# ABSTRACT

| | |
|---|---|
| Title of Dissertation: | CHARACTERIZATION OF POLYUBIQUITIN CHAINS BY MASS SPECTROMETRY |
| | Amanda Elizabeth Lee, Doctor of Philosophy, 2016 |
| Dissertation directed by: | Professor Catherine Fenselau, Department of Chemistry and Biochemistry |

The work outlined in this dissertation will allow biochemists and cellular biologists to characterize polyubiquitin chains involved in their cellular environment by following a facile mass spectrometric based workflow. The characterization of polyubiquitin chains has been of interest since their discovery in 1984. The profound effects of ubiquitination on the movement and processing of cellular proteins depend exclusively on the structures of mono and polyubiquitin modifications anchored or unanchored on the protein within the cellular environment. However, structure-function studies have been hindered by the difficulty in identifying complex chain structures due to limited instrument capabilities of the past.

Genetic mutations or reiterative immunoprecipitations have been used previously to characterize the polyubiquitin chains, but their tedium makes it difficult to study a broad ubiquitinome. Top-down and middle-out mass spectral based

proteomic studies have been reported for polyubiquitin and have had success in characterizing parts of the chain, but no method to date has been successful at differentiating all theoretical ubiquitin chain isomers (ubiquitin chain lengths from dimer to tetramer alone have 1340 possible isomers). The workflow presented here can identify chain length, topology and linkages present using a chromatographic-time-scale compatible, LC-MS/MS based workflow.

To accomplish this feat, the strategy had to exploit the most recent advances in top-down mass spectrometry. This included the most advanced electron transfer dissociation (ETD) activation and sensitivity for large masses from the orbitrap Fusion Lumos. The spectral interpretation had to be done manually with the aid of a graphical interface to assign mass shifts because of a lack of software capable to interpret fragmentation across isopeptide linkages. However, the method outlined can be applied to any mass spectral based system granted it results in extensive fragmentation across the polyubiquitin chain; making this method adaptable to future advances in the field.

# CHARACTERIZATION OF POLYUBIQUITIN CHAINS BY MASS SPECTROMETRY

by

Amanda Elizabeth Lee

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Catherine Fenselau, Chair
Professor Alan Jay Kaufman, Dean's Representative
Professor David Fushman
Professor Neil Blough
Professor John Ondov

# Dedication

This work is dedicated to my parents, Carol and Buzz Lee. All good things I have ever accomplished are a direct results of their parenting. Thank you for being the strongest people I have ever known and for building a better life for your children.

# Acknowledgements

I would first like to recognize my advisor, Catherine Fenselau, for believing in me to take on this project and for giving me the tools necessary to tackle any difficulty I faced on it. Along with the chief of the lab, I must thank all the past and present members of the lab who have given me direction (personally and professionally): Dr. Joe Cannon (for pioneering the project), Dr. Meghan Burke (for mentorship), Yeji Kim (for defining hard work), Lucía Geis (for her brainz), and Sitara Chauhan and Kate Adams (for keeping me "sane"!).

This project would not have been successful without support from Dr. David Fushman. There are many members of the Fushman Lab that provided sample I used in my work and advice on handling ubiquitin including Carlos Castañeda, Emma Dixon, Meredith Miller, Tanuja Kashyap, and Dulith Abeykoon.

I would also like to acknowledge the guidance I received from Dr. Yan Wang on how to properly run the instruments and for all her constructive advice on how to present and think about my project.

I would also like to thank the Wolf Pack (Kim Huynh, Steve Wolf, Marcus Carter, Shweta Ganapati, Tessy Thomas, and Sitara Chauhan (double credit!)), my comrades-in-doctorate, for commiserating with me and helping me, not just get through this process, but enjoy my time here. Also to my editor-in-chiefs Mary Ivory and Dipavo Banerjee, thank you for your advice, patience, and time!

Finally I would like to thank my family for all their support and encouragement; especially my parents, Carol and Buzz, my sisters, Jennifer and Christine, and my brothers-in-law, Michele and Pierre.

**Table of Contents**

# List of Tables

# List of Figures

# List of Abbreviations

**ACN**: Acetonitrile

**AGC**: Automated Gain Control

**CID**: Collision induced dissociation

**Da:** Dalton

**DNA:** Deoxyribonucleic acid

**ESI:** Electrospray ionization

**ETD:** Electron transfer dissociation

**ETciD:** Electron transfer dissociation with supplemental activation using CID

**EThcD:** Electron transfer dissociation with supplemental activation using HCD

**FA:** Formic acid

**FT:** Fourier transform

**FT-ICR:** Fourier transform ion cyclotron resonance

**-GG:** Glycinylglycine

**HCD:** Higher-energy collisional dissociation

**HPLC:** High pressure liquid chromatography

**IP:** Immunoprecipitation

**IRM:** Ion routing multipole

**kDa:** Kilodalton

**LC:** Liquid chromatography

**LC-MS:** Liquid chromatography mass spectrometry

**LC-MS/MS:** Liquid chromatography tandem mass spectrometry

**LTQ:** Linear trap quadrupole

**MALDI:** Matrix assisted laser desorption ionization

**MDSC:** Myeloid derived suppressor cell

**MS:** Mass spectrometry

**MWAC:** Microwave assisted acid cleavage

**MWCO:** Molecular weight cutoff

**NCE:** Nominal collisional energy

**NMR:** Nuclear magnetic resonance

**PPM:** Part per million

**PTM:** Post translational modification

**RF:** Radio frequency

**RP-HPLC:** Reversed phase high pressure liquid chromatography

**SILAC:** Stable Isotope Labeling with Amino acids in Cell culture

**TOF:** Time of flight

**Ub**: Ubiquitin

**UbiCRest:** Ubiquitin chain restriction

**UHPLC:** Ultra high pressure liquid chromatography

**UbL:** Ubiquitin-Like

# Chapter 1: Introduction

*Ubiquitin Chains*

Ubiquitin (Ub) is a 76 amino acid protein that is found ubiquitously in eukaryotic cells, hence its name. There are also many Ub-like (UbL) proteins in a range of eukaryotic cell types including Rub1 (AKA Nedd8 in mammals)[1] and SUMO,[2] which have complimentary, but different sequences and functions compared to Ub.[3] Prokaryotes are also known to carry a UbL, called ThiS, which has been shown to have similar functions to Ub.[4] The fact that Ub and its sequence are so highly conserved suggests its importance for cellular proliferation.[5]

The alteration of a protein after it has been translated is called a post translational modification (PTM). Ub can be attached to proteins, including itself, as a PTM, through the formation of an isopeptide bond between its carboxyl terminus (G76) and the ε-amine of a lysine (K) residue or the N-terminus of the initial methionine (M1) in a process called ubiquitination.[6] (Figure 1.1)

$M_1$QIFV$K_6$TLTG$K_{11}$TITLEVEPSDTIENV$K_{27}$A$K_{29}$IQD$_{32}$$K_{33}$EGIPPDQQRLIFAG$K_{48}$Q
LEDGRTLSDYNIQ$K_{63}$ESTLHLVLRLRGG

Figure 1.1. Human ubiquitin sequence, which contains 76 amino acid residues. All potential isopeptide linkage sites (lysines) and the initial methionine are highlighted in cyan.

Ubiquitination can occur in different ways: a single Ub can be attached at a single site (monoubiquitination) (Figure 1.2a) or multiple sites (multiubiquitination), (Figure 1.2b) or a chain of internally linked Ub can be anchored to a protein (polyubiquitination.) (Figure 1.2d) Various polyUbs are also present in the cell unanchored (Figure 1.2c). The possible linkage sites for intra Ub chain formation are M1, K6, K11, K27, K29, K33, K48, and K63 (highlighted in cyan in Figure 1.1). PolyUb is difficult to characterize because a single Ub chain can be made of all one linkage type (homotypic) (Figure 1.2c) or different linkages (mixed) (Figure 1.2d)[7] It is even possible -and seen in vivo- to have a branched polyUb;[8] a Ub with multiple isopeptide linked Ubs attached (Figure 1.2e). Thus the potential intracellular complexity of these polyUb chains makes proteomic analysis immensely challenging.

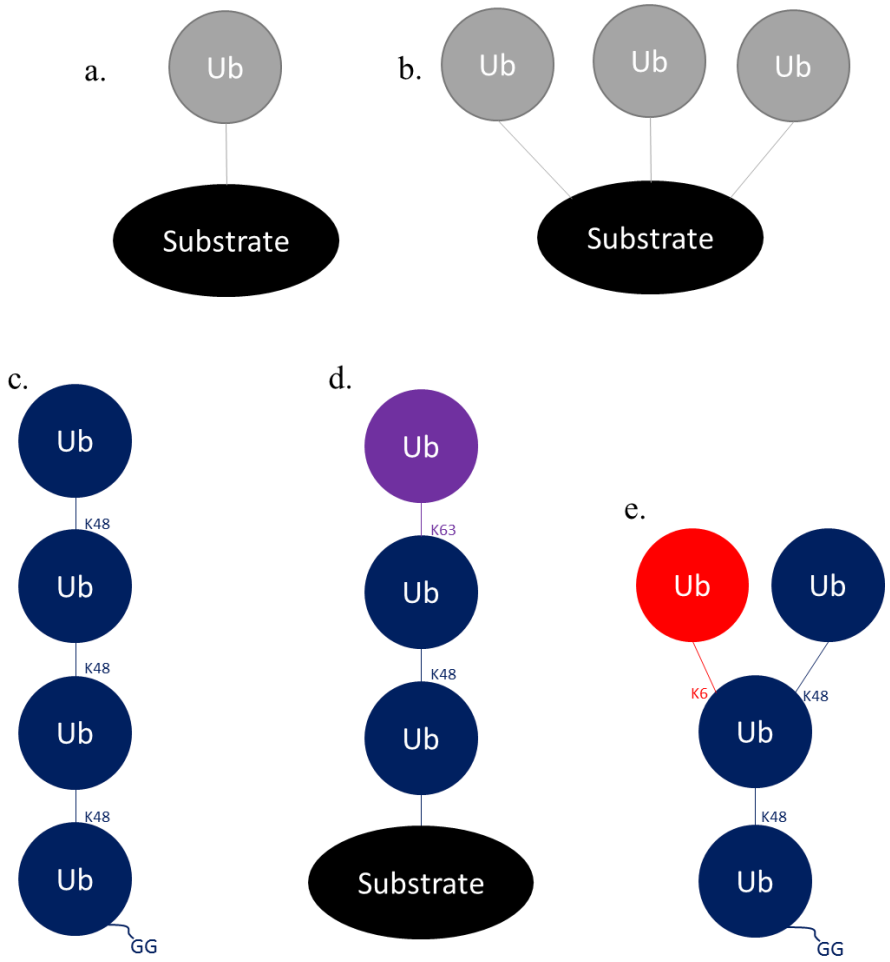



Figure 1.2. The many types of ubiquitination laid out in ball-and-stick models. Monoubiquitination (single: a. and multi: b.) and polyubiquitination (homotypic: c., heterotypic: d., and branched: e.) are shown anchored (a., b., and c.) and unanchored (c. and e.) to a substrate protein.

Anchored and unanchored chains have been known to play a major role in DNA repair,[9] protein degradation,[5,10] cancer morphology,[11] protein kinase activity,[12] redox regulation,[13] and aggresome degradation.[14] This myriad of cellular responses is thought to be directed by different structures of Ub polymers depending on chain

shape and linkages present (Figure 1.3)[15] and yet, characterization and quantification of the ubiquitinome has been limited.[7,16] Current characterization techniques do not allow for the full identification of the different possible structures and linkages of polyUb, thus limiting the information cellular biologists can obtain about the function of Ub in the cell.[17]

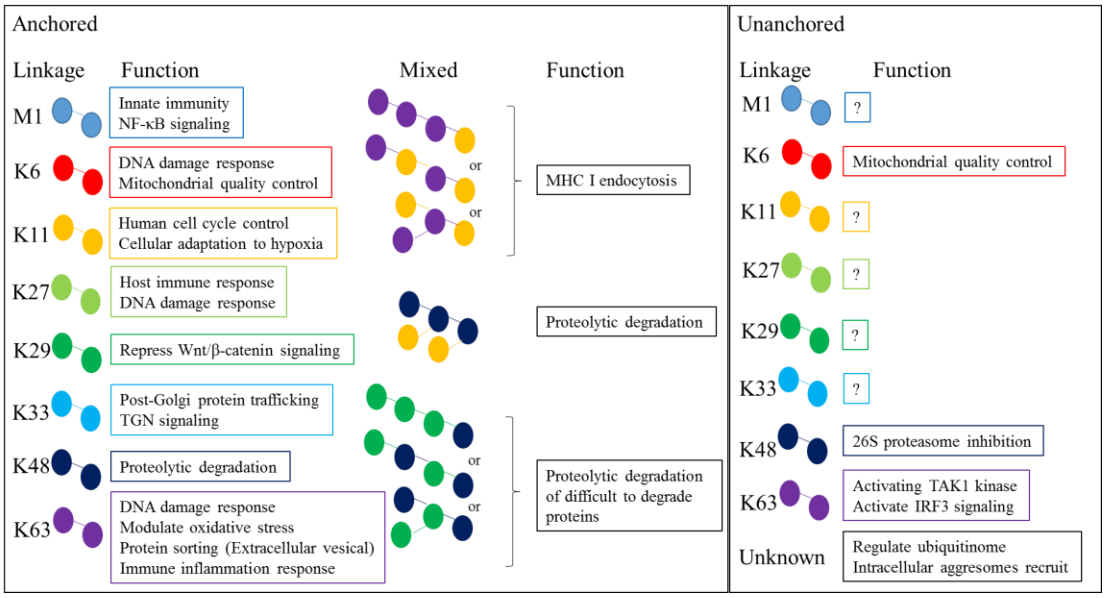



Figure 1.3. Outline of the different linkages and shapes of polyUb and their known functions. (Adapted and updated from reference 15)

The scientific community has developed a few methods for identification and characterization of Ub linkages and, very generally, shape. One method which does not use mass spectrometry directly is the use of antibodies. Antibodies are often used

to isolate, and/or confirm the presence of, specific protein substrates from samples. There are many anti-Ub antibodies, and some have linkage specificity. However immunoprecipitation does not always reveal the number of Ub moieties (unless coupled with intact mass analysis) or complete linkage other than the epitope. If there is a mixed linkage, for example, then the epitope of one of the linkages will match the antibody, but the other linkages in the polyUb maybe unaccounted for in the antibody based analysis alone.[18] There is not yet an antibody for every linkage type either, so some linkages are impossible to identify by this method. It has also been shown that not all antibodies isolate the same cohort, which would introduce unexpected variables.[19] There are antibodies which can isolate unanchored Ubs[20] and some that isolate polyUbs,[3] but more specific features (such as branching) cannot be elucidated.

Similar to antibody work, proteins called deubiquitinases (DUBs) have been used to reveal the structure of polyUbs without using mass spectrometry. DUBs remove ubiquitin isopeptide bonds either generally or at specific linkages. A process called UbiCRest (ubiquitin chain restriction)[21] has been shown to effectively identify linkages present in heterotypic chains. Some topological determination has been accomplished using DUBs.[22] These methods are extremely useful and have been implemented successfully *in vitro* experiments.[23] However, the DUBs currently available do not cover all the possible Ub linkages, and the DUBs that have specificity have not been tested to confirm their specificity in all chain lengths and topologies.[22] This is an uncertainty that fragmentation resulting from tandem mass spectrometry could overcome since the linkage present would produce unique fragments matching that linkage only.

Many of the studies shown in Figure 1.3 have elucidated the linkages present and shown their importance in cellular functions using sequence mutations.[24] By converting specific K residues to R, the general shape of the Ub can be maintained, but the conjugating abilities are completely lost. This means, for example, that if K63 linkages are suspected to be involved in a process, researchers can genetically mutate the cells to produce K63R mutant Ubs and see if the functions being studied are perturbed.[13,25] In this way, scientists have definitively defined functions without any uncertainty. However, these studies must disrupt the normal Ub pathway, which may have unpredicted consequences.[26]

Two specific studies that attempted K to R mutations for topological determination came to the conclusion that they were unable to determine the exact topology of the polyUb chain using mutations. The authors suggest a chain shape and determined the linkages involved, but no more was definitively concluded.[8,27] Another K to R mutation experiment successfully demonstrated that K11 branched off K48 chains *in vivo* and that these chains allowed for a more efficient proteolysis.[28] However, all these processes involved multiple experiments where separate mutations were needed. These mutations not only involve more invasive biochemical techniques *in vivo*, but also require a lot of time in cell culture and genetic mutation experiments. To map the entire ubiquitinome would be a gargantuan biochemical undertaking using any one of the methods mentioned above.

*Mass Spectrometric Methods for Ubiquitin Determination*

Bottom-Up Proteomics

Cellular systems have many different enzymes that cleave peptide bonds on proteins. These enzymes have been isolated and are now commercially available for general use. Trypsin is one of the most common cleavage enzymes; it cleaves specifically at arginine (R) and lysine (K) residues making it possible to predict the product peptides from large protein repositories. When these peptides are coupled with a LC-MS/MS analysis, the method is called bottom-up proteomics. The bottom-up refers to the fact that the peptides must be pieced back together after identification by MS/MS and matched with a protein that has the same potential peptide products and sequence identified by MS/MS.[29]

Many laboratories use trypsin cleavage to locate ubiquitination on proteins because it leaves a -GG tag on the ε-amine of the target protein's lysine after trypsin cleaves at R74 of Ub. When peptides from the conjugated protein are sequenced by tandem mass spectrometry (MS/MS) and bioinformatics, the additional mass from the -GG tag, 114.043 Da, can be recognized and located.[30,31,32,33] However, this technique shows neither the length nor the linkage pattern of the ubiquitination side chain, but can show all the linkages present (as Ub peptides with -GG tags) and the location of the ubiquitination on the substrate protein.[34] (Figure 1.4)


Figure 1.4. Visual representation of an ubiquitinated protein undergoing tryptic digestion. Cleavage at R74 of the Ub leaves a signature mass addition of 114.043 Da onto the modified lysine (K*) due to the -GG remnant. (Adapted from ref 31)

Many ubiquitinomal features have been interrogated using bottom-up -GG tagging. This method allows for a facile shot-gun proteomic study to reveal Ub levels in the cell and to quickly identify the sites. This method allows for quantitative measurements which can compare the amount of ubiquitination compared to other PTMs.[35,36] For example, through -GG tagging, it was found that total Ub is 486.4 pmol/mg (or 0.42%) total protein of the cellular cargo in HEK293 cells.[36] It was also determined that monoUb was the most abundant form of around 60%,[37] but that the cellular levels of mono and polyUb change depending on the system. PolyUb reached at most 14% of the total Ub in the MEF cell line.[37,36]

Bottom-up proteomics has also allowed for quantification of the linkages present in the cell. By using SILAC (Stable Isotope Labeling with Amino acids in Cell culture) with -GG tagged peptides the relative abundance of the linkages were found to be K11 (35%), K48 (30%), K63 (11%), K6 (11%), K27 (7%), K29 (4%), K33 (1%), and M1 (1%).[38] Knowing the linkages present is important to show that

the complexity of the Ub pool cannot be reduced to a few highly abundant linkages. A method is needed that can correctly assign all linkages present. Bottom-up proteomics has been able to find all the Lys linkages, but it is limited in knowing their relationship to each other and the chain topology, which has been shown to have different effects in the cell.[8,27,28]

## Top-Down Proteomics

An alternative approach is to identify proteins directly from the cell without digesting to peptides, and is thus called top-down proteomics. Top-down proteomics can be very beneficial for the identification of a multitude of PTMs because the whole protein is kept together, and consequently, the potential to map co-PTMs (multiple PTMS on the same protein) becomes much more probable. For example, a bottom-up experiment could find a peptide with an acetylation and then a different peptide of the same protein with a phosphorylation site. How would a bottom-up user tell if the acetylation and the phosphorylation were happening together on the protein or if the protein had multiple proteoforms? It is also possible to lose modifications in bottom-up because they are a smaller sub-set of a more common PTM on the same peptide. Common data dependent tandem mass spectrometric experiments may not collect data on these lower abundant peptides.[39,40] With sensitive top-down proteomics, proteoforms and isoforms can be completely characterized.

Top-down is limited, however, by the instrumentation's fragmentation ability and mass analyzer's sensitivity/resolving power. There must be a high enough

resolution to separate $^{13}C$ isotopes, as seen in Figure 1.5 for deconvolution and charge state definition. Also the larger the peptides or proteins being subjected to fragmentation, the more inefficient the fragmentation tends to be. This is because, in collisionally induced dissociation, (CID and HCD) the energy can be dispersed across bonds and rotations and the larger the precursor molecule, the more dispersion possible.[41] In electron capture/transfer dissociation (ECD/ETD), the reaction is generally more efficient for precursor ions with higher charge states, which, in electrospray ionization, means a larger intact mass.[42,43] Very large proteins are still limited in ECD/ETD fragmentation when the electron transferred only results in charge state reduction. Supplemental collisional energy has been successful in enabling more efficient ETD reactions and will be implemented in this work.[44]

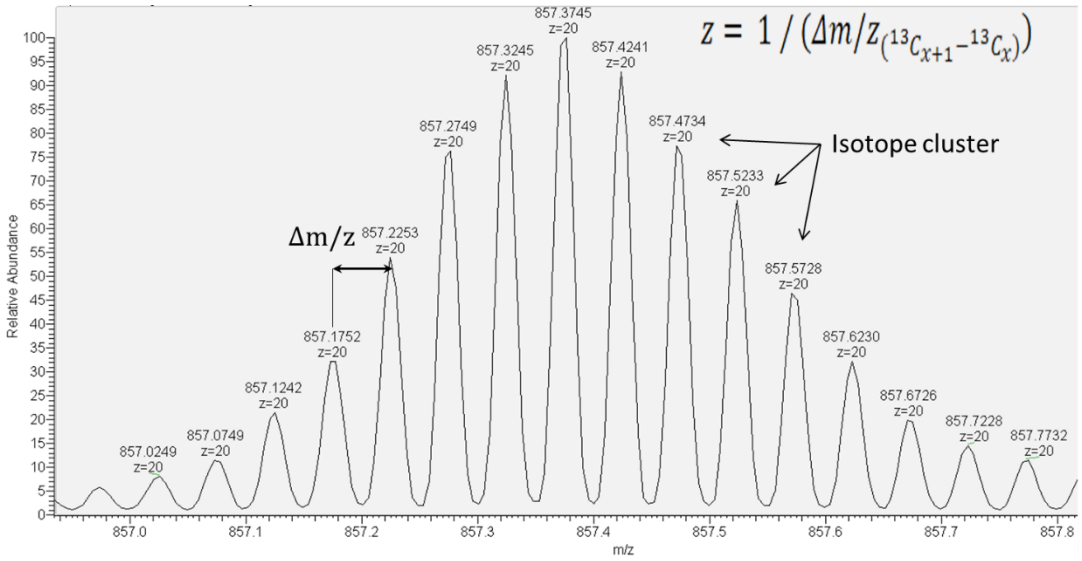



Figure 1.5. Example spectrum of an intact Ubiquitin dimer linked at K63. The isotope cluster represented above is for the charge state 20. The inverse of the difference between two adjacent m/z isotope peaks ($^{13}C_{X+1}$-$^{13}C_X$), shown between the arrows, is the equation used to determine the charge state (z).

Each Ub adds ~8500 Da and the mass of a protein conjugated to a long polyUb will quickly exceed the current upper mass range of many top-down tandem mass spectrometers used in a chromatographic time-scale. Also each addition is of the same chemistry and thus makes polyUb of different lengths difficult to separate. The linkage location changes the shape of some Ub dimers,[45,46] but may not affect their retention in the typical stationary phases used and so they may not separate easily. The LC-MS/MS instrumentation is available to detect larger masses, and is becoming more common in laboratories. This along with improved column chemistries coupled with UHPLC systems can taper these issues.

Middle-Out Proteomics

Instead of using one extreme (bottom-up) or the other (top-down), it is possible to use a chemical reaction that can give the benefits of both. A well-documented chemical reaction that cleaves selectively at aspartate (D) is microwave accelerated acid cleavage (MWAC).[46] D residues are less common than the cleavage sites of tryptic cleavage methods (K and R), so it leaves larger peptides and will not remove common PTMs on the proteins, leading to this method being called middle-out.[47,48,49,50]

Another common method for creating larger middle-out-sized peptides is to run minimal enzymatic cleavage. Trypsin can be limited in its digestion by mutations of the enzyme and/or lowering the reaction time, referred to here as minimal trypsinolysis.[51] Acid hydrolysis can also be limited by the time the reaction is allowed to proceed.[52] Both options will create peptides that retain more information on the proteins' overall structure, due to missed-cleavage events leaving longer sequences (>2000 Da). Both these reactions can create a more telling polypeptide for polyUb chains in particular.[51,49,52,53]

Trypsin cleaves Ub preferentially at the R74 leaving -GG tags on any modified lysine residue.[54] If the time is limited, cleavage at R74 will be one of the few which occurs. This means that any isopeptides that are branched might be seen as multiple -GG tags on a single Ub peptide.[51] (Figure 1.6) The convenience of this middle-out method is that the -GG tags can be searched as variable modifications using well established programs. Limitations of minimal trypsinolysis is that the C-terminal -GG connecting the Ub chain to the substrate protein is lost for every chain

type, thus losing definition of the potential attachment sites of the polyUb chain in a mixture. Also, this creates a problem with more complex, unusual structures of polyUb, which have branched and unbranched sections (as in Figure 1.2e). In brief, if all the R74 are cleaved in the chain, there would be no way to tell on which Ub moiety within a chain the branched and unbranched points were located. (Figure 1.6)

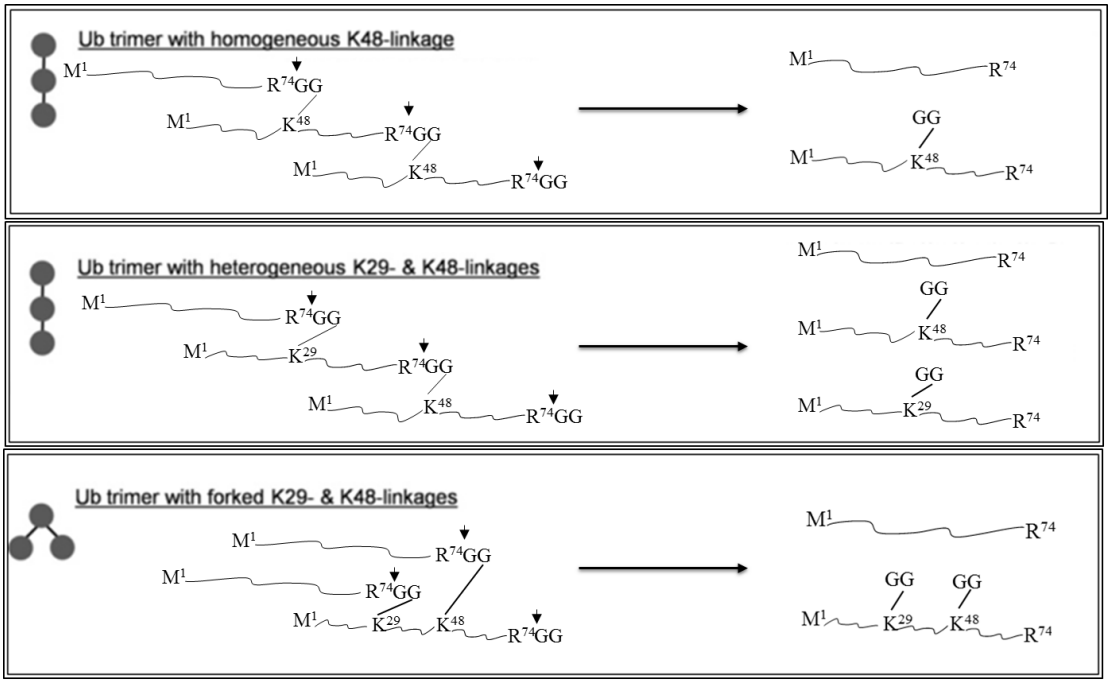



Figure 1.6. Visual representation of minimal tryptic digestion of ubiquitin chains that cleaves preferentially at R74 and can lead to structural classification. (Adapted from ref 33)

Time-limited MWAC on polyUb results in many peptides, which can be used for polyUb chain analysis. (Figure 1.7)[49] The goal of using acid hydrolysis is to retain the C-terminus at all branch points. By doing this, a study of a complex polyUb mixture can prove the Ub chain is unanchored or could show the exact spot on the conjugated protein that the chain is anchored. This conjugation information is lost in trypsin analysis.[7] Another benefit is that certain Ub-like (UbL) proteins (i.e. Nedd8) also produce -GG tags when digested with trypsin, but when using acid hydrolysis, even complete acid hydrolysis, there is no more overlap in the attached isopeptide sequence.[55]

The main limit of acid hydrolysis is that the reaction must be tailored to create incomplete cleavage, producing a large mixture of polypeptides which will lower the signal of the target peptide. A simple example of this is the loss and retention of the D on the termini of the peptides in acid cleavage. Some peptides will retain the D on the N-terminus, on the C-terminus, or none. This will split the signal of the peptide into four different possible masses, complicating not only the separation and detection of the digestion products, but the interpretation of the data collected. (Figure 1.7)

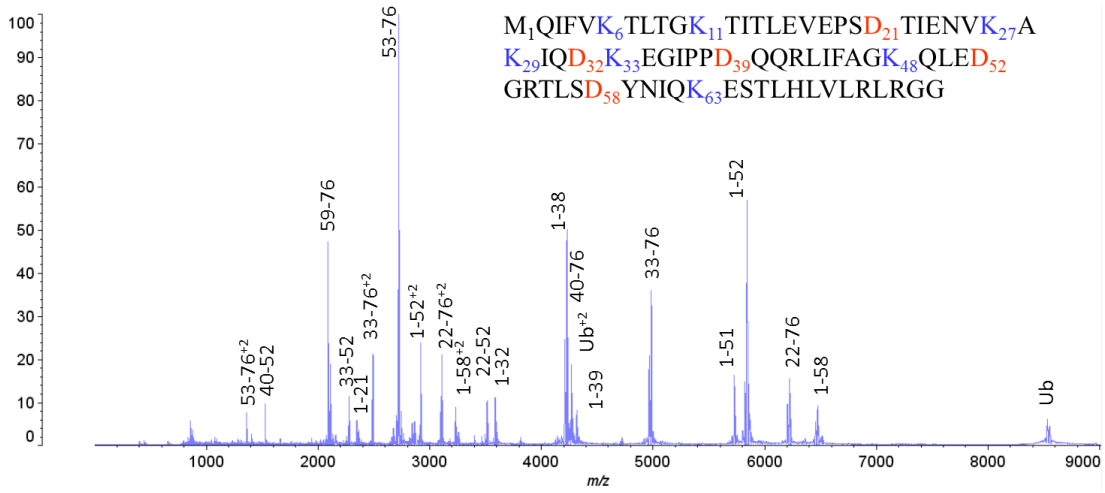



Figure 1.7. Product spectrum of monoubiquitin hydrolyzed with acetic acid at $140^0$C for 60 sec in microwave assisted acid hydrolysis. Peptides are labeled with the amino acid number which corresponds with the aspartic acid residues (D) colored red in the sequence above.

