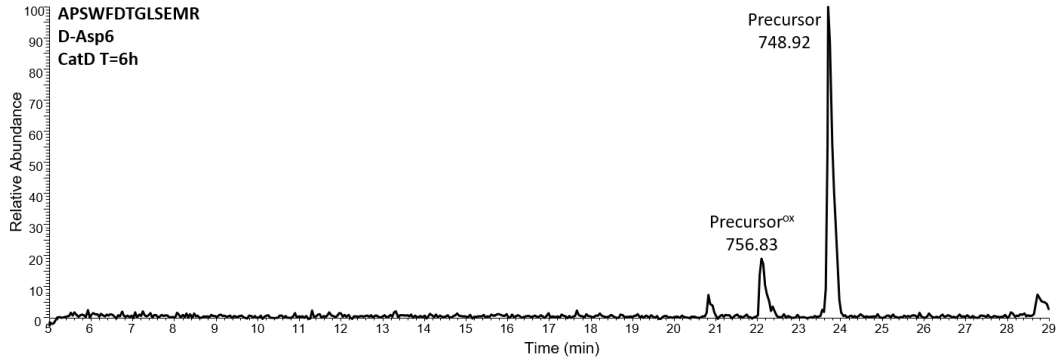


Figure 2.S17. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

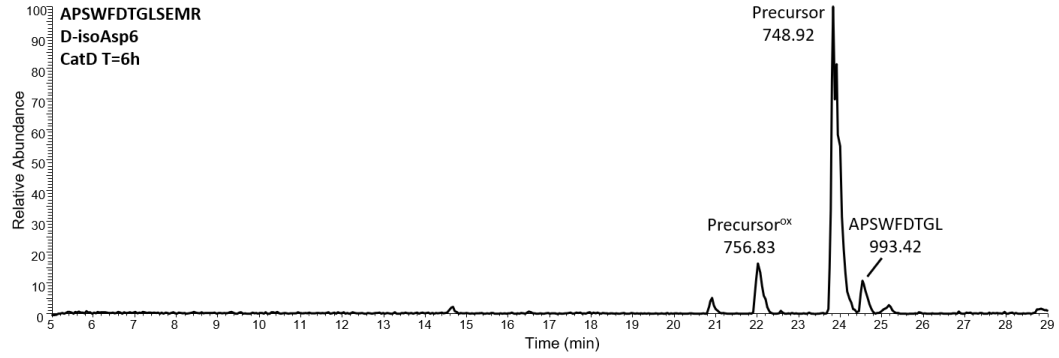


Figure 2.S18. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

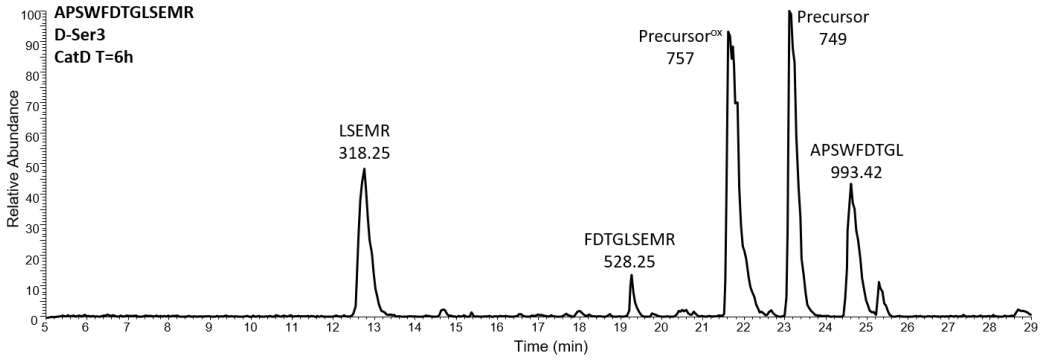


Figure 2.S19. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

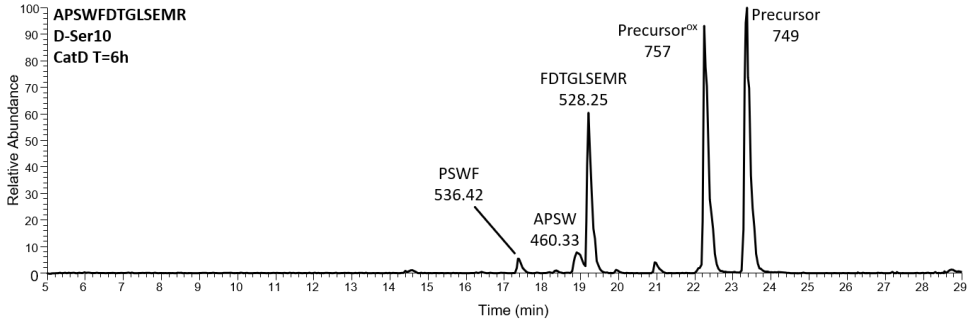


Figure 2.S20. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

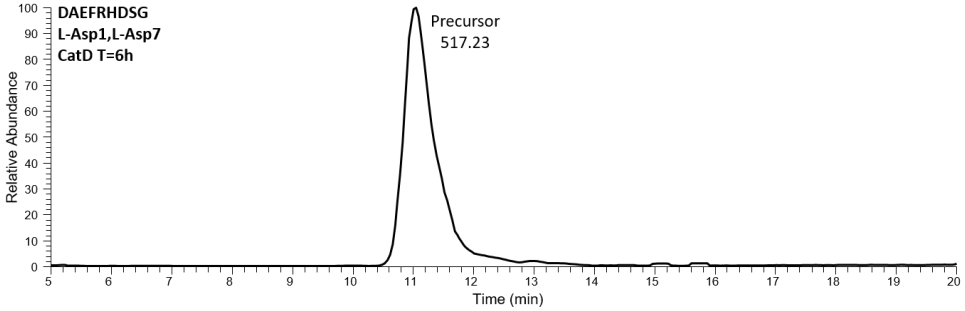


Figure 2.S21. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

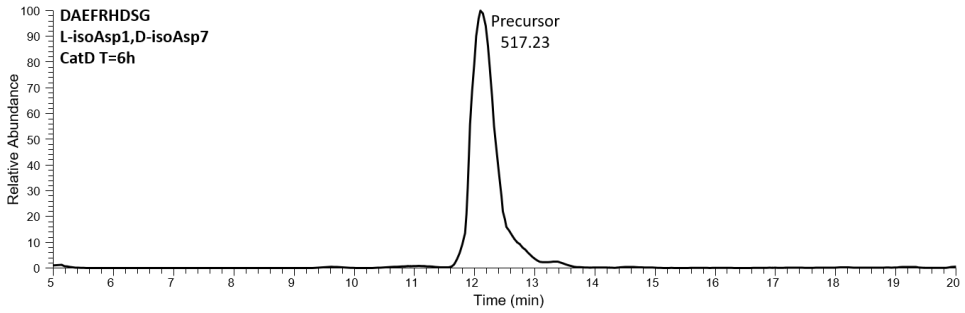


Figure 2.S22. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

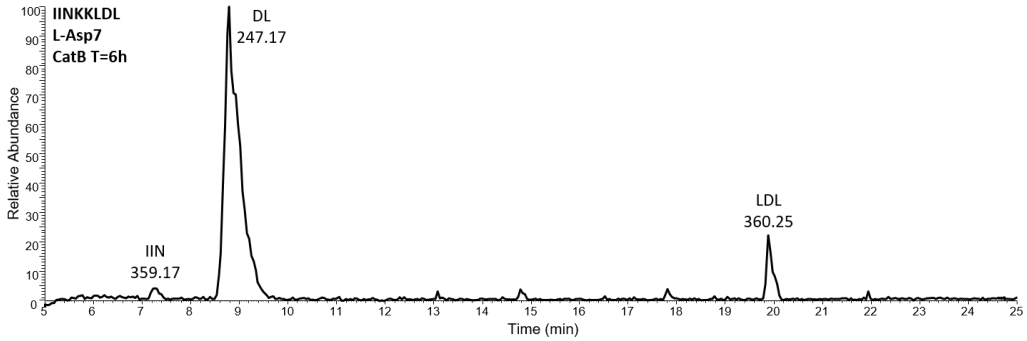


Figure 2.S23. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

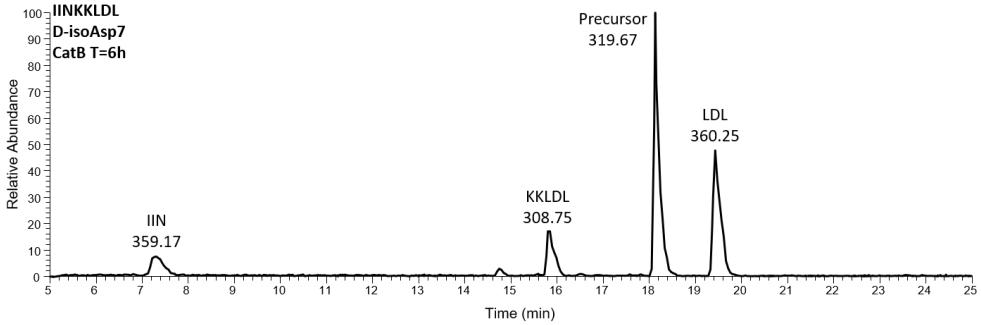


Figure 2.S24. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

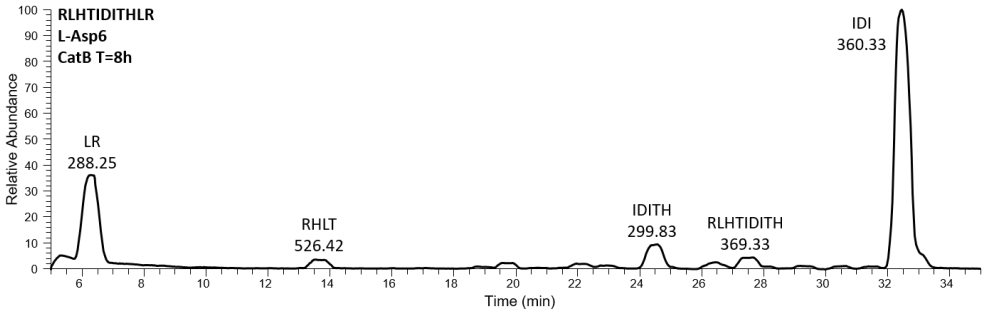


Figure 2.S25. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

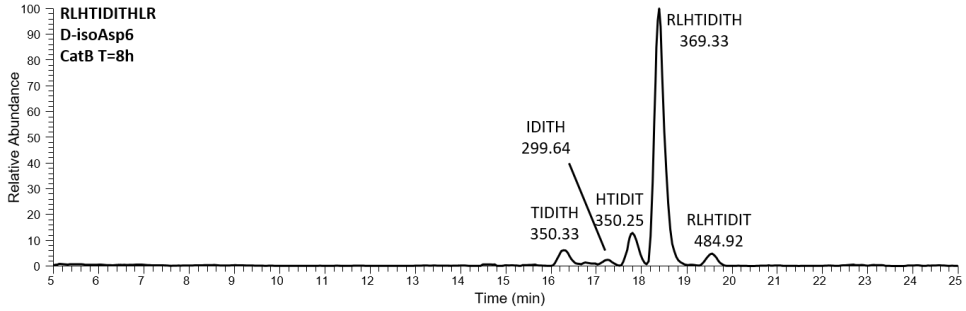


Figure 2.S26. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

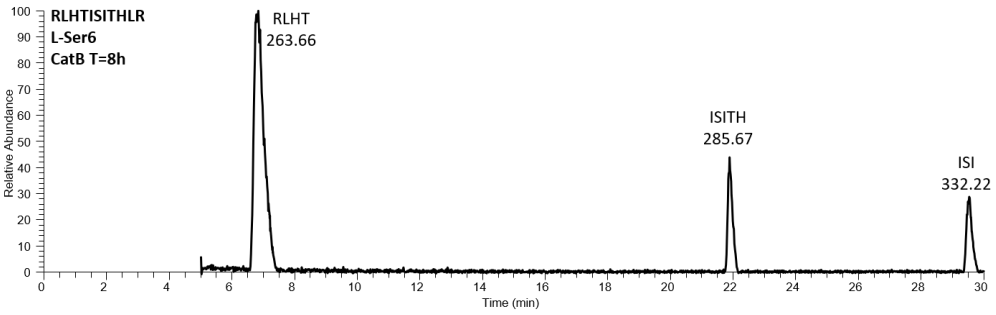


Figure 2.S27. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

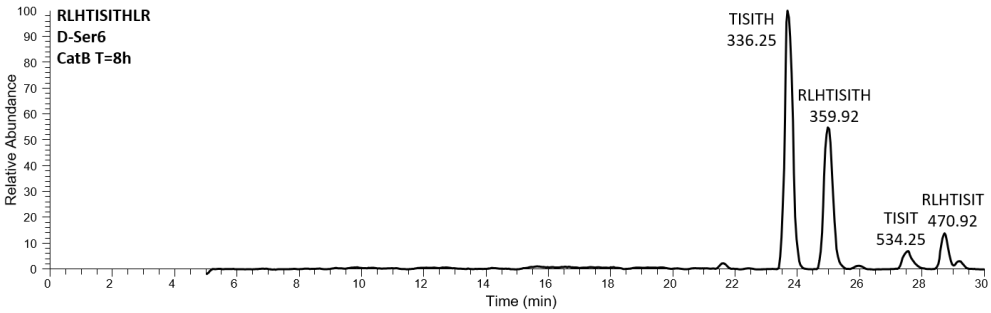


Figure 2.S28. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

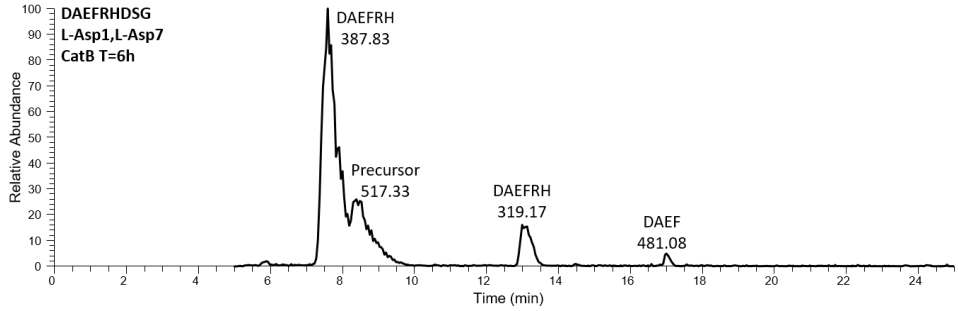


Figure 2.S29. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

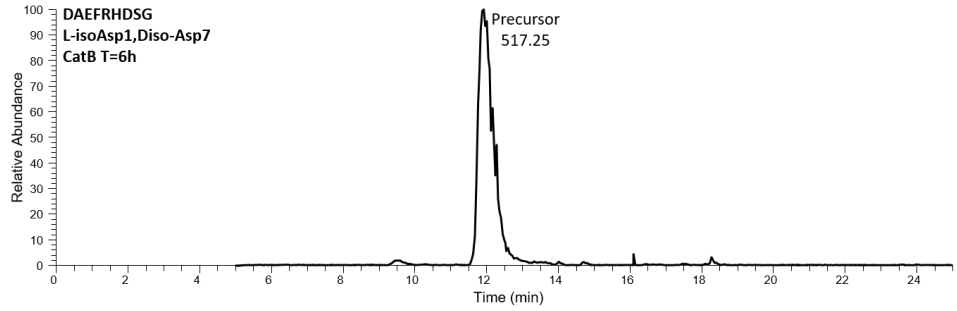


Figure 2.S30. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

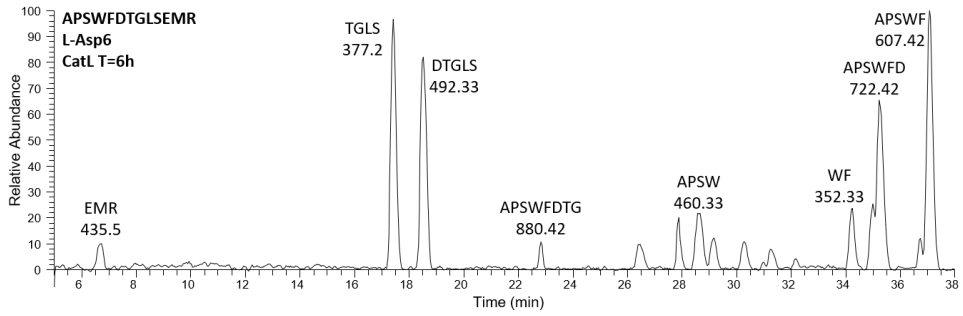


Figure 2.S31. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

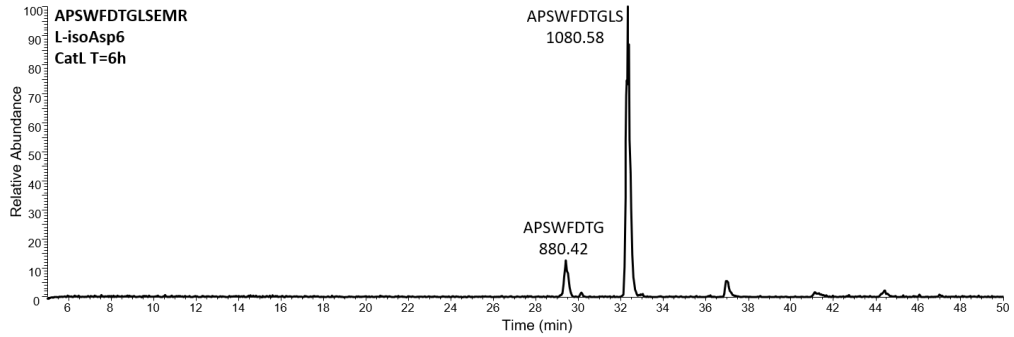


Figure 2.S32. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

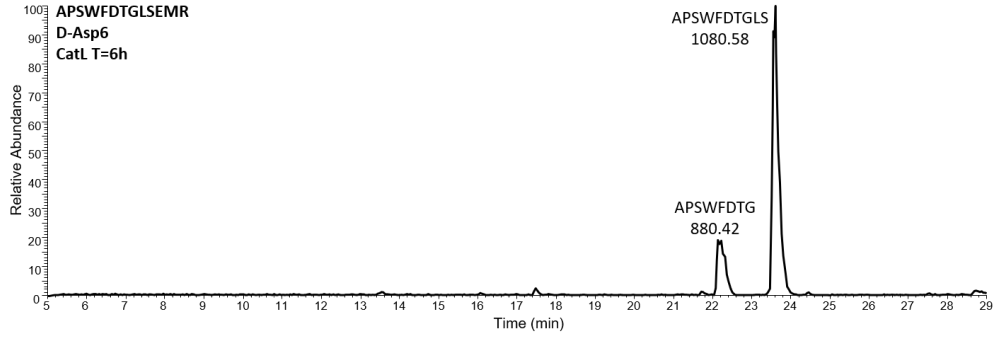


Figure 2.S33. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

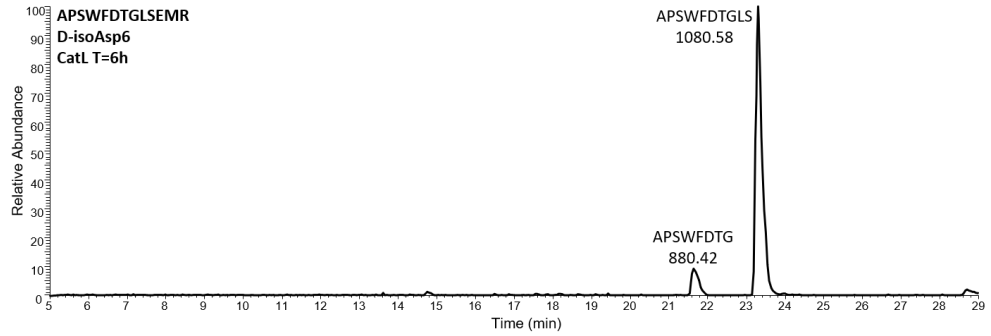


Figure 2.S34. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

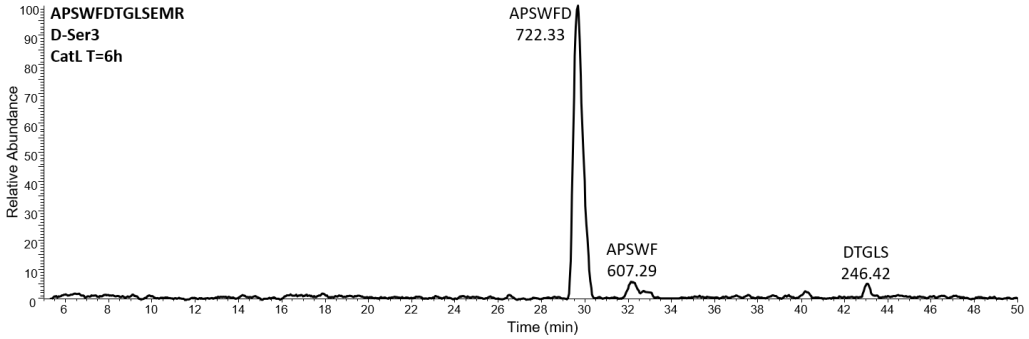


Figure 2.S35. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

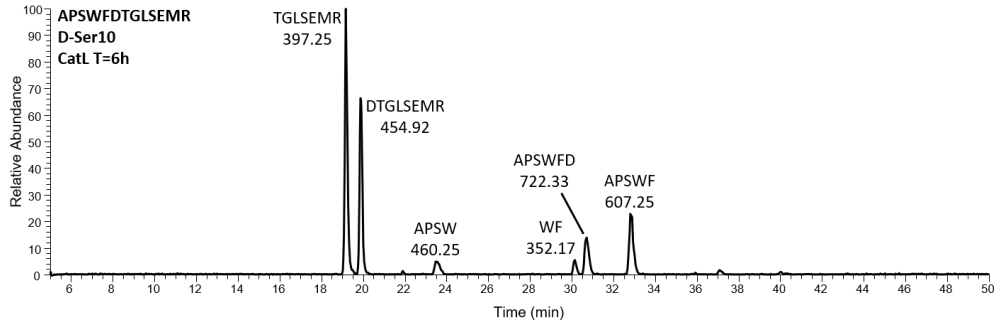


Figure 2.S36. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

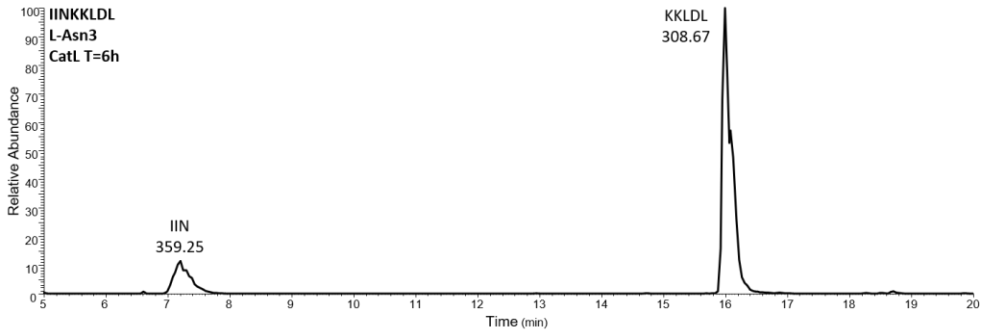


Figure 2.S37. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

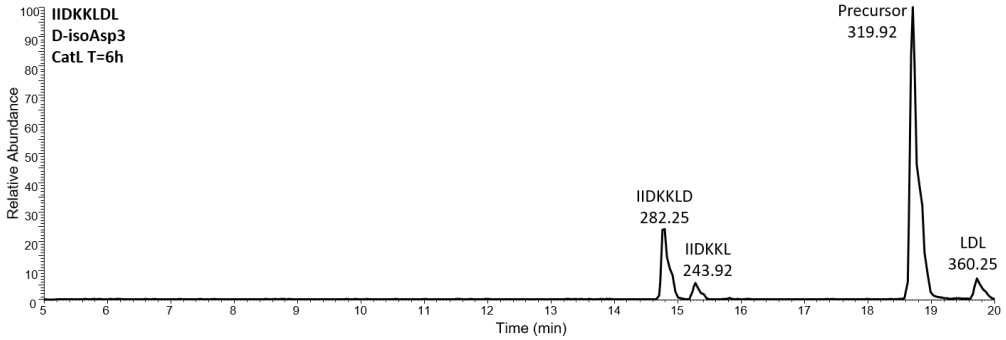


Figure 2.S38. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

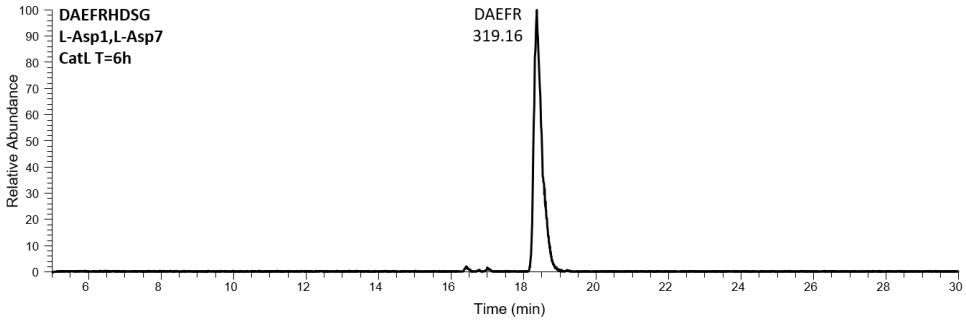


Figure 2.S39. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

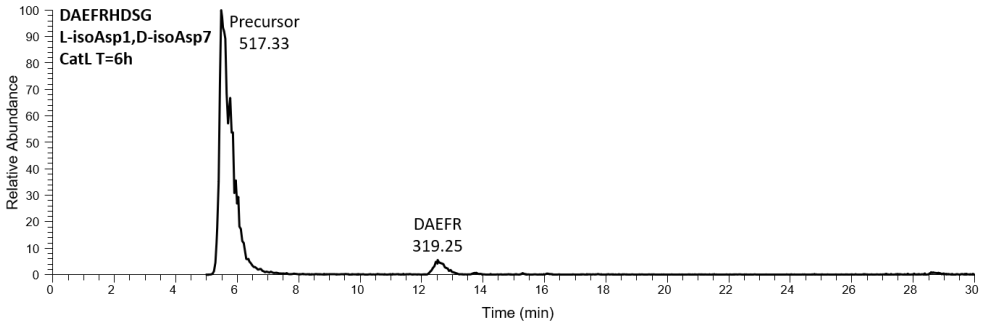


Figure 2.S40. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

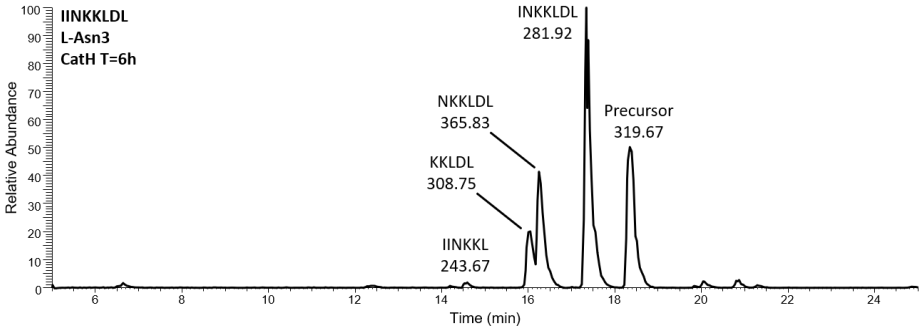


Figure 2.S41. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

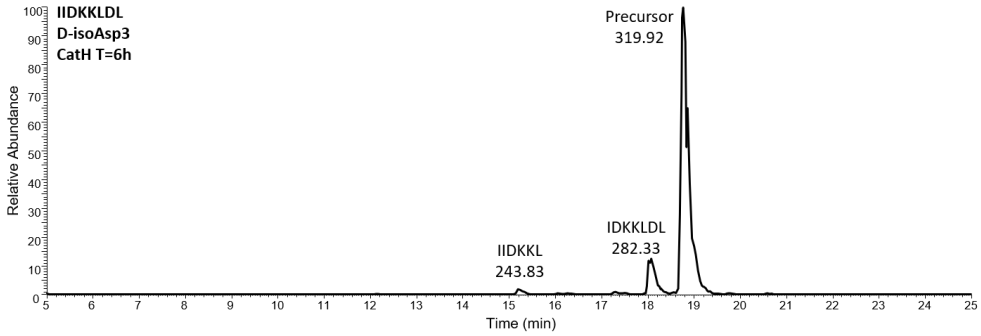


Figure 2.S42. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

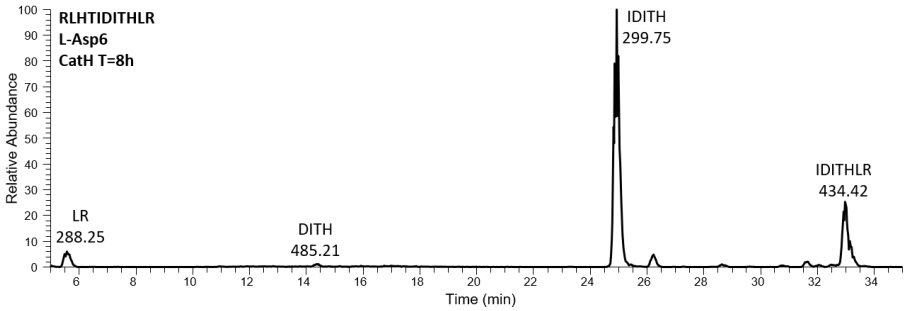


Figure 2.S43. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

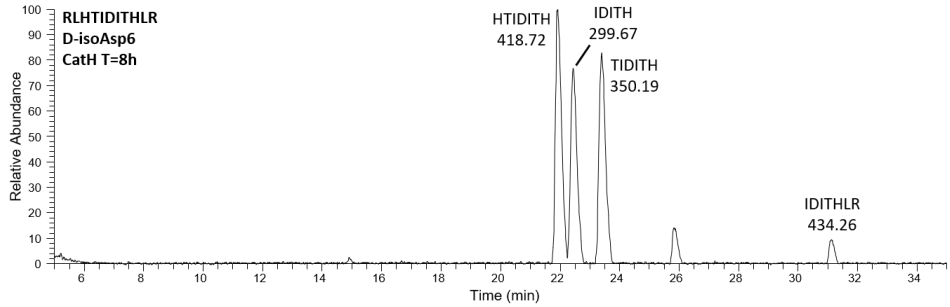


Figure 2.S44. LC-MS total ion chromatogram for peptide isomeric form listed in upper left corner.

