MS Cleavable Crosslinkers and Tags for Protein Structural Analysis

by

Billy M. Clifford-Nunn

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Chemistry)

in The University of Michigan

2013

Doctoral Committee

Professor Philip C. Andrews, Chair

Associate Professor Kristina I. Hakansson

Professor Janine R. Maddock

Professor Mark E. Meyerhoff

Assistant Professor Brandon T. Ruotolo

© Billy M Clifford-Nunn

All Rights Reserved

2013

Acknowledgements

Five years ago, almost to the day as I'm writing this, Brian Coppola gave a room full of incoming graduate students a goal: to get out. Clearly, those words and that goal have stuck with me these past five years. However, I would not have reached that goal without many people. In fact, so many people have touched my life and taught me things these past five years that I cannot name them all.

First, to my mentors, Dr. Phil Andrews and Dr. Janine Maddock, I am so thankful to have worked with you. You are both two of the most generous and brilliant people I know and it has been a privilege to work with you. You have taught me so much, (some things which I did not want to know) that I am truly grateful. Thanks for letting me choose what projects I wanted to pursue, pushing me when I needed to be pushed, coming into the lab and working with me when I needed help or just an extra pair of hands to make a deadline, sending me home when I was sick, and telling me to go take a nap when I was overtired and cranky. I really couldn't have asked for better mentors.

Thank you very much to my committee members Dr. Mark Meyerhoff, Dr. Brandon Ruotolo, and Dr. Kristina Hakansson for giving advice, offering suggestions, reading this thesis, and helping me to grow as a scientist. Your input has been invaluable and I wish I would have sought your guidance more frequently.

I am especially thankful for the interdisciplinary nature of this work and therefore the collaborations I participated in. Thank you to those who allowed me to be a guest in their laboratory and use their instruments. Kristina Hakansson allowed me to use the FTICR in her laboratory but I could not have successfully completed experiments without the expertise of Bo Wang. Brandon Ruotolo allowed me to use the Synapt in his laboratory and both Yueyang Zhong and Linjie Han were key in getting successful data from that instrument. Linjie, I really do appreciate your long evenings working with me as I collected the last data for my thesis. Thank you.

All my synthetic work was completed in the laboratory of Dr. Hollis Showalter who has to be one of the most amazing people I've ever met. He welcomed me into his lab at a time when I really didn't know much about synthesis. After all, the two semesters of organic chemistry that you take in college does not make a synthetic chemist. However, Hollis spent a lot of time with me, reminding me of basic techniques, teaching me how to search the synthetic literature, and showing me things in the synthesis lab that you don't know unless you have the years of experience that he has. I learned a great deal from Hollis and I'm very thankful for the time he invested in me. I will admit that I have two things about our interactions that I regret. First, that a chemical I was working with, DCC, caused us both to have a nasty allergic reaction, and second, that the time was so short.

Even though I was hesitant at the idea of rotations when I came to Michigan, I truly learned a lot from them and they contributed significantly to my professional development. The most important thing I learned was that when you truly enjoy doing research, it doesn't matter what project you work on. My rotation mentors Dr. Mark Meyerhoff, Dr. David Lubman, and Dr. Katrin Karbstein taught me about what was expected of me as a graduate student and I

enjoyed the research in each of their labs. It would have been a privilege to work with any of one of them.

Thank you so much to my undergraduate mentors Dr. Robert Stack, Marina Ionina-Prasov, and Monique Wilhelm. I may not have even considered graduate school without their influences. In fact, I probably would have ended up going to medical school. Thank you so much for showing me the light!

The people I worked with every day had a huge impact on the completion of this thesis. After all, when science isn't going well, the people working at the bench next to yours are often the ones that help get you through it. The long nights of preparing for conferences and lab presentations, group meetings that last longer than three hours, and mutual commiseration bring people together. I am very thankful and blessed to have worked in three different labs and have had great lab mates in all of them, people that I consider friends. Ryszard Zielke, Nicolette Ognjanovski, Daniel Sheidy, Lance Shultz, Katharyn Cochrane, Valerie Forsyth – it was each one of you who made me look forward to coming to the Maddock lab each day. Andrews lab members Eric Simon, Pratik Prasik, Donna Veine, Chunchao Zhang, Hye Kyong Kweon, Maureen Kachman, and Yan Zhang, you each have taught me things over the past five years and I thank you for that. Thanks to Bryan E. Smith for writing algorithms to help me analyze crosslinking data. Here's a big thank you to the Showalter lab members who answered my questions and offered suggestions when synthesis was just not working. Yafei Jin, Rod Sorenson, Allen Brooks, Fardokht Abulwerdi – thank you so much. And a special thank you goes to Angela Walker and Lolita Piersimoni – without your advice, help, and open ears, I wouldn't have made it through these last months with the little sanity I have left. Thank you all.

Although my education work is not included in this thesis, I am very grateful to those of you who have encouraged me in that area. Kathleen Nolta – someday I hope I'm as well liked and respected, and have as high an attendance record as you do. Joseph Krajcik and Brian Coppola – thanks for being such charismatic and entertaining teachers that you persuaded me to get a master's degree in secondary science education. Thank you Betsy Davis for mentoring me and jumping right into the midst into my chemistry education project. Amy Gottfried: I have had such fun with you and learned so much from you. Thank you for the teaching experiences you have given me. Thank you for encouraging me not to give up even when I really wanted to. Thank you for being both a mentor and a friend.

I could not have completed this work without my friends surrounding me, cheering me on. To the people that I have bounced ideas off of, vented with things were going poorly and celebrated with ice cream when things were going well, thank you. Gwen Anderson, Leah Makley, and Ashley Brant, your encouragement will not be forgotten and is very much appreciated. I am especially appreciative of my church friends for always encouraging me to lean on God. Shelly Turner, Serry Gardner, and Jodi Latocki, thank you. I am most thankful to God for giving me the abilities to do science, the persistence to keep going, the faith to believe things will work out, and for giving me rest when the way got too difficult.

Last, but certainly not least is my family. Thanks for putting up me working ridiculously long hours, putting chemicals in the refrigerator that I forget I had in my pocket, writing presentations during holidays, and pipetting in front of the TV because I had to work but I just wanted to be home. Thanks for listening to me talk and showing interest in my work even if you really may not have had much interest. To my husband, you have been amazingly understanding and supportive as I finished up the writing and the fact that you married me during this process

speaks to your character and patience. I love you very much. Mom, thank you for always encouraging me to meet my goals and thank you for the sacrifices you made so I could make those goals. Without you, I wouldn't be here. Even the birds have to have wind beneath their winds and you truly have been mine. I have now finished what I started, just as you asked.

Table of Contents

List of Figures

List of Tables

List of Schema

List of Appendices

List of Abbreviations

3D: Three dimensional

αCHA: α-cyano-4-hydroxycinnamic acid

Aβ: β amyloid

ACN: Acetonitrile

ADH: Alcohol dehydrogenase

CCS: Collision cross-section

CFP: Cyan fluorescent protein

CID: Collision induced dissociation

CIU: Collision induced unfolding

CYC: Cytochrome c

Da: Dalton

DABCO: 1,4-diazabicyclo[2.2.2]octane

DABCO Tag: 1-(4-((2,5-dioxopyrrolidin-1-yl)oxy)-4-oxobutyl)-1,4-diazabicyclo[2.2.2]octan-1-ium

DC: Direct current

DC4: DABCO Crosslinker containing four carbon atoms in each spacer arm. 1,4-Bis(4-((2,5-dioxopyrrolidin-1-yl)oxy)-4-oxobutyl)-1,4-diazabicyclo[2.2.2]octane-1,4-diium bromide

DC5: DABCO Crosslinker containing five carbon atoms in each spacer arm. 1,4-Bis(5-((2,5-dioxopyrrolidin-1-yl)oxy)-5-oxopentyl)-1,4-diazabicyclo[2.2.2]octane-1,4-diium bromide.

DC6: DABCO Crosslinker containing six carbon atoms in each spacer arm. 1,4-Bis(6-((2,5-dioxopyrrolidin-1-yl)oxy)-6-oxohexyl)-1,4-diazabicyclo[2.2.2]octane-1,4-diium bromide

DCC: Dicylcohexylcarbodiimide

DCU: Dicyclohexylurea

DNA: Deoxyribonucleic acid

DHB: 2,5-Dihydroxybenzoic acid

ECD: Electron capture dissociation

ETD: Electron transfer dissociation

EDAC: 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride

EI: Electron impact ionization

EM: Electron microscopy

ESI: Electrospray ionization

eV: Electron volt

FRET: Fluorescence resonance energy transfer

FTICR: Fourier Transform Ion Cyclotron Resonance

GDH: Glutamate dehydrogenase

GFP: Green fluorescent protein

GluC: Endoproteinase GluC

GST: Glutathione S Transferase

HCD: Higher energy C-trap dissociation

HDX: Hydrogen-deuterium exchange

HPLC: High performance liquid chromatography

IM: Ion Mobility

IRMPD: Infrared multiple photon dissociation

ISD: In-source decay

KLH: Hemocyanin from keyhole limpet

MALDI: Matrix-Assisted Laser Desportion Ionization

MeDABCO Tag: 1-4-((2,5-dioxopyrrolidin-1-yl)oxy)-4-oxobutyl)-4-methyl-1,4-diazabicyclo[2.2.2]octane-1,4-diium bromide iodide

MES: 2-(N-morpholino)ethanesulfonic acid

MS: Mass Spectrometry

MS/MS: Tandem mass spectrometry

MS^n: Tandem mass spectrometry, n represents the number of mass measurements made

NHS esters: N-hydroxysuccinimide esters

NMR: Nuclear Magnetic Resonance Spectroscopy

PC1: Bis(2,5-dioxopyrrolidin-1-yl) 2,2'-(piperazine-1,4-diyl)diacetate

PDB: Protein databank

PK: Pyruvate kinase

PPI: Protein-protein interaction

PPM: parts per million

PSD: Post-source decay

PTM: Post-translational modification

RF: Radio frequency

RNA: Ribonucleic acid

SAXS: Small angle x-ray scattering

SDS: Sodium dodecyl sulfate

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SPR: Surface plasmon resonance

SPW: Surface plasmon wave

SRIG: Stacked ring ion guide

T wave: Traveling wave

TAP: Tandem affinity purification

TPCK: L-1-tosylamido-2-phenylethyl chloromethyl ketone

TFA: Trifluoroacetic acid

TROSY: Transverse relaxation optimized spectroscopy

Y2H: Yeast Two Hybrids

YFP: Yellow fluorescent protein

Abstract

Chemical crosslinking combined with mass spectrometry is a powerful tool for analyzing protein interactions, however, there are several challenges associated with the method. Proteins are frequently crosslinked at lysine residues which makes the proteins more difficult to digest, resulting in large, branched peptides that can be difficult to ionize and fragment. Data interpretation is difficult because there are two overlapping series of ions from each peptide in the MS/MS spectrum. The first goal of this thesis was to synthesize, evaluate, and apply new crosslinkers to overcame several of these challenges.

The first crosslinker synthesized, PC1, was a cyclic tertiary amine and exhibited several of the desired properties upon initial evaluation with model peptides. However, upon application to the tetrameric protein aldolase, little fragmentation of the peptide backbone was needed for crosslinked peptide identification. The structure of PC1 was then modified to make the crosslinker more labile and the crosslinkers DC4, DC5, and DC6 were synthesized.

DC4, DC5, and DC6 each contained two intrinsic positive charges due to the presence of two quaternary amines. DC4-crosslinked peptides were characterized and facile fragmentation was observed at both sites of positive charge. This yielded sets of ions differing by the mass loss from the quaternary amines, and produced a mobile proton needed for backbone fragmentation. Further CID of the fragments consistently provided spectra suitable for peptide identification. DC4 was applied to the protein aldolase and several crosslinked peptides were found that were in agreement with the crystal structures. Residues that were deeply buried within the crystal structure were not observed as modified.

DC4-crosslinked proteins were analyzed by ion-mobility mass spectrometry. Crosslinked proteins had a larger CCS and required higher energies to unfold and to dissociate, indicating that crosslinkers stabilize the quaternary and tertiary structures of proteins. Crosslinking with DC4 altered the fragmentation pathway because different subunits were dissociated from the tetramer of crosslinked avidin than unmodified avidin. Furthermore, when sufficient energy was applied to fragment the tetramer into peptides, crosslinked avidin resulted in a much richer series of unique peptides compared to noncrosslinked avidin.

Chapter 1

Introduction

Protein structure is critical to its function. For example, the fibrous nature of collagen is necessary for it to function as a structural protein[1-3], while enzymes require specific conformations for substrate binding[4, 5], and activity[6, 7]. From primary structure to quaternary structure, deviations at any level can have detrimental effects on protein function which can lead to disease. A key example is the tetrameric protein hemoglobin[8, 9]. A single mutation in the primary structure from a glutamic acid to a valine residue on one of the polypeptide chains causes the protein to form fibers[10-12]. This change in the quaternary structure is the primary cause of the disease sickle cell anemia[10-18].

There is a significant body of literature detailing the structure and function of single proteins[19-23]; however proteins often interact and exist in complexes[24-26] of multiple proteins[27-30], deoxyribonucleic acid (DNA) molecules[31-34], ribonucleic acid (RNA) molecules[35-38] and/or metal ions or small molecule cofactors[39-44] as in the examples of ATP synthase[45], the ribosome[46], and the DNA replisome[47]. In fact, it often is the intact complex that is required for biological function e.g. the fully assembled ribosome is necessary for translation as opposed to the individual RNAs and proteins. Therefore, knowledge about the assembly and structure of protein complexes is critical to increase our understanding of molecular biology, disease, and to potentially develop drug targets. The literature describing the structure and function of most complexes is increasing at a much slower rate compared to the literature describing individual

subunits[48, 49]. This is primarily because current methods typically used to study protein structure do not work as well for large complexes[50]. Thus, the overall purpose of this thesis was to develop a method that could be used for large scale analysis of protein complexes and protein interactions.

There are several existing methods used to study protein structure and protein-protein interactions (PPIs), each with advantages and disadvantages[51] as shown in Tables 1.1, 1.2, and 1.3. The majority of structures deposited in the protein databank (PDB) are derived from x-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), and cryo-electron microscopy (cryo-EM), although small angle x-ray scattering (SAXS) (Table 1.1) and to a lesser extent, fluorescence resonance energy transfer (FRET) are also used to elucidate protein structure. Protein-protein interactions have mostly been determined using yeast-two hybrids (Y2H) and coimmuno-precipitation (CO-IP) experiments. However surface plasmon resonance (SPR), protein microarrays, and the previously mentioned techniques such as FRET and NMR have been used to obtain interaction data with varying levels of success (Table 1.2). Purification of protein complexes may be difficult so techniques such as tandem affinity purification (TAP) have been developed to purify complexes. After purification, the identities of proteins within complexes are regularly analyzed by mass spectrometry (MS). MS is a commonly used detection method for CO-IPs, SPR and protein microarrays.

In the past few decades, mass spectrometric methods have been advanced beyond just a detection method and have become amenable to analysis of protein structure and complexes. Techniques such as native mass spectrometry, ion mobility mass spectrometry (IMS) or a combination of mass spectrometry with hydrogen-deuterium exchange (HDX) or chemical

crosslinking have been used to yield valuable structural insights, particularly with large complexes that cannot be analyzed by other methods (Table 1.3).

Table 1.1. A comparison of the four commonly used methods to determine protein structure

Technique	**Benefits**	**Limitations**	**Information Gained**	**Routine Resolution**
X-ray Crystallography	• Atomic Resolutions	• Crystals required • Large amount (milligrams) of pure material needed	• 2°, 3°, 4° structure • Some PPIs	Atomic (< 2Å)
NMR	• Atomic Resolution • In solution • Small amounts of mid range solution concentrations needed	• Size limited (<50 kDa) • Sample must be homogeneous	• Protein folding • Kinetic data • 2°, 3°, 4° structure • Some PPIs	Atomic (< 2Å)
Cryo-EM	• In solution • Mid resolution – amino acids not assigned • Transient complexes analyzable	• Difficult sample preparation – sample needs to be as pure and homogeneous as possible. • MW > 200 kDa • Diameter < 700Å	• Shape of intact complex • 2°, 3°, 4° structure	Atomic (> 4Å)
SAXS	• In solution • Large particle size accessible	• Low resolution • Pure protein needed • High concentration of protein needed (1 mg/ml)	• Overall structural features • Stoichiometry of complex • Extent of protein folding	12-15 Å

Table 1.2. A comparison of the commonly used methods to determine protein-protein interactions

Technique	Benefits	Limitations	Information Gained
Y2H	• High-throughput • Membrane proteins accessible • No special equipment required	• High false positive rate • High false negative rate	• Interaction partners • Regions of interactions
Co-IP	• Proteins in biologically relevant states • Special equipment not necessarily required	• Antibody to protein of interest required • High false positive rate • High false negative rate • One protein's interacting partners studied at a time	• Interacting partners
FRET	• In vivo application • In cell localization	• False positives and negatives • Low sensitivity due to background • Limited computational tools • Requires pair wise analysis of protein fusions	• Binary interacting partners • Protein conformational changes
SPR	• Real time measurements	• False positive and negatives dues to immobilization on a surface • Only measuring surface behavior	• Interaction data • Real time association rates
Protein Microarrays	• Many interacting partners screened at once • Many types of partners	• Requires purification of the protein of interest • Nonspecific binding causes false positives • Limited sensitivity • False negatives	• Multiple types of interacting partners (DNA, RNA, protein, small molecule, etc)
TAP	• High throughput screening of protein complexes	• High false positive rate • High false negative rate • One protein's interacting partners studied at a time • Lose weak interactions	• Interacting partners

Table 1.3. A comparison of mass spectrometric methods for determining protein structure and interactions.

Technique	Advantages	Disadvantages	Information Gained
Native MS	• Fast, sensitive method • Little material needed • Stoichiometry of complexes obtained	• Dissociation is not random • Limited dissociation	• Complex stoichiometry • Identities of complex components • Mass of complex
HDX	• Relatively fast and sensitive • Little material needed	• High rates of back exchange – lowers sensitivity • Deuterium scrambling • Labor intense data analysis • Not high throughput	• Info on 2°, 3°, 4° structure • Protein folding info
Ion Mobility	• Sensitive method • Shape and size of complex • Little material required	• Ionization conditions to maintain native structures must be developed for each complex • Limited computational tools • Time consuming data analysis • Not high throughput	• Complex conformations - size and shape • Extent of protein folding • Stoichiometry of complex
Crosslinking	• Covalently bonds PPI in place • Strong purification methods can be used • Sensitive method • High-throughput potential • Little material required	• Chemical crosslinking limits peptide fragmentation • Limited computational tools • Time consuming data analysis	• Direct PPIs • Approximate site of interaction • Puts constraints on 3° and 4° structure

1.1 Methods of Studying Protein Structure.

1.1.1 X-Ray Crystallography.

Currently, over 85% of the structures in the PDB are derived using x-ray crystallography (PDB, http://www.pdb.org/). These structures range from the early work of the 1950s on simple proteins such as myoglobin[52] and hemoglobin[9] to more recent work involving large complexes such as ribosomes[46] and viruses[53]. In x-ray crystallography, a protein must first be crystallized so that the protein molecules are in a three dimensional ordered array. The crystals are then bombarded with x-rays, which are refracted when they strike an area of electron density. The intensity and angle of refraction can be measured and this information is used to construct an electron density map[54]. The intensity of diffraction data from one molecule is too weak to analyze, so in an x-ray crystallography experiment, the sample must be crystallized. This results in a three dimensional ordered array of molecules that are then bombarded with x-rays. The intensities and angles of the diffracted x-rays are reconstructed into an electron density map which is useful for determining the mean positions of the atoms in the crystal. This high resolution method provides a superior level of detail about protein structure and can provide information about protein-protein interactions[55-57].

A major limitation of this technique is the need for crystals. A large amount of homogenous, highly concentrated, very pure material is needed to make suitable crystal. It can be very difficult to isolate sufficient quantities of the pure protein, particularly for larger complexes that are more dynamic in nature. Moreover, obtaining protein crystals is not trivial. The crystallization process often requires several days or weeks and may require hundreds of trials with different crystallization conditions. This is difficult because the complex must be stable over the crystallization period. Furthermore, if the protein has flexible regions or if the

complex is dynamic, usable crystals may not be obtained or the data generated may be inconclusive. Although protein interaction data can be obtained using x-ray crystallography, that data can be considered questionable[58, 59] because it is unclear whether the proteins would interact in solution or whether the contacts observed are due to the crystallization process[58].

1.1.2 Nuclear Magnetic Resonance Spectroscopy. NMR is another powerful technique for studying protein complexes and interactions [60-64]. Each peak in an NMR spectrum represents a nucleus in a different chemical environment. Therefore, a one dimensional NMR spectrum of proteins can yield information about the extent of protein folding[65, 66]. However, a one dimensional spectrum is not sufficient for determining more intricate structural details because there are too many atoms to be resolved. To increase spectral resolution multidimensional NMR was developed[64]. Data are obtained that include the distances between atoms up to 6Å away (both through space and through bonds) and bond angles. These data, combined with the protein sequence, can be input into programs that calculate a structure consistent with those restraints. Spectra of select small proteins have been sufficiently resolved to determine their structures using 2D analysis[67-69] but, most proteins require additional resolving power. To achieve the resolution necessary for larger proteins with more atoms, higher dimensional analysis using multiple nuclei[70], transverse relaxation-optimized spectroscopy (TROSY)[71-73], isotope labeling and/or a combination of the three is required. Cryo-probes and higher strength magnets also have contributed to the accessibility of protein NMR. The use of these techniques has facilitated the analysis of very large complexes[72-74], such as the binding site of GroES onto GroEL[74] and the structure of malate synthase G, an 82kDa enzyme[72]. However, these types of studies require a high frequency magnet, labeling with ^{2}H, and the data analysis is very demanding, making these types of analyses difficult and far from routine.

There are several benefits to using NMR for determine protein structure. First, NMR of protein complexes is usually performed in solution, making NMR currently the only in-solution technique that is routinely capable of producing atomic resolution protein structures. Second, in most cases, all parts of the protein are observed, regardless of whether the protein is structured or unstructured. Unstructured regions can complicate the data, but unlike x-ray crystallography, the data is interpretable.

NMR is frequently used to determine protein interaction surfaces and binding sites[60, 63]. When a protein binds to another protein or a ligand, the chemical environment at the binding site changes and thus the chemical shifts of the nuclei are altered and detected. Depending on the NMR being performed, the complex is analyzed over time (minutes to several days) which means weaker interactions (Kd ~ mM) between two proteins can be observed.[75]

The main limitations of using NMR to analyze protein complexes are the size of the complex and the amount and purity of sample needed. Generally, proteins larger than 50kDa are too large for routine analysis although some exceptions have been reported[72, 74]. With higher field magnets and cooled sample detection systems, routine analysis can be completed on 100 µl of a 50µM isotopically labeled sample[76]. However, isotopic labeling is expensive and is not always possible. Without isotopic labeling, a protein concentration greater than 1mM is needed for a successful NMR study. High purity samples are required and the high concentration needed is problematic for insoluble proteins, and labor-intensive for soluble proteins. However methods for in-cell NMR structure determinations are being developed[77-81].

1.1.3 Cryo-Electron Microscopy. Cryo-EM is an in-solution technique useful for obtaining mid resolution structures of large complexes in a native state[82]. In a cryo-EM experiment, a sample is rapidly frozen to cryogenic temperatures before water has a chance to

form ice crystals [83]. A beam of electrons is aimed at the sample and the electrons either pass through the sample without striking anything or hit the sample and are diffracted. The electrons are detected and the energy and the angle at which they hit the detector can be converted into an image. This technique typically produces low resolution images of an intact complex, although improvements are being made and images with atomic resolution are becoming more routine[84-86]. Atomic resolution of 4Å is sufficient to fit the protein backbone onto the image, but higher resolution is required to confidently assign amino acid residues to a structure. Several structures have been determined at ~3.5Å resolution[87-89] and in the case of the human adenovirus, structural features were revealed in the cryo-EM images[87, 90] that were not resolved in the x-ray crystal structure[91]. Also in the adenovirus study, the amino acids needed for several protein-protein interactions were identified[90].

Cryo-EM is useful because it can be used to fit individual components generated from another high resolution technique onto the intact complex structure.[92-95] An additional benefit is that transient complexes can be analyzed because the rapid freezing maintains the solution phase conformation of the complex[89, 96]. In cryo-electron tomography, the sample can be tilted at multiple angles and images taken to get three dimensional (3D) information[97-99]. The sample can only be titled 60-70° before it is too thick to obtain an image, but then computer algorithms are used to compile all the data and generate a full 3D shape[100, 101].

Sample preparation is a key factor in achieving publishable data in cryo-EM experiments and can be troublesome. First, the sample must be thin enough that the electrons can pass through the sample as opposed to being absorbed. Second, the sample must be frozen quickly enough that cubic ice is not formed. If ice is present in the sample, the ice will absorb all the

electrons and make the sample useless. Third, purer samples generally produce better data, although papers are beginning to be published detailing structures from inside the cell[102].

As with any technique, there are several limitations to using cryo-EM. One limitation to EM is that the beam of electrons can obliterate the sample, destroying its structure. This is overcome in cryo-EM by reducing the number of electrons, but it results in an image with low contrast, making it more difficult to distinguish structural features[83, 103]. To partially solve this problem, many images are taken and the data is averaged across all the images recorded. Because the images are averaged, the particles being analyzed must be structurally homogeneous or the data will be misleading. Therefore, harsh purification methods that can cause structural deformity or partial dissociation of a complex can be problematic with cryo-EM. Particle size is also a limitation. To achieve high resolution structures, a particle with a molecular weight larger that 200 kDa is needed, but particles with a diameter greater than 700 Å are problematic[82]. Cryo-EM is not routinely used for determining protein-protein interactions. In the human adenovirus study, the amino acids needed for several protein-protein interactions were able to be identified because of the higher resolution achieved[90]. Until at least 3.5 Å resolution is commonly obtained, protein interaction studies in the literature using this approach will remain scarce.

1.1.4 Small angle-x-ray scattering. SAXS is similar to x-ray crystallography in that x-rays are aimed at the sample and upon colliding with electrons, the x-rays scatter and are detected[104]. However, there are two important differences between the two techniques. SAXS is a solution based technique, so when the x-rays strike the sample, they are hitting a single molecule as opposed to a crystal that contains repeated units of single molecules. Since molecules in solution are in multiple conformations rather than an orderly array (as in a crystal),

the scattered x-rays detected are an average of all conformations. This results in a much lower resolution structure than produced by x-ray crystallography[105]. In x-ray crystallography, the majority of x-rays are measured, regardless of the angle that the x-rays were diffracted, but in SAXS, only x-rays that have been diffracted 10° or less are recorded.

SAXS is a useful low resolution technique (12-15Å) for studying protein complexes. X-rays are not as likely to destroy the sample as the electron beam used in cryo-EM experiments, and larger particle sizes can be analyzed unlike in NMR. Mostly, SAXS is used for identifying aggregated versus unfolded proteins, and overall structural characteristics and stoichiometry of proteins and complexes[106]. Data from SAXS can be combined with other data (x-ray crystal structures, PPI data etc) and modeling completed to yield more detail regarding protein structure[107-110]. For example, DNA gyrase A subunit is composed of two subunits; one has a known crystal structure and the other subunit does not. SAXS data of the DNA gyrase A subunit, combined with the known crystal structure, and modeling was used to determine the envelope of the complex and therefore the arrangements of the domains[108]. SAXS is not used to identify protein-protein interactions, although algorithms exist to predict PPIs from SAXS data[111].

The limitations to SAXS include the low resolution, large amount of protein needed, and computational tools needed. The low resolution limits the information obtainable by SAXS, although useful data can still be obtained as discussed in the preceding paragraph. High concentrations of relatively pure protein are required (1 mg/ml) for SAXS experiments[106] which can be problematic for insoluble proteins. The protein of interest must be the major component of the solution and have all higher molecular weight species removed, but buffer components are not critical. In fact, the components needed to solubilize membrane proteins (i.e. detergents) are

compatible with SAXS, making SAXS a good technique for studying membrane proteins and complexes[112, 113]. The lack of computational tools to analyze data has been a significant problem in the past. Algorithms are now available for analysis of SAXS data[114-116], but algorithms that combine SAXS data with data from other techniques are still needed[117].

1.2 Methods of Studying Protein Interactions.

1.2.1 Yeast Two-Hybrids. The Y2H screen is a high-throughput method of studying multiple binary interactions simultaneously *in vivo*. One of the potential interacting partners (the bait) is genetically engineered with the DNA-binding region of a transcription factor responsible for transcribing a reporter gene that is not expressed in the cell normally, originally the *lacZ* gene[118]. The second interacting partner (the prey) is genetically engineered with the activation domain of the same transcription factor. Without both the DNA binding domain and the activation domain, the reporter gene (i.e. *lacZ*) will not be expressed. If the bait and the prey interact, the two parts of the transcription factor will be in close proximity. This allows for activation and transcription of the reporter gene, which causes a specific phenotype in the cell line. Y2H screenings have been used in large scale yeast studies to reveal thousands of potential binary interactions[119, 120] and multiple adaptations have been made to the original Y2H system[118] to yield new information regarding PPIs. Using these modified Y2Hs, more than one bait can be used simultaneously[121], inhibitors of PPIs can be identified[122], membrane proteins can be studied[123] and in PPIs with large proteins, the region of interaction can be determined[124].

The major limitations to Y2H screens are the high false positive and false negative rates.[125] For example, the yeast two-hybrid system has been used to determine PPIs of

mammalian proteins[126, 127], however, the mammalian protein may be post-translationally modified differently in yeast as opposed to the parent organism. Post-translational modifications (PTMs) can greatly affect PPIs and lead to false positives or negatives. Additionally, depending on the Y2H chosen, certain PPIs may not be detectable. In the original Y2H screens, membrane proteins and proteins that could not transverse the nuclear membrane were not accessible[118], although there are now Y2H methods to analyze membrane proteins[123]. Because both the bait and the prey are fusion proteins, the location of the reporter protein can prevent PPIs from occurring, either by causing protein misfolding or through steric hindrance[128]. False positives can arise because of nonspecific interactions and because of biologically irrelevant conditions. The bait and prey proteins may be expressed at a higher level or may be localized in an alternative cellular compartment. Because of these reasons, it has been common practice to verify protein interactions observed in yeast two-hybrid screens with alternative and complimentary methods.[125]

1.2.2 Coimmunoprecipitation. CoIP is routinely used to determine interacting partners of a particular protein[129-131]. First, cell lysates are treated with antibodies that recognize the protein of interest. The protein of interest binds to the antibody and then the antibodies are precipitated from the remaining cell lysate, along with the antigen and the proteins bound to the antigen. The antibody is then washed several times, the bound proteins eluted, and analyzed by immunoblotting or mass spectrometry[132]. One of the key advantages to CoIP experiments is that proteins are mostly in their biologically relevant states. The cell expresses the proteins at appropriate cellular levels, post-translationally modifies them, and assembles the complexes, so when a complex is purified and analyzed, the results should be similar to *in vivo* studies.

Although CoIP has been used successfully to identify numerous PPIs, the location of interaction is not identified and there are several limitations to this method. One requirement to this method is an antibody to the protein of interest. Generating antibodies to numerous proteins can be time consuming and expensive, so often the proteins are genetically engineered to contain a tag or epitope to which a commercial antibody exists, i.e. glutathione s-transferase (GST)[133, 134]. Depending on the protein, the epitope tag can cause protein misfolding or can prevent interactions, leading to false positives and negatives. The precipitation and washing steps in CoIP experiments cause a bias towards strong interactions because weak interacting partners may be washed away. False positives can also occur if proteins are in proximity to each other due to cell lysis but not in the cell itself or if the antibody recognizes the coprecipitated protein[132].