*Fragmentation of Large Peptides/Proteins*

The fragmentation experimentalists seek is a process in which a peptide backbone is broken in a predictable and reliable manner. A precursor MS is acquired and a peptide/protein m/z is selected from the available ions for fragmentation. Once fragmentation on the isolated m/z occurs, the fragment ions are scanned to produce a product ion spectrum. These fragments can be used as a ladder to piece together the amino acid sequence. The most common and useful fragmentation patterns are labeled a, b, c, x, y, and z (Figure 1.8), which form from different fragmentation techniques.

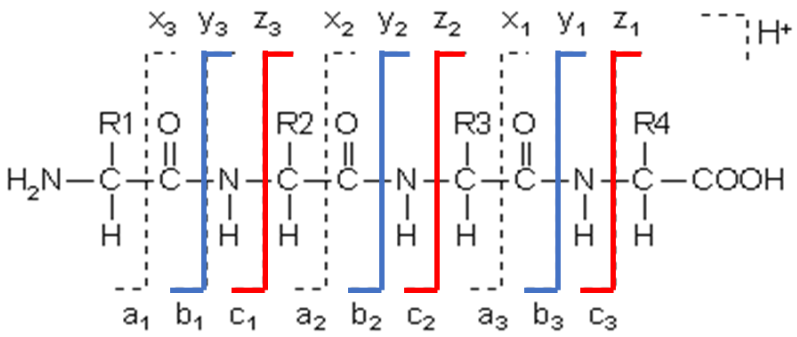



Figure 1.8. All potential fragmentation paths shown on a tetrapeptide. (www.matrixscience.com)

Collisionally Induced Dissociation

The most common technique for inducing fragmentation of peptides/proteins is collisionally induced dissociation (CID). In CID, the precursor ion is moved into the collision cell where it is activated by multiple low-energy collisions with an inert gas (Ne or Ar). The dominant fragmentation pattern seen after CID is a series of b and y ions. (Figure 1.8 in blue)[43] The activation energy of this technique is limited by the mass of both collision partners; the larger the precursor ion, the less effective the fragmentation will be.[55]

Another collisional activation technique is higher-energy collisional dissociation (HCD).[57] As its name suggests, HCD is a higher energy fragmentation technique, comparable to CID. The main difference lies in the fact that the energy of the collision can be as much as 100 times larger in HCD compared to CID.[57] This is accomplished by performing the collision in a multipole with the ability to reach

higher potentials. The type of high energy collision discussed in this dissertation is specific to ThermoFisher Scientific instruments.

### Electron Transfer Dissociation

Electron transfer dissociation (ETD) is a fragmentation technique that requires a reagent anion to transfer an electron to a precursor cation to induce fragmentation.[40,58] ETD was conceptualized as a way to allow electron capture dissociation (ECD) to occur in mass analyzers other than an FTICR.[59,58] Briefly, ECD involves the capture of an electron released from a heated filament source to react with multiply changed cation peptides. Both reactions result in a non-ergodic pathway, which means that fragmentation does not involve intramolecular vibrational energy redistribution.[59,58] ETD generally produces c and z ions (Figure 1.8 in red), as that is the bond where the radical reaction centers, and ETD has been reported as more effective at fragmenting higher mass peptides then CID or HCD.[43,60] The higher the charge on the peptide/protein, the faster the electron transfers and the reaction occurs, and thus a lower reaction time is required for ions with larger charge states. Because the reaction is non-ergodic, the side-chains and PTMs on proteins or peptides will remain intact while the N–Cα bond preferentially breaks.[40,41]

Sometimes fragmentation does not occur as readily and the major ion present in the product spectrum will be the charge reduced ion radical form of the precursor ion. Thus the base peak becomes the charge reduced precursor ion and the fragment ions tend to be very low in relative abundance. It is possible and beneficial to couple

ETD with HCD or CID to induce fragmentation.[43,61,62,63] By doing so, the ETD non-ergodic reaction is induced with only slight fragmentation from the supplemental CID or HCD.[64]

*Instrumentation*

Orbitrap LTQ-XL

The first generation of the orbitrap instrumentation introduced by ThermoFisher Scientific was the orbitrap LTQ-XL. (Figure 1.9) This instrument uses the resolving power of a standard orbitrap (maximum of 100,000 at 200 m/z)[65] to enable the study of large peptides and small proteins. The precursor ions enter the orbitrap by first being collected and correctly oriented in the C-trap. (Figure 1.9) The mass/charge ratio is measured on an orbitrap by dividing a constant (k) (found for each orbitrap specifically) by the frequency of the ions' movement along the length of the orbitrap (z-axis) (ω). (Eq. 1.1)

$$\frac{m}{z} = \frac{k}{\omega^2} \qquad \text{Eq. 1.1}$$

Once a precursor ion is selected for fragmentation, that ion is moved into the linear ion trap (for CID and ETD) or a multipole (for HCD) where the respective fragmentation occurs. (Figure 1.9) From there the mass analysis on the product ions could be accomplished via the orbitrap (high resolution) or the linear ion trap (low resolution). This instrument facilitated confident identifications in large proteomic

studies.[66,67] The orbitrap LTQ-XL was, however, limited in its ability to fragment large proteins.[68] The orbitrap LTQ-XL was optimized to run collisionally induced dissociation (CID) as the main fragmentation technique. However, CID is known to be less efficient at fragmentation of larger masses.[43]

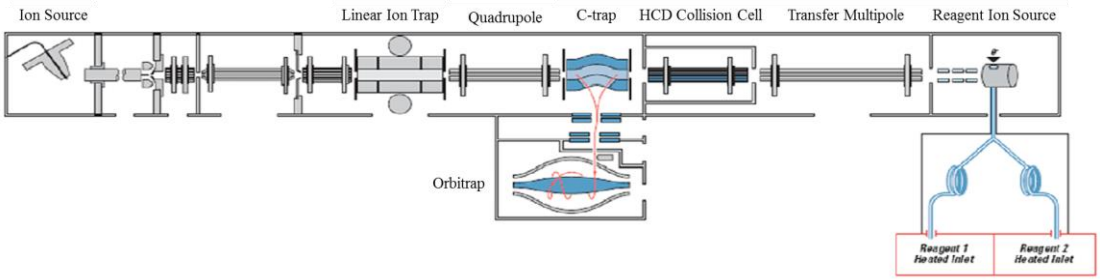



Figure 1.9. Layout of the orbitrap LTQ-XL. (http://planetorbitrap.com)

### Orbitrap Fusion Lumos Tribrid

The latest innovation from ThermoFisher Scientific is the orbitrap Fusion Lumos Tribrid mass spectrometer. As with all orbitrap mass analyzers, the resolution is incredibly high. However, the Fusion Lumos contains an ultra-high field orbitrap (unlike the orbitrap LTQ which had a standard orbitrap) acquiring a resolving power of up to 450,000 at 200 m/z with increased sensitivity and speed of acquisition.[65,68]

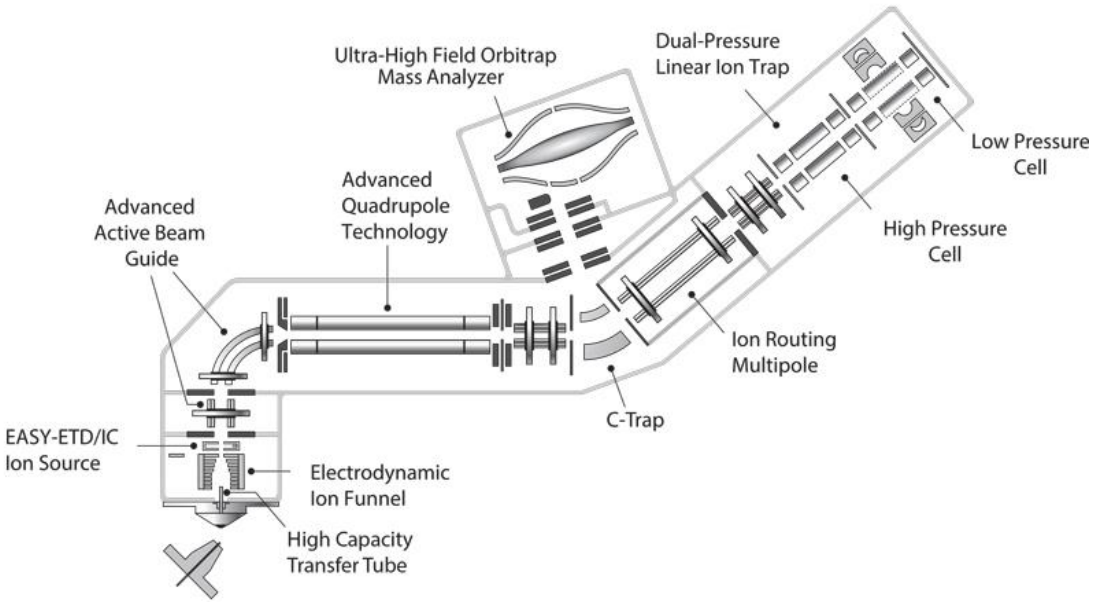



Figure 1.10. Layout of the orbitrap Fusion Lumos Tribrid (http://planetorbitrap.com)

The orbitrap Fusion has other improvements than just the orbitrap mass analyzer compared to the first generation. One of the most notable for top-down and middle-out analyses is more efficient pumps; these provide lower pressure in the IRM (ion routing multipole), orbitrap, and C-trap, which allows for larger masses to transfer more efficiently into the orbitrap for analysis. It also has a wider transfer tube, which allows for higher sensitivity, especially of high mass molecules. Another improvement that allows for more efficient top-down experiments is the use of a segmented linear ion trap (LTQ) (which was first introduced in the orbitrap Velos). The segmentation allows for more ions to enter the ETD reaction cell while keeping the reactive anion separate before they mix and the reaction takes place. As seen in Figure 1.10, there is a separation of the high pressure and low pressure LTQ, allowing for higher pressure in the reaction chamber, which increases ETD and CID efficiency

compared to that of the XL. To further improve ETD capabilities (which are necessary for our top-down experiments), the ETD source was moved from the back (Figure 1.9) to the front of the instrument (Figure 1.10). In the creation of the fluoranthene anion, there is also cation and neutral products produced. By running the fluoranthene product though the active beam guide, the anion can be separated and brought into the reaction cell free of any other species.[68]

*Bioinformatics*

Bottom-up bioinformatics searches that match fragmentation data with large protein repositories is a well-established medium.[69] These programs are able to search for –GG tagged ubiquitination sites on peptides, on a chromatographic time-scale.[33,34] As discussed previously, using bioinformatics in bottom-up proteomics has resulted in most of the qualitative and quantitative information available now for the ubiquitinome.

Middle-out spectra resulting from missed-cleavage are more difficult for search engines' algorithms to decipher. The processing power required to search and match the complex peptide mixtures from time-limited middle-out cleavage is available in new software such as Proteome Discoverer (PD) and ProSightPC (ThermoFisher).[70,71,72] However, the software still necessitates confirmation of the isopeptide linkage by manual curation.[51,52]

A few programs are available to process top-down proteomic data.[70,73,74] The programs can either help the user by defining common PTMs[73] (not Ub) or give mass

differences and allow the user to define the PTM (with suggestions from the program) by manual curation.[70,75,74] The former would be required for Ub identification.

One important limit of top-down proteomic algorithms is that limited fragmentation to completely characterize a PTM will dispute any identification's statistical significance and render the identification unusable;[76] this is especially true in analytes with isopeptide bonds. There is currently software that can accommodate for disulfide bonds,[77] however there is not an algorithm for dissecting intact polyUb isopeptide bonds. The difficulty lies in the fact that, for polyUb, there are isomeric subunits, which will give off almost entirely the same fragment ions. This leaves only a small set of ions that can be used to differentiate the identical theoretical structures. By using a graphical viewer to show the fragmentation present, the isopeptide linkages can be identified and bolstered by computer supported manual interpretation.

*Objectives*

Ubiquitination is a common post translational modification (PTM) which is traditionally discovered as a -GG tag on a lysine residue of a peptide after trypsinolysis. The small mass tag, a product of cleavage at R74 on Ub leaving glycinylglycine (-GG) isopeptide linked to a lysine, is used as a variable modification in proteomic search engines. These identifications can define the location of mono or polyUbs on target proteins, but tells nothing about the Ub chain itself. Characterization of the Ub chain topology is an ill-defined branch of proteomics, but is important for determination of the function of the ubiquitinome. It has been shown by methods beside mass spectrometry that the linkages within an Ub chain have

different functionalities within the cell. Not only that but very specific linkages with difficult to decipher topologies have been shown to have important functions. The methods are normally very tedious, as they require protein mutations within a cell line, or the use of many different antibodies or DUBs. It is difficult to determine both the linkages and the chain topology present in one experiment. To date there has not been a study that could map structures of the polyUbs in a facile manner. The work outlined in this dissertation will show that, by the use of mass spectrometry, the chain structure, including both linkages present and the chain topology can be elucidated in simple workflows.

# Chapter 2: Mass Spectrometric Analysis of Ub Dimers (Adapted from Ref 52)

*Introduction*

Ubiquitin (Ub) dimers are the most simple and robust form of polyUb.[78] Despite their simplicity, Ub dimers are reported as kinase activity activators when anchored on NEMO (NF-κB essential modulator),[79] they control specific hydrolysis and enzymatic activities,[80] and they have been found to have multiple specific DUBs. This suggests dimers have multiple relevant biological activities. Dimers have also been reported as unanchored *in vivo* (Figure 2.1).[81] Currently, functionality is not known for unanchored dimers; they have been suggested to be "building blocks" for the formation of larger chains.[81,82] The Ub that has a free C-terminus (unanchored seen with the -GG tail) or is attached to a substrate protein (anchored) is called the proximal Ub, labeled with a P. The Ub that has no other Ub attached to it, except by its own C-terminus, is called the distal Ub labeled with a D (Figure 2.1). For anchored chains, an S (for substrate) can simply be added within the subscript to represent the addition to the C-terminus of the moiety (i.e. $P_S$ Ub for proximal Ub conjugated to a substrate protein).

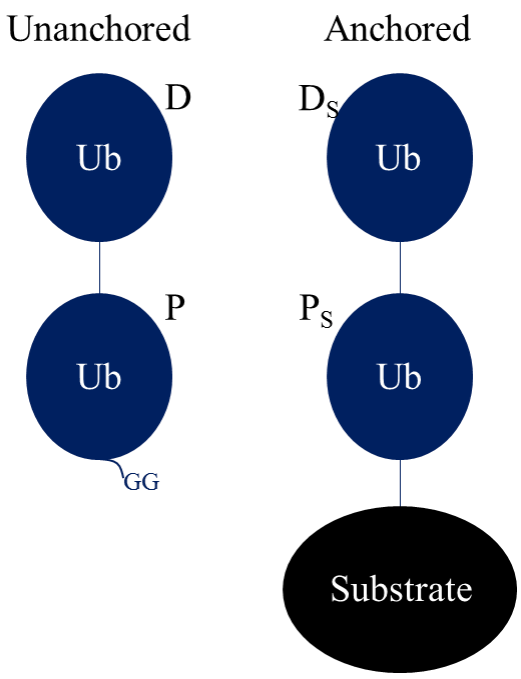



Figure 2.1. Ball-and-Stick representation of unanchored (left) and anchored (right) Ub dimers.

NMR and crystallography studies show that the different linkages within ubiquitin dimers create different overall quaternary structures. These differing structures have been assigned as the reason for the linkage based functionality.[83] For example, the two most abundant Ub isopeptide linkages, K48 and K63, form two distinct inter-subunit topologies.[84] K48 linked dimers form a closed conformation where the interactions keeping the two Ub moieties within close proximity is a hydrophobic patch including residues L8, I44, and V70.[85] This is in contrast to the topology of K63 dimers, which have an open conformation with little to no interaction between the subunits.[45]

Top-Down Approach

As stated in the first chapter, one of the most difficult aspects of ubiquitin chains' structural determination by mass spectrometry comes from the fact that the amino acid sequence is repeated in all the moieties within the chain. This creates many fragments that are redundant for each moiety and are thus not useful for determination of the chain's linkage (Figure 2.2). Complexity of the chain increases as the chain length increases (due to more possible topological features and possible linkage combinations), making analysis more difficult. Thus by starting with dimers, we can explore what aspects of the simplest example are unique, which will help build the method to longer and more complex structures.

As seen in Figure 2.2, the distal and proximal ubiquitin have the exact same sequence. The resulting redundant ions are largely ignored for any structural analysis. The ions that are unique and informative are starred in Figure 2.2. Specifically for unanchored chains, the C-terminus of the proximal Ub will be free. The series of z ions (seen in Figure 2.2 starred in gold), from the free C-terminus are unique and can tell the location of the linkage on the proximal Ub.

The same principle can be applied to the anchored dimer, except uniqueness is found in the z or y ion fragments on the proximal after the mass addition of the substrate protein (determined by top-down proteomic searches) is added to the C-terminus.

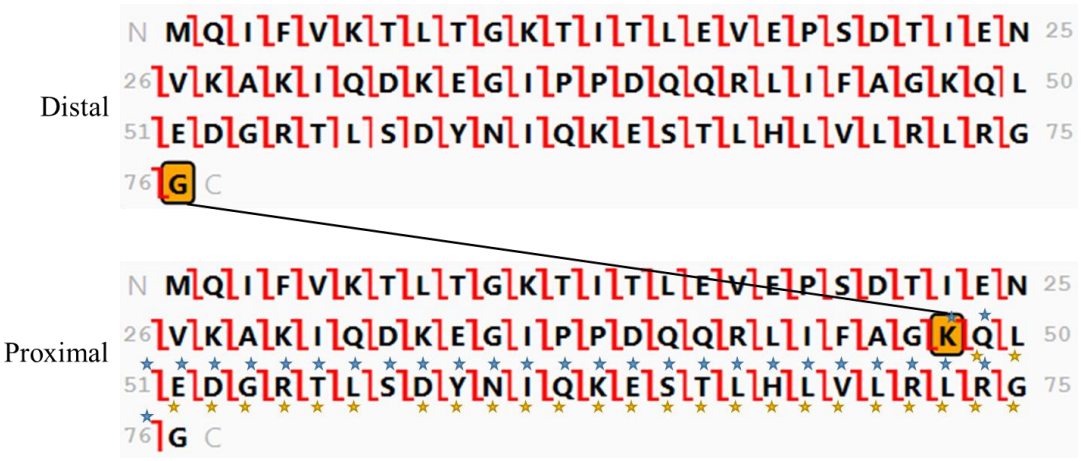



Figure 2.2. Theoretical full fragmentation pattern of c and z ions resulting from ETD fragmentation of Ub–$^{48}$Ub. The gold stars represent unique z ions and the blue stars represent the unique c ions.

Middle-Out Approach

To eliminate some of the redundancies that result from two identical moieties, it is possible to truncate the Ub dimer to a product that will give more unique fragment ions. This will also lower the intact mass, making it easier to detect and fragment on less efficient mass spectrometers, which give poor fragmentation data for the large intact masses of dimers.[81] More unique fragments will make linkage site identification simpler. (Figure 2.3) However, by truncating even just the distal Ub, information on the length of the polyUb chain is lost. There is also a plethora of different peptides created when using time-limited reactions. Thus a more complex mixture of ions must be separated using pre-MS/MS HPLC. Furthermore, the mixture created is of peptides similar in sequence and chemical properties, as they are just

fragments of Ub, making separation more difficult. If the length is known (perhaps by using middle-out as supplemental to a top-down analysis), then a MWAC middle-out strategy can be implemented without losing structural information.

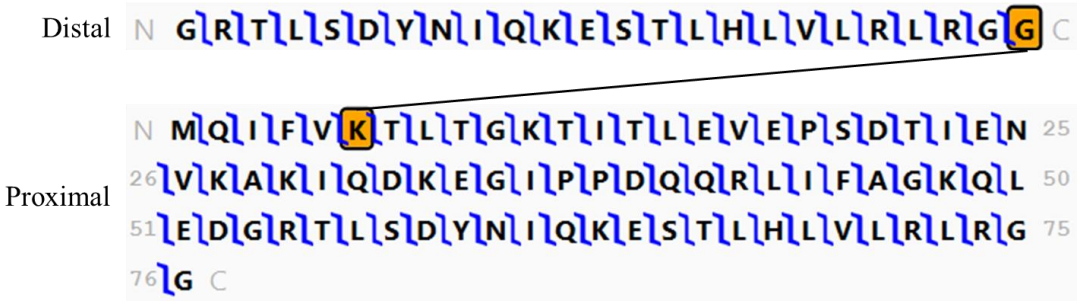



Figure 2.3. Theoretical full fragmentation pattern of b and y ions resulting from CID fragmentation of a K6-linked diUb. In this example of a truncated, branched peptide, all fragments present are unique. Truncation is achieved by time-limited MWAC.

The goal of the experiments reported in this chapter is to completely characterize the linkages present in a set of seven different isopeptide-linked Ub dimers. This will be done using the most current instrumentation available for top-down proteomics. This chapter will also discuss a method compatible with the use of older instrumentation that is not able to produce viable top-down results. The overarching goal is to set up a protocol that, along with characterizing Ub dimers, can be used as a stepping stone to characterize the more complex longer polyUb chains.

*Methods and Materials*

**Synthesis of Ubiquitin Dimers.** Ubiquitin polymers were prepared either chemically (K6, K11, K27, K29, K33 and Rub1-Ub and Ub-Rub1)[87] or enzymatically (K48, K63)[88] by members of the Fushman Laboratory.

**Middle-Out Microwave-Assisted Acid Cleavage**. Diubiquitin samples were diluted to 0.1mg/ml in a 100 µL 12.5% acetic acid solution and digested at $140^0$C using 300W microwave energy varying times in a CEM Discover microwave (Matthews, NC). Acidic solvent was removed by lyophilization in a FreeZone 2.5 Plus from Labconco Corporation (Kansas City, MO) before further analysis. These conditions have been previously reported to hydrolyze proteins with high selectivity at Asp residues,[50] yielding peptides that produce searchable CID spectra.[72]

**Middle-Out MALDI Analysis.** A Kratos Axima CFR MALDI-TOF MS (Shimadzu Biosciences, Columbia, MD) was used in linear mode to acquire mass spectra. One hundred scans were integrated per spectrum. Laser voltage was 84V and the matrix used was α-cyano-4-hydroxycinnamic acid at 10mg/ml in 70/30/0.1 acetonitrile/water/TFA. The 0.1mg/ml protein sample and the matrix solution were mixed 1:1 by volume for MALDI analysis.

**LC-MS/MS of Middle-Out Peptides**. After lyophilization, samples were diluted with HPLC grade water with 0.1% formic acid back to 100 µL. 5 µL of this solution was then injected, concentrated and desalted on a C8 trapping column (0.5×3 mm, Agilent Technologies) for 5 minutes before being separated through a pepSil C8 column (Column Technology Inc., Fremont, CA) with a flow rate of 300 nL/min and a gradient of 18-22% in 120 min (Solvent A: 97.5% HPLC grade water, 2.5% ACN,

and 0.1% formic acid; Solvent B: 97.5% ACN, 2.5% HPLC grade water and 0.1% formic). Analysis was performed in reverse-phase using a 2D nanoHPLC system (Shimadzu BioSciences, Columbia, MD) interfaced to an LTQ-orbitrap XL mass spectrometer (Thermo-Fisher, San Jose, CA). Precursor masses were acquired with a resolution of 60000 while fragment ions were acquired with a resolution of 30000 all within the orbitrap mass analyzer. Fragmentation was accomplished by CID with collision energy normalized at 35% NCE. The four most intense multiply charged precursor ions calculated from the intact peptide mass were targeted and fragmented in each cycle.

**Intact protein analysis by LC-MS/MS**. Intact dimers were diluted in Solvent A (97.5% water, 2.5% ACN and 0.1% formic acid) to 0.03 mg/mL The chromatography was performed using an Ultimate 3000 ultra-high performance liquid chromatograph (ThermoFisher Scientific, San Jose, CA) interfaced to an orbitrap Fusion Lumos Tribrid mass spectrometer (ThermoFisher Scientific). Two µL of intact sample were injected, concentrated, and desalted on a PepSwift Monolith trap (200 µm x 5 mm) for 5 mins before separation on a ProSwift RP-4H column (100 µm x 25 cm) (ThermoFisher Scientific) with a linear gradient of 30% to 60% solvent B (75% ACN, 25% water, 0.1% formic acid) over 15 mins. The potential for in-source fragmentation was set to 30V. Precursor and fragment ion masses were acquired with a resolution of 120,000 at 200 m/z. Fragmentation was triggered in data-dependent mode by electron transfer supported by chemical ionization (ETciD) with a 6 msec ETD reaction time and supplemental activation at 10% normalized CID.

**Interpreting the Spectra.** Precursor and fragmentation ions were deconvoluted using Xtract 3.0 (ThermoFisher Scientific). Fragment ions from the most intense m/z precursor ions selected in data-dependent mode were combined and then matched against the sequence of monoubiquitin. ProSightLite (http://prosightlite.northwestern.edu/) graphical interface[75] was used with a 5 ppm mass tolerance for top-down results and 20 ppm mass tolerance for middle-out results. The strategy utilized the monoubiquitin sequence as a template to assess the fragmentation patterns of each of the Ub moieties present in the dimers. ProSight Lite allows for custom mass additions to any amino acid in the template sequence. ProSightLite also identifies ions as c and z, formed by ETD, in red, and b and y, formed by CID, in blue.

## *Results and Discussion*

### Top-Down Strategy

Unanchored natural diUb with all seven isopeptide linkages were obtained and analyzed to determine if mass spectrometry could be used to determine the linkages present. To interpret the fragmentation spectrum and determine the linkage site, the number of Ub moieties first must be determined, as a different method must be applied to each polyUb chain length (discussed in subsequent chapters). Ub dimers were initially identified by their intact mass. All intact Ub molecular masses were matched within 5 ppm of the theoretical mass. (17101.22 Da) (Table 2.1)

Table 2.1. Intact mass analysis of each isopeptide-linked diUb. Theoretical mass of the all-natural dimers is 17101.22 Da.

| Ub isopeptide linkage | Observed mass (Da) | Mass Difference (ppm) |
|---|---|---|
| Ub–$^{63}$Ub | 17101.30 | 0.6 |
| Ub–$^{48}$Ub | 17101.24 | 0.6 |
| Ub–$^{33}$Ub | 17101.26 | 3.5 |
| Ub–$^{29}$Ub | 17101.24 | 1.2 |
| Ub–$^{27}$Ub | 17101.28 | 2.3 |
| Ub–$^{11}$Ub | 17101.21 | 1.2 |
| Ub–$^{6}$Ub | 17101.23 | 4.7 |

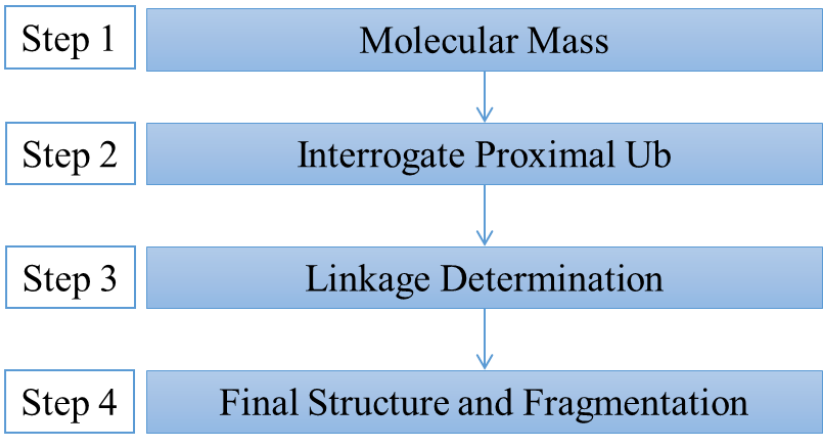


Figure 2.4. Workflow designed to characterize unanchored Ub dimers.