References

[1] Kaushik, S.; Cuervo, A. M. Proteostasis and Aging. *Nat. Med.* **2015**, *21* (12), 1406–1415.

[2] Mizushima, N. Autophagy: Process and Function. *Genes Dev.* **2007**, *21* (22), 2861–2873.

[3] Dikic, I. Proteasomal and Autophagic Degradation Systems. *Annu. Rev. Biochem.* **2017**, *86* (1), 193–224.

[4] Luzio, J. P.; Hackmann, Y.; Dieckmann, N. M. G.; Griffiths, G. M. The Biogenesis of Lysosomes and Lysosome-Related Organelles. *Cold Spring Harb. Perspect. Biol.* **2014**, *6* (9), a016840–a016840.

[5] Bissa, B.; Beedle, A.; Govindarajan, R. Lysosomal Solute Carrier Transporters Gain Momentum in Research. *Clin. Pharmacol. Ther.* **2016**, *100* (5), 431–436.

[6] Kiselyov, K.; Jennings Jr., J. J.; Rbaibi, Y.; Chu, C. T. Autophagy, Mitochondria and Cell Death in Lysosomal Storage Diseases. *Autophagy* **2007**, *3* (3), 259–262.

[7] Platt, F. M.; Boland, B.; van der Spoel, A. C. Lysosomal Storage Disorders: The Cellular Impact of Lysosomal Dysfunction. *J. Cell Biol.* **2012**, *199* (5), 723–734.

[8] 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.

[9] Jansson, E. T. Strategies for Analysis of Isomeric Peptides. *J. Sep. Sci.* **2018**, *41* (1), 385–397.

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

[11] Reissner, K. J.; Aswad, D. W. Deamidation and Isoaspartate Formation in Proteins: Unwanted Alterations or Surreptitious Signals? *Cell. Mol. Life Sci.* **2003**, *60* (7), 1281–1295.

[12] 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.

[13] 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.

[14] 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.

[15] Zhu, J. X.; Doyle, H. A.; Mamula, M. J.; Aswad, D. W. Protein Repair in the Brain, Proteomic Analysis of Endogenous Substrates for Protein L-Isoaspartyl Methyltransferase in Mouse Brain. *J. Biol. Chem.* **2006**, *281* (44), 33802–33813.

[16] Yamamoto, A; Takagi, H.; Kitamura, D.; Tatsuoka, H.; Nakano, H.; Kawano, H.; Kuroyanagi, H.; Yahagi, Y.; Kobayashi, S.; Koizumi, K.; et al. Deficiency in Protein L-Isoaspartyl Methyltransferase Results in a Fatal Progressive Epilepsy. *J. Neurosci.* **1998**, *18* (6), 2063–2074.

[17] Kim, E.; Lowenson, J. D.; MacLaren, D. C.; Clarke, S.; Young, S. G. Deficiency of a Protein-Repair Enzyme Results in the Accumulation of Altered Proteins, Retardation of Growth, and Fatal Seizures in Mice. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (12), 6132–6137.

[18] Kreil, G. D-AMINO ACIDS IN ANIMAL PEPTIDES. *Annu. Rev. Biochem.* **1997**, *66* (1), 337–345.

[19] 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.

[20] 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* (5), 1343–1352.

[21] Koike, M. Cathepsin D Deficiency Induces Lysosomal Storage with Ceroid Lipofuscin in Mouse CNS Neurons. *Neurosci. Res.* **2000**, *38* (18), S29.

[22] Felbor, U.; Kessler, B.; Mothes, W.; Goebel, H. H.; Ploegh, H. L.; Bronson, R. T.; Olsen, B. R. Neuronal Loss and Brain Atrophy in Mice Lacking Cathepsins B and L. *Proc. Natl. Acad. Sci.* **2002**, *99* (12), 7883–7888.

[23] Papassotiropoulos, A.; Bagli, M.; Kurz, A.; Kornhuber, J.; Förstl, H.; Maier, W.; Pauls, J.; Lautenschlager, N.; Heun, R. A Genetic Variation of Cathepsin D Is a Major Risk Factor for Alzheimer's Disease. *Ann. Neurol.* **2000**, *47* (3), 399–403.

[24] Ciechanover, A.; Kwon, Y. T. Degradation of Misfolded Proteins in Neurodegenerative Diseases: Therapeutic Targets and Strategies. *Exp. Mol. Med.* **2015**, *47* (3), e147–e147.

[25] Wolfe, D. M.; Nixon, R. A. Autophagy Failure in Alzheimer's Disease and Lysosomal Storage Disorders : A Common Pathway To Neurodegeneration? **2018**, 237–257.

[26] Nixon, R. A.; Yang, D. S. Autophagy Failure in Alzheimer's Disease-Locating the Primary Defect. *Neurobiol. Dis.* **2011**, *43* (1), 38–45.

[27] Nixon, R. A. Amyloid Precursor Protein and Endosomal–lysosomal Dysfunction in Alzheimer's Disease: Inseparable Partners in a Multifactorial Disease. *FASEB J.* **2017**, *31* (7), 2729–2743.

[28] Roher, A. E.; Lowenson, J. D.; Clarke, S.; Wolkow, C.; Wang, R.; Cotter, R. J.; Reardon, I. M.; Zurcherneely, H. A.; Heinrikson, R. L.; Ball, M. J.; et al. Structural Alterations in the Peptide Backbone of Beta-amyloid Core Protein May Account for its Deposition and Stability in Alzheimer's-disease. *J. Biol. Chem.* **1993**, *268* (5), 3072–3083.

[29] Wu, Y.; Zhou, J.; Fishkin, N.; Rittmann, B. E.; Sparrow, J. R. Enzymatic Degradation of A2E, a Retinal Pigment Epithelial Lipofuscin Bisretinoid. *J. Am. Chem. Soc.* **2011**, *133* (4), 849–857.

[30] Bohley, P.; Seglen, P. O. Proteases and Proteolysis in the Lysosome. *Experientia* **1992**, *48* (2), 151–157.

[31] Müller, S.; Dennemärker, J.; Reinheckel, T. Specific Functions of Lysosomal Proteases in Endocytic and Autophagic Pathways. *Biochim. Biophys. Acta - Proteins Proteomics* **2012**, *1824* (1), 34–43.

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

[33] Szendrei, G. I.; Prammer, K. V.; Vasko, M.; Lee, V. M.-Y.; Otvos, L. The Effects of Aspartic Acid-Bond Isomerization on in Vitro Properties of the Amyloid β-Peptide as Modeled with N-Terminal Decapeptide Fragments. *Int. J. Pept. Protein Res.* **2009**, *47* (4), 289–296.

[34] McDermott, J. R.; Gibson, A. M. Degradation of Alzheimer's Beta-Amyloid Protein by Human Cathepsin D. *Neuroreport* **1996**, *7* (13), 2163–2166.

[35] Ballatore, C.; Lee, V. M. Y.; Trojanowski, J. Q. Tau-Mediated Neurodegeneration in Alzheimer's Disease and Related Disorders. *Nat. Rev. Neurosci.* **2007**, *8* (9), 663–672.

[36] Hasegawa, M.; Morishima-Kawashima, M.; Takio, K.; Suzuki, M.; Titani, K.; Ihara, Y. Protein Sequence and Mass Spectrometric Analyses of Tau in the Alzheimer's Disease Brain. *J. Biol. Chem.* **1992**, *267* (24), 17047–17054.

[37] Solé-Domènech, S.; Rojas, A. V.; Maisuradze, G. G.; Scheraga, H. A.; Lobel, P.; Maxfield, F. R. Lysosomal Enzyme Tripeptidyl Peptidase 1 Destabilizes Fibrillar Aβ by Multiple Endoproteolytic Cleavages within the β-Sheet Domain. *Proc. Natl. Acad. Sci.* **2018**, 201719808.

[38] Fuchs, S. M.; Raines, R. T. Pathway for Polyarginine Entry into Mammalian Cells. *Biochemistry* **2004**, *43* (9), 2438–2444.

[39] Adams-Cioaba, M. A.; Krupa, J. C.; Xu, C.; Mort, J. S.; Min, J. Structural Basis for the Recognition and Cleavage of Histone H3 by Cathepsin L. *Nat. Commun.* **2011**, *2* (1), 197.

[40] Shimizu, T.; Matsuoka, Y.; Shirasawa, T. Biological Significance of Isoaspartate and Its Repair System. *Biol. Pharm. Bull.* **2005**, *28* (9), 1590–1596.

[41] Zirah, S.; Kozin, S. A.; Mazur, A. K.; Blond, A.; Cheminant, M.; Ségalas-Milazzo, I.; Debey, P.; Rebuffat, S. Structural Changes of Region 1-16 of the Alzheimer Disease Amyloid β-Peptide upon Zinc Binding and in Vitro Aging. *J. Biol. Chem.* **2006**, *281* (4), 2151–2161.

[42] 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*, 131–141.

[43] Xiao, G.; Bondarenko, P. V. Identification and Quantification of Degradations in the Asp–Asp Motifs of a Recombinant Monoclonal Antibody *J. Pharm. Biomed. Anal.* **2008**, *47* (1), 23–30.

[44] Rehder, D. S.; Chelius, D.; McAuley, A.; Dillon, T. M.; Xiao, G.; Crouse-Zeineddini, J.; Vardanyan, L.; Perico, N.; Mukku, V.; Brems, D. N.; et al. Isomerization of a Single Aspartyl Residue of Anti-Epidermal Growth Factor Receptor Immunoglobulin γ2 Antibody Highlights the Role Avidity Plays in Antibody Activity. *Biochemistry* **2008**, *47* (8), 2518–2530.

[45] Sadakane, Y.; Yamazaki, T.; Nakagomi, K.; Akizawa, T.; Fujii, N.; Tanimura, T.; Kaneda, M.; Hatanaka, Y. Quantification of the Isomerization of Asp Residue in Recombinant Human AA-Crystallin by Reversed-Phase HPLC. *J. Pharm. Biomed. Anal.* **2003**, *30* (6), 1825–1833.

[46] Robinson, N. E.; Robinson, A. B. Molecular Clocks. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (3), 944–949.

[47] Fezoui, Y.; Teplow, D. B. Kinetic Studies of Amyloid β-Protein Fibril Assembly. *J. Biol. Chem.* **2002**, *277* (40), 36948–36954.

[48] Toyama, B. H.; Savas, J. N.; Park, S. K.; Harris, M. S.; Ingolia, N. T.; Yates, J. R.; Hetzer, M. W. Identification of Long-Lived Proteins Reveals Exceptional Stability of Essential Cellular Structures. *Cell* **2013**, *154* (5), 971–982.

[49] De Kimpe, L.; van Haastert, E. S.; Kaminari, A.; Zwart, R.; Rutjes, H.; Hoozemans, J. J. M.; Scheper, W. Intracellular Accumulation of Aggregated Pyroglutamate Amyloid Beta: Convergence of Aging and Aβ Pathology at the Lysosome. *Age (Omaha).* **2013**, *35* (3), 673–687.

[50] Kozin, S. A.; Mitkevich, V. A.; Makarov, A. A. Amyloid-β Containing Isoaspartate 7 as Potential Biomarker and Drug Target in Alzheimer's Disease. *Mendeleev Commun.* **2016**, *26* (4), 269–275.

[51] Hu, X.; Crick, S. L.; Bu, G.; Frieden, C.; Pappu, R. V.; Lee, J.-M. Amyloid Seeds Formed by Cellular Uptake, Concentration, and Aggregation of the Amyloid-Beta Peptide. *Proc. Natl. Acad. Sci.* **2009**, *106* (48), 20324–20329.

[52] Kozin, S. A.; Cheglakov, I. B.; Ovsepyan, A. A.; Telegin, G. B.; Tsvetkov, P. O.; Lisitsa, A. V.; Makarov, A. A. Peripherally Applied Synthetic Peptide IsoAsp7-Aβ(1-42) Triggers Cerebral β-Amyloidosis. *Neurotox. Res.* **2013**, *24* (3), 370–376.

[53] Kulikova, A. A.; Cheglakov, I. B.; Kukharsky, M. S.; Ovchinnikov, R. K.; Kozin, S. A.; Makarov, A. A. Intracerebral Injection of Metal-Binding Domain of Aβ Comprising the Isomerized Asp7 Increases the Amyloid Burden in Transgenic Mice. *Neurotox. Res.* **2016**, *29* (4), 551–557.

[54] Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. A Structural Model for Alzheimer's -Amyloid Fibrils Based on Experimental Constraints from Solid State NMR. *Proc. Natl. Acad. Sci.* **2002**, *99* (26), 16742–16747.

[55] Mawuenyega, K. G.; Kasten, T.; Sigurdson, W.; Bateman, R. J. Amyloid-beta Isoform Metabolism Quantitation by Stable Isotope-Labeled Kinetics. *Anal. Biochem.* **2013,** *440*, 56–62.

[56] Mawuenyega, K. G.; Sigurdson, W.; Ovod, V.; Munsell, L.; Kasten, T.; Morris, J. C.; Yarasheski, K. E.; Bateman, R. J. Decreased Clearance of CNS -Amyloid in Alzheimer's Disease. *Science* **2010,** *330*, 1774–1774.

[57] Luhrs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Dobeli, H.; Schubert, D.; Riek, R. 3D Structure of Alzheimer's Amyloid- (1-42) Fibrils. *Proc. Natl. Acad. Sci.* **2005**, *102* (48), 17342–17347.

[58] Dan, A.; Takahashi, M.; Masuda-Suzukake, M.; Kametani, F.; Nonaka, T.; Kondo, H.; Akiyama, H.; Arai, T.; Mann, D. M.; Saito, Y.; et al. Extensive Deamidation at Asparagine Residue 279 Accounts for Weak Immunoreactivity of Tau with RD4 Antibody in Alzheimer's Disease Brain. *Acta Neuropathol. Commun.* **2013**, *1* (1), 54.

[59] Watanabe, A.; Takio, K.; Ihara, Y. Deamidation and Isoaspartate Formation in Smeared Tau in Paired Helical Filaments. Unusual Properties of the Microtubule-Binding Domain of Tau. *J. Biol. Chem.* **1999**, *274* (11), 7368–7378.

[60] Tompa, P. Intrinsically Unstructured Proteins. *BioEssays* **2003**, *25* (9), 847–855.

CHAPTER 3: Proteolysis of Amyloid Beta by Lysosomal Enzymes as a Function of Fibril Morphology

Abstract

Aggregation of amyloid-beta (Aβ) into extracellular plaques is a well-known hallmark of Alzheimer's disease (AD). Similarly, autophagic vacuoles, autophagosomes, and other residual bodies within dystrophic neurites, though more difficult to detect, are characteristic features of AD. To explore the potential intersection between these observations, we conducted experiments to assess whether Aβ fibril formation disrupts proteolysis by lysosomal enzymes. Fibrils constituted from either Aβ 1-40 or Aβ 1-42 were grown under both neutral and acidic pH. The extent of proteolysis by individual cathepsins (L, D, B, and H) was monitored by both thioflavin T fluorescence and liquid-chromatography combined with mass spectrometry. The results show that all Aβ fibril morphologies are resistant to cathepsin digestion, with significant amounts of undigested material remaining for samples grown in either neutral or acidic pH. Further analysis revealed that the neutral-grown fibrils are proteolytically resistant throughout the sequence, while the acid-grown fibrils prevented digestion primarily in the C-terminal portion of the sequence. Fibrils grown from Aβ 1-42 are generally more resistant to degradation compared to Aβ 1-40. Overall, the results indicate that Aβ fibrils formed in the neutral pH environments found in intracellular or extracellular spaces may pose the greatest difficulty for complete digestion by the lysosome, particularly when the fibrils are comprised of Aβ 1-42.

Introduction

Autophagy is a critical process needed to clear cellular waste and free up resources for reuse or energy production. Within this framework, autophagy delivers peptides and proteins to the lysosome where they are digested into constituent amino acids, forming a crucial cog in the gears that drive proteostasis.[1] Target substrates can be gathered from inside the cell or from the extracellular space through endocytosis.[2] Due to variety of pathways leading to the lysosome, substrates can be subjected to many conditions and environments prior to fusion with a lysosome. Additionally, the endo/lysosomal system utilizes acidic compartments for delivery and degradation of substrates, further expanding the range of different environments that may be experienced prior to degradation.[3] Failure of the endo/lysosomal system can lead to a variety of complications, including a class of diseases known as lysosomal storage disorders. Lysosomal storage is most frequently caused by hydrolase dysfunction, which leads to the accumulation of undigested substrates and eventual failure of the organelle.[4] Autophagic disruption has also been associated with Alzheimer's disease (AD) due to the hallmark observation of lysosomal storage.[5,6]

Inside the lysosome, hydrolases known as cathepsins degrade peptides and proteins into the constituent amino acids which are then released by transporter proteins back into the cytosol.[7] While a few members of the cathepsin family are exopeptidases which cleave from the termini, the majority are endopeptidases which cleave somewhere in the middle of the sequence.[8] Among the endopeptidases, the most abundant and active are cathepsin L (catL, all cathepsins will be abbreviated similarly) and catD.[9] Studies have

demonstrated that knocking out either catD or catL induces pathology and death in mice within 4 weeks.[10,11] The most abundant lysosomal exopeptidase, catB, recognizes and binds to the C-terminus (as well as the C-terminal mimic L-isoAsp)[12] and removes two amino acids at a time as dipeptides. The complementary aminopeptidase, catH, works from the N-terminus to remove one amino acid at a time. Although catB and catH are primarily exopeptidases, they both possess a secondary endopeptidase activity.[13,14] Knockouts of catB also cause shortened lifespans[10], while knockouts of catH lead to significantly reduced levels of important neurotransmitter peptides.[15] These studies demonstrate that while the cathepsins perform similar functions, each enzyme is individually vital to maintain proteostasis.

Amyloid beta (Aβ) is a peptide of typically 39-42 residues known to aggregate into fibrils,[16] similar to other amyloid proteins such as human islet amyloid polypeptide[17] and CsgA.[18] As target of autophagy, Aβ (including fibrils) can be trafficked from both intracellular and extracellular spaces to the lysosome.[19] Aβ fibrils have been shown to impede digestion by proteases such as trypsin[20] and bovine brain proteases.[21] In a related system, α-synuclein aggregates were found to resist degradation by catL.[22] These previous experiments suggest that amyloid fibrils are resistant in varying degrees to degradation by proteases. Aβ spontaneously assembles into fibril structures in solution via side-chain and backbone interactions. Aβ fibrils are comprised primarily of stacked beta-sheet structures stabilized by hydrophobic interactions along the middle and C-terminal region of the sequence.[23,24] Additionally, Aβ fibrils are often polymorphic and comprised of strands that vary in the molecular arrangement of constituent peptides.[25,26]

NMR experiments have shown that accessible morphologies include beta sheet layers composed of dimeric or trimer assemblies of Aβ.[27,28,29] Notably for both of these morphologies, the beta-sheet core remains tightly bound while the N-terminal residues from 1-14 are disordered.

The pH during fibril formation also impacts fibril morphology. For example, fibrils formed in acidic environments differ from those formed at neutral pH and time course measurements of fibril formation show that aggregation occurs faster at lower pH.[30] Importantly, acidic conditions present in the lysosome could facilitate alternate morphologies for fibrils formed therein.[31] The structural difference caused by lower pH is attributed to varying protonation states of histidine residues in the sequence.[32] While other ionizable groups are unaffected by a shift from cellular to lysosomal pH, the pKa value for the imidazole group comprising the sidechain of histidine lies in the middle of the relevant pH range. As a consequence, histidine is likely to become protonated in acidic compartments. The formation of amyloid beta fibrils is driven by hydrophobic interactions between the peptide chains, and as such can potentially be altered by hydrophilic charges.[33] The resulting differences between fibrils due to His protonation was observed using transmission electron microscopy, revealing markedly different morphology distributions.[34] By substituting alanine for histidine residues, it was then demonstrated that protonation of sidechains for His6, His13, and His14 significantly affects fibril structure by disfavoring amyloid sheets in the N-terminal half of the sequence. In the body, the majority of Aβ is located extracellularly, however it is known to accumulate in the lysosome of neuronal cells during the process of autophagy.

Herein, we examine incubations of fibrils grown at cellular and acidic pH with lysosomal cathepsins to evaluate their ability to degrade these structures. Analysis was performed via thioflavin T (ThT) fluorescence to measure the general extent of degradation, and proteolytic products were also quantitatively assessed with liquid-chromatography/mass spectrometry. Significant differences were observed in the amount of degradation as a function of the pH used to grow the fibrils and whether the fibrils were composed of Aβ40 or Aβ42. Further examination of these digestions with mass spectrometry revealed differences in both the identified sequences and length of peptides remaining after incubation. By plotting these products as a function of intensity, we were able to map proteolytically resistant regions as a function of fibril composition and pH during formation.

Experimental Procedures

Fibril Formation

Lyophilized Aβ powder was purchased from Anaspec. The samples were disentangled via ammonium hydroxide treatment.[35] 100 μg aliquots of each peptide were dissolved in 50 μL of 0.1% ammonium hydroxide solution (w/v) with sonication and monomeric peptide was confirmed by ThT, followed by dilution to 1 mL with either 50 mM tris pH 7.2 or acetate pH 5 buffer for fibril growth. Amyloid beta 1-42 aliquots were fibrilized at 25 μM while amyloid beta 1-40 were fibrilized at 100 μM to start fibril growth after brief agitation. Fibrils were grown for 5 days at 37°C and checked by ThT fluorescence to confirm fibril presence.

Cathepsin Incubations

Aliquots containing Aβ were digested by cathepsins in acetate buffer pH 5, with 1 mM ethylenediaminetetraacetic acid (EDTA) and 500 μM dithiothreitol (DTT) to prevent active site oxidation. For each digestion 0.4 μg of enzyme were incubated with 20 μg of amyloid beta for a 1:50 enzyme:substrate ratio (w/w). A control sample was set up for each digestion with no enzyme added. Incubations occurred over an 18-hour period at 37°C to allow for maximum digestion potential. Digestions were quenched by dilution with 200 mM tris before immediate fluorescent measurements.

Fluorescence Measurements

The presence of beta-sheet rich aggregates was examined by ThT assay. Samples were diluted to 2 μM in 200 mM tris buffer with 6 μM ThT. Emission scans were performed on a QuantaMaster-400 fluorimeter using an excitation wavelength of 440 nm and an emission wavelength of 485 nm.

LC-MS Analysis

Samples were analyzed on a Thermo Fisher Ultimate 3000 RSLCnano System interfaced with a Thermo Fisher Velos Pro Orbitrap using an electrospray ionization (ESI) source. Peptides were separated on a capillary column packed in-house with C18 3 μm resin using a Shotgun Proteomics Inc high pressure vessel. Mobile phase A was water 0.1% formic acid and mobile phase B was 80% acetonitrile in water with 0.1% formic acid. Nano-ESI was performed using a spray voltage of 2.1 kV with an S-lens value of 65.

Results and Discussion

Amyloid beta 1-40 (Aβ40) and 1-42 (Aβ42) were incubated at pH 5 and pH 7.2 to produce stable fibrils with varying morphologies. Importantly, fibrils formed at different pH do not interconvert if the pH is shifted after fibrils have formed.[36] These fibrils were then digested over an 18-hour period in separate experiments by cathepsin L, D, B, and H, which represent two crucial endopeptidases, one carboxypeptidase, and one aminopeptidase. Although the pH was varied to create fibrils with differing morphology, the digestion experiments for both types of fibrils were conducted at identical pH corresponding to the optimal value for each cathepsin. The extent of digestion was coarsely measured by ThT fluorescence intensity, as shown in Fig. 3.1a for the digestion of neutral Aβ42 by catL. ThT is a fluorescent dye used to measure the presence of protein and peptide aggregates due to its ability to fluoresce intensely when bound to beta-sheet rich structures.[37] The fractional intensity of fluorescence remaining after digestion represents the amount of remaining beta-sheet rich structures present in the sample after proteolysis. The fluorescence data from all cathepsin digestions is compiled in Fig 3.1b. For the endopeptidases, catL digested more fibril relative to catD. For the acid-grown fibrils (in green), catL reduced the amount of fluorescence by >80% from the initial level. However, for the neutral-grown fibrils less digestion was observed, particularly for Aβ42, which only exhibited a ~30% reduction in fluorescence intensity. Similarly, catD reduced ThT fluorescence less for neutral-grown fibrils less, and yielded almost no change for the neutral Aβ42 fibrils.