1.2.3 Tandem Affinity Purification. A modification of CoIP, in this method an affinity tag (TAP) is engineered into the protein of interest that allows the protein to be selectively extracted from cell lysates, along with any proteins bound to the original protein[135-139]. The proteins are then analyzed and identified, typically by mass spectrometry. This method has been used to successfully purify and identify hundreds of protein complexes in yeast[140], including proteins that did not have previous functional annotations[24].

TAP can be used in large scale studies for high throughput screening of protein interactions but has high false positive and false negative rates.[137, 141] In the yeast studies, some known interactions were not observed. The tags attached to the protein of interest may prevent protein interactions, as well as affect protein expression, folding, and function[51]. It is also possible that after the protein folds, the tag may no longer even be accessible to the antibody, preventing protein complex isolation. Although the TAP technique allows for identification of

all the proteins isolated, the presence of multiple protein complexes or interactions cannot be determined. Another disadvantage is that TAP is a stringent purification method, so weaker interactions may not be observed.

1.2.4 Fluorescence Resonance Energy Transfer. FRET is very useful to determine if two specific proteins interact. The two proteins of interest are each genetically engineered to contain different chromophores, usually fluorescent. Typically variations of green fluorescent protein (GFP)[142, 143], yellow fluorescent protein (YFP)[144] and cyan fluorescent protein (CFP)[145, 146] are used. The proteins are then irradiated with the wavelength of light that one of fluorophores absorbs at which causes that fluorophore to have more energy and be in an excited state. If three specific criteria are met, then energy can be transferred from one fluorophore to the other[147-150]. First, the fluorophores must be within 10nm of each other. Second, the emission spectrum of the donor fluorophore must significantly overlap with the absorption spectrum of the acceptor. Lastly, the electromagnetic dipoles of the donor emission and the acceptor absorption must be aligned properly for the energy transfer to occur. Because FRET experiments are done in solution, the orientations of the fluorophores are randomized over the time of the experiment and this criterion is met[151]. The energy is then transferred from one fluorophore to another, causing the acceptor fluorophore to be in an excited state. When the acceptor fluorophore returns to the ground state, it fluoresces and either the decrease in fluorescence of the first fluorophore or the increase in fluorescence of the second fluorophore is measured. The amount of energy transferred is inversely dependent on the distance between the two fluorophores, therefore proteins that interact and are close to each other will transfer energy more efficiently than those proteins that are not in close proximity[149].

FRET has been used successfully to measures changes in protein conformations, for example, when a ligand binds or when as a complex completes its function[152]. There are a multitude of studies where FRET has been used to study binary interactions of proteins *in vivo*[153-155]. With the combination of FRET and microscopy, FRET can also be used for the localization of interactions within the cell[156-159], making this a truly powerful technique.

Albeit powerful, there are several limitations to FRET that must be considered. First, FRET is labor intensive because each protein must be genetically engineered with fluorophores. The throughput is low, resolution is limited, and the analysis of PPIs is mostly of binary interactions. Biologically irrelevant data may be produced because the presence of the fluorophores can actually change the structure of the protein by preventing protein folding or preventing the interaction between the two proteins. The significant overlap of the donor emission spectrum with the acceptor absorption spectrum that is required for FRET is also problematic because it generates background fluorescence that is detected. Computational methods exist to partially account for this[160, 161], but it is still problematic.

1.2.5 Surface Plasmon Resonance. SPR is based on the principle that electrons in metals have the ability to undergo charge-density oscillations known as plasmons[162-164]. When a metal surface is hit with plane polarized light, surface plasmon waves (SPWs) are generated and the majority of light is reflected off the surface[165, 166]. If that light hits the surface at a very specific angle known as the SPR angle, some of the light energy will interact with the SPWs, producing surface plasmon resonance, and that light will not be reflected. The SPR angle can be calculated because it will be the angle at which the least amount of light is reflected[167, 168]. The SPR angle is dependent on the metal film, and on the thickness and refractive index of whatever is in contact with the metal film[168]. If there is a change in the SPR angle, that change can be

quantified and used to calculate the refractive index or thickness of the surface next to the metal film[169]. To determine protein interactions, a protein of interest can be bound to the metal film and then potential binding partners can be flowed over the protein of interest. If the proteins interact, the refractive index will change, thus changing the SPR angle. The more proteins that bind, the bigger the change in the SPR angle and binding affinity can be estimated.

One of the main benefits to using is SPR is that affinity of interactions and association rates can be measured in real time in real time[170]. Very small amount of material (μg) is consumed and it is not necessary to label the molecules. Because this technique is based on the refractive index of the surface next to the metal, many types of interactions can be studied including protein-protein[171], protein-nucleic acid[172] and protein-ligand[173, 174]. Recently, groups have started working on imaging and high-throughput studies as the next step in SPR[175-177].

One limitation to SPR is due to the fact that one of the analytes must be immobilized onto a surface. This immobilization can cause the protein to behave differently, either to due to restriction of the movement required for analyte binding or from blocking the binding site altogether. Plasmons have a maximum intensity at the metal surface that decays exponentially, making SPR a surface technique that can only probe 200nm from the surface. The behaviors of the analytes in bulk solution are not measured. Also proper controls must be completed to ensure that the analytes are not interacting with the metal surface.

1.2.6 Protein microarrays. In a protein microarray, a surface with many immobilized molecules (i.e. proteins) has a solution a solution containing analytes of interest flowed over it. Any analytes that are nonspecifically bound are removed in a wash step. Specific interactions can be detected by fluorescence or mass spectrometry or SPR. The intensity of the signal is

proportional to the number of molecules bound and thus gives an indication of the strength of the interaction. The sample immobilized on the surface could be an antibody[178-180], a small molecule drug, a nucleic acid[181-183], cell lysate[184], or a protein or peptide[185, 186]. The analyte that is flowed over the surface could be a labeled antibody or an unknown sample, therefore a variety of interacting partners can be screened.

A key benefit of protein microarrays is that many proteins and can be screened at once for many types of interacting partners. However, this requires purification of many interacting partners which can be time consuming. To make microarrays higher throughput, chemically synthesized peptides instead of intact proteins can be used[187]. Sensitivity and specificity are challenges with protein microarrays. Nonspecific binding of a protein to the surface and to the protein of interest must be controlled for. False positives are produced by proteins that are in great excess and nonspecifically bind the protein of interest. These can produce more signal than a true interacting partner that is very low in abundance. False negatives can occur whenever an analyte is immobilized on a surface in such a way that its interacting site is altered or unavailable.

1.3 Mass Spectrometry

As discussed in the preceding section, there are multiple methods that can elucidate protein structures and interactions. With CoIPs, TAP, and protein microarrays, mass spectrometry is often the detection method of choice due to its sensitivity, speed, and accuracy[137, 188-193]. However, mass spectrometry can also be used as more than a detection method. There are mass spectrometry methods that provide complementary information or

analyses of samples that cannot be obtained by other methods. Because mass spectrometry was such a critical part of this thesis, it will be discussed in detail below.

Mass spectrometry is a technique in which the mass to charge ratio (m/z) of an analyte is measured. A broad range of analytes can be assessed using different mass spectrometers as long as the analyte is charged and in the gas phase. Several mass spectrometers were used in this work and each has its advantages and disadvantages. There are three major components to a mass spectrometer as shown in Figure 1.1: the ionization source, the analyzer, and the detector. The ionization source is utilized to convert the sample from its current matrix into the gas phase (if needed) and ionizes the molecules. The analyzer then separates the ions based on their m/z, and lastly, the detector measures the intensity or abundance of those ions. Methods of detection will not be discussed here, but multiple ionization methods and mass analyzers were used in this thesis work and each will be discussed in further detail.

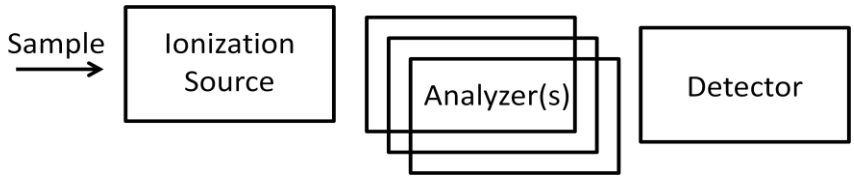


Figure 1.1. Components of a mass spectrometer. The three main components of a mass spectrometer include the ionization source, the analyzer, and finally the detector. Multiple analyzers can be utilized for tandem mass spectrometry experiments.

1.3.1 Ionization Methods. Since the early 1900s when J.J. Thomson developed the first mass spectrometer[194-196], development of ionization methods has been an important area in mass spectrometry research. Electron impact (EI) ionization was the main ionization method used through the 1960s[197, 198] and is still used in many instruments today. EI is a high energy ionization method and can cause the analyte to extensively fragment during ionization. This fragmentation is beneficial for identifying what functional groups are present in small organic

molecules, i.e. loss of 15 m/z suggests the presence of a methyl group. However, EI is not utilized for the analysis of peptides and proteins because of the challenges of getting large biomolecules into the gas phase and because smaller peptides, which can be ionized, fragment too extensively to provide useful information. Therefore to analyze these types of molecules, softer ionization methods had to be developed that did not result in such extensive fragmentation. Many ionization mechanisms were developed that were not suitable to protein work because they are still too harsh. It was not until the late 1980s that concurrently John Fenn developed electrospray[199] (ESI) and the combined work of Tanaka with Hillenkamp and Karas yielded matrix-assisted laser desorption ionization[200, 201] (MALDI). These two methods are the two most commonly used methods for ionizing proteins and are considered as soft ionization methods that form stable ions without extensive fragmentation during ionization.

1.3.1.1 Electrospray Ionization. In ESI, the analyte is present in either an acidic or basic solution which is pumped through a small capillary as shown in Figure 1.2. An electric field is applied between the capillary and the instrument inlet. Because of this electric field charge accumulates at the end of the capillary. With all the charge being localized in one area, there is a large amount of charge repulsion which causes the formation of a Taylor cone and then charged droplets are sprayed into the air. Gas and heat are used to evaporate the solvent from the charged droplets. As the droplets get smaller and more charged, eventually they explode into the gas phase because the charge repulsion is greater than the surface tension of the droplet.[202-204] This ionization method is known as a "soft" ionization method because fragmentation of intact proteins and peptides is minimal. It was later discovered that not only was electrospray a "soft" ionization method that did not result in fragmentation of covalent bonds, but also noncovalent interactions could be maintained[205, 206].

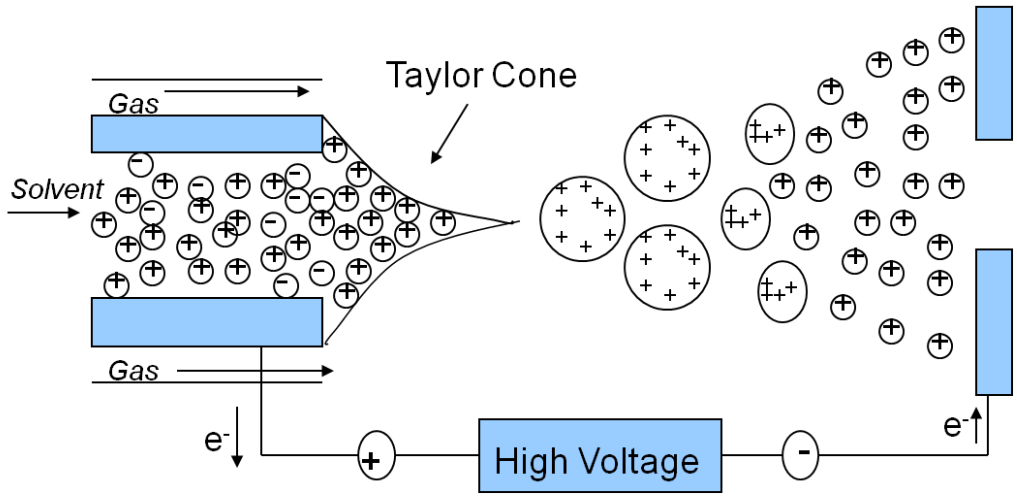


Figure 1.2. **A cartoon of electrospray ionization**. The solvent and analyte are passed through a capillary that has an electric field applied to it. This causes an accumulation of charge on the tip of the capillary and due to charge repulsion the solution forms a Taylor cone and then bursts into droplets. The solvent evaporates from the droplets, producing smaller and more charged droplets that eventually explode into the gas phase.

A key characteristic of ESI is that multiply charged ions are produced[207, 208]. This is beneficial for several reasons. Certain mass analyzers can only scan a limited m/z range so large singly charged analytes cannot be detected[202]. However, large analytes with multiple charges have smaller m/z values and may be in the range detectable by the mass analyzer. Smaller m/z ions are also easier to detect[209] and tandem mass spectrometry experiments are improved when multiply charged ions are selected[210]. One disadvantage to having multiply charged ions is that one ion can have more than one charge state and therefore multiple peaks in the same spectrum. The signal from one analyte is spread out over the multiple charge states, which lowers the signal to noise ratio. ESI is also very sensitive to solution conditions. When high amounts of salts are present, the salts get ionized instead of the analyte of interest, complicating or dominating the spectrum[211, 212].

The electrospray process as described above is dependent on the characteristics of the sample liquid, but also on other parameters such as flow rate, and the electric field at the

capillary[213, 214]. Standard ESI sources use flow rates in the microliter per minute range. To desolvate those volumes of sample, a heated desolvation gas is often utilized[215]. When spraying intact proteins, the higher flow rate combined with the use of a heated desolvation gas causes the protein to become denatured. After ionization and desolvation, many of the ions do not enter into the mass spectrometer and are lost. To solve this problem, nanoelectrospray (nESI) was developed[216]. nESI uses a smaller diameter capillary, nanoliter per minute flow rates or no flow rate at all, very little desolvation gas, and a smaller potential across the capillary compared to ESI. These parameter changes maintain protein structure that is more native-like in the gas phase which is a critical component for native mass spectrometry[217, 218] as discussed in section 1.4.1.

1.3.1.2 Matrix-Assisted Laser Desorption Ionization. In MALDI, the analyte must be dissolved in a solution with a matrix that absorbs electromagnetic radiation very strongly at the wavelength of the instrument's laser[219]. A drop of the solution is deposited onto a stainless steel plate and allowed to thoroughly dry. The solvent evaporates causing the matrix and the analyte to cocrystallize. The sample is usually placed under a vacuum and irradiated by the laser[200, 220] as shown in Figure 1.3. MALDI is not completely understood, however, the matrix absorbs the energy emitted by the laser and is rapidly heated, causing the sublimation of the matrix and analyte.[202] Proton transfer from the matrix to the analyte can occur in the solid phase, during heating, during sublimation, or subsequently during desolvation from a variety of mechanisms.[221, 222]

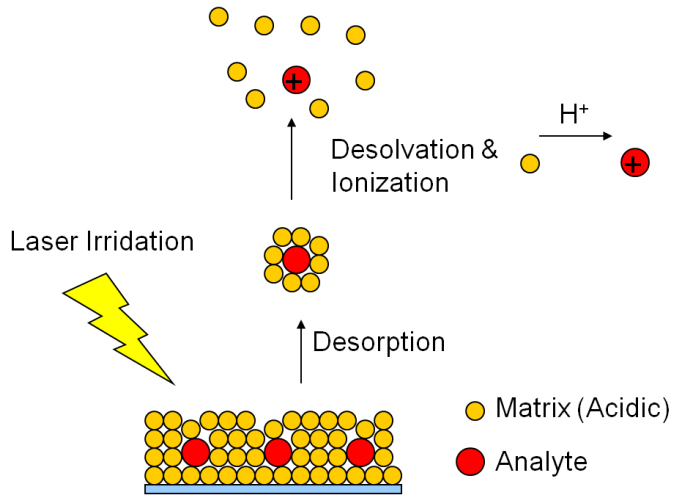


Figure 1.3. A cartoon depicting MALDI. The matrix and analyte are cocrystallized on a stainless steel target. The crystals are irradiated by a laser which causes desorption of the crystals.

Unlike ESI, MALDI mainly forms singly charged ions[200]. If the concentration of an ion is high enough or if an entire protein is ionized, multiply charged ions are formed, but they are the minority. Recently there has been work to obtain multiply charged ions more routinely in MALDI[223, 224] but tandem mass spectrometry experiments requiring multiply charged ions cannot be routinely completed using a MALDI source. Reproducibility between laser shots can be a problem depending on the quality of the matrix crystals but MALDI is much less sensitive to salts and other contaminants compared to ESI. The signal to noise ratio can be higher than ESI because each analyte has only one charge state.

1.3.2 Mass Analyzers. Once the analyte is in the gas phase and is ionized, it is then transferred into the mass analyzer. Mass analyzers are used for separating ions based on their mass to charge ratio and therefore, are critical parts of the instrument. An ideal mass spectrometer would have high resolution, high mass accuracy, high sensitivity, large dynamic range, fast response time, and be inexpensive. Not all of these criteria, however, can be simultaneously achieved. Time of flight (TOF) mass analyzers are beneficial because they can

measure very high m/z values, and thus are well suited to analyze large protein complexes[225-227]. They have midrange resolving capabilities and price and are capable of high mass accuracy[228, 229]. Quadrupoles and ion traps can be cheaper instruments but most do not scan to very high m/z values because of the nature of the analyzer[230] and m/z values higher than 4,000 are usually not observed. The mass accuracy of these analyzers tends to be in the low hundreds of parts per million (ppm) range but these are used frequently for proteomics work. Orbitrap[231] and Fourier Transform Ion Cyclotron Resonance (FTICR)[232, 233] mass spectrometers are useful because they have very high mass accuracy, (< 5 ppm) and the resolution is much greater than the other mass analyzers. The high mass accuracy and resolving power are needed to distinguish between two components with similar mass, for example glutamate and lysine residues or between acetyl lysine and tryimethyllysine[234]. However, these instruments have a limited mass range and are more expensive to initially purchase and maintain. Table 1.4 summarizes the differences between these four mass analyzers. Many mass spectrometers now are hybrid mass spectrometers that include multiple analyzers so that multiple types of experiments can be completed on a single instrument; examples include triple quadrupoles[235], TOF-TOFs, and Q-TOFs[236].

Table 1.4. **A comparison of mass analyzers.**

	Quadrupole	Ion Trap	TOF	FTICR	Orbitrap
Mass Limit (m/z)	4,000	4,000	>1,000,000	4,000	4,000
Resolution (m/z 1000)	2,000	4,000	20,000	100,000	60,000
Mass Accuracy (ppm)	100	100	10	<1.2	<3

1.3.3 Tandem Mass Spectrometry. Mass spectrometry is initially used to measure the mass to charge ratio of an analyte, however, structural information of the analyte is not obtained from this data. When an analyte is activated, fragmentation can occur and yield product ions in a

tandem mass spectrometry or MS/MS experiment[237]. The product ions are then analyzed by an additional mass analyzer. Depending on the activation method used, specific types bonds will be cleaved, which yields valuable information about the structure of the analyte. Those second generation product ions can be further isolated and fragmented in an MS^n experiment, where n represents the generation of ions analyzed.

1.3.4 Activation Methods. Many activation methods have been developed over the past 50 years, but the most commonly used methods for peptide analysis and the methods that were used in this thesis include collision induced dissociation (CID)[238], higher energy c-trap dissociation (HCD)[239], electron capture dissociation (ECD)[240], infrared multiphoton dissociation (IRMPD)[241] and in-source decay (ISD). When peptides are fragmented, they usually fragment along the peptide backbone as shown in Figure 1.4[242, 243]. Which bond is cleaved depends on the activation method. During a tandem mass spectrometry experiment, many molecules are simultaneously being fragmented, thus cleavage will not occur at only one position. Cleavage will occur at only one type of bond, for example, at the amide bond, but every amide bond along the backbone can be cleaved as shown in Figure 1.5. Therefore, the difference between two of one type of ion is a mass of an amino acid residue. The amino acid sequence of the peptide can be determined by finding the sequence of ions separated by masses of the amino acids[244]. (Figure 1.5) It is beneficial to use multiple activation methods to get several types of ions, which yields more structural information.

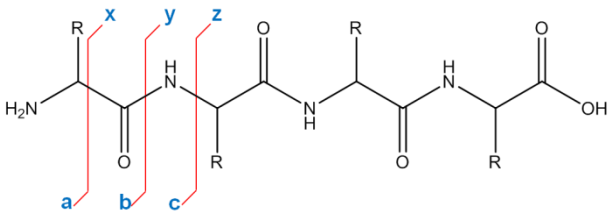


Figure 1.4. Fragmentation of peptides. When peptides are activated, they fragment along the peptide backbone. When cleavage occurs at the amide bond, the fragment containing the c-terminus is known as a y-ion and the fragment containing the n-terminus is known as a b-ion. The other two pairs of ions are produced from cleavages at different bonds along the backbone as shown.

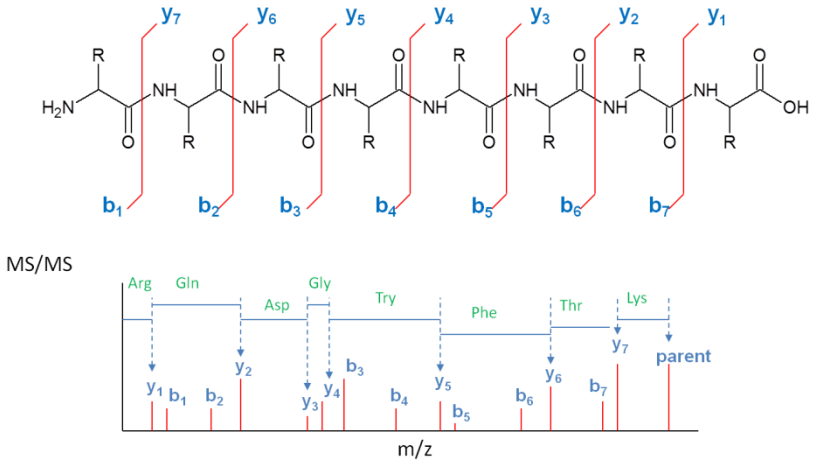


Figure 1.5. Peptide sequencing. Using one type of activation method, fragmentation occurs primarily at the amide bond of the peptide, creating multiple ions that differ in mass by an amino acid residue mass. A sequence of those ions can be found in the spectrum which is used to identify the sequence of the peptide.

CID is the most commonly used activation method today[245-247] and is the workhorse for many types of proteomic experiments. In most CID experiments, the ions to be fragmented are guided by an electric field into a collision cell containing an inert gas such as argon or nitrogen. Over time, the ions will collide with the inert gas which causes the ion's kinetic energy to be converted into internal energy. This excess of internal energy results in fragmentation of the precursor[248]. Depending on the initial kinetic energy of the precursor, CID can be categorized

as low energy or high energy. CID on a TOF-TOF mass spectrometer is known as high energy CID because the collision voltages are in the kiloelectron volt (keV) range[249] while on an orbitrap or FTICR, the energy is in the 10s of electronvolt (eV) range[250].

When positively charge peptides are being activated, primarily b-, y-, and occasionally a-ions are produced after fragmentation using CID[244]. A mechanism has been suggested for this fragmentation pattern known as the mobile proton model[210]. Although the most basic sites are likely to be the location of positive charge, the amide nitrogen is a potential site of positive charge. When the amide nitrogen is protonated, the amide bond is weakened and fragmentation occurs at that bond[251] leading to b- and y- ions. Intramolecular proton transfers can occur so that fragmentation occurs all along the peptide backbone[210, 252, 253]. Hydrogen bonding between the amide nitrogen of peptide backbone and a localized site of positive charge is also sufficient to weaken the amide bond, and cause fragmentation upon gas phase activation.

Photodissociation occurs when ions absorb photons. In IRMPD, an IR laser generates low energy photons. The precursor ions must absorb hundreds to thousands of low energy photons to increase the internal energy of the ion enough for dissociation to occur. Upon dissociation, peptides produce mainly b- and y-ions[241]. IRMPD is typically done in trapping instruments, i.e. FTICRs because of the capability of those instruments to store ions for longer periods of time. The ions need to be stored for a long enough period of time to absorb many photons. A key benefit to using IRMPD is that the energy added to the precursor ions can be carefully controlled and the efficiencies of dissociation may be higher than typical CID[241].

ECD utilizes a very different mechanism for dissociation[240]. Multiply charged ions are irradiated with low energy electrons. The precursor ions capture the electrons, which lowers the

overall ion charge, raising the internal energy which causes dissociation. Dissociation occurs faster than the randomization of the electron over the entire protein or peptide, so c- and z- ions are produced (Figure 1.4). ECD is a lower energy technique and has been especially useful in phosphoproteomics work[254-256] to maintain and identify the site of phosphorylation.

As previously discussed, ESI and MALDI are known as soft ionization techniques, meaning the structure of the analyte remains intact. Although this is mostly true, fragmentation can still occur during ionization. When this occurs, it is known as in-source decay (ISD). Ionization conditions such as voltages and temperatures can be adjusted to make ISD occur more often in ESI[257-260] but this type of fragmentation occurs mostly in MALDI[261]. In certain mass spectrometers, usually TOFs, fragmentation can occur after ionization when the ion is in the flight tube. This post-source decay (PSD) occurs on metastable ions or ions that are stable enough to leave the ion source but still have excess energy resulting in fragmentation[249, 262, 263]. When ions undergo PSD, they are observed at the m/z of precursor, but with significantly decreased resolution. Ions resulting from ISD are observed at the m/z of the daughter ion, not the precursor. ISD generally produces z- and c- type fragments[264] and has been used to fragment intact proteins[265]. Typically, ISD has a low efficiency of fragmentation, although this is somewhat matrix dependent[265].

1.4 Mass Spectrometry for Protein Structure and Interactions

Multiple ionization methods, mass analyzers and activation methods make mass spectrometry a versatile technique useful for characterizing many types of analytes. This versatility extends into the analysis of protein structure and protein interactions. Mass spectrometry is useful for analyzing structures and complexes that other methods cannot access. There are several mass spectrometry methods used for analysis of these complex problems.

Native mass spectrometry, ion mobility mass spectrometry (IMS), hydrogen-deuterium exchange (HDX) and radical foot printing, and chemical crosslinking are examples of these methods discussed below.

1.4.1 Native Mass Spectrometry. Over the past few decades, native mass spectrometry has become increasingly used for studying protein interactions. In this method, an entire complex is introduced into the gas phase under very gentle spray conditions which retain a native-like structure.[205, 266, 267] The mass of the intact structure is obtained which reveals stoichiometry information of a complex and further dissociation experiments can yield information on the identity of the components in the complex[268-271]. Huge complexes have been analyzed using these methods and data regarding proteins structure[272-274], oligonucleotide complexes[275, 276], and protein-ligand complexes have been obtained[277-279]. A major limitation with dissociation experiments is that one subunit, usually a small one at the outer edge of the complex, tends to dissociate predominantly. Dissociation occurs more efficiently with increasing charge, but when the first dissociation occurs, the small subunit captures a significant amount of the charge, making further dissociations difficult[280]. As an example, in the case of the tetradecamer GroEL, after two activation events where two monomers were dissociated, the charge of the complex was too small to allow any further dissociation[281]. Limitations do include the need for the sample to be of high quality, without protein degradation[282] and for optimization of ionization conditions retaining the native-like state which can be labor intensive. The concentration of the protein needed is often high, but unlike traditional x-ray crystallography approach only small amounts of protein are needed. Protein flexibility and heterogeneity are not limiting factors. Native mass spectrometry is also not limited by molecular weight of the complex unlike in NMR.

1.4.2 Hydrogen-Deuterium Exchange. HDX is frequently used to determine which parts of a protein are exposed to solvent and which are inaccessible because of protein structure, folding or interactions[190]. Protons found on the amide backbone of a protein and protons on the side chains of certain amino acids are exchangeable protons, meaning water molecules surrounding the protein can accept a proton at the same time hydronium ions in solution can donate a proton[283]. The rate at which this exchange occurs is proportional to the acidity of the proton and to the amount of solvent accessibility that particular proton has. Because of these reasons, amide protons have rates that can be easily measured by mass spectrometry while the protons of amino acid side chains do not. Also, the exchangeable protons that are buried due to the protein's structure or within the interacting surfaces of a complex will not be exchanged as frequently as a proton that is solvent exposed[284, 285]. When deuterated water is used as a solvent, as the exchange is occurring, the exchangeable protons that are exposed to the deuterated water will be replaced with deuterons and the protein will increase in mass accordingly. In an HDX experiment, the sample will be incubated with deuterated water for a specific amount of time and then analyzed to determine the extent of deuterium incorporation. Mass spectrometry can then be used to analyze the amount of deuterium incorporation into the entire protein, and fragmentation can localize the region of deuterium incorporation. To determine the specific residues with higher rates of exchange, the protein is proteolytically digested into peptides with pepsin (an enzyme active at pH 2) and then analyzed by mass spectrometry. Usually high performance liquid chromatography (HPLC) is used to separate the peptides prior to mass spectrometry. This entire process must be completed very quickly because the protons that have been exchanged with deuterons can back-exchange with the solvents used for HPLC and MS[286]. The mass of the peptide can yield information as to how

many deuterium atoms are incorporated into that peptide, but usually tandem mass spectrometry is used to determine the exchange rates of each amide bond. These HDX methods have been used successfully to gain structural or folding information on many proteins including cytochrome c[287, 288], aldolase[289, 290], cellular retinoic acid-binding protein I[291], IGF1 binding protein[292], and ERK2[293].

The main drawback with HDX experiments is the high rate of back-exchange which varies from 10-50% and have detrimental effects on data quality[294]. The rate of hydrogen exchange is dependent upon pH and temperature, and therefore to minimize further exchange during proteolytic digestion and mass spectrometry, the pH is usually lowered to 2-3, and the temperature to 0°C[285, 295]. These deviations from biologically relevant conditions often results in complete denaturation of the protein[294] and back-exchange still occurs, although at a much slower rate. The second major problem with HDX experiments is with deuterium scrambling. CID is the most commonly used method for ion fragmentation, however, recent reports have shown that the CID process can result in scrambling of the deuterium atoms because of the mobile proton[296-298]. This greatly reduces the reliability of HDX data. To overcome this problem, several groups have started using electron based fragmentation techniques. Electron based fragmentation techniques are not dependent on a mobile proton for fragmentation, therefore, ECD or ETD data are much more reliable without deuterium scrambling[299, 300]. Higher order structural information such as stoichiometry cannot be determined by HDX because there is no difference in the exchange rate of an area protected due to monomeric folding versus tetrameric folding. HDX is not a high throughput method and is not routinely completed *in vivo*.

1.4.3 Ion Mobility Mass Spectrometry (IM-MS). IM-MS utilizes ion mobility spectrometry prior to mass spectrometry by first separating the ions by size, shape, and charge in

the ion mobility cell, then further separating them by the mass to charge ratio in the mass spectrometer. The separation of ions by ion mobility takes approximately of 10-60 ms and to accumulate one mass spectrum takes approximately 100 μs[301]. For each ion mobility separation accomplished, many mass spectra are acquired and a two dimensional plot showing drift time versus m/z is produced. Ion mobility provides an additional measure of separation, but most importantly, IM-MS provides information about the size and shape of an ion through collision cross-section (CCS) measurements. Ions that have the same m/z ratio but have different sizes or conformations will have different CCSs, therefore different conformations of ions can be distinguished[302]. For example, β-amyloid (Aβ) is a peptide that when it oligomerizes causes debilitating diseases like Alzheimer's[303]. IM-MS has been crucial in the characterization of different Aβ oligomeric conformations, mechanism of their formation, and in developing targets to inhibit formation[304-307], all of which is necessary progress to determine the cause of toxicity, and ultimately treatment and prevention.