Once the polyUb is determined to be a dimer by its mass (Step 1: Table 2.1), a specifically designed workflow can be followed to determine all aspects of the chain's structure (Figure 2.4). Step 2, interrogate the proximal Ub occurs next. In Step 2, the fragmentation pattern of Ub is confirmed (to avoid the rare case that the intact mass is the same as that of another protein). For unanchored dimers, the y or z ions from the proximal Ub will be unique to a certain linkage. Practically, this means that wherever the y or z ions from the unmodified sequence of Ub drop off is where the linkage should be located in the proximal moiety of the dimer. Since there is only one Ub moiety in the dimers that has a Lys-linked isopeptide, this should be the only topological feature that needs addressing. In future chapters, there will be many more complex steps needed to determine topology. Step 3, linkage determination, aims to confirm the linkage that is suggested in Step 2. By adding the mass of the distal Ub to the N-terminus, fragmentation will appear that specifically represents only the linkage. By using all the information built by the previous steps, Step 4 is the visualization of the final structure using the fragmentation available.

The theoretical idea of Step 2 is visualized here in Figure 2.5. All the isopeptide-linked dimers can be seen to have nearly complete fragmentation coverage of the Ub sequence with the b and c ions, suggesting monoUb is the correct sequence. The y and z ions tell a different story. The C-terminal fragment ions (y and z) consistently stop before the theoretical mass addition of the distal Ub on the proximal Ub. For example, in Figure 2.5g (Ub–$^{6}$Ub), y and z ions can be followed all the way past K11 (eliminating that Lys as having a mass addition) and up to L8, stopping completely before K6.


Figure 2.5. Visualization of Step 2, interrogate proximal Ub, in the unanchored Ub workflow. All predicted mass addition locations are boxed.

Step 3 can then be implemented to confirm the finding in Step 2. What this means practically is that, if the mass addition of the distal Ub is placed on the M1 (the N-terminal) of the proximal Ub, the only fragment ions that appear should be in support of the Lys that is suggested as the linkage site in Step 2 (Figure 2.6). For example, for the K48 dimer, the y and z ions stop before K48, suggesting it as the correct isopeptide linkage location (Figure 2.5b), however, this prediction is based only on the absence of one type of ion (y or z). To confirm the linkage at K48 the fragmentation pattern was assessed when the distal mass is added to the M1. Figure 2.6b shows that all the redundant fragment ions disappear and what is left are the fragments that only support the linkage location. In the K48 example, no fragments

supporting other Lys residues can be seen, and the c ions formed start after K48 (at E51). Therefore all the unique ions support only K48. This is true for all the linkages studied, confirming each one as the linkage predicted by the synthesis (Figure 2.6).

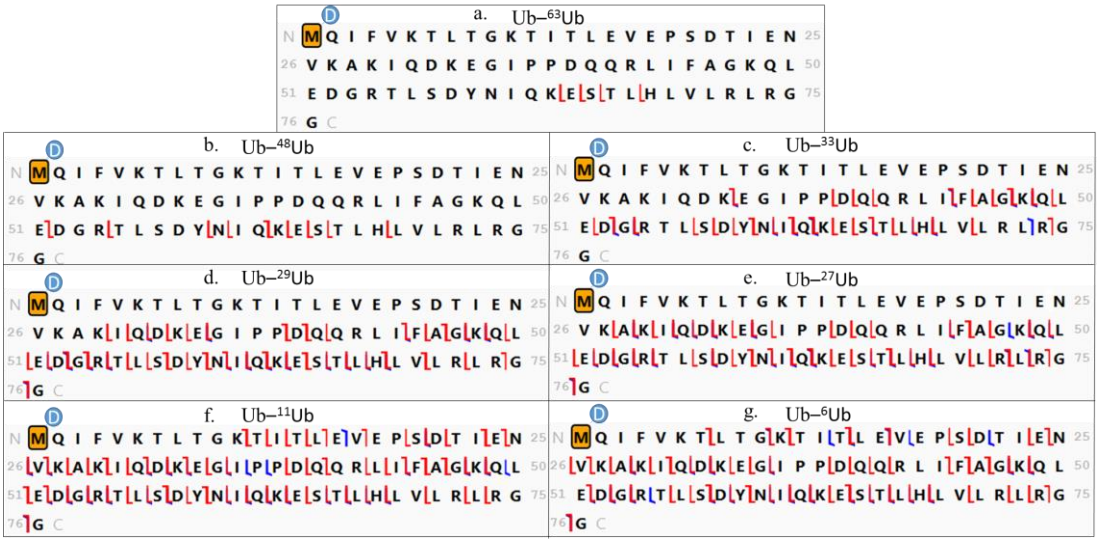



Figure 2.6. Step 3, linkage determination, visualized for all seven isopeptide linkages of diUbs. The initial methionine is highlighted to indicate the trial addition of the mass of the distal Ub (8541.6056 Da), which is represented by a blue ball. Fragmentation can be seen in support of each theorized linkage location as labeled.

Once the linkage location is confirmed, Step 4, Final structure and fragmentation can be visualized by placing the distal Ub on the predicted Lys and linking the two Ub moieties together in the proper orientation. (Figure 2.7)


Figure 2.7. Visualization of Step 4, final structure and fragmentation. Each Ub dimer is represented with the correct theoretical linkage based on the synthesis.

Middle-Out Strategy

To perform top-down protein analysis requires instrumentation that has high enough sensitivity and fragmentation capability to fragment large intact masses of proteins. If this is not available, or if a top-down experiment did not provide enough fragmentation for a confident identification of a polyUb chain, that work can be supplemented with a middle-out workflow. In the case of Ub dimers, MWAC reproducibly produced peptides that retained the C-terminus of the proximal Ub and truncated the proximal and distal Ubs to produce a peptide that was lower in mass and thus easier to characterize. (Figure 2.8)

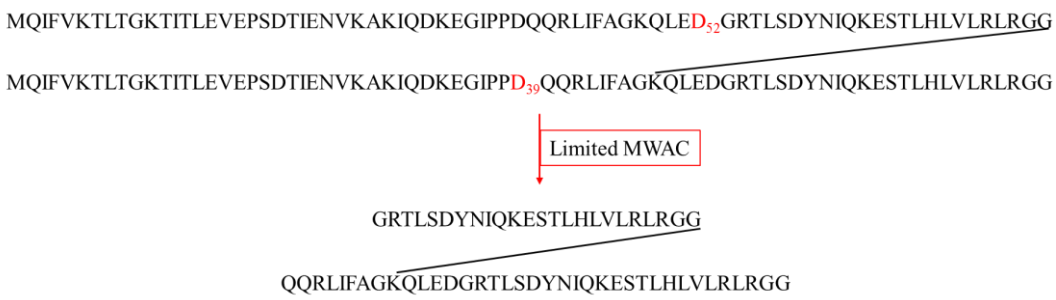



Figure 2.8. Truncation of an intact Ub dimer linked at K48 to a middle-out peptide. Mass of the intact dimer, 17101.22 Da, is reduced to 6942.77 Da after limited MWAC after digestion which only cleaves certain Asp residues.

MWAC is limited to produce missed cleavage events by lowering the time of the reaction. The complete digestion of diUb takes 15 mins, but nearly complete digestion can also be seen in as little as 3 mins. Different short time trails were attempted to see which produced different truncated peptides (Figure 2.9). A 30 sec digestion produced many differently massed peptides according the MALDI-TOF analysis, and, relative to the 20 and 10 sec runs, there was very little intact Ub left over. Thus, a 30 sec digestion time was chosen for time-limited MWAC.

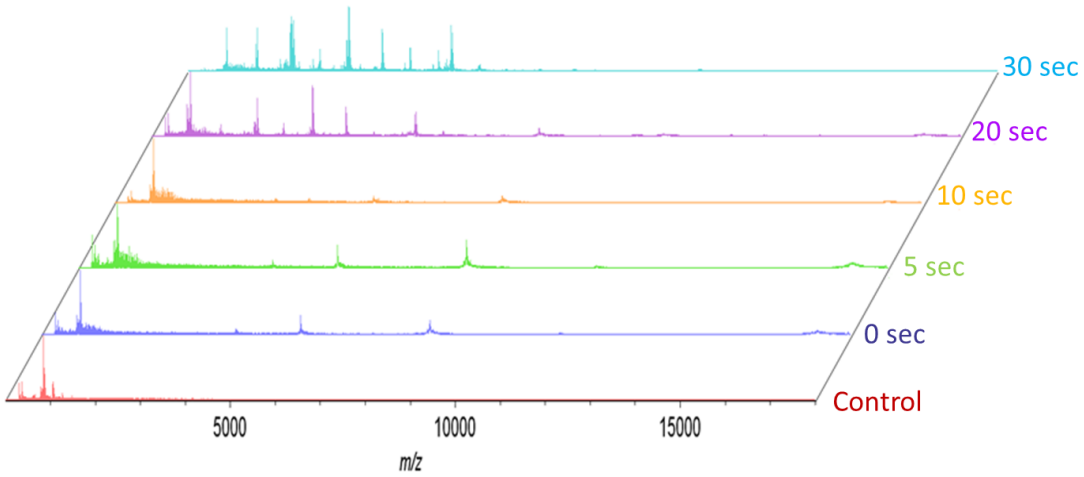



Figure 2.9. Time-trials for MWAC of a K63 linked Ub dimer. The control in red is the spectrum of the matrix, α-CHCA, and the rest are spectra of products from increasing times of MWAC. At time 0 sec we can see the $MH^{+}$ at approximately 17100 Da, the $MH^{+2}$ at approximately 8550 Da, the $MH^{+3}$ at approximately 5700 Da, and the $MH^{+4}$ at approximately 4300 Da.

Precursor mass analysis on the orbitrap LTQ-XL revealed high mass accuracy (all errors fell within 11 ppm) for each of the truncated middle-out peptides of interest. (Table 2.2) An interesting note on the MWAC products is that the truncated peptides selected here are not all unique to only one linkage with the exception of the K63 linkage product. For example, the truncated peptide selected for K48 in Table 2.2 can also be seen in the hydrolysis of K63 linked dimer, but the hydrolysis product for K63 would not be seen in a hydrolysis of K48. (Figure 2.10) This means that the intact mass of the peptide cannot definitively identify the linkage as K48 (or K33, K29, K27, K11, K6), but must be coupled to fragmentation data in tandem mass spectrometry to identify the linkage. The only acceptation to this rule is K63 linkage, whose mass is unique to the linkage.[49]

Table 2.2. Sequences and the theoretical and observed monoisotopic masses of the truncated branched peptides analyzed from the all-native K-linked ubiquitin dimers separately.

| Sequences of the Truncated Target Peptides | Mass (Da) | |
|---|---|---|
| | Theor. | Obser. |
| GRTLSDYNIQKESTLHLVLRLRGG \| K63 GRTLSDYNIQKESTLHLVLRLRGG | 5432.98 | 5433.04 |
| GRTLSDYNIQKESTLHLVLRLRGG \| K48 QQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG | 6942.77 | 6942.83 |
| GRTLSDYNIQKESTLHLVLRLRGG \| K33 KEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG | 7696.17 | 7696.23 |
| GRTLSDYNIQKESTLHLVLRLRGG \| K29 TIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG | 8935.85 | 8935.89 |
| GRTLSDYNIQKESTLHLVLRLRGG \| K27 TIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG | | |
| GRTLSDYNIQKESTLHLVLRLRGG \| K11 MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG | 11267.09 | 11267.03 |
| GRTLSDYNIQKESTLHLVLRLRGG \| K6 MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG | | 11267.12 |

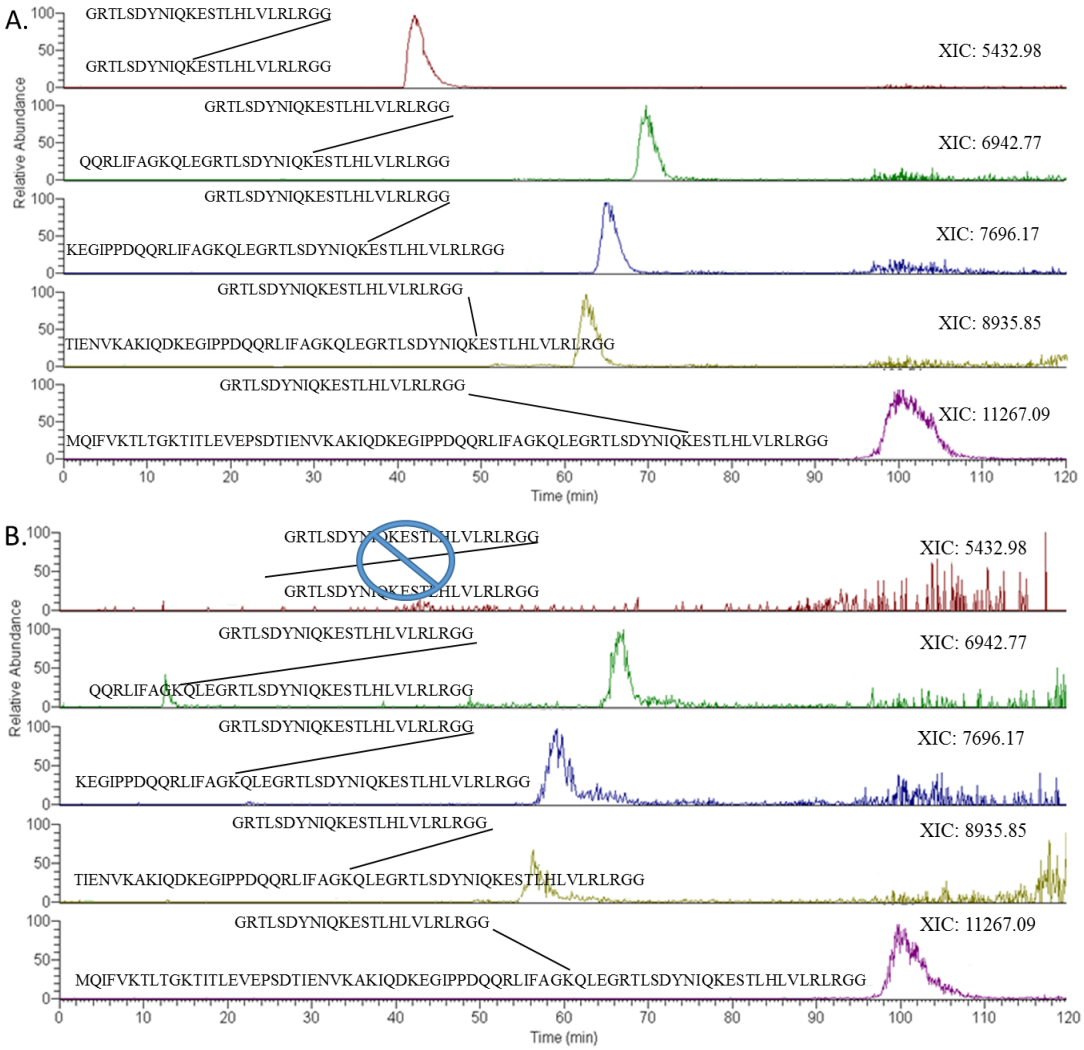



Figure 2.10. The extracted ion chromatograms (XIC) for A. hydrolysis products of K63 linked dimer and B. hydrolysis products of K48 linked dimer. Sequences of the truncated peptides of interest represent the masses selected for the XIC.

To properly characterize the linkages, CID fragmentation was used on targeted m/z values. The resulting fragmentation patterns are seen in Figure 2.11. Most of the linkages were identified correctly to their theoretical linkage site. In the

case of K27, K29, and K11, linkages were constricted to only two possible linkage sites. The K27 and K29 residues are so close that it was difficult to obtain fragmentation to distinguish them on the first generation orbitrap. They could be differentiated in complete acid hydrolysis seen in Appendix Table 1 and Figure 1. A lack of fragmentation density was not an issue with the reported top-down work, as K6, K11, K27 and K29 could be differentiated in dimers using the more advanced instrumentation.

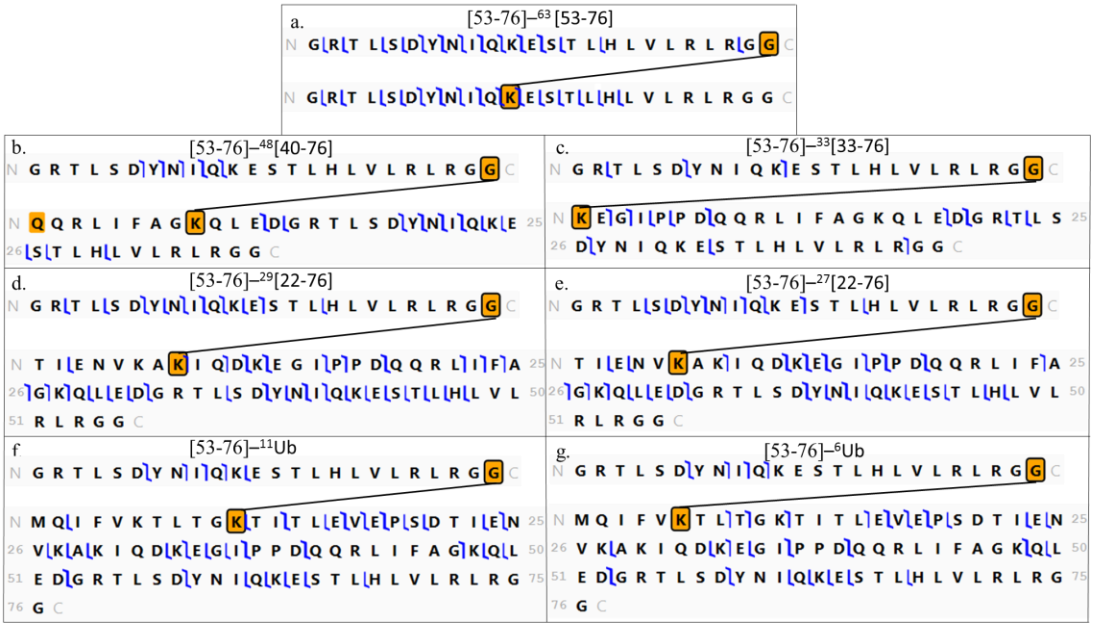



Figure 2.11. Fragmentation patterns for selected time-limited MWAC products for all isopeptide-linked Ub dimers. Lysine residues highlighted in gold with a black box surround are the sites of the isopeptide linkages. The gold highlighted Q in b. represents the loss of 17.02655 from N-terminal Glutamine (Q).

*Summary*

Ub dimers can be completely characterized on a chromatographic time scale by LC-MS/MS using a top-town proteomic workflow. To accomplish this, it is necessary to take advantage of recent advances in highly efficient fragmentation and mass analyzers that are capable of high mass detection. A facile workflow directed specifically at characterizing the linkages of Ub dimers was created, which takes advantage of these advances. However, if this equipment is unavailable, characterization can be accomplished by middle-out analysis.

# Chapter 3: Mass Spectrometric Analysis of Ub Trimers (Adapted from Ref 53)

*Introduction*

Ubiquitin (Ub) trimers have been reported *in vivo* at low relative abundance and are not a common length for polyUb studies.[33] Trimers were reported as active, potent, and specific activators of RIG-I,18[81] and K48 linked trimers were reported to have greater proteolytic ability than dimers, but less than tetramers.[86] More specific functions have not been elucidated.

Trimers may not be as prevalent or active in cells compared to dimers and tetramers, but developing a method to characterize trimer is vital to conceptualize a facile strategy to characterize longer chains that are readily found in cells. Compared to dimers, trimers have an extra ubiquitin moiety, and so must be considered differently because distinctive topologies, along with all 8 linkages locations, are possible.[53] The triUb chain can form theoretically two different topologies: unbranched and branched. (Figure 3.1) The unbranched chain in Figure 3.1 has one proximal and one distal Ub, like the dimers along with a new and unique moiety called the endo Ub. (Figure 3.1 left) The endo Ub in trimers is defined as a Ub moiety with a single Ub on its C-terminus and another Ub on one of its Lys residues. The branched chain has a proximal Ub and two distal Ubs. (Figure 3.1 right)

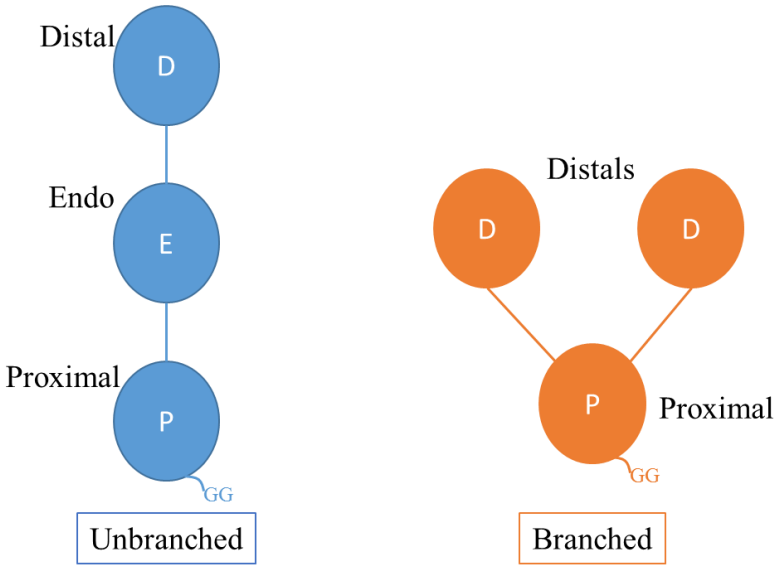



Figure 3.1. The two general topologies available for trimers are unbranched (left) and branched (right). The unbranched chain has three distinct moieties, a proximal, distal and endo, whereas the branched has no endo, only a proximal and two distals.

In an unknown trimer, there would be 92 possible combinations of the different topologies and linkages (64 linkage possibilities for unbranched and 28 for branched).[33] When all the moieties result in almost the same fragmentation, the question becomes, what fragmentation can we use to determine the topology and the linkage? It would be impossible to tell the difference between a branched Ub and an unbranched Ub by almost all the fragment ions that appear. The difference between the unbranched and branched is the presence of an endo Ub.[23]

Top-Down Approach

To differentiate the two available topologies, fragmentation must find which is truly unique to only one topology. This fragmentation is seen in the endo Ub of the unbranched Ub–$^{6}$Ub–$^{63}$Ub between K63 and the C-terminus (G76) shown in Figure 3.2 in green. The masses of the fragmentation within this region is diagnostic for the endo Ub because only the proximal Ub is attached to the C-terminus in an unbranched Ub, but in a branched, there would be no moiety that has only one Ub attached to the C-terminus of another moiety. Both distals in the branched Ub have a proximal and the other distal mass linked to the C-terminus; thus two moieties are added (Figure 3.2). This is also true for fragmentation toward the N-terminus, where the endo Ub containing moiety would have fragment masses in the same region with only one Ub added to the N-terminal.

Shown in Figure 3.2 the fragment ions marked with purple and gold stars are unique to the linkages present, but not the topology (branched vs. unbranched). The gold starred fragments are unique to the linkage on the proximal Ub and would be the same for the branched and unbranched chains. This is the same principle as Step 2 in Figure 2.4. The purple starred fragments show the fragment ions that are unique for the linkage in the proximal for the branched, and in the endo Ub for the unbranched. (Figure 3.2) This chapter's aim is to prove this theoretical concept and show a mass spectrometric method researchers can follow for their own work.

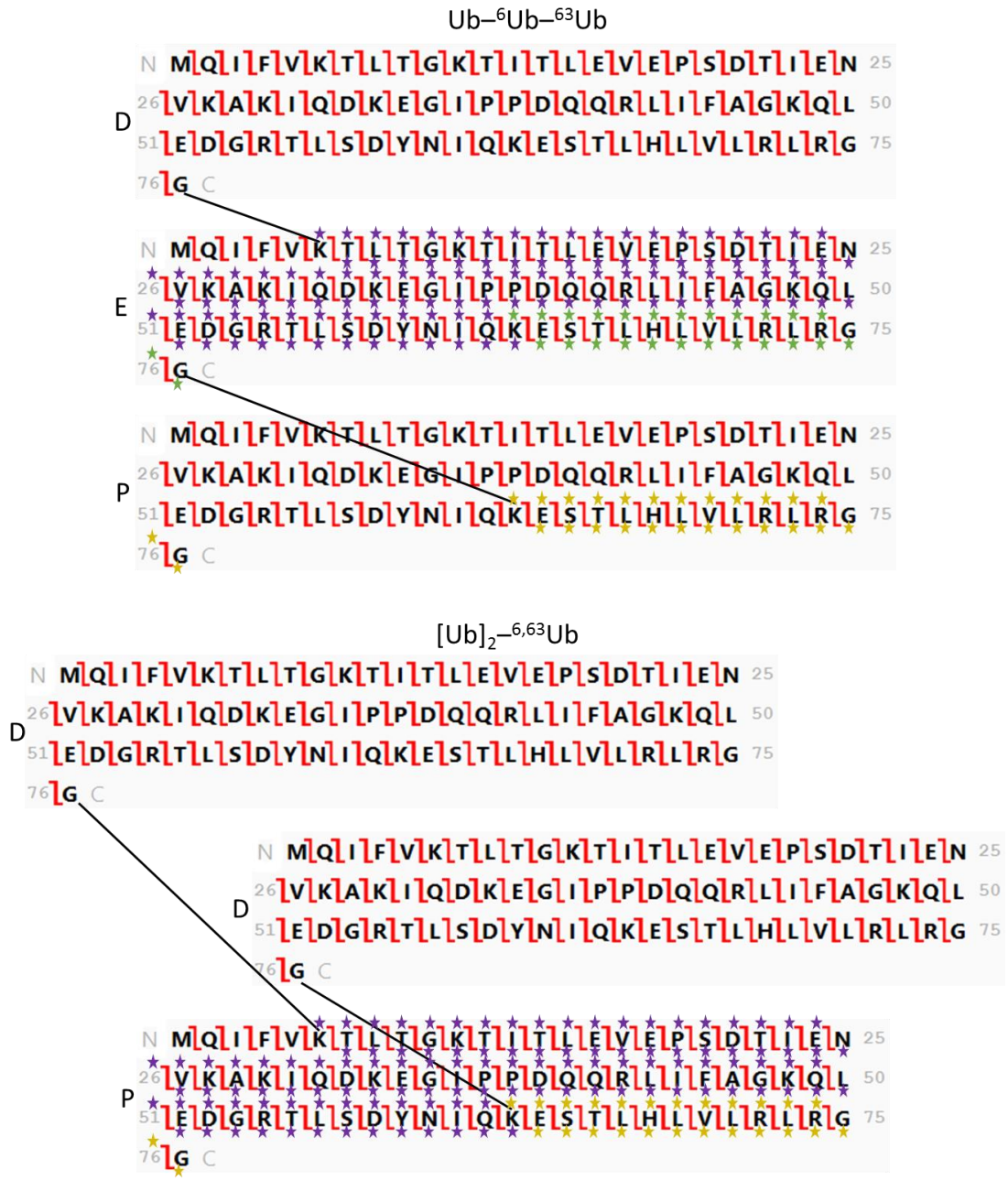



Figure 3.2. Theoretical full fragmentation pattern for two Ub trimers that differ only in topology. Color coding shows the fragmentation that is unique to the linkages present, but the same in both (purple and gold) and fragmentation that is different between the topologies (green). The top image is that of a unbranched Ub–$^{6}$Ub–$^{63}$Ub and bottom, a branched [Ub]$_{2}$–$^{6,63}$Ub.

Middle-Out Approach

If top-down methods are unable to differentiate the topologies, or if there isn't enough fragmentation to determine the linkages present, middle-out methods may prove useful. If hydrolysis occurs on only the distal Ub moieties at D52 (colored red in Figure 3.3), then the chain can be truncated to produce a smaller more easily fragmented peptide (each truncation at D52 in the distal Ub results in a reduction of 5816.2 Da). The unbranched chain will be reduced from 25624.8 Da to 19808.6 Da. (Figure 3.3, Ub–$^{6}$Ub–$^{63}$Ub) Better fragmentation may make it easier to identify the endo Ub, proving the topology. On the branched chains both distal Ubs can be truncated reducing the mass from 25642.8 Da to 13974.6 Da. (Figure 3.3 $[Ub]_2$–$^{6,63}$Ub) By doing so the sequence on the proximal Ub is distinctive and there are unique fragments, which can be used to more easily assess the linkages present.