Endopeptidases operate by binding to several amino acids on either side of the peptide bond targeted for hydrolysis, and as such are potentially sensitive to differences in substrate backbone structure in either direction of the surrounding sequence region. In contrast, catB and catH act primarily as exopeptidases and have the strongest interactions with residues to one side of the targeted peptide bond. For catH, digestion of Aβ42 and Aβ40 yielded similar results and reductions in ThT fluorescence were not significant for either acidic or neutral fibrils. This suggests that catH was unable to access the amyloid region lending ThT fluorescence, although it is unclear whether any portion of the N-terminus was digested. For catB, which attacks from the C-terminal side, significantly more ThT fluorescence was retained by the Aβ42 fibrils for both acidic and neutral fibers. Indeed, neutral or acidic had little effect on the catB results, suggesting that differences in fibril structure likely occur near the N-terminus (as discussed in the introduction). Although there are clearly differences in the ThT results for our various test conditions, more detailed information would likely facilitate greater understanding.

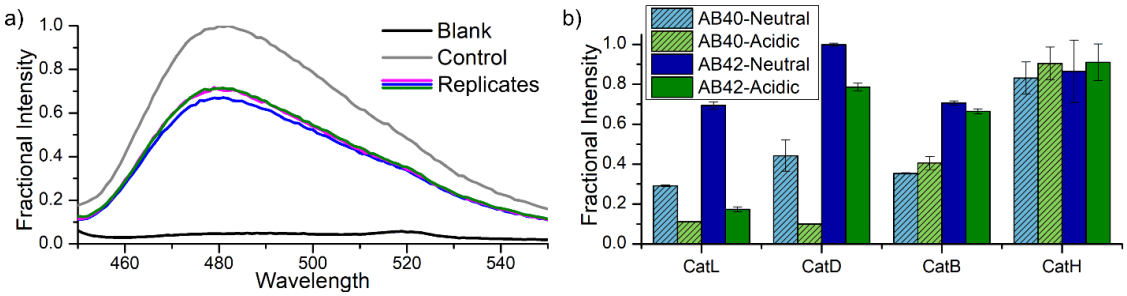


Figure 3.1. (a) Three replicates for ThT fluorescence following digestion of Aβ42-Neutral by catL. (b) Compiled fluorescence data for all cathepsin incubations. The dark blue peak for CatL derives from the data in (a).

To more precisely determine the outcome of each cathepsin digestion, the samples were analyzed with a combination of liquid-chromatography and mass spectrometry. These experiments were able to identify the precise peptides remaining following proteolysis and defibrilization. Raw chromatograms of the digestion of Aβ42 with catL are shown in Fig. 3.2a and 2b for neutral and acidic fibrils, respectively. Notable differences in retention times and relative intensities are apparent, suggesting that the two experiments generated considerably different peptide profiles. This possibility is confirmed by MS analysis, which is illustrated schematically in Fig. 3.2c. Each identified sequence from the chromatograms in Fig. 3.2a/b are displayed as individual lines in Fig. 3.2c, where the length of each line maps out the corresponding peptide sequence in relation to the full sequence shown in the middle of the diagram. The sequences are also ordered by intensity, with more intense peptides displayed closer to the full Aβ42 sequence. The full list of peptides can also be found in Tables S1 and S2 in the supporting information. Peptides located in the C-terminal portion of Aβ42 were found for both acidic and neutral fibrils, indicating resistance to catL digestion in this region. However, for the neutral fibrils, a variety of peptides were found from the N-terminal region, including many peptides that contained the N-terminus itself. These results suggest that the amyloid forming region is resistant to digestion for both acidic and neutral fibrils, but that the N-terminal region is accessible and susceptible to digestion for acidic fibrils. The data from Fig. 3.2c can be more succinctly summarized with additional analysis as shown in Fig. 3.2d and 2e. In Fig. 3.2d, histograms of residual peptide length are shown for four different experiments: neutral Aβ40, acidic Aβ40, neutral Aβ42, and

acidic Aβ42. Aβ42 digestion products skew more towards the longer sequence lengths in general while Aβ40 digestion products are clustered in smaller length peptides. Notably, the 26-30 length bin contains only intensity from neutral-grown fibrils of Aβ40 and Aβ42. When compared to the catL digestion of monomeric Aβ42 collected in Table S3 and displayed as histograms in Fig. 3.S1, a striking difference is observed. In the monomeric digest, nearly 80% of the peptide products are of lengths less than 5 amino acids long, and there are no peptides observed greater than 15 residues long. This contrast in peptide lengths demonstrates a substantial obstruction is occurring in the proteolysis of Aβ fibrils.

To visually display a summary of the regions with the most resistance to proteolysis, sequences with a length of at least 11 amino acids were used to generate an intensity map as shown in Fig. 3.2e. The fractional intensity value (y-axis) for each residue (x-axis) was calculated by adding up the intensity of all peptides containing the residue and then dividing by the intensity of all peptides. Only peptides longer than 11 residues were included. Accordingly, a fractional intensity value of 0.5 means that the residue is present in 50% of the total peptide intensity that was observed. Examination of the solid lines (derived from Aβ42 data) reveals a quantitative assessment in excellent agreement with the semi-quantitative representation shown in Fig. 3.2c. Again, differences in product profiles between various digestions are apparent. The C-terminal region is resistant to proteolysis in all cases. Interestingly, neutral Aβ42 fibrils are the most resistant in the N-terminal region. For Aβ40, less difference is noted between digestion of acidic and neutral fibrils. These results are consistent with the fluorescence data shown in Fig. 3.1b,

where the neutral-grown fibrils display a higher amount of intensity after digestion in both Aβ40 and Aβ42, demonstrating an increase in proteolytic resistance.

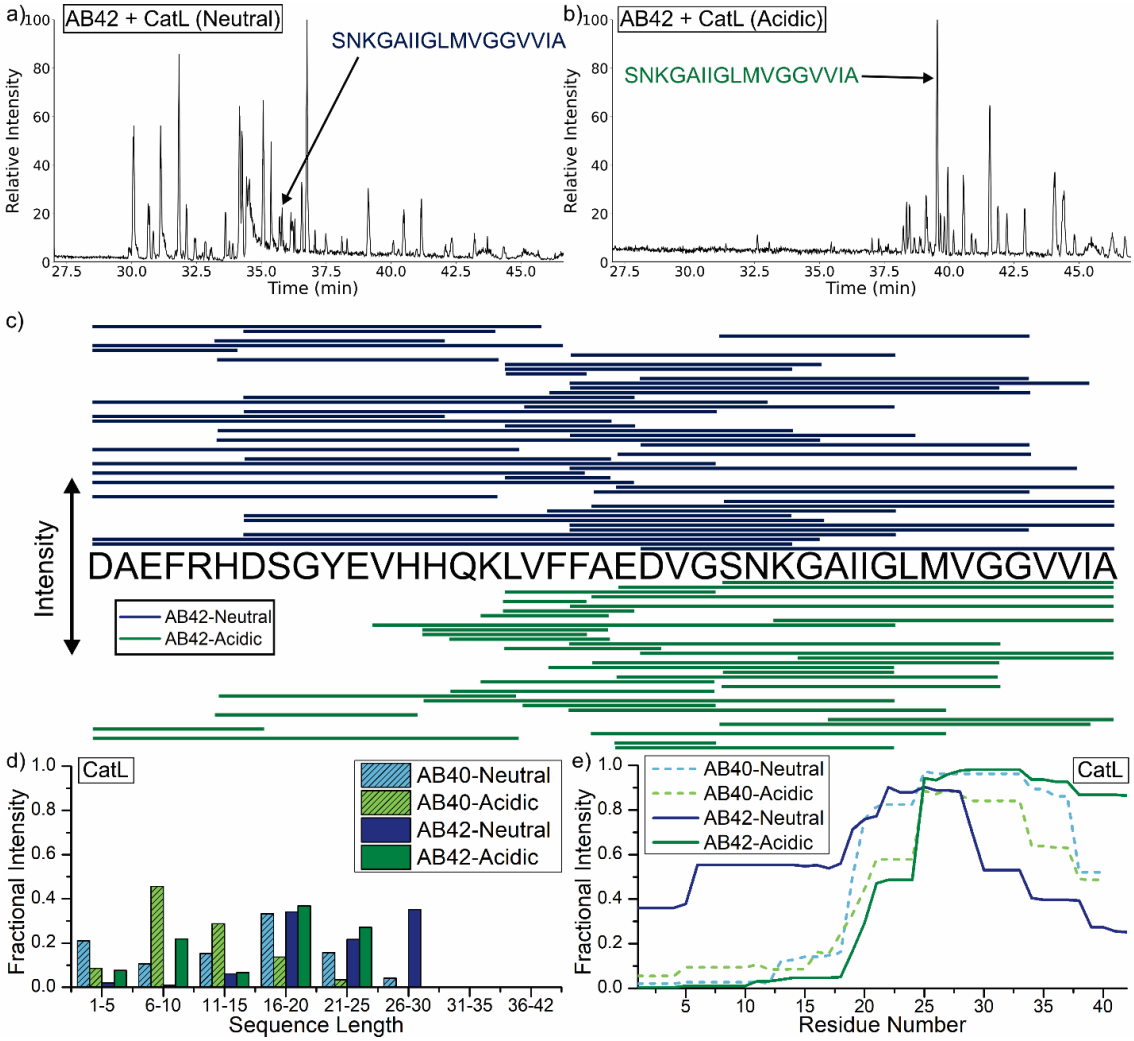


Figure 3.2. (a) Raw chromatogram for the digestion of Aβ42-neutral and (b) Aβ42-acidic by catL. The proteolytic product Aβ 16-42 is shown in both chromatograms as a reference. (c) Compiled line diagram of the identified peptides from the chromatograms. Lines indicate which part of the full sequence comprises the proteolytic product. Products are ordered by intensity, with the most intense products listed closer to the full sequence. (d) Bar plot showing the length of the identified product peptides. (e) Plot of the residue intensity among the total intensity of all peptides with a length greater than 10, representing areas resistant to proteolysis.

The same analysis was performed on digestions with the endopeptidase catD, as shown in Fig. 3.3. The peptide lengths shown in Fig. 3.3a consist of a tighter spread than those seen with catL digestions. Interestingly, the neutral-grown Aβ42 products are almost entirely sequences with lengths of 16-20 residues. Although most of the intensity from Aβ40 digestions is found in the 16-20 length as well, some peptides were also identified with lengths greater than 20. The absence of shorter peptides may suggest that fewer acceptable binding sites exist for catD or that it is less capable of digesting fibrils in general. The catD digestion of monomeric Aβ42 is illustrated in Table S4 and displayed in Fig. 3.S2. The distribution of intensities suggests that catD is able to more easily bind and degrade sequences in the monomer form, populating the 6-10 and 11-15 bins with around 65% of the product intensity. Additionally, the 1-5 length bin contains a small portion of the intensity, while none is present in either Aβ42 fibril digestion. The peptide intensity map shown in Fig. 3.3b reveals some similar trends to the previous catL digestions. All digestions showed a resistant region in the C-terminal half of the sequence, extending from around Lys16 to Met35. In the neutral-grown Aβ42, nearly half of the peptides originating from the N-terminal half of the sequence survive digestion. The Aβ40 digests for both acidic and neutral fibrils also yielded more N-terminal peptide intensity. The fluorescence data for these samples shown in Fig. 3.1b is higher for both digestions of neutral-grown Aβ, which indicates that the resistant residues in the N-terminal half may be contributing to the formation of greater amounts of stacked beta-sheet content involved in ThT binding.

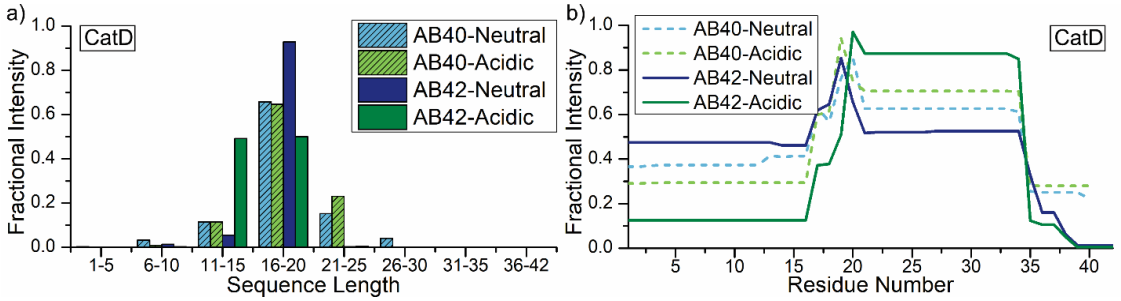


Figure 3.3. (a) Bar plot of the sequence length of identified proteolytic products from catD incubations. (b) Plot of the residue intensity among the total intensity of all peptides with a length greater than 10.

Identical analyses for results from catB and catH are shown in Fig. 3.4. These enzymes act primarily as exopeptidases, though both possess some endopeptidase activity. Examining the data for the catB digests in Fig. 3.4a reveals a wide spread of peptide lengths. The Aβ42 neutral fibril distribution is skewed towards longer lengths, while the Aβ40 neutral fibrils are distributed more towards the center of the distribution. Acidic fibrils from both Aβs populate bimodal distributions, favoring longer and shorter peptides. The peptide intensity map is shown in Fig. 3.4b. Although catB is a carboxypeptidase which cleaves from the C-terminus, it is unable to progress very far before encountering resistance. Indeed, only the acidic fibrils from Aβ40 reveal any cleavage at the C-terminus that is nearly complete. Ironically, most of the degradation for catB takes place due to secondary endopeptidase activity in the N-terminal region.

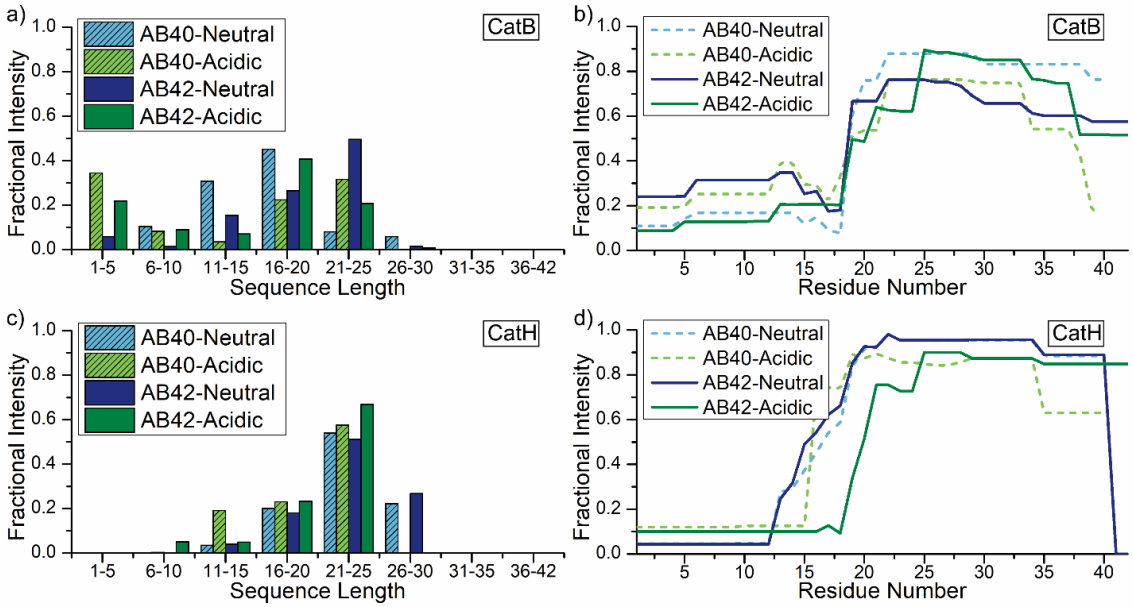


Figure 3.4. (a) Bar plot of the sequence length of identified proteolytic products from exopeptidase catB and (c) catH incubations. (b) Plot of the residue intensity among the total intensity of all peptides with a length greater than 10 for catB incubations and (d) catH incubations.

In the length histograms for catH shown in Fig. 3.4c, the majority of all peptide intensity is observed in peptides of length 21-25. CatH is an aminopeptidase that preferentially cleaves one amino acid at a time and is likely to produce products that will not be retained by LC. Notably, peptides of the longest observed length (26-30) were only recorded for the neutral-grown fibrils. In the residue intensity plot for catH shown in Fig. 3.4d, the N-terminal region is similarly digested for all four experiments. The primary differences in Fig. 3.4d relate to Aβ42. The acidic fibrils allow greater penetration in the N-terminal region, while an unexpected cleavage at the C-terminus is observed for the neutral fibrils. This C-terminal cleavage goes essentially to completion as with the catD digests and is more efficient than the C-terminal cleavage observed for

catB. The smaller differences between these digestion profiles are similarly reflected in the smaller differences between the fluorescence intensity shown in Fig. 3.1b.

Two commonly observed fibril morphologies for both Aβ40/42 are comprised of either dimeric layers (PDB: 2NAO) or trimeric layers (PDB: 2M4J) of beta sheets as shown in Fig. 3.5. In the dimer form, the N-terminal region is a disordered strand that extrudes from the fibril core. This morphology exposes the N-terminus and is consistent with the high digestion levels that we observed for all cathepsins. In particular, little resistance was encountered in the N-terminal region for catH, although the acidic fibrils appear to have greater exposure, particularly for Aβ42. His13 and His14 bridge the transition between the N-terminal tail and more organized beta-sheet. When protonated, the hydrophilic side-chains of His may lead to additional disorder. Additionally, the C-terminal tail is exposed for the dimeric fibril, making it easier for the cathepsins to cleave (as is observed in almost all of the cathepsin incubations). The digestion results match the expected protection afforded by the dimeric structure rather well, suggesting that a significant fraction of the fibrils present in our digestion may be comprised of similar dimeric sheets.

In trimeric form, the N-terminal region is ordered but still represents the most solvent exposed region of the peptide, protecting the beta sheet and C-terminus that are obscured with the core. This structure would be more easily digested near the N-terminus as well, and protonated histidine side-chains would also contribute to looser binding and more solvent exposure in the acid-grown fibril samples. This structure would be less available for enzymatic binding as it nears the bend region from residues 19-28, which is consistent

with the sloped increase of residue intensity observed for all cathepsin incubations in the same region. However, the buried C-terminal tail would be expected to make digestion difficult from this side of the peptide. Given that we observe digestion of C-terminal residues, particularly for catD, this morphology is not likely the dominant in form present in our fibrils.

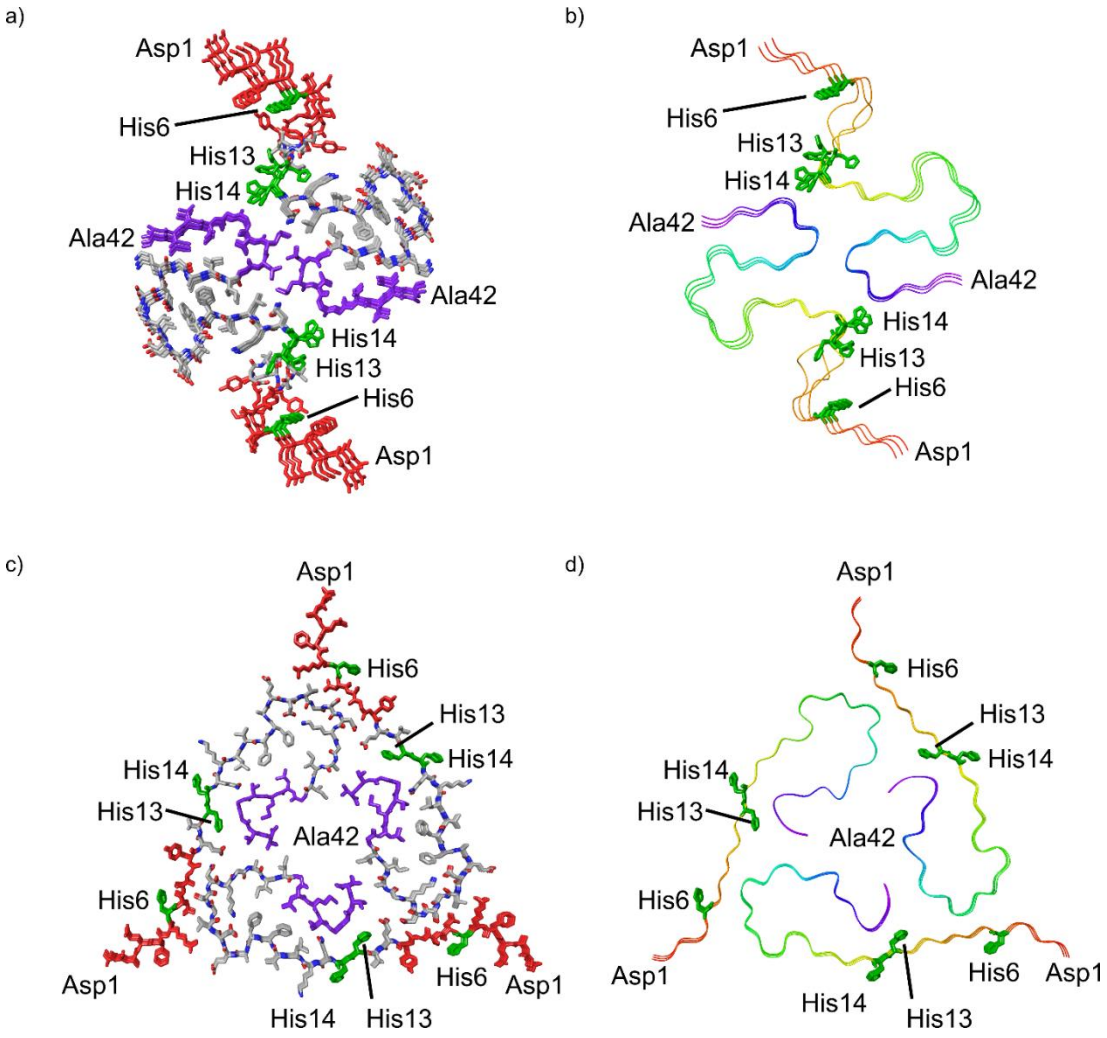


Figure 3.5. a) Tube representation and b) ribbon representation of Aβ 1-42 dimer fibril (PDB: 2NAO). The N-terminal region is shown by highlighting residues 1-10 in red and protrudes in a disordered fashion from the fibril core. The C-terminal region is shown by highlighting residues 33-40/42 in purple with exposed tail sticking out from the hydrophobic core. Relevant histidine residues are shown in green. c) Tube representation and d) ribbon representation of Aβ 1-40 trimer fibril (PDB: 2M4J). The N-terminal region is exposed with some accessibility of the following residues until the bend region.

Conclusions

We have examined the influence of fibril formation on the proteolysis of both Aβ40 and Aβ42 in a detailed and quantitative fashion. It is clear that fibril formation interferes with proteolysis by the major lysosomal cathepsins in every case. The differential digestion obtained for fibrils formed at either neutral or acidic pH confirms the likelihood that such fibrils have distinct structures, mostly related to the N-terminal portion of the sequence. Overall, our results suggest that fibrils composed of Aβ42 and formed at neutral pH will present the greatest difficulty for digestion within the lysosome. In contrast, monomeric Aβ is easily digested by cathepsins and appears unlikely to contribute to lysosomal pathology. However, our results suggest that it is possible for amyloid fibrils to contribute to AD pathology and the lysosomal storage observed in the disease by simply evading degradation.

Acknowledgements.

The authors gratefully acknowledge funding from the National Institute on Aging, R01AG066626.

Supporting Information

Sequence	Mass	Charges	Score	Intensity
DVGSNKGAIIGLMVGGVVIA	1869.039	2	154.1	186460000
DAEFRHDSGYEVHHQKLVFFAEDVGSNK	3260.528	4;5	102.72	185020000
DAEFRHDSGYEVHHQKLVFFAEDVGSNKG	3317.549	4;5	77.646	183940000
DSGYEVHHQKLVFFAEDVGSNKGAIIG	2916.441	3	83.247	127920000
FAEDVGSNKGAIIGLMVGG	1833.93	2	208.51	109090000
FAEDVGSNKGAIIGLMVGGVVIA	2216.188	2	139.64	65668000
DSGYEVHHQKLVFFAEDVGSNKG	2562.214	4	76.573	45466000
DSGYEVHHQKLVFFAEDVGSNK	2505.193	4	81.967	42945000
FFAEDVGSNKGAIIG	1523.762	2	142.08	35732000
AEDVGSNKGAIIGLMVGGVVIA	2069.119	2	193.51	34654000
SNKGAIIGLMVGGVVIA	1597.923	2	188.27	32499000
DAEFRHDSGYEVHHQK	1953.872	4	97.767	30940000
AEDVGSNKGAIIGLMVGG	1686.861	2	121.21	30536000
EDVGSNKGAIIGLMVGGVVIA	1998.082	2	229.56	30105000
DAEFRHDSGYEVHHQKLVFFAE	2660.241	4	61.862	26636000
LVFFA	595.337	1	76.733	25365000
DAEFRHDSGYEVHHQKLVFF	2460.161	4	93.495	21574000
FAEDVGSNKGAIIGLMVGGVV	2032.066	2	126.56	15270000
DAEFRHDSGYEVHHQKLVFFAEDVG	2931.358	4	60.278	13334000
DSGYEVHHQKLVFFA	1775.863	3	138.63	13269000
EDVGSNKGAIIGLMVGG	1615.824	2	231.07	12187000
DAEFRHDSGYEVHHQKL	2066.956	4	94.532	12092000
DVGSNKGAIIGLMVGGVV	1684.918	2	132.86	11718000
HDSGYEVHHQKLVFFAEDVGSNKG	2699.273	4	91.337	11586000
FAEDVGSNKGAIIGL	1489.778	2	160.86	10622000
HDSGYEVHHQKLVFFAEDVGSNK	2642.251	4	92.289	9911600
LVFFAE	724.3796	1	78.156	9415300
DAEFRHDSGYEVHHQKLVFFA	2531.198	4	93.768	9161400
DAEFRHDSGYEVHH	1697.718	3	170.66	8034900
DSGYEVHHQKLVFFAEDVG	2176.023	3	80.719	7750500
VFFAEDVGSNKGAIIG	1622.83	2	142.89	7344800
DAEFRHDSGYEVHHQKLVFFAEDVGSN	3132.433	4	75.539	7082400
DSGYEVHHQKLVFFAE	1904.906	3	116.9	6684200
FFAEDVGSNKGAIIGLMVGG	1980.998	2	154.1	6287600
FAEDVGSNKGAIIGLMVG	1776.908	2	98.592	5833700
VFFAEDVGSNK	1211.582	2	225.87	5769000

FAEDVGSNKGAIIGLMVGGVVI	2145.15	2	152	5416300
DVGSNKGAIIGLMVGG	1486.781	2	176.78	4348700
LVFF	524.2999	1	83.876	4066000
LVFFAEDVGSNK	1324.666	2	246.86	3731800
LVFFAEDVGSNKG	1381.688	2	279.77	2709000
HDSGYEVHHQK	1335.596	3	166.07	2663700
FAEDVGSNKGAIIG	1376.694	2	131.42	2660800
DAEFRH	773.3457	2	100.27	2656200
DAEFRHDSGYEVHHQKLVF	2313.093	4	94.544	2630600
HDSGYEVHH	1079.442	3	125.81	1802200
SNKGAIIGLMVGG	1215.665	2	164.97	1692200
DSGYEVHHQK	1198.537	3	96.668	956900
DAEFRHDSGYEVHHQKLV	2166.024	4	67.646	789670

Table 3.S1. Peptide sequences identified in the digest of neutral-grown Aβ42 fibrils by catL. Raw data shown in Fig. 2a and compiled in Fig. 2c.