An ion mobility mass spectrometer has the same components as a regular mass spectrometer – an ionization source, analyzer, and detector, but prior to the analyzer, there must also be a component that separates ions by mobility. In most instruments used for biomolecule work, the ionization source is an electrospray[308-311] or nano-electrospray source which introduces the complex into the gas phase in a native-like state[205, 266, 267]. Analyzers can vary[312-314] but TOFs are well suited to the analysis of intact protein complexes because the range of m/z is not limited. There are four types of ion mobility spectrometers that have been combined with mass spectrometers: aspirator[314], drift time[315], traveling wave (T-wave)[316], and differential[317]. The majority of data on multi-protein complexes using IM-MS have been

produced using T-wave instruments as were the data in this thesis, therefore only T-wave ion mobility instruments will be discussed in further detail.

In an IM-MS experiment, once the complex is ionized, it is injected into an ion mobility cell. The cell is under the influence an electric field causing the complex to travel through the cell however, the cell contains inert neutral gas molecules which collide with the complex as it travels. Larger ions will collide more frequently with the neutral gases in the cell resulting in a larger CCS compared to more compact ions with a similar mass.[318] In the past, ion mobility has had limited sensitivity for two main reasons. As the ions traveled through the drift cell, they radially diffused sufficiently enough that many ions did not enter the mass analyzer and subsequently get detected. Secondly sensitivity was limited by the low duty cycle because the second pulse of ions could not enter the drift tube until the first set has been detected. One way that to solve these problems is to use a traveling wave ion mobility cell.

In T-wave ion guides, the applied electric field is not constant. Instead, the ions are injected into a series of ring electrodes known as a stacked ring ion guide (SRIG). In a SRIG each ring electrode has a radio frequency (RF) voltage applied to it and adjacent rings have the opposite phase RF voltage. This causes the ions to stay within the radial dimensions of the ring[319] and when a gas is present, as in the case of an ion mobility cell, the axial motion of the ions is also stopped or severely limited. The ions are moved through the SRIG by applying a brief direct current (DC) voltage to a pair of rings in the SRIG. DC voltage is applied to the next set of rings in the SRIG creating a traveling wave of voltage[320]. The ions are moved by the DC through the SRIG. When the ions collide with the neutral gas molecules in the ion mobility cell, their progress through the SRIG is slowed or stopped, which allows the DC voltage to move past the ions. This event will happen more often with ions with larger CCS values (and a given m/z),

therefore these ions will have longer drift times and are separated by CCS. To determine the CCS of an unknown ion, calibration of the instrument must be completed using known drift times and CCSs of known complexes[321].

IM-MS has been used to determine the size of complexes, the stoichiometry in a complex[322, 323], and to estimate dissociation constants following tandem mass spectrometry experiments[324]. However, one of the key issues in IM-MS is the relationship between the gas phase structures and the solution phase structures. The electrospray process, loss of surrounding solvent, and the subsequent analyses can cause the protein structure to be affected. Ionized proteins have Coulombic repulsions between the charges on the protein which cause unfolding but bulk solvent is useful for mediating those charges through hydrogen bonding and electrostatic interactions. The surrounding solvent contributes to the structure of a protein when it's in solution[325-327], but during an IM-MS experiment, most of the bulk solvent is removed and the protein can then easily unfold. The energies required to move ions from the source into the instrument can also cause proteins to lose their quaternary structure. Whenever unfolding occurs, the protein can stay unfolded or morph into new structures that are compact, but not equal to the native structure[328, 329]. To combat this issue, efforts are being made to maintain and stabilize the in-solution structure in the gas phase. For example, additives are added to maintain the folded structure - either in the gas-phase[330] or before ionization[331, 332].

Currently, the IM-MS analysis of protein complexes is not a high throughput method. Ionization conditions that retain a compact structure require significant amounts of time to develop and there are limited computational tools, making data analysis labor intensive and time consuming[333]. Sensitivity is improving and now only small volumes of a complex of low micromolar concentrations are needed. This is requires much less material than x-ray

crystallography or NMR, but the sensitivity is not as good as FRET or other mass spectrometric techniques.

1.4.4 Chemical Crosslinking. Crosslinking protocols have been in use since the early 1950s[334-337] to study protein structure, and since the early 1960s to study structures of protein with DNA[338-340] and RNA[341, 342]. In a crosslinking experiment a protein or complex is subjected to conditions (usually addition of a crosslinking reagent or exposure to ultraviolet (UV) light) which cause covalent bonds to be formed between adjacent parts of the protein or complex. In early protein crosslinking experiments, samples were analyzed by sodium-dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)[343] or western blots[335]. SDS-PAGE is useful for approximating the molecular weight of a denatured protein. If higher molecular weights than the monomers are observed after crosslinking, then those higher molecular weights are attributed to two subunits being adjacent to each other in a complex[343].

Crosslinking has increased in popularity over the past decade, both to study protein quaternary structure and to study protein-protein interactions. The advances in mass spectrometry have fueled the surge in crosslinking studies because MS is a sensitive detection method for analyzing crosslinking data. The combination of chemical crosslinking with mass spectrometry has several benefits compared to other technologies. After crosslinking, further purification steps can be taken to isolate the protein of interest without losing weaker binding partners because crosslinking covalently locks protein interactions in place[344]. Interacting partners can be determined and spatial information can be obtained. Relatively small quantities of protein are required, and data can be rapidly generated from crosslinking experiments. Quantification of protein-protein interactions can also be completed using a combination of light and heavy isotope crosslinking reagents[345]. Therefore, crosslinking combined with mass

spectrometry is a sensitive and well-suited technique to analyze protein quaternary structure and protein-protein interactions in a quantitative and high-throughput manner.

Chemical crosslinking combined with mass spectrometry has been used in several key recent studies. Juri Rappsilber and coworkers used crosslinking to further analyze RNA polymerase II (Pol II) with initiation factors bound[346]. The crystal structure of Pol II was available[347], but the structure of the complex with the appropriate initiation factors was not known. In their paper, Rappsilber crosslinked Pol II (a 12 subunit protein) with a 3 subunit transcription factor, TFIIF and observed hundreds of crosslinks. These crosslinks allowed them to predict the overall architecture of TFIIF and to suggest its location on the structure of Pol II. These data, combined with other data in the literature allowed the authors to suggest a mechanism for TFIIF function in suppressing non-specific DNA binding[346]. James Bruce's laboratory has completed several studies using crosslinking reagents *in vivo* to analyze protein interactions[348-352]. In their initial study, Bruce and coworkers crosslinked whole cells and then did CoIPs for two proteins crucial for anaerobic respiration of metal hydroxides and oxides (OmcA and MrtC) in *Shewanella oneidensis*[351]. When using crosslinking, several new proteins were observed that appeared to be associated with the OmcA-MtrC interaction network.

One of the first and most commonly used crosslinkers is BS3 (Figure 1.6). BS3 and its analogs are commercially available from multiple sources and have been used in several studies to analyze protein structure[353-357] including the RNA Polymerase II study previously discussed[346]. BS3 is a water soluble, amine reactive crosslinker. The behavior of BS3-crosslinked peptides in the mass spectrometer has been extensively characterized[358-361] and multiple algorithms and software programs have been developed to analyze crosslinked datasets[345, 362-364]. Heavy isotope versions of BS3 and its analogs are also commercially

available and many studies use a combination of the light and heavy versions to facilitate identification of the crosslinked peptides[344, 365-370]. The heavy crosslinkers are usually deuterated, which significantly increases the expense of the experiment and can affect chromatographic separation.[344, 345, 371, 372] Additional programs are available to analyze data utilizing the combination of the light and heavy crosslinkers[372-374]. However, even with the software programs available, data analysis is still difficult, the false positive rates and high false negative rates remain high, and the intrinsic challenges associated with chemical crosslinking persist.

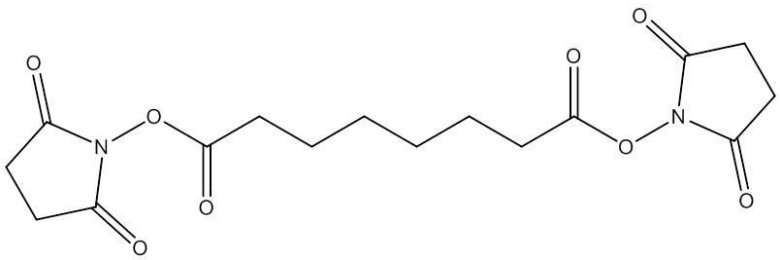


Figure 1.6. Structure of BS3. BS3 one of the most commonly used, amine reactive crosslinkers.

A number of challenges have been traditionally associated with mass spectrometric analysis of crosslinked peptides.[344, 345, 375, 376] Sample complexity is increased by the presence of side reactions, intracrosslinked peptides (type 1 – see Figure 1.7) and dead-end reactions (type 0 in which one end of the crosslinker reacts with protein and the other end with a nucleophile (e.g., water)).[360, 377] True crosslinked peptides (type 2 crosslinks) in a proteolytic digest are generally very low in abundance relative to noncrosslinked peptides. Tandem mass spectra of crosslinked peptides often undergo diminished fragmentation relative to linear peptides by collision-induced dissociation or are inherently complex due to overlapping b- and y- ion series from both peptides.[376, 378]

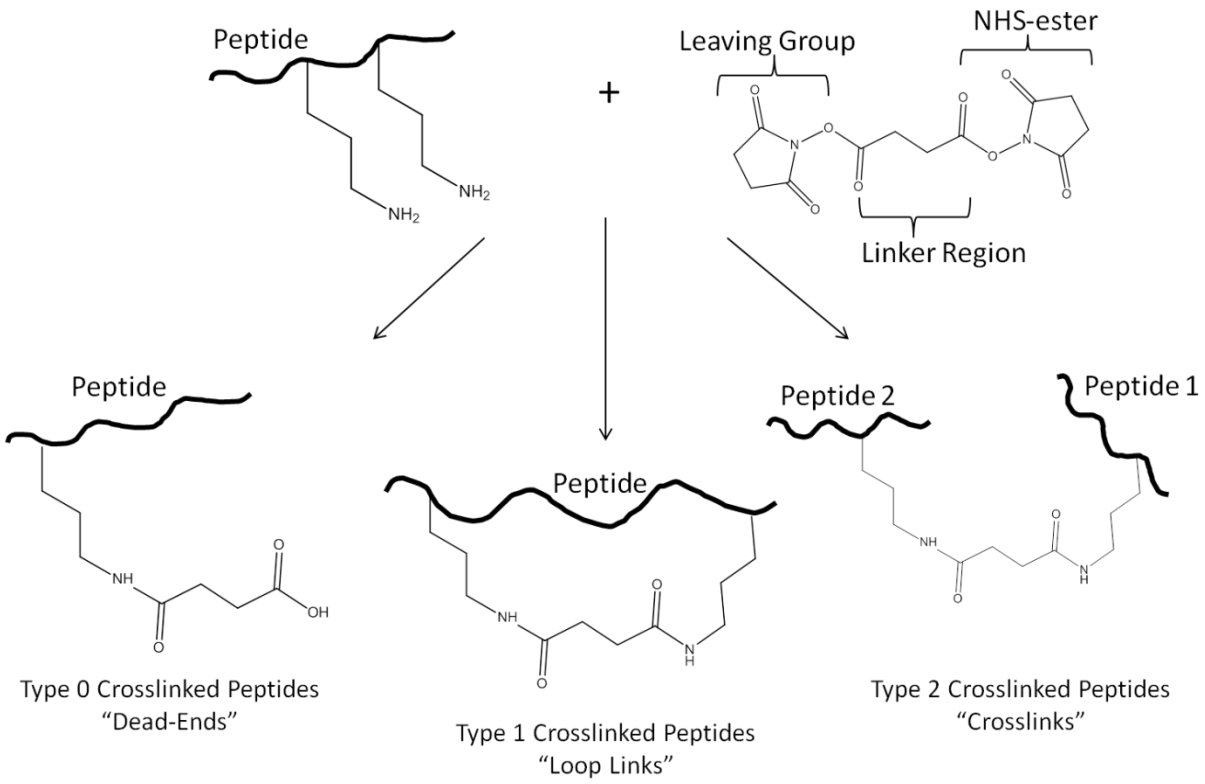


Figure 1.7. Products from crosslinking reactions. During a crosslinking reaction, three types of modifications can occur, resulting in three types of peptide products. Dead-end products (Type 0, left) result when the crosslinking reagent reacts on one end with protein and does not couple on the other end or reacts with a small molecule (e.g., water) on the other end. Loop links (type 1, middle) occur when both ends of the crosslinking reagent react within the protein but no proteolytic cleavage occurs between the two sites. When proteolytic cleavage occurs between the two sites or the sites are on different proteins, an authentic crosslinked peptide (type 2, right) is the result.

The nature of the crosslinking chemistry can also introduce challenges. For example, N-hydroxysuccinimide (NHS) ester is a common reactive group used to crosslink proteins and it reacts primarily with free amines such as the epsilon nitrogen in lysyl residues. Other nucleophilic residues such as seryl, threonyl and tyrosyl residues react to a much lesser extent[359, 379]. At physiological pH, lysine residues are protonated, but after crosslinking with NHS esters, they are less likely to retain the positive charge. This change in charge can significantly affect

protein structure. Proteins crosslinked at lysyl residues also have fewer sites available for tryptic digestion, resulting in larger peptides. These larger peptides may not fragment as effectively by collision-induced decay (CID) and they may have complex fragmentation patterns making the spectra difficult to identify and interpret[375, 379]. Because of these reasons, data analysis is currently a major bottleneck to high throughput analyses of crosslinked proteins. Even with the algorithms currently available, most crosslinked spectra are manually validated to lower the high false positive rate[380].

To overcome these obstacles, crosslinking reagents have been developed commercially and by various research groups. Properties of these new crosslinkers include a broader range of chemical reactivities, the addition of affinity ligands and, improved fragmentation properties in MS/MS[351, 352, 369, 381-391]. Crosslinking reagents have been developed that include maleimide groups to react specifically with cysteine residues[392], or nonspecific reactive groups to react with a range of residues[393-396]. Free cysteine residues are not as abundant in proteins as lysine residues which suggests maleimide reactive groups have less opportunity to result in crosslinks than the NHS esters. Nonspecific crosslinkers can theoretically produce more crosslinks because they can react with many residues. Data analysis from experiments using nonspecific crosslinkers is even more difficult and the amount of any one crosslinked peptide is reduced. In general, nonspecific crosslinkers tend not to be water soluble because they contain aryl azides, benzophenone derivatives or other organic moieties. The organic solvents required to solubilize these crosslinkers can directly affect protein interactions and structures, leading to experimental artifacts. There are, however, several crosslinkers that contain biotin affinity tags[385, 397, 398] or that utilize azide and alkyne chemistry[391, 399] to purify crosslinked peptides. Most of these affinity methods also target the type 1 crosslinked peptides (dead-ends) which are in large excess

over the type 2 crosslinked peptides, and thus the informative peptides still constitute a minor population of the enriched peptides. The addition of these affinity tags also contributes to the insolubility problem.

To overcome the poor fragmentation properties of crosslinked peptides, both chemically cleavable and mass spectrometry cleavable crosslinking reagents have been developed[400-403]. Chemical cleavage of the crosslinked peptides prior to mass spectrometry leads to ambiguity regarding which two cleaved peptides were crosslinked, particularly in complex mixtures. More desirable are crosslinkers that fragment efficiently in MS/MS mode to yield two major fragment ions corresponding to the component peptides which can subsequently be subjected to MS^3 for identification. Thus, MS/MS provides a clear precursor/products relationship indicating unambiguously that the two product peptides were covalently crosslinked. Several recent efforts have addressed the design of these MS-cleavable crosslinking reagents, leading to new crosslinkers with improved functionality,[352, 378, 404-411] including electron transfer dissociation (ETD) cleavable reagents[369, 412] but further improvements are still needed.

A major bottleneck with crosslinking combined with mass spectrometry is identification of the rare crosslinked peptides. This effort has been facilitated by use of a mixture of heavy and light isotopically labeled crosslinking reagents and chemical tags[368]. Mass analysis of the peptides identifies ions separated by the distinctive mass corresponding to the number of hydrogen atoms substituted by deuterium[344, 366, 367, 413]. Dead-end peptides also produce pairs of ions and computational schemes have been developed which filter these spectra out as part of the identification process[365, 371, 372]. Crosslinked peptides (type 2) that produce marker ions upon fragmentation facilitate the identification of the spectra and to our knowledge few crosslinkers have been identified to do so, specifically with type 2 peptides.[358, 360]

Numerous efforts have been made to develop new crosslinkers including several key contributions. James Bruce and colleagues have developed a set of crosslinkers[313, 406, 409], one of which is shown in Figure 1.8A. These crosslinkers are amine reactive, mass spectrometry cleavable, membrane permeable, and contain a biotin tag for affinity purifying crosslinked peptides, thus overcoming several of the major problems associated with crosslinking studies. These crosslinkers have been successfully used to probe the membrane proteome of *Shewanella oneidensis* MR-1[414] including the OmcA and MrtC interaction network[351]. More recently, Bruce and colleagues have used their crosslinkers to probe interactions *in vivo*[415-417] and have identified many potential interacting partners. Two major drawbacks to these types of crosslinkers should be considered. First, these crosslinkers are not water soluble and the organic solvent needed to solubilize the crosslinkers may be causing false interactions. Second, the arm spans of these crosslinkers are so long that a higher proportion of nonspecific interactions may be observed.

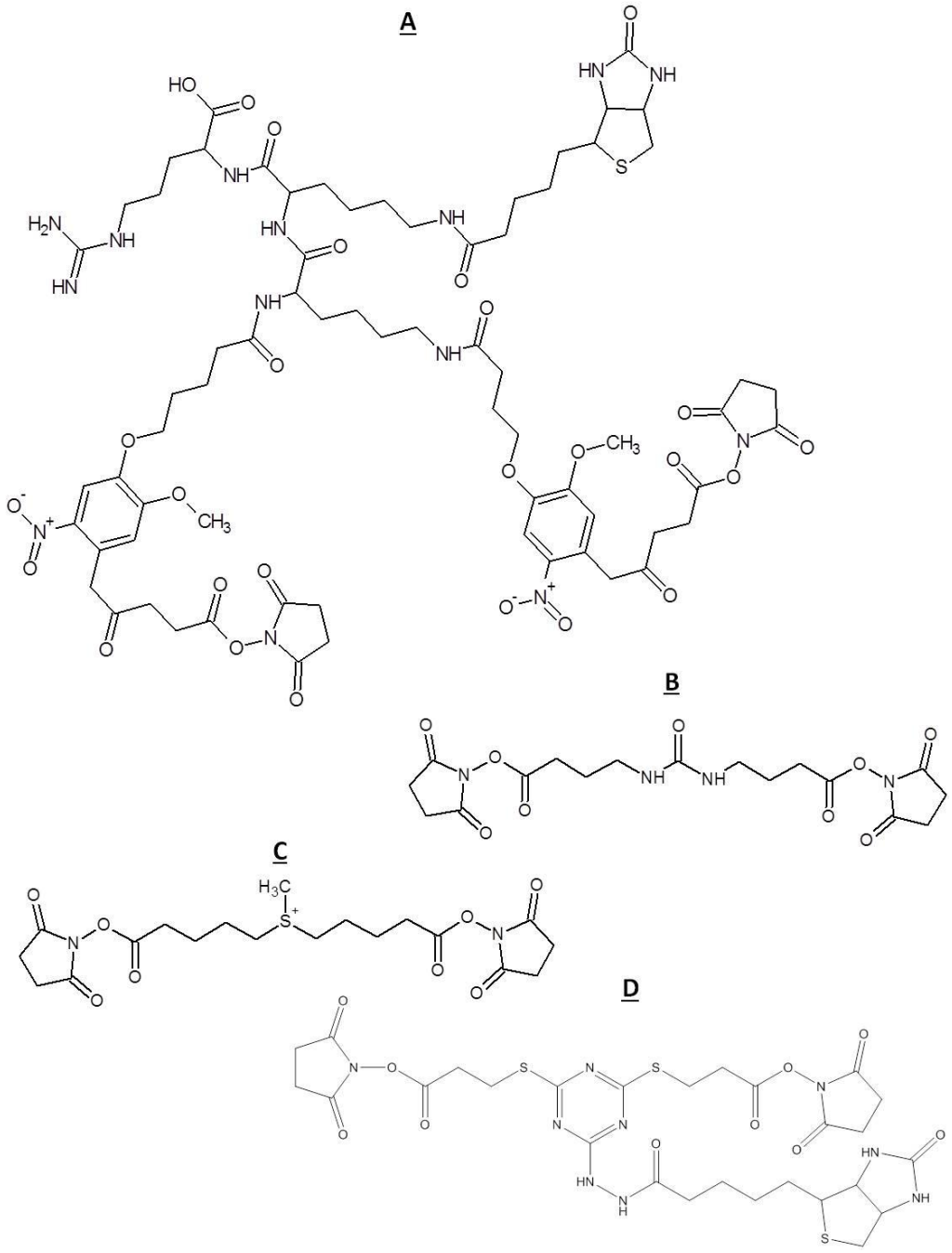


Figure 1.8. Structures of recently developed novel crosslinkers. A sample of the novel mass spectrometry cleavable crosslinkers that have been recently designed for protein structure and interaction[397, 403, 407, 414].

Andrea Sinz and coworkers have designed several mass spectrometry cleavable crosslinkers[383, 387, 410], the most recent of which is shown in Figure 1.8B[418]. This urea based crosslinker is amine reactive, mass spectrometry cleavable, and produces characteristic neutral

losses upon fragmentation in a MALDI-TOF-TOF mass spectrometer which facilitates identification of the crosslinked peptides. Gavin Reid and coworkers designed a sulfonium crosslinker[407] (Figure 1.8C) with interesting properties. Multiple crosslinkers were published by Petrotchenko and coworkers[404]; Figure 1.8D contains the structure of the most recent[397]. Each of these reagents is mass spectrometry cleavable, but they are not soluble in water and after crosslinking with free amines, they remove a positive charge on the protein. These details can affect protein structure (as previously discussed) and should be addressed in crosslinker design.

Balancing the properties of the reagents to obtain good reactivity in solution and stability in MS mode while maximizing fragmentation of the crosslinker for a variety of peptides in MS/MS mode is challenging. Many of the crosslinking studies use model proteins but publications analyzing crosslinks of more complex systems are only recently starting to appear, in part because of computational limitations in data analysis and also because reagents that work for model peptides do not necessarily work well for the diverse peptides derived from intact proteins[346, 351]. The purpose of this thesis was to develop new crosslinkers to overcome the current difficulties and then apply them to both model systems and more unknown complexes.

1.5 Dissertation Research Summary

The first goal of this dissertation was to design a crosslinker which overcame the intrinsic difficulties associated with combining chemical crosslinking and mass spectrometry. The first crosslinker, PC1, was designed and synthesized. The synthesis of PC1 was completed by the Vahlteich Medicinal Chemistry Core at the University of Michigan, all other synthetic work was completed by me. In Chapter 2 of this thesis, the fragmentation of PC1 was extensively characterized acetylated model peptides in multiple mass spectrometers with several

fragmentation mechanisms. The initial results suggested that PC1 had several of the characteristics needed for a successful crosslinker. The PC1 crosslinked model peptides cleaved via mass spectrometry and generated marker ions to facilitate identification of the crosslinked peptides. Next, the crosslinker was applied to a more complex system, the tetrameric protein aldolase, and the fragmentation was found to be insufficient.

In Chapter 3, the goal was to adjust the design of PC1 to make the crosslinked peptides more labile in the mass spectrometer. The central moiety of PC1 was a piperazine ring and the initial goal was to methylate the piperazine nitrogen atoms. However, the products from the methylation reactions were not identifiable. Instead I synthesized a longer crosslinker containing 1,4-diazabicyclo[2.2.2]octane as the central moiety. Once the initial synthesis was developed, three crosslinkers of increasing lengths DC4, DC5, and DC6, and two protein tags were synthesized.

In Chapter 4, the crosslinkers DC4, DC5, and DC6 were characterized. First, DC4 was applied to the acetylated model peptides and the fragmentation observed on a MALDI-TOF-TOF and on an electrospray orbitrap mass spectrometer. Initial fragmentation events resulted in cleavage on either side of the positive charges so crosslinked peptides are identified as pairs of ions separated by defined masses. The structures of the component peptides were then able to be robustly determined by MS^3 because their fragmentation products rearrange to generate a mobile proton. The fragmentation behavior of DC5, and DC6 was also characterized. Activation of DC4-, DC5-, and DC6-crosslinked peptides resulted in a similar fragmentation pattern where cleavage occurred at the sites of positive charges, but to different extent. Peptides crosslinked with DC5 were less labile than peptides crosslinked with DC4, and DC6 crosslinked peptides were even less labile. Because DC4 was the more labile crosslinker and therefore more likely to

be useful, it was further applied to the tetrameric protein aldolase and a bottom-up analysis was completed. Several crosslinked peptides were observed that were in agreement with the known flexibility and crystal structure of aldolase. Residues that were deeply buried and therefore inaccessible to crosslinking were not observed as modified. The results from Chapter 4 were very promising because we had developed a crosslinker that is mass spectrometry cleavable, very soluble in water, retains the positive charge on the protein due to the quaternary amines, and fragments in such a way that a computational algorithm could be written to identify crosslinked peptides from a complex mixture.

The bottom-up approach was so successful and DC4 such a labile crosslinker, that in Chapter 5, I combined DC4 crosslinking with ion mobility mass spectrometry in a top-down approach. It was thought that crosslinking would alleviate some of the challenges associated with top-down mass spectrometry. Each of the four crosslinkers (PC1, DC4, DC5, DC6) were applied to the homotetrameric protein avidin and then analyzed by IM-MS. Compared to the noncrosslinked samples, the collisional cross-section (CCSs), mass, and charge of crosslinked avidin increased. The CCSs for the crosslinked avidin were closer to the in-solution size which was partially due to large amounts of salt adducts.

MS/MS experiments were used to create the collision induced unfolding and dissociation profiles of the DC4-crosslinked avidin. The crosslinked multimers exhibited greater stability than the uncrosslinked avidin, both to unfold and to dissociate. This was likely caused by a combination of salt adducts and the addition of crosslinks. After the dissociation patterns were analyzed, it was clear that crosslinking altered the fragmentation pathways of the crosslinked avidin. Two key observations were made regarding the dissociation profiles. First, after dissociation, a more symmetric distribution of charge was observed than with typical CID of

protein complexes. Secondly, the peptide fragments observed after CID of the DC4-crosslinked avidin tetramer were significantly different than the unmodified avidin. These observations have potential implications to the field of top-down mass spectrometry and are discussed in more detail in chapter 5.

In the last chapter, Chapter 6, I summarize the results of this work, and put it in context of the current literature. There are several remaining questions to be answered and the experiments necessary to answer them are discussed. Lastly, I discuss future directions.

1.6 References:

[1] B. Brodsky, J. A. M. Ramshaw, *Matrix Biology* **1997**, *15*, 545.
[2] R. Z. Kramer, J. Bella, P. Mayville, B. Brodsky, H. M. Berman, *Nature Structural Biology* **1999**, *6*, 454.
[3] G. N. Ramachandran, G. Kartha, *Nature* **1954**, *174*, 269.
[4] J. Monod, J.-P. Changeux, F. Jacob, *Journal of Molecular Biology* **1963**, *6*, 306.
[5] J. Monod, J. Wyman, J.-P. Changeux, *Journal of Molecular Biology* **1965**, *12*, 88.
[6] E. Haber, C. B. Anfinsen, *Journal of Biological Chemistry* **1962**, *237*, 1839.
[7] C. B. Anfinsen, *Science* **1973**, *181*, 223.
[8] M. F. Perutz, H. Muirhead, J. M. Cox, L. C. Goaman, F. S. Mathews, E. L. McGandy, L. E. Webb, *Nature* **1968**, *219*, 29.
[9] M. F. Perutz, M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will, A. C. North, *Nature* **1960**, *185*, 416.
[10] B. C. Wishner, K. B. Ward, E. E. Lattman, W. E. Love, *Journal of Molecular Biology* **1975**, *98*, 179.
[11] W. E. Love, P. M. D. Fitzgerald, J. C. Hanson, W. W. Royer, W. M. Ringle, in *Biochemical and clinical aspects of hemoglobin abnormalities* (Ed.: W. S. Caughey), Academic Press, New York, **1978**, pp. 165.
[12] E. A. Padlan, W. E. Love, *Journal of Biological Chemistry* **1985**, *260*, 8280.
[13] V. M. Ingram, *Nature* **1957**, *180*, 326.
[14] L. Pauling, H. A. Itano, S. J. Singer, I. C. Wells, *Science* **1949**, *110*, 543.
[15] M. F. Perutz, A. M. Liquori, *Nature (London)* **1968**, *219*, 902.
[16] W. E. Love, P. M. D. Fitzgerald, J. C. Hanson, W. E. Royer, in *Development of therapeutic agents for sickle cell disease* (Eds.: J. Rosa, Y. Beuzard, J. Hercules), Elsevier, North-Holland, Amsterdam, **1979**, pp. 65.
[17] E. A. Padlan, W. E. Love, *Journal of Biological Chemistry* **1985**, *260*, 8272.
[18] J. W. Harris, *Proceedings of the Soceity for Experimental Biology and Medicine* **1950**, *75*, 197.
[19] J. Sygusch, D. Beaudry, M. Allaire, *Biochemistry* **1987**, *84*, 7846.
[20] H. Pan, D. L. Smith, *Biochemistry* **2003**, *42*, 5713.
[21] I. A. Rose, E. L. O'Connell, A. H. Mehler, *The Journal of Biological Chemistry* **1965**, *240*, 1758.
[22] D. Tomar, T. Khan, R. R. Singh, S. Mishra, S. Gupta, A. Surolia, D. M. Salunke, *PLoS ONE* **2012**, *7*, e43522.

[23] G. Hagelueken, H. Huang, B. R. Clarke, T. Lebl, C. Whitfield, J. H. Naismith, *Molecular Microbiology* **2012**, n/a.
[24] A. C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Huertier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Searphin, B. Kuster, G. Neubauer, G. Superti-Furga, *Nature* **2002**, *415*, 141.
[25] A. C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M. A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. hudak, A. M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, G. Superti-Furga, *Nature* **2006**, *440*, 631.
[26] Y. Ho, a. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S. L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutillier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A. R. Willems, H. Sassi, P. A. Nielsen, K. J. Ramuseen, J. R. Andersen, L. E. Johansen, L. H. Hensen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B. D. Sorensen, J. Matthisen, R. C. Hendrickson, F. Gleeson, T. Pawson, M. F. Moran, D. Durocher, M. Mann, C. W. Hogue, D. Figeys, M. Tyers, *Nature* **2002**, *415*, 180.
[27] B. Alberts, *Cell* **1998**, *92*, 291.
[28] R. W. Hendrix, *Journal of Molecular Biology* **1979**, *129*, 375.
[29] K. Braig, Z. Otwinowski, R. Hegde, D. C. Boisvert, A. Joachimiak, A. L. Horwich, P. B. Sigler, *Nature* **1994**, *371*, 578.
[30] J. Li, J. Wang, J. Wang, Z. Nawaz, J. Liu, M., J. Qin, J. Wong, *The EMBO Journal* **2000**, *19*, 4342.
[31] N. Quade, C. Mendonca, K. Herbst, A. K. Heroven, C. Ritter, D. W. Heinz, P. Dersch, *Journal of Biological Chemistry* **2012**.
[32] E. Pohl, R. K. Holmes, W. G. J. Hol, *Journal of Molecular Biology* **1999**, *292*, 653.
[33] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251.
[34] K. Luger, T. J. Rechsteiner, A. J. Flaus, M. M. Y. Waye, T. J. Richmond, *Journal of Molecular Biology* **1997**, *272*, 301.
[35] S. C. Lim, M. W. Bowler, T. F. Lai, H. Song, *Nucleic Acids Research* **2012**.
[36] S. R. Price, P. R. Evans, K. Nagai, *Nature* **1998**, *394*, 645.
[37] J. M. Ryter, S. C. Schultz, *The EMBO Journal* **1998**, *17*, 6819.
[38] M. A. Convery, S. Rowsell, N. J. Stonehouse, A. D. Ellington, I. Hirao, J. B. Murray, D. S. Peabody, S. E. V. Phillips, P. G. Stockley, *Nature Structural Biology* **1998**, *5*, 133.
[39] B. A. C. Ackrell, B. Cochran, G. Cecchini, *Archives of Biochemistry and Biophysics* **1989**, *268*, 26.
[40] V. Yankovskaya, R. Horsefield, S. Törnroth, C. Luna-Chavez, H. Miyoshi, C. Léger, B. Byrne, G. Cecchini, S. Iwata, *Science* **2003**, *299*, 700.
[41] C. MacMunn, *Philosophical Transactions of the Royal Society of London* **1886**, *177*, 267.
[42] D. Keilin, *Proceedings of the Royal Society of London, Series B* **1925**, *98*, 312.
[43] R. E. Dickerson, T. Takano, D. Eisenberg, O. B. Kallai, L. Samson, A. Cooper, E. Margoliash, *Journal of Biological Chemistry* **1971**, *246*, 1511.
[44] E. W. Ainscough, A. G. Bingham, A. M. Brodie, W. R. Ellis, H. B. Gray, T. M. Loehr, J. E. Plowman, G. E. Norris, E. N. Baker, *Biochemistry* **1987**, *26*, 71.
[45] J. P. Abrahams, A. G. W. Leslie, R. Lutter, J. E. Walker, *Nature* **1994**, *370*, 621.
[46] M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. D. Cate, H. F. Noller, *Science* **2001**, *292*, 883.