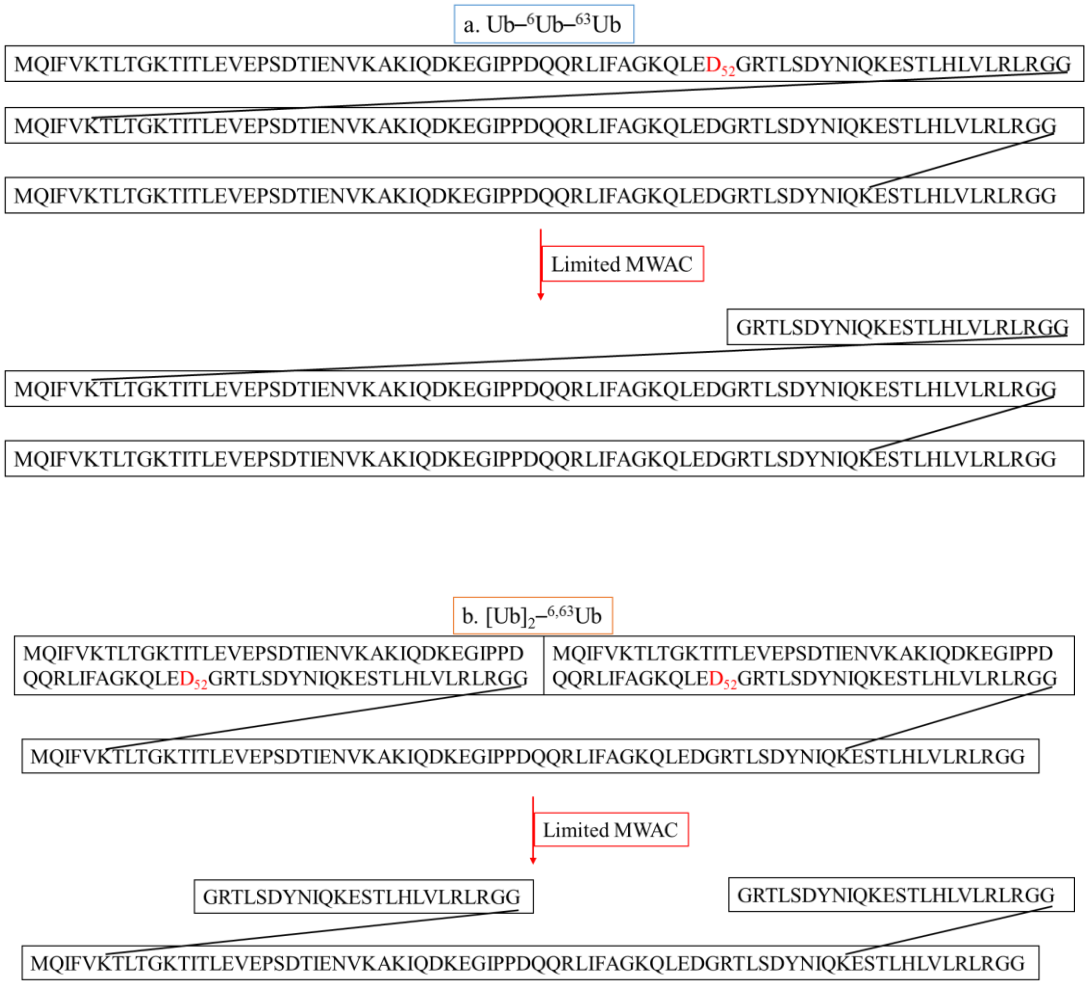



Figure 3.3. Ideal truncation of a. the unbranched (Ub–$^{6}$Ub–$^{63}$Ub) and b. branched ([Ub]$_2$–$^{6,63}$Ub) triUb topologies using limited MWAC.

*Methods and Materials*

**Synthesis of Ubiquitin Trimers**. Ub–$^{33}$Ub–$^{33}$Ub and [Ub]$_2$–$^{11,33}$Ub were assembled chemically through silver-mediated ligation of an activated Ub to the selectively deprotected lysine of the other Ub.[88,89] [Ub]$_2$–$^{6,48}$Ub (E1, Ubch7, and NleL), [Ub]$_2$–$^{11,63}$Ub (E1, Ube2s, MMS2, and Ubc13),[46] Ub–$^{48}$Ub–$^{48}$Ub (E1 and E2-25K),[85] and Ub–$^{63}$Ub–$^{48}$Ub (E1, MMS2, and Ubc13, E2-25K)[23] were generated enzymatically.[85,45] All chains were produced by members of the Fushman Laboratory.

**Trimer Analysis by LC-MS/MS with ESI.** Intact trimers were diluted in solvent A (97.5% water, 2.5% ACN and 0.1% formic acid) to 0.03 mg/mL The chromatography was performed using an Ultimate 3000 ultra-high performance liquid chromatograph (ThermoFisher Scientific, San Jose, CA) interfaced to an orbitrap Fusion Lumos Tribrid mass spectrometer (ThermoFisher Scientific). Three µL of intact sample were injected, concentrated, and desalted on a PepSwift Monolith trap (200 µm x 5 mm) for 5 mins before separation on a ProSwift RP-4H column (100 µm x 25 cm) (ThermoFisher Scientific) with a gradient of 20% to 40% solvent B (75% ACN, 25% water, 0.1% formic acid) over 15 mins. The potential for in-source fragmentation was set to 10V. Precursor and fragment ion masses were acquired with a resolution of 120,000. Fragmentation was triggered in data dependent mode by electron transfer supported by chemical ionization (EThcD) with a 6 msec ETD reaction time and supplemental activation at 10% normalized HCD (optimized as Appendix Table 2). To obtain high density fragmentation and high sensitivity, top-

down parameters provided by ThermoFisher Scientific were used with minimal altercations.

**Interpreting the Spectra.** Precursor and fragmentation ions were deconvoluted using Xtract 3.0 (ThermoFisher Scientific). Fragment ions from the top m/z precursor ions selected in data dependent mode were combined and then matched against the sequence of monoubiquitin using ProSight Lite (http://prosightlite.northwestern.edu/) with a 10 ppm mass tolerance. In our strategy, the monoubiquitin sequence is used as a template to assess the fragmentation patterns of each of the Ub moieties present in the trimer. ProSight Lite allows for custom mass additions to any amino acid in the template sequence. Masses equivalent to one or two Ub moieties were added, and changes in fragmentation patterns assigned by ProSight Lite were used to assign the topology of each trimer as branched or unbranched chain. This is discussed in detail in the Results and Discussion section. Finally, linkage sites were assigned by inspection of fragmentation patterns assigned to the monoubiquitin template. ProSight Lite also identifies ions as c and z formed primarily by ETD, and b and y formed primarily by HCD.

**Microwave-Assisted Acid Cleavage**. Ubiquitin trimers were diluted to 0.15 mg/mL in 12.5% acetic acid and digested for 60 sec at $140^0$C using 300 W of power in a CEM Discover microwave (Matthews, NC). These conditions have been previously determined to produce partial cleavage of polyubiquitins at Asp residues (Appendix Table 3).[49,52] Digested trimers were lyophilized and resuspended in solvent A (97.5% water, 2.5% ACN and 0.1% formic acid) at 0.1 mg/mL for chromatography performed using the LC-MS/MS system specified above. Five µl of

each digested sample were injected, concentrated, and desalted on a Zorbax C8 trap (0.5X3 mm, Agilent Technologies) for 5 mins before being separated on a Zorbax C8 column (3.5 µm, 150 mm X 75 µm, Agilent Technologies, Santa Clara, CA) with a 500 nL/min flow rate and a gradient of 30% to 37% solvent B (solvent B: 75% ACN, 25% water, 0.1% formic acid) over 45 mins. Spectra were acquired and processed as described above.

## *Results and Discussion*

### Top-Down Strategy

PolyUb chains were interrogated to find the topology and linkages present using one mass spectrometry based workflow. To determine and test a workflow, a set of six triUbs (Figure 3.4) were obtained. An orbitrap Fusion Lumos mass spectrometer is shown to provide and record extensive fragmentation in high mass proteins.[90] However, in highly repetitive isopeptide Ub moieties, it is also necessary to be able to interpret the spectrum. Using a graphical interface to visualize the fragmentation, a facile and novel workflow was established to interpret the trimers.

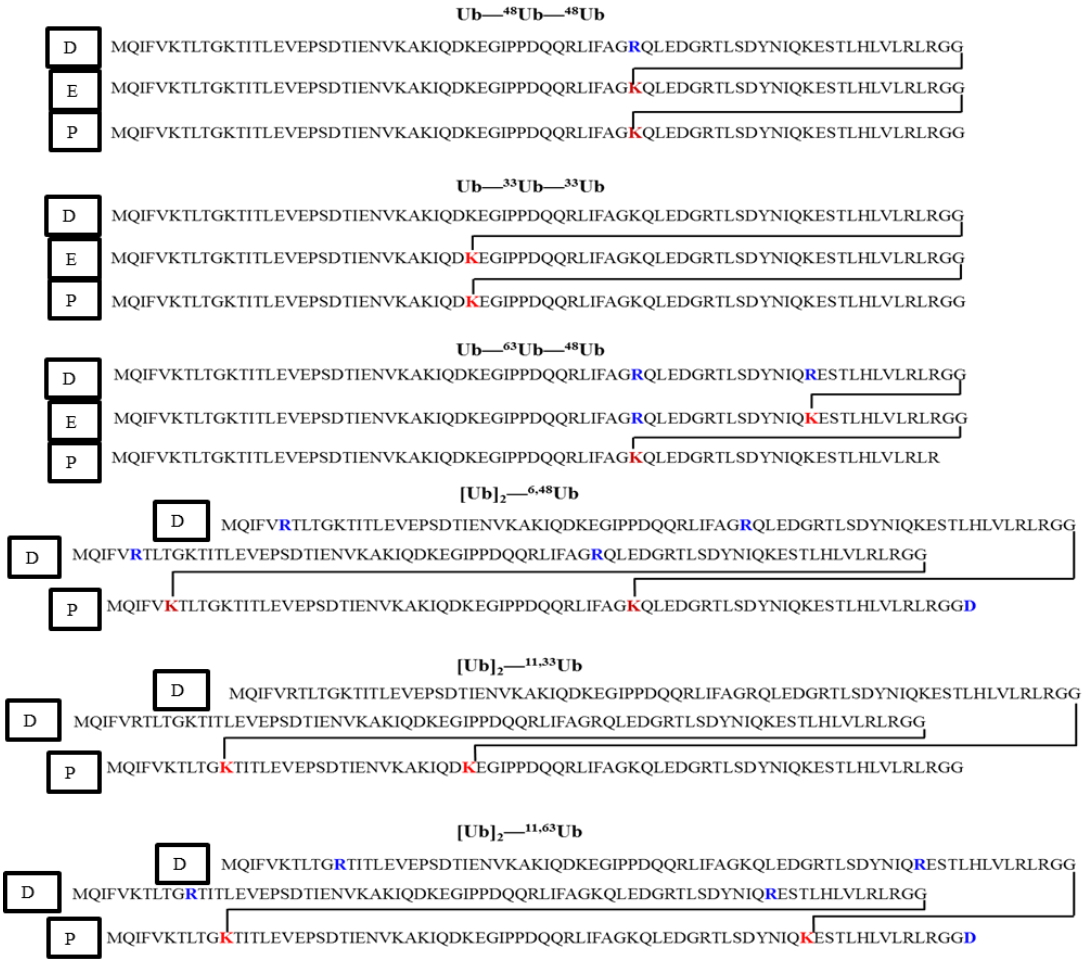



Figure 3.4. Sequence and connectivity of each unbranched (top 3) and branched (bottom 3) standard trimer. From top to bottom trimers present are Ub—$^{48}$Ub—$^{48}$Ub, Ub—$^{33}$Ub—$^{33}$Ub, Ub—$^{63}$Ub—$^{48}$Ub, [Ub]$_2$—$^{6,48}$Ub, [Ub]$_2$—$^{11,33}$Ub, and [Ub]$_2$—$^{11,63}$Ub. Residues in red represent the modified lysines. Residues in blue represent mutations made for the synthesis of the trimer from K to R. Ub—$^{63}$Ub—$^{48}$Ub also is missing residues G75 and G76 which could not be highlighted. Ubiquitin moieties are labeled with a D for distal, E for endo, or P for proximal.

The workflow designed for Ub trimers has many similar steps to that of the dimers. However there is a very vital step added to determine the topology of the chain, which dimers did not have to address. (Figure 3.5)

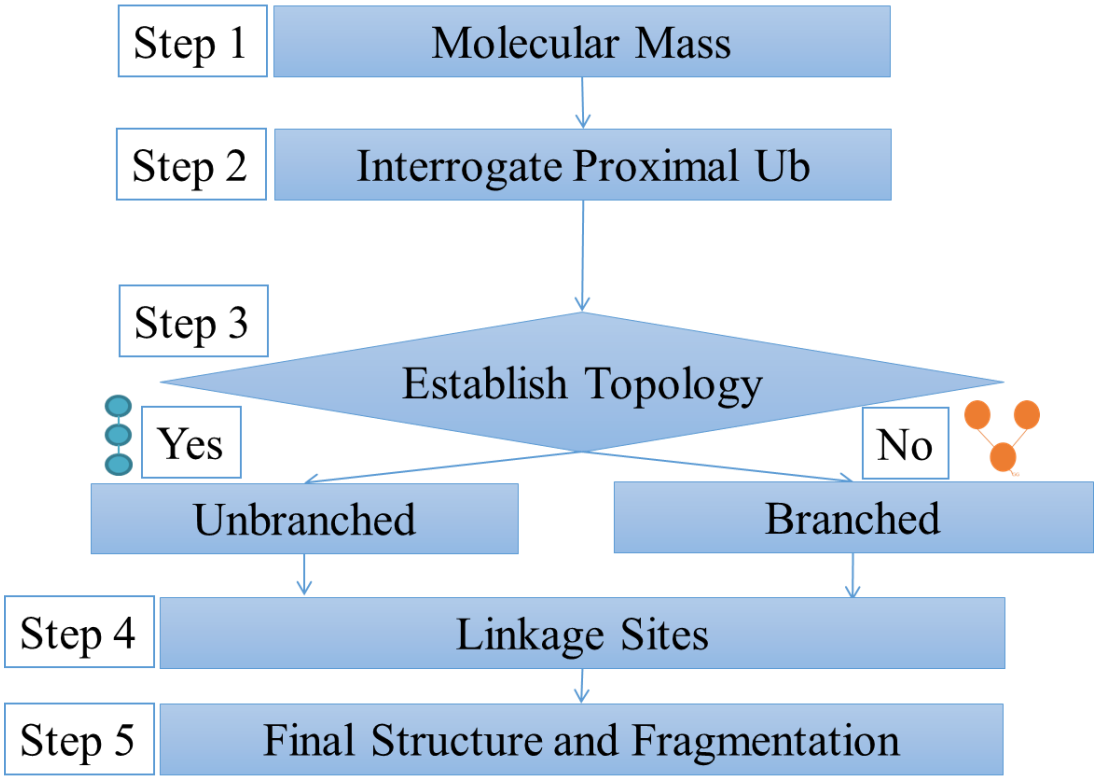



Figure 3.5. Workflow designed specifically to characterize Ub trimers with each step described to the left.

The path for this workflow is specific to trimers, (Figure 3.5) and thus Step 1, "Molecular Mass" determination, must confirm the chain as a trimer. The fragmentation pattern can then be matched to a sequence of monoUb and can be

further interrogated using ProSight Lite, which allows for custom mass additions to any amino acid in the sequence. In the "Proximal" step (Step 2) we follow the z ions from the free C-terminus of the proximal Ub. Where the fragmentation terminates is where one of (or the only) mass addition on the proximal Ub occurs. To determine if there are two Ub masses (unbranched chains) or one Ub mass (branched chains) added to that particular lysine, we move on to the "Establish Topology" step (Step 3). Does the addition of a single Ub mass to the C-terminus and another to the N-terminus create diagnostic c or z fragmentation between E63 and G76? If yes, then the chain topology is unbranched, if no, then it is branched; as demonstrated in Figure 3.2. Once the topology is known the "Linkage Sites" can be determined in Step 4 by adding the Ub moieties to a specific Lys residue, the fragmentation pattern that emerges will determine the linkages present. Once the linkage is known Step 5, the "Final Structure and Fragmentation", can be visualized and put together and the chain, completely characterized.

Table 3.1. Calculated and measured molecular masses of six triUbs.

| Trimer | Calc. Intact Mass (Da) | Exp. Intact Mass (Da) | Mass Diff. (ppm) |
|---|---|---|---|
| Ub–$^{48}$Ub–$^{48}$Ub | 25670.83 | 25670.86 | 0.9 |
| Ub–$^{33}$Ub–$^{33}$Ub | 25642.83 | 25642.87 | 1.5 |
| Ub–$^{63}$Ub–$^{48}$Ub | 25612.81 | 25612.8 | 0.14 |
| $[Ub]_2$–$^{6,48}$Ub | 25869.88 | 25869.79 | 3.4 |
| $[Ub]_2$–$^{11,33}$Ub | 25642.83 | 25642.77 | 2.2 |
| $[Ub]_2$–$^{11,63}$Ub | 25869.88 | 25870.01 | 5 |

High-resolution mass spectrometry was used to identify each isomer's intact mass within 5 ppm. (Table 3.1) Once the sample is recognized as a Ub trimer by Step 1, extensive fragmentation is required to determine all other features of the chain. In Step 2 all the fragment ions are matched using ProSight Lite to the sequence of the proximal Ub (Figure 3.6). High fragmentation density will support the tentative assignment as an Ub. In Step 2, a linkage site on the proximal Ub is elucidated. Because the only free C-terminus is in the proximal Ub, any z ions characterized on the monoubiquitin template must be formed from the proximal Ub. This series of z fragment ions will be terminated on the template when a mass addition occurs in the sequence (Figure 3.6). This is exactly the same concept as Step 2 from the dimer workflow (Figure 2.4). The only difference is whether we are adding two Ub masses (for an unbranched chain), or a single Ub mass (for a branched).

When using ETD fragmentation with supplemental activation (by HCD in this chapter), fragmentation is not expected to occur readily from the supplemental activation.[61,43] Because HCD is known to produce internal fragmentation,[64] the present work uses high mass accuracy in the fragment ions and only plots the c and z ions from ETD. The non-ergodic fragment products of ETD will give more accurate and correct assignments.

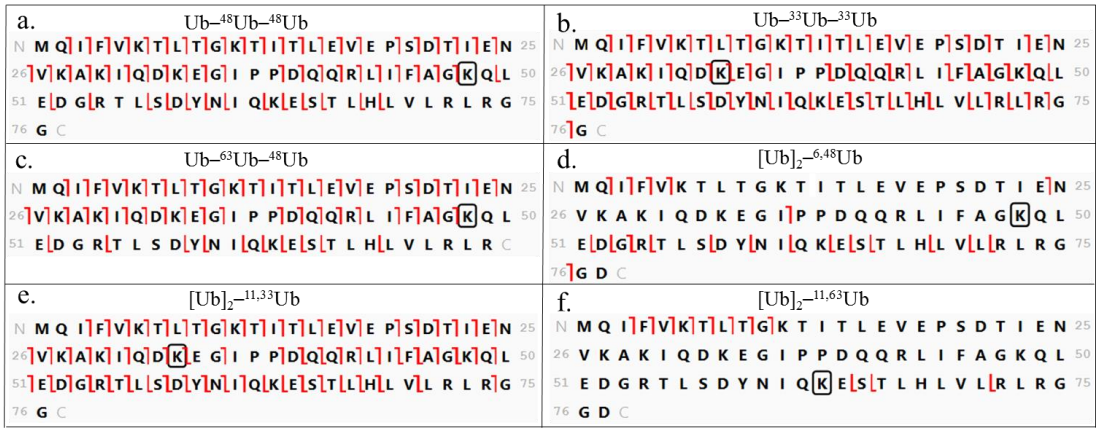



Figure 3.6. Initial matches (Step 2) against the monoubiquitin template of fragment ions from six tri-Ub chains. Both c/z ions (red) are plotted. The predicted isopeptide location is boxed in black.

Almost no information can be gathered about the linkage or topology using c ions, because all three N-termini of a native triUb chain can produce the same c ions. In this study unmutated distal Ubs will produce a redundant and indistinguishable set of c ions (Figure 3.6b and e) whereas the synthetic chains with mutations on the distal Ubs have c ions end at the first mutation. (Figure 3.6a, c, d, and f) All sequence variants are highlighted in Figure 3.4. However, even with mutations, the c ions that are present do serve to confirm the sample as a Ub.

Once the sequence is matched to Ub the question becomes: what is the topology of the chain? (Step 3) To answer this question the structural difference between the branched and the unbranched trimers must be understood. An important difference is shown in Figure 3.1 and 3.2. The unbranched trimer contains an “endo” Ub, a Ub which carries isopeptide bonds at both the C-terminus and on one of its Lys

residues, which the branched isomer does not. Diagnostic fragment ions (Figure 3.2 green) can prove the presence of an endo Ub, thus an unbranched topology, and by their absence a branched topology (Step 3, Figure 3.7). In this step a new template is established using ProSight Lite in which the mass of a proximal moiety is added at G76. ProSight Lite is then used to map fragment ions against the modified template. Ions that are unique to an endo Ub are formed by c and z fragments between K63 and the C-terminus that carry the mass of the distal/proximal Ub respectively. (Figure 3.7). If diagnostic c and z ions -- formed by amide bond cleavage between K63 and G76 (Boxed in Figure 3.7) which carry the mass of the distal (c ions) or proximal (z ions) Ub -- are observed the trimer is unbranched. (Figure 3.7a, b, and c) If c/z diagnostic ions are not observed, then the trimer is likely branched. (Figure 3.7 d, e, and f)

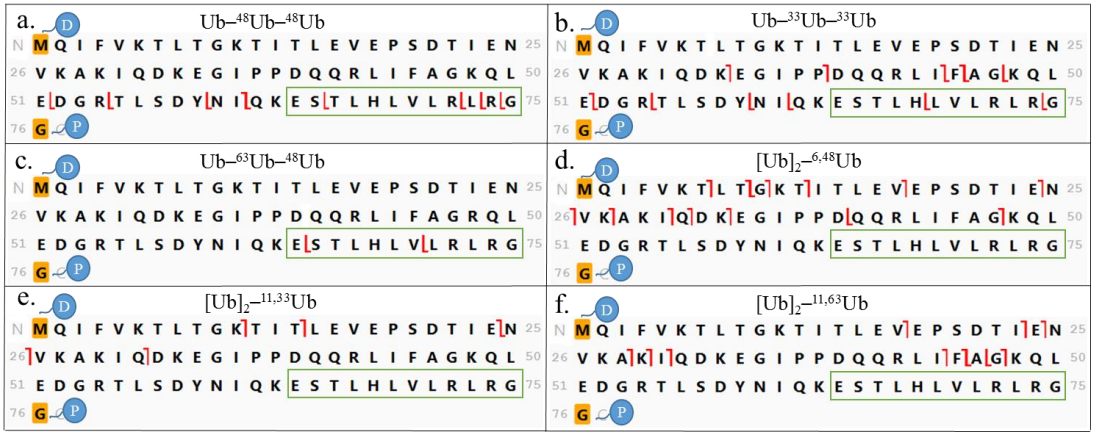



Figure 3.7. Visual representation of Step 3 in which fragmentation patterns of the endo ubiquitin are interrogated. Sites of trial additions of the proximal Ub are highlighted and have a ball and stick representation of the proximal Ub. For this step, if c or z ion fragments are seen in the green bracketed area, the chain is unbranched, and if there are no c or z ion fragments, it is branched.

After the topology of the trimer is characterized as unbranched or branched, the linkage locations are verified in Step 4 by inspection of the fragmentation pattern on a topologically correct trimeric template. A simple approach, which avoids tedious iterative addition of one or two Ub masses at each lysine in the proximal and then endo moieties, is to add the appropriate mass to the N-termini (M1) and trace amide bond cleavage to the point where the fragmentation stops (Figure 3.8 and 3.9). Since the topology is known, the mass addition indicated from Step 2 on the proximal Ub will be one (branched) or two Ub (unbranched). For example, in the case of the Ub–$^{48}$Ub–$^{48}$Ub unbranched chain, z fragmentation in the proximal moiety is observed to occur up to L50, indicating that the mass of diUb should be added to K48 (Figure

3.8a). The absence of contradictory ions will confirm this linkage. An analogous approach is then applied to the endo Ub in the Ub–$^{48}$Ub–$^{48}$Ub example. After adding the mass of a monoUb to the C-terminus of the endo template, formation of ions assigned as z is seen not to occur beyond D52. This indicates that the endo Ub is also modified at K48. For confirmation, if the distal monoUb mass is added to the N-terminus no change is observed in this z fragmentation pattern. This unique fragmentation is labeled in Appendix Figure 2 in the combined product ion mass spectrum. Despite the large masses of the modifications, this strategy is supported by ProSight Lite in a manner similar to the way that the mass increment of a classical –GG tag is handled by conventional bottom up software.

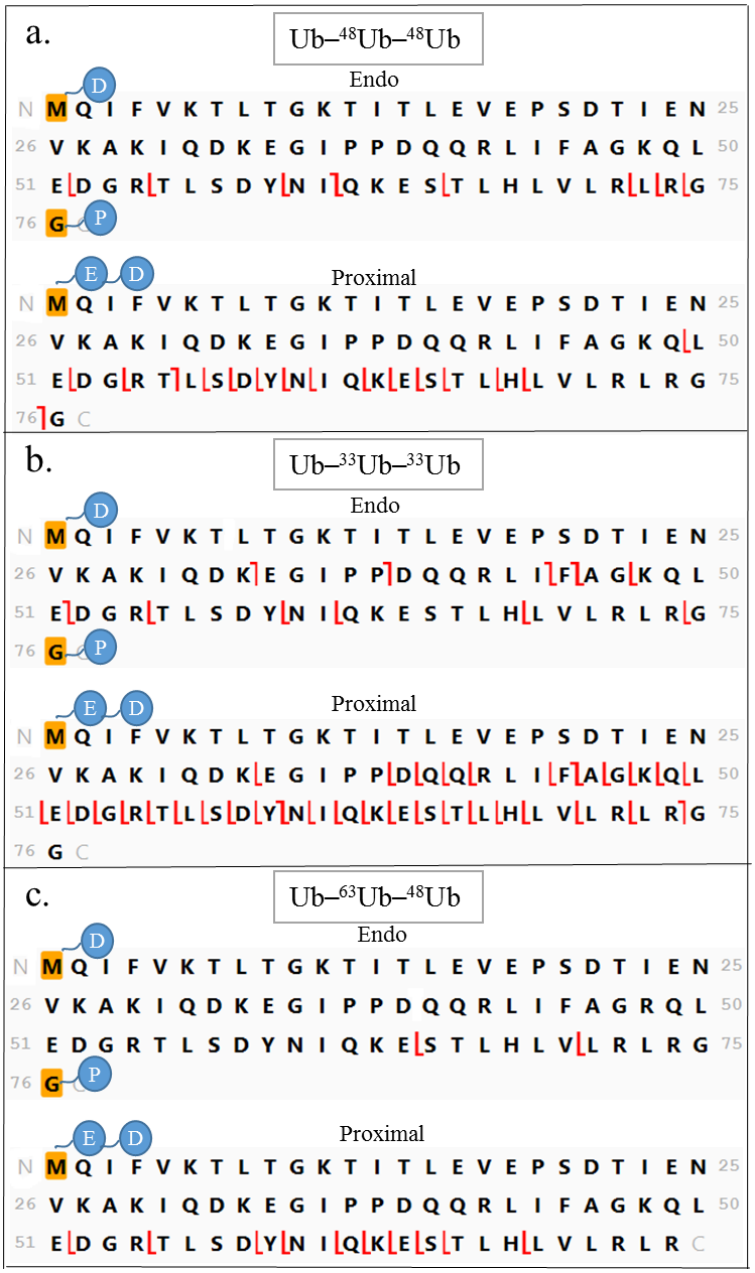



Figure 3.8. Visual representation of Step 4, linkage determination for the endo and proximal moieties in chain trimers. The endo (top) and proximal (bottom) sequences for each unbranched chain. Sites of trial addition are highlighted in the endo Ub for monoUb mass and in the proximal Ub for diUb.

If Step 3 indicates that the chain is branched, the linkage locations can be determined through a similar process. Now the proximal Ub is modified by two distal Ub moieties. Using the branched synthetic standard $[Ub]_2$–$^{11,33}$Ub as an example, Step 2 will already have shown a Ub addition at K33 (Figure 3.6e) due to the absence of z ions formed after E34. In a generalized approach, which should confirm this observation and find the remaining linkage, a template was constructed in which the mass of a monoUb is added to the N-terminus, just as in the unbranched chain determination, and the mass of another Ub is added to K63 (i.e. the closest linkage site to the C-terminus) (Figure 3.9b). The fragmentation pattern can then be used to determine where the c ions and z ions start and end. Again the fragmentation pattern shown in Figure 3.7b is consistent with the assignment of one linkage at K33, because in figure 3.7b, z ions are only observed before F46 and in Figure 3.9b c ions end at Q31. Confirming c ions are observed only after K11 in the sequence of the proximal ubiquitin, and the fragmentation pattern revealed by the new template localizes the second Ub addition at the amino terminus, at either K6 or K11.

Using this strategy the structure of the $[Ub]_2$–$^{6,48}$Ub trimer was unambiguously defined (Figure 3.9), while in the $[Ub]_2$–$^{11,63}$Ub trimer the top-down strategy localized the linkage site toward the N-terminus, but could not distinguish K6 and K11. In the case where complete fragmentation is reported, and in the linkage determining step, all fragments end before K11 in the ambiguous cases, then the assignment at K11 would be stronger. As it is reported here, fragmentation is not complete enough to defend the linkage location at K11 without any ambiguity. In the

two ambiguous cases reported here, middle-out analysis was used as a supplementary technique to assign the position of attachment (see section Middle-Out Strategy).