Sequence	Mass	Charges	Score	Intensity
SNKGAIIGLMVGGVVIA	1597.923	2	193.62	128650000
EDVGSNKGAIIGLMVGGVVIA	1998.082	2	230.96	47933000
LVFFAEDVG	995.4964	1;2	187.14	34551000
AEDVGSNKGAIIGLMVGGVVIA	2069.119	2;3	188.63	30876000
LVFF	524.2999	1	83.876	30674000
FAEDVGSNKGAIIGLMVGGVVIA	2216.188	2;3	137.73	27233000
LVFFAE	724.3796	1	78.156	12010000
KLVFFA	723.432	2	86.627	9795900
KGAIIGLMVGGVVIA	1396.848	2	113.62	7653200
VHHQKLVFFAEDVGSNKGAIIG	2365.254	4	60.542	6277000
HQKLVFFA	988.5494	2	88.734	5780700
HQKLVFF	917.5123	2	105.42	5674500
QKLVFFA	851.4905	2	107.73	5408800
FAEDVGSNKGAIIGLMVG	1776.908	2	196.02	5319700
LVFFAED	839.4065	1	97.64	5136900
DVGSNKGAIIGLMVGGVVIA	1869.039	2	171.99	4744300
GAIIGLMVGGVVIA	1268.753	2	209.69	4611200
AEDVGSNKGAIIGLMVG	1629.84	2	131.79	4379700
FFAEDVGSNKGAIIG	1523.762	2	119.74	4129000
SNKGAIIG	758.4287	2	87.754	4082500
EDVGSNKGAIIGLMVG	1558.803	2	180.66	3825900
KLVFFAEDVG	1123.591	2	117.89	3698900
SNKGAIIGLMVG	1158.643	2	145.46	3635100
QKLVFFAEDVG	1251.65	2	141.52	2448600
HDSGYEVHHQKL	1448.68	3	152.54	2176100
HQKLVFFAEDVGSNKGAIIG	2129.127	3	79.92	1929400
VFFAEDVG	882.4123	1	94.163	1859300
FAEDVGSNKGAIIGLM	1620.818	2	138.24	1800200
HDSGYEVH	942.3832	2	157.99	1415100
AIIGLMVGGVVIA	1211.731	2	184.16	1413700
SNKGAIIGLMVGGVVI	1526.885	2	138.42	1200600
DAEFRHD	888.3726	2	126.28	1149100
AEDVGSNKGAIIGLM	1473.75	2	171.67	1018900
DAEFRHDSGYEVHHQKL	2066.956	4	87.447	888530
EDVG	418.17	1	65.829	720700
EDVGSNKGAIIG	1158.588	2	120.99	692690

Table 3.S2. Peptide sequences identified in the digest of acid-grown Aβ42 fibrils by catL. Raw data shown in Fig. 2b and compiled in Fig. 2c.

Sequence	Mass	Charges	Score	Intensity
GVVIA	457.29	1	52.856	72661000
LMVGG	475.2465	1	52.856	354050
VGGVV	429.2587	1	52.856	1587100
GVVI	386.2529	1	57.455	2529800
LMVG	418.225	1	57.455	43695000
DVGSNKGAIIG	1029.546	2	57.96	688040
AIIG	372.2373	1	62.263	16345000
FFAEDVG	783.3439	1	69.39	1058000
KGAIIG	557.3537	1	72.338	4239000
KLVF	505.3264	1	76.689	9315000
HDSGYEVH	942.3832	2	77.748	4763900
HQKLV	623.3755	1;2	81.525	5110900
DAEF	480.1856	1	83.129	727750
LVFF	524.2999	1	83.876	11474000
FAEDVG	636.2755	1	86.344	39443000
DAEFRH	773.3457	2	88.734	2328000
FAEDVGSNK	965.4454	2	94.302	808550
DAEFR	636.2867	1;2	96.041	43190000
HQKLVF	770.4439	2	99.228	7677400
FAEDVGSN	837.3505	1	107.32	4062600
DSGYEVH	805.3243	1;2	118.06	3361800
AEDVGSNKGAIIG	1229.625	2	150.46	5898400
DAEFRHDSGYE	1324.532	2;3	152.54	19251000
EDVGSNKGAIIG	1158.588	2	152.97	5194000
FAEDVGSNKG	1022.467	2	172.59	2013500
FAEDVGSNKGAIIG	1376.694	2	188.96	26072000

Table 3.S3. Peptide sequences identified in the digest of monomeric Aβ42 by catL

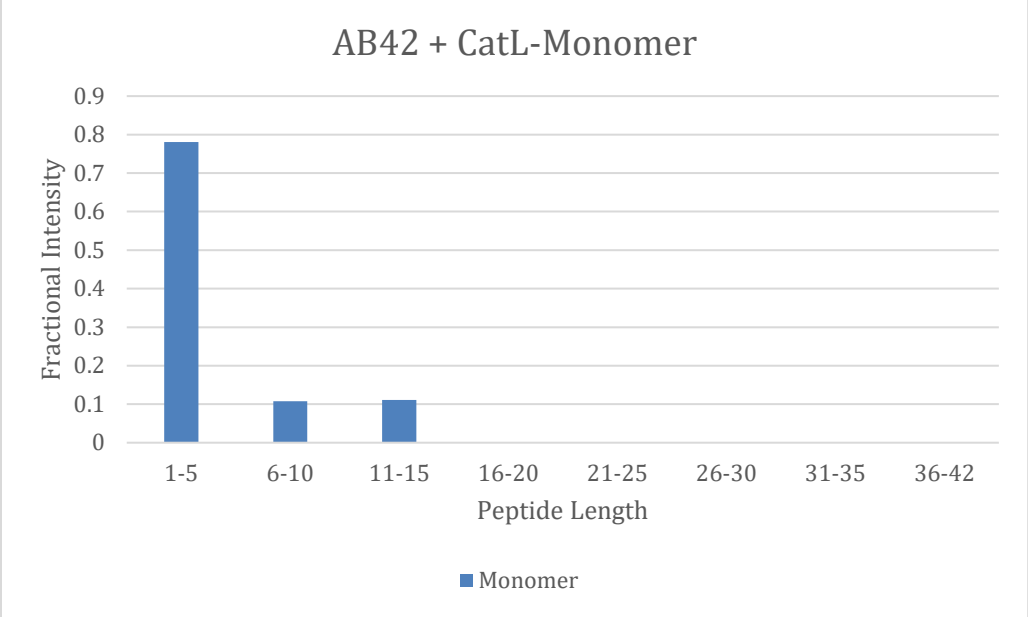


Fig. 3.S1. Bar plot showing the length of the identified product peptides from the digestion of monomeric Aβ42 by catL.

Sequence	Mass	Charges	Score	Intensity
DAEFRHDSGYEVHHQKLVFFA	2531.198	4	61.477	187720000
EVHHQKLVFF	1282.682	3	66.299	70117000
EVHHQKLVF	1135.614	3	66.429	132460000
AEFRHDSGYEVHHQKL	1951.929	4	68.283	23688000
LMVGG	475.2465	1	69.708	42017000
DAEFRHDSGYEVHH	1697.718	4	75.088	46610000
DAEFRHDSGY	1195.489	2;3	79.974	17745000
DAEFRHDSGYEVHHQK	1953.872	4	87.363	4587300
DAEFRHDSGYEVHHQ	1825.777	4	87.824	26314000
DAEFRHDSGYEVHHQKLV	2166.024	4	91.657	170920000
DAEFRHDSGYEVHHQKL	2066.956	4;5	94.339	6924200000
VGSNKGAIIG	914.5185	2	97.635	4584000
DAEFRHDSGYEVHHQKLVFF	2460.161	3;4;5	101.3	1.5882E+10
DAEFRHDSGYE	1324.532	3	102.52	7974700
FRHDSGYEVHHQKLVF	1997.986	4	105.19	32877000
EVHHQKL	889.477	2	107.74	518320
FAEDVGSNKGAIIGLMVGG	1833.93	2	111.72	258920000
VGSNKGAIIGL	1027.603	2	112.71	172900000
AEDVGSNKGAIIGLMVGG	1686.861	2	118.61	139540000
DAEFRHDSGYEVHHQKLVF	2313.093	3;4	120.32	1.1339E+10
RHDSGYEVHHQKLVF	1850.918	4	120.53	23684000
FAEDVGSNKGAIIGLMVG	1776.908	2	120.59	169540000
AEDVGSNKGAIIG	1229.625	2	121.36	180090000
AEFRHDSGYEVHHQKLVF	2198.066	4	122.33	246080000
GSNKGAIIGL	928.5342	2	124.08	26630000
AEDVGSNKGAIIGL	1342.709	1;2	130.01	2068900000
FAEDVGSN	837.3505	1	133.81	16657000
DAEFRHD	888.3726	2	137.18	6093400
FAEDVGSNKGAIIG	1376.694	2	139.31	1624100000
SNKGAIIGL	871.5127	2	148.04	129080000
FAEDVGSNKGAIIGLM	1620.818	2	149.41	711770000
DAEFRHDSGYEVH	1560.659	3	159.44	38405000
FAEDVGSNKGAIIGL	1489.778	2	191.33	1.3544E+10
FAEDVGSNKGA	1093.504	2	194.94	93101000
EDVGSNKGAIIGL	1271.672	2	210.64	269300000

Table 3.S4. Peptide sequences identified in the digest of monomeric Aβ42 by catD.

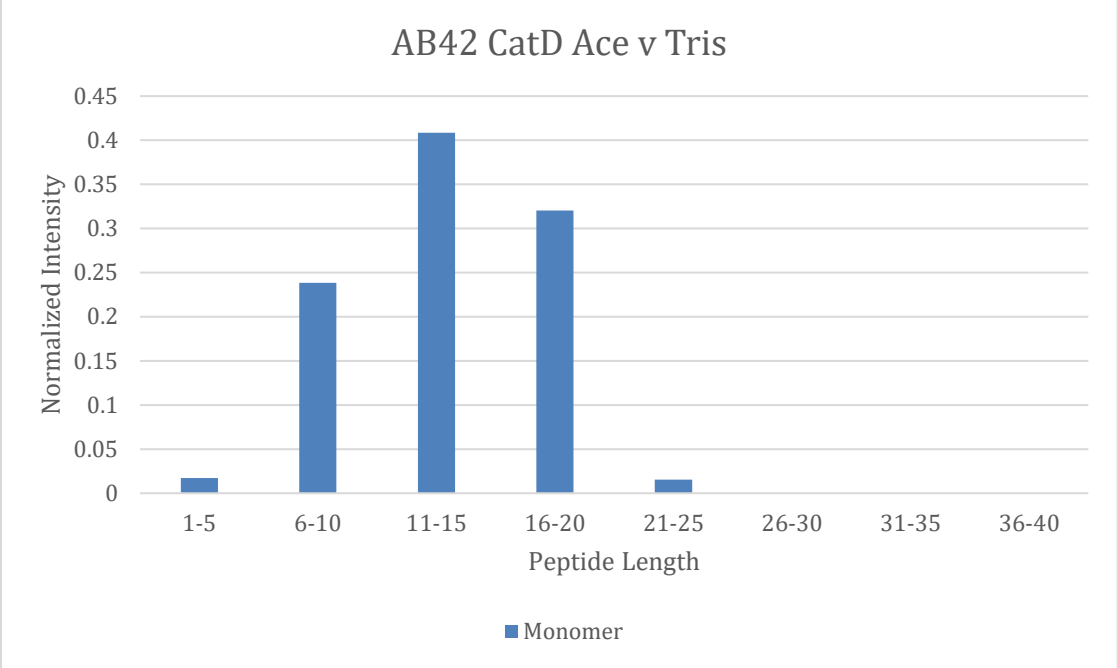


Fig. 3.S2. Bar plot showing the length of the identified product peptides from the digestion of monomeric Aβ42 by catD.

References

[1] Kaushik, S.; Cuervo, A.M. Proteostasis and Aging. *Nat. Med.* **2015**, *21(12)*, 1406-1415.

[2] Mizushima, N. Autophagy: Process and Function. *Genes Dev.* **2007**, *21(22)*, 2861-2873.

[3] Luzio, J.P.; Hackmann, Y.; Dieckmann, N.M.G.; Griffiths, G.M. The Biogenesis of Lysosomes and Lysosome-Related Organelles. *Cold Spring Harbor Perspect. Biol.* **2014**, *6(9)*, a016840-a016840.

[4] Kiselyov, K.; Jennings, J.J., Jr.; Rbaibi, Y.; Chu, C.T. Autophagy, Mitochondria, and Cell Death in Lysosomal Storage Diseases. *Autophagy* **2007**, *3(3)*, 259-262.

[5] Wolfe, D.M.; Nixon, R.A. Autophagy Failure in Alzheimer's Disease and Lysosomal Storage Disorders: A Common Pathway To Neurodegeneration? *Autophagy of the Nervous System*; World Scientific, 2018; pp 237-257.

[6] 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*, 1387-1395.

[7] Bissa, B.; Beedle, A.; Govindarajan, R. Lysosomal Solute Carrier Transporters Gain Momentum in Research. *Clin. Pharmacol. Ther.* **2016**, *100(5)*, 431-436.

[8] Turk, V.; Stoka, V.; Vasiljeva, O.; Renko, M.; Sun, T.; Turk, B.; Turk, D. Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochim. Biophys. Acta* **2012**, *1824*, 68-88.

[9] Turk, B.; Turk, D.; Turk, V. Lysosomal cysteine proteases: More than scavengers. *Biochim. Biophys. Acta.* **2000**, *1477*, 98-111.

[10] Felbor, U.; Kessler, B.; Mothes, W.; Goebels, H.H.; Ploegh, H.L.; Bronson, R.T.; Olsen, B.R. Neuronal Loss and Brain Atrophy in Mice Lacking Cathepsin B and L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99(12)*, 7883-7888.

[11] Koike, M. Cathepsin D Deficiency Induces Lysosomal Storage with Ceroid Lipofuscin in Mouse CNS Neurons. *Neurosci. Res.* **2000**, *38(18)*, S29.

[12] 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*, 606-611.

[13] Quraishi, I.; Nägler, D.K.; Fox, T.; Sivaraman, J.; Cygler, M.; Mort, J.S.; Storer, A.C. The occluding loop in cathepsin B defines the pH dependence of inhibition by its propetide. *Biochemistry* **1999**, *38(16)*, 5017-5023.

[14] Kirschke, H.; Langner, J.; Wideranders, B.; Ansorge, S.; Bohley, P.; Hanson, H. *Acta. Biol. Med. Ger.* **1977**, *36*, 185-199.

[15] Lu, W.D.; Funkelstein, L.; Toneff, T.; Reinheckel, T.; Peters, C.; Hook, V. Cathepsin H functions as an aminopeptidase in secretory vesicles for production of enkephalin and galanin peptide neurotransmitters. *J. Neurochem.* **2012**, *122*, 512-522.

[16] Ghiso, J.; Frangione, B. Amyloidosis and Alzheimer's disease. *Adv. Drug Delivery Rev.* **2002**, *54(12)*, 1539-1551.

[17] Lam, Y.P.Y.; Wootton, C.A.; Hands-Portman, I.; Wei, J.; Chiu, C.K.C.; Romero-Canelon, I.; Lermyte, F.; Barrow, M.P.; O'Connor, P.B. Determination of the Aggregate Binding Site of Amyloid Protofibrils Using Electron Capture Dissociation Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020**, *31(2)*, 267-276.

[18] Wang, H.; Shu, Q.; Rempel, D.L.; Frieden, C.; Gross, M.L. Understanding curli amyloid-protein aggregation by hydrogen-deuterium exchange and mass spectrometry. *Int. J. Mass Spectrom.* **2016**, *420*, 16-23.

[19] Ciechanover, A.; Kwon, Y.T Degradation of misfolded proteins in neurodegenerative diseases: therapeutic targets and strategies. *Exp. Mol. Med.* **2015**, *47*, e147.

[20] Kheterpal, I.; Williams, A.; Murphy, C.; Bledsoe, B.; Wetzel, R. Structural Features of the Aβ Amyloid Fibril Elucidated by Limited Proteolysis. *Biochemistry* **2001**, *40(39)*, 11757-11767.

[21] Chauhan, V.; Sheikh, A.M.; Chauhan, A.; Spivack, W.D.; Fenko, M.D.; Malik, M.N. Fibrillar amyloid beta-protein inhibits the activity of high molecular weight brain protease and trypsin. *J.Alzheimers Dis.* **2005**, *7(1)*, 37-44.

[22] McGlinchey, R.P.; Dominah, G.A.; Lee, J.C. Taking a Bite Out of Amyloid: Mechanistic Insights into α-Synuclein Degradation by Cathepsin L. *Biochemistry* **2017**, *56(30)*, 3881-3884.

[23] Colvin, M.T.; Silvers, R.; Ni, Q.Z.; Can, T.V.; Sergeyev, I.; Rosay, M.; Donovan, K.J.; Michael, B.; Wall, J.; Linse, S.; Griffin, R.G. Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils. *J. Am. Chem. Soc.* **2016**, *138*, 9663-9674.

[24] Zhang, Y.; Rempel, D.L.; Zhang, J.; Sharma, A.K.; Mirica, L.M.; Gross, M.L. Pulsed hydrogen-deuterium exchange mass spectrometry probes conformational changes in amyloid beta (Aβ) peptide aggregation. *PNAS*, **2013**, *110(36)*, 14604-14609.

[25] Colletier, J.-P.; Langanowsky, A.; Landau, M.; Zhao, M.; Soriaga, A.B.; Goldschmidt, L.; Flot, D.; Cascio, D.; Sawaya, M.R.; Eisenberg, D. Molecular Basis for amyloid- β polymorphism. *PNAS*, **2011**, *108(41)*, 16938-16943.

[26] Fändrich, M.; Nyström, S.; Nilsson, K.P.R.; Böckmann, A.; LeVine III, H.; Hammarström, P. Amyloid fibril polymorphism: a challenge for molecular imaging and therapy. *J. Intern. Med.* **2018**, *283*, 218-237.

[27] Lu, J.-X.; Qiang, W; Yau, W.-M.; Schwieters, C.D.; Meredith, S.C.; Tycko, R. Molecular Structure of β-Amyloid Fibrils in Alzheimer's Disease Brain Tissue. *Cell*, **2013**, *154*, 1257-1268.

[28] Wälti, M.A.; Ravotti, F.; Arai, H.; Glabe, C.G.; Wall, J.S.; Böckmann, A.; Güntert, P.; Meier, B.H.; Riek, R. Atomic-resolution structure of a disease-relevant Aβ(1-42) amyloid fibril. *PNAS*, **2016**, E4976-E4984.

[29] Colvin, M.T.; Silvers, R.; Ni, Q.Z.; Can, T.V.; Sergeyev, I.; Rosay, M.; Donovan, K.J.; Michael, B.; Wall, J. Linse, S.; Griffin, R.G. Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils. *J. Am. Chem. Soc.* **2016**, *138*, 9663-9674.

[30] Su, Y.; Chang, P.-T. Acidic pH promotes the formation of toxic fibrils from β-amyloid peptide. *Brain Res.* **2001**, *893*, 287-291.

[31] Zheng, L.; Cedazo-Minguez, A.; Hallback, M.; Jerhammar, F.; Marcusson, J.; Terman, A. Intracellular distribution of amyloid beta peptide and its relationship to the lysosomal system. *Transl. Neurodegener.* **2012**, *1(19)*, 1-7.

[32] Shi, H.; Li, H.; Gong, W.; Gong, R.; Qian, A.; Lee, J.Y.; Guo, W. Structural and Binding Properties on Aβ Mature Fibrils Due to the Histidine Tautomeric Effect. *ACS Chem. Neurosci.* **2019**, *10*, 4612-4618.

[33] Paravatsu, A.K.; Petkova, A.T.; Tycko, R. Polymorphic Fibril Formation by Residues 10-40 of the Alzheimer's β-Amyloid Peptide. *Biophys. J.* **2006**, *90*, 4618-4629.

[34] Brännström, K.; Islam, T.; Sandblad, L.; Olofsson, A. The role of histidines in amyloid β fibril assembly. *FEBS Lett.* **2017**, *591*, 1167-1175.

[35] Ryan, T.M.; Caine, J.; Mertens, H.D.T.; Kirby, N.; Nigro, J.; Breheney, K.; Waddington, L.J.; Streltsov, V.A.; Curtain, C.; Masters, C.L. Roberts, B.R. Ammonium

hydroxide treatment of A beta produces an aggregate free solution suitable for biophysical and cell culture characterization. *PeerJ.* 1:e73, **2013**.

[36] Wood, S.J.; Maleeff, B.; Hart, T.; Wetzel, R. Physical, Morphological, and Functional Differences between pH 5.8 and 7.4 Aggregates of the Alzheimer's Amyloid Peptide Aβ. *J. Mol. Biol.* **1996**, *256*, 870-877.

[37] Biancalana, M.; Koide, S. Molecular Mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta, Proteins and Proteomics.* **2010**, *1804(7)*, 1405-1412.

CHAPTER 4: A Two-Trick Pony: Lysosomal Protease Cathepsin B Possesses Surprising Ligase Activity

Abstract

Cathepsin B is an important protease within the lysosome, where it helps recycle proteins to maintain proteostasis. It is also known to degrade proteins elsewhere but has no other known functionality. However, by carefully monitoring peptide digestion with liquid chromatography and mass spectrometry, we observed synthesis of novel peptides during cathepsin B incubations. This ligation activity was explored further with a variety of peptide substrates to establish mechanistic details and was found to operate through a two-step mechanism with proteolysis and ligation occurring separately. Further explorations using varied sequences indicated increased affinity for some substrates, though all were found to ligate to some extent. Finally, experiments with a proteolytically inactive form of the enzyme yielded no ligation, indicating that the ligation reaction occurs in the same active site but in the reverse direction of proteolysis. These results clearly establish that cathepsin B can act as both a protease and ligase, although protease action eventually dominates over longer periods of time.

Introduction

Cathepsin B (catB) is a cysteine protease and one of several members of the cathepsin family, which are localized primarily in lysosomes. The cathepsins serve to digest proteins and peptides sent to the lysosome for recycling, breaking down substrates into their constituent amino acids which are released for future synthesis or energy production. Like

most of the cathepsin family, catB possesses endoprotease activity, allowing it to cleave between internal residues in a protein sequence. However, catB is best known for its exopeptidase activity, where it preferentially removes two amino acids at a time from the C-terminus of a peptide substrate. This uncommon dual functionality among the cathepsins is due to the presence of an occluding loop comprised of 20 amino acids which blocks one end of the active site and restricts the space to accommodate only two residues past the catalytic cysteine.[1] This unique two-residue slot is shown adjacent to the active site and catalytic residue of catB in Fig. 4.1.[2,3] In addition to steric allowances, the occluding loop contains two histidine residues at positions 110 and 111 that bind to the properly positioned C-terminal carboxylate and facilitate exopeptidase activity.

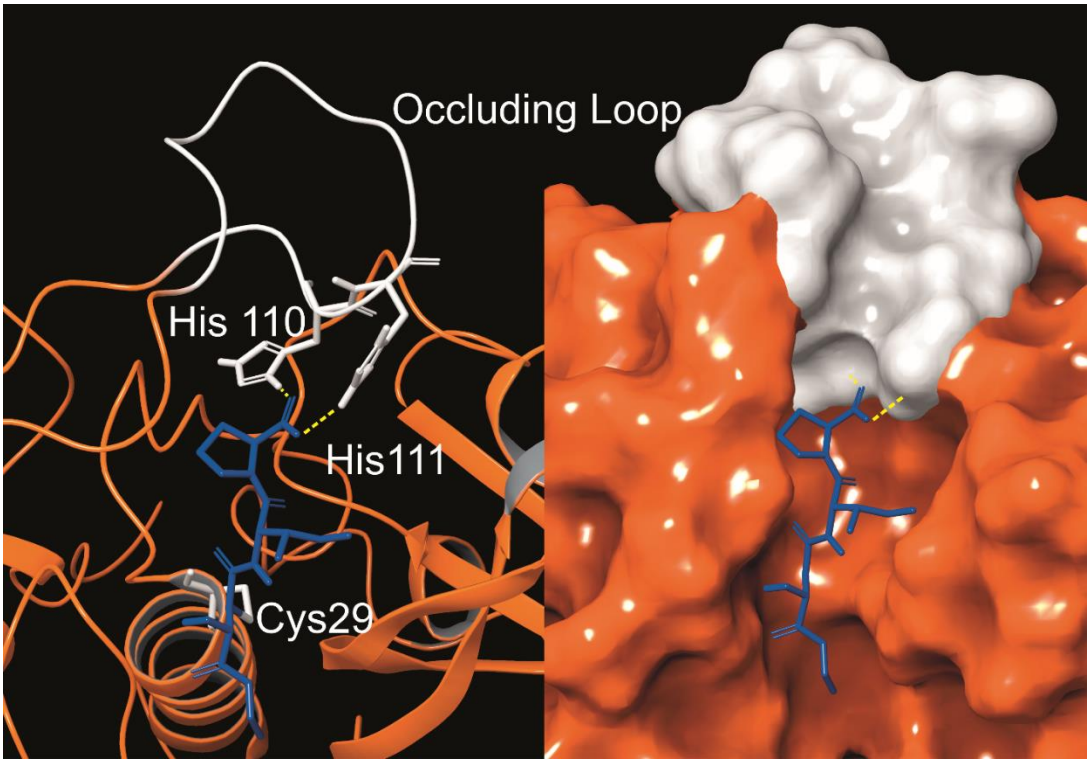


Figure 4.1. The active site of catB shown with a bound substrate inhibitor in ribbon representation (left) and molecular surface representation (right). The occluding loop and catalytic cysteine are shown in white, and the substrate is shown in blue. His110 and His111 from the occluding loop bind to the C-terminus of a substrate to favor carboxypeptidase activity. Images constructed from PDB ID:1CSB.