[47] Y. Shamoo, T. A. Steitz, *Cell* **1999**, *99*, 155.
[48] S. Cusack, *Current Opinion in Structural Biology* **1999**, *9*, 66.
[49] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Research* **2000**, *28*, 235.
[50] C. V. Robinson, A. Sali, W. Baumeister, *Nature* **2007**, *450*, 973.
[51] A. Gingras, R. Aebersold, B. Raught, *Journal of Physiology* **2005**, *563*, 11.
[52] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, D. C. Phillips, *Nature* **1958**, *181*, 662.
[53] P. Plevka, R. Perera, J. Cardosa, R. J. Kuhn, M. G. Rossmann, *Science* **2012**, *336*, 1274.
[54] A. W. Hull, *Journal of the American Chemical Society* **1919**, *41*, 1168.
[55] M. S. Smyth, J. H. J. Martin, *Molecular Pathology* **2000**, *53*, 8.
[56] Z. Du, J. K. Lee, S. Fenn, R. Tjhen, R. M. Stroud, T. L. James, *RNA* **2007**, *13*, 1043.
[57] T. Ritschel, C. Atmanene, K. Reuter, A. Van Dorsselaer, S. Sanglier-Cianferani, G. Klebe, *Journal of Molecular Biology* **2009**, *393*, 833.
[58] B. Kobe, G. Guncar, R. Buchholz, T. Huber, B. Maco, N. Cowieson, J. L. Martin, M. Marfori, J. K. Forwood, *Biochemical Society Transactions* **2008**, *36*, 1438.
[59] J. Janin, in *Protein-Protein Complexes* (Ed.: M. Zacharias), Imperial College Press, London, **2010**, pp. 1.
[60] E. Zuiderweg, R.P., *Biochemistry* **2002**, *41*, 1.
[61] C. Dominguez, M. Schubert, O. Duss, S. Ravindranathan, F. H. T. Allain, *Progress in Nuclear Magnetic Resonance Spectroscopy* **2011**, *58*, 1.
[62] T. Madl, F. Gabel, M. Sattler, *Journal of Structural Biology* **2011**, *173*, 472.
[63] R. W. Montalvao, A. Cavalli, X. Salvatella, T. L. Blundell, M. Vendruscolo, *Journal of the American Chemical Society* **2008**, *130*, 15990.
[64] A. Kumar, G. Wagner, R. R. Ernst, K. Wuethrich, *Journal of the American Chemical Society* **1981**, *103*, 3654.
[65] M. Bieri, A. H. Kwan, M. Mobli, G. F. King, J. P. Mackay, P. R. Gooley, *FEBS Journal* **2011**, *278*, 704.
[66] H. J. Dyson, P. E. Wright, *Chemical Reviews* **2004**, *104*, 3607.
[67] P. J. Folkers, G. M. Clore, P. C. Driscoll, J. Dodt, S. Kohler, A. M. Gronenborn, *Biochemistry* **1989**, *28*, 2601.
[68] P. C. Driscoll, A. M. Gronenborn, L. Beress, G. M. Clore, *Biochemistry* **1989**, *28*, 2188.
[69] A. Yee, X. Chang, A. Pineda-Lucena, B. Wu, A. Semesi, B. Le, T. Ramelot, G. M. Lee, S. Bhattacharyya, P. Gutierrez, A. Denisov, C.-H. Lee, J. R. Cort, G. Kozlov, J. Liao, G. Finak, L. Chen, D. Wishart, W. Lee, L. P. McIntosh, K. Gehring, M. A. Kennedy, A. M. Edwards, C. H. Arrowsmith, *Proceedings of the National Academy of Sciences* **2002**, *99*, 1825.
[70] G. M. Clore, A. M. Gronenborn, *Current Opinion in Chemical Biology* **1998**, *2*, 564.
[71] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proceedings of the National Academy of Sciences* **1997**, *94*, 12366.
[72] V. Tugarinov, W. Y. Choy, V. Y. Orekhov, L. E. Kay, *Proceedings of the National Academy of Sciences* **2005**, *102*, 622.
[73] M. Pellecchia, P. Sebbel, U. Hermanns, K. Wuthrich, R. Glockshuber, *Nature Structural & Molecular Biology* **1999**, *6*, 336.
[74] J. Flaux, E. B. Bertelsen, A. L. Horwich, K. Wuthrich, *Nature* **2002**, *418*, 207.
[75] A. Marintchev, D. Frueh, G. Wagner, in *Methods in Enzymology, Vol. Volume 430* (Ed.: L. Jon), Academic Press, **2007**, pp. 283.
[76] A. H. Kwan, M. Mobli, P. R. Gooley, G. F. King, J. P. Mackay, *FEBS Journal* **2011**, *278*, 687.
[77] Y. Ito, P. Selenko, *Current Opinion in Structural Biology* **2010**, *20*, 640.

[78] Z. Serber, V. Dotsch, *Biochemistry* **2001**, *40*, 14317.
[79] M. M. Dedmon, C. N. Patel, G. B. Young, G. J. Pielak, *Proceedings of the National Academy of Sciences* **2002**, *99*, 12681.
[80] D. S. Burz, K. Dutta, D. Cowburn, A. Shekhtman, *Nature Methods* **2006**, *3*, 91.
[81] T. Sakai, H. Tochio, T. Tenno, Y. Ito, T. Kokubo, H. Hiroaki, S. M., *Journal of Biomolecular NMR* **2006**, *36*.
[82] Z. H. Zhou, in *Advances in Protein Chemistry and Structural Biology, Vol. Volume 82* (Eds.: J. L. Steven, B. V. V. Prasad), Academic Press, **2011**, pp. 1.
[83] K. A. Taylor, R. M. Glaeser, *Science* **1974**, *186*, 1036.
[84] J. D. Yoder, J. O. Cifuente, J. Pan, J. M. Bergelson, S. Hafenstein, *J Virol* **2012**.
[85] T. F. Lerch, J. K. O'Donnell, N. L. Meyer, Q. Xie, K. A. Taylor, S. M. Stagg, M. S. Chapman, *Structure* **2012**, *20*, 1310.
[86] T. A. Bharat, N. E. Davey, P. Ulbrich, J. D. Riches, A. de Marco, M. Rumlova, C. Sachse, T. Ruml, J. A. Briggs, *Nature* **2012**, *487*, 385.
[87] J. Liu, M. J. Bartesaghi, G. Borgnia, S. Subramaniam, *Nature* **2008**, *455*, 109.
[88] M. Wolf, R. L. Garcea, N. Grigorieff, S. C. Harrison, *Proceedings of the National Academy of Sciences* **2010**, *107*, 6298.
[89] W. Zhang, M. Kimmel, C. M. Spahn, P. A. Penczek, *Structure* **2008**, *16*, 1770.
[90] H. Liu, L. Jin, S. B. Koh, I. Atanasov, S. Schein, L. Wu, Z. H. Zhou, *Science* **2010**, *329*, 1038.
[91] V. S. Reddy, S. K. Natchiar, P. L. Stewart, G. R. Nemerow, *Science* **2010**, *329*, 1071.
[92] M. Topf, A. Sali, *Current Opinion in Structural Biology* **2005**, *15*, 578.
[93] G. Pintilie, W. Chiu, *Biopolymers* **2012**, *97*, 742.
[94] J. Esquivel-Rodriguez, D. Kihara, *Journal of Physical Chemistry B* **2012**, *116*, 6854.
[95] S. Mulepati, A. Orr, S. Bailey, *Journal of Biological Chemistry* **2012**, *287*, 22445.
[96] M. E. W. Janssen, E. Kim, H. Liu, L. M. Fujimoto, A. Bobkov, N. Volkmann, D. Hanein, *Molecular Cell* **2006**, *21*, 271.
[97] C. E. Hsieh, M. Marko, J. Frank, C. A. Mannella, *J Struct Biol* **2002**, *138*, 63.
[98] J. Cope, J. Heumann, A. Hoenger, in *Current Protocols in Protein Science*, John Wiley & Sons, Inc., **2001**.
[99] M. Cyrklaff, A. Linaroudis, M. Boicu, P. Chlanda, W. Baumeister, G. Griffiths, J. Krijnse-Locker, *PLoS ONE* **2007**, *2*, e420.
[100] M. Stolken, F. Beck, T. Haller, R. Hegerl, I. Gutsche, J. M. Carazo, W. Baumeister, S. H. W. Scheres, S. Nickell, *Journal of Structural Biology* **2011**, *173*, 77.
[101] M. F. Schmid, C. R. Booth, *Journal of Structural Biology* **2008**, *161*, 243.
[102] J. O. Ortiz, F. Brandt, V. R. F. Matias, L. Sennels, J. Rappsilber, S. H. W. Scheres, M. Eibauer, F. U. Hartl, W. Baumeister, *The Journal of Cell Biology* **2010**, *190*, 613.
[103] R. M. Glaeser, *Physics Today* **2008**, *61*, 48.
[104] A. Guinier, *Comptes rendus hebdomadaires des séances de l'Académie des sciences* **1938**, *206*, 1374.
[105] D. A. Jacques, J. Trewhella, *Protein Science* **2006**, *19*, 642.
[106] G. L. Hura, A. L. Menon, M. Hammel, R. P. Rambo, F. L. Poole II, S. E. Tsutakawa, F. E. Jenney Jr, S. Classen, K. A. Frankel, R. C. Hopkins, S.-j. Yang, J. W. Scott, B. D. Dillard, M. W. W. Adams, J. A. Tainer, *Nature Methods* **2009**, *6*, 606.
[107] M. Bada, D. Walther, B. Arcangioli, S. Doniach, M. Delarue, *Journal of Molecular Biology* **2000**, *300*, 563.
[108] L. Costenaro, J. G. Grossmann, C. Ebel, A. Maxwell, *Structure* **2005**, *13*, 287.

[109] E. Gherardi, S. Sandin, M. V. Petoukhov, J. Finch, M. E. Youles, L.-G. Ofverstedt, R. N. Miguel, T. L. Blundell, G. F. Vande Woude, U. Skoglund, D. I. Svergun, *Proceedings of the National Academy of Sciences* **2006**, *103*, 4046.
[110] O. S. Rosenberg, S. Deindl, R. J. Sung, A. C. Nairn, J. Kuriyan, *Cell* **2005**, *123*, 849.
[111] S. J. Kim, C. Dumont, M. Gruebele, *Biophys J* **2008**, *94*, 4924.
[112] R. Zhang, P. A. Marone, P. Thiyagarajan, D. M. Tiede, *Langmuir* **1999**, *15*, 7510.
[113] Z. He, V. M. Garamus, S. S. Funari, M. Malfois, R. Willumeit, B. Niemeyer, *Journal of Physical Chemistry B* **2002**, *106*, 7596.
[114] D. I. Svergun, M. V. Petoukhov, M. H. Koch, *Biophysical Journal* **2001**, *80*, 2946.
[115] M. V. Petoukhov, P. V. Konarev, A. G. Kikhney, D. I. Svergun, *Journal of Applied Crystallography* **2007**, *40*, S223.
[116] D. I. Svergun, *Biophysical Journal* **1999**, *76*, 2879.
[117] J. Lipfert, S. Doniach, in *Annual Review of Biophysics and Biomolecular Structure, Vol. 36*, Annual Reviews, Palo Alto, **2007**, pp. 307.
[118] S. Fields, O. K. Song, *Nature* **1989**, *340*, 245.
[119] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, J. M. Rothberg, *Nature* **2000**, *403*, 623.
[120] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, Y. Sakaki, *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 4569.
[121] I. G. Serebriiskii, V. Khazak, E. a. Golemis, *Journal of Biological Chemistry* **1999**, *274*, 17080.
[122] J. Huang, S. L. Schreiber, *Proceedings of the National Academy of Sciences* **1997**, *94*, 13396.
[123] I. Stagljar, C. Korostensky, N. Johnsson, S. te Heesen, *Proceedings of the National Academy of Sciences* **1998**, *95*, 5187.
[124] I. G. Serebriiskii, O. Mitina, E. N. Pugacheva, E. Benevolenskaya, E. Kotova, G. G. Toby, V. Khazak, W. G. Kaelin, J. Chernoff, E. A. Golemis, *Genome Research* **2002**, *12*, 1785.
[125] M. Zhou, T. Veenstra, *Proteomics* **2007**, *7*, 2688.
[126] X. Yu, X. X. Liu, T. D. Liu, K. Hong, J. Lei, R. F. Yuan, J. H. Shao, *Digestive Diseases and Sciences* **2012**, *57*, 2347.
[127] J. E. Jong, K. W. Jeong, H. Shin, L. R. Hwang, D. Lee, T. Seo, *Cancer Letters* **2012**, *324*, 109.
[128] A. Bruckner, C. Polge, N. Lentze, D. Auerbach, U. Schlattner, *International Journal of Molecular Sciences* **2009**, *10*, 2763.
[129] H. R. Horvitz, *J Mol Biol* **1974**, *90*, 727.
[130] A. M. Verhagen, P. G. Ekert, M. Pakusch, J. Silke, L. M. Connolly, G. E. Reid, R. L. Moritz, R. J. Simpson, D. L. Vaux, *Cell* **2000**, *102*, 43.
[131] Y. Q. Li, J. Ren, W. H. Yu, Q. Li, H. Kuwahara, L. Yin, K. L. Carraway, D. Kufe, *Journal of Biological Chemistry* **2001**, *276*, 35239.
[132] E. M. Phizicky, S. Fields, *Microbiological Reviews* **1995**, *59*, 94.
[133] R. B. Birge, J. E. Farjardo, C. Reichman, S. E. Shoelson, Z. Songyang, L. C. Cantley, H. Hanafusa, *Molecular Cellular Biology* **1993**, *13*, 4648.
[134] P. H. Warne, P. R. Vicianna, J. Downward, *Nature* **1993**, *364*, 352.
[135] J. Freed, J. Smith, P. Li, A. Greene, *Proteomics* **2007**, *7*, 2371.
[136] T. Glatter, A. Wepf, R. Aebersold, M. Gstaiger, *Molecular Systems Biology* **2009**, *5*, 1.
[137] A. Gingras, M. Gstaiger, B. Raught, R. Aebersold, *Nature Reviews Molecular Cellular Biology* **2007**, *8*, 645.
[138] J. Rappsilber, S. Siniossoglou, E. Hurt, M. Mann, *Analytical Chemistry* **2000**, *72*, 267.
[139] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Seraphin, *Nature Biotechnology* **1999**, *17*, 1030.

[140] N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili, J. F. Greenblatt, *Nature* **2006**, *440*, 637.
[141] J. Miernyk, J. Thelen, *Plant Journal* **2008**, *53*, 597.
[142] D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, M. J. Cormier, *Gene* **1992**, *111*, 229.
[143] M. V. Matz, K. A. Lukyanov, S. A. Lukyanov, *Bioessays* **2002**, *24*, 953.
[144] T. Nagai, K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba, A. Miyawaki, *Nat Biotechnol* **2002**, *20*, 87.
[145] M. A. Rizzo, G. H. Springer, B. Granada, D. W. Piston, *Nature Biotechnology* **2004**, *22*, 445.
[146] H. W. Ai, N. C. Shaner, Z. Cheng, R. Y. Tsien, R. E. Campbell, *Biochemistry* **2007**, *46*, 5904.
[147] T. Forster, *Naturwissenschaften* **1946**, *6*, 166.
[148] T. Forster, in *Modern quantum chemistry, Vol. 3* (Ed.: O. Sinanoglu), Academic Press, Nwe York, **1965**, pp. 93.
[149] L. Stryer, *Annual Reviews in Biochemistry* **1978**, *47*, 819.
[150] J. R. Lakowicz, *Principles of fluorescence spectroscopy, Vol. xxvi*, Springer, New York, **2006**.
[151] R. N. Day, M. W. Davidson, *Bioessays* **2012**, *34*, 341.
[152] W. M. Shih, Z. Gryczynski, J. R. Lakowicz, J. A. Spudich, *Cell* **2000**, *102*, 683.
[153] L. S. He, D. P. Olson, X. L. Wu, T. S. Karpova, J. G. McNally, P. E. Lipsky, *Cytometry Part A* **2003**, *55A*, 71.
[154] M. C. Overton, K. J. Blumer, *Current Biology* **2000**, *10*, 341.
[155] I. A. Tonaco, J. W. Borst, S. C. de Vries, G. C. Angenent, R. G. Immink, *J Exp Bot* **2006**, *57*, 33.
[156] B. J. Bacskai, B. Hochner, M. Mahaut-Smith, S. R. Adams, B. K. Kaang, E. R. Kandel, R. Y. Tsien, *Science* **1993**, *260*, 222.
[157] H. Y. Li, E. K. O. Ng, S. M. Y. Lee, M. Kotaka, S. K. W. Tsui, C. Y. Lee, K. P. Fung, M. M. Y. Waye, *Journal of Cellular Biochemistry* **2001**, *80*, 293.
[158] Y. X. Wang, E. L. Botvinick, Y. H. Zhao, M. W. Berns, S. Usami, R. Y. Tsien, S. Chien, *Nature* **2005**, *434*, 1040.
[159] A. Hoppe, S. Seveau, J. Swanson, *Cellular Microbiology* **2009**, *11*, 540.
[160] C. Berney, G. Danuser, *Biophysical Journal* **2003**, *84*, 3992.
[161] A. Periasamy, H. Wallrabe, Y. Chen, M. Barroso, in *Biophysical Tools for Biologists, Vol 2: In Vivo Techniques, Vol. 89* (Eds.: J. J. Correia, H. W. Detrich), Elsevier Academic Press Inc, San Diego, **2008**, pp. 569.
[162] H. Raether, *Physics of Thin Films* **1977**, *9*, 145.
[163] B. D. Ratner, D. G. Castner, in *Surface analysis: The principal techniques* (Ed.: J. C. Vickerman), John Wiley & Sons, New York, **1997**.
[164] P. B. Garland, *Quarterly Reviews of Biophysics* **1996**, *29*, 91.
[165] F. Caruso, M. J. Jory, G. W. Bradberry, J. R. Sambles, D. N. Furlong, *Journal of Applied Physics* **1998**, *83*, 1023.
[166] J. Homola, S. S. Yee, G. Gauglitz, *Sensors & Actuators B: Chemical* **1999**, *54*, 3.
[167] L. Levesque, B. E. Paton, *Applied Optics* **1997**, *36*, 7199.
[168] E. Stenberg, B. Persson, H. Roos, C. Urbaniczky, *Journal of Colloid Interface Science* **1991**, *143*, 513.
[169] H. E. de Bruijn, R. P. H. Kooyman, J. Greve, *Applied Optics* **1990**, *29*, 1974.

[170] N. J. Geddes, A. S. Martin, F. Caruso, R. S. Urquhart, D. N. Furlong, J. R. Sambles, K. A. Than, J. A. Edgar, *Journal of Immunological Methods* **1994**, *175*, 149.
[171] A. Rogstam, S. Linse, A. Lindqvist, P. James, L. Wagner, T. Berggard, *The Biochemical Journal* **2007**, *401*, 353.
[172] P. S. Katsamba, S. Park, I. A. Laird-Offringa, *Methods* **2002**, *26*, 95.
[173] W. Huber, F. Mueller, *Curr Pharm Des* **2006**, *12*, 3999.
[174] I. Navratilova, M. Dioszegi, D. G. Myszka, *Anal Biochem* **2006**, *355*, 132.
[175] C. Boozer, G. Kim, S. Cong, H. Guan, T. Londergan, *Current Opinion in Biotechnology* **2006**, *17*, 400.
[176] J. S. Yuk, H.-S. Kim, J.-W. Jung, S.-H. Jung, S.-J. Lee, W. J. Kim, J.-A. Han, Y.-M. Kim, K.-S. Ha, *Biosensors and Bioelectronics* **2006**, *21*, 1521.
[177] R. L. Rich, D. G. Myszka, *Anal Biochem* **2007**, *361*, 1.
[178] S. P. Lal, R. I. Christopherson, C. G. dos Remedios, *Drug Discovery Today* **2002**, *7*, S143.
[179] M. F. Templin, D. Stoll, M. Schrenk, P. C. Traub, C. F. Vohringer, T. O. Joos, *Trends in Biotechnology* **2002**, *20*, 160.
[180] D. S. Wilson, S. Nock, *Angewandte Chemie International Edition in English* **2003**, *42*, 494.
[181] H. Petach, L. Gold, *Current Opinion in Biotechnology* **2002**, *13*, 309.
[182] M. Schaeferling, S. Schiller, H. Paul, M. Kruschina, P. Pavlickova, M. Meerkamp, C. Giammasi, D. Kambhampati, *Electrophoresis* **2002**, *23*.
[183] S. Weng, K. Gu, P. W. Hammond, P. Lohse, C. Rise, R. W. Wagner, M. C. Wright, R. G. Kuimelis, *Proteomics* **2002**, *2*, 48.
[184] C. P. Paweletz, L. Charboneau, V. E. Bichsel, N. L. Simone, T. Chen, J. W. Gillespie, M. R. Emmert-Buck, M. J. Roth, I. E. Petricoin, L. A. Liotta, *Oncogene* **2001**, *20*, 1981.
[185] G. MacBeath, S. L. Schreiber, *Science* **2000**, *289*, 1760.
[186] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, M. Snyder, *Science* **2001**, *293*, 2101.
[187] A. Kaushansky, J. E. Allen, A. Gordus, M. A. Stiffler, E. S. Karp, B. H. Chang, G. MacBeath, *Nature Protocols* **2010**, *5*, 773.
[188] T. Berggard, S. Linse, P. James, *Proteomics* **2007**, *7*, 2833.
[189] J. Benesch, B. Ruotolo, D. Simmons, C. V. Robinson, *Chemistry Reviews* **2007**, *107*, 3544.
[190] L. Konermann, X. Tong, Y. Pan, *Journal of Mass Spectrometry* **2008**, *43*, 1021.
[191] A. Heck, *Nature Methods* **2008**, 927.
[192] J. Loo, *Mass Spectrometry Reviews* **1997**, *16*, 1.
[193] W. Yang, H. Steen, M. Freeman, *Proteomics* **2008**, *8*, 832.
[194] J. J. Thomson, *Proceedings of the Royal Society* **1913**, *A89*, 1.
[195] J. J. Thomson, *Philosophical Magazine* **1897**, *44*, 293.
[196] J. J. Thomson, *Philosophical Magazine* **1912**, *24*, 209.
[197] A. O. Nier, *Review of Scientific Instruments* **1947**, *18*, 398.
[198] A. J. Dempster, *Physical Review* **1921**, *18*, 415.
[199] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, *Science* **1989**, *246*, 64.
[200] F. Hillenkamp, M. Karas, A. Ingeldoh, B. Stahl, in *Biological Mass Spectrometry* (Eds.: A. L. Burlingame, J. A. McCloskey), Elsevier, Amsterdam, **1990**, p. 49.
[201] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, T. Matsuo, *Rapid Communications in Mass Spectrometry* **1988**, *2*, 151.
[202] E. de Hoffmann, V. Stroobant, *Mass Spectrometry Principles and Applications*, Wiley, West Sussex, England, **2007**.
[203] K.-Y. Li, H. Tu, A. K. Ray, *Langmuir* **2005**, *21*, 3786.

[204] P. Kebarle, U. H. Verkerk, *Mass Spectrometry Reviews* **2009**, *28*, 898.
[205] B. Ganem, Y. T. Li, J. D. Henion, *Journal of the American Chemical Society* **1991**, *113*, 6294.
[206] B. Ganem, Y. T. Li, J. D. Henion, *Journal of the American Chemical Society* **1991**, *113*, 7818.
[207] T. R. Covey, R. F. Bonner, B. I. Shushan, J. Henion, *Rapid communications in mass spectrometry : RCM* **1988**, *2*, 249.
[208] M. Mann, C. K. Meng, J. B. Fenn, *Analytical Chemistry* **1989**, *61*, 1702.
[209] J. Sunner, G. Nicol, P. Kebarle, *Analytical Chemistry* **1988**, *60*, 1300.
[210] A. R. Dongre, J. L. Jones, A. Somogyi, V. H. Wysocki, *Journal of the American Chemical Society* **1996**, *118*, 8365.
[211] U. H. Verkerk, P. Kebarle, *J Am Soc Mass Spectrom* **2005**, *16*, 1325.
[212] J. R. Chapman, *Mass spectrometry of protein and peptides*, Humana Press, New Jersey, NJ, USA, **2000**.
[213] P. Kebarle, L. Tang, *Analytical Chemistry* **1993**, *65*, 972A.
[214] M. G. Ikonomou, A. T. Blades, P. Kebarle, *Analytical Chemistry* **1991**, *63*, 1989.
[215] S. Banerjee, S. Mazumdar, *International Journal of Analytical Chemistry* **2012**, *2012*.
[216] M. S. Wilm, M. Mann, *International Journal of Mass Spectrometry and Ion Processes* **1994**, *136*, 167.
[217] J. A. Loo, R. R. O. Loo, H. R. Udseth, C. G. Edmonds, R. D. Smith, *Rapid Communications in Mass Spectrometry* **1991**, *5*, 101.
[218] M. C. Fitzgerald, I. Chernushevich, K. G. Standing, C. P. Whitman, S. B. H. Kent, *Proceedings of the National Academy of Sciences* **1996**, *93*, 6851.
[219] F. Hillenkamp, M. Karas, R. C. Beavis, B. T. Chait, *Analytical Chemistry* **1991**, *63*, A1193.
[220] M. Karas, F. Hillenkamp, *Analytical Chemistry* **1988**, *60*, 2299.
[221] R. Zenobi, R. Knochenmuss, *Mass Spectrometry Reviews* **1998**, *17*, 337.
[222] R. Knochenmuss, R. Zenobi, *Chemistry Reviews* **2003**, *103*, 441.
[223] S. Trimpin, E. D. Inutan, T. N. Herath, C. N. McEwen, *Molecular & Cellular Proteomics* **2010**, *9*, 362.
[224] V. Frankevich, J. Zhang, M. Dashtiev, R. Zenobi, *Rapid Communications in Mass Spectrometry* **2003**, *17*, 2343.
[225] I. Chernushevich, A. N. Krutchinsky, W. Ens, K. G. Standing, in *44th ASMS Conference on Mass Spectrometry and Allied Topics*, Portland, OR, **1996**, p. 751.
[226] I. V. Chernushevich, W. Ens, K. G. Standing, P. C. Loewen, M. C. Fitzgerald, S. B. H. Kent, R. C. Werlen, M. Lankinen, X.-J. Tang, C. F. Brewer, S. Saha, in *43rd ASMS Conference on Mass Spectrometry and allied Topics*, Atlanta, GA, **1995**, p. 1327.
[227] X.-J. Tang, C. F. Brewer, S. Saha, I. Chernushevich, W. Ens, K. G. Standing, *Rapid Communications in Mass Spectrometry* **1994**, *8*, 750.
[228] W. Stephens, *The Bulletin of the American Physical Society* **1946**, *21*, 22.
[229] W. C. Wiley, I. H. McLaren, *Review of Scientific Instruments* **1955**, *26*, 1150.
[230] R. E. Kaiser, J. N. Louris, J. W. Amy, R. G. Cooks, D. F. Hunt, *Rapid Communications in Mass Spectrometry* **1989**, *3*, 225.
[231] Q. Z. Hu, R. J. Noll, H. Y. Li, A. Makarov, M. Hardman, R. G. Cooks, *Journal of Mass Spectrometry* **2005**, *40*, 430.
[232] M. B. Comisarow, A. G. Marshall, *Chemical Physics Letters* **1974**, *25*, 282.
[233] A. G. Marshall, C. L. Hendrickson, G. S. Jackson, *Mass Spectrom Rev* **1998**, *17*, 1.
[234] C. C. Zhang, A. K. Walker, R. Zand, M. A. Moscarello, J. M. Yan, P. C. Andrews, *Journal of Proteome Research* **2012**, *11*, 4791.
[235] R. A. Yost, C. G. Enke, *Journal of the American Chemical Society* **1978**, *100*, 2274.
[236] G. L. Glish, D. E. Goeringer, *Analytical Chemistry* **1984**, *56*, 2291.