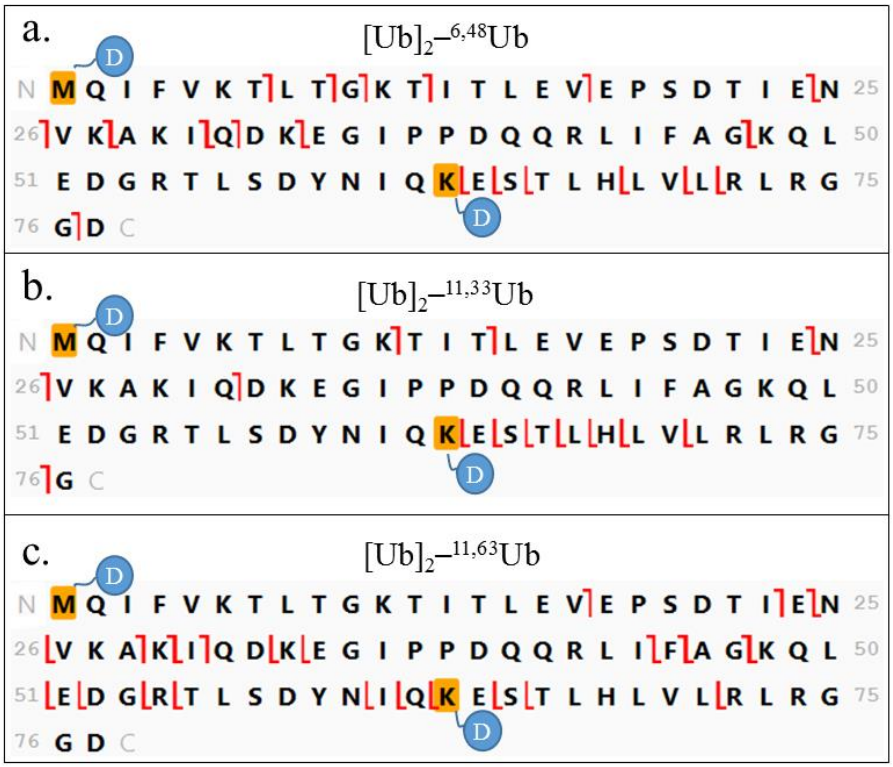



Figure 3.9. Visual representation of Step 4, linkage determination for the branched trimers. Linkage sites need to be determined only for the proximal Ub (shown). The site of trial addition is highlighted gold and distal moieties are represented with ball and stick cartoons.

After all linkage sites are confirmed, a final image can be put together (Step 5). Fragmentation density should be the highest when mapped against this final correct structure. Thus Step 5 provides final confirmation of the correct assignment. Final images are shown in Figure 3.10 for the six isomeric trimers studied.

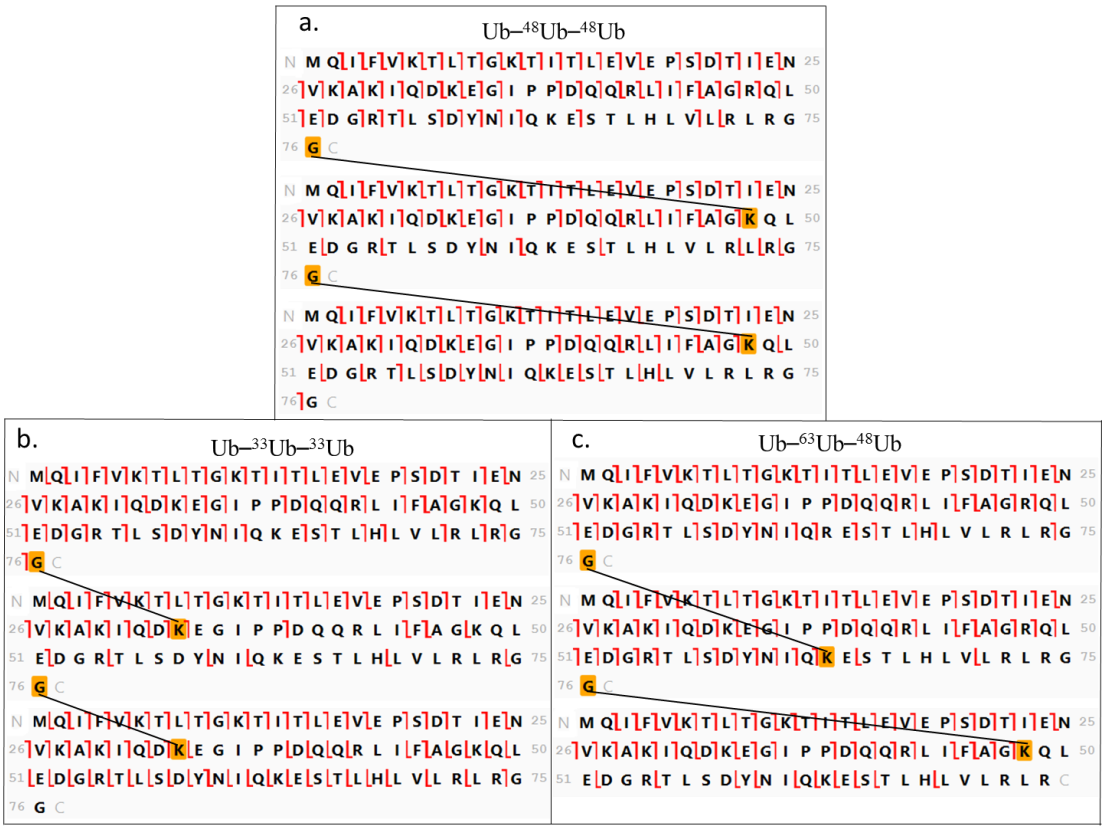



Figure 3.10. Visual representation of Step 5, the final images after complete characterization.

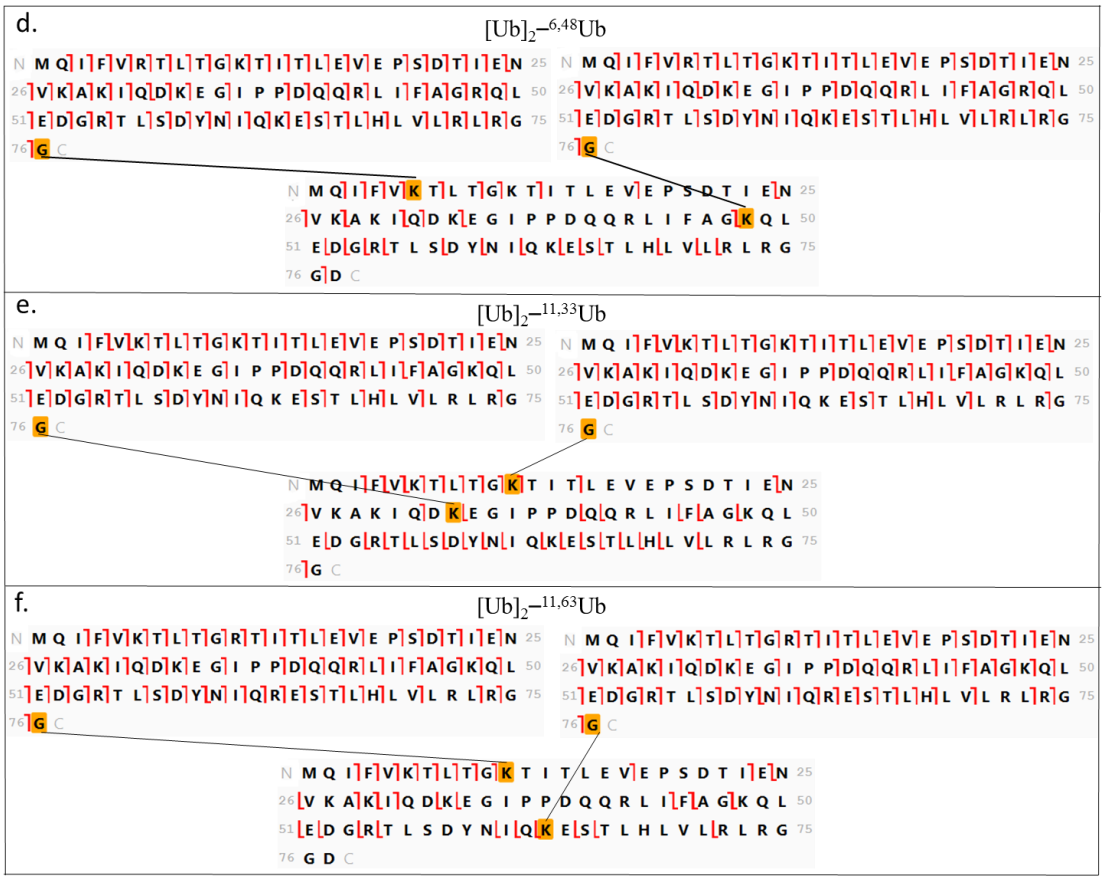



Figure 3.10 (*Continued*). Visual representation of Step 5, the final images after complete characterization.

Middle-Out Strategy

The unbranched and branched trimers can be analyzed by acid cleavage to confirm the identifications made by top-down analysis. Two of the unbranched topologies were assumed in the top-down analysis with only one diagnostic ion in the K63 to G76 region. To further prove the topology as unbranched (and branched), MWAC can be employed, which will truncate the chains and proved further fragmentation.

The ideal truncation that, like in the dimers, retains all the isopeptide linkages and C-termini, can be seen in Figure 3.3. The peptides produced are all cleaved on the distal Ub at D52 only. For the branched Ub chains, this means there are two moieties truncated, and for the unbranched, only one. (Figure 3.3) It was shown that, for the trimers to produce the ideal peptides it required 60 sec of digestion time at 140 $^{0}$C and 300 W power. (Appendix Table 3)

Extensive fragmentation is seen in all branched (Figure 3.11) and unbranched (Figure 3.12) middle-out peptides. The unbranched moieties will exhibit the same diagnostic ions that were used to elucidate the topology in the top-down analysis. These fragments are boxed in green in Figure 3.11. The middle-out results confirm the top-down work, with even more diagnostic ions per trimer. Fragmentation for the unbranched triUbs in top-down was extensive enough to confidently designate all linkages present.

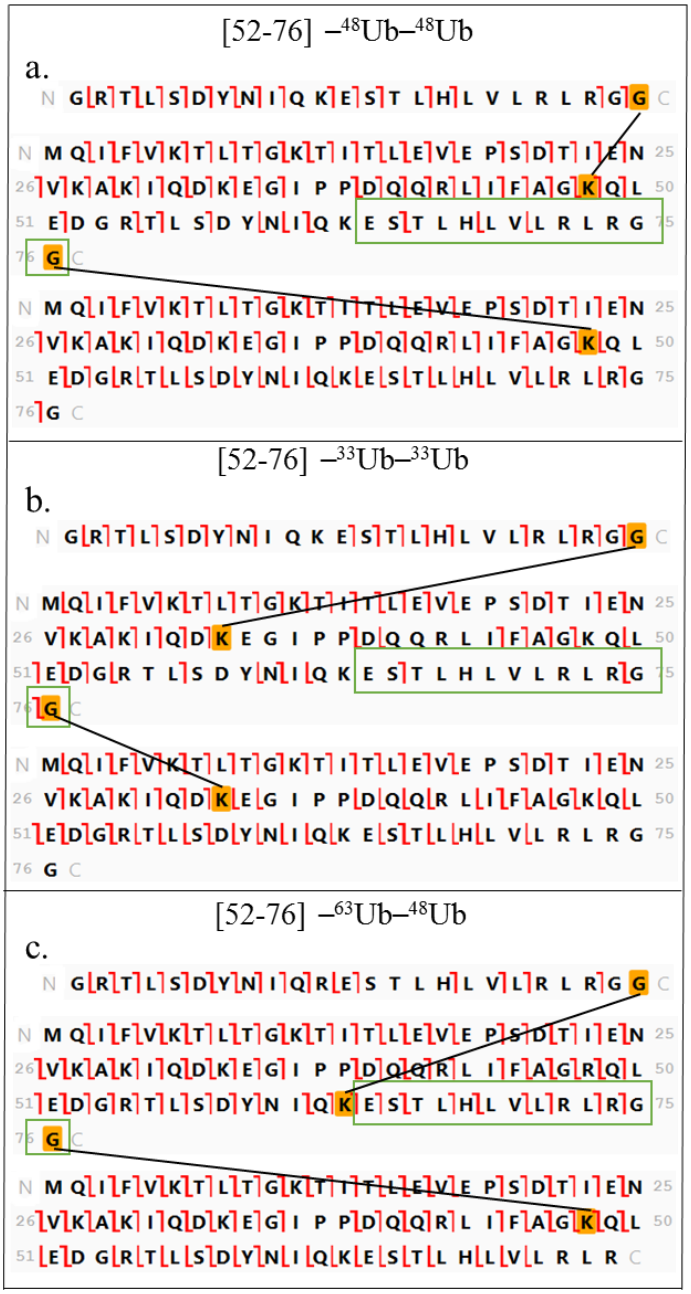



Figure 3.11: Final fragmentation pattern seen for straight chain trimers of ubiquitin after 60 sec time-controlled acetic acid hydrolysis to produce middle-down peptides. Boxed in green are the diagnostic ions proving linearity of the chain as with the top-down protocol.

MWAC could be used to resolve the sites of attachment in the branched trimers $[Ub]_2$–$^{11,33}$Ub and $[Ub]_2$–$^{11,63}$Ub. (Figure 3.12b. and c.) Truncation on only the distal Ubs at D52 produces a peptide that no longer produces duplicate c ions from the distal and proximal moieties. The c ions from the proximal moiety are unique in this middle-out peptide. Thus, the c and z ions seen in Figure 3.12b are unique to the proximal moiety and distinguish attachment at K11 from K6.


Figure 3.12: Final fragmentation pattern seen for peptides unique to branched trimers of Ub after time-controlled acetic acid hydrolysis to produce middle-down peptides.

*Summary*

Structures and functions of polyUbs are not well correlated yet, because polyUb modifications are difficult to decipher. This chapter suggests a robust strategy which provides that structural information. EThcD mass spectra of six synthetic ubiquitin trimers (multiply branched proteins with molecular masses exceeding 25600Da) were examined using an orbitrap Fusion Lumos instrument to determine how top-down mass spectrometry could be used to characterize the trimeric chain topology and linkage sites in a single, facile workflow. The efficacy of this method relies on the formation, detection, and interpretation of extensive fragmentation. In cases where fragmentation is not extensive enough, middle-out methods were employed and resulted in complete characterization of all six isomeric chains. With improvements in top-down instrumentation, fragmentation should not limit this workflow, making it applicable to future improvements in instrumentation and methods.

# Chapter 4: Mass Spectrometric Analysis of Ub Tetramers (Adapted from Ref 91)

*Introduction*

Tetrameric ubiquitin (tetraUb) is one of the most studied linkage lengths of the ubiquitinome. This is partially due to its potency in early proteolysis studies compared to other lengths[93]. K63 linked tetraUb has been shown to play a role in antiviral signaling[94] and in tumor necrosis factor (TNF) signaling[95]. However, studies are not well equipped to characterize the lengths of polyUbs present due to lack of methods to determine the lengths present *in vivo* above trimers in large proteomic studies.[77]

TetraUb chains can form 1240 different isomers. As with all Ub chain lengths, each isopeptide linkage can be attached at 1 of 7 different lysines or can be linked at the initial methionine (M1) and the chain can comprise homogeneous or mixed linkages. Without considering specific linkage sites there are four general topologies that a tetrameric chain can take (Figure 4.1). Each of the four topologies has a different number of total isomers possible based on isopeptide linkage location. The all unbranched has 512 possible linkage combinations, the branched α-endo has 224, the branched proximal has 448 (from the two iterations of the dimer and monomer attachments), and the all branched has 56 (totaling 1240 isomers).

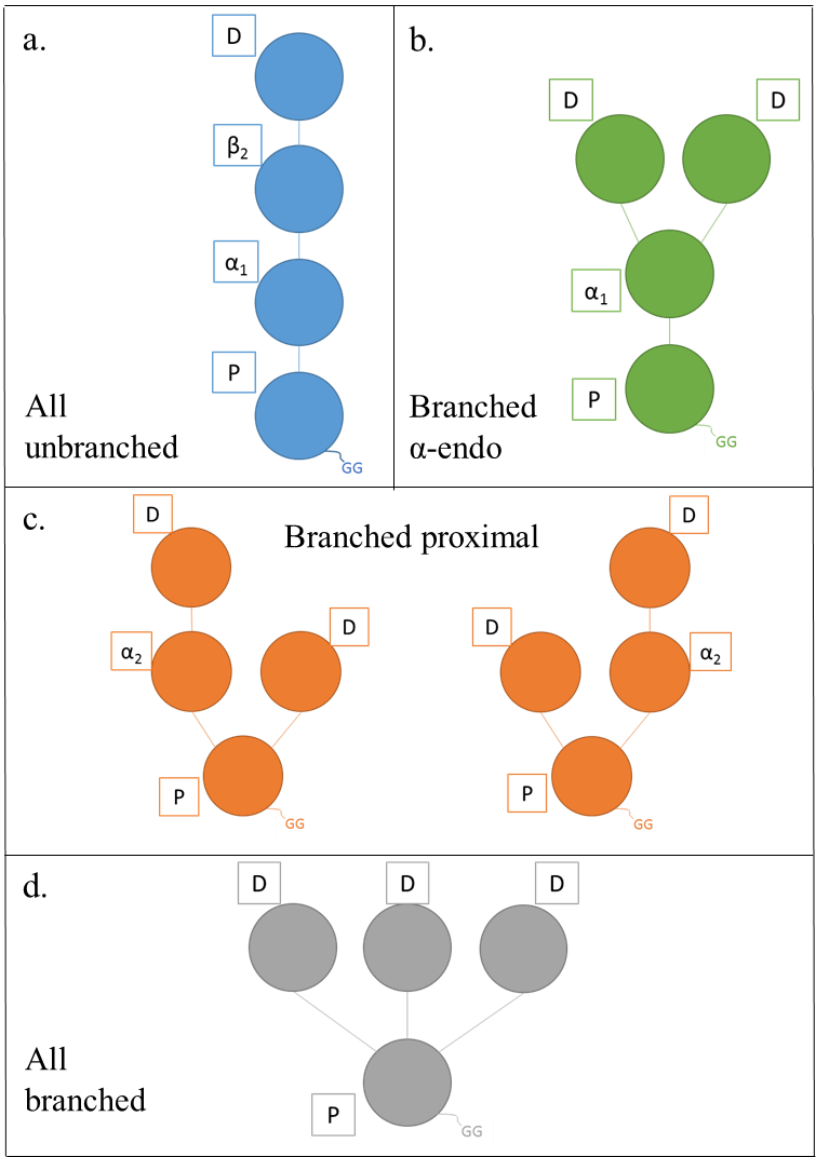



Figure 4.1. The four general topologies available to tetramers. P labels the proximal Ub, D the distal Ub, $\alpha_{1/2}$ is the α-endo Ub, and $\beta_2$ is the β-endo Ub.

To facilitate characterization by mass spectrometry the different moieties in the polymer must be distinguished. Traditionally the Ub with a free C-terminus is called the proximal Ub, designated as P in Figure 4.1. The Ub attached most distant

from the proximal Ub, or with no isopeptide linkage other than on its own C-terminus, is known as the distal (D) moiety. Intermediate Ubs are termed endo.[23,53] In this mass-based analysis of tetramers there are two different types of endo Ubs. The endo Ub attached by its C-terminus directly to a lysine on the proximal Ub is defined as an α-endo Ub. In unbranched tetramers the α-endo Ub has only the proximal mass attached to its C-terminus and is designated with a subscript 1 (i.e. $\alpha_1$-endo Ub). In the general case ($\alpha_n$-endo Ub) the subscript n designates the number of Ubs attached directly and indirectly to the C-terminus of the endo moiety in question. The endo Ub attached by its C-terminus directly to a lysine on the α-endo Ub is defined as a β-endo Ub. The masses of two or three Ub moieties are attached to the β-endo Ub's C-terminus in a tetramer. Thus it is designated with a subscript of 2 or higher (i.e. $\beta_2$-endo Ub). This nomenclature can be extended to higher polyubiquitins.

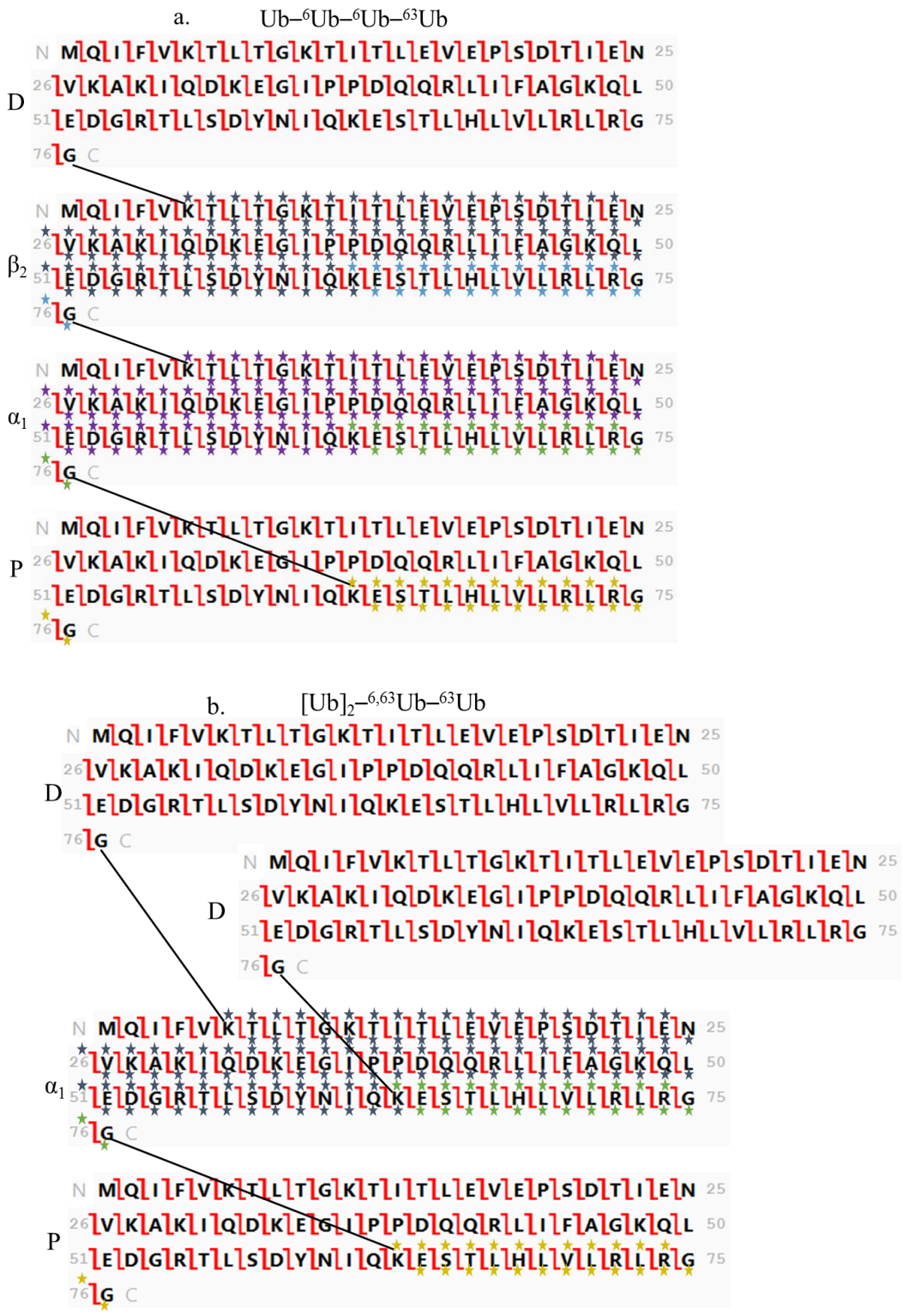



Figure 4.2. Three different topologies with the similar linkages are shown where stars highlight the unique fragments that can distinguish the topologies and linkages.

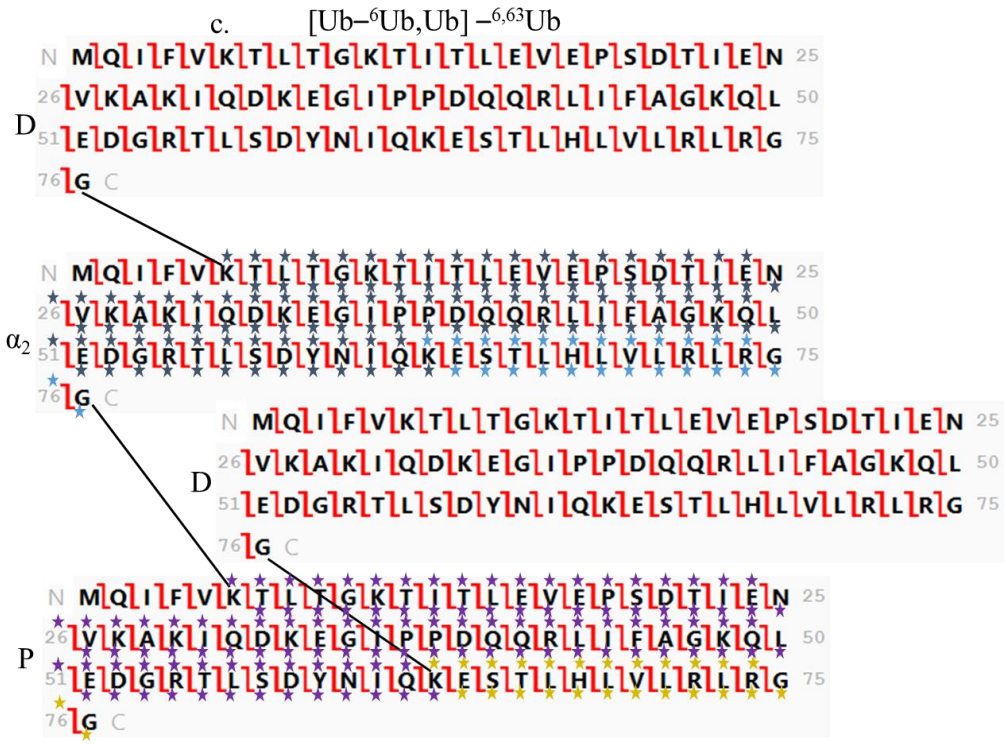



Figure 4.2 (*Continued*). Three different topologies with the similar linkages are shown where stars highlight the unique fragments that can distinguish the topologies and linkages.

In Figure 4.2, complete fragmentation patterns for three theoretical tetramers are shown with all the characteristically relevant fragment ions starred. All other fragments are shared with the distal Ubs and thus cannot be used for structural determination. Unique fragmentation is not unique between topologies (the same colored star fragments would represent the same masses in a mass spectrum), but instead is those fragments are unique to the Ub moiety within tetramer being

interrogated. For example, the fragments starred in gold are unique to the proximal Ub, but are not unique between tetramers; each proximal Ub shown has the same fragments starred in gold. (Figure 4.2a, b, c)

To determine the topology of the chain, diagnostic ions (green and light blue stars) must be used. (Figure 4.2) Though at least one of these groups of diagnostic ions is present in each example tetramer, the combination of diagnostic ions changes depending on the orientation of the moieties. For example the spectrum represented in Figure 4.2a has fragment masses matching the theoretical masses of both the green and light blue diagnostic ions, and so can be defined as the all unbranched topology (also seen in Figure 4.1a). Contrastingly, figure 4.2b has only masses matching the diagnostic ions for the $\alpha_1$-endo Ub, suggesting a different structure entirely (seen in Figure 4.1b). Thus, different topologies can be traced by the combination of the diagnostic ions in green and light blue present in one tetramer study. This is extremely similar to the diagnostic ion strategy from the trimer analysis, but with more variability in topology, and another set of diagnostic ions to consider.

Similarly to the trimer study, the ions starred in purple and dark blue map the linkage locations on different Ub moieties. These ions cannot be used to distinguish the topology of the chain nor can they alone tell which moiety they are present on. This is demonstrated by the fragments with purple stars, where, in Figure 4.2a the same mass would be mapped on the $\alpha_1$-endo Ub and on the proximal Ub of Figure 4.2c.The characterization strategy presented in this chapter must employ all of these fragments to completely characterize the tetraUb chains in simple comprehendible manner.

The objective of the study covered in this chapter is to develop a structured workflow for interpreting top-down mass spectra of unanchored tetraUbs to ascertain the topology and linkage sites, and to test and demonstrate this approach across all tetramer topologies. The strategy is tested on six synthetic standards whose chemical structures are shown in Figure 4.3. The strategy requires extensive fragmentation across the branched polypeptides, provided here by electron transfer dissociation on an orbitrap Fusion Lumos mass spectrometer.

*Methods and Materials*

**Synthesis of Ubiquitin Tetramers.** All ubiquitin tetramers were assembled from the respective recombinant Ub monomers using linkage-specific enzymes as described[23,45,85,88] or, in case of $[Ub]_{3-}^{6,27,48}Ub$, by combining this methodology with a nonenzymatic chain assembly approach[87] by members of the Fushman laboratory.

**LC-MS/MS.** Intact tetramers were diluted to 0.03 mg/mL in Solvent A (97.5% water, 2.5% ACN and 0.1% formic acid). The chromatography was performed using an Ultimate 3000 ultra-high performance liquid chromatograph (ThermoFisher Scientific, San Jose, CA) interfaced to a orbitrap Fusion Lumos Tribrid mass spectrometer (ThermoFisher Scientific). Five µL was injected, concentrated and desalted on a PepSwift Monolith trap (200 µm x 5 mm) for 5 min at 99% Solvent A before separation on a ProSwift RP-4H column (100 µm x 25 cm) (ThermoFisher Scientific) with a gradient of 30% to 50% solvent B (75% ACN, 25% water, 0.1% formic acid) over 15 min. Source fragmentation was set to 10%.