In addition to the lysosome, catB is found in a variety of cellular locations where it is involved in proteolysis. One secondary role is the activation of enzymes such as trypsin, β-Galactosidase, and renin after synthesis of the respective proenzymes.[4] In antigen presenting cells, catB is thought to be involved in pruning sequences along with other cathepsins in endosomes before selection by the major histocompatibility complex.[5] CatB has also been implicated in apoptosis, where its release into the cytosol after lysosomal membrane permeabilization cleaves the apoptotic protein Bid into its active form.[6] Another role catB plays outside the lysosome is degradation of the extracellular matrix, often to

promote tumor growth in cancer.[7] Like other lysosomal proteases, catB is most active at acidic pH, however it can function at near-neutral pH in the cytosol.[8] Due to the loss of salt-bridges required for stabilization of the occluding loop, catB operates primarily as an endopeptidase at higher pH. These secondary roles for catB demonstrate its versatility but do not reveal any known function outside proteolysis.

While peptide bonds are facile to cleave using proteases or chemical methods such as acid hydrolysis[9], synthesis of new bonds in aqueous solution is a more difficult challenge. Due to their high stability, carboxylic acids are not naturally reactive enough for amines to attack and spontaneously form a peptide bond. Instead targets must be modified into a more reactive form first. Biologically, new peptides are formed in the ribosome using an active ester as a target for a nucleophilic amine attack, binding the N-terminal amine from an amino acid to the carbonyl from the C-terminus of the growing peptide chain.[10] In vitro methods commonly exploit this same type of ester-amine reaction, most famously in the solid-phase peptide synthesis and NHS-ester reactions.[11] Ester-amine reactions are non-specific to the termini however, and unlike ribosomal synthesis which controls the reaction sterically, any free amine such as a lysine can form an incorrect linkage. For this reason, these reactions require careful protection and subsequent de-protection of the side-chains to increase the reaction efficiency. Native chemical ligation is an alternative to standard ester-based reactions, utilizing a thioester and a cysteine instead.[12] It has the advantage of operating orthogonally to canonical amino acids, so no protection of peptide side-chains is required. However, this method still requires chemical modification to produce a reactive

C-terminal thioester group which is not found naturally in peptides, and the reaction is also limited to sites with a cysteine residue.

In biology, ligation of existing peptides is performed by enzymes known as ligases, which catalyze connections between the N-terminus and C-terminus.[13] These ligases are rare however, and typically highly specific in their natural use, often for antigen preparation for the immune system such as in the case of Butelase 1.[14] Some proteolytic enzymes with known ligase activity have been engineered to increase the favorability of the ligation reaction.[15] In the most notable case, residues of the active site of subtilisin were modified through mutagenesis, leading to a significant increase in ligation efficiency when used with activated esters.[16,17] Trypsin has also been recently engineered into a more efficient ligase known as Trypsiligase, although it is limited to a specific sequence motif for coupling.[18] In both of these enzymes, modified substrates containing esters or thioesters have been used to produce ligation products in a high yield.[19] These ligases have been used for site-specific reactions such as the labelling of free N-termini in complex samples which does not produce side-products at lysine residues unlike chemical means.[20,21]

Herein, we examine incubations of catB with synthetic peptide sequences via liquid chromatography and mass spectrometry (LC-MS) to evaluate its ability to perform previously unreported ligation reactions between peptide substrates. Ligated products were detected in multiple incubations, and we explored the sequence specificity for the reaction by substituting amino acids with different functional groups. Additional experiments using a mutated and non-proteolytic form of the enzyme revealed the role of the active site residues and provide further information about the ligation mechanism.

Experimental Procedures

Peptide Synthesis

All peptides were synthesized following an accelerated FMOC-protected solid-phase peptide synthesis protocol.[11] Following synthesis, peptides were purified on a Varian ProStar HPLC using a Phenomenex Jupiter Proteo C12 4 μm 90 Å 250 mm × 4.6 mm column and then validated by LC-MS. Purified peptides were stored frozen in water and vacuum-dried before reconstitution in appropriate buffers.

Cathepsin B Incubations

Cathepsin B for initial incubations was purchased from Athens Research. Purified peptides were dissolved in 50 mM acetate buffer pH 5.5 to a final concentration of 0.5 mg/mL with 1 mM EDTA and 1 mM DTT to prevent active site oxidation. Timepoint aliquots were removed after 2 hours and diluted to a final concentration of 5 μM with 0.2% TFA solution to pH quench the enzymatic reaction. Digested samples were frozen until analysis.

LC-MS Analysis

Samples were analyzed on an Agilent 1100 HPLC interfaced with a ThermoFisher LTQ using an electrospray ionization source. During HPLC, peptides were separated on a ThermoScientific BetaBasic C-18 3 μM, 150 x 2.1 mm column using water with 0.1% formic acid for mobile phase A, and acetonitrile with 0.1% formic acid for mobile phase B. Samples were separated with a gradient of 5-70% phase B over 32.5 minutes.

Results and Discussion

While evaluating LC-MS results for the stepwise proteolysis of a synthetic peptide APSWFDTGLSEMR, an unusual product corresponding to excision of an internal sequence (APSWFDTGMR) was observed as shown in Fig. 4.2a. This result appeared to suggest that the lysosomal protease catB had removed the internal residues LSE from the peptide or had otherwise cleaved the peptide multiple times and stitched the ends back together. Careful evaluation confirmed that the APSWFDTGMR product was not present in the initial sample and must have been produced during incubation, as shown by the data in Fig. 4.S1. The identity of this product was further examined by collision-induced dissociation (CID), yielding characteristic b and y ions covering most of the sequence as shown in Fig. 4.2b. Importantly, the b_8 and b_9 ions map out the unexpected C-terminal linkage, definitively confirming the proposed ligation product. Generation of this new ligated peptide product requires initial cleavage of peptide bonds 8 and 11, followed by creation of a new peptide bond between Gly8 and the methionine in MR. The act of rejoining peptide fragments, or apparent peptide ligase activity, has not been previously reported for catB. Surprisingly, the ligation product comprises almost 20% of the detected peaks after 2 hours of incubation, indicating that ligation can be reasonably favorable under appropriate conditions.

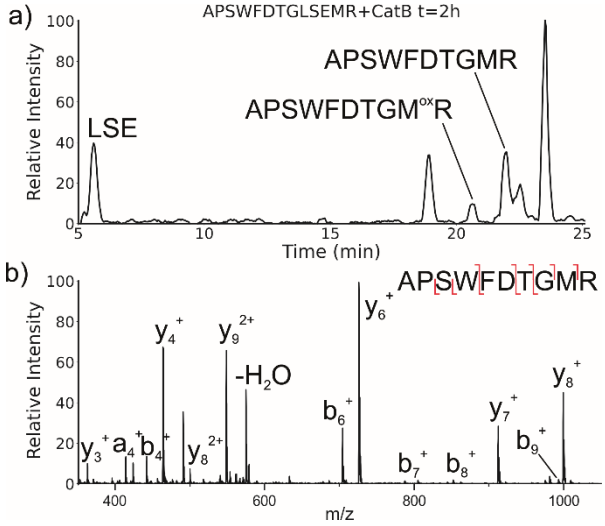
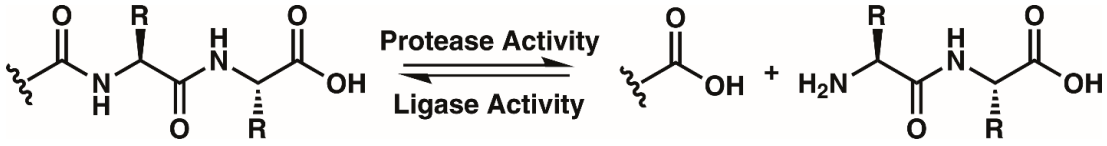


Figure 4.2. (a) LC chromatogram for the incubation of APSWFDTGLSEMR with catB. The labelled fragment APSWFDTGMR is produced by protease cleavage first followed by ligation of the dipeptide MR onto the proteolytic fragment APSWFDTG. (b) CID fragmentation of the ligated peak APSWFDTGMR.

While this behavior has previously gone unnoticed in catB, dual protease/ligase ability has been observed in other enzymes such as Subtiligase, Butelase 1, and the enzyme Legumain.[22] Legumain participates in the MHC antigen cleaving pathway but can also ligate cyclic peptides that are involved in plant defense mechanisms. Unlike other dual-activity enzymes however, Legumain ligation operates through a reactive succinimide not directly associated with the proteolytic residue. Instead, the succinimide lies directly adjacent to the histidine of the catalytic triad, which allows it to share the same active site as the proteolytic machinery. For catB, ligase activity appears to target dipeptides such as MR, which may suggest involvement of the same occluding loop and catalytic site responsible for exopeptidase action. Ligase behavior would simply constitute reversal of the protease action as illustrated in Scheme 4.1.



Scheme 4.1. The protease activity of CatB removes two amino acid residues as a dipeptide from the C-terminus of a substrate, while the ligase activity attaches a dipeptide to the C-terminus of a substrate.

To elucidate whether the internal removal and ligase activity occurs in one step or multiple steps, two synthetic sequences, APSWFDTGLSEMR and GPSWFDTGLSEGR were incubated with catB as shown in Fig. 4.3a. While a one-step process would yield the original sequences minus the internal fragment, a two-step process could lead to scrambling of the C-terminal appendages. The color-coded sequences shown in Fig. 4.3 confirm that scrambling occurred, suggesting that the two C-terminal amino acids are first removed and released, followed by subsequent reattachment following removal of LSE. This illustrates not only that the process occurs in two steps, but also that it is not sequence-specific to the MR portion of the original APSWFDTGLSEMR peptide. Quantitation of the homogeneous ligated products which are comprised of only the original sequence versus the heterogeneous products comprised of mixed sequences is shown in Fig. 4.3b. The

amount of homogeneous ligation is significantly higher, perhaps due to the nearby localization of both necessary pieces after proteolysis.

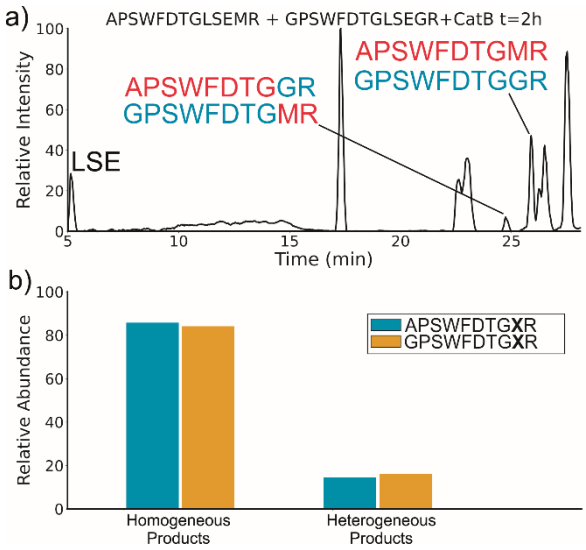


Figure 4.3. (a) LC chromatogram of the incubation of APSWFDTGLSEMR and GPSWFDTGLSEGR with CatB. Multi-colored labels indicate products formed by ligation of fragments from different starting sequences. (b) Relative quantitation of the homogeneous versus heterogeneous ligation products.

To further explore the sequence specificity, modifications were made to the C-terminal residue to determine if it operates as a key factor in the ligase activity (see Fig. 4.4, derived from data in Fig. 4.S2). The sequences APSWFDTGLSEMR and GPSWFDTGLSEGR were incubated with the synthetic dipeptides GK and GA, to evaluate whether arginine or any basic residue is required. In this incubation, ligation was detected for both the lysine and alanine substitutions, indicating that neither arginine nor a basic residue is a necessary element in ligation. With this knowledge, an incubation of both dipeptides with a new sequence, VFFAEDVGSNK, was performed and is shown in Fig. 4.5. In this incubation, the full sequence with appended dipeptide GA was detected, as well as the proteolytic

fragment VFF with ligated dipeptides GA and GK in separate peaks. These fragments, especially the appended full sequence, demonstrate that the previous internal remove of LSE before ligation is a coincidental action resulting from competing protease activity, and that proteolysis is not required for ligase activity. The low abundance of these fragments also shows that while the ligase activity ostensibly works for a variety of sequences, it appears to be affected by substrate affinity.

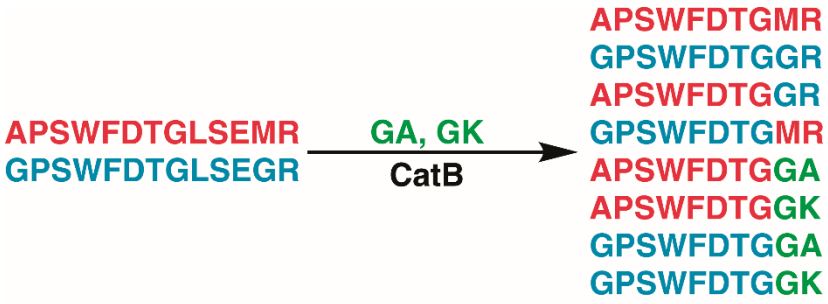


Figure 4.4. Incubation of APSWFDTGLSEMR and GPSWFDTGLSEGR with dipeptides GA and GK results in a series of heterogeneous and homogeneous ligation products. Raw LC chromatogram shown in Supporting Information Figure 4.S1.

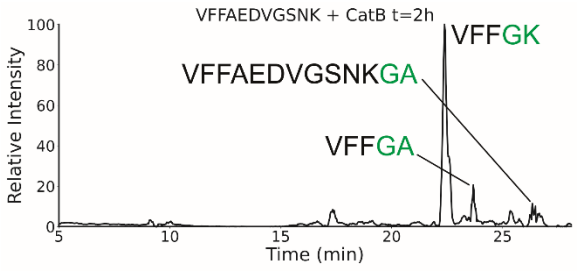


Figure 4.5. Extracted ion chromatogram for the incubation of VFFAEDVGSNK with CatB. Ligation is detected on proteolytic fragments as well as the full intact sequence.

This illustrates why ligase activity may have gone unnoticed previously, as the initial incubation substrate APSWFDTG appears to have a high affinity for ligase activity, allowing for easy detection while other ligation products are harder to detect. The

competing protease activity also serves to break down any ligated products, and by the end of typical incubations lasting 12 hours or more, these sequences have been fully digested, erasing the evidence. No fragments larger than two residues were attached to the C-terminus of a peptide in any incubation, further implying that the occluding loop may play a role in ligation.

Experiments with isomeric substitutions were conducted to further probe the limits of catB ligation. An incubation of catB with the L-isoAsp isomer of APSWFDTGLSEMR resulted in the ligated fragment FDTGMR, as shown in Fig. 4.6a. The identity of the fragment was confirmed with CID fragmentation as shown in Fig. 4.S3. This ligation product is also obtained by removal of the internal LSE sequence, however this product does not appear in the incubation with the canonical sequence shown in Fig. 4.2, which does not show any abundant cleavage at the W-F bond. W-F cleavage results seemingly from the recognition of the isoAsp side-chain as a C-terminus. The altered connectivity of L-isoAsp reduces the side-chain length by one carbon while similarly lengthening the backbone connection. This new orientation produces a structural analogue of a normal C-terminal carboxylate group, as illustrated in Fig. 4.6b. When this C-terminal mimic is bound by catB, the FD residues from APSWFDTGLSEMR sit in the pocket normally reserved for the last two residues of a sequence, and cleavage occurs at the W-F bond as it would to remove a dipeptide. This unintended behavior has not been previously reported for catB but has been noted for other carboxypeptidases.[23]

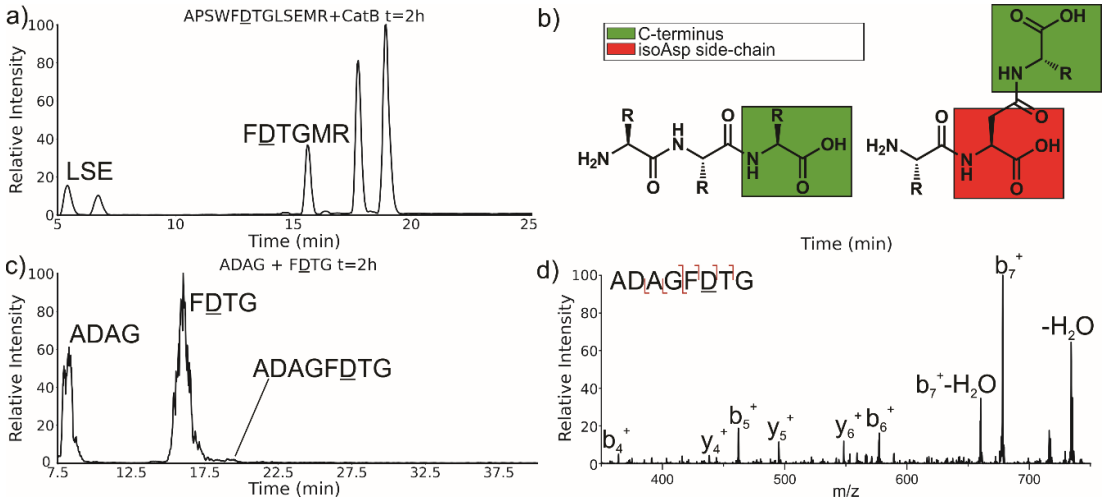


Figure 4.6. (a) LC chromatogram of the incubation of APSWFDTGLSEMR. Incubation produces the ligated fragment FDTGMR which has been cleaved near the L-isoAsp residue due to recognition by catB. (b) Structures of a canonical peptide (left) and L-isoAsp containing peptide (right). Due to the altered connectivity of L-isoAsp, the side-chain structurally resembles a C-terminus and can be recognized by catB as a binding target. (c) LC chromatogram for the incubation of ADAG and FDTG with catB. The incubation produces the ligated fragment ADAGF(isoAsp)TG, CID fragmentation shown in (d).

The ability of catB to recognize L-isoAsp as a C-terminal residue suggests that it could also serve as a site for ligation. In order to explore this possibility, two sequences were incubated together with catB, ADAG and FDTG, modelled after the ligated fragment from Fig. 4.6a. Indeed, after a period of incubation followed by LC-MS analysis, the ligated fragment ADAGFDTG was identified in a small quantity as shown in Fig. 4.6c. This peptide is the first ligation product that does not correspond to attachment of a simple dipeptide. In this case, the first two residues of FDTG take on the role of a dipeptide and are ligated together with ADAG. These results reveal that a peptide with an L-isoAsp in the second position can be attached to another peptide regardless of the size limitation normally imposed by the occluding loop. As L-isoAsp can be easily repaired to L-Asp by the repair enzyme Protein L-isoaspartyl methyltransferase[24], this method of utilizing L-

isoAsp to ligate longer sequences could be used to produce longer canonical peptide sequences. While Butelase 1 has been previously shown to work on peptides with epimerized residues[25], this is the first report of a ligase recognizing an isoAsp residue. These specificities indicate the catB could be used as a complement to N-terminal ligases, since dipeptides and isoAsp-modified peptides will only be added to the C-terminus and regular peptides will not be recognized by the occluding loop.

The implied participation of the occluding loop by preferential ligation of dipeptides (or dipeptide mimics) suggests that the same active site may be used for both proteolysis and ligation in catB. To probe whether the same active site is required for ligation, incubations were performed with a mutated form of catB. The C29A mutant lacks the catalytic cysteine required for proteolysis, as demonstrated previously.[26] Incubation of catB-C29A with APSWFDTG and the dipeptide MR is shown in Fig. 4.7a. This incubation did not yield any ligated APSWFDTGMR products when investigated by LC-MS analysis. Within the same timeframe, wild type catB was able to produce the necessary fragments (APSWFDTG and MR) via proteolysis of APSWFDTGLSEMR, and then ligate them together again into the target peptide APSWFDTGMR shown in Fig. 4.7b. The results in Fig. 4.7b also represent a replicate of those shown in Fig. 4.2a, with the exception of a completely different source of catB. Overall, the results indicate that the proteolytic cysteine is necessary for ligation and provide evidence that ligation operates by reversal of the proteolytic mechanism.

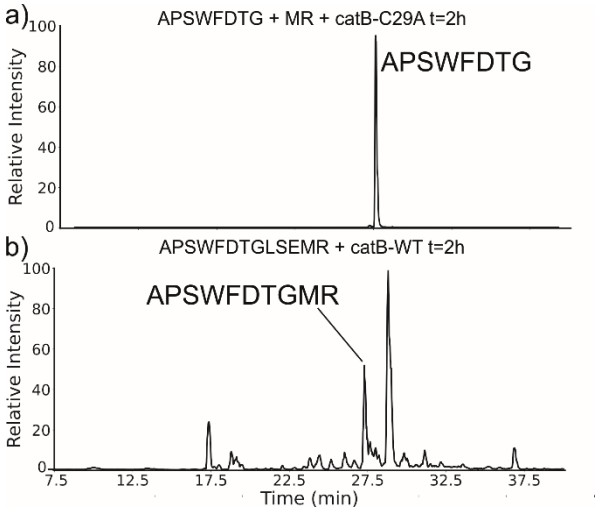


Figure 4.7. (a) LC chromatogram of the incubation of APSWFDTG and peptide MR with catB-C29A mutant. No amount of ligated product was detected. (b) Incubation of the peptide APSWFDTGMR with wild type catB produces the ligated peak APSWFDTGMR in the same timeframe.

Conclusions

In summary, unique ligase activity was identified in catB, which was previously only recognized as a protease. Explorations of this reaction revealed that it has broad sequence tolerance though varied substrate affinity. Dipeptides are preferentially coupled to the C-terminus in the ligation, but we demonstrate that catB can produce sequences of any length by utilizing an isoAsp residue as a C-terminal mimic. A highly specific ligation reaction such as this could be used to produce specific target sequences, to couple protein domains together, or to add modifications such as fluorescent tags to existing peptides without any supporting chemical reactions or protecting groups. Future experiments could be performed to optimize the ligase yield of catB by mutagenesis of the residues in close proximity to the active-site, as in the case of previously engineered enzymes, or through the use of modified peptide substrates. More extensive substrate libraries could also be

screened with catB to determine relative sequence propensities for ligation, opening up new possibilities for targeted substrate design for in vitro and in vivo ligations.

Funding and additional information

The authors are grateful for funding from the NIH (1R01AG066626-01 for RRJ and 1R35GM137901-01 for YZ).

Conflict of Interest: The authors declare no conflicts of interest in regard to this manuscript.

Supporting Information

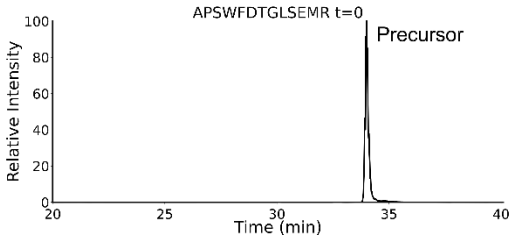


Figure 4.S1. LC chromatogram for the initial timepoint of APSWFDTGLSEMR + CatB. No peaks were present except for the precursor peptide.

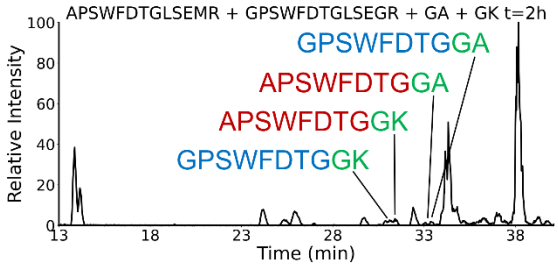


Figure 4.S2. LC chromatogram for the incubation of APSWFDTGLSEMR and GPSWFDTGLSEGR with dipeptides GA and GK with CatB. Raw data for the results shown in Figure 4.4.

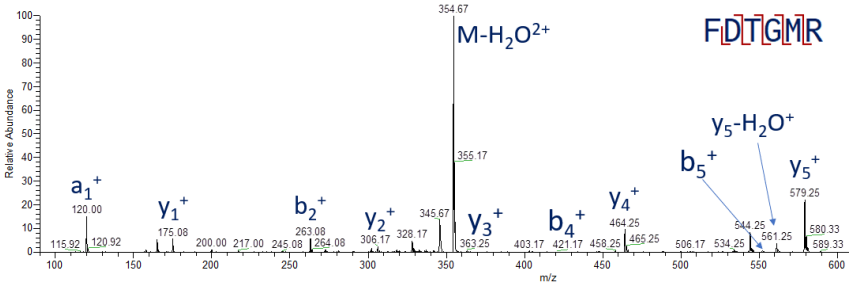


Figure 4.S3. CID fragmentation spectrum for the ligation product FDTGMR.

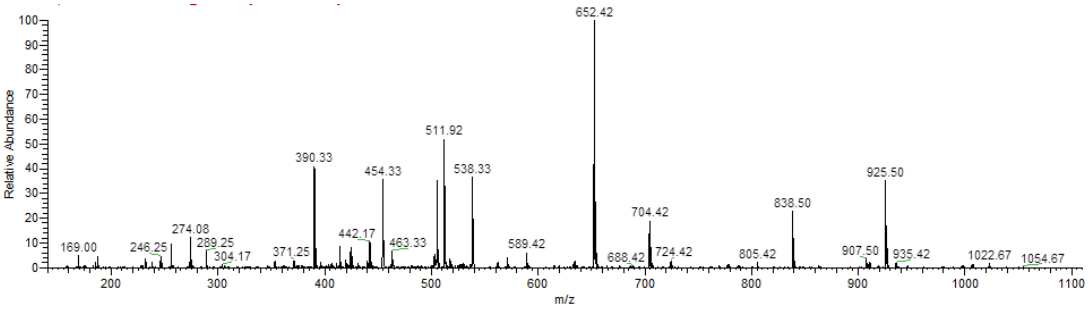


Figure 4.S4. CID fragmentation spectrum for the ligation product APSWFDTGGR shown in Figure 4.3.

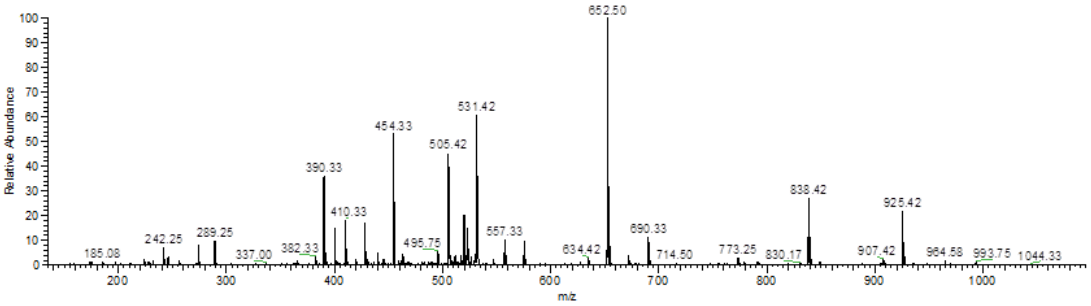


Figure 4.S5. CID fragmentation spectrum for the ligation product GPSWFDTGGR shown in Figure 4.3.