[237] J. H. Futrell, C. D. Miller, *Review of Scientific Instruments* **1966**, *37*, 1521.
[238] K. R. Jennings, *International Journal of Mass Spectrometry and Ion Physics* **1968**, *1*, 227.
[239] J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, M. Mann, *Nature Methods* **2007**, *4*, 709.
[240] R. A. Zubarev, N. L. Kelleher, F. W. McLafferty, *Journal of the American Chemical Society* **1998**, *120*, 3265.
[241] D. P. Little, J. P. Speir, M. W. Senko, P. B. Oconnor, F. W. McLafferty, *Analytical Chemistry* **1994**, *66*, 2809.
[242] K. Biemann, *Biomedical and Environmental Mass Spectrometry* **1988**, *16*, 99.
[243] P. Roepstorff, J. Fohlman, *Biomedical Mass Spectrometry* **1984**, *11*, 601.
[244] H. Steen, M. Mann, *Nat Rev Mol Cell Biol* **2004**, *5*, 699.
[245] S. A. McLuckey, *Journal of the American Society for Mass Spectrometry* **1992**, *3*, 599.
[246] A. K. Shukla, J. H. Futrell, *J Mass Spectrom* **2000**, *35*, 1069.
[247] K. R. Jennings, *International Journal of Mass Spectrometry* **2000**, *200*, 479.
[248] K. Levsen, *Fundamental aspects of organic mass spectrometry*, Verlag Chemie, Weinheim; New York, **1978**.
[249] B. Spengler, *Journal of Mass Spectrometry* **1997**, *32*, 1019.
[250] L. Sleno, D. A. Volmer, *Journal of Mass Spectrometry* **2004**, *39*, 1091.
[251] A. L. McCormack, A. Somogyi, A. R. Dongre, V. H. Wysocki, *Analytical Chemistry* **1993**, *65*, 2859.
[252] G. Tsaprailis, H. Nair, Á. Somogyi, V. H. Wysocki, W. Zhong, J. H. Futrell, S. G. Summerfield, S. J. Gaskell, *Journal of the American Chemical Society* **1999**, *121*, 5142.
[253] B. Paizs, S. Suhai, *Mass Spectrometry Reviews* **2005**, *24*, 508.
[254] S. M. M. Sweet, C. M. Bailey, D. L. Cunningham, J. K. Heath, H. J. Cooper, *Molecular & Cellular Proteomics* **2009**, *8*, 904.
[255] B. Macek, M. Mann, J. V. Olsen, in *Annual Review of Pharmacology and Toxicology, Vol. 49*, Annual Reviews, Palo Alto, **2009**, pp. 199.
[256] D. L. Swaney, C. D. Wenger, J. A. Thomson, J. J. Coon, *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 995.
[257] J. A. Loo, H. R. Udseth, R. D. Smith, J. H. Futrell, *Rapid Communications in Mass Spectrometry* **1988**, *2*, 207.
[258] J. A. Loo, C. G. Edmonds, R. D. Smith, *Science* **1990**, *248*, 201.
[259] J. A. Loo, C. G. Edmonds, R. D. Smith, *Analytical Chemistry* **1991**, *63*, 2488.
[260] J. A. Loo, C. G. Edmonds, R. D. Smith, *Analytical Chemistry* **1993**, *65*, 425.
[261] E. Sachon, G. Clodic, T. Blasco, G. Bolbach, *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 1880.
[262] B. Spengler, D. Kirsch, R. Kaufmann, R. J. Cotter, *Rapid Communications in Mass Spectrometry* **1991**, *5*, 198.
[263] B. Spengler, D. Kirsch, R. Kaufmann, E. Jaeger, *Rapid Commun Mass Spectrom* **1992**, *6*, 105.
[264] R. S. Brown, J. J. Lennon, *Analytical Chemistry* **1995**, *67*, 3990.
[265] K. Demeure, L. Quinton, V. Gabelica, E. De Pauw, *Analytical Chemistry* **2007**, *79*, 8678.
[266] Z. Hall, C. Robinson, *Journal of the American Society for Mass Spectrometry* **2012**, *23*, 1161.
[267] S. K. Chowdhury, V. Katta, B. T. Chait, *Journal of the American Chemical Society* **1990**, *112*, 9012.
[268] A. A. Rostom, M. Sunde, S. J. Richardson, G. Schreiber, S. Jarvis, R. Bateman, C. M. Dobson, C. V. Robinson, *Proteins* **1998**, *Suppl 2*, 3.
[269] K. Pagel, S. J. Hyung, B. T. Ruotolo, C. V. Robinson, *Anal Chem* **2010**, *82*, 5363.
[270] J. L. Benesch, B. T. Ruotolo, F. Sobott, J. Wildgoose, A. Gilbert, R. Bateman, C. V. Robinson, *Anal Chem* **2009**, *81*, 1270.
[271] H. Zhang, W. Cui, J. Wen, R. E. Blankenship, M. L. Gross, *Anal Chem* **2011**, *83*, 5598.

[272] F. Zal, F. Chausson, E. Leize, A. Van Dorsselaer, F. H. Lallier, B. N. Green, *Biomacromolecules* **2002**, *3*, 229.
[273] H. Ramstrom, S. Sanglier, E. Leize-Wagner, C. Philippe, A. Van Dorsselaer, J. Haiech, *J Biol Chem* **2003**, *278*, 1174.
[274] M. W. Pinkse, C. S. Maier, J. I. Kim, B. H. Oh, A. J. Heck, *J Mass Spectrom* **2003**, *38*, 315.
[275] A. A. Rostom, P. Fucini, D. R. Benjamin, R. Juenemann, K. H. Nierhaus, F. U. Hartl, C. M. Dobson, C. V. Robinson, *Proc Natl Acad Sci U S A* **2000**, *97*, 5185.
[276] C. L. Hanson, P. Fucini, L. L. Ilag, K. H. Nierhaus, C. V. Robinson, *J Biol Chem* **2003**, *278*, 1259.
[277] J. Kool, N. Jonker, H. Irth, W. M. A. Niessen, *Analytical and Bioanalytical Chemistry* **2011**, *401*, 1109.
[278] N. Tahallah, R. H. Van Den Heuvel, W. A. Van Den Berg, C. S. Maier, W. J. Van Berkel, A. J. Heck, *J Biol Chem* **2002**, *277*, 36425.
[279] R. H. van den Heuvel, A. H. Westphal, A. J. Heck, M. A. Walsh, S. Rovida, W. J. van Berkel, A. Mattevi, *J Biol Chem* **2004**, *279*, 12860.
[280] J. L. P. Benesch, *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 341.
[281] E. van Duijn, D. A. Simmons, R. H. H. van den Heuvel, P. J. Bakkes, H. van Heerikhuizen, R. M. A. Heeren, C. V. Robinson, S. M. van der Vies, A. J. R. Heck, *Journal of the American Chemical Society* **2006**, *128*, 4694.
[282] E. van Duijn, *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 971.
[283] K. Linderstrøm-Lang, *Chem. Soc. Spec. Publ.* **1955**, *2*, 1.
[284] A. Berger, K. Linderstrøm-Lang, *Archives of Biochemistry and Biophysics* **1957**, *69*, 106.
[285] R. S. Molday, Englande.Sw, Kallenrg, *Biochemistry* **1972**, *11*, 150.
[286] H. M. Zhang, G. M. Bou-Assaf, M. R. Emmett, A. G. Marshall, *J Am Soc Mass Spectrom* **2009**, *20*, 520.
[287] M. F. Jeng, S. W. Englander, *Journal of Molecular Biology* **1991**, *221*, 1045.
[288] T. D. Wood, R. A. Chorush, F. M. Wampler, D. P. Little, P. B. Oconnor, F. W. McLafferty, *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 2451.
[289] D. L. Smith, Y. Z. Deng, Z. Q. Zhang, *Journal of Mass Spectrometry* **1997**, *32*, 135.
[290] Y. H. Deng, D. L. Smith, *Biochemistry* **1998**, *37*, 6256.
[291] S. J. Eyles, J. P. Speir, G. H. Kruppa, L. M. Gierasch, I. A. Kaltashov, *Journal of the American Chemical Society* **2000**, *122*, 495.
[292] H. Ehring, *Analytical Biochemistry* **1999**, *267*, 252.
[293] A. N. Hoofnagle, K. A. Resing, E. J. Goldsmith, N. G. Ahn, *Proceedings of the National Academy of Sciences* **2001**, *98*, 956.
[294] I. A. Kaltashov, C. E. Bobst, R. R. Abzalimov, *Analytical Chemistry* **2009**, *81*, 7892.
[295] Y. Bai, J. S. Milne, L. Mayne, S. W. Englander, *Proteins: Structure, Function, and Genetics* **1993**, *17*, 75.
[296] R. S. Johnson, D. Krylov, K. A. Walsh, *Journal of Mass Spectrometry* **1995**, *30*, 386.
[297] A. G. Harrison, T. Yalcin, *International Journal of Mass Spectrometry and Ion Processes* **1997**, *165–166*, 339.
[298] X. J. Tang, P. Thibault, R. K. Boyd, *Analytical Chemistry* **1993**, *65*, 2824.
[299] K. D. Rand, M. Zehl, O. N. Jensen, T. J. D. Jorgensen, *Analytical Chemistry* **2009**, *81*, 5577.
[300] R. R. Abzalimov, D. A. Kaplan, M. L. Easterling, I. A. Kaltashov, *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 1514.
[301] T. E. Angel, U. K. Aryal, S. M. Hengel, E. S. Baker, R. T. Kelly, E. W. Robinson, R. D. Smith, *Chemical Society Reviews* **2012**, *41*, 3912.
[302] D. E. Clemmer, R. R. Hudgins, M. F. Jarrold, *Journal of the American Chemical Society* **1995**, *117*, 10141.

[303] G. M. Shankar, S. Li, T. H. Mehta, A. Garcia-Munoz, N. E. Shepardson, I. Smith, F. M. Brett, M. A. Farrell, M. J. Rowan, C. A. Lemere, C. M. Regan, D. M. Walsh, B. L. Sabatini, D. J. Selkoe, *Nat Med* **2008**, *14*, 837.
[304] H. Amijee, C. Bate, A. Williams, J. Virdee, R. Jeggo, D. Spanswick, D. I. C. Scopes, J. M. Treherne, S. Mazzitelli, R. Chawner, C. E. Eyers, A. J. Doig, *Biochemistry* **2012**, *51*, 8338.
[305] S. L. Bernstein, T. Wyttenbach, A. Baumketner, J.-E. Shea, G. Bitan, D. B. Teplow, M. T. Bowers, *Journal of the American Chemical Society* **2005**, *127*, 2075.
[306] H. L. Cole, J. M. D. Kalapothakis, G. Bennett, P. E. Barran, C. E. MacPhee, *Angewandte Chemie-International Edition* **2010**, *49*, 9448.
[307] M. I. Iurascu, C. Cozma, N. Tomczyk, J. Rontree, M. Desor, M. Drescher, M. Przybylski, *Analytical and Bioanalytical Chemistry* **2009**, *395*, 2509.
[308] S. J. Valentine, J. G. Anderson, A. D. Ellington, D. E. Clemmer, *The Journal of Physical Chemistry B* **1997**, *101*, 3891.
[309] B. Ells, D. A. Barnett, K. Froese, R. W. Purves, S. Hrudey, R. Guevremont, *Analytical Chemistry* **1999**, *71*, 4747.
[310] L. M. Matz, H. H. Hill, *Analytical Chemistry* **2001**, *73*, 1664.
[311] D. P. Smith, K. Giles, R. H. Bateman, S. E. Radford, A. E. Ashcroft, *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 2180.
[312] C. S. Hoaglund, S. J. Valentine, D. E. Clemmer, *Analytical Chemistry* **1997**, *69*, 4156.
[313] X. Tang, J. E. Bruce, H. H. Hill, *Rapid Communications in Mass Spectrometry* **2007**, *21*, 1115.
[314] A. Adamov, J. Viidanoja, E. Karpanoja, H. Paakkanen, R. A. Ketola, R. Kostiainen, A. Sysoev, T. Kotiaho, *Review of Scientific Instruments* **2007**, *78*.
[315] M. J. Cohen, F. W. Karasek, *Journal of Chromatographic Science* **1970**, *8*, 330.
[316] K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose, R. H. Bateman, *Rapid Communications in Mass Spectrometry* **2004**, *18*, 2401.
[317] I. A. Buryakov, E. V. Krylov, E. G. Nazarov, U. K. Rasulev, *International Journal of Mass Spectrometry and Ion Processes* **1993**, *128*, 143.
[318] E. A. Mason, M. E.W., *Transport properties of ions in gases*, John Wiley & Sons, New York, USA, **1988**.
[319] D. Gerlich, in *Advances in Chemical Physics*, John Wiley & Sons, Inc., **2007**, pp. 1.
[320] S. D. Pringle, K. Giles, J. L. Wildgoose, J. P. Williams, S. E. Slade, K. Thalassinos, R. H. Bateman, M. T. Bowers, J. H. Scrivens, *International Journal of Mass Spectrometry* **2007**, *261*, 1.
[321] B. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S. J. Hyung, C. V. Robinson, *Nature Protocols* **2008**, *3*, 1139.
[322] S. C. Wang, A. Politis, B. N. Di, V. N. Bavro, S. J. Tucker, P. J. Booth, N. P. Barrera, C. V. Robinson, *Journal of the American Chemical Society* **2010**, *132*, 15468.
[323] A. Y. Park, C. V. Robinson, *Critical Reviews in Biochemistry and Molecular Biology* **2011**, *46*, 152.
[324] J. T. Hopper, N. J. Oldham, *J Am Soc Mass Spectrom* **2009**, *20*, 1851.
[325] J. Novotný, R. Bruccoleri, M. Karplus, *Journal of Molecular Biology* **1984**, *177*, 787.
[326] D. Eisenberg, A. D. McLachlan, *Nature* **1986**, *319*, 199.
[327] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577.
[328] K. Breuker, F. W. McLafferty, *Proceedings of the National Academy of Sciences* **2008**, *105*, 18145.
[329] E. R. Badman, S. Myung, D. E. Clemmer, *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 1493.
[330] D. Bagal, E. N. Kitova, L. Liu, A. El-Hawiet, P. D. Schnier, J. S. Klassen, *Analytical Chemistry* **2009**, *81*, 7801.
[331] L. Han, S.-J. Hyung, B. T. Ruotolo, *Faraday Discussions* **2013**.

[332] L. Han, S.-J. Hyung, J. J. S. Mayers, B. T. Ruotolo, *Journal of the American Chemical Society* **2011**, *133*, 11358.
[333] Y. Zhong, S.-J. Hyung, B. T. Ruotolo, *Expert Reviews in Proteomics* **2012**, *9*, 47.
[334] P. Alexander, M. Fox, K. A. Stacey, L. F. Smith, *Biochem J* **1952**, *52*, 177.
[335] F. Wold, in *Methods in Enzymology, Vol. Volume 11* (Ed.: C. H. W. Hirs), Academic Press, **1967**, pp. 617.
[336] F. C. Hartman, F. Wold, *Biochemistry* **1967**, *6*, 2439.
[337] W. G. Niehaus Jr, F. Wold, *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1970**, *196*, 170.
[338] J. Marmur, L. Grossman, *Proceedings of the National Academy of Sciences of the United States of America* **1961**, *47*, 778.
[339] K. C. Smith, *Biochemical and Biophysical Research Communications* **1962**, *8*, 157.
[340] P. D. Lawley, P. Brookes, *Journal of Molecular Biology* **1967**, *25*, 143.
[341] M. Y. Feldman, *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* **1967**, *149*, 20.
[342] V. D. Axelrod, M. Y. Feldman, Chuguev, II, A. A. Bayev, *Biochimica Et Biophysica Acta* **1969**, *186*, 33.
[343] G. E. Davies, G. R. Stark, *Proceedings of the National Academy of Sciences of the United States of America* **1970**, *66*, 651.
[344] J. Seebacher, P. Mallick, N. Zhang, J. S. Eddes, R. Aebersold, M. H. Gelb, *Journal of Proteome Research* **2006**, *9*, 2770.
[345] J. Y. Lee, L. Lachner, J. Nunnari, B. Phinney, *journal of Proteome Research* **2007**, *6*.
[346] Z. A. Chen, A. Jawhari, L. Fischer, C. Buchen, S. Tahir, T. Kamenski, M. Rasmussen, L. Lariviere, J.-C. Bukowski-Wills, M. Nilges, P. Cramer, J. Rappsilber, *The EMBO Journal* **2010**, *29*, 717.
[347] K.-J. Armache, S. Mitterweger, A. Meinhart, P. Cramer, *Journal of Biological Chemistry* **2005**, *280*, 7131.
[348] J. E. Bruce, *Proteomics* **2012**, *12*, 1565.
[349] L. Yang, C. Zheng, C. R. Weisbrod, X. Tang, G. R. Munske, M. R. Hoopmann, J. K. Eng, J. E. Bruce, *Journal of Proteome Research* **2012**, *11*, 1027.
[350] C. Zheng, L. Yang, M. R. Hoopmann, J. K. Eng, X. Tang, C. R. Weisbrod, J. E. Bruce, *Molecular & Cellular Proteomics* **2011**, *10*.
[351] H. Zhang, X. Tang, G. R. Munske, N. Zakharova, L. Yang, C. Zheng, M. A. Wolff, N. Tolic, G. A. Anderson, L. Shi, M. J. Marshall, J. K. Frederickson, J. E. Bruce, *Journal of Proteome Research* **2008**, *7*, 1712.
[352] H. Zhang, Z. Tang, G. R. Munske, N. tolic, G. A. Anderson, J. E. Bruce, *Molecular & Cellular Proteomics* **2009**, *8*, 409.
[353] G. H. Dihazi, A. Sinz, *Rapid Communications in Mass Spectrometry* **2003**, *17*, 2005.
[354] S. Jennebach, F. Herzog, R. Aebersold, P. Cramer, *Nucleic Acids Research* **2012**, *40*, 5591.
[355] J. Pettelkau, T. Schroder, C. H. Ihling, B. E. S. Olausson, K. Kolbel, C. Lange, A. Sinz, *Biochemistry* **2012**, *51*, 4932.
[356] F. Herzog, A. Kahraman, D. Boehringer, R. Mak, A. Bracher, T. Walzthoeni, A. Leitner, M. Beck, F.-U. Hartl, N. Ban, L. Malmstrom, R. Aebersold, *Science (New York, N.Y.)* **2012**, *337*, 1348.
[357] R. Fritzsche, C. H. Ihling, M. Götze, A. Sinz, *Rapid Communications in Mass Spectrometry* **2012**, *26*, 653.
[358] L. F. A. Santos, A. H. Iglesias, F. C. Gozzo, *Journal of Mass Spectrometry* **2011**, *46*, 742.
[359] S. Madler, C. Bich, D. Touboul, R. Zenobi, *Journal of Mass Spectrometry* **2009**, *44*, 694.
[360] A. H. Iglesias, L. F. A. Santos, F. C. Gozzo, *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 557.

[361] S. P. Gaucher, M. Z. Hadi, M. M. Young, *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 395.
[362] A. Panchaud, P. Singh, S. A. Shaffer, D. R. Goodlett, *Journal of Proteome Research* **2010**, *9*, 2508.
[363] T. Walzthoeni, M. Claassen, A. Leitner, F. Herzog, S. Bohn, F. Forster, M. Beck, R. Aebersold, *Nature Methods* **2012**, *9*, 901.
[364] W. Li, H. A. O'Neill, V. H. Wysocki, *Bioinformatics (Oxford, England)* **2012**, *28*, 2548.
[365] A. Maiolica, D. Cittaro, D. Borsotti, L. Sennels, C. Ciferri, C. Tarricone, A. Musacchio, J. Rappsilber, *Journal of Molecular Cellular Proteomics* **2007**, *6*, 2200.
[366] C. J. Collins, B. Schilling, M. Young, G. Dollinger, R. K. Guy, *Bioorganic & Medicinal Chemistry Letters* **2003**, *13*, 4023.
[367] D. R. Muller, P. Schindler, H. Towbin, U. Wirth, H. H. Voshol, S. Hoving, M. O. Steinmetz, *Analytical Chemistry* **2001**, *73*, 1927.
[368] E. V. Petrotchenko, J. J. Serpa, C. H. Borchers, *Analytical Chemistry* **2010**, *82*, 817.
[369] C. H. Ihling, A. Schmidt, S. Kalkhof, D. Schultz, C. Stingl, K. Mechtler, M. Haack, A. Beck-Sickinger, D. Cooper, A. Sinz, *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 1100.
[370] V. Redeker, J. Bonnefoy, J.-P. L. Caer, S. Pemberton, O. Laprevote, R. Melki, *The FEBS Journal* **2010**, *277*, 5112.
[371] Y. J. Lee, *Molecular Biosystems* **2008**, *4*, 816.
[372] O. Rinner, J. Seebacher, T. Walzthoeni, L. N. Mueller, M. Beck, A. Schmidt, M. Mueller, R. Aebersold, *Nature Methods* **2008**, *5*, 315.
[373] E. V. Petrotchenko, C. H. Borchers, *Bmc Bioinformatics* **2010**, *11*.
[374] C. A. G. Soderberg, W. Lambert, S. Kjellstrom, A. Wiegandt, R. P. Wulff, C. Mansson, G. Rutsdottir, C. Emanuelsson, *PLoS ONE* **2012**, *7*.
[375] A. Sinz, *Mass Spectrometry Reviews* **2006**, *25*, 663.
[376] J. W. Back, A. F. Hartog, H. L. Dekker, A. O. Muijsers, L. J. d. Koning, L. d. Jong, *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 222.
[377] C. L. Swaim, J. B. Smith, D. L. SMith, *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 736.
[378] E. Soderblom, B. Bobav, J. Cavannagh, M. Goshe, *Rapid Communications in Mass Spectrometry* **2007**, *21*, 3395.
[379] S. Kalkhof, A. Sinz, *Analytical Biochemistry* **2008**, *392*, 305.
[380] P. Singh, S. Shaffer, A. Scherl, C. Holman, R. Pfuetzner, T. Larson Freeman, S. Miller, P. Hernandex, R. Appel, D. Goodlett, *Analytical Chemistry* **2008**, *80*, 8799.
[381] Q. Zhang, E. Crosland, D. Fabris, *Analytica Chimica Acta* **2008**, *627*, 117.
[382] P. Novak, G. H. Kruppa, *European Journal of Mass Spectrometry* **2008**, *14*, 355.
[383] F. Krauth, C. H. Ihling, H. H. Ruttinger, A. Sinz, *Rapid Communications in Mass Spectrometry* **2009**, *23*, 2811.
[384] Z. Zhang, C. B. Post, D. L. Smith, *Biochemistry* **1996**, *35*, 779.
[385] S. Kang, L. Mou, J. Lanman, S. Velu, W. J. Brouillette, P. E. Prevelige Jr, *Rapid Communications in Mass Spectrometry* **2009**, *23*, 1719.
[386] N. Fujii, R. B. Jacobsen, N. L. Wood, J. S. Schoeniger, R. K. Guy, *Bioorganic & Medicinal Chemistry Letters* **2004**, *14*, 427.
[387] A. Sinz, S. Kalkhof, C. H. Ihling, *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 1921.
[388] M. Trester-Zedlitz, K. Kamada, S. K. Burley, D. Fenvo, B. T. Chait, T. W. Muir, *Journal of the American Chemical Society* **2003**, *125*, 2416.
[389] G. H. Hur, J. L. Meier, J. Baskin, J. A. Codelli, R. c. Bertozzi, M. A. Marshiel, M. D. Burkart, *Chemical Biology* **2009**, *16*, 372.

[390] M. A. Nessen, g. Kramer, J. Back, J. M. Baskin, L. E. J. Smeenk, L. J. de Koning, J. H. van Maarseveev, L. de Jong, C. R. Bertozzi, H. Himstra, C. G. de Koster, *Journal of Proteome Research* **2009**, *8*, 3702.
[391] S. M. Chowdhury, X. Du, N. Tolic, S. Wu, R. J. Moore, M. U. Mayer, R. D. Smith, J. N. Adkins, *Analytical Chemistry* **2009**, *81*, 5524.
[392] S. Yoshitake, M. Imagawa, E. Ishikawa, Y. Niitsu, I. Urushizaki, M. Nishiura, R. Kanazawa, H. Kurosaki, S. Tachibana, N. Nakazawa, H. Ogawa, *Journal of Biochemistry* **1982**, *92*, 1413.
[393] I. Forne, J. Ludwigsen, A. Imhof, P. B. Becker, F. Mueller-Planitz, *Molecular & Cellular Proteomics* **2012**, *11*.
[394] G. Dorman, G. D. Prestwich, *Biochemistry* **1994**, *33*, 5661.
[395] Y. Sadakane, Y. Hatanaka, *Analytical Sciences* **2006**, *22*, 209.
[396] S. M. Sharkady, J. M. Nolan, *Nucleic Acids Research* **2001**, *29*, 3848.
[397] E. V. Petrotchenko, J. J. Serpa, C. H. Borchers, *Molecular & Cellular Proteomics* **2011**, *10*.
[398] J. Luo, J. Fishburn, S. Hahn, J. Ranish, *Molecular & Cellular Proteomics* **2012**, *11*.
[399] C. H. Sohn, H. D. Agnew, J. E. Lee, M. J. Sweredoski, R. L. J. Graham, G. T. Smith, S. Hess, G. Czerwieniec, J. A. Loo, J. R. Heath, R. J. Deshaies, J. L. Beauchamp, *Analytical Chemistry* **2012**, *84*, 2662.
[400] D. Vellucci, A. Kao, R. M. Kaake, S. D. Rychnovskv, L. Huang, *Journal of the American Society for Mass Spectrometry* **2010**, *21*, 1432.
[401] G. J. King, A. Jones, B. Kobe, T. Huber, D. Mouradov, D. A. Hume, I. L. Ross, *Analytical Chemistry* **2008**, *80*, 5036.
[402] E. V. Petrotchenko, V. K. Olkhovik, C. H. Borchers, *Molecular & Cellular Proteomics* **2005**, *4*, 1167.
[403] M. Q. Muller, F. Dreiocker, C. H. Ihling, M. Schafer, A. Sinz, *Analytical Chemistry* **2010**, *82*, 6958.
[404] E. V. Petrotchenko, K. Xiao, J. Cable, Y. Chen, N. V. Dokholyan, C. H. Borchers, *Molecular & Cellular Proteomics* **2009**, *8*, 273.
[405] M. W. Gardner, L. A. Vasicek, S. Shabbir, E. V. Anslyn, J. S. Brodbelt, *Analytical Chemistry* **2008**, *80*, 4807.
[406] X. Tang, G. R. Munske, W. Siems, J. E. Bruce, *Analytical Chemistry* **2005**, *77*, 311.
[407] Y. Lu, M. Tanasova, B. Borhan, G. E. Reid, *Analytical Chemistry* **2008**, *80*, 9279.
[408] T. T. Kasper, J. W. Back, M. Vitale, A. F. Hartog, W. Roseboom, L. J. De Koning, J. H. van Maarseveev, A. O. Muijsers, C. G. De Koster, L. De Jong, *ChemBioChem* **2007**, *8*, 1281.
[409] L. Yang, X. Tang, C. R. Weisbrod, G. R. Munske, J. K. Eng, P. D. von Haller, N. K. Kaiser, J. E. Bruce, *Analytical Chemistry* **2010**, *82*, 3556.
[410] M. Q. Muller, F. Dreiocker, C. H. Ihling, M. Schafer, A. Sinz, *Analytical Chemistry* **2010**, *45*, 880.
[411] F. Dreiocker, M. Q. Muller, A. Sinz, M. Schafer, *Journal of Mass Spectrometry* **2009**, *45*, 178.
[412] M. J. Trnka, A. L. Burlingame, *Molecular & Cellular Proteomics* **2010**, *9*, 2306.
[413] F. Chu, S. Mahrus, C. S. Craik, A. L. Burlingame, *Journal of the American Chemical Society* **2006**, *128*, 10362.
[414] X. Tang, W. Yi, G. R. Munske, D. P. Adhikari, N. L. Zakharova, J. E. Bruce, *Journal of Proteome Research* **2006**, *6*, 724.
[415] H. Zhang, X. Tang, G. R. Munske, N. Tolic, G. A. Anderson, J. E. Bruce, *Molecular & Cellular Proteomics* **2009**, *8*, 409.
[416] L. Yang, C. Zheng, C. R. Weisbrod, X. Tang, G. R. Munske, M. R. Hoopmann, J. K. Eng, J. E. Bruce, *Journal of Proteome Research* **2011**, *11*, 1027.
[417] C. Zheng, L. Yang, M. R. Hoopmann, J. K. Eng, X. Tang, C. R. Weisbrod, J. E. Bruce, *Molecular & cellular proteomics : MCP* **2011**, *10*, M110.006841.

[418] M. Q. Muller, J. J. Zeiser, F. Dreiocker, A. Pich, M. Schafer, A. Sinz, *Rapid Communications in Mass Spectrometry* **2011**, *25*, 155.

Chapter 2

PC1: Synthesis and Evaluation of a Piperazine Containing Crosslinker

The first major goal of this thesis involved the design and synthesis of new crosslinking reagents to overcome several of the intrinsic difficulties found with commercially available crosslinkers discussed in Chapter 1. The goal was to design a crosslinking reagent that would react specifically and completely with known functional groups, cleave upon CID fragmentation and produce marker ions, be water soluble, and contain an affinity tag for purification of crosslinked peptides. N-hydroxysuccinimide (NHS) esters are frequently used as a reactive group in crosslinking reagents because they are susceptible to nucleophilic attack, particularly by primary amines such as those found in lysyl residues and the N-termini of proteins [1]. NHS esters were chosen as the reactive group for these crosslinkers for these reasons. In addition their chemistry is well known and their relatively rapid reaction with water makes them self-quenching reagents.

2.1 Design of a Novel Crosslinker

Crosslinked peptides may either fragment poorly or produce complex MS/MS spectra because there are two overlapping series of b- and y- ions[1, 2]. This makes data analysis and identification of crosslinked peptides difficult but CID cleavable crosslinkers could improve the fragmentation efficiency and simply the spectra. Therefore, one of the most important characteristics to include in the design of the crosslinker was that crosslinked peptides would

fragment well by CID. Tertiary amines and cyclic amines are known to undergo facile fragmentation in CID, have interesting rearrangement potential, and good solubilities in water. The energies necessary to induce fragmentation of these structures are comparable to the energy required for peptide backbone fragmentation, so both fragmentation events can occur during the same MS/MS experiment. Crosslinking reagents [3], as well as commercial isobaric tags such as iTRAQ[4], have been designed to include these structural features as shown in Figure 2.1. Previous work (unpublished) in the Andrews lab demonstrated that the iTRAQ reagent is very antigenic and the lab is in possession of antibodies raised against iTRAQ labeled proteins. Because of these reasons, it was hypothesized that a crosslinker with a structure incorporating tertiary amines would be CID cleavable, would produce marker ions, and might be antigenic. Thus the synthesis of structure of PC1 was proposed. (Figure 2.2) The structure of this crosslinker was not deemed to be water soluble, but if PC1 behaved as expected, the NHS-esters could be changed to sulfo-NHS esters to improve solubility. The Vahlteich Medicinal Chemistry Core was commissioned to synthesize PC1 and then I evaluated its utility via several assays.

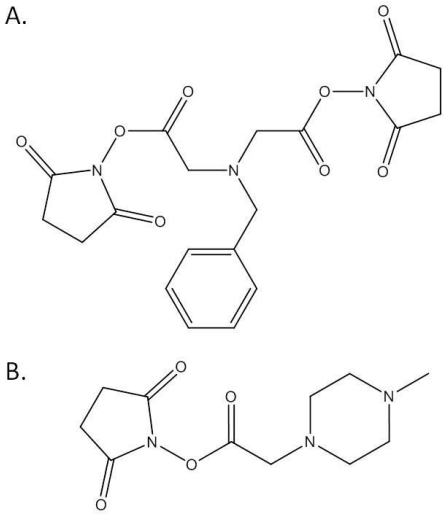


Figure 2.1. Tertiary amine containing reagents for protein chemistry. Both the crosslinker in (A)[3] and the tagging reagent (iTRAQ) in (B)[4] are cleavable by low energy CID.

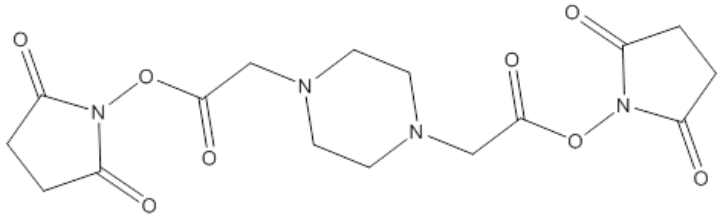


Figure 2.2. Structure of PC1. PC1, the first crosslinker synthesized and evaluated in this work.

2.2 SDS-PAGE and Western blot Analysis:

The first goal was to verify that PC1 was a viable crosslinker for multimeric proteins. Fructose 1,6-bisphosphate aldolase (aldolase) was chosen for testing purposes because it is a structurally well-characterized enzyme composed of four subunits of approximately 39kDa each[5]. PC1 was applied to aldolase at a range of ratios of crosslinker to protein and then the extent of crosslinking was analyzed via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 2.3). The objective was to add increasing amounts of crosslinker to maximize the amount of the biologically relevant tetramer while minimizing the amount of higher order structures because those structures represented nonspecific interactions. As the ratio of crosslinker to protein increased, bands began to appear at higher molecular weights, indicating the monomers were covalently attached. At a ratio of 500:1, the most prevalent band resulted from the tetramer, but higher molecular weight bands (nonbiologically relevant structures) were observed. The lanes crosslinked with lower ratios of PC1 had significant amounts of the monomer, trimer, and dimer, however the 50:1 ratio appeared to have the highest amount of tetramer with the lowest amount of higher order structures. The concentration of aldolase in the cell is approximately 200 ng/mL[6] and the concentration of

protein used prior to crosslinking was 5 µg/µl. At such high concentrations of protein, it was possible that the aldolase in solution was actually in those in higher order structures and not just as a tetramer. Those structures were still not biologically relevant but simultaneously lowering the concentration of aldolase and maintaining the same ratio of protein to crosslinker should reduce the amount of higher order crosslinking. The directions for commercially available crosslinkers suggested the use of a crosslinker to protein ratio of 10-50:1 for best results (www.piercenet.com). Those suggested ratios were similar to the ratios observed here that produce the least nonbiologically relevant structures. From these data, it was apparent that PC1is able to effectively crosslink aldolase as efficiently as other known crosslinkers.