Precursor and fragment ion masses were acquired with a resolution of 120,000 at m/z 200 using "intact protein mode" with 1 mtorr ion routing multipole (IRM) pressure. The radio frequency of the C-trap was set to 30%. Data dependent MS/MS was carried in top-N mode with a precursor list of m/z values calculated for each tetramer. Isolated parents ions were fragmented using electron transfer dissociation supplemented with collisionally induced dissociation (ETciD) with a 3 msec ETD reaction time and supplemental activation at 10% normalized CID and averaging 20 μscans. Reaction time was lowered compared to the dimers and trimers to accommodate for the increase in mass. Lower reaction times are seen to improve fragmentation in higher massed proteins,[60] however this work saw little to no difference in spectra acquired with three or six msec.

**Processing the Spectra.** Precursor and fragment ions were deconvoluted using Xtract 3.0 (ThermoFisher Scientific). Fragment ions were matched against modified sequences of monoubiquitin using ProSight Lite[75] with a mass tolerance equal to or less than 4 ppm. ProSight Lite classifies fragment ions as a,b,c,x,y and z and provides a probability for each modified structure based on fragmentation. Because our polymers contain structural redundancy an analysis of unique fragments was required and these were assigned manually.

*Results and Discussion*

Six tetramers were synthesized for this study, representing the different topologies seen in Figure 4.1. Three unbranched tetramers were studied (Figure 4.1a), Ub–$^{48}$Ub–$^{48}$Ub–$^{48}$Ub, Ub–$^{63}$Ub–$^{63}$Ub–$^{63}$Ub and Ub–$^{63}$Ub–$^{6}$Ub–$^{63}$Ub (with chain nomenclature as described in reference 23). A branched topology in which an $\alpha_1$-endo Ub carries two distal Ubs, $[Ub]_2$–$^{6,48}$Ub–$^{48}$Ub, is represented in Figure 4.1b. Tetramer [Ub–$^{63}$Ub][Ub]–$^{6,63}$Ub with two isopeptide linked lysines on the proximal Ub is shown in Figure 4.1c. In this case one branch comprises a distal moiety and the other branch contains an $\alpha_2$-endo Ub linked to another distal unit. Finally, a tetramer $[Ub]_3$–$^{6,27,48}$Ub with three distal Ubs linked to the proximal Ub was synthesized, shown in Figure 4.1d.

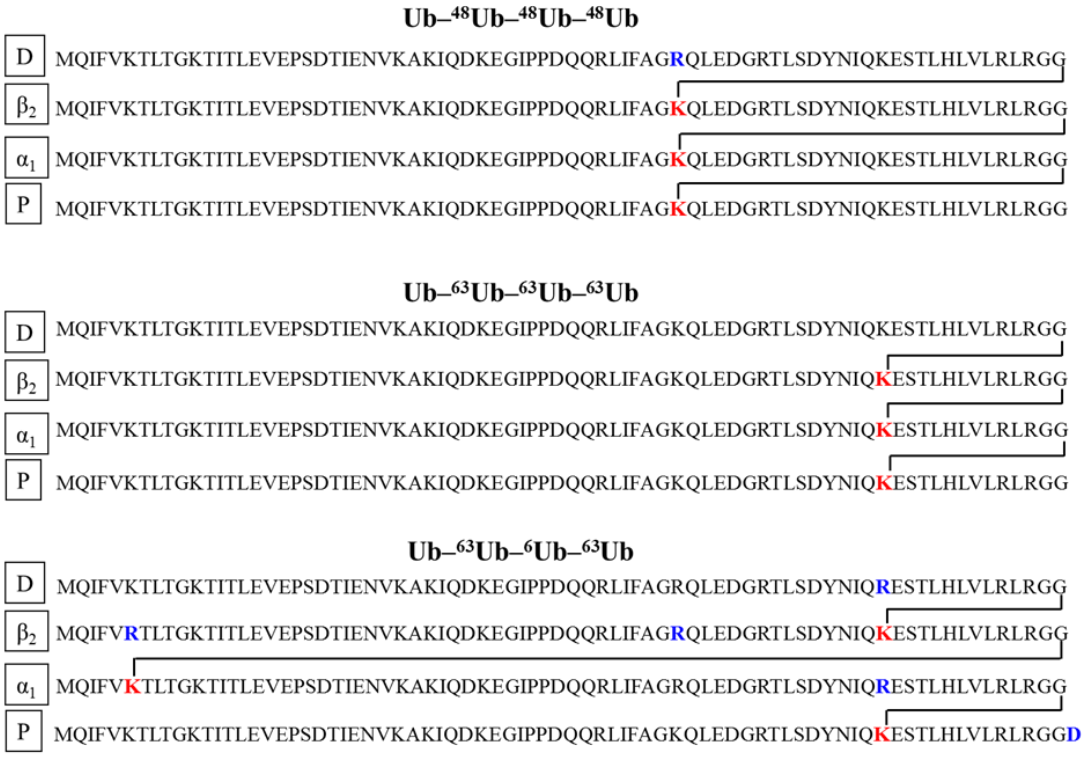



Figure 4.3. Sequence and connectivity of each unbranched tetramer. From top to bottom tetramers presented are Ub–$^{48}$Ub–$^{48}$Ub–$^{48}$Ub, Ub–$^{63}$Ub–$^{63}$Ub–$^{63}$Ub, and Ub–$^{63}$Ub–$^{6}$Ub–$^{63}$Ub. Residues in red represent the modified lysines. Residues in blue represent mutations made for the synthesis of the tetramer. Ubiquitin moieties are labeled with a D for distal Ub, $\alpha_{1 \text{ or } 2}$ for the α-endo Ubs attached to the proximal moiety, $\beta_2$ for the β-endo Ub once removed from the proximal moiety, and P for proximal Ub.

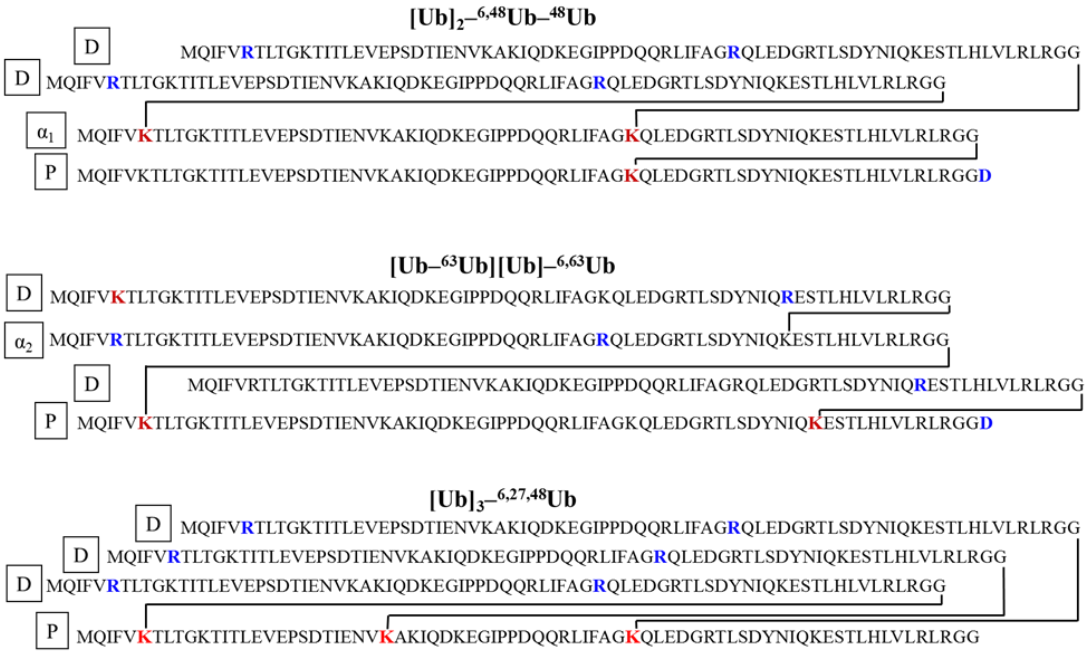



Figure 4.3 (*Continued*). Sequence and connectivity of each branched tetramer. From top to bottom tetramers presented are $[Ub]_2$–$^{6,48}$Ub–$^{48}$Ub, [Ub–$^{63}$Ub][Ub]–$^{6,63}$Ub, and $[Ub]_3$–$^{6,27,48}$Ub. Residues in red represent the modified lysines. Residues in blue represent mutations made for the synthesis of the tetramer. Ubiquitin moieties are labeled with a D for distal Ub, $\alpha_{1 \text{ or } 2}$ for the α-endo Ubs attached to the proximal moiety, and P for proximal Ub.

The workflow presented here (Figure 4.4) extends the strategy used for triUbs,[53] in which the monoubiquitin sequence is used as a template to assess the fragmentation patterns of each of the Ub moieties present in the tetramer. ProSight Lite allows for custom mass additions to any amino acid in the template sequence,[75] and masses equivalent to one, two or three Ub moieties are added and tested. The

resulting fragmentation patterns are used to assign the topology of each tetramer. Specific linkage sites are assigned by inspection of fragmentation patterns assigned to each monoubiquitin template. Only c and z ions are considered on the templates, because b and y ions were found to introduce uninformative complexity, for example, miss-matching assignments corresponding to internal fragment ions.[44,61]

Ubiquitin tetramers can present four different topologies, compared to two topologies for trimers. Thus an extra step has been incorporated to assign the tetraUb topology. In the case of trimers, the difference between branched and unbranched isomers is revealed by the presence or absence of an endo ubiquitin. In the case of the tetramer, three of the four potential topologies have one or two endo moieties, which again provide a path for differentiation. Figure 4.1 shows the four different topologies.

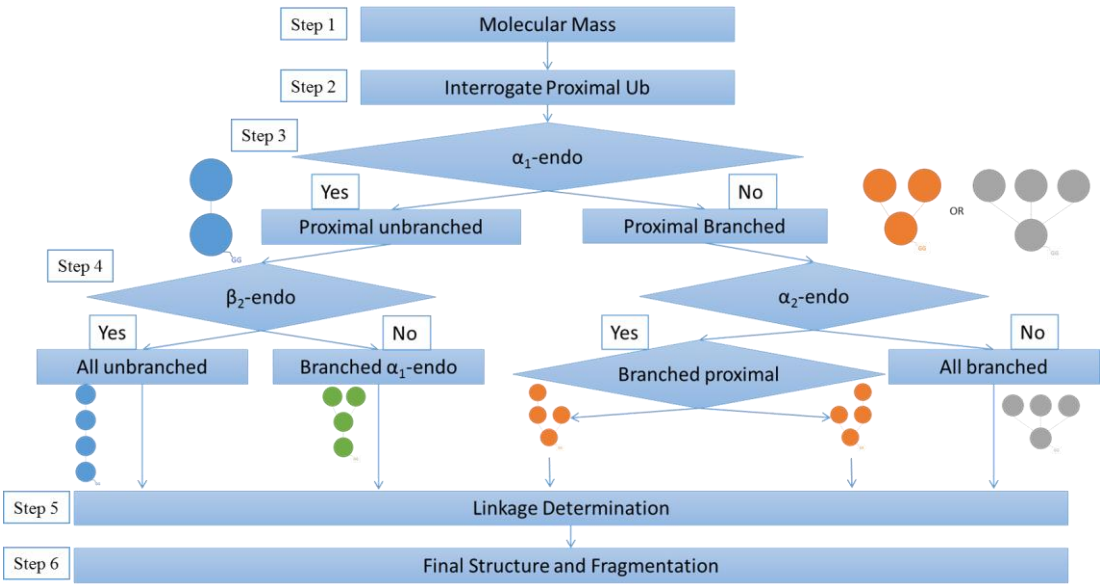



Figure 4.4. Simplified workflow developed for interrogation of the topology and linkages present in ubiquitin tetramers.

This workflow is specific to tetrameric Ub chain lengths, and is slightly different from the top-down workflows reported previously for ubiquitin dimers and trimers. Therefore Step 1, molecular mass determination (Figure 4.4), is necessary before moving on to any subsequent step. The intact masses of our standard compounds were all confirmed as tetramers within 2 ppm mass error (Table 4.1). In Step 2 the spectrum is mapped onto a proximal moiety to confirm it as the spectrum of an unanchored polyUb with a free C-terminus. The proximal Ub can also be interrogated in this step to recognize the attachment site closest to the C-terminus. Steps 3 and 4 then test for $\alpha_1$-endo or $\alpha_2$-endo and $\beta_2$-endo subunits based on diagnostic ions. Once the presence/absence and nature of any endo Ubs are known, a general topological shape can be assigned. Once the full topology is determined the

linkage sites are identified by inspection of the fragmentation patterns assigned to templates comprising each subunit with appropriate modifications (Step 5). In Step 6 the deduced structure is confirmed as that with the highest number of unique fragments.

Table 4.1. The theoretical and experimental masses for each of the six isomeric tetramers is shown.

| Tetramer | Calc. Intact Mass (Da) | Exp. Intact Mass (Da) | Mass Diff. (ppm) |
|---|---|---|---|
| Ub-$^{48}$Ub-$^{48}$Ub-$^{48}$Ub | 34212.44 | 34212.47 | 0.9 |
| Ub-$^{63}$Ub-$^{63}$Ub-$^{63}$Ub | 34184.43 | 34184.47 | 1.2 |
| Ub-$^{63}$Ub-$^{6}$Ub-$^{63}$Ub | 34411.48 | 34411.55 | 2 |
| [Ub]$_2$-$^{6,48}$Ub-$^{48}$Ub | 34411.48 | 34411.48 | 0 |
| [Ub-$^{63}$Ub,Ub]-$^{6,63}$Ub | 34411.48 | 34411.52 | 1.1 |
| [Ub]$_3$-$^{6,27,48}$Ub | 34352.47 | 34352.53 | 1.7 |

Proximal moieties with free carboxyl groups are used as templates in Step 2. As illustrated in Figure 4.5, fragmentation was mapped with high density for all our synthetic tetramers. This confirms each sample as a polyubiquitin. This fragmentation pattern also provides information on the site of the mass addition (linkage site) closest to the C-terminus just as in the previous chapters. All the z ions mapped can be formed only by cleavage in the unmodified portion of the proximal Ub. Based on the extent of formation of z ions back through the sequence, supported modified lysines are boxed in black. Because the distal Ub can produce c ions by cleavage near its C-

terminus, any c ions seen here in this initial match could be redundant and cannot be used for sequence or topological shape information. Subsequent steps are needed to determine the overall topology and the number of mass additions on the proximal Ub.

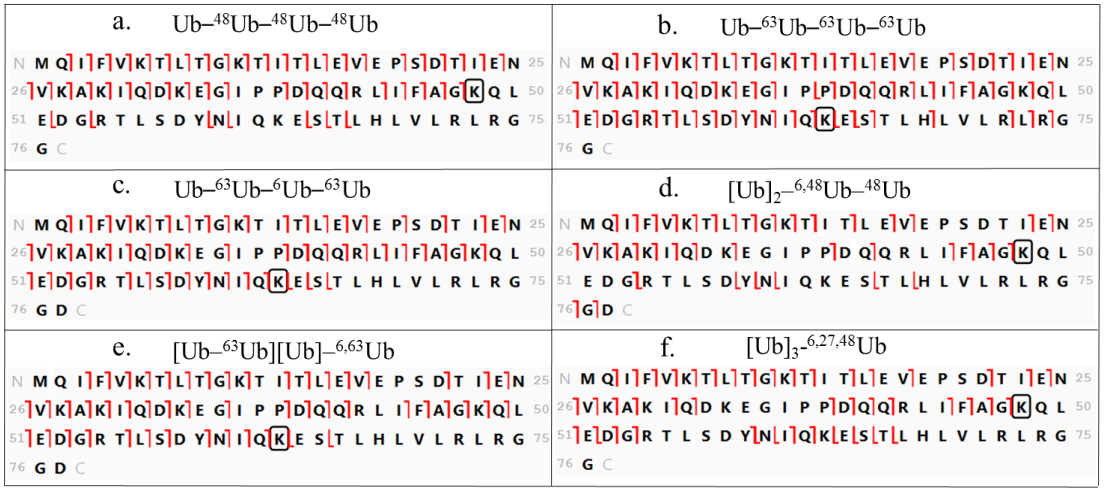



Figure 4.5. Matches (Step 2) of fragment ions in spectra of the proximal moiety of the six ubiquitin tetramers against monoubiquitin templates. Structures are indicated in each panel.

Step 3 is the first step in determining the topology. In this step, diagnostic ions are sought that are unique to any $\alpha_1$-endo subunits present. Such diagnostic ions are mapped to the region between K63 and the C-terminus when the mass of a single Ub has been added to the C-terminus and two Ub masses have been added to the N-terminus as shown in Figure 4.6. The diagnostic ions are boxed in Figure 4.6. Ions formed by cleavage between K63 and G76 will be mapped only when the C-terminus

is bonded to another Ub moiety. If there is branching on the proximal Ub, no fragment ions will be mapped in this region, indicating that the subunit defined as $\alpha_1$-endo Ub is not present (Figure 4.6e and f). Ions formed in any other part of the sequence cannot differentiate if the isopeptide linkage is on the C-terminus or on a lysine. (Figure 4.2) The structures that do not have the $\alpha_1$-endo Ub are the all branched $[Ub]_3$–$^{6,27,48}$Ub (Figure 4.6f) and the branched proximal [Ub–$^{63}$Ub][Ub]–$^{6,63}$Ub as expected. (Figure 4.6e)

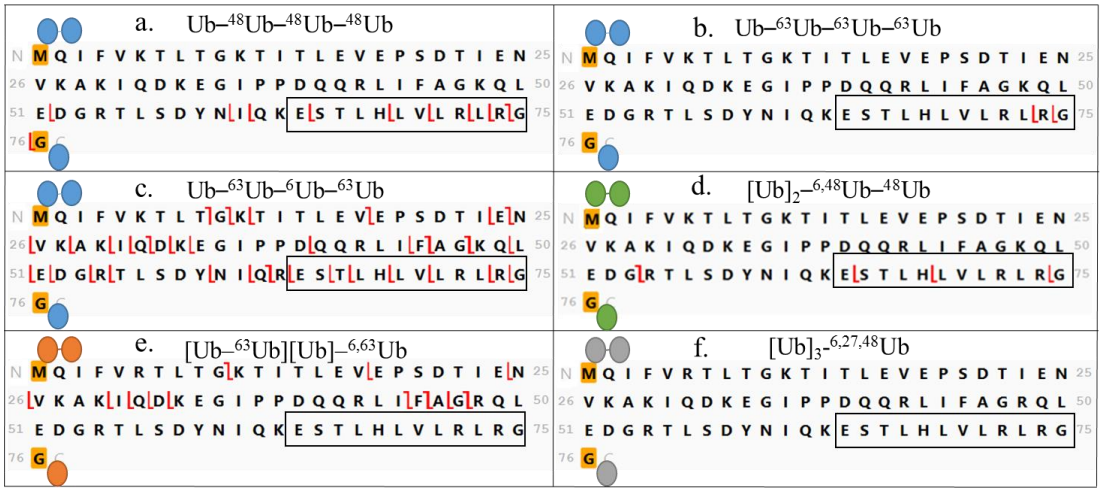



Figure 4.6. Visual representation of Step 3 in which fragment ions characteristic of an $\alpha_1$-endo moiety are sought. Sites of trial additions of masses of one and two Ubs are highlighted, with addition of a single Ub mass at G76 and the mass of two Ubs at M1. At this step, if any ions are confirmed in the boxed 64-76 region, the presence of a $\alpha_1$-endo Ub subunit is confirmed.

Step 4 is the final step needed to determine the topology of an unknown tetramer and it is applied to either group of tetramers defined by Step 3, tetramers with or without an α-endo subunit. This step is similar to Step 3, but now the masses added to the C- and N-terminus are switched (Figure 4.7). This defines $\alpha_2$-endo and $\beta_2$-endo Ub, because they have two Ub masses added to the C-terminus. By looking only for fragments mapped between K63 and the C-terminus, we eliminate contributions from isomers with different connectivity. All the tetramers expected to have a $\alpha_2$-endo or $\beta_2$-endo Ub have the diagnostic ions (Figure 4.7a, b, c, and e), and these ions are missing in the spectra of standards that do not contain either of these subunits (Figure 4.7d and f). The correct diagnostic ions are detected in spectra of all the synthetic standards and topological classifications are achieved. Knowledge of the topological shape is necessary to move on to determining linkage sites since each topology requires trial modifications at different positions to map fragment ions (Step 5).

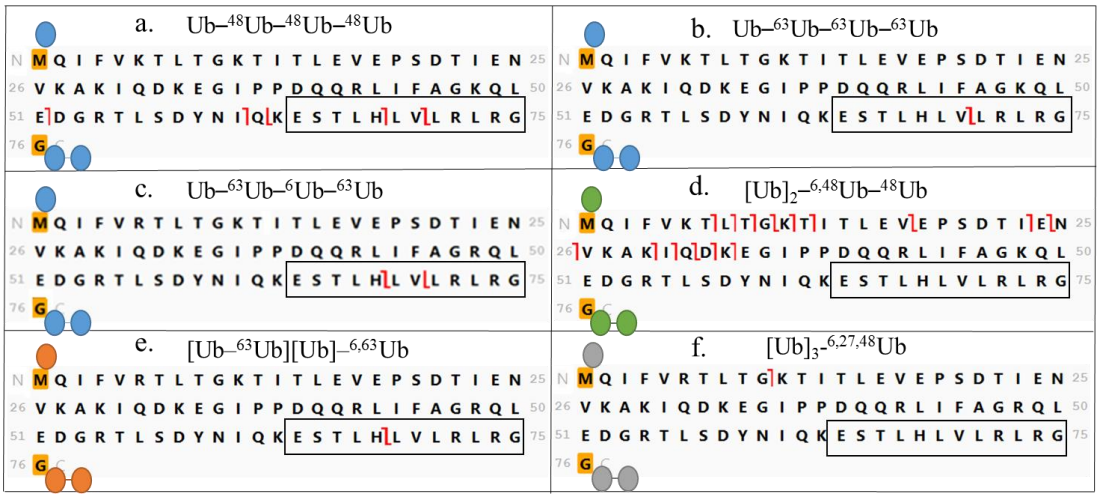



Figure 4.7. Visual representation of Step 4 in which fragment ions characteristic of $\alpha_2$-endo or $\beta_2$-endo moieties are sought. Sites of trial additions of masses of one and two Ubs are highlighted, with addition of a single Ub mass (a circle) at M1 and the mass of two Ubs (two circles) at G76. The presence of either subunit is confirmed if c or z ions are found to be formed by fragmentation within the boxed region, 64-76.

After the topology of the tetramer is characterized, information can be sought about the linkage sites. One approach is to reiteratively add Ub masses at each lysine in the template and accept the isomer that maps the most unique fragment ions. Uniqueness was determined manually in this study. In a more general approach, the appropriate Ub masses are added to the N- and C-termini of each subunit (Figures 4.8-4.11) and the fragmentation mapped on the correct template. Templates of each subunit (carrying trial modifications dictated by topology) in each of the three unbranched tetramers are shown in Figure 4.8. In each panel, fragmentation can be seen to proceed from the carboxyl terminus back toward, but not beyond, the first

modified lysine encountered. The same pattern is observed in top-down MS/MS spectra of ubiquitin trimers.[53] As seen in Figure 4.8c this is particularly informative for the K6 linkage site in the $\alpha_1$-endo moiety of Ub–$^{63}$Ub–$^{6}$Ub–$^{63}$Ub. (Informative fragment ions are labeled for Ub–$^{63}$Ub–$^{6}$Ub–$^{63}$Ub in Appendix Figure 3)

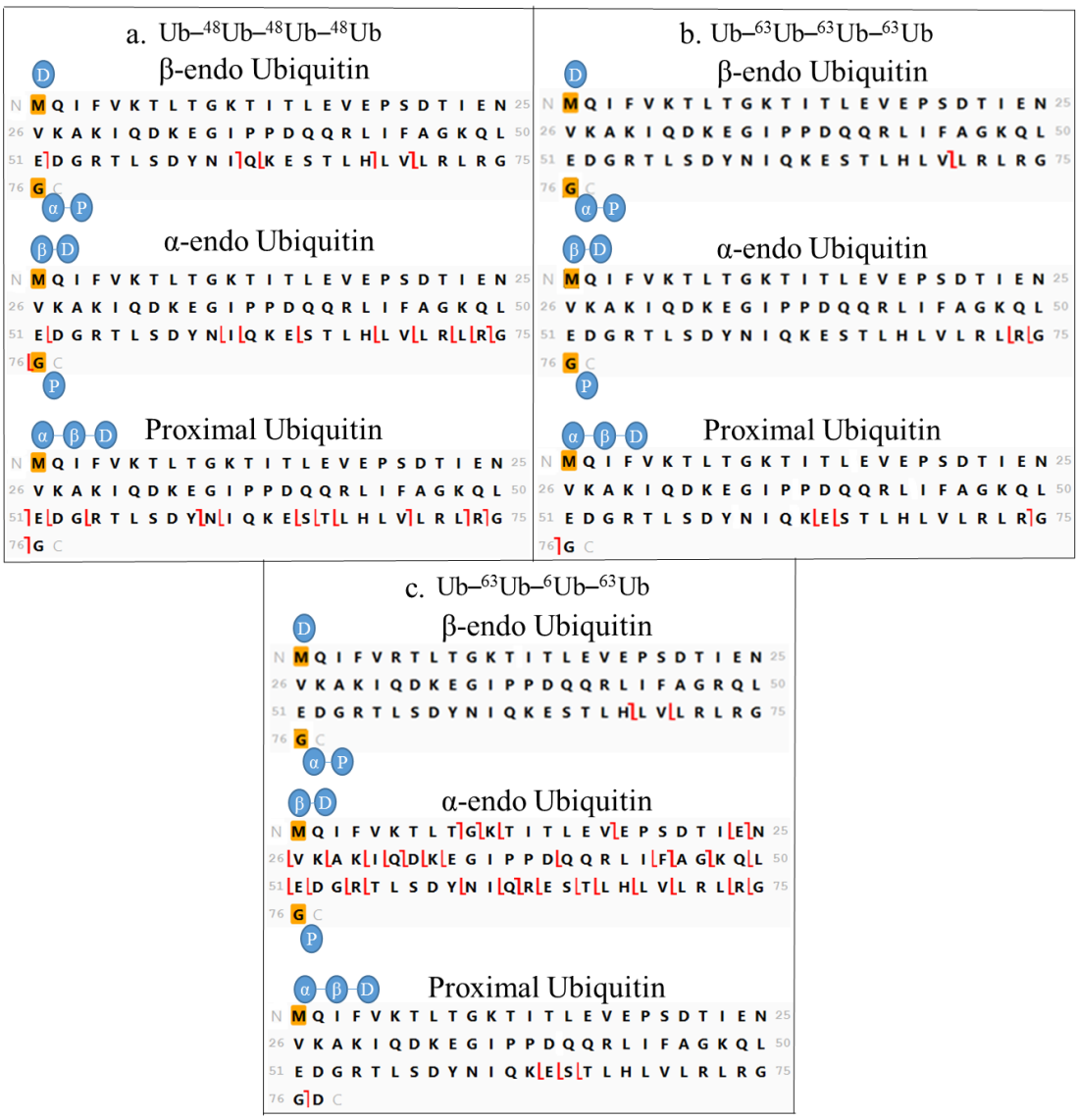



Figure 4.8. Linkage determination, Step 5, visualized for the unbranched tetramers. The tetramer and the subunits under review are identified in each panel. In the template for each subunit the sites of trial additions of the masses of one, two or three Ubs are highlighted in gold and the trial modifications are represented by circles.

In tetramers with branched $\alpha_1$-endo topology (Figure 4.1b), linkage sites must be determined in only two subunits, the proximal and α-endo moieties. Determination of the linkage site on the proximal Ub proceeds analogously to that of the proximal moieties in the all unbranched tetramers illustrated in Figure 4.8; the mass of three Ub moieties is added to the N-terminus and the template is inspected for z ion formation from the carboxyl terminus back into the chain. As seen in Figure 4.9b this fragmentation proceeds past K63 and terminates before K48. Thus the α-endo Ub is proposed to be attached to K48 in the proximal Ub. The α-endo Ub contains two linkage sites in this topology and must also be interrogated. Here the template is constructed (Figure 4.9a) by adding the mass of one distal Ub on the N-terminus, the mass of the proximal Ub on the C-terminus, and one Ub on K63 (the lysine closest to the C-terminus). Observation of a series of c ions formed by fragmentation beyond K6 but not in residues 1-5 locates one linkage site at K6. The combination of c ions and z ions localizes the second attachment to K48 or K63. In Figure 4.6d we can see that there is fragmentation up to K48 (c ion on G53 and z ion on R54) on the α-endo Ub, which eliminates K63 as a possible linkage assignment, definitively allocating the attachment to K48.