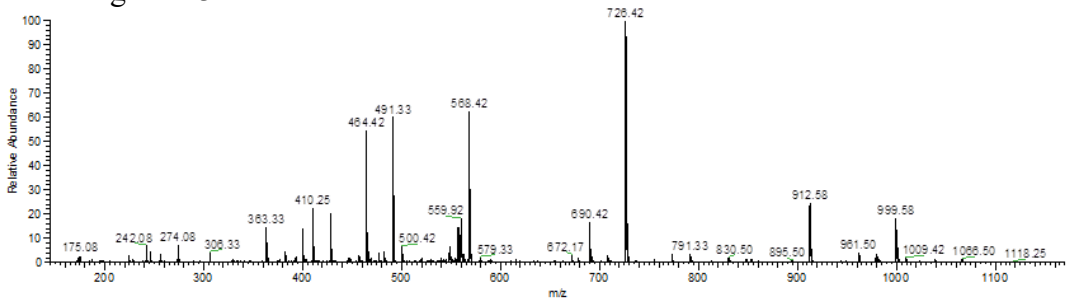


Figure 4.S6. CID fragmentation spectrum for the ligation product GPSWFDTGMR shown in Figure 4.3.

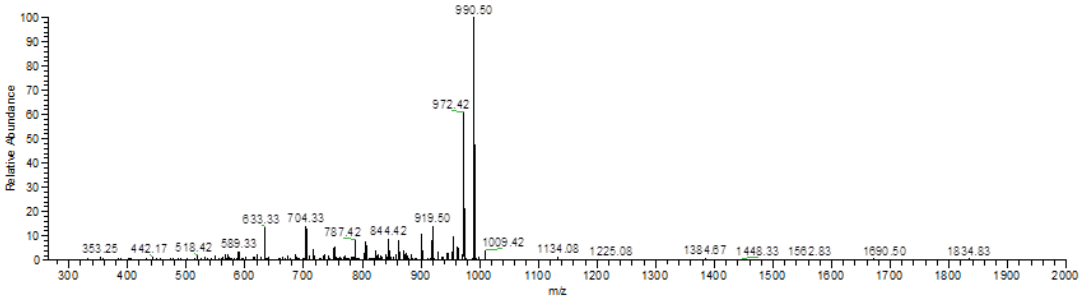


Figure 4.S7. CID fragmentation spectrum for the ligation product APSWFDTGGA shown in Figure 4.4.

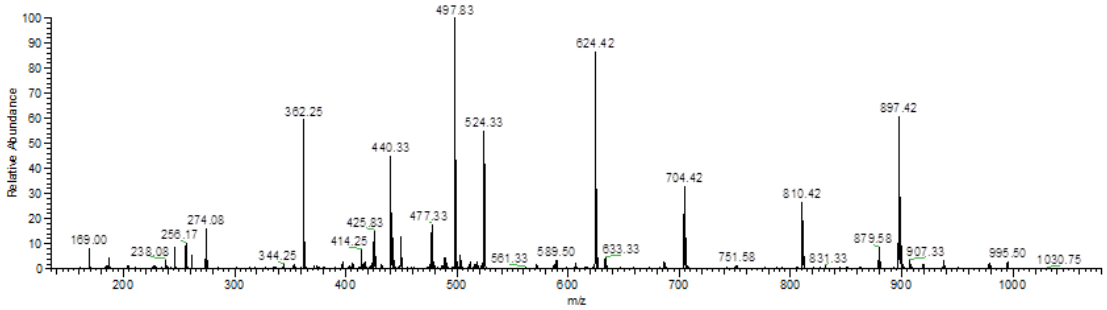


Figure 4.S8. CID fragmentation spectrum for the ligation product APSWFDTGGK shown in Figure 4.4.

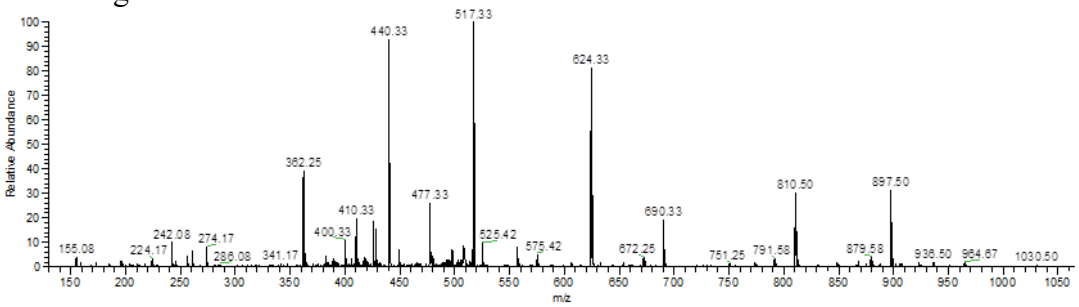


Figure 4.S9. CID fragmentation spectrum for the ligation product GPSWFDTGGK shown in Figure 4.4.

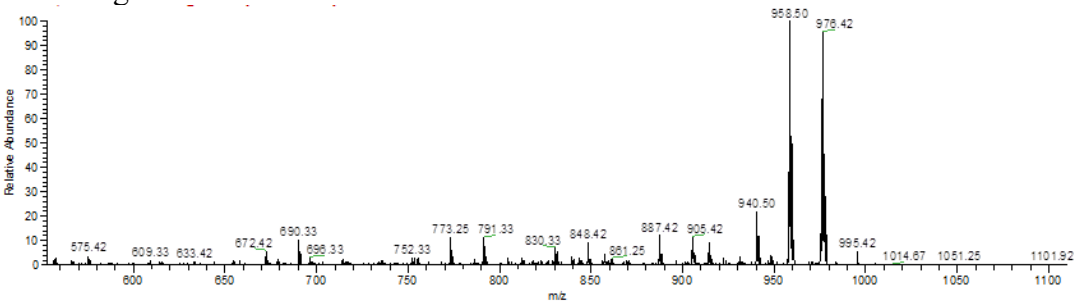


Figure 4.S10. CID fragmentation spectrum for the ligation product GPSWFDTGGA shown in Figure 4.4.

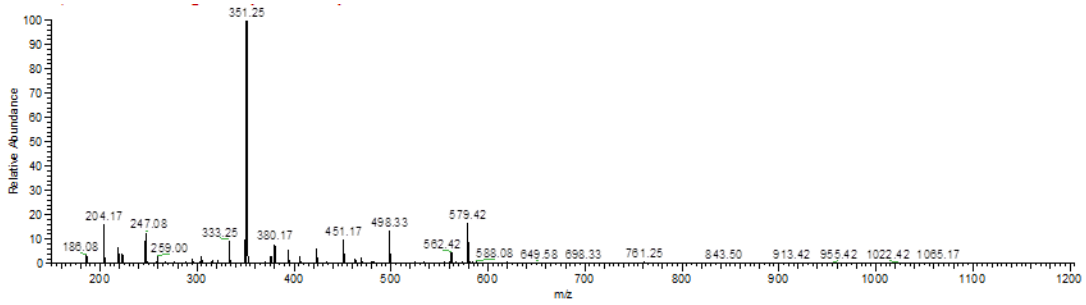


Figure 4.S11. CID fragmentation spectrum for the ligation product VFFGA shown in Figure 4.5.

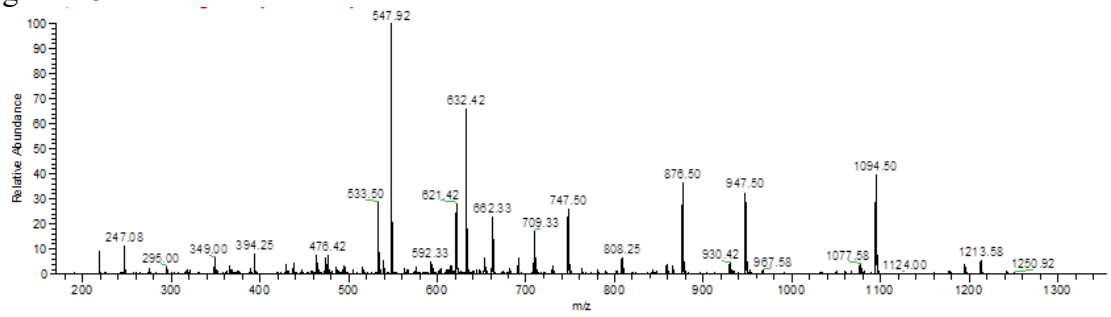


Figure 4.S12. CID fragmentation spectrum for the ligation product VFFAEDVGSNKGA shown in Figure 4.5.

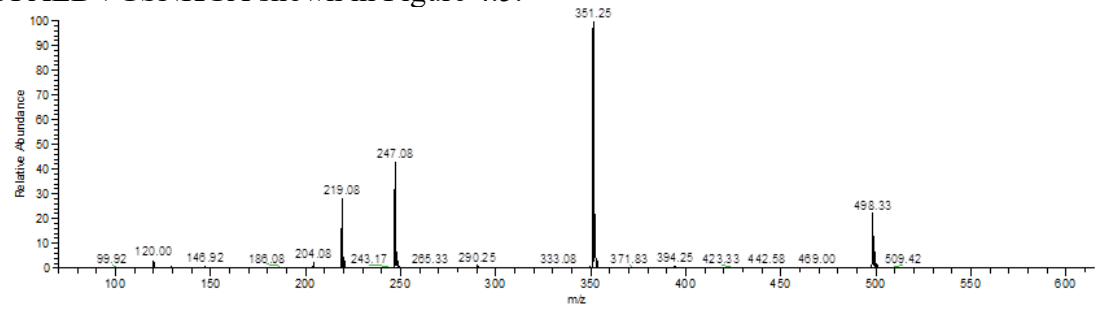


Figure 4.S13. CID fragmentation spectrum for the ligation product VFFGK shown in Figure 4.5.

References

[1] Illy, C.; Quraishi, O.; Wang, J.; Purisima, E.; Vernet, T.; Mort, J.S. Role of the Occluding Loop in CatB Activity. *J. Biol. Chem.* **1996**, *272(2)*, 1197-1202.

[2] Greenspan, P.D.; Clark, K.L.; Tommasi, R.A.; Cowen, S.D.; Mcquire, L.W.; Farley, D.L.; Van Duzer, J.H.; Goldberg, R.L.; Zhou, H.; Du, Z.; Fitt, J.J.; Coppa, D.E.; Fang, Z.; Macchia, W.; Zhu, L.; Capparelli, M.P.; Goldstein, R.; Wigg, A.M.; Doughty, J.R.; S Bohacek, R.; Knap, A.K. Identification of Dipeptidyl Nitriles as Potent and Selective Inhibitors of Cathepsin B Through Structure-Based Drug Design. *J. Med. Chem.* **2001**, *44*, 4524-4534.

[3] Turk, D.; Podobnik, M.; Popovic, T.; Katunuma, N.; Bode, W.; Huber, R.; Turk, V. Crystal Structure of Cathepsin B Inhibited with CA030 at 2.0 Angstrom Resolution: A Basis for the Design of Specific Epoxysuccinyl Inhibitors. *Biochemistry* **1995**, *34* (14), 4791–4797.

[4] Berdowska, I. Cysteine Proteases as Disease Markers. *Clinica Cimica Acta*. **2004**, *342*, 41-69.

[5] Turk, V.; Turk, B.; Turk, D. Lysosomal Cysteine Proteases: Facts and Opportunities. *EMBO J.* **2001**, *20(17)*, 4629-4633.

[6] Blomgran, R.; Zheng, L.; Stendahl, O. Cathepsin-cleaved Bid Promotes Apoptosis in Human Neutrophils Via Oxidative Stress-Induced Lysosomal Membrane Permeabilization. *J. Leukocyte Biol.* **2007**, *81*, 1213-1223.

[7] Jedeszko, C.; Sloane, B.F. Cysteine cathepsins in human cancer. *Biol. Chem.* **2004**, *385*, 1017-1027.

[8] Driessen, C.; Lennon-Duménil, A.; Ploegh, H.L. Individual Cathepsins Degrade Immune Complexes Internalized By Antigen-Presenting Cells Via Fcγ Receptors. *Eur. J. Immunol.* **2001**, *31*, 1592-1601.

[9] Smith, R.M. and Hansen, D.E. The pH-Rate Profile for the Hydrolysis of a Peptide Bond. *J. Am. Chem. Soc.* **1998**, *120(35)*, 8910-8913.

[10] Trobro, S. and Åqvist, J. Mechanism of peptide bond synthesis on the ribosome. *PNAS*. **2005**, *102(35),* 12395-12400.

[11] Hood, C. A.; Fuentes, G.; Patel, H.; Page, K.; Menakuru, M.; Park, J. H. Fast conventional Fmoc solidphase peptide synthesis with HCTU. *J. Pept. Sci.* **2008**, *14(1)*, 97–101.

[12] Dawson, P.E.; Muir, T.W.; Clark-Lewis, I.; Kent, S.B. Synthesis of proteins by native chemical ligation. *Science*. **1994**, *266*, 776-779.

[13] Noike, M.; Matsui, T.; Ooya, K.; Sasaki, I.; Ohtaki, S.; Hamano, Y.; Maruyama, C.; Ishikawa, J.; Satoh, Y.; Ito, H.; Morita, H.; Dairi, T. A peptide ligase and the ribosome cooperate to synthesize the peptide pheanomycin. *Nat. Chem. Bio.* **2015**, *11*, 71-78.

[14] Nguyen, G. K. T.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J.P. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Bio.* **2014**, *10*, 732-740.

[15] Nuijens, T.; Toplak, A.; Schmidt, M.; Ricci, A.; Cabri, W. Natural Occurring and Engineerd Enzymes of Peptide Ligation and Cyclization. *Front. Chem.* **2019**, *7(829)*, 1-8.

[16] Chang, T.K.; Jackson, D.Y.; Burnier, J.P.; Wells, J.A. Subtiligase: A tool for semisynthesis of proteins. *PNAS*. **1994**, *91*, 12544-12548.

[17] Weeks, A.M. and Wells, J.A. Subtiligase-Catalyzed Peptide Ligation. *Chem. Rev.* **2020**, *120(6)*, 3127-3160.

[18] Liebscher, S.; Schöpfel, M.; Aumüller, T.; Sharkhuukhen, A.; Pech, A.; Höss, E.; Parthier, C.; Jahreis, G.; Stubbs, M.T.; Bordusa, F. N-terminal protein modification by substrate activated reverse proteolysis. *Angew. Chem. Int. Ed.* **2014**, *53(11)*, 3024-3028.

[19] Tan, X.; Yang, R.; Liu, C.F. Facilitating Subtiligase-Catalyzed Peptide Ligation Reactions using Peptide Thioester Substrates. *Org. Lett.* **2018**, *20*, 6691-6694.

[20] Wildes, D. and Wells, J.A. Sampling the N-terminal proteome of human blood. *PNAS.* **2010**, *107(10)*, 4561-4566.

[21] Nguyen, G.K.T.; Cao, Y.; Wang, W.; Fa Liu, C.; Tam, J.P. Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide. *Angew. Chem. Int. Ed.* **2015**, *127*, 15920-15924.

[22] Dall, E.; Brandstetter, H. Mechanistic and Structural Studies on Legumain Explain its Zymogenicity, Distinct Activation Pathways, and Regulation. *PNAS*. **2013**, *110(27)*, 10940-10945.

[23] Murray, E.D.; Clarke, S. Synthetic Peptide Substrates for the Erythrocyte Protein Carboxyl Methyltransferase. Detection of a New Site of Methylation at Isomerize L-Aspartyl Residues. *J. Biol. Chem.* **1984**, *259(17)*, 10722-10732.

[24] Lowenson, J.D. and 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*, 5985-5995.

[25] Nguyen, G.K.T.; Hemu, X.; Quek, J.P.; Tam, J.P. Butelase-Mediated Macrocyclization of D-Amino-Acid-Containing Peptides. *Angew. Chem. Int. Ed.* **2016**, *55*, 12802-12806.

[26] Dai, Z.; Cheng, Q.; Zhang, Y. Rational Design of a Humanized Antibody Inhibitor of Cathepsin B. *Biochemistry*. **2020**, *59(14)*, 1420-1427.

CHAPTER 5: Efficient Isothiocyanate Modification of Peptides Facilitates Structural Analysis by Radical-Directed Dissociation

Abstract

Radical-directed dissociation (RDD) is a powerful technique for structural characterization of peptides in mass spectrometry experiments. Prior to analysis, a radical precursor must typically be appended to facilitate generation of a free radical. To explore the use of a radical precursor that can be easily attached in a single step, we conducted experiments to characterize the modification and use of iodophenylisothiocyanate in a "click" reaction. The reaction was found to be selective for amine residues, producing an iodophenylthiourea modification in high yield. Photodissociation yields were characterized at 266 nm and 213 nm for the 2-, 3-, and 4-iodo isomers of the precursor and found to be highest for the 4-iodo isomer in nearly all cases. Radical and collisional fragmentation spectra revealed unique losses for the tag and ab initio calculations identified a mechanism for tag loss involving proton transfer within the thiourea group. Examination of RDD data revealed 4-iodobenzoic acid, 4-iodophenylthiourea, and 3-iodo-tyrosine yield similar fragments for any given peptide, although differences in fragment intensity are noted. Iodophenylisothiocyanate labeling in combination with RDD can be used to differentiate isomeric amino acids within peptides, which should facilitate simplified evaluation of isomers present in complex biological samples.

Introduction

Mass spectrometry (MS) is a powerful technique for characterization of biomolecules due to its ability to accurately measure the mass of intact analytes and obtain further information following fragmentation. A variety of fragmentation methods are available, with differing chemical mechanisms creating fragments in each. The most commonly employed fragmentation method is collision-induced dissociation (CID), which involves heating molecules through collisions with a neutral gas.[1] These collisions impart small amounts of energy into the molecule which build up until the lowest energy dissociation pathway is accessible, at which point fragmentation occurs. In peptides, this method produces primarily b- and y-ions via cleavage of the amide bond. Photodissociation can be employed to impart energy into the analyte via absorbed photons. In the longer wavelength region, infrared multiphoton dissociation (IRMPD) heats a peptide slowly and results in a similar spectrum to CID[2], while ultraviolet photodissociation (UVPD) can be employed at different UV wavelengths to generate diverse fragments due to the ability to access higher energetic pathways.[3,4,5]

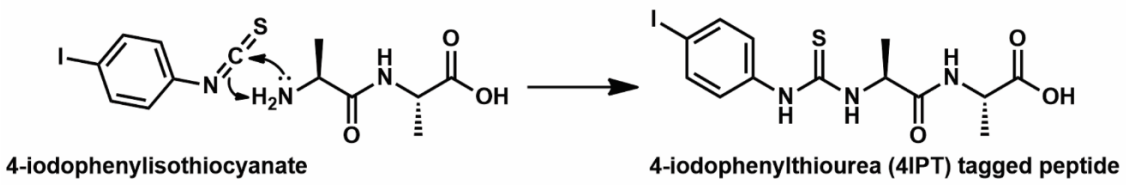
As opposed to these methods which operate by imparting energy into even-electron peptides, distinct fragmentation pathways are observed for activation of odd-electron, radical analytes. In electron-transfer dissociation (ETD) or electron-capture dissociation (ECD), a hydrogen abundant radical is initially created by the incoming electron.[6,7] Dissociation of this fragile species produces c- and z-ions that are not typically observed in even-electron fragmentation spectra. Radical-directed dissociation (RDD) also generates odd-electron species, but operates through a different mechanism. In RDD, a

radical is produced through homolytic cleavage of a labile bond, yielding a hydrogen-deficient molecule that is also susceptible to fragmentation.[8] Photocleavage of a carbon-iodine bond can cleanly produce such radicals with quantitative yields.[9] Suitable carbon-iodine bonds can be incorporated by iodination of tyrosine residues.[10] Alternatively, 4-iodobenzoic acid (4IB) can be appended to the N-terminus or a lysine residue through an NHS-ester reaction.[9,11] In both cases, the iodine atom is coupled directly to a phenyl ring that acts as a chromophore and increases absorption of incident light, increasing the yield of photodissociation. Other methods for radical generation include the use of CID-cleavable tags such as TEMPO[12], nitric oxide[13], and peroxycarbamate.[14]

In a hydrogen-deficient peptide, the radical can migrate to other sites by abstracting a hydrogen atom, relocating the hydrogen to the original radical location and generating a radical in the new site. Addition of collisional energy results in unique fragmentation through bond rearrangement with the radical to produce a-, x-, and z-ions as well as side-chain losses.[15] Radical migration takes place between sites which are spatially proximal, a property that is independent of peptide sequence. The mechanism for radical migration lends RDD high structural sensitivity and the ability to identify isomers in lipids,[16] glycans,[17] and peptides[18,19] based on differences in fragmentation intensities among the various forms.

RDD typically requires covalent modification of analytes prior to analysis. Although not commonly employed with mass spectrometry, isothiocyanate functional group tagging of peptides and proteins is among the family of “click” reactions characterized by rapid reaction time in a one pot synthesis and high yield.[20] Isothiocyanate reacts with

both amines (as illustrated in Scheme 5.1) and thiols, although the thiol product is less stable and the modification is reversible.[21] Isothiocyanates are naturally occurring in some dietary vegetable species and have been studied for their native reactivity with protein functional groups.[22,23] Historically, the Edman degradation procedure was ubiquitously used to sequence proteins by capitalizing on the reaction between protein N-terminal amines and isothiocyanates.[24] The most common modern use of isothiocyanates in protein research is the fluorescent labelling of protein amines utilizing labels such as fluorescein isothiocyanate (FITC)[25,26] or rhodamine isothiocyanate[27]. Isothiocyanate reactions have been employed to attach sulfonic acid groups to protein and peptide digests to aid with gas-phase peptide sequencing in mass spectrometry.[28,29] Isothiocyanate derivatization has additionally been implemented to enhance separation in liquid chromatography coupled to mass spectrometry.[30]



Scheme 5.1. Concerted mechanism for the reaction of a free amine with the isothiocyanate functional group to produce a thiourea linkage.

Herein, we characterize isothiocyanate modified peptides for suitability in RDD experiments. Photodissociation yields at 213 nm and 266 nm were measured as a function of iodine position. Fragmentation following radical initiation was characterized, revealing both typical RDD products and fragments unique to the isothiocyanate tag. A mechanism supported by B3LYP/6-31+G(d) calculations that can account for the unique fragments is

proposed. The effectiveness of isothiocyanate modified peptides for RDD-based structural characterization and differentiation of isomers is evaluated.

Experimental Procedures

Materials

4-iodophenylisothiocyanate was purchased from Oakwood Products Inc. 2- and 3-iodophenylisothiocyanate were purchased from Fisher Scientific. All iodine-containing reagents were used without further purification. RRLIEDNEYTARG was purchased from Bachem Inc. RQpsVELHSPQSLPR was purchased from Anaspec Inc. All other peptides were synthesized using an accelerated FMOC-protected solid-phase peptide synthesis protocol[31] using FMOC-protected amino acid residues purchased from Anaspec Inc.

Modification Reaction

Peptides were dissolved in a 1:1 ratio of 50 mM borate buffer pH 8.5:acetonitrile to a concentration of 250 μM in 40 μL. Iodophenyisothiocyanate was dissolved in acetonitrile to a concentration of 50 mM and 0.2, 1, or 2 μL were added to the peptide for a 1:1, 1:5, or 1:10 peptide:isothiocyanate ratio for 30 minutes. Samples were desalted using a Michrom Bioresources MacroTrap peptide trap and redissolved in 50:50 H_2O:ACN with 0.1% formic acid to a final concentration of 10 μM for electrospray mass spectrometry.

Mass Spectrometry

Electrospray ionization (ESI) at 4 kV spray voltage was used for analysis of peptides in Orbitrap and LTQ mass spectrometers. UVPD at 213 nm was conducted using an Orbitrap Velos Pro with an HCD cell modified with a quartz window for photoactivation

for 50 ms with an FQSS 213-Q4 CryLas laser. UVPD at 266 nm was performed in an LTQ ion trap mass spectrometer modified with a quartz window to allow for photoactivation with an Nd:YAG laser (Continuum, Santa Clara, CA).

Results and Discussion

In this study, model synthetic peptides were incubated with commercially available iodophenylisothiocyanates to modify free amines located at either the N-terminus or a lysine side-chain as shown in Scheme 5.1. The outcome of this reaction is the attachment of an iodophenyl radical precursor to a peptide via a thiourea linker. The resulting peptide contains a phenyl chromophore attached to a photolabile iodine atom that can be irradiated with 266 nm or 213 nm light to produce homolytic cleavage and generate a site-specific radical. Once generated, the radical can migrate to spatially proximal sites throughout the sequence and cleave bonds to produce structurally-dependent fragments. Reaction yields are shown in Fig. 5.1a for the incubation of 3-iodophenylisothiocyanate with RRLIEDNEYTARG as a function of stoichiometric excess and temperature. The yields are estimated based on ion counts from all contributing charge states for the modified and unmodified peptides. The additional iodophenyl group in the modified peptides is hydrophobic, which may enhance ionization and lead to slight over-representation of the modified form.[32] The reported yields therefore represent an upper limit, but are valid for comparison of varying conditions. After 30 minutes of incubation at room temperature, little modification (9.12%) is observed even with 10-fold excess isothiocyanate. However, increasing the temperature slightly to 37°C resulted in a

significantly higher yield (95.71%) for the same 1:10 ratio. Notably, at 37°C even the 1:1 ratio was able to achieve over 70% yield. Isolation and irradiation of the modified peptide is shown in Fig. 5.1b, where application of 213 nm light produces a site-specific homolytic cleavage of the carbon-iodine bond, generating a radical ion. To examine the effect of iodine position on subsequent dissociation products, we modified a series of peptides with the 2-, 3-, and 4-iodophenylisothiocyanates. The respective photodissociation yields following activation at 266 nm are shown in Fig. 5.1c. There is a significant difference observed between the isomers, and the 4-iodo isomer is found to produce the greatest relative yield for all of the examined sequences. The 3-iodo isomer was found to produce a higher or similar yield as the 2-iodo isomer in most cases. The variation in yields between the different sequences is likely due to the presence of additional chromophores found in the aromatic side-chains of tyrosine, phenylalanine, and tryptophan. These chromophores are capable of transferring energy if the peptide structure places them in proximity to the 4IPT tag. The photodissociation yield may also differ based on proximity to charged sites, which can lead to shifts in the absorption profile.[33] Photodissociation yields obtained following 213 nm excitation are shown in Fig. 5.1d. Smaller differences are observed at 213 nm, but the 4-iodo isomer is generally slightly favored at this wavelength as well.