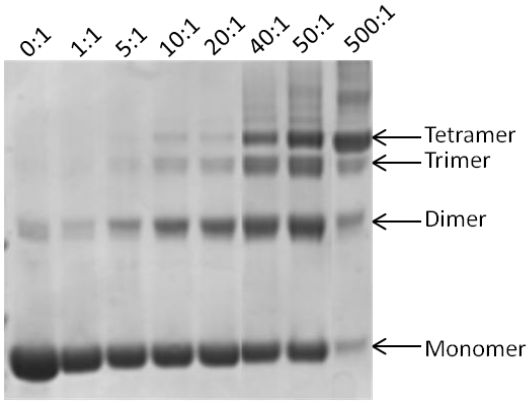


Figure 2.3. SDS-PAGE of PC1 crosslinked aldolase. Aldolase was crosslinked with increasing ratios of PC1 and analyzed via SDS-PAGE. The ratios of PC1 to aldolase monomer are shown at the top of the gel.

The crosslinker PC1 was designed to be antigenic, therefore the next goal was to evaluate its antigenic properties. The Andrews lab had previously raised antibodies against iTRAQ labeled protein and because the structure of PC1 was very similar to the tagging reagent iTRAQ, PC1 crosslinked proteins could cross react with an iTRAQ antibody. Therefore the ability of those antibodies to cross react with PC1 crosslinked proteins (aldolase) was tested via western

blot (Figure 2.4). There was a strong signal from the iTRAQ labeled aldolase in lane 3 (Figure 2.4), however PC1 crosslinked proteins in lane 2 (Figure 2.4) did not produce a signal which suggested that PC1 crosslinked proteins were not cross reactive with the iTRAQ antibody. Previous work in the Andrews lab with the iTRAQ antibody suggested that the iTRAQ antibody was very sensitive, therefore a only very small amount of labeled protein was needed to see signal. Even though the structure of PC1 is similar to the iTRAQ reagent, there are considerable differences which could cause the antibody to be significantly less sensitive to PC1 crosslinked proteins compared to iTRAQ labeled proteins. To account for this expected difference in reactivity, a larger amount of PC1 labeled protein compared to iTRAQ labeled protein was loaded on the gel. Even with the additional PC1 crosslinked protein present, there was no observed signal on the western blot. This indicates that the iTRAQ antibodies were not reactive enough with PC1 crosslinked proteins but PC1 may be antigenic if coupled to a protein and used as an antigen.

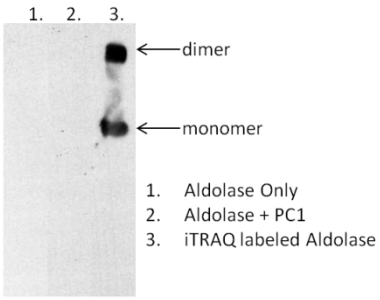


Figure 2.4. Western blot of PC1 crosslinked aldolase. Western blot of Aldolase crosslinked with PC1 probed with an antibody raised against an iTRAQ labeled protein.

2.3 MS Fragmentation of Model PC1-Crosslinked Peptides

2.3.1 MALDI-TOF-TOF Analysis

One of the most important characteristics in the design of this new crosslinker was that the crosslinked peptides were cleavable by mass spectrometry. To investigate the behavior of

PC1 crosslinked peptides in the mass spectrometer, N-terminally acetylated peptides containing an internal lysine and a C-terminal arginine were crosslinked. The lyophilized peptides were reconstituted at a very high concentration and PC1 was added to force crosslinking. The products from the crosslinking reaction were then analyzed via MALDI-TOF MS. (Figure 2.5A) The most abundant peaks were the result of dead-end reactions, but a significant amount of crosslinked peptide was observed. The crosslinked peptides were then fragmented via CID in a MALDI-TOF-TOF (2.5B) and several features were noted. Several b- and y-ions necessary for sequencing were observed, but one ion, $\mathbf{b_6}/b_7$ dominates the spectrum. (The b_6 ion in one peptide and the b_7 ion from the second peptide were the same muss and couldn't be differentiated.) This observation was explained from a mobile proton model perspective[7]. The dominant $\mathbf{b_6}/b_7$ ion resulted from loss of the C-terminal arginine residue on either peptide. In this crosslinked peptide, the arginine residues were the most basic sites. Although the nitrogen atoms in the crosslinker were basic, they were less basic than the arginine side chain and the N-termini were acetylated[8, 9]. Therefore, the majority of singly charged ions from this crosslinked peptide would be protonated at the arginine side chain. That proton was still mobile, but would be localized mostly on the arginine residue. The closest atom to the protonated arginine side chain that could accept a proton was the amide nitrogen between the C-terminal arginine and the next amino acid. When the proton shifted onto that nitrogen, the amide bond was weakened and then cleaved to produce the dominant $\mathbf{b_6}/b_7$ ion. The proton could continue to shift down the peptide backbone which resulted in the other b- and y- ions observed, but because of the high proton affinity of the arginine residues, those ions were produced less frequently. If this explanation were true, CID of the same crosslinked peptide with multiple charges should not have a dominant $\mathbf{b_6}/b_7$ ion and this was investigated in section 2.3.2.

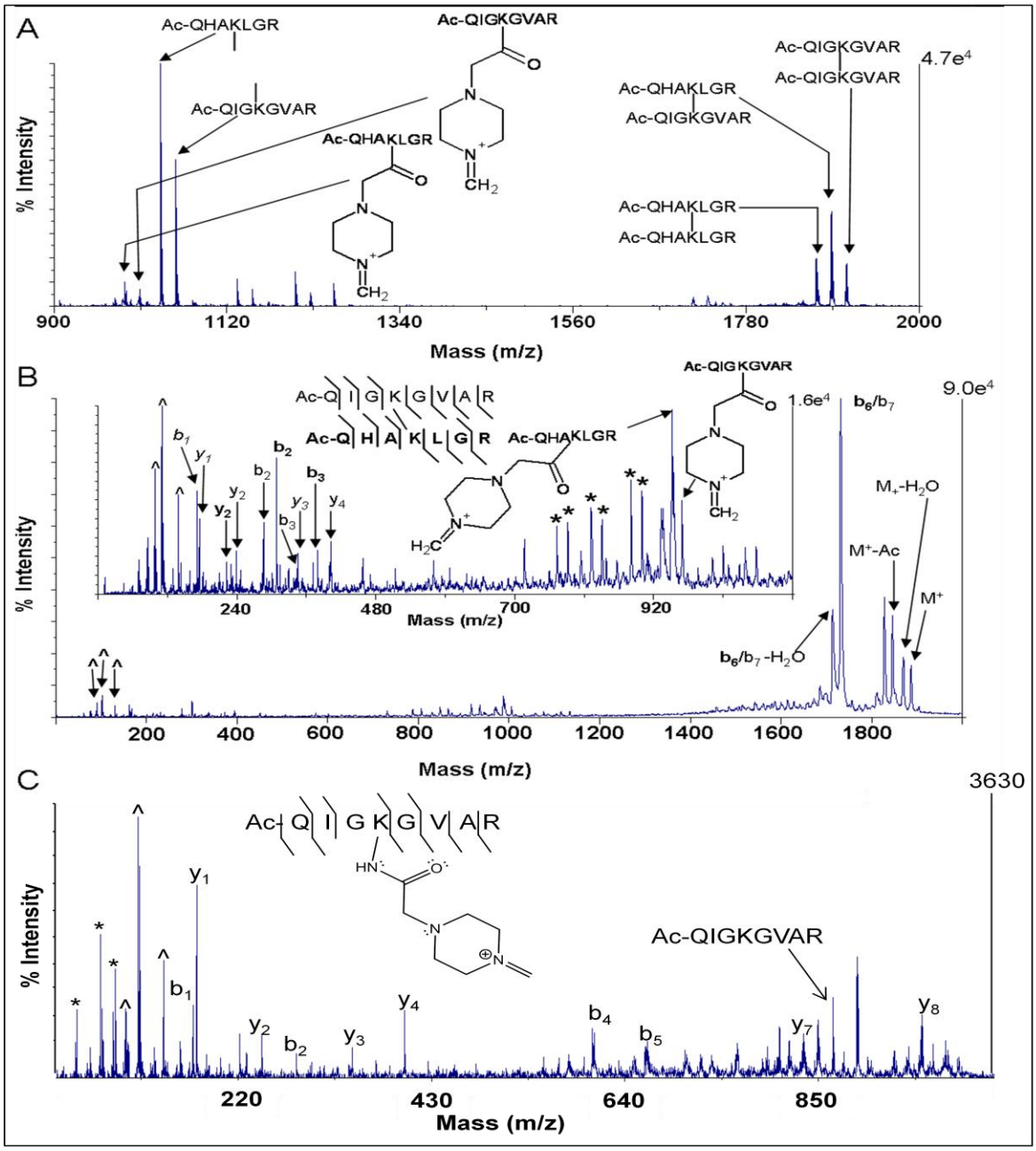


Figure 2.5. **CID on a MALDI-TOF-TOF of a PC1-crosslinked peptide.** (A) MALDI-TOF Mass spectrum of a crosslinking reaction between acetylated peptides QHAKLGR and QIGKGVAR. Dead-end reactions, crosslinked peptides and in source decay products from the crosslinked peptides are indicated. (B) MS/MS on a MALDI-TOF-TOF of precursor 1887.8 which was determined to be Ac-QHAKLGR crosslinked to Ac-QIGKGVAR. (C) MS/MS of the ISD product precursor 1008.4. ^ represents the marker ions observed in spectra from crosslinked peptides at m/z 99.1, 111.1, 139.1. Immonium ions are marked with an asterisk. Ions in bold font result from the cleavage of the peptide.

No b- and y- ion pairs were observed in the MS/MS of this crosslinked peptide, but this was peptide dependent. Alternatively, it could be due to the fact that this peptide was crosslinked directly in the middle of the sequence as it has been shown that branched peptides fragment less efficiently[10], and the peptide shown here was branched directly in the middle of the sequence. An additional model peptide that was crosslinked near the C-terminus fragmented much more efficiently (Figure 2.6) which would support the hypothesis that peptides crosslinked in the middle fragment less efficiently than peptides crosslinked at their termini. Data acquired in subsequent experiments were used to validate this hypothesis (section 2.4).

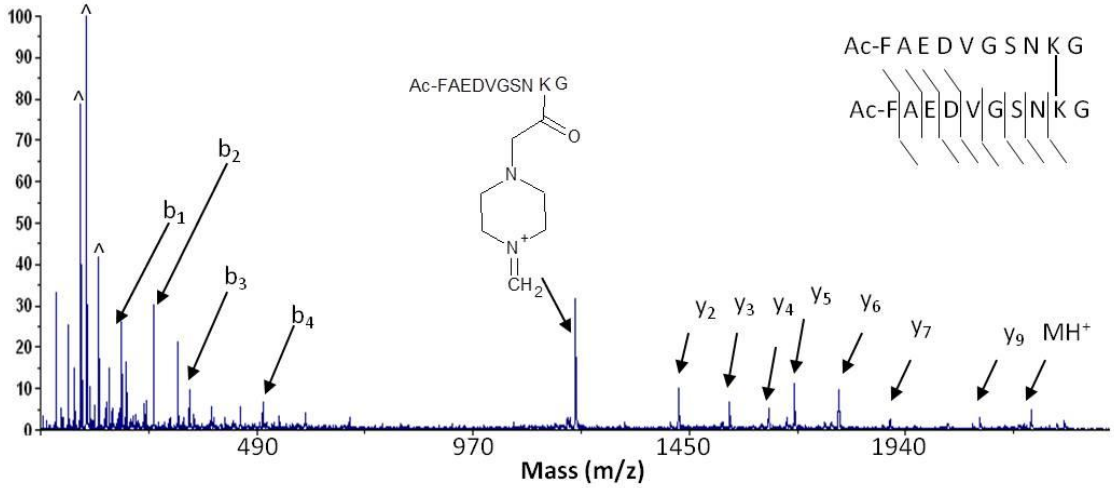


Figure 2.6. CID on a MALDI-TOF-TOF of a PC1-crosslinked peptide crosslinked near the C-termini. Peptides crosslinked at the C-termini exhibited more extensive fragmentation. Peaks marked with a carat represent the low mass marker ions.

Another noteworthy characteristic in the MS/MS of crosslinked peptides (Figure 2.5B) was in the low mass region of the spectrum where three marker ions were consistently observed corresponding to m/z values of 99.1, 111.1 and 139.1. Because there are no common ions that appear at those specific m/z values, the combination of the three marker ions facilitated a confident identification of the spectrum as a crosslinked peptide. Potential structures of these marker ions and mechanisms for obtaining the structures were proposed (Figure 2.7). This

mechanism required that the amide formed between the crosslinker and the peptide must be protonated. Then all three marker ions resulted from the same initial fragmentation event. The lone pair of electrons on the piperazine nitrogen was delocalized out of the ring, resulting in a carbon bond cleavage which produces carbon monoxide and one original unprotonated peptide. The remaining peptide contained the now charged piperazine ring and additional rearrangements lead to the fragments shown in Figure 2.7.

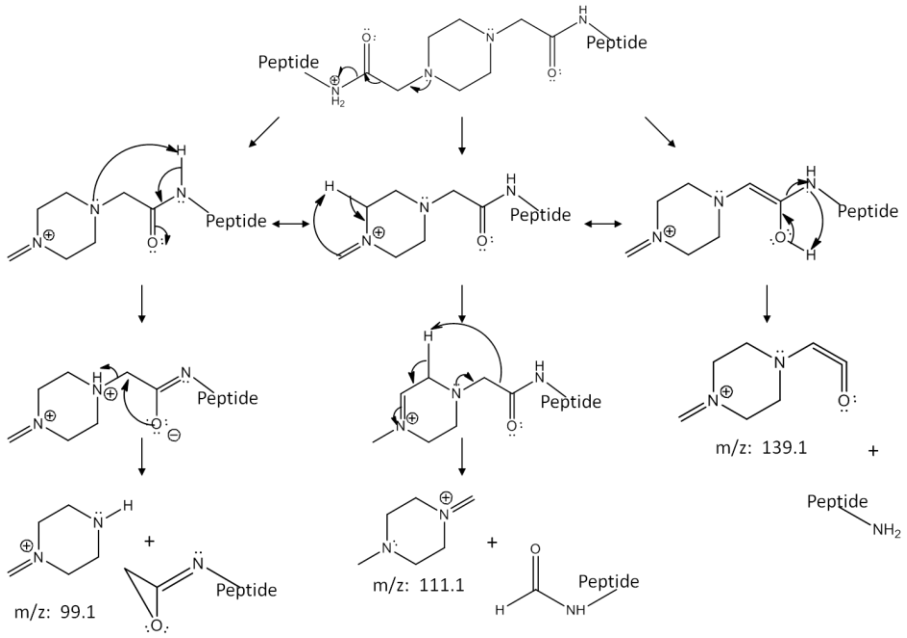


Figure 2.7. PC1 fragmentation mechanism. Suggested fragmentation mechanism of a PC1 crosslinked peptide.

In the MS spectrum of the crosslinked sample, m/z values were observed that corresponded to each peptide containing the charged piperazine ring – the initial fragmentation event which resulted in the three marker ions. MS/MS experiments on those precursors supported the identification of those peaks (Figure 2.5C) and clearly demonstrated that fragmentation occurred within the crosslinking moiety. Because the peptides with the piperazine ring were observed in the MS, the original crosslinked peptide must have been fragmented by in source decay (ISD). Further fragmenting a product ion from an MS/MS experiment is known as

MS^3, however in true MS^3, two precursor isolation and activation events occur. First, all the masses are observed without fragmentation, and one of those masses is isolated and activated to yield MS/MS data. Then, a precursor is isolated and activated from the MS/MS data to yield MS^3 data. With PC1 crosslinked peptides, the first fragmentation event was ISD and only one precursor was isolated, so pseudo MS^3 was a more appropriate description for the fragmentation. However, based on the proposed structure of the ISD peaks, the charge was localized and no mobile proton was available, so pseudo MS^3 cleavage must have occurred through a charge remote fragmentation mechanism[11]. The three marker ions were also observed in the MS/MS of the masses from ISD which was expected based on the proposed mechanisms in Figure 2.7.

Tandem mass spectrometry of the dead-end product (Figure 2.8) yielded a marker ion clearly visible at m/z 157.1 (Figure 2.9) which was produced in a manner analogous to the crosslinked peptides. The presence of this ion unambiguously identified this spectrum as being from a dead-end peptide and a complete series of b- and y-ions allowed for the sequencing of this peptide.

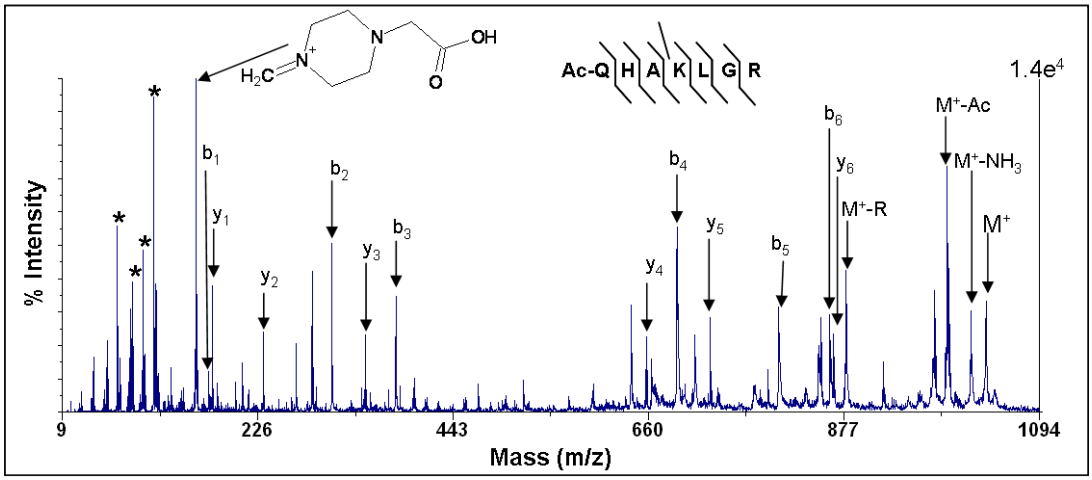


Figure 2.8. CID of a PC1-dead-end peptide on a MALDI-TOF-TOF. MS/MS spectrum of precursor 1035.5 in the MALDI-TOF-TOF. The complete series of b- and y- ions are observed, along with a marker ion at m/z 157.2 that facilitates identification of this spectrum as fragmentation of a dead-end peptide. Peaks marked with an asterisk are immonium ions from amino acids in the peptide.

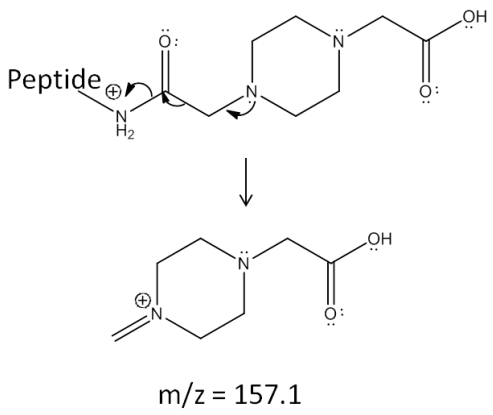


Figure 2.9. **PC1 dead-end fragmentation mechanism.** Fragmentation mechanism and the resulting marker ion from CID of a PC1 modified dead-end peptide.

2.3.2 FTICR Analysis of PC1 Crosslinked Peptides

CID fragmentation of PC1 crosslinked peptides was also evaluated on a Bruker FTICR instrument (Figure 2.10A). Fragmentation within the crosslinking moiety was observed as with the MALDI-TOF-TOF experiments. CID on both instruments produced a similar set of fragments, but when the 3+ charge state was fragmented using the FTICR the $\mathbf{b_6}/b_7$ ion was not observed and many ions were of comparable intensity, resulting in a more even distribution of ions. This supported the earlier argument that the dominant $\mathbf{b_6}/b_7$ ion was due to the single charge and preferential cleavage of the arginine residue. Fragmentation using ECD resulted in homolytic cleavage within the crosslinking moiety; however, as opposed to CID, cleavage occurred at the site of the piperazine ring as shown in Figure 2.10B. The mechanism of ECD fragmentation is significantly different[12-14] than the mobile proton model with CID[7], so it was reasonable that different bonds within the crosslinker were cleaved. Most of the expected c- and z- ions were observed with good intensity so that the ECD spectrum resulted in a more confident identification of the crosslinked peptide compared to the CID spectrum. Upon IRMPD fragmentation, cleavage within the crosslinking moiety was not observed (Figure 2.10C). The

lower energy used for fragmentation was not high enough to induce fragmentation within the crosslinking moiety but there still were a significant number of reasonably intense b- and y-ions. The distinct marker ions that were observed in MALDI-TOF-TOF CID were not observed using the FTICR. One possible explanation was the energy used for CID in the MALDI-TOF-TOF was an order of magnitude more than in the FTICR (2 keV vs. ~200 eV).

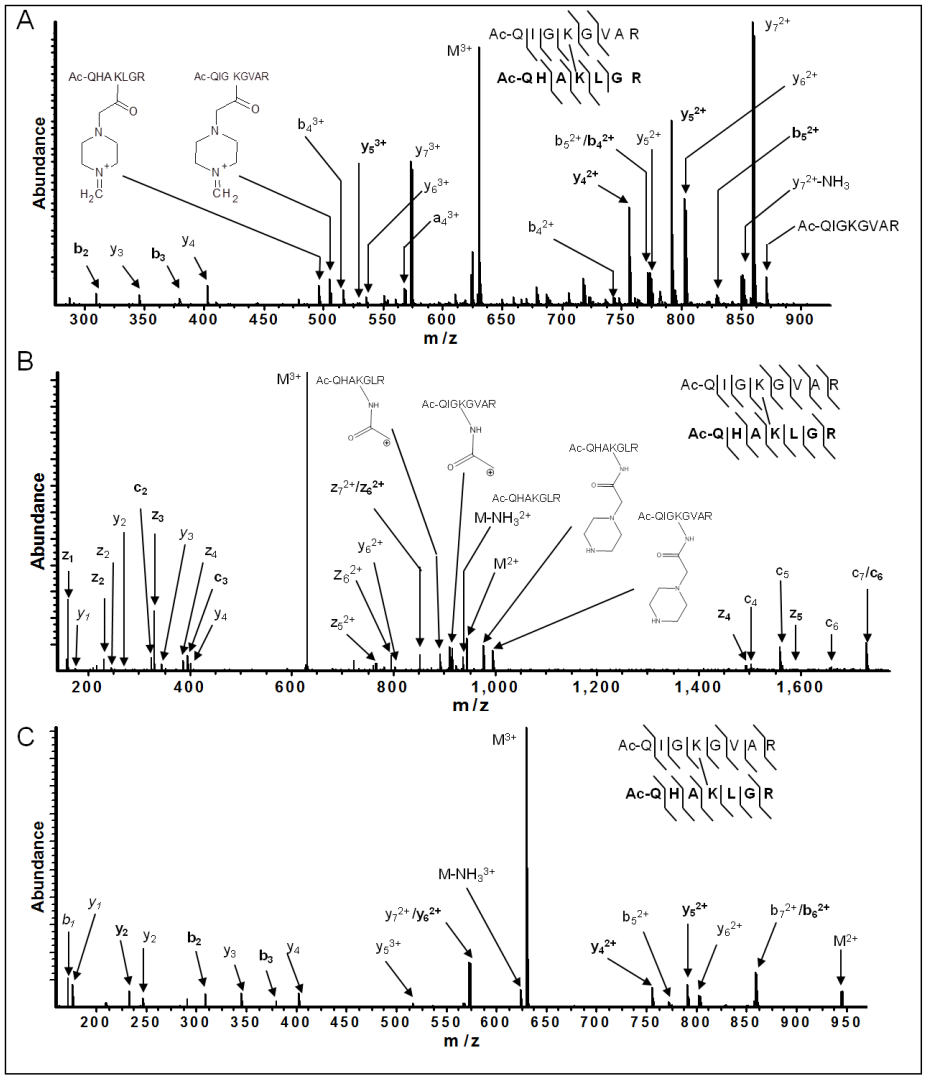


Figure 2.10. **PC1-crosslinked peptide fragmentation in an FTICR.** CID fragmentation in the FTICR of crosslinked peptide (precursor m/z 629.6). Fragmentation within the crosslinker moiety was observed in CID (A) and ECD (B) fragmentation but not IRMPD (C). However all three fragmentation methods provided sufficient fragmentation to identify the crosslinked peptides.

2.3.3 LTQ-Orbitrap Analysis

The fragmentation of crosslinked peptides was also evaluated in the LTQ-Orbitrap by CID (Figure 2.11A) and HCD (Figure 2.11B). Both methods of fragmentation resulted in fragmentation within the crosslinking moiety as was also observed with CID on both the MALDI-TOF-TOF and the FTICR. Sufficient b- and y- ions were observed to facilitate the identification of the crosslinked peptide. The marker ions observed by MALDI-TOF-TOF were not observed via HCD or CID on the Orbitrap. The Orbitrap is an electrospray instrument, the marker ions could have been multiply charged which would have resulted in m/z values below the mass range accessible to the instrument. On the MALDI-TOF-TOF, the ions were singly charged and of good intensity suggesting that if the ions were formed in the Orbitrap, they should have been observed. A more likely possibility is that the higher energy in MALDI-TOF-TOF was required for the rearrangements needed to produce the marker ions.

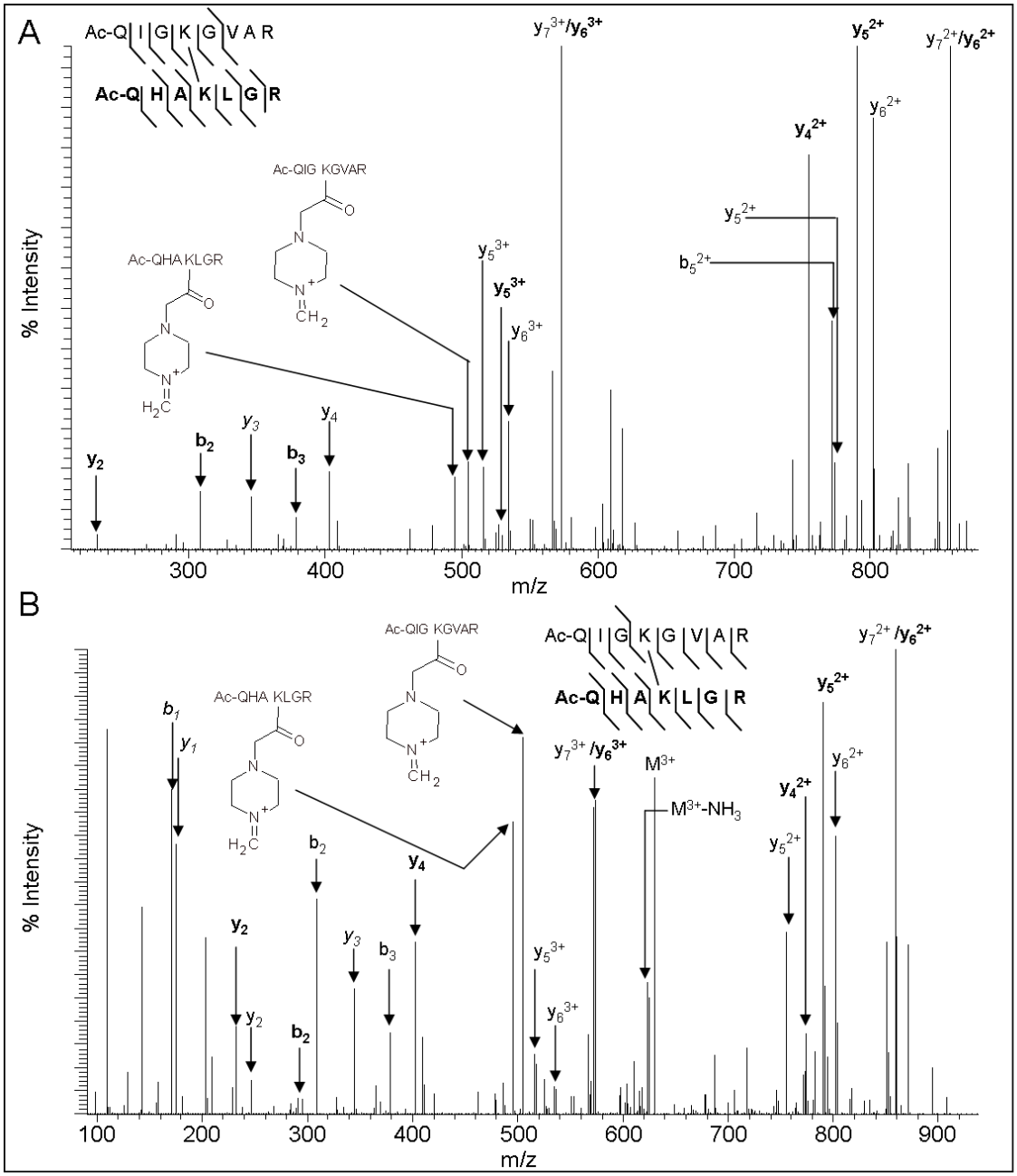


Figure 2.11. **PC1-crosslinked peptide fragmentation in an LTQ-Orbitrap.** LTQ-Orbitrap FTICR fragmentation of Ac-QHAKLGR crosslinked with Ac-QIGKGVAR, precursor 629.6 via CID (A), HCD (B).

2.4 Application of PC1 to the Tetrameric Protein Aldolase

After characterization with model peptides, PC1 was then applied to aldolase, digested with multiple enzymes, and manually analyzed for crosslinked peptides on the MALDI-TOF-TOF. Unmodified peptides and multiple dead-end peptides were observed as shown in Table 2.1. The dead-end peptides behaved as expected and fragmented into b- and y- ions as well as the marker ion (Figure 2.12A). Several ions were then identified as potentially crosslinked peptides but upon CID fragmentation, few fragments were observed other than the three low mass marker ions (Figure 2.12B). This confirmed that the spectrum was either from a crosslinked peptide or an ISD fragment from a crosslinked peptide; however, sufficient fragmentation of the crosslinked peptides was not achieved to allow identification by either de novo sequencing or database searching. These results indicated fragmentation of the PC1 crosslinker via MALDI-TOF-TOF was not adequate for broader application to peptide mixtures.

Table 2.1. Aldolase peptides observed after crosslinking with PC1 and MALDI-TOF-TOF analysis. Stars indicate modified residues and when there are more than one set of residues loop-linked within the same peptide.