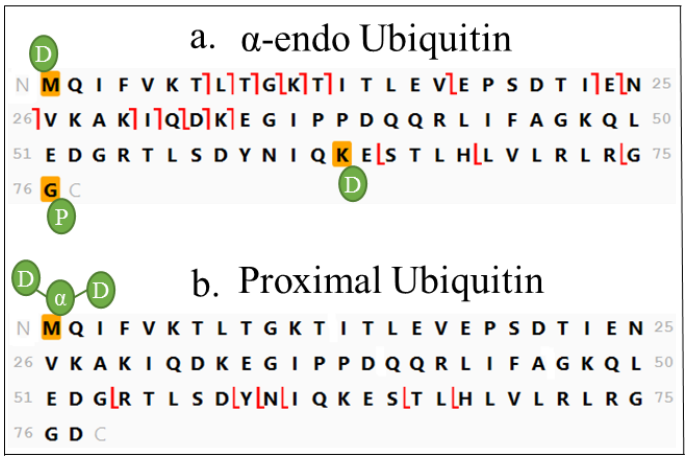



Figure 4.9. Linkage determination, Step 5, visualized for the $\alpha_1$-endo Ub tetramer $[Ub]_2$–$^{6,48}$Ub–$^{48}$Ub. Sites of trial additions are highlighted and the number of Ub masses added are indicated by circles.

In the tetramer topology group called branched proximal (Figure 4.1c) the proximal Ub carries two linkage sites. One of these comprises a monoUb moiety and the other a diUb moiety. Analysis of the proximal moiety in Step 2 (Figure 4.5c) clearly assigned one branch at K63. To discern whether the dimer or the monomer was linked at K63, dimer and monomer masses were added alternately to M1 or K63 and the template that allowed assignment of more unique fragment ions was chosen as the correct orientation. This comparison is illustrated in Figures 4.10a and 4.10b, where it can be seen that localization of the dimer mass at the N-terminus and the monomer at K63 provides more fragmentation. This template also allows K6 to be assigned as the attachment site for diUb, distinguished by c fragmentation at G10.

To initiate template-based analysis for the $\alpha_2$-endo Ub, the mass of two Ub is added to the C-terminus and the mass of the distal Ub is added to the N-terminus.

(Figure 4.10c) The fragmentation pattern supports linkage at K63, and this assignment is further confirmed in Step 6.

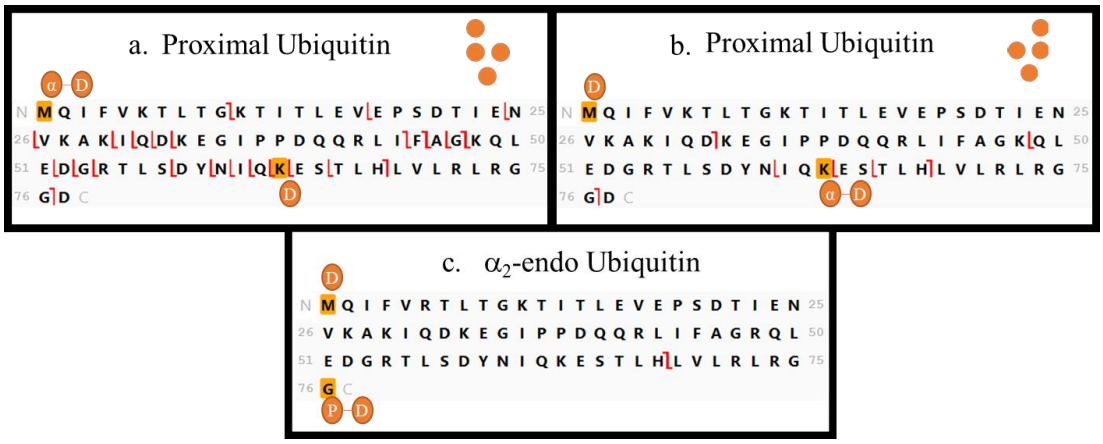



Figure 4.10. Linkage determination, Step 5, visualized for [Ub–$^{63}$Ub][Ub]–$^{6,63}$Ub. The two proximal templates represent two isomeric structures. Trial linkage sites are highlighted and the number of Ub masses added are represented by circles.

The last topological category to consider in interpreting the spectrum of an unknown ubiquitin tetramer is that in which the proximal Ub carries three monoubiquitin moieties attached at three different sites (called here the all branched tetramer). Examination of the template for the proximal moiety in Step 2 (Figure 4.5f) indicates that formation of z ions proceeds from G76 back past K63 to terminate before K48, indicating that K48 is a linkage site. Alternating addition of two Ubs and a single Ub moiety to the M1 and K63 distal Ubs (Figure 4.11a-b) provides support for K6 as a linkage site. No information is available for the third linkage site of the

synthetic standard (K27). The absence of fragmentation suitable for this last assignment is confirmed when the spectrum is mapped onto the correct structure in Figure 4.12f. When mapped, confirming fragmentation is seen for linkages at K48 (by Step 2 and Step 6) and K6 (by Step 5). The last moiety is limited to linkage at K11, K27, K29, or K33. This workflow is compatible with any future developments in activation that will provide the missing fragmentation.

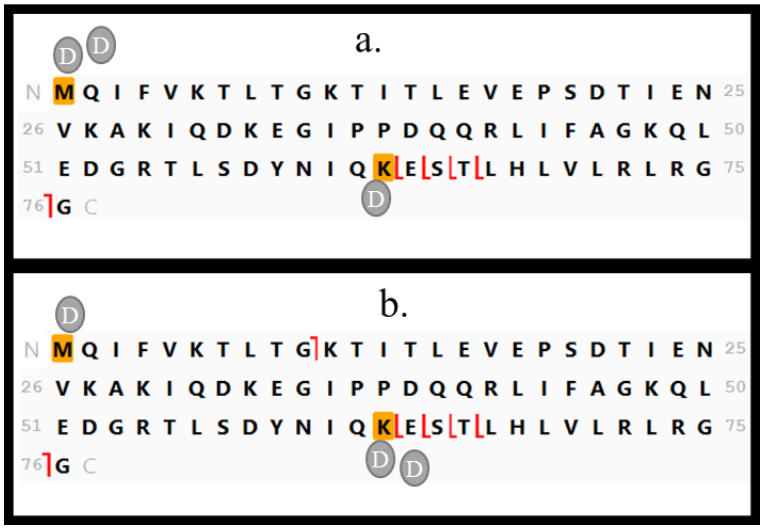



Figure 4.11. Template maps of fragmentation in the spectrum of $[Ub]_{3-}^{6,27,48}Ub$. a. and b. Trial attachment sites are highlighted and the number of ubiquitin masses added is shown as circles.

After the structure of each of the synthetic standards is characterized using the strategy outlined above, a final evaluation is made in Step 6. In Figure 4.12 the fragment ions recorded in the MS/MS spectrum are assigned to structures deduced

from the synthetic standards. These can be compared with patterns assigned to incorrect structures. The correct structure should have a high fragment density and the highest number of unique fragments. All structures shown in Figure 4.12 yielded more extensive unique fragmentation patterns than those of incorrect alternatives.


Figure 4.12. Visual representation of Step 6, in which fragmentation images are assigned to each of the final tetramer structures. The name of the tetramer is shown in the panel. Highlighted lysine and glycine residues are part of isopeptide linkages.

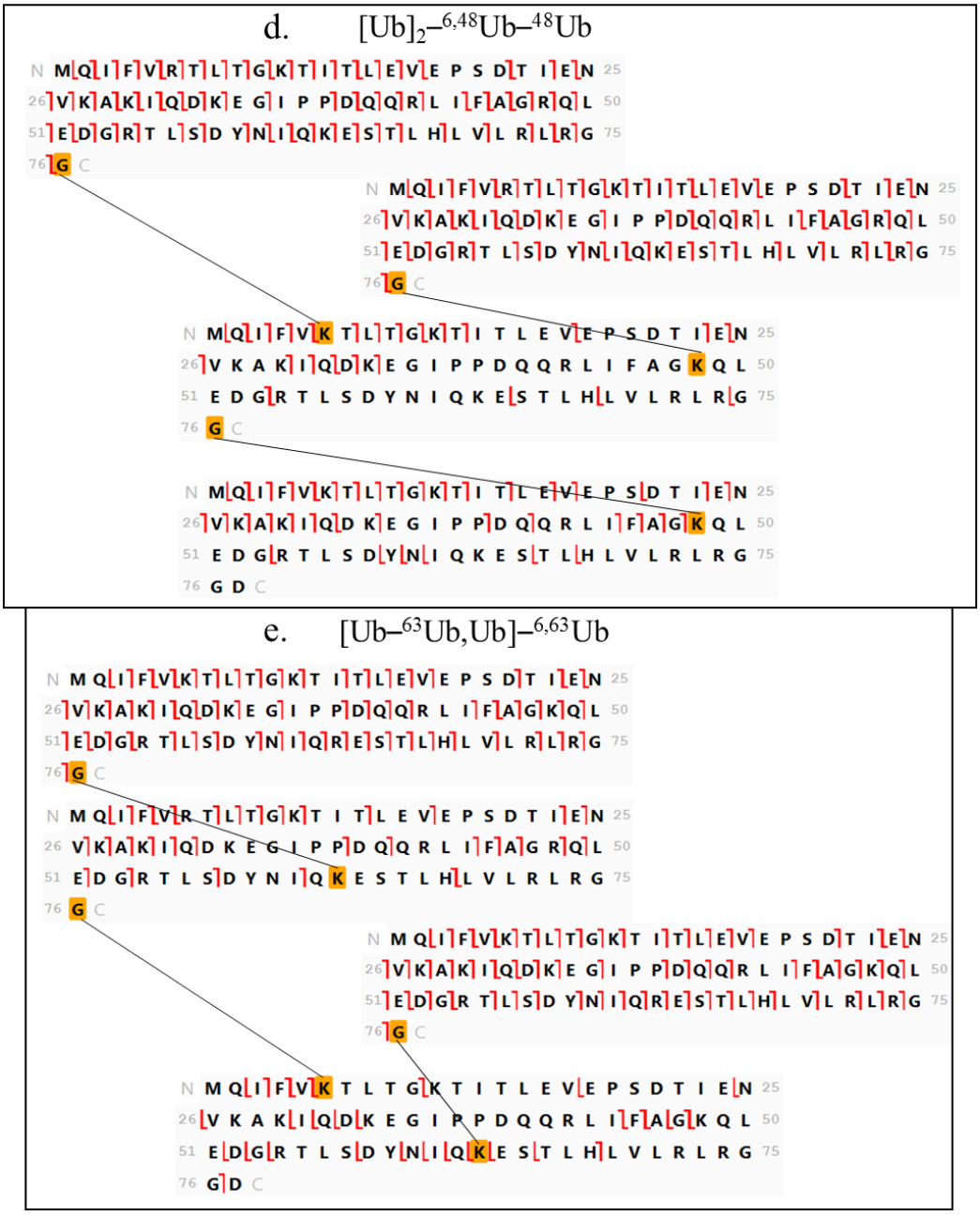



Figure 4.12 (*Continued*). Visual representation of Step 6, in which fragmentation images are assigned to each of the final tetramer structures. The name of the tetramer is shown in the panel. Highlighted lysine and glycine residues are part of isopeptide linkages.

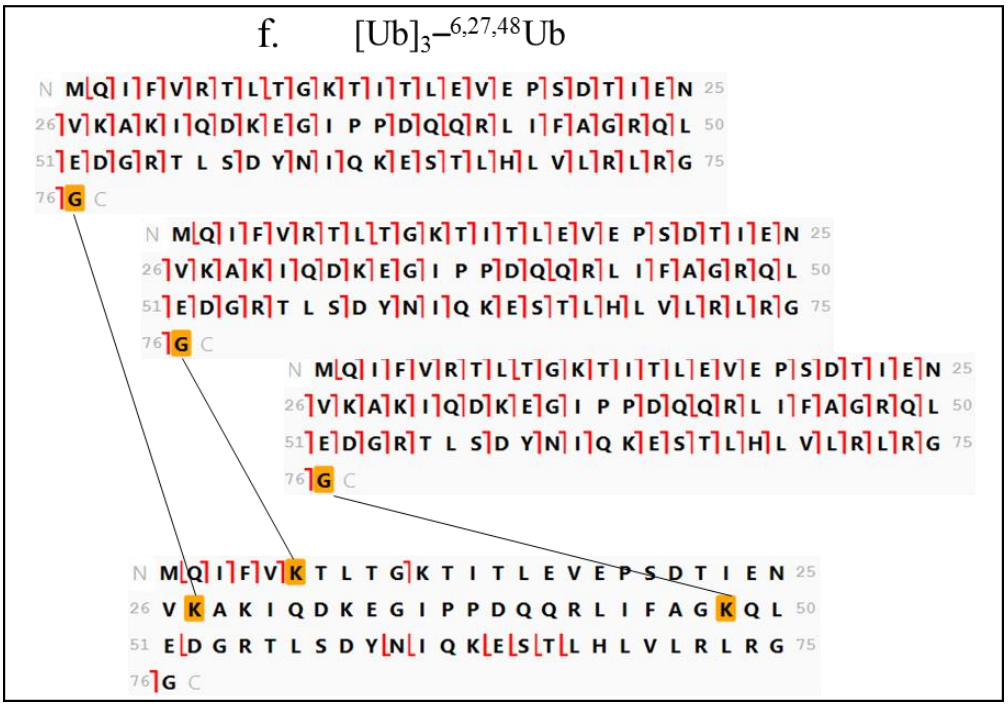



Figure 4.12 (*Continued*). Visual representation of Step 6, in which fragmentation images are assigned to each of the final tetramer structures. The name of the tetramer is shown in the panel. Highlighted lysine and glycine residues are part of isopeptide linkages.

*Summary*

The top-down mass spectrometric workflow presented in this study correctly classifies Ub tetramer chains into one of four possible topologies. Assignment of the topology permits fragmentation to be visualized on an appropriate template and to allow linkage sites to be identified on each ubiquitin moiety within the chain. The success of the approach to interpretation of the MS/MS spectra depends on achieving extensive fragmentation. Sufficient fragmentation was recorded using ETciD on an orbitrap Fusion Lumos to assign topology to all the standards analyzed and to assign linkage sites in three of the four topologic groups of unanchored ubiquitin tetramers. The most highly branched isomer fragmented most poorly. The workflow presented will be even more efficient as activation techniques continue to be developed which provide complete fragmentation of heavy proteins. This is a workflow that can also be readily modified for ubiquitin-like polymers (e.g. SUMO) and mixed polymers by adjusting the masses and the sequences added to the graphical interface.

# Chapter 5: Conclusions and Future Prospective

By using the most advanced instrumentation in top-down proteomics, this work has provided a novel and facile strategy for mass spectral characterization of all polyUb chains. The location of isopeptide linkages within, and the length of, polyUb chains have been experimental linked to a plethora of different cellular functions. To determine these features, researchers have developed a multitude of different methods, not all of which require mass spectrometry. These methods have not been successful in differentiation of all possible topologies and linkages of polyUb chains. The lack of a facile method within the community that can differentiate all polyUb isomers has led to the focus of this doctoral work.

PolyUb chains, made of the small protein ubiquitin (Ub), are found throughout eukaryotic cells. Their diverse functions have been experimental linked to the length, linkages, and topology of the chain. To divulge the relationship between the chain's features and the functions produced, labs can use only a few methods. The most prominent method has been to mutate the Ub chains at specific Lys residues, which disrupts chain formation at the mutated Lys, and report if the function being studied was effected. Another common method for polyUb characterization is to digest the chain in trypsin, resulting in peptides with a bioinformatics compatible search for the PTM by –GG tags at 114.043 Da. However, both these methods will only show linkages involved, not the length and only mildly can interrogate the

topology of the chain. Another method was developed and is commercially available which divulges the length, linkage location, and to a degree, the topology by using deubiquitinases DUBs (UbiCRest); yet it requires completely pure sample and multiple experiments involving multiple DUBs.

This work has produced a strategy which can be adapted to any fragmentation technique or mass analyzer that is powerful enough to produce viable results in top-down proteomics. Outlined in the previous pages are strategies for characterizing Ub dimers, trimers, and tetramers. Each workflow combined was able to differentiate a total of 1340 possible isomers (8 from dimers, 92 from the trimers, and 1240 from the tetramers). The general tenet of the workflow can be applied to any chain length. The addition of more moieties does not hinder the strategy, in fact, additional steps can simply be added to the workflows presented to create a scheme for any chain length.

In figure 5.1, a workflow is presented for pentamers, whose mass lies at the analytical limit of even the most advanced instrumentation in top-down proteomics. However, the workflow can be followed if the fragmentation is present. In an all unbranched pentamer linked homogenously at K48 (Figure 5.2), the instrumentation and methodology used for the tetramers is seen to completely characterize the pentaUb chain. All moieties have diagnostic ions used to characterize the unbranched topology (fragments seen between K63 and G76). Even at the analytical limit, the workflow can be applied and diagnostic ions can still be seen on a chromatographic time-scale.

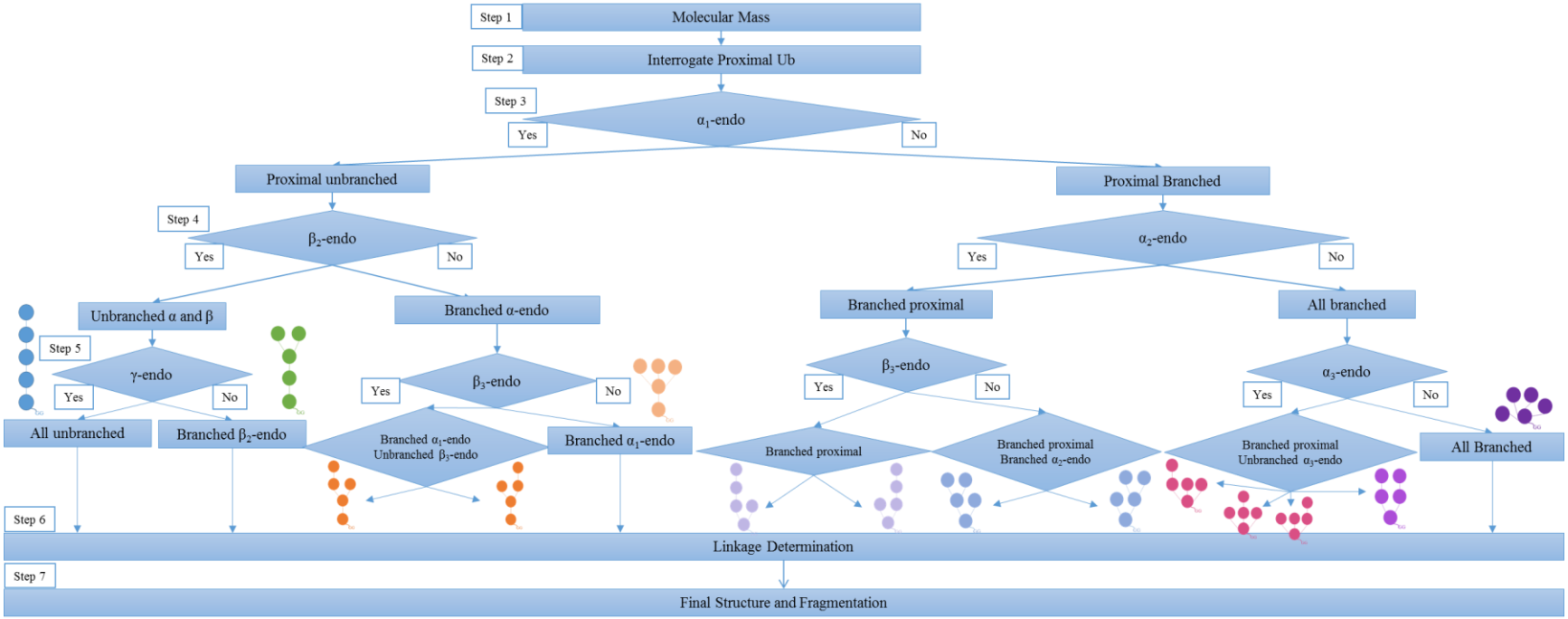



Figure 5.1. Workflow developed for the characterization of Ub pentamers.

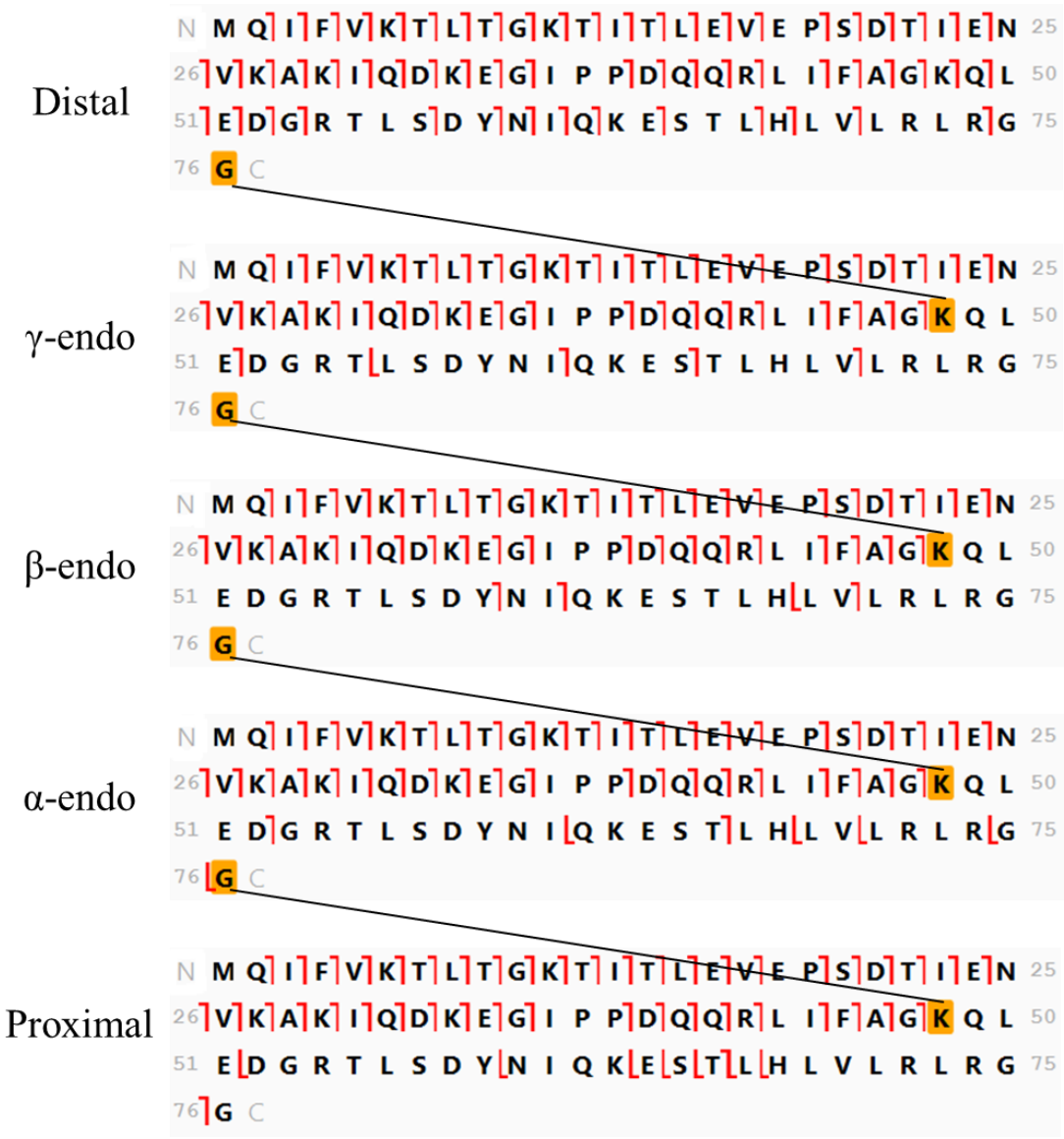



Figure 5.2: Final structure and fragmentation (Step 7) for Ub–$^{48}$Ub–$^{48}$Ub–$^{48}$Ub–$^{48}$Ub, the K48-linked unbranched Ub pentamer. Fragmentation and deconvolution was accomplished via the same process as the tetramers, however two parameters were adjusted. In this sample, 400 ng of pentamer were injected and the fragment ions were mapped with 0.1 Da error.

The method outlined in this work is applicable to any future advances in instrumentation and fragmentation. It could also be adapted to bioinformatics software to digitize this process. A workflow can be created for longer polyUbs along the same tenet as the examples outlined in this dissertation. At present there is no method which can so readily characterize polyUb chains in their entirety. Current research in cellular biology would greatly benefit from the advances proposed in this work. Definition of the ubiquitinome has the potential to outline and define the mechanisms at work in protein turn-over, DNA repair mechanisms, and eventually indicate drug targets as, Ub is involved in many vital cellular functions (and malfunctions).

# Appendices

Appendix Table 1. Sequences and the theoretical and observed monoisotopic masses of the fully truncated branched peptides analyzed from the seven all native K-linked diUbs.

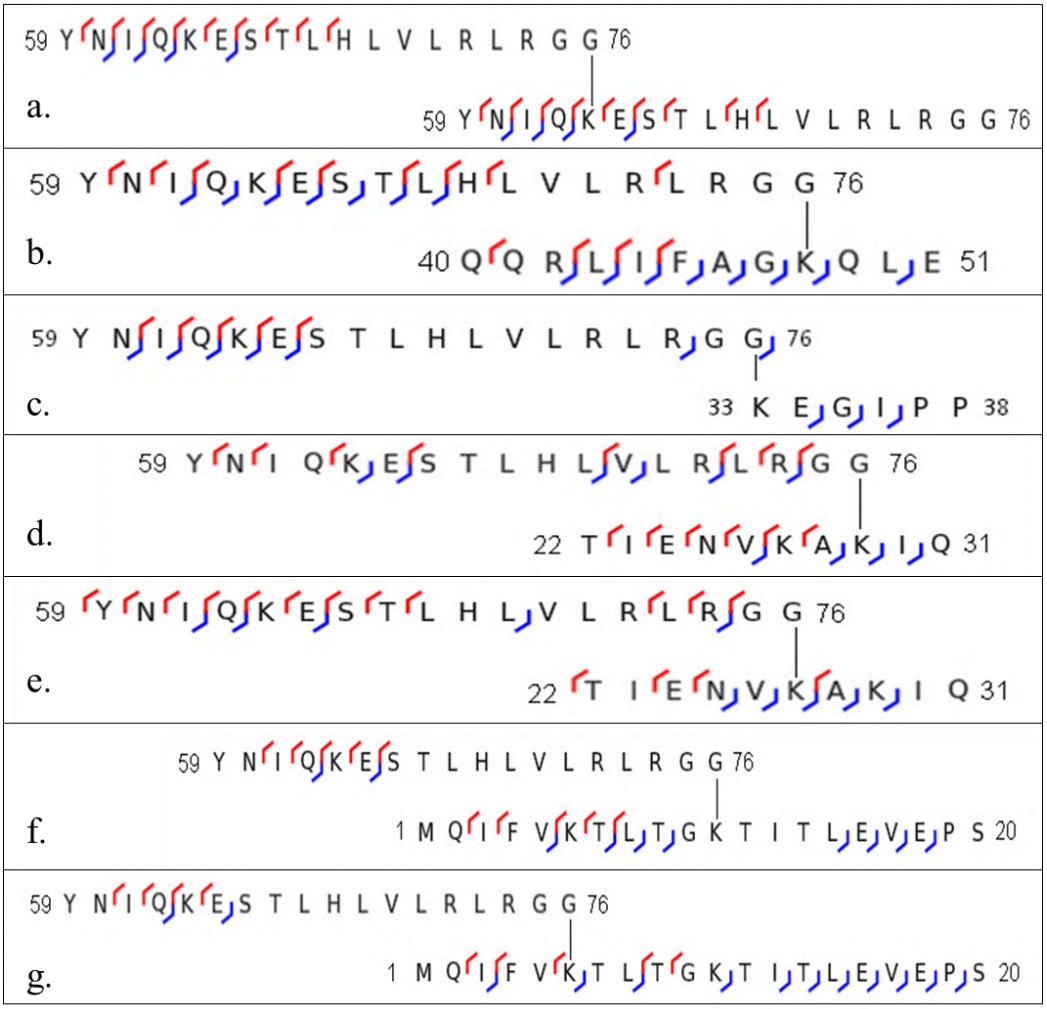
| Linkage | Sequences of the Fully Truncated Target Peptides | Mass (Da) | |
|---|---|---|---|
| | | Theor. | Obser. |
| K63 | YNIQKESTLHLVLRLRGG<br>\|<br>YNIQKESTLHLVLRLRGG | 4174.35 | 4174.36 |
| K48 | YNIQKESTLHLVLRLRGG<br>\|<br>QQRLIFAGKQLE | 3490.95 | 3490.96 |
| K33 | YNIQKESTLHLVLRLRGG<br>\|<br>KEGIPP | 2717.52 | 2717.53 |
| K29 | YNIQKESTLHLVLRLRGG<br>\|<br>TIENVKAKIQ | 3220.83 | 3220.85 |
| K27 | YNIQKESTLHLVLRLRGG<br>\|<br>TIENVKAKIQ | | |
| K11 | YNIQKESTLHLVLRLRGG<br>\|<br>MQIFVKTLTGKTITLEVEPS | 4312.39 | 4312.41 |
| K6 | YNIQKESTLHLVLRLRGG<br>\|<br>MQIFVKTLTGKTITLEVEPS | | |

Appendix Table 2. Optimization of fragmentation parameters for top-down analysis.