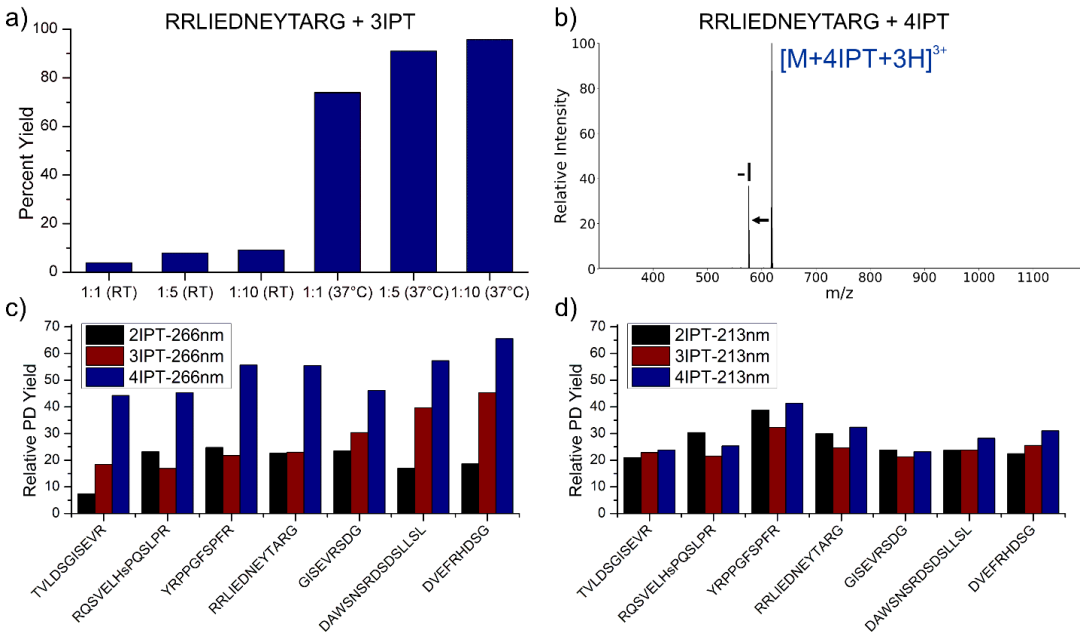


Figure 5.1. a) Reaction yields based on MS data for modification of the peptide RRLIEDNEYTARG with isothiocyanate. Temperature exerts a greater effect than stoichiometric excess. b) Irradiation with 213 nm light produces site-specific cleavage of the iodine atom to generate a radical species. c) Photodissociation yields for a series of peptides irradiated with 266 nm light. d) Photodissociation yields for peptides irradiated with 213 nm light.

Following loss of iodine, collisional activation of the re-isolated radical yields the RDD spectrum, as shown for DVEFRHDSG modified with 4IPT in Fig. 5.2a. A series of a- and z-ions are observed as well as partial and full side-chain losses which are typical of RDD spectra. Additionally, we observed two unique losses which correspond to partial-tag (-pt) and full-tag (-ft) losses of the 4IPT. RDD of the deprotonated species for this peptide is shown in Fig. 5.S1, displaying typical radical fragments in addition to the partial and full tag losses. As opposed to the proton-driven mechanism in CID, the overall peptide charge does not play a central role in radical fragmentation and RDD is relatively unaffected.[34] The CID spectrum of the non-radical peptide shown in Fig. 5.2b also

displays abundant loss of the partial- and full-tag fragments, indicating that these losses are primarily a result of vibrational heating. Further isolation and fragmentation of the partial-tag loss is shown in Fig. 5.2c. In this spectrum, no fragment corresponding to the full-tag loss is observed, indicating that the partial-tag loss is not the first step towards the full-tag loss, rather each is generated independently. CID of the re-isolated full-tag-loss fragment is shown in the top spectrum of Fig. 5.2d, compared to fragmentation of the unmodified peptide in the bottom spectrum. The two spectra are nearly identical, suggesting that the full-tag loss leaves behind the original sequence without discernable rearrangement. Similarly favorable fragmentation has been observed previously in other urea and thiourea containing compounds.[35,36,37]

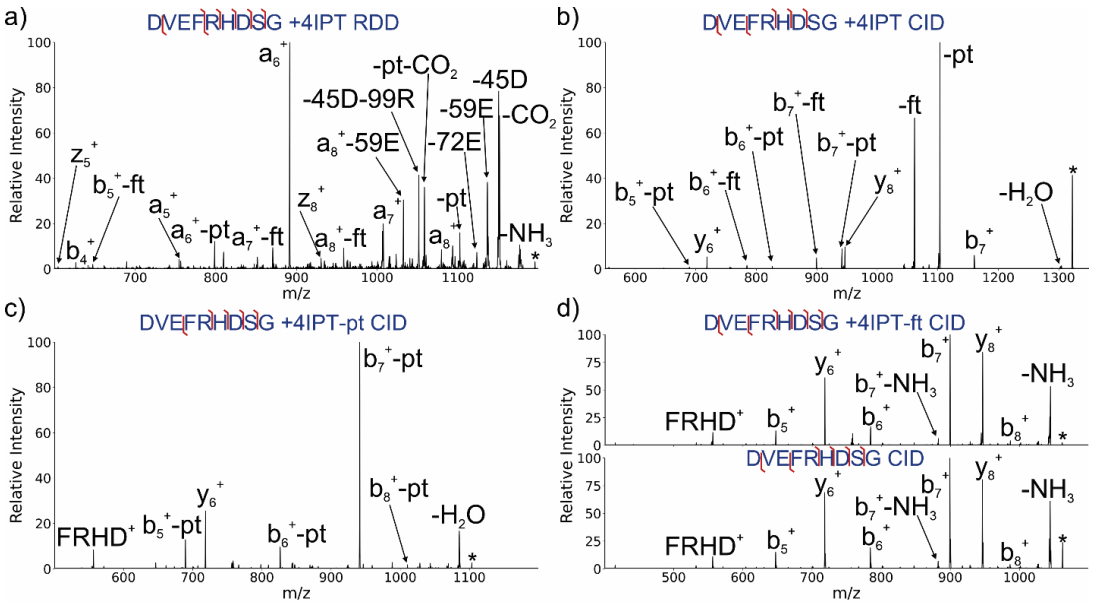
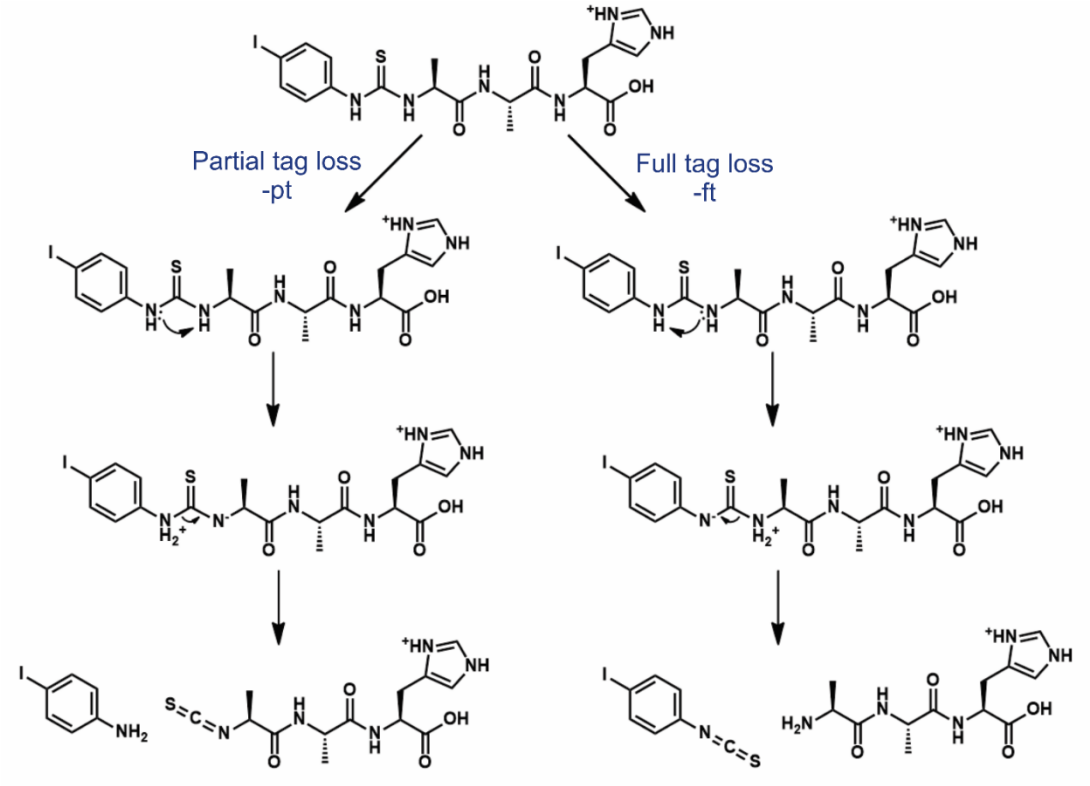


Figure 5.2. a) RDD fragmentation of DVEFRHDSG modified with 4IPT produces a-ions and side-chain losses. Unique losses of the partial tag (-pt) and full tag (-ft) and observed in some fragments. b) CID fragmentation of DVEFRHDSG modified with 4IPT produces b/y-ions and partial/full tag loss. c) CID fragmentation of the partial tag loss does not lead to full tag loss. d) CID fragmentation of the full tag loss (top) matches CID of the unmodified peptide (bottom).

A proposed mechanism for both unique losses is shown in Scheme 5.2. For the partial-tag loss, the first nitrogen of the thiourea abstracts a hydrogen from the second. This results in dissociation of the bond connecting the first nitrogen with the thiourea carbon, and loss of the iodophenyl group with an amine. The sulfur and carbon from the thiourea group remain on the peptide and form an isothiocyanate group at the N-terminus. In a complementary mechanism for the full tag loss, the second nitrogen deprotonates the first, resulting in loss of the full tag with the two terminal groups swapped. The iodophenyl group in this mechanism is lost as an isothiocyanate, reforming the initial reactant that was used for modification. The peptide ends up with a terminal amine, regenerating the original peptide structure. As the original structure is reformed, the fragmentation matches the unmodified peptide, as illustrated in Fig. 5.2d.



Scheme 5.2. Proposed mechanisms for the fragmentation pathways of the partial tag and full tag losses observed in the RDD and CID spectra of IPT modified peptides.

Energy minimization and transition state calculations on a model structure (methyl-thiourea-methyl) were carried out using the B3LYP/6-31+G(d) level of theory. Transition states were confirmed by the presence of a single imaginary frequency that corresponded to transfer of a hydrogen atom, as shown in Fig. 5.3a. An analogous structure was also optimized for urea, which yielded a similar transition state (see Fig. 5.S2). The relative energetics for reactants/transition state/products are shown in Fig. 5.3b. While the transition states for urea and thiourea are structurally quite similar, the change in energy for the urea pathway is lower at +103 kJ/mol above the reactant compared to +181 kJ/mol for thiourea. The final products for each are analogous as well, and the calculations suggest that both pathways are somewhat endothermic.

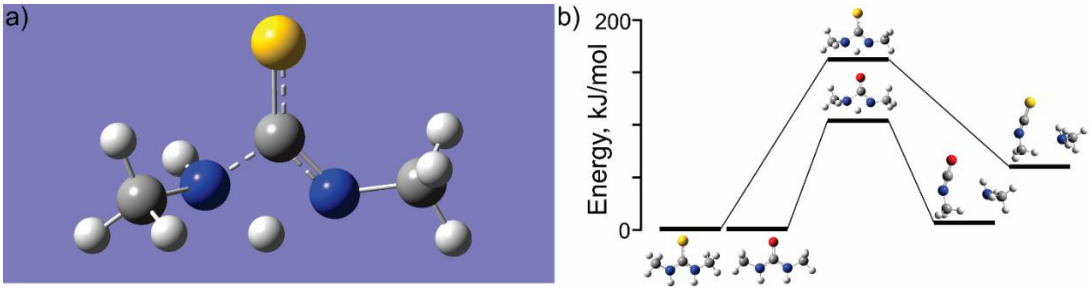


Figure 5.3. a) Transition state calculated using Gaussian with the B3LYP/6-31+G(d) level of theory for the proposed mechanism of partial and full tag losses involving the thiourea linker. A hydrogen is transferred between nitrogen atoms, leading to dissociation of one carbon-nitrogen bond. b) Energy diagram for the overall dissociation reaction of the thiourea linker and analogous urea linker.

A comparison of RDD performed with the 4IPT tag versus the previously used 4IB tag and iodo-tyrosine modifications is shown in Fig. 5.4. In all spectra the common RDD fragments of a-, x-, and z-ions as well as abundant side-chain losses are observed. Additionally, the c_9- and z-1 ions are observed in all cases which are produced by a unique radical mechanism at threonine residues.[15] Some variation in fragment intensity is

observed between the tags, most noticeably in the -87R^+ and -106Y peaks. Apparent intensity differences are evident in the a_9 and x_{10} fragments as well, which are both reduced in intensity in the 4IPT spectrum. The iodo-tyrosine spectrum is the only one which did not contain the x_{10} ion or the b_6 ion. With the iodo-tyrosine radical precursor, the radical starts in a remote location from the N-terminus, and this change is similarly reflected in the fragmentation. These examples illustrate the structural sensitivity of RDD, where differences in the precursor composition, radical starting location, or gas phase structure can lead to significant differences in the resulting spectra. The most notable difference between the spectra is found in the presence of the partial tag loss which occurs as the most abundant peak in the 4IPT tag spectrum. In the 4IB spectrum where the tag is attached via an amide group, no analogous loss of the tag is observed. This difference may provide further evidence that the mechanism of tag loss is occurring through hydrogen transfer between the nitrogen atoms of the thiourea group, which cannot occur in an amide group that contains only one nitrogen atom.

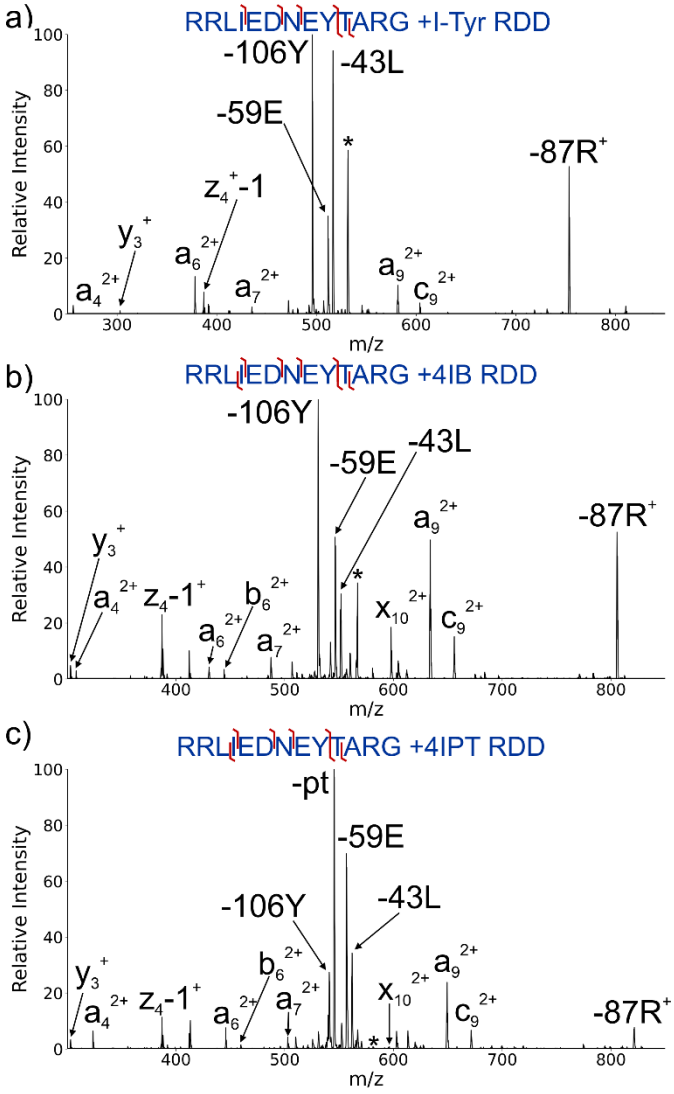


Figure 5.4. a) RDD spectrum of RRLIEDNEYTARG utilizing an iodo-Tyrosine modification compared to b) RDD performed with the 4IB tag and c) RDD performed with the 4IPT tag. The same major fragments are observed in both spectra with some variations in intensity, with the exception of the b_6 and x_{10} fragments which are not observed in a). The 4IPT spectrum in c) additionally contains a partial tag (-pt) loss that is not observed with 4IB.

One of the most advantageous applications of RDD is structural characterization, due to the sensitivity of radical migration to three-dimensional structure. Comparison of RDD spectra for peptide pairs where a single amino acid has been isomerized is shown in Fig.

5.5. For GISEVRSDG (with either L-Asp or L-isoAsp), the resulting RDD spectra vary most in the abundances of the a8 and -CO_2 losses. To quantitatively assess differences between the spectra, an R_{isomer} score is calculated from the two fragments that change intensity the most between spectra. The ratio of these two fragments in each spectrum are taken and divided to yield a ratio of ratios.[38] A significance threshold R_{isomer} score of 2.4 has been established previously with 4IB.[39] For Fig. 5.5a, an R_{isomer} score of 6.40 is obtained from the 4IPT RDD spectra for the GISEVRSDG isomers. In Fig. 5b, the sequence HFSPEELK contains the L-Ser and D-Ser epimers, differing in inversion of one chiral bond. Here, the glutamic acid sidechain and partial tag loss fragments change significantly between spectra with a score or 5.16. While these epimers represent smaller modifications to the structure compared to L-isoAsp, they are still able to be discriminated using radical fragmentation. A variety of peptides were tested and the resulting R_{isomer} scores are displayed in Table 1. 4IPT is able to differentiate these isomers of Asp, Ser, His, and Glu, demonstrating the versatility of RDD isomer analysis. Additional comparisons to 4IB scores are shown for two sequences at the bottom. For these sequences, 4IPT reaches a higher score, but both tags are able to discriminate between the isomer modifications.

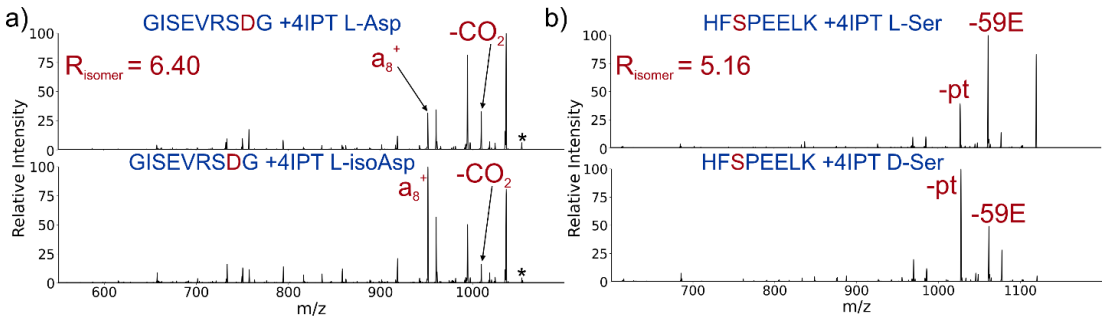


Figure 5.5. a) RDD examination of GISEVRSDG between L-Asp form (top) and L-isoAsp form (bottom). A score of 6.40 exceeds the threshold of 2.40 to confidently discriminate between isomer forms. b) RDD examination of WFDTGK epimers L-Asp (top) and D-Asp (bottom).

Sequence	Isomer 1	Isomer 2	R_{isomer} Score
GISEVRS**D**G	L-Asp8	L-isoAsp8	6.40
WF**D**TGK	L-Asp3	D-Asp3	3.96
DVEFRH**D**SG	L-Asp7	L-isoAsp7	6.16
DAWSNSR**D**SDSLLSL	L-Asp8	L-isoAsp8	4.56
TVLDSGIS**E**VR	L-Glu9	D-isoGlu9	5.15
HF**S**PEELK	L-Ser3	D-Ser3	5.00
IQTGL**D**ATHAER	L-Asp6	D-Asp6	3.30
HFSPEDLTVK	L-His1	D-His1	5.16
IQTGL**D**ATHAER (4IB)	L-Asp6	D-Asp6	2.67
HFSPEDLTVK (4IB)	L-His1	D-His1	3.25

Table 5.1. R_{isomer} scores of tested peptides containing amino acid isomers indicating the difference between fragmentation spectra. Raw data is included in the supporting information.

Conclusions

We have characterized the modification of peptides with a new radical precursor using a commercially available iodophenylisothiocyanate that easily forms covalent bonds with amines. This modification was found to be well-suited for RDD experiments, producing the typical fragments expected of a radical peptide while maintaining sufficient structural sensitivity to allow for isomer identification. The tag is partially labile and can be lost with sufficient vibrational excitation, producing unique mass losses that could be used to confirm the identity of modified peptides in complex samples. Although isocyanate-reactive reagents have been widely employed in solution-phase studies of proteins, they have found less use in mass spectrometry or proteomics. The ease of modification and stability of the resulting products in both solution and the gas phase suggest that isocyanate coupling may be an excellent candidate for other applications requiring covalent modification of peptides or other biomolecules for mass-spectrometry based experiments.

Acknowledgements

The authors gratefully acknowledge funding from the National Science Foundation (1904577).

Supporting Information

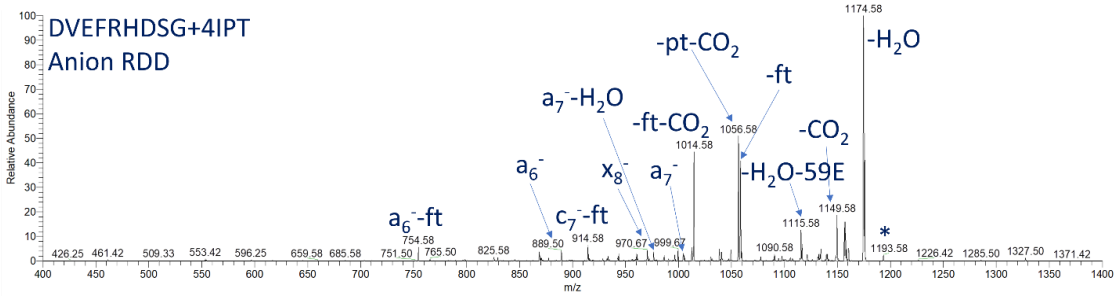


Figure 5.S1. Anionic RDD spectrum for the peptide DVEFRHDSG + 4IPT. Typical radical fragmentation is produced as RDD is not a proton-driven mechanism.

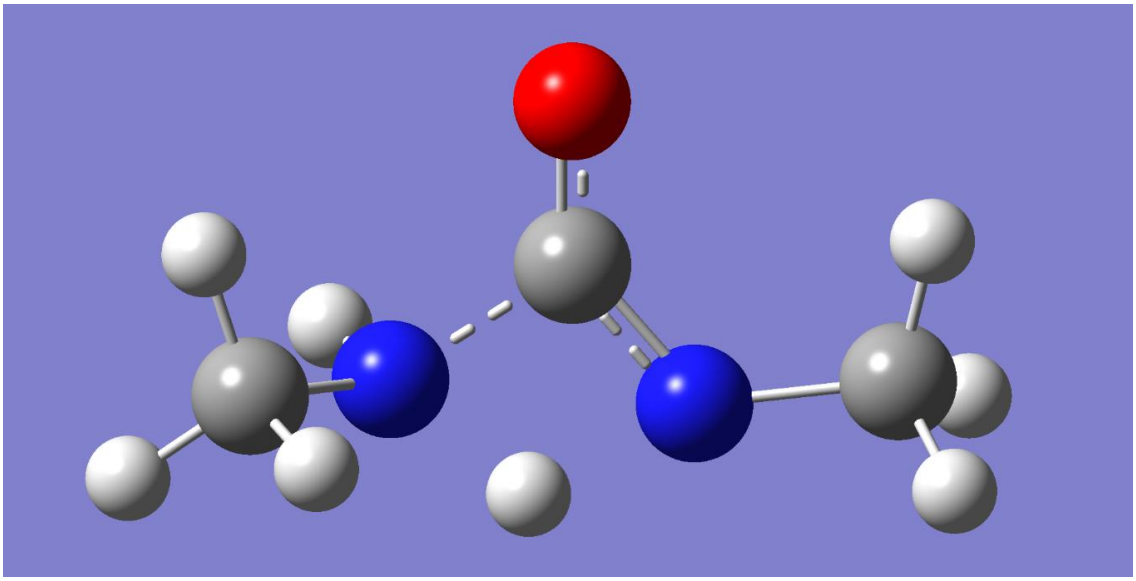


Figure 5.S2. Transition state calculated using Gaussian with the B3LYP/6-31+G(d) level of theory for the proposed mechanism of partial and full tag losses involving the urea linker

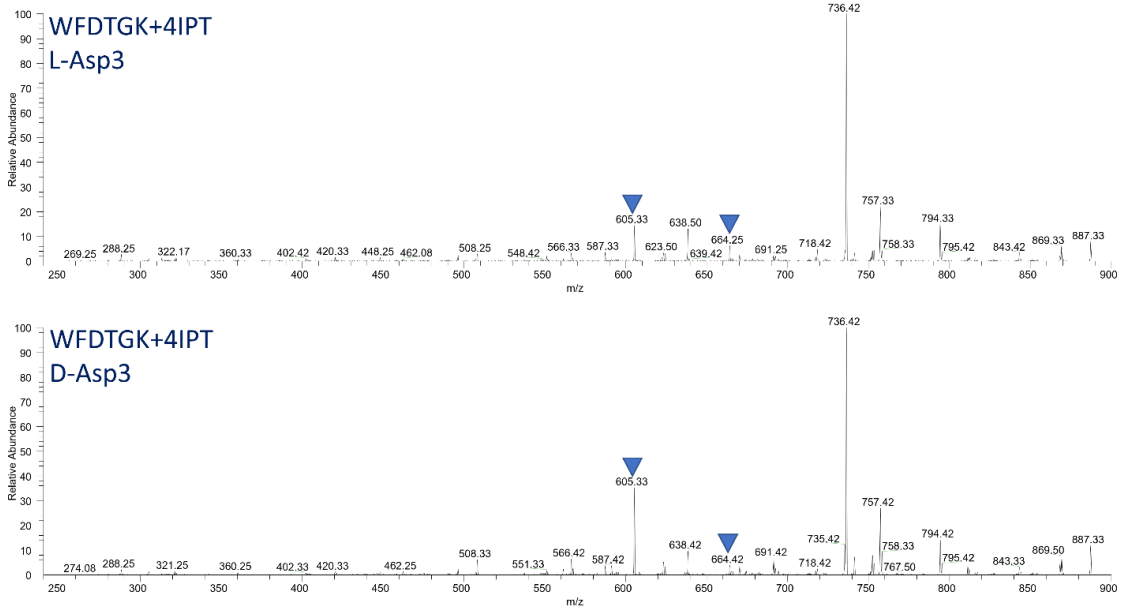


Figure 5.S3. RDD comparison of WFDTGK L-Asp (Top) and D-Asp (Bottom). R_{isomer} = 3.96