Sequence	m/z	Modification
ALQASALKAWGGKKENLKAAQEEYVKR	3706.21	Loop-Link & 3 Dead-ends
*PHSHPALTPEQK*K*ELSDIAHR	2741.64	Loop-link & Dead-end
FHETLYQK*ADDGRPFPQVIK	2573.52	Dead-end
VDK*GVVPLAGTNGETTTQGLDGLSER	2798.56	Dead-end
IVAPGKGILAADESTGSIAKR	2422.55	Dead-end
YQK*ADDGGRPFPQVIK*SK	2345.41	Dead-end
K*ENLK*AAQEEYVK*R	2258.22	Dead-end
YQK*ADDGRPFPQVIK	1946.15	Dead-end
IGEHTPSALAIMENANVLAR	2107.16	None
HACTQKYSHEEIAMATVTALR	2361.24	Deamidated
ALSDHHIYLEGTLLKPN	1920.79	None
AMDNHPQQTQS	1274.68	Deamidation & Oxidation
QLLLTADDRVNPCIGGVIL	2010.28	None
EHTPSALAIMENANVLAR	1918.99	None

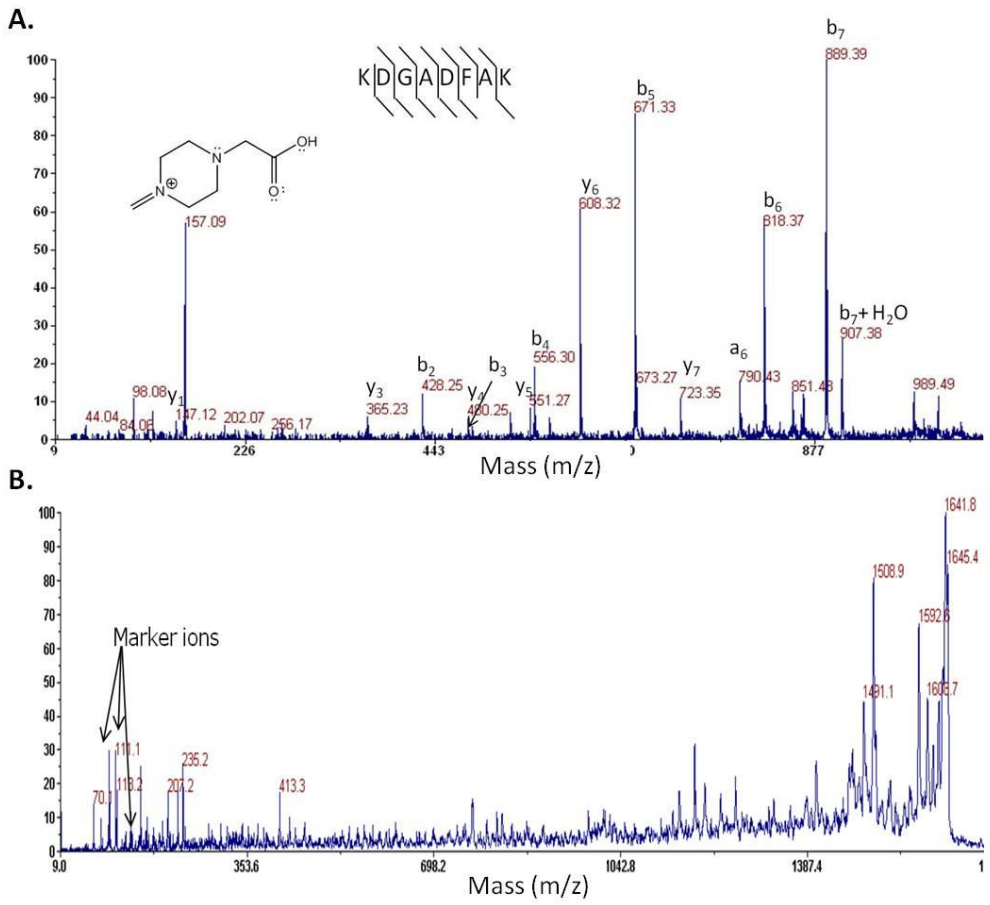


Figure 2.12. CID of PC1-modified peptides from aldolase. (A) Aldolase peptides that were dead-ends fragmented into b- and y- ions as well as the characteristic marker ion at m/z 157. (B) The marker ions clearly indicated this peptide was either crosslinked or an ISD peak from a crosslinked peptide. However, there are insufficient b- and y- ions to sequence the peptide.

Similar aldolase digests were run on an ESI-LTQ-Orbitrap and several dead-end and loop-link peptides were observed (Table 2.2). Several more dead-ends and loop-links were observed in the Orbitrap as opposed to the TOF-TOF, potentially due to their size. The dead-end peptides observed in the Orbitrap were large and multiply charged. MALDI primarily generates singly charged ions and because fragmentation (and detection) efficiency in the MALDI-TOF-TOF decreases with increasing m/z value increases, the large dead-end peptides may not have been observed or may not have fragmented efficiently enough to be identified.

Table 2.2. Aldolase peptides observed after crosslinking with PC1 and analysis on an ESI-LTQ-Orbitrap. Stars indicate modified residues and when there are more than one set of residues loop-linked within the same peptide, numbered stars indicate which are residues are crosslinked to each other.

Sequence	m/z	Modification
GGVVGIK*VDK*GVVPLAGTNGETTTQGLDGLSER	1131.6	Loop-link
ALQASALK*AWGGK*K	798.46	Loop-link
ALQASALK*^{1}AWGGK*^{1}K*^{2}ENLK*^{2}AAQEEYVK	1055.57	Loop-link
SKGVVGIK*VDK*GVVPLAGTNGETTTQGLDGLSER	902	Loop-link
ADDGRPFPQVIK*SK*GGVVGIK	584	Loop-link
IVAPGK*GILAADESTGIAK*R	1032	Loop-link
GGVVGIK*	407.2	Dead-end
VLAAVYK*	474.2	Dead-end
ALQASALK*	493.3	Dead-end
GGVVGIK*VDK	578.4	Dead-end
ALQASALK*AWWGK	742.9	Dead-end
ENLK*AAQEEYVK	803.4	Dead-end
DGADFAK*WR	625.3	Dead-end
GILAADESTGSIAK*R	558.3	Dead-end
IVAPGK*GILAADESTGSIAK	1041.5	Dead-end
VDK*GVVPLAGTNGETTTQGLDGLSER	933.4	Dead-end
SK*GGVVGIK*	606.8	Dead-end
AAQEEYVK*R	606.8	Dead-end
IVAPGK*GILAADESTGSIAK*R	808.1	Dead-end
K*DGADFAK	518.2	Dead-end
YASICQQNGIVPIVEPEILPDGDHDLK*R	1102.5	Dead-end
GGVVGIK*VDK*GVVPLAGTNGETTTQGLDGLSER	1198.6	Dead-end
GILAADESTGSIAK*	758.8	Dead-end
ENLK*AAQEEYVK*R	643.3	Dead-end
VNPCIGGVILFHETLYQK	677.9	None
ALANSLACQGIK	538.2	None
RTVPPAVTGVTFLSGGQSEEEASINLNAINK	1067.5	None
TVPPAVTGVKLSGGQSEEEASINLNAINK	1015.2	None
RLQSIGTENTEENR	823.9	None
VDKGVVPLAGTNGETTTQGLDGLSER	872.1	None
AAQEEYVKR	547.3	None
SKGGVVGIK	422.7	None
GVVPLAGTNGETTTQGLDGLSER	1136.5	None
IGEHTPSALAIMENANVLAR	1054.5	None
YSHEEIAMATVTALR	564.6	None
DGADFAK	362.1	None
VLAAVYK	382.2	None
ALQASALK	401.2	None
ELSDIAHR	470.7	None
QLLLTADDR	522.7	None
GILAADESTGSIAKR	66.8	None
ADDGRPFPQVIK	671.8	None
LQSIGTENTEENR	745.8	None

On the Orbitrap, the dead-end peptides exhibited the typical fragmentation patterns (data not shown). Although the fragmentation of loop-link peptides was not characterized with model peptides, loop-links were expected to fragment into b- and y- ion pairs. Unlike with crosslinked peptides, two peptides were not being simultaneously fragmented so the spectra were greatly simplified and easily identified. With loop-links, cyclic peptides are formed so fragmentation within the crosslinker was expected but the aldolase loop-link peptides did not exhibit fragmentation within the crosslinker (Figure 2.13).

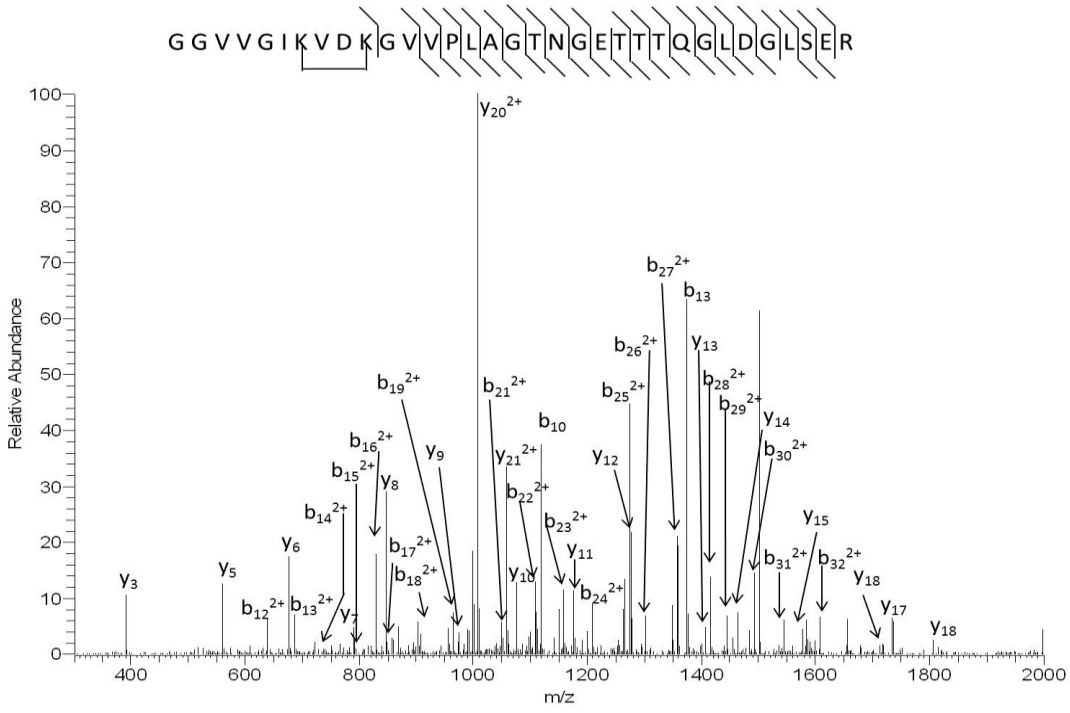


Figure 2.13. CID on the ESI-LTQ-Orbitrap of a PC1-loop-linked peptide from aldolase. This spectrum has sufficient b- and y- ions to identify the peptide; however no ions were observed on the N-terminal side of the loop-link and no fragmentation within the crosslinker was observed.

Crosslinked peptides from aldolase were not identified on the Orbitrap but the spectra may have been too complex to manually analyze and identify crosslinked peptides. A peptide's

maximum charge state can be estimate by adding up the number of basic residues (lysine, arginine, histidine) and the N-terminus[15]. Most tryptic peptides will have at least one basic residue plus the N-terminus so they exist primarily as +2 or +3 charge states. Crosslinked tryptic peptides contain two N-termini and at least two basic residues, so they should average higher charge states. In the Orbitrap data, many MS/MS spectra were collected on peptides with +4 and +5 charge states. These large, multiply charged species could represent the PC1 crosslinked peptides, but the spectra were too complex to be manually interpreted.

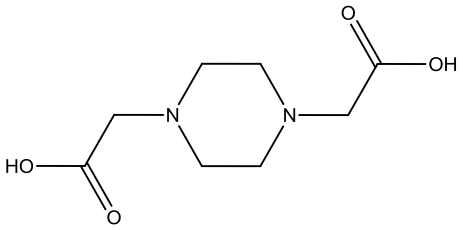
Several observations were made that lead to the conclusion that higher energy is needed to induce fragmentation within the crosslinker. In the MALDI-TOF-TOF data the marker ion from dead-end peptides was a dominant ion (the marker ion resulted from fragmentation within the crosslinker) in the spectra of both acetylated and non-acetylated peptides. However, when the acetylated peptides were fragmented in the Orbitrap, the marker ion was very low in intensity. Also, three marker ions for crosslinked peptides were consistently observed with the MALDI-TOF-TOF that were not observed with the ESI-Orbitrap. The suggested mechanisms (Figure 2.7) for obtaining those ions utilized intermediate structures that would have required higher energy. The collision energy in the TOF-TOF (2 keV) is significantly higher than in the Orbitrap (200eV), so it was possible that higher energy was required for crosslinker fragmentation. The model peptides were analyzed via IRMPD which is an even lower energy fragmentation mechanism and no fragmentation within the crosslinker was observed. From these observations, it was concluded that PC1 crosslinked peptides did not fragment well enough to aid in crosslinked peptide identification for routine use.

2.5 Conclusions:

Upon initial analysis, acetylated PC1 crosslinked peptides fragmented well via CID, HCD, IRMPD, and ECD. CID of those peptides on a MALDI-TOF-TOF produced marker ions unique to PC1 crosslinked peptides which would be especially useful for identifying a spectrum as being produced from a crosslinked peptide. However, the facile fragmentation of the model acetylated peptides contrasted with the insufficient fragmentation of the larger peptides from aldolase. The lack of fragmentation (MALDI-TOF-TOF) or complicated fragmentation (ESI-LTQ-Orbitrap) of PC1 crosslinked peptides demonstrated that PC1 did not meet the first criterion in the design of a new crosslinker. As this was a critical requirement, additional crosslinkers were designed and synthesized as discussed in Chapter 3.

2.6 Materials and Methods:

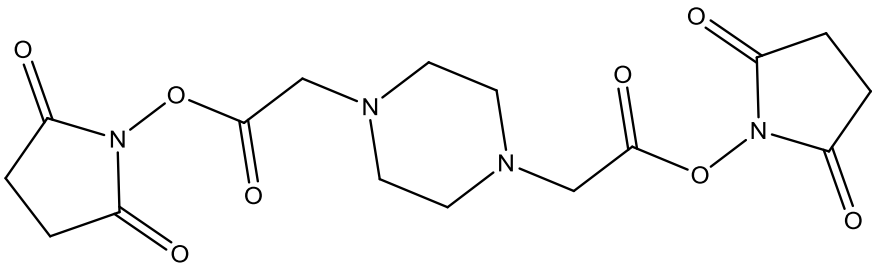
Materials. N-terminally acetylated peptides QIGKGVAR, QHAKLGR, were synthesized by the Biomedical Research Core Facility at the University of Michigan Medical School using standard Fmoc-based methods, a PTI Symphony synthesizer and purification by reversed-phase HPLC. All peptides were at least 90% pure and had the expected intact masses and fragmentation patterns. All reagents and solvents were commercially available and used without further purification unless specified. Melting points were determined in open capillary tubes on a Laboratory Devices Mel-Temp apparatus and are uncorrected. The NMR spectra were recorded on a Bruker instrument at 500 MHz for ^{1}H and 125 MHz for ^{13}C spectra. Chemical shift values are recorded in δ units (ppm). Mass spectra were recorded on a Micromass LCT Time-of-Flight mass spectrometer with an electrospray ionization source in positive ion mode.



1

2,2'-(Piperazine-1,4-diium-1,4-diyl)diacetate (1).

Piperazine (2.0 g, 23.2 mmol) and bromoacetic acid (6.65 g, 47.8 mmol) were dissolved in NaOH (10%, 50 ml). The resulting solution was stirred for 1 hour at room temperature. Colorless needles (sodium 2,2'-(piperazine-1,4-diyl)diacetate) were formed and collected by filtration. The obtained needles were dissolved in water and the solution adjusted to pH 5-6 with concentrated HCl. The white precipitate was collected by filtration and dried in vacuo to give 2,2'-(piperazine-1,4-diium-1,4-diyl)diacetate (4.26 g, 91%). ^{1}H NMR (D_2O) δ 3.87 (s, 4H; 2 x CH_2), 3.61 (s, 8H; 4 x CH_2); ^{13}C NMR ($CDCl_3$) δ 168.5, 56.9, 49.1; MS (ES^+), m/z 203.0 $[M + H]^+$; MS (ES^-), m/z 201.0 $[M - H]^-$; M.P. 250-252^{o}C.



2

Bis(2,5-dioxopyrrolidin-1-yl) 2,2'-(piperazine-1,4-diyl)diacetate (2).

To a suspension of **(1)** (2.59 g, 12.8 mmol) in dichloromethane (50 ml) was added thionyl chloride (3.7 ml, 51.2 mmol) over 10 mins at room temperature. The resulting suspension was heated at reflux for 4 hours. Solvent and excess reagents were removed under reduced pressure

to give a white solid. A cooled solution of N-hydroxysuccinimide (NHS, 4.42 g, 38.4 mmol) and N,N-diisopropylethylamine (10.7 ml, 76.8 mmol) in dichloromethane (50 ml) was added to the solid. The resulting suspension was allowed to stir at room temperature for 12 hours. Dichloromethane (500 ml) was added to the reaction mixture. This organic solution was washed with water (3 x 300 ml), saturated brine (1 x 300 ml), dried over Na_2SO_4 and concentrated to give an off-white solid as crude product (1.26 g). The crude product was triturated twice in dichloromethane to give pure bis(2,5-dioxopyrrolidin-1-yl) 2,2'-(piperazine-1,4-diyl)diacetate as a white solid (950 mg, 20 %). R_f 0.04 (100% EtOAc); ^{1}H NMR ($CDCl_3$) δ 3.61 (s, 4H; 2 x CH_2), 2.87 (s, 8H; 4 x CH_2), 3.77 (s, 8H; 4 x CH_2); ^{13}C NMR ($CDCl_3$) δ 168.8, 165.3, 56.4, 52.2, 25.6.; m/z 397.1 $[M + H]^+$; M.P. decomposed at 168-170^{o}C.

Crosslinking Reactions. Directly prior to all aldolase experiments, 10 mM PC1 was prepared in 100 mM HEPES pH 7.0. For SDS-PAGE 10μg aldolase (2 μl of 5 μg/μl) were used for each lane. PC1 to aldolase monomer was added over a range of ratios from 1-500:1. Then 100 mM HEPES pH 7.0 was added to keep the total buffering capacity and the concentration of aldolase constant. For mass spectrometry analysis, a 40 molar excess of PC1 per aldolase monomer was used to crosslink 5 μg/μl aldolase solution. The model peptide work was completed using custom designed and synthesized peptides. The Biomedical Research Core Facility synthesized, purified, and lyophilized the peptides we designed. Those peptides were to be positively charged at neutral pH, contain an internal lysine, and a C-terminal arginine. Peptide solutions were prepared (~5 μg/μl in water) and then reacted with a 5 molar excess of 5 mM PC1, which was prepared immediately before use in 50% acetonitrile, 100 mM HEPES, pH 7.0. Reactions were carried out at room temperature for at least 30 mins and up to two hours.

SDS-PAGE and Western Blot. Gradient gels (4-12% Bis Tris) were purchased from Invitrogen. Crosslinked aldolase samples (10 μg in 20 μl) were combined with 5 μl 5X SDS loading dye (0.1 M Tris-HCl, pH 6.8, 20% glycerol, 2% SDS, 0.02% bromophenol blue, 0.1M dithiothreitol (DTT)) and heated for 10 mins at 95°C. Samples were loaded onto the gels which were then electrophoresed 150V until the bromophenol tracking dye ran off the gel. Gels for direct visualization were then submerged in Coomassie G-250 (0.1% Coomassie G-250, 1.28 M ammonium sulfate, 36% Methanol, 3% phosphoric acid) overnight at room temperature and destained for several days in water.

For western blots, gels, whatman paper, and nitrocellulose were submerged in transfer buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, 10% methanol) for five mins. The proteins were transferred for two hours in a semi-dry transfer apparatus using constant current at an amperage of 0.8 times the area of the gels being transferred as recommended by the manufacturer (Hoefer). The membranes were blocked in 5% dry milk prepared in 0.1% Tween, 137 mM NaCl, 2.7mM KCl, 10 mM Na_2HPO_4, 1.76 mM KH_2PO_4, pH 7.4 (PBST) for one hour at room temperature, shaking. Next, the membranes were treated with the anti-iTRAQ as primary antibody diluted 1:10,000 in 5% milk prepared in PBST for one hour at room temperature. The membranes were washed three times with PBST, once for 15 mins, and twice for five mins before the secondary antibody was added. The secondary goat anti rabbit, peroxidase conjugated antibody was purchased from Pierce, prepared in a 1:10,000 dilution of antibody to 5% milk prepared in PBST. Membranes were incubated in secondary for 1 hour at room temperature and the wash steps repeated as after the primary antibody. Chemilluminescence was used as the detection method. An equal volume of a solution of 2.5 mM luminal, 0.295 mM Loumaric acid, 100 mM Tris, pH 8.5 was combined with 2% hydrogen

peroxide, 100 mM Tris pH 8.5 and applied directly to the membranes. Kodak blue basic film was exposed to the membranes for an appropriate amount of time and then processed in an Kodak M35A X-OMAT processor.

Sample Preparation. Prior to mass spectrometry, the crosslinked synthetic peptides were dried using a vacuum centrifuge. The dry peptides were reconstituted in 10 μL 0.1% trifluoroacetic acid and Millipore C_{18} ZipTips were used to desalt the fractions according to the manufacturer's instructions. After crosslinking intact aldolase, the pH was adjusted to 8.0 with 100 mM triethylammonium bicarbonate and then the solution was digested with L-1-tosylamido-2-phenylethyl chloromethyl ketone (TPCK) treated trypsin from Worthington Biomedical Corporation at a 10:1 mass ratio (aldolase/trypsin) overnight at 37°C. The tryptic digest was further digested with Endoprotease GluC (50:1) overnight at 37°C.

MALDI-TOF-TOF. The crosslinked synthetic peptides were manually spotted onto a MALDI target using 5 mg/ml 2,5-dihydroxybenzoic acid (DHB), 5 mg/ml α-cyano-4-hydroxycinnamic acid (αCHA) in 50% acetonitrile 0.1% trifluoracetic acid (TFA). For the PC1 crosslinked aldolase, the digests were separated using an HPLC equipped with a Gilson 811C dynamic mixer, Gilson 307 and 305 pumps, Scientific Systems Model CP-21 Lo Pulse, Amerisham Biosystems 759A absorbance detector set at 214nm, and Isco Foxy 200 fraction collector with a C_{18} PLRP-S 5 micron 100 Angstrom column. Solvent A was 5% acetonitrile, 0.1% TFA while solvent B was 90% acetonitrile 0.1% TFA. The flow rate was set to 1 ml/min with the following binary gradient: 0 min, 6.5% B, 14 min, 6.5% B, 17 min, 18% B, 62.4 min, 40%B, 77.8 min, 50% B, 83.4min, 70% B, 84.8 min, 100% B, 86.2 min, 100% B, 87.6 min, 6.5% B, 96 min, 6.5% B. One min fractions were collected and then dried using a Jouan speedvap. Those fractions were reconstituted in 20 μL of 0.1% TFA, sonicated 30 mins or more

and manually spotted onto a MALDI target using 5 mg/ml DHB, 5 mg/ml α-cyano-4-hydroxycinnamic acid (αCHA) in 50% acetonitrile 0.1% trifluoracetic acid (TFA).

Both the standard peptides and the crosslinked aldolase digest spectra were acquired in positive reflector mode in the 4800 TOF/TOF mass spectrometer (Applied Biosystems Sciex). MS spectra were acquired from 600 m/z to 3,500 m/z in each fraction. CID fragmentation occurred using a collision energy of 2 keV and atmosphere was used as the collision gas in the medium gas setting. The eight most abundant peaks in each well were automatically selected for CID fragmentation.

ESI-LTQ-Orbitrap. The crosslinked standard peptides were analyzed with a hybrid linear quadrupole ion trap–orbitrap mass spectrometer (ThermoFisher Scientific, Inc. model LTQ-Orbitrap XL, San Jose, CA, USA). They were introduced by infusion using a TriVersa Nanomate nanospray ionization source from Advion BioSciences (Ithaca, NY, USA). Peptides were delivered with a pneumatic displacement pressure of 0.45 psi and an applied spray voltage of 1.63 kV. The heated capillary temperature was set at 200 ºC. The aldolase digests were separated on a house made C_{18} column, 12 cm long, 5 μm, 300 Å, 99 μm column ID using the same gradient described above for MALDI-TOF-TOF. Monoisotopic precursor ions were selected for multi stage tandem mass spectrometric acquisition using an isolation window of 3.0 (m/z units) and excitation energy settings of 35 for both MS^2 and 35 for MS^3 spectra. All CID spectra were collected in FT (Orbitrap) mode for high mass accuracy (> 10 ppm) and peak resolution (30000 at *m/z* 400).

FTICR. The crosslinked standard peptides were analyzed with an actively shielded 7 Tesla quadrupole-FTICR mass spectrometer (APEX-Q, Bruker Daltonics, Billerica, MA). The

samples were directly infused using a syringe pump at a flow rate of 70 µl/hour into an electrospray ionization source (Apollo II, Bruker Daltonics). The source was operating in positive ion mode, used a voltage of -3.8 kV and had a counter flow of nitrogen gas at 240°C to desolvate the ESI droplets. Ions within a mass window of 5-10 Da were collected in the first hexapole, then transferred through the mass selective quadrupole to a second hexapole before being transferred via high-voltage ion optics and trapped in the ICR cell by gated trapping. This ion accumulation sequence was repeated up to five times to maximize precursor signal. All FTICR spectra were acquired using XMASS (version 6.1, Bruker Daltonics) using 512 K data points summed over 64 scans. The MIDAS analysis software was used to process the FTICR data. Three fragmentation methods were utilized in the FTICR. For ECD, multiply charged ions were irradiated 30-80ms with electrons generated from an indirectly heated hollow dispenser cathode at a bias voltage of 0.01-0.50 V. The lens electrode in front of the hollow cathode was kept at 1.0 V. To achieve fragmentation via IRMPD the ions were irradiated for 100-300 ms by 10.6 µm photons at 10 W laser power (25W CO2 laser, Synrad, Mukilteo, WA). In CAD, argon was pulsed into the ICR cell until a pressure reading of 10-7 mbar (gauge calibrated for N_2) was achieved. Simultaneously a radio frequency was applied to the cell that differed from the precursor ion's cyclotron frequency by -1000 Hz.

2.7 References

[1] A. Sinz, *Mass Spectrometry Reviews* **2006**, *25*, 663.

[2] S. Kalkhof, A. Sinz, *Analytical Biochemistry* **2008**, *392*, 305.

[3] J. W. Back, A. F. Hartog, H. L. Dekker, A. O. Muijsers, L. J. d. Koning, L. d. Jong, *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 222.

[4] P. L. Ross, Y. Huang, N., J. Marchese, N., B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D. J. Pappin, *Molecular & Cellular Proteomics* **2004**, *3*, 1154.

[5] A. Maurady, A. Zdanov, D. d. Moissac, D. Beaudry, J. Sygusch, *The Journal of Biological Chemistry* **2002**, *277*, 9474.

[6] M. Takashi, Y. Zhu, Y. Nakano, K. Miyake, K. Kato, *Urological Research* **1992**, *20*, 307.

[7] A. R. Dongre, J. L. Jones, A. Somogyi, V. H. Wysocki, *Journal of the American Chemical Society* **1996**, *118*, 8365.
[8] A. G. Harrison, *Mass Spectrometry Reviews* **1997**, *16*, 201.
[9] A. G. Harrison, T. Yalcin, *International Journal of Mass Spectrometry and Ion Processes* **1997**, *165–166*, 339.
[10] O. Jahn, H. Tezval, J. Spiess, K. Eckart, *International Journal of Mass Spectrometry* **2003**, *228*, 527.
[11] M. L. Gross, *International Journal of Mass Spectrometry and Ion Processes* **1992**, *118*, 137.
[12] R. A. Zubarev, N. A. Kruger, E. K. Fridriksson, M. A. Lewis, D. M. Horn, B. K. Carpenter, F. W. McLafferty, *Journal of the American Chemical Society* **1999**, *121*, 2857.
[13] R. A. Zubarev, D. M. Horn, E. K. Fridriksson, N. L. Kelleher, N. A. Kruger, M. A. Lewis, B. K. Carpenter, F. W. McLafferty, *Analytical Chemistry* **2000**, *72*, 563.
[14] F. Tureček, *Journal of the American Chemical Society* **2003**, *125*, 5954.
[15] T. R. Covey, R. F. Bonner, B. I. Shushan, J. Henion, *Rapid communications in mass spectrometry : RCM* **1988**, *2*, 249.

Chapter 3

Synthesis of Mass Spectrometry Cleavable Crosslinkers and Tagging Reagents Containing Quaternary Amines

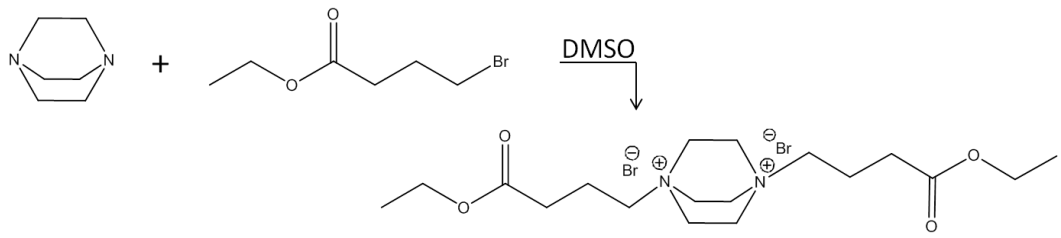
3.1 Development of Synthetic Pathways for NHS Crosslinkers Containing Quaternary Amines

As described in Chapter 2, PC1 crosslinked peptides did not fragment as efficiently as anticipated, and water solubility was poor. To address these issues minor structural modifications were incorporated into the structure that would both alleviate the water solubility and provide more facile cleavage by CID. The original plan was to methylate the nitrogen atoms of the piperazine ring of PC1 thus introducing quaternary amines in the ring structure. However, attempts to methylate PC1 using common methylating agents such as methyl iodide or methyl triflate did not produce the desired product. The NMR spectra were very complicated, and mass spectrometry data suggested the presence of the starting material and some monomethyl adduct, but not the desired dimethylated product.

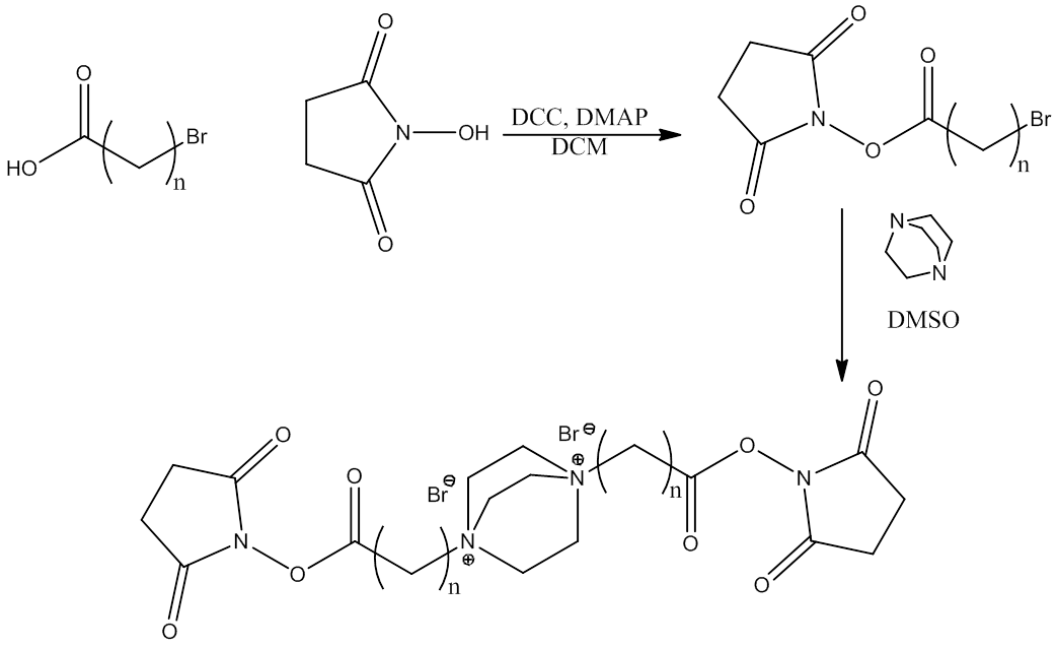
The synthetic route was changed to couple N,N-dimethyl piperazine with ethyl 2-bromoacetate and the ethyl ester, 1,4-bis(2-ethyoxy-3-oxoetheyl)-1,4-dimethylpiperazine-1,4,diium bromide was successfully formed. Unfortunately, further attempts to convert the ethyl ester into an NHS ester failed. A previous report in the literature described selective

displacements with 2,5-dioxopyrrolidin-1-yl 2-bromoacetate[1], however, I was unsuccessful in applying this to the coupling of N,N-dimethyl piperazine. The NMR spectra for this product were extremely complicated and I was unable to determine if the correct product was being formed. The complex NMR may have been due to the existence of product(s) displaying different conformations of the piperazine ring. To alleviate this possibility and to facilitate synthesis, 1,4-diazabicyclo[2.2.2]octane (DABCO) was used as a starting material. The selective displacement with DABCO produced only monoaddition of the 2,5-dioxopyrrolidin-1-yl 2-bromoacetate followed by hydrolysis of the NHS-ester to a carboxylic acid as observed by mass spectrometry (data not shown).