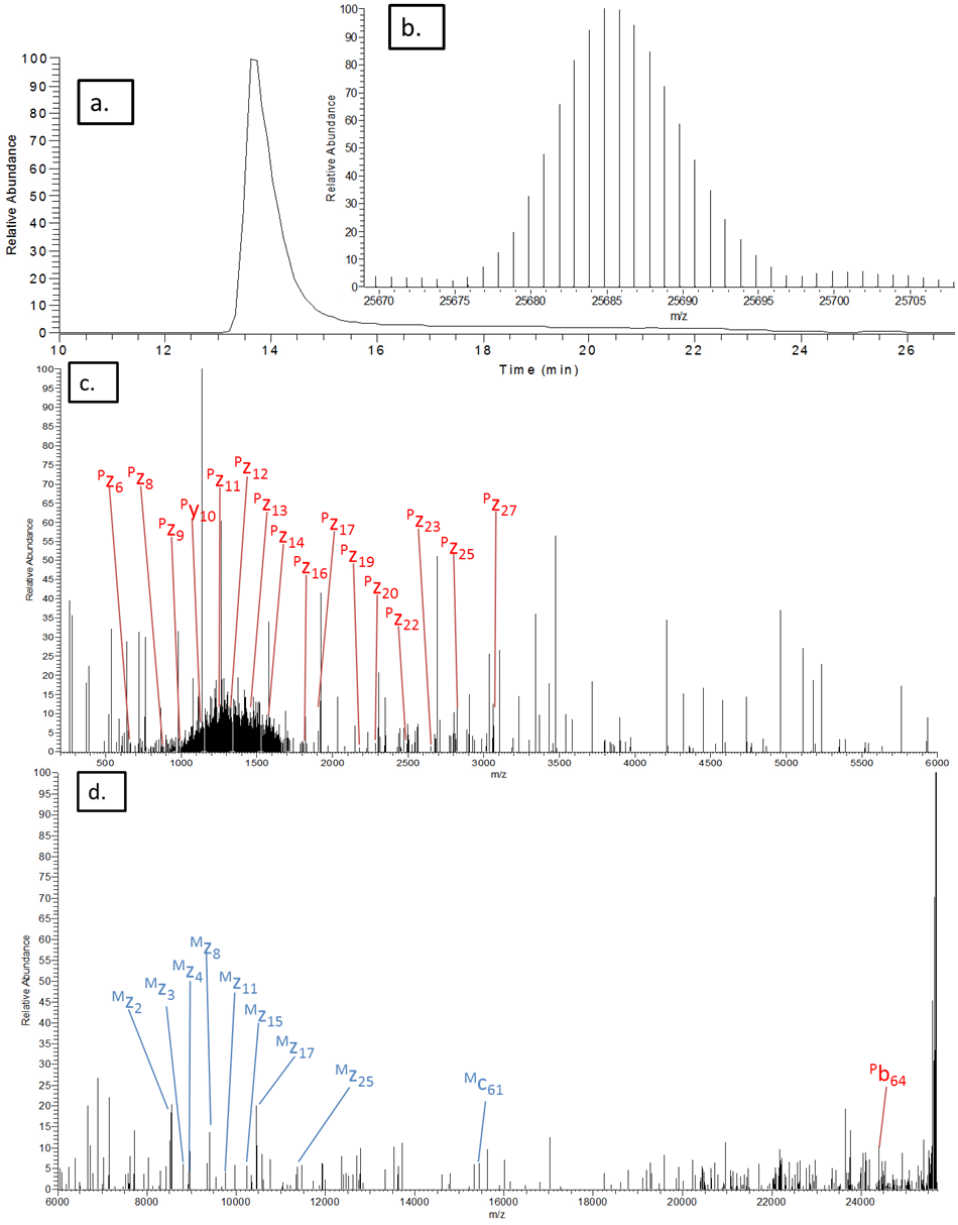
| | Ubiquitin 8.5 kDa | | Myoglobin 16 kDa | | Carbonic Anhydrase 29 kDa | |
|---|---|---|---|---|---|---|
| Fragmentation Technique | Percent Coverage | b/c/y/z-ions | Percent Coverage | b/c/y/z-ions | Percent Coverage | b/c/y/z-ions |
| EThcD 6 msec 10%SA | 95 | 14/61/29/59 | 70 | 1/76/15/76 | 21 | 7/23/6/21 |
| EThcD 12msec 10%SA | 95 | 12/59/35/57 | 77 | 3/74/22/75 | 18 | 6/18/6/24 |
| EThcD 25msec 10%SA | 96 | 20/61/37/58 | 73 | 2/67/22/70 | 16 | 1/19/1/21 |
| | | | | | | |
| CID 25% | 81 | 45/0/54/2 | 45 | 24/5/61/5 | 10 | 10/0/18/0 |
| CID 35% | 87 | 40/1/59/4 | 40 | 19/4/55/5 | 15 | 16/2/20/0 |
| CID 50% | 87 | 28/1/61/7 | 29 | 12/0/36/5 | 16 | 18/2/22/1 |
| | | | | | | |
| HCD 12% | 56 | 27/1/37/3 | 37 | 26/7/49/1 | 17 | 25/2/18/0 |
| HCD 25% | 84 | 22/1/56/4 | 19 | 7/1/22/4 | 10 | 12/0/15/3 |
| HCD 40% | 35 | 10/1/15/2 | 3 | 2/1/0/2 | 3 | 0/0/8/2 |

Appendix Table 3. Increase seen in the relative abundance of the desired peptide product for trimers from time-limited acid hydrolysis when switching from 30sec to 60 sec digestion time.

| Ubiquitin trimer peptide | Ratio of abundance (60sec/30sec) |
|---|---|
| [53-76]–$^{48}$Ub–$^{48}$Ub | 0.8 |
| [53-76]–$^{33}$Ub–$^{33}$Ub | 1.9 |
| [53-76]$_2$–$^{6,48}$Ub | 3 |
| [53-76]$_2$–$^{11,33}$Ub | 2.6 |
| [53-76]$_2$–$^{11,63}$Ub | 6.7 |

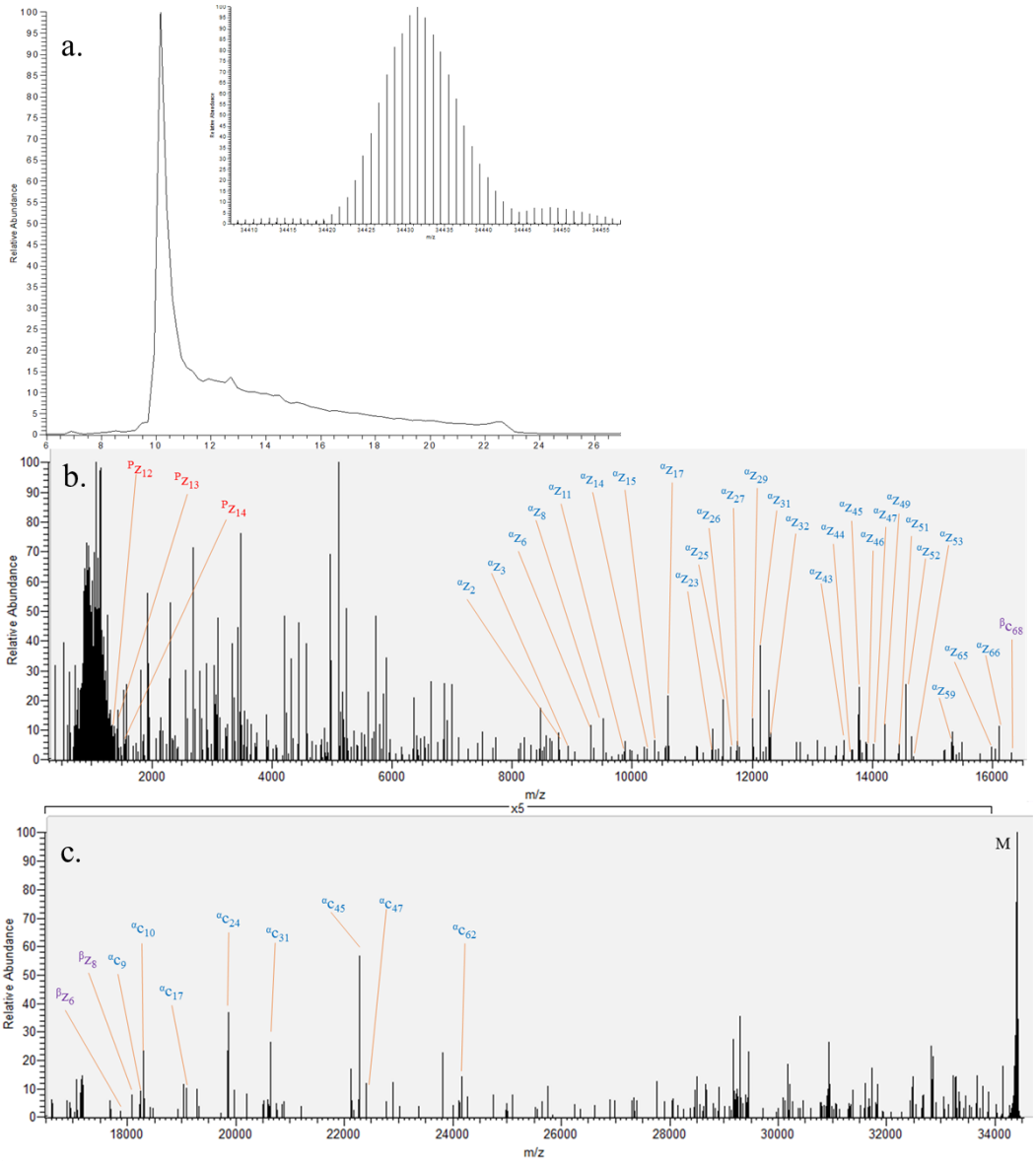



Appendix Figure 1. Fragmentation by CID observed in all fully truncated K-linked dimers. Linkages presented are K63 (a.), K48 (b.), K33 (c.), K29 (d.), K27 (e.), K11 (f.), and K6 (g.). In the case of this figure only, y ions are shown in red and b ions are shown in blue.




Appendix Figure 2. A representative LC-MS/MS visualized for Ub-$^{48}$Ub-$^{48}$Ub. The chromatogram is shown in a. from 8min to 27mins. Peak at 13.6 mins represents the elution of the trimer. The deconvoluted intact mass's isotope cluster is shown in b. The top 5 m/z were deconvoluted and plotted in c. (200-6000 Da) and d. (6000-25700 Da). The labeled masses in red are from the proximal ubiquitin (they also have a

superscript P) and they represent to ions used in Step 2 to distinguish that there is a mass addition to K48 on the proximal ubiquitin. In blue are shown the diagnostic ions from the endo ubiquitin (also shown with a superscript E) which show the chain is unbranched in Step 3 and the ions showing the linkage is K48 on the endo ubiquitin are also labeled.




Appendix Figure 3. LC-MS/MS spectrum of Ub–$^{63}$Ub–$^{6}$Ub–$^{63}$Ub.

a. The chromatogram is shown from 6 min to 28 min.

b. The deconvoluted isotope cluster of the molecular ions is plotted from 34,410-34,455 Da.

c. MS/MS spectrum obtained when the five most abundant isotope peaks were deconvoluted, merged and plotted from 200-16,500 Da (top) and 16,500-25,700 Da (bottom).

The units are the same on the two Y axes. Ions represented by peaks labeled P in red originate from fragmentation in the proximal ubiquitin. Peaks representing diagnostic ions from the $\alpha_1$-endo Ub are labeled $\alpha$ in blue. Peaks labeled $\beta$ in purple represent diagnostic ions formed from cleavages in the $\beta_2$-endo Ub. The base peak labeled M represents precursor ions not fragmented by ETciD.

# Bibliography


1. Jones, J.; Wu, K.; Yang, Y.; Guerrero, C.; Nillegoda, N.; Pan, Z.-Q.; Huang, L.A. *J. of Proteome Res.* **2008**, *7*, 1274–1287.
2. Bruderer, R.; Tatham, M. H.; Plechanovova, A.; Matic, I.; Garg, A.K.; Hay, R.T. *EMBO Rep.* **2011**, *12*, 142–148.
3. Hjerpe, R.; Aillet, F.; Lopitz-Otsoa, F.; Lang, V.; England, P.; Rodriguez, M. S. *EMBO Rep*. **2009**, *10*, 1250–1258.
4. Xu, X.; Niu, Y.; Liang, K.; Wang, J.; Li, X.; Yang, Y. *Biochemical and Biophysical Research Communications*. **2015**, *459*, 240–245.
5. Varshavsky, A. *Annu. Rev. Biochem.* **2012**, *81*, 167–176.
6. Finley, D.; Chau, V. *Annu. Rev. Cell Bio.* **1991**, *7*, 25–69.
7. Low, T.Y.; Magliozzi, R.; Guardavaccaro, D.; Heck, A.J.R. *Proteomics*. **2012**, *13*, 526–537.
8. Goto, E.; Yamanaka, Y.; Ishikawa, A.; Aoki-Kawasumi, M.; Mito-Yoshida, M.; Ohmura-Hoshino, M.; Matsuki, Y.; Kajikawa, M.; Hirano, H.; Ishido, S. *Journal of Biological Chemistry*. **2010**, *285*, 35311–35319.
9. Chen, P.-C.; Na, C.H.; Peng, J. *Amino Acids.* **2012**, *43*, 1049–60.
10. Sylvestersen, K.B.; Young, C.; Nielsen, M.L. *Curr. Opin. Chem. Biol.* **2013**, *17*, 49–58.
11. Vogel, R.I.; Coughlin, K.; Scotti, A.; Iizuka, Y.; Anchoori, R.; Roden, R.B.; Marastoni, M.; Bazzaro, M. *Oncotarget*. **2015**, *6*, 4159–4170.

12. Xia, Z.-P.; Sun, L.; Chen, X.; Pineda, G.; Jiang, X.; Adhikari, A.; Zeng, W.; Chen, Z.J. *Nature*. **2009**, *461*, 114–119.

13. Silva, G.M.; Finley, D.; Vogel, C. *Nat. Struct. Mol. Biol.* **2015**, *22*, 116–23.

14. Hao, R.; Nanduri, P.; Rao, Y.; Panichelli, R.S.; Ito, A.; Yoshida, M.; Yao, T.-P. *Molecular Cell*. **2013**, *51*, 819–828.

15. Akutsu, M.; Dikic, I.; Bremm, A. *J. Cell Sci.* **2016,** 875–880.

16. Nguyen, L.K.; Dobrzyński, M.; Fey, D.; Kholodenko, B.N. *Front. Physiol.* **2014**, *5*, 4.

17. Mayne, J.; Starr, A.E.; Ning, Z.; Chen, R.; Chiang, C.-K.; Figeys, D. *Analytical Chemistry Anal. Chem.* **2014**, *86*, 176–195.

18. Newton, K.; Matsumoto, M.L.; Wertz, I.E.; Kirkpatrick, D.S.; Lill, J.R.; Tan, J.; Dugger, D.; Gordon, N.; Sidhu, S. S.; Fellouse, F.A.; Komuves, L.; French, D.M.; Ferrando, R.E.; Lam, C.; Compaan, D.; Yu, C.; Bosanac, I.; Hymowitz, S.G.; Kelley, R.F.; Dixit, V.M. *Cell.* **2008**, *134* (4), 668–678.

19. Gilda, J.E.; Ghosh, R.; Cheah, J.X.; West, T.M.; Bodine, S.C.; Gomes, A.V. *PLoS One* **2015**, *10* (8), 1–18.

20. Reyes-Turcu, F.E.; Horton, J.R.; Mullally, J.E.; Heroux, A.; Cheng, X.; Wilkinson, K.D. *Cell.* **2006**, *124* (6), 1197–1208.

21. Hospenthal, M.K.; Mevissen, T.E.T.; Komander, D. *Nat. Protoc.* **2015**, *10* (2), 349–361.

22. Hospenthal, M.K.; Freund, S.M.V.; Komander, D. *Nat. Struct. Mol. Biol.* **2013**, *20* (5), 555–565.

23. Nakasone, M.A.; Livnat-Levanon, N.; Glickman, M.H.; Cohen, R.E.; Fushman, D. *Structure* **2013**, *21* (5), 727–740.

24. Kristariyanto, Y.A.; Abdul Rehman, S.A.; Campbell, D.G.; Morrice, N.A.; Johnson, C.; Toth, R.; Kulathu, Y. *Mol. Cell* **2015**, 1–12.

25. Saito, K.; Horikawa, W.; Ito, K. *PLoS Genet.* **2015**, *11* (4), 1–18.

26. Liu, K.; Lyu, L.; Chin, D.; Gao, J.; Sun, X.; Shang, F.; Caceres, A.; Chang, M.-L.; Rowan, S.; Peng, J.; Mathias, R.; Kasahara, H.; Jian, S.; Taylor, A. *Proc. Natl. Acad. Sci. U. S. A.* **2015**.

27. Boname, J.M.; Thomas, M.; Stagg, H.R.; Xu, P.; Peng, J.; Lehner, P.J. *Traffic* **2010**, *11* (2), 210–220.

28. Meyer, H.J.; Rape, M. *Cell* **2014**, *157* (4), 910–921.

29. Denis, N.J.; Vasilescu, J.; Lambert, J.-P.; Smith, J.C.; Figeys, D. *Proteomics* **2007**, *7* (6), 868–874.

30. Thorne, A.W.; Sautiere, P.; Briand, G.; Crane-Robinson, C. *EMBO J.* **1987**, *6* (4), 1005–1010.

31. Peng, J.; Schwartz, D.; Elias, J.E.; Thoreen, C.C.; Cheng, D.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S.P. *Nat. Biotechnol.* **2003**, *21* (8), 921–926.

32. Xu, P.; Peng, J. *Biochim. Biophys. Acta* **2006**, *1764* (12), 1940–1947.

33. Xu, P.; Peng, J. *Anal. Chem.* **2008**, *80* (9), 3438–3444.

34. Kim, W.; Bennett, E.J.; Huttlin, E.L.; Guo, A.; Li, J.; Possemato, A.; Sowa, M.E.; Rad, R.; Rush, J.; Comb, M.J.; Harper, J.W.; Gygi, S.P. *Mol. Cell* **2011**, *44* (2), 325–340.

35. Gerber, S.A.; Rush, J.; Stemman, O.; Kirschner, M.W.; Gygi, S.P. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (12), 6940–6945.

36. Kaiser, S.E.; Riley, B.E.; Shaler, T.A.; Trevino, R.S.; Becker, C.H.; Schulman, H.; Kopito, R.R. *Nat. Methods* **2011**, *8* (8), 691–696.

37. Clague, M.J.; Heride, C.; Urbé, S. *Trends Cell Biol.* **2015**, 1–10.

38. Xu, P.; Duong, D.M.; Seyfried, N.T.; Cheng, D.; Xie, Y.; Robert, J.; Rush, J.; Hochstrasser, M.; Finley, D.; Peng, J. *Cell* **2009**, *137* (1), 133–145.

39. Geis-Asteggiante, L.; Dhabaria, A.; Edwards, N.; Ostrand-Rosenberg, S.; Fenselau, C. *Int. J. Mass Spectrom.* **2015**, *378*, 264–269.

40. Chi, A.; Bai, D.L.; Geer, L.Y.; Shabanowitz, J.; Hunt, D.F. *Int. J. Mass Spectrom.* **2007**, *259* (1-3), 197–203.

41. Coon, J. J. *Anal. Chem.* **2009**, *81* (9), 3208–3215.

42. Ge, Y.; Lawhorn, B.G.; ElNaggar, M.; Strauss, E.; Park, J.H.; Begley, T.P.; McLafferty, F.W. *J. Am. Chem. Soc.* **2002**, *124* (4), 672–678.

43. Frese, C.K.; Altelaar, A.F.M.; Hennrich, M.L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A.J.R.; Mohammed, S. *J. Proteome Res.* **2011**, *10* (5), 2377–2388.

44. Frese, C.K.; Altelaar, A.F.M.; Van Den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A.J.R.; Mohammed, S. *Anal. Chem.* **2012**, *84* (22), 9668–9673.

45. Varadan, R.; Assfalg, M.; Haririnia, A.; Raasi, S.; Pickart, C.; Fushman, D. *J. Biol. Chem.* **2004**, *279* (8), 7055–7063.

46. Castañeda, C.A.; Kashyap, T.R.; Nakasone, M.A.; Krueger, S.; Fushman, D. *Structure.* **2013**, *21*, 1168–1181.

47. Li, A.; Sowder, R.C.; Henderson, L.E.; Moore, S.P.; Garfinkel, D.J.; Fisher, R.J. *Anal. Chem.* **2001**, *73* (22), 5395–5402.

48. Li, J.; Shefcheck, K.; Callahan, J.; Fenselau, C. *Int. J. Mass Spectrom.* **2008**, *278* (2-3), 109–113.

49. Cannon, J.; Nakasone, M.; Fushman, D.; Fenselau, C. *Anal. Chem.* **2012**, *84*, 10121–10128.

50. Swatkoski, S.; Gutierrez, P.; Ginter, J.; Petrov, A.; Dinman, J.D.; Edwards, N.; Fenselau, C. *J. Proteome Res.* **2007**, *6* (11), 4525–4527.

51. Valkevich, E.M.; Sanchez, N.A.; Ge, Y.; Strieter, E.R. *Biochemistry* **2014**, *53* (30), 4979–4989.

52. Lee, A.E.; Castañeda, C.A.; Wang, Y.; Fushman, D.; Fenselau, C. *J. Mass Spectrom.* **2014**, *49* (12), 1272–1278.

53. Lee, A.E.; Geis-Asteggiante, L.; Dixon, E.K.; Kim, Y.; Kashyap, T.R.; Wang, Y.; Fushman, D.; Fenselau, C. *J. Mass Spectrom.* **2016**, *51* (4), 315–321.

54. Burke, M.C.; Wang, Y.; Lee, A.E.; Dixon, E.K.; Castaneda, C.A; Fushman, D.; Fenselau, C. *Anal. Chem.* **2015**, *87* (16), 8144–8148.

55. Osula, O.; Swatkoski, S.; Cotter, R.J. *J. Mass Spectrom.* **2012**, *47* (5), 644–654.

56. Dass, C. Fundamentals of Contemporary Mass Spectrometry; John Wiley & Sons: Hoboken, NJ, 2007.

57. Olsen, J.V; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nat. Methods* **2007**, *4* (9), 709–712.

58. Syka, J.E.P.; Coon, J.J.; Schroeder, M.J.; Shabanowitz, J.; Hunt, D.F. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528–9533.

59. Zubarev, R.; Kelleher, N.L.; McLafferty, F.W. *J. Am. Chem. Soc* **1998**, *120* (16), 3265–3266.

60. Sweredoski, M.J.; Moradian, A.; Raedle, M.; Franco, C.; Hess, S. *Anal. Chem.* **2015**, *87* (16), 8360–8366.

61. Sobott, F.; Watt, S.J.; Smith, J.; Edelmann, M.J.; Kramer, H.B.; Kessler, B.M. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (9), 1652–1659.

62. Moradian, A.; Franco, C.; Sweredoski, M.J.; Hess, S. *J. Anal. Sci. Technol.* **2015**, *6* (1), 21.

63. Zhao, P.; Viner, R.; Teo, C.F.; Boons, G.-J.; Horn, D.; Wells, L. *J. Proteome Res.* **2011**, *10* (9), 4088–4104.

64. Swaney, D.L.; McAlister, G.C.; Wirtala, M.; Schwartz, J.C.; Syka, J.E.P.; Coon, J.J. *Anal. Chem.* **2007**, *79* (2), 477–485.

65. Zubarev, R. a; Makarov, A. *Anal. Chem.* **2013**, *85* (11), 5288–5296.

66. Meierhofer, D.; Wang, X.; Huang, L.; Kaiser, P. *J Proteome Res.* **2008**, *7* (10), 4566–4576.

67. Burke, M.C.; Oei, M.S.; Edwards, N.J.; Ostrand-rosenberg, S.; Fenselau, C. **2014**.

68. Eliuk, S.; Makarov, A. *Annu. Rev. Anal. Chem.* **2015**, *8* (1), 61–80.

69. Yates, J.R.; Eng, J.K.; McCormack, A.L.; Schieltz, D. *Anal. Chem.* **1995**, *67* (8), 1426–1436.

70. Zamdborg, L.; LeDuc, R. D.; Glowacz, K.J.; Kim, Y.-B.; Viswanathan, V.; Spaulding, I.T.; Early, B.P.; Bluhm, E.J.; Babai, S.; Kelleher, N.L. *Nucleic Acids Res.* **2007**, *35* (Web Server issue), W701–W706.

71. Bonet-Costa, C.; Vilaseca, M.; Diema, C.; Vujatovic, O.; Vaquero, A.; Omeñaca, N.; Castejón, L.; Bernués, J.; Giralt, E.; Azorín, F. *J. Proteomics* **2012**, *75* (13), 4124–4138.

72. Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C. *J Proteome Res.* **2010**, 3886–3890.

73. Liu, X.; Sirotkin, Y.; Shen, Y.; Anderson, G.; Tsai, Y.S.; Ting, Y.S.; Goodlett, D.R.; Smith, R.D.; Bafna, V.; Pevzner, P. a. *Mol. Cell. Proteomics*, **2012**, *11* (6). 1-13.

74. Guner, H.; Close, P.L.; Cai, W.; Zhang, H.; Peng, Y.; Gregorich, Z.R.; Ge, Y. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (3), 464–470.

75. Fellers, R.T.; Greer, J.B.; Early, B.P.; Yu, X.; Leduc, R.D.; Kelleher, N.L.; Thomas, P. M. *Proteomics* **2015**, 1235–1238.

76. Boyne, M.T.; Pesavento, J.J.; Mizzen, C. A.; Kelleher, N.L. *J. Proteome Res.* **2006**, *5* (2), 248–253.

77. Xu, H.; Zhang, L.; Freitas, M. a. *J. Proteome Res.* **2008**, *7* (1), 138–144.

78. Swaney, D.L.; Rodríguez-mias, R.A; Villén, J. *Embo* **2015**, *16* (9), 1–14.

79. Kulathu, Y.; Komander, D. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (8), 508–523.

80. Setsuie, R.; Sakurai, M.; Sakaguchi, Y.; Wada, K. *Neurochem. Int.* **2009**, *54* (5-6), 314–321.

81. Strachan, J.; Roach, L.; Sokratous, K.; Tooth, D.; Long, J.; Garner, T. P.; Searle, M. S.; Oldham, N.J.; Layfield, R. *J. Proteome Res.* **2012**, *11* (3), 1969–1980.

82. Van Nocker, S.; Vierstra, R.D. *J. Biol. Chem.* **1993**, *268* (33), 24766–24773.

83. Andersen, K.A.; Martin, L.J.; Prince, J.M.; Raines, R.T. *Protein Science* **2014**, 1–34.

84. Tenno, T.; Fujiwara, K.; Tochio, H.; Iwai, K.; Morita, E.H.; Hayashi, H.; Murata, S.; Hiroaki, H.; Sato, M.; Tanaka, K.; Shirakawa, M. *Genes Cells*, **2004**, *9*, 865–875.

85. Varadan, R.; Walker, O.; Pickart, C.; Fushman, D. *J. Mol. Biol.* **2002**, *324* (4), 637–647.

86. Deverauxf, Q.; Ustrellf, V.; Pickart, C.; Rechsteiner, M. *The J. of Biological Chem*. **1994**, *5*, 7059–7061.

87. Castañeda, C.A.; Liu, J.; Chaturvedi, A.; Nowicka, U.; Cropp, T.A.; Fushman D. *J. Am. Chem. Soc*. **2011**, *133*, 17855-17868.

88. Castañeda, C.A.; Liu, J.; Kashyap, T.R.; Singh, R.K.; Fushman, D.; Cropp, T.A. *Chem. Communications*, **2011**, *47* (7), 2026-2028.

89. Dixon, E.K.; Castañeda, C.A.; Kashyap, T.R.; Wang, Y.; Fushman, D. *Bioorg. Med. Chem*. **2013**, ***21*** (12), 3421-3429.

90. Senko, M. W.; Remes, P. M.; Canterbury, J.D.; Mathur, R.; Song, Q.; Eliuk, M.; Mullen, C.; Earley, L.; Hardman, M.; Blethrow, J.D.; Bui, H.; Specht, A.;

Lange, O.; Denisov, E.; Makarov, A.A.; Horning, S.; Zabrouskov, V. *Anal. Chem.* **2013**, *85*, 11710−11714.

91. Lee, A.E.; Geis-Asteggiante, L.; Dixon, E.K.; Miller, M.; Wang, Y.; Fushman, D.; Fenselau, C. *J. Mass Spectrom.* **2016**, In Press.

92. Amerik, A.Y.; Swaminathan, S.; Krantz, B.A.; Wilkinson, K. D.; Hochstrasser, M. *EMBO J.* **1997**, *16* (16), 4826–4838.

93. Jiang, X.; Kinch, L.N.; Brautigam, C. A.; Chen, X.; Du, F.; Grishin, N. V.; Chen, Z.J. *Immunity* **2012**, *36* (6), 973–959.

94. Gerlach, B.; Cordier, S. M.; Schmukle, A. C.; Emmerich, C. H.; Rieser, E.; Haas, T.L.; Webb, A. I.; Rickard, J.A.; Anderton, H.; Wong, W. W.-L.; Nachbur, U.; Gangoda, L.; Warnken, U.; Purcell, A.W.; Silke, J.; Walczak, H. *Nature* **2011**, *471*, 591–596.