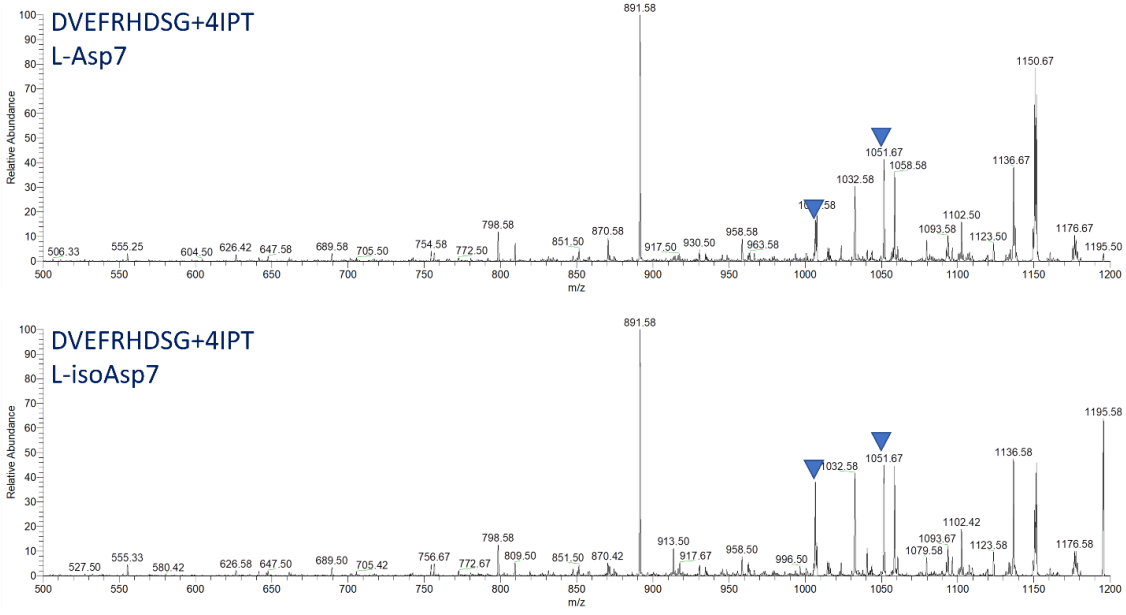


Figure S4. RDD comparison of DVEFRHDSG+4IPT L-Asp (Top) and L-isoAsp (Bottom). R_{isomer} = 6.16

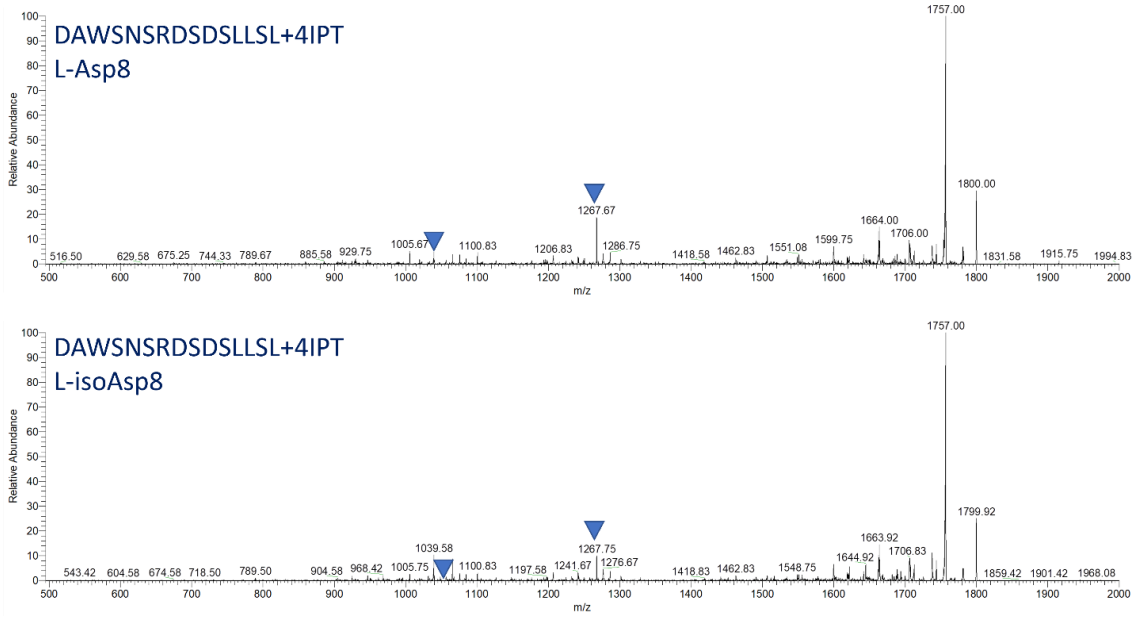


Figure 5.S5. RDD comparison of DAWSNSRDSDSLLSL+4IPT L-Asp (Top) and L-isoAsp (Bottom). R_{isomer} = 4.56

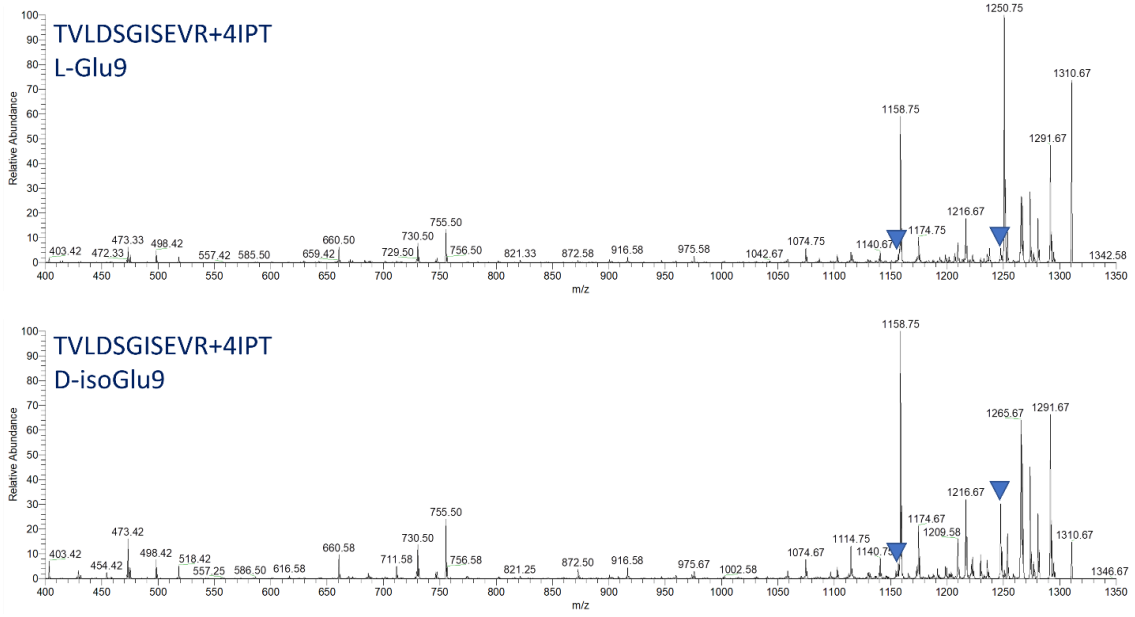


Figure 5.S6. RDD comparison of TVLDSGISEVR+4IPT L-Glu (Top) and D-isoGlu (Bottom). R_{isomer} = 5.15

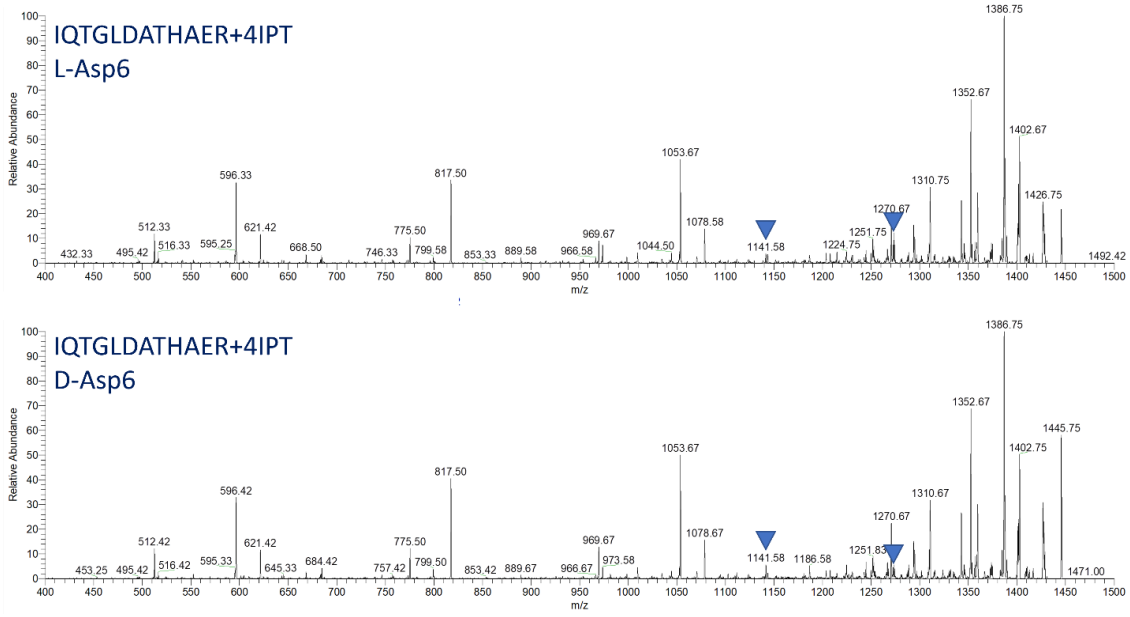


Figure 5.S7. RDD comparison of IQTGLDATHAER+4IPT L-Asp (Top) and D-Asp (Bottom). R_{isomer} = 3.30

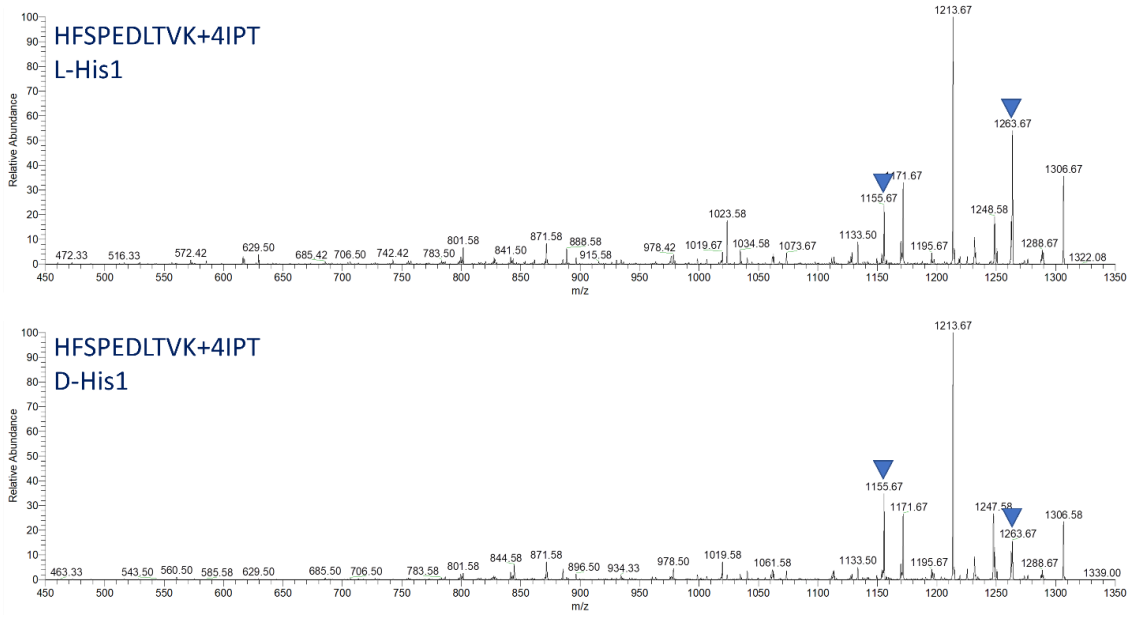


Figure 5.S8. RDD comparison of HFSPEDLTVK+4IPT L-His (Top) and D-His (Bottom). R_{isomer} = 5.16

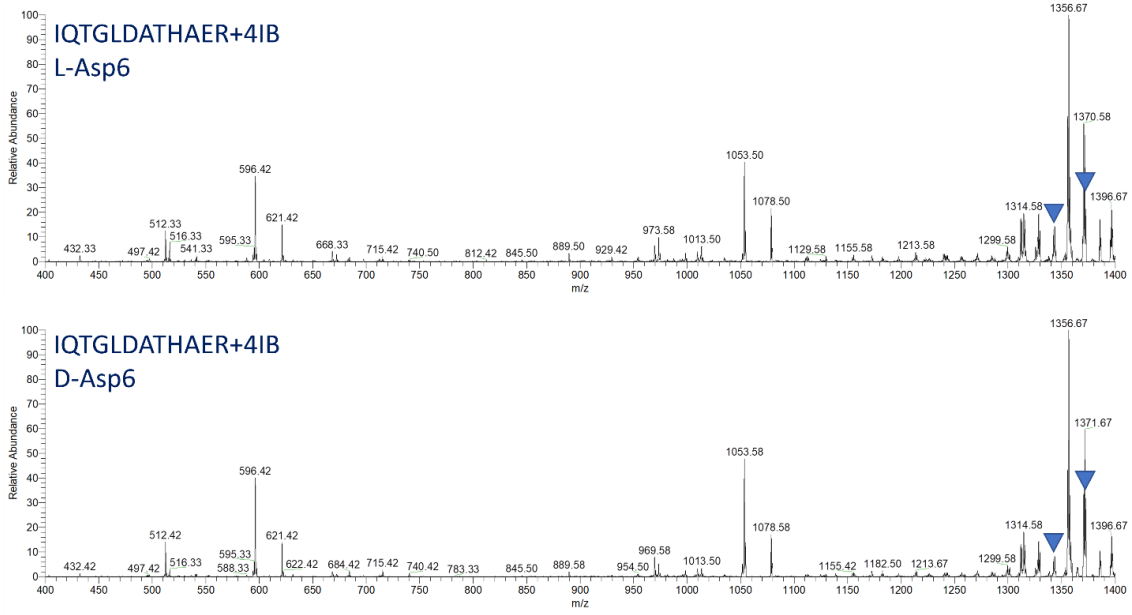


Figure 5.S9. RDD comparison of IQTGLDATHAER+4IB L-Asp (Top) and D-Asp (Bottom). R_{isomer} = 2.67

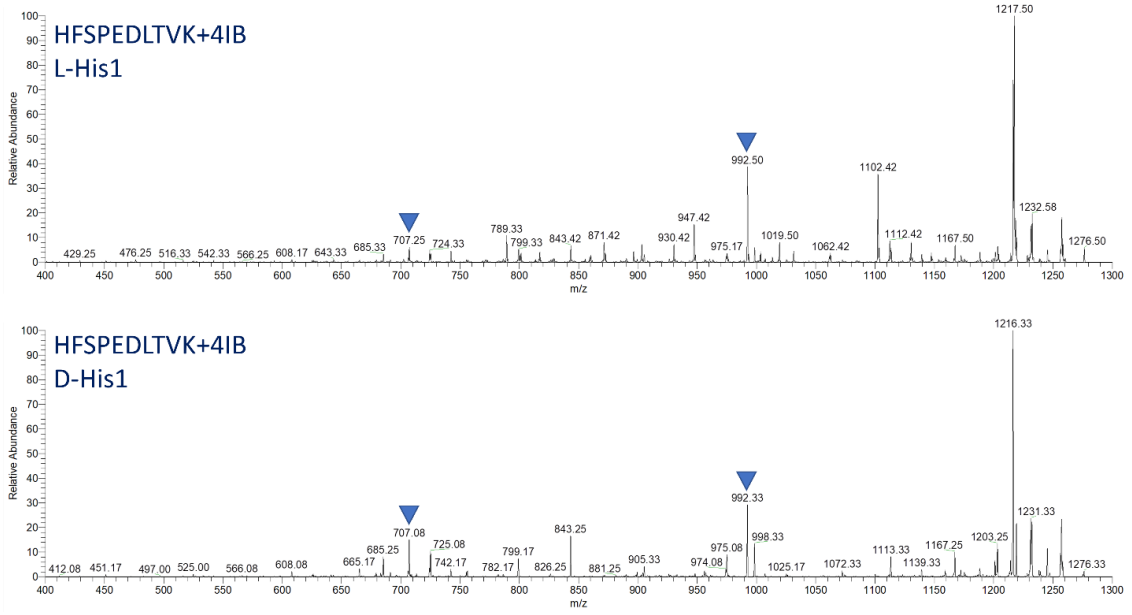


Figure 5.S10. RDD comparison of HFSPEDLTVK+4IB L-His (Top) and D-His (Bottom). R_{isomer} = 3.25

References

[1] Wells, J.M.; McLuckey, S.A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol.* **2005**, *402*, 148-185.

[2] Crowe, M.C.; Brodbelt, J.S. Infrared Multiphoton Dissociation (IRMPD) and Collisionally Activated Dissociation of Peptides in a Quadrupole Ion Trap with Selective IRMPD of Phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1581-1592.

[3] Yeh, G.K.; Sun, Q.; Meneses, C.; Julian, R.R. Rapid peptide fragmentation without electrons, collisions, infrared radiation, or native chromophores. *J. Am. Soc. Mass Spectrom.* **2009**, *20(3)*, 385-393.

[4] Girod, M.; Sanader, Z.; Vojkovic, M.; Antoine, R.; MacAleese, L.; Lemoine, J.; Bonacic-Koutecky, V.; Dugourd, P. UV Photodissociation of proline-containing peptide ions: insights from molecular dynamics. *J. Am. Soc. Mass Spectrom.* **2015**, *26(3)*, 432-443.

[5] Morgan, J.W.; Hettick, J.M.; Russell, D.H. Peptide sequencing by MALDI 193-nm photodissociation TOF MS. *Methods Enzymol.* **2005**, *402*, 186-209.

[6] Moore, B. N.; Ly, T.; Julian, R. R. Radical Conversion and Migration in Electron Capture Dissociation. *J. Am. Chem. Soc.* **2011**, *133* (18), 6997–7006.

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

[8] Tureček, F.; Julian, R.R. Peptide Radicals and Cation Radicals in the Gas Phase. *Chem. Rev.* **2013**, *113*, 6691-6733.

[9] Ly, T; Zhang, X.; Sun, Q.; Moore, B.; Tao, Y.; Julian, R.R. Rapid, quantitative, and site specific synthesis of biomolecular radicals from a simple photocaged precursor. *Chem. Commun.* **2011**, *47*, 2835-2837.

[10] Ly, T.; Julian, R.R. Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. *J. Am. Chem. Soc.* **2008**, *130*, 351-358.

[11] 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.

[12] Lee, M.; Kang, M.; Moon, B.; Oh, H.B. Gas-phase peptide sequencing by TEMPO-mediated radical generation. *Analyst* **2009**, *134*, 1706-1712.

[13] Hao, G.; Gross, S.S. Electrospray Tandem Mass Spectrometry Analysis of S- and N-Nitrosopeptides: Facile Loss of NO and Radical-Induced Fragmentation. *J. Am. Soc. Mass Spectrom*. **2006**, *17*, 1725-1730.

[14] Masterson, D.S.; Yin, H.; Chacon, A.; Hachey, D.L.; Norris, J.L.; Porter, N.A. Lysine Peroxycarbamates: Free Radical-Promoted Peptide Cleavage. *J. Am. Chem. Soc.* **2004**, *126*, 720-721.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Lyon, Y.A.; Sabbah, G.M.; Julian, R.R. Differences in α-Crystallin isomerization reveal the activity of protein isoaspartyl methyltransferase (PIMT) in the nucleus and cortex of human lenses. *Exp. Eye Res.* **2018**, *171*, 131-141.

[20] Planas, O.; Gallavardin, T.; Nonell, S. A novel fluoro-chromogenic click reaction for the labelling of proteins and nanoparticles with near IR theranostic agents. *Chem. Commun.* **2015**, *15*, 5586-5589.

[21] Wilderspin, A.F.; Green, N.M. The Reaction of Fluorescein Isothiocyanate with Thiols: A Method for Assay of Isothiocyanates. *Anal. Biochem.* **1983**, *132*, 449-455.

[22] Nakamura, T.; Abe-Kanoh, N.; Nakamura, Y. Physiological relevance of covalent protein modification by dietary isothiocyanates. *J. Clin. Biochm. Nutr.* **2018**, *62(1)*, 11-19.

[23] Mi, L.; Pasqua, A.J.D.; Chung, F.-L. Proteins as binding targets of isothiocyanates in cancer prevention. *Carcinogenesis* **2011**, *32(10)*, 1405-1413.

[24] Han, K.K.; Belaiche, D.; Moreau, O.; Briand, G. Current Developments in Stepwise Edman Degradation of Peptides and Proteins. *Int. J. Biochem.* **1985**, *17(4)*, 429-445.

[25] George, G.; Geetha, M.; Appukuttan, P.S. Antigen-Induced Activation of Antibody Measured by Fluorescence Enhancement of FITC Label at Fc. *J. Fluoresc.* **2015**, *25*, 1493-1499.

[26] Kim, B.S.; Oh, J.M.; Kim, K.S.; Seo, K.S.; Cho, J.S.; Khang, G.; Lee, H.B.; Park, K.; Kim, M.S. BSA-FITC-loaded microcapsules for in vivo delivery. *Biomaterials* **2009**, 902-909.

[27] Hu, S.; Jiang, H.; Zhu, J.; Wang, J.; Wang, S.; Tang, J.; Zhou, Z.; Liu, S.; Shen, Y. Tumor-specific fluorescence activation of rhodamine isothiocyanate derivatives. *J. Controlled Release* **2021**, *330*, 842-850.

[28] Chen, P.; Nie, S.; Mi, W.; Wang, X.-C.; Liang, S.-P. De novo sequencing of tryptic peptides sulfonated by 4-sulfophenyl isothiocyanate for unambiguous protein identification using post-source decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 191-198.

[29] Gargis, S.R.; Heath, H.E.; Heath, L.S.; LeBlanc, P.A.; Simmonds, R.S.; Abbott, B.D.; Timkovich, R.; Sloan, G.L. Use of 4-Sulfophenyl Isothiocyanate Labeling and Mass Spectrometry to Determine the Site of Action of the Streptococcolytic Peptidoglycan Hydrolase Zoocin A. *Appl. Environ. Microbiol.* **2009**, *75(1)*, 72-77.

[30] Santa, T. Isothiocyanates as derivatization reagents for amines in liquid chromatography/electrospray ionization-tandem mass spectrometry. *Biomed. Chromatogr.* **2010**, *24(9)*, 915-918.

[31] 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(1)*, 97-101.

[32] Shuford, C.M.; Muddiman, D.C. Capitalizing on the hydrophobic bias of electrospray ionization through chemical modification in mass spectrometry-based proteomics. *Expert Rev. Proteomics* **2011**, *8(3)*, 317-323.

[33] Kirk, B.B.; Trevitt, A.J.; Blanksby, S.J.; Tao, Y.; Moore, B.N.; Julian, R.R. Ultraviolet Action Spectroscopy of Iodine Labeled Peptides and Proteins in the Gas Phase. *J. Phys. Chem. A* **2013**, *117*, 1228-1232.

[34] Moore, B.; Sun, Q.; Hsu, J.C.; Lee, A.H.; Yoo, G.C.; Ly, T.; Julian, R.R. Dissociation Chemistry of Hydrogen-Deficient Radical Peptide Anions. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 460-468.

[35] Iacobucci, C.; Piotrowski, C.; Rehkamp, A.; Ihling, C.H.; Sinz, A. The First MS-Cleavable, Photo-Thiol-Reactive Cross-Linker for Protein Structural Studies. *J. Am. Soc. Mass Spectrom.* **2018**, *30*, 139-148.

[36] Falvo, F.; Fiebig, L.; Dreiocker, F.; Wang, R.; Armentrout, P.B.; Schäfer, M. Fragmentation reactions of thiourea- and urea-compounds examined by tandem MS-, energy-resolved CID experiments, and theory. *Int. J. Mass Spectrom.* **2012**, *330-332*, 124-133.

[37] Müller, M.Q.; Dreiocker, F.; Ihling, C.H.; Schäfer, M.; Sinz, A. Cleavable Cross-Linker for Protein Structure Analysis: Reliable Identification of Cross-Linking Products by Tandem MS. *Anal. Chem.* **2010**, *82*, 6958-6968.

[38] 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*, 10598-10609.

[39] 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.

CHAPTER 6: Concluding Remarks

The intricacies of protein structure and the inevitable accumulation of damaging alterations cause a host of complications in biochemical systems. While some modifications are caused by a change in chemical composition, other prevalent sources of age-related damage arise from rearrangement of three-dimensional structure without an accompanying chemical exchange that can be easily detected. While the underlying makeup remains the same, shape is arguably one of the most important features in biochemistry as evidenced by the wide variety of analogous bioactive compounds and further established by the active site configurations of enzymes responsible for cellular function . As protein structure is uniquely susceptible to alterations, understanding structures and their consequences is critically necessary.

Briefly, we have taken advantage of the powerful capabilities of mass spectrometry to separate and characterize the products of biochemical interactions between altered protein structures and lysosomal enzymes. This information allowed us to understand the extent of obstruction in proteolysis caused by changes in primary structure as outlined in chapter two. These tests provided insight into the sensitivity of enzymatic processes both in vitro and in vivo to non-canonical amino acids, and allowed us to make connections between age-related structural modifications and autophagic failure which is observed in Alzheimer's disease.

This understanding was expanded through the study of enzymatic digestion of higher order structure modifications in chapter three. This further expanded our understanding of the effect of structure, revealing how canonical amino acids can still arrange into larger

forms which interfere with normal processes. By utilizing the power of product identification in mass spectrometry, we were able to identify regions of the amyloid beta peptide which proved most problematic in the digestion of higher order fibril structures. The comparison of different product profiles additionally allowed us to make insights into the structural morphologies produced under different starting conditions. Future work looking at the effects of isomer and fibril modifications on cellular systems and function would provide valuable information that draws connections between the autophagic failure and the long-term effects on biological systems. A drawdown of resources normally freed by autophagy would affect other aspects such as protein synthesis, mitosis, and cell viability, making it important to study the effect of structural modifications in a larger context.

Further examination of enzymatic interactions revealed a unique ligase behavior which was highly dependent on underlying substrate structure as outlined in chapter 4. By exploring the structural dependence of this reaction, we revealed the presence of a secondary enzyme activity that operated by reversal of its primary mode of action. A unique interaction between the enzyme and an isomerized residue additionally revealed another effect of substrate isomerization outside of proteolysis. We were able to capitalize on this interaction to expand the capability of the ligase behavior beyond its inherent dipeptide-ligase reaction. As ligation can be used by cellular systems that prepare antigens, future studies into cathepsin B may reveal previously unknown roles that it performs outside of the lysosome. Additionally, the unique recognition mechanism that occurs between cathepsin B and L-isoAsp may be present and undiscovered in other

enzymes which are not typically incubated with isomers in a laboratory setting, but may encounter them in an aging cell.

Finally, we followed up these investigations by developing a tool to enable easy modification and analysis of structural modification for future studies. This work established the efficacy of a new radical precursor based on the isothiocyanate reaction which could modify peptides in high yield. We characterized this tag in radical chemistry fragmentation that is highly sensitive to structure, and comparison of different radical fragmentation spectra allowed us to differentiate peptide isomers. Isothiocyanates are used in the modification of proteins in biology, but are not widely used in mass spectrometry. This new tag enables easy and rapid modification of proteins with a radical precursor, which reduces barriers for use of RDD in proteomics and structural experiments where sensitivity to modifications is important.

Ultimately, we have utilized mass spectrometry as a powerful tool to identify and characterize protein structure. Our studies make important connections between structural modifications that occur with age and the biochemical processes that occur in the lysosome. The connections made between structural modifications and lysosomal enzymatic behavior provide a meaningful foundation from which to build new ideas in the sphere of aging research.