For the purposes of developing reaction conditions, the selective displacement between DABCO and ethyl 2-bromoacetate was attempted as shown Scheme 3.1. This reaction produced the desired di-addition product in good yields. Using the same conditions with the 2,5-dioxopyrrolidin-1-yl 2-bromoacetate, the di-addition product was observed, but the NHS product appeared to be undergoing hydrolysis to the dicarboxylic acid under these conditions. The introduction of two quaternary amines in close proximity to the NHS-esters appeared to greatly increase their reactivity. To avoid this problem, the distance between the NHS-ester and the quaternary amine was increased using a four carbon spacer arm which resulted in the desired di-addition product, DC4, in good yield (74%) (See scheme 3.2). Similarly, the selective displacement reactions were successful using 5- and 6-carbon spacer arms (DC5, and DC6 respectively).



Scheme 3.1. **Test reaction to work out the selective displacement reaction conditions.**



DC4: n = 3, DC5: n = 4, DC6: n = 5

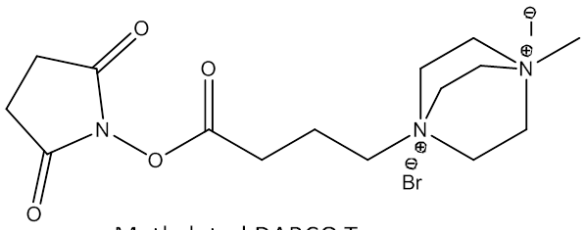
Scheme 3.2. **Successful synthetic route to preparing three crosslinking reagents, DC4, DC5, DC6.**

The synthesis of the crosslinker DC3, where n =2 in Scheme 3.2 was attempted in this work, but was unsuccessful. The intermediate for DC3, 2,5-dioxopyrrolidin-1-yl 3-bromopropanoate was successfully prepared, however, upon the addition of DABCO, an elimination reaction occurred, yielding 2,5-dioxopyrrolidin-1-yl acrylate. To circumvent this, the nucleophilic addition of DABCO with acrylic acid [2] was attempted and resulted in the desired dibetaine. All further attempts to obtain

the NHS ester by coupling with di-(N-succinimidyl) carbonate[3] or di-(N-succinimidyl) oxalate[4] failed. One possibility yet to be explored is to try converting the dibetaine to an acid chloride which could be reacted with NHS to give the NHS ester.

3.2 Tagging Reagents

Concurrently, tagging reagents were synthesized (Figure 3.1). These reagents were synthesized for multiple reasons. The tagging reagents were expected to be helpful in evaluating the fragmentation patterns with the crosslinker DC4 which was expected to have interesting charging and fragmentation characteristics due to the constrained quaternary amines. Because of the intrinsic positive charges, these reagents could increase the net charge on both peptides and proteins. Increasing the charge on peptides and proteins is known to improve fragmentation efficiency, particularly for ECD[5, 6]. The ionization efficiency could also be increased because the peptides would already be ionized due to the intrinsic positive charges. The Andrews laboratory has a history of studying membrane topologies,[7] and new water soluble tagging reagents would be helpful for that purpose. Both a singly-charged DABCO tag and a doubly-charged version were synthesized.



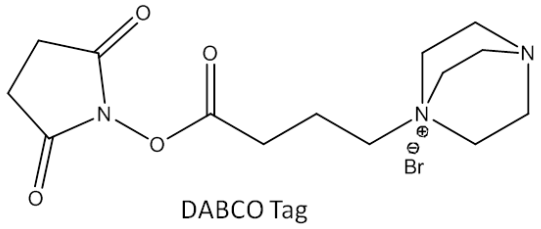


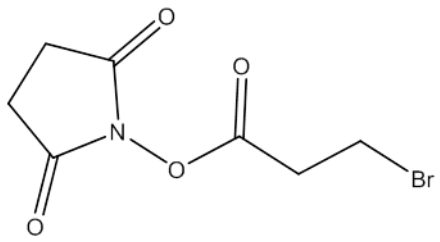
Figure 3.1. Quaternary tagging reagents synthesized in this work. The top tag had two quaternary amines and will be referred to as MeDABCO tag throughout this thesis. The bottom tag has one quaternary amine and will be referred to as DABCO tag.

3.3 Conclusions:

In summary, the chemistry demonstrated here is well suited to the development of crosslinkers and tags having related structures and may be useful in future studies. Three crosslinkers and two tagging reagents were synthesized with expected novel fragmentation and charge characteristics. Not unexpectedly, the proximity of the quaternary amines to the NHS ester was very important in determining the reactivity of the NHS ester. The impact of the quaternary amines prevented preparation of the shortest crosslinker, DC2, containing only one methylene group between the quaternary amine and the NHS ester. It was likely that the quaternary amines were inductively removing electron density from the carbonyl groups of the NHS esters, making them more susceptible to nucleophilic attack. For this reason, I was unable to isolate the crosslinker DC2. Increasing the length between the quaternary amines decreased the reactivity and the crosslinkers DC4, DC5, and DC6 were isolated.

3.4 Materials and Methods:

All reagents and solvents were commercially available and used without further purification. Melting points were determined in open capillary tubes on a Laboratory Devices Mel-Temp apparatus and were uncorrected. The NMR spectra were recorded on a Bruker instrument at 500 MHz for ^{1}H and 125 MHz for ^{13}C spectra. Chemical shift values were recorded in δ units (ppm). Mass spectra were recorded on a Micromass LCT Time-of-Flight mass spectrometer with an electrospray ionization source. All spectra were shown in Appendix A.

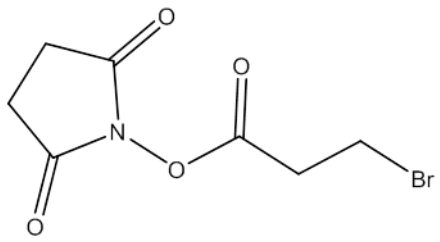


1

2,5-Dioxopyrrolidin-1-yl 3-bromopropanoate (1).

Similarly described by Elliot [8] **(1)** was prepared by combining 4.97g (43.1 mmol) N-hydroxysuccinimide (NHS), 8.90 g (43.1 mmol) dicyclohexylcarbodiimde (DCC,), and 6.60 g (43.1 mmol) 3-bromopropionic acid in 1.2 L ethyl acetate and stirred at room temperature overnight. The resulting dicyclohexylurea (DCU) precipitate was removed by filtration, washed with ethyl acetate and the filtrate stripped to an oil under reduced pressure. The oil sat at room temperature for 20 mins and additional DCU precipitated from the oil which was then filtered again. Excess solvent evaporated under reduced pressure to a solid. The solid was triturated

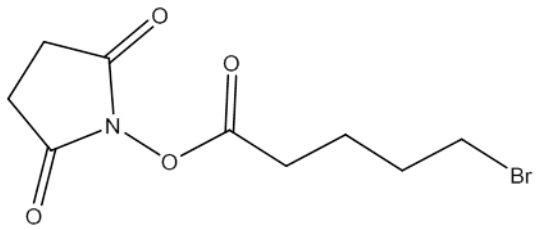
with ethanol, filtered, and dried to yield 5.712g (53%) of a white solid. ^{1}H NMR δ 3.6 (t, 2H), 3.2 (t, 2H), 2.8 (s, 4H). ^{13}C NMR 170.64, 167.24, 34.5, 26.69, 25.92.



2

2,5-Dioxopyrrolidin-1-yl 4-bromobutanoate (2).

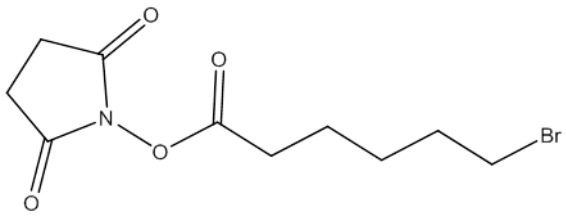
Compound **2** was prepared as previously described by Almirante et al. with minor modifications [9]. Briefly, a mixture of 4-bromobutyric acid (24 g, 144 mmol), NHS (19.8 g, 172mmol), DCC (35.5 g, 172 mmol), 4-(dimethylamino)pyridine (DMAP, 3.5 g, 28.6 mmol) and dichloromethane (300 ml) was stirred overnight at room temperature. Precipitated *N*,*N*-dicyclohexylurea was filtered off, washed with dichloromethane, and the filtrate was concentrated to an oil that was purified by flash silica gel chromatography eluting with 30:70 ethyl acetate:hexanes. Approximately 200 ml fractions were collected and each was concentrated and assayed by ^{1}H NMR. Fractions containing pure product were combined to give **3** (23.3 g, 61%) as a white solid: mp: 58-64 °C; ^{1}H NMR (d_6DMSO) δ 3.6 (t, 2H), 2.8 (m, 6H), 2.2 (m, 2H); ^{13}C NMR (d_6DMSO) δ 170.64, 168.64, 33.38, 29.47, 27.97, 25.92.



3

2,5-Dioxopyrrolidin-1-yl 5-bromopentanoate (3).

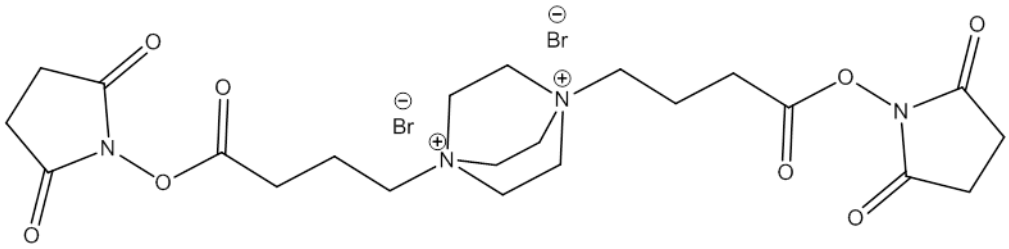
A mixture of 12.3g (65 mmol) of 5-bromovaleric acid, 8.97 g (78 mmol) NHS, 15 g (78 mmol) of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDAC), and 1.58 g (13 mmol) of DMAP was stirred overnight in 200 ml of dichloromethane at room temperature. The solution was washed with water (2x) and the combined washes then back-extracted with dichloromethane (2x). The combined organic layers were dried over magnesium sulfate and concentrated in vacuo to an oil that was diluted with 2-propanol to induce crystallization. The resultant precipitate was triturated with 2-propanol, collected by filtration and dried to yield 9.60 g (53%) of **3** as a white solid; mp: 81-84°C: ^{1}H NMR in d^{6}DMSO δ 3.5 (t, 2H), 2.8 (s, 4H), 2.7 (t, 2H), 1.9 (m, 2H), 1.7 (m, 2H). ^{13}C NMR 170.72, 169.29, 34.77, 31.54, 29.76, 25.91, 23.43.



4

2,5-Dioxopyrrolidin-1-yl 6-bromohexananoate (4).

The same method as described above to prepare (**3**) was used to prepare (**4**) but with 13.2 g (67 mmol) of 6-bromohexanoic acid, 9.3 g (81.3 mmol) of NHS, 15.6 g (81.3 mmol) of EDAC (81.3 mmol), and 2.02 g (6.4 mmol) of DMAP in 200 ml of dichloromethane as starting materials. Again, 2-propanol was used to induce crystallization and the resultant precipitate was triturated with 2-propanol, collected by filtration and dried to yield 13.65 g (70%) of (**3**) as a white solid; ^{1}H NMR (d^{6}DMSO) δ 3.5 (t, 2H), 2.8 (s, 4H), 2.7 (t, 2H), 1.8 (m, 2H), 1.6 (m, 2H), 1.4 (m, 2H). ^{13}C NMR 170.72, 169.38, 35.27, 32.14, 30.45, 27.13, 25.91, 23.84. M.P. 68-70°C.

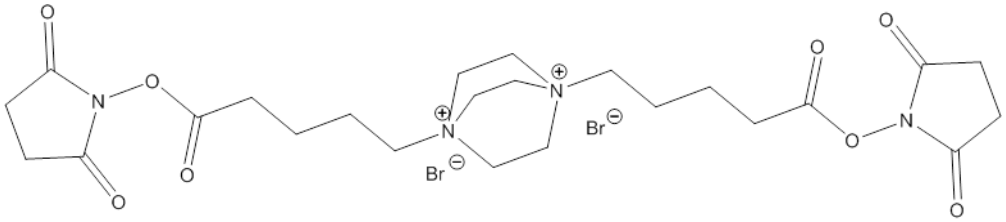


5

1,4-Bis(4-((2,5-dioxopyrrolidin-1-yl)oxy)-4-oxobutyl)-1,4-diazabicyclo[2.2.2]octane-1,4-diium bromide (5).

A mixture of NHS ester (**2**) (2 g, 7.5 mmol), 1,4-diazabicyclo[2.2.2]octane (DABCO; 403 mg, 3.6 mmol) , and DMSO (6 ml) was stirred at room temperature for one week. Acetonitrile

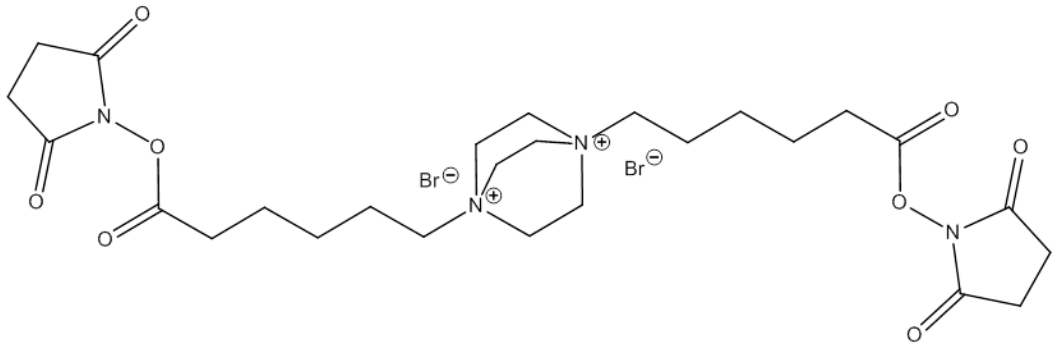
was added to precipitate the product which was collected by filtration, washed with acetonitrile, and dried to yield **5** (1.7 g, 74 %) as a white solid: mp 230 °C (dec); ^{1}H NMR (d_6DMSO) δ 3.9 (s, 6H), 3.6 (t, 4H), 2.8 (m, 12H), 2.1 (m, 4H); ^{13}C NMR (d_6DMSO) δ 170.64, 168.64, 62.4, 51.08, 27.42, 25.94, 17.55; MS (ES^+) m/z 559.2, 561.2 $(M^{2+}Br^-)^+$.



6

1,4-Bis(5-((2,5-dioxopyrrolidin-1-yl)oxy)-5-oxopentyl)-1,4-diazabicyclo[2.2.2]octane-1,4-diium bromide (6).

A mixture of 584 mg (2.1 mmol) of **(3)**, 112 mg (1 mmol) of DABCO, in 2 ml of DMSO was stirred for two weeks at room temperature. The compound was precipitated with 2-propanol to quantitatively yield a white solid: ^{1}H NMR δ 3.94 (t, 12H), 3.6 (t, 4H), 2.8 (s, 12H), 1.8 (m, 4H), 1.69 (m, 4H); ^{13}C NMR (d_6DMSO) δ: 170.7, 169.1, 63.1, 50.9, 29.9, 25.9, 21.5, 20.9. MS: (ES^+) m/z: 587.2, 589.2 $(M^{2+}Br^-)^{1+}$.

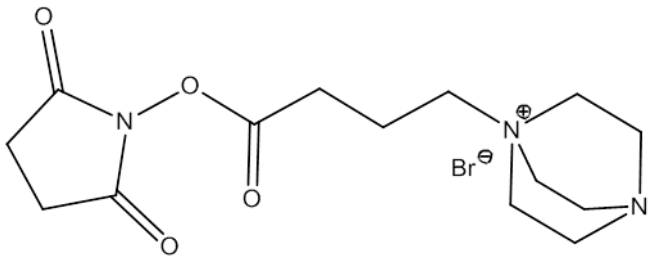


7

1,4-Bis(6-((2,5-dioxopyrrolidin-1-yl)oxy)-6-oxohexyl)-1,4-diazabicyclo[2.2.2]octane-1,4-diium bromide (7).

A mixture of 613 mg (2.1 mmol) of **(4)**, 112 mg (1 mmol) of 1,4-diazabicyclo[2.2.2] octane (DABCO), and 2 ml of DMSO was stirred for two weeks at room temperature. An equal

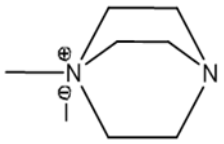
volume of isopropanol was added to induce crystallization and the mixture was allowed to sit for two days. The resulting white crystals were filtered and dried. The yield was quantitative. ^{1}H NMR δ 3.9 (s, 12H), 3.6 (t, 4H), 2.8 (m, 12H), 1.8 (m, 4H), 1.6 (m, 4H), 1.0 (t, 4H); ^{13}C NMR (d_6DMSO) δ: 170.7, 169.3, 63.4, 50.8, 30.3, 25.9, 25.0, 24.1, 21.3**.** MS: (ES^+): m/z 615.2, 617.2 ($M^{2+}Br^-$).



8

1-(4-((2,5-dioxopyrrolidin-1-yl)oxy)-4-oxobutyl)-1,4-diazabicyclo[2.2.2]octan-1-ium bromide (8).

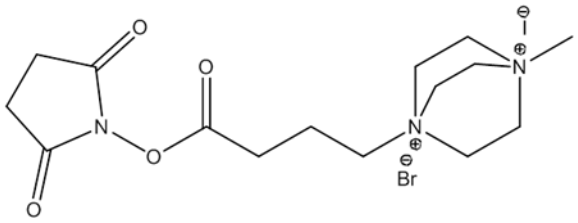
A solution of 3.0 g (11.3 mmol) (**2**) and 1.15 g (10.3 mmol) DABCO in 10 ml acetonitrile was stirred overnight at 78°C. The precipitate was then filtered, washed with acetonitrile, and dried to yield 2.83 g (73%) of (**8**). ^{1}H NMR δ 3.32 (m, 8H), 3.02 (t, 6H), 2.84 (m, 6H), 2.1 (m, 2H); MS: ES^+ m/z 296.0 (M^{1+}); ^{13}C NMR (d_6DMSO) δ 170.68, 168.75, 62.22, 52.14, 45.14, 27.68, 25.93, 17.24; M.P. 230°C (dec)



9

1-methyl-1,4-dizabicyclo[2.2.2]octan-1-ium iodide (9).

To methylate DABCO, the method of Kazock [10] was used. DABCO (4.48 g – 40 mmol) was dissolved in 50 ml ethyl acetate and stirred on ice. 2.5 ml (40.1 mmol) of methyl iodide was added drop wise over 20 mins. A white precipitate formed and the suspension was allowed to stir for ten mins. The white solid was filtered, washed with ethyl acetate and dried to yield 7.97 g (63%). ^{1}H NMR δ 3.25 (t, 6H), 3.02 (t, 6H), 2.97 (s, 3H) MS: 127.1 (M^{1+}); ^{13}C NMR (d_6DMSO) δ 53.6, 51.2, 45.3.



10

1-(4-((2,5-dioxopyrrolidin-1-yl)oxy)-4-oxobutyl)-4-methyl-1,4-diazabicyclo[2.2.2]octane-1,4-diium bromide iodide (10).

A solution of 2.62g (**9)** (10.3 mmol) with 3.00 g of (**2)** (11.3 mmol) in 10 ml acetonitrile was heated at 78°C overnight. The resulting white precipitate was filtered, washed with acetonitrile and dried to yield 2.5 g (47%). ^{1}H NMR δ3.91 (m, 12H) 3.57 (t, 2H), 3.28 (t, 3H), 2.89 (m, 6H), 2.1 (m, 2H) MS: m/z 390.1, 392.1 $(M+Br)^{1+}$; ^{13}C NMR (d_6DMSO) δ 170.6, 168.6, 62.5, 52.9, 51.8, 50.9, 27.4, 25.9, 17.5 ; M.P. 261°C (dec)

3.5 References

[1] D. Tong, P. Mu, Q. Dong, B. Zhao, W. Liu, J. Zhao, L. Li, T. Zhou, J. Wang, S. Guodong, *Colloids and Surfaces B* **2007**, *58*, 61.

[2] A. O. Kazantsev, S. A. Kazakov, K. V. Shirshin, S. M. Danov, V. L. Krasnov, *Chemistry of Heterocyclic Compounds* **1998**, *34*, 484.

[3] T. Kometani, T. Fitz, D. S. Watt, *Tetrahedron Letters* **1986**, *27*, 919.

[4] J. Guo, N. Xu, Z. Li, S. Zhang, J. Wu, D. H. Kim, M. S. Marma, Q. Meng, H. Cao, X. Li, S. Shi, L. Yu, S. Kalachikov, J. J. Russo, N. J. Turro, J. Ju, *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 9145.

[5] D. M. Horn, K. Breuker, A. J. Frank, F. W. McLafferty, *Journal of the American Chemical Society* **2001**, *123*, 9792.

[6] R. A. Zubarev, N. L. Kelleher, F. W. McLafferty, *Journal of the American Chemical Society* **1998**, *120*, 3265.

[7] X. Chen, P. J. Ulintz, E. S. Simon, J. A. williams, P. C. Andrews, *Molecular & Cellular Proteomics* **2008**, *7*, 2323.

[8] R. D. Elliott, R. W. Brockman, J. A. Montgomery, *Journal of Medicinal Chemistry* **1986**, *29*, 1052.

[9] N. Almirante, M. Ferrario, E. Ongini, in *PCT int. Appl.*, **2006**.

[10] J.-Y. Kazock, M. Taggougui, B. Carre, P. Willmann, D. Lemordant, *Synthesis* **2007**, *24*, 3776.

Chapter 4

Evaluation of DC4 as a Mass Spectrometry Cleavable Crosslinker

Chemical crosslinkers have a long history of useful applications in chemistry and biochemistry, particularly with application to biopolymer structures. The application of chemical crosslinkers to protein structure has increased potential when coupled with mass spectrometry but current commercial crosslinkers have a number of limitations when used with mass spectrometry. In the past decade there has been an increase in research involving chemical crosslinking combined with mass spectrometry and several crosslinkers have been developed to address one or more of the inherent difficulties associated with this technique[1-3]. The purpose of this thesis was to develop a crosslinker that addressed several of those challenges. Crosslinked peptides have been reported to fragment poorly via CID[4-6], consistent with my observations. For this reason and to overcome some other limitations of current crosslinking reagents, a mass spectrometry cleavable crosslinker was designed that upon initial CID would readily cleave within the crosslinking moiety, allowing the individual peptides released in MS^2 to be identified via MS^3. Additional desired properties included generation of distinctive marker ions for the various reaction products, a method to enrich crosslinked peptides, excellent solubility, ease of synthesis, good reactivity, self-quenching, and good stability in dry form. In Chapter 3 of this thesis, the crosslinker DC4 (Figure 4.1) was designed and synthesized. In this

chapter (Chapter 4), DC4 was assessed to determine which of the desired properties it possessed. This work has been published in *The Journal for the American Society for Mass Spectrometry*[7].

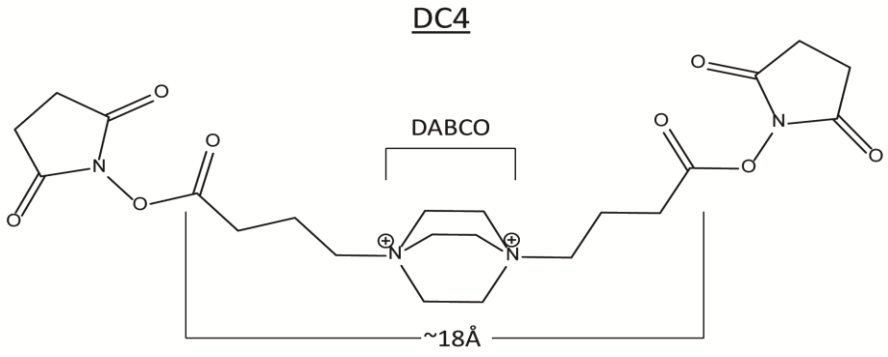


Figure 4.1. Structure of DC4. The crosslinker synthesized was named DC4 and spanned approximately 18 Angstroms.

4.1 MS Fragmentation of Model DC4-Crosslinked Peptides

To determine how DC4-crosslinked peptides fragmented in the mass spectrometer, model peptides were crosslinked and analyzed. N-terminally acetylated peptides QIGKGVAR and QHAKLGR were prepared in solution concentrated enough to force the ternary reaction of two peptides with DC4 through mass action. As expected, crosslinked peptides, dead-end reaction products, and in-source decay cleavage products from both the dead-end reaction and the crosslinked peptide products were observed in the initial mass spectrum on a MALDI-TOF-TOF-MS. No peptides containing unmodified lysyl residues were observed. Although there were two intrinsic positive charges in the crosslinking moiety, the mass of the crosslinked peptide represents a loss of a proton to yield the singly charged ion $(M-H)^+$ in MALDI-TOF-MS. No evidence of a doubly charged crosslinked peptide existed in the MALDI spectra. This may have been due to formation of an internal ion pair between the carboxyl group of the C-terminus (and Asp/Glu residues in other peptides) and the quaternary amine[8, 9]. CID of the crosslinked

peptide resulted in two fragments, 1050.6 and 938.5, both from cleavage at the sites of intrinsic positive charge (Figure 4.2A). The fragment at 1050.6 Da was due to cleavage at the intrinsic positive charge on the nitrogen containing ring (DABCO), retaining the DABCO moiety. The second fragment resulted from a rearrangement reaction which eliminated the DABCO ring. The same two cleavage products were visible in the original mass spectrum resulting from in-source decay. Pseudo MS^3 (MS^2 of the ISD product) of the elimination product (938.5) produced a spectrum with a robust series of b- and y- ions from which the component peptide was identified. (Figure 4.2B) Rearrangement of quaternary amines has been previously described by He and Reilly[10], a similar mechanism involving a sulfonium crosslinker was proposed by Lu et al,[11] and the proposed mechanism for DC4 was shown here in Figure 4.3. This mechanism was consistent with the data presented and importantly, it mobilizes an existing proton on the rearranged product when the original precursor was not protonated. This mobile proton can then facilitate backbone fragmentation yielding the b- and y- ion series.

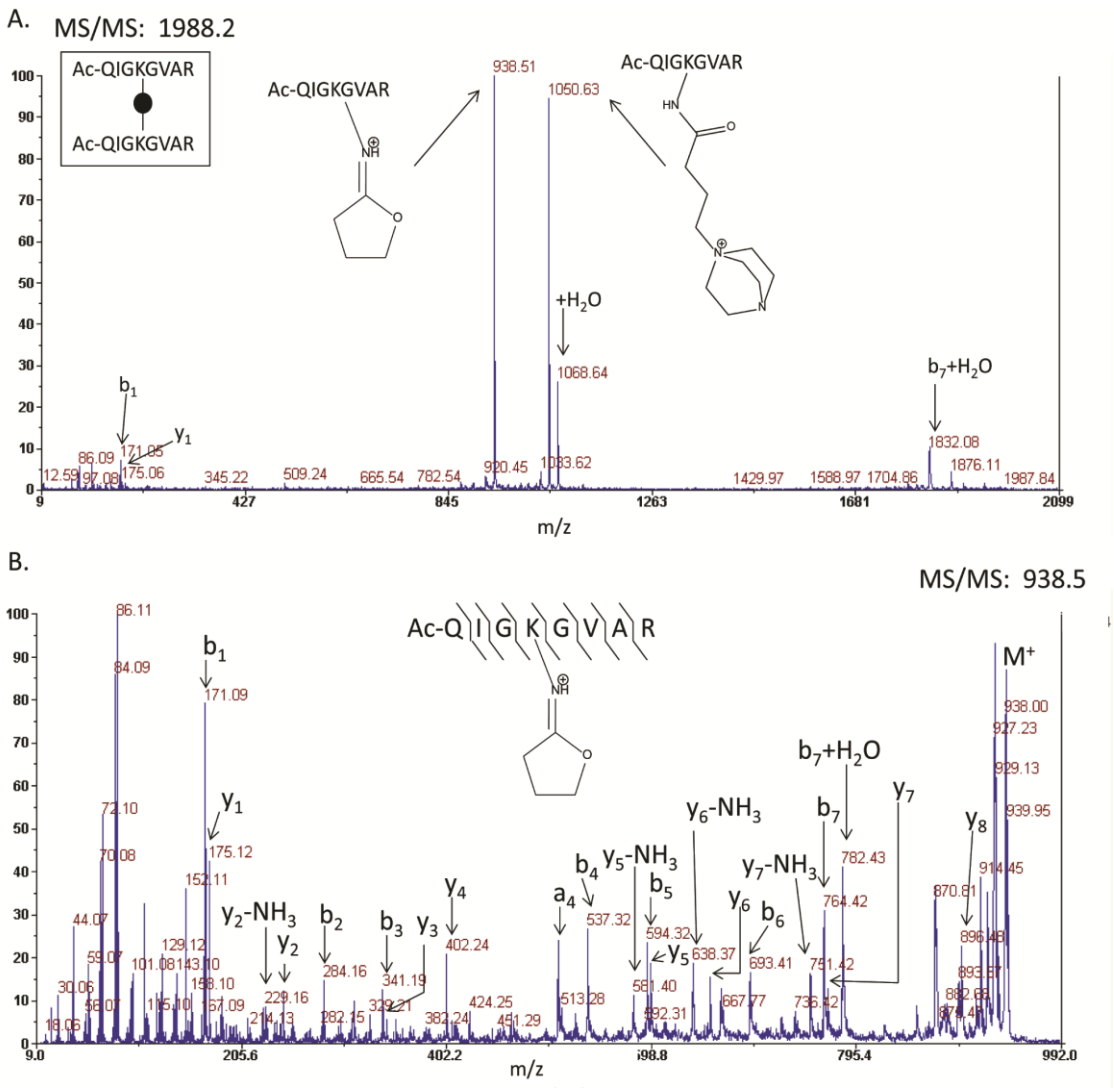


Figure 4.2. CID fragmentation of DC4-crosslinked model peptides on a MALDI-TOF-TOF (A) CID fragmentation on the MALDI-TOF-TOF of the crosslinked synthetic peptide m/z 1988.2 resulted in two major cleavage events at the intrinsic positive charges which were also observed in the first mass spectrum via in-source decay. (B) Pseudo MS^3 of the fragment without an intrinsic positive charge produced a robust series of b- and y-ions to sequence the peptide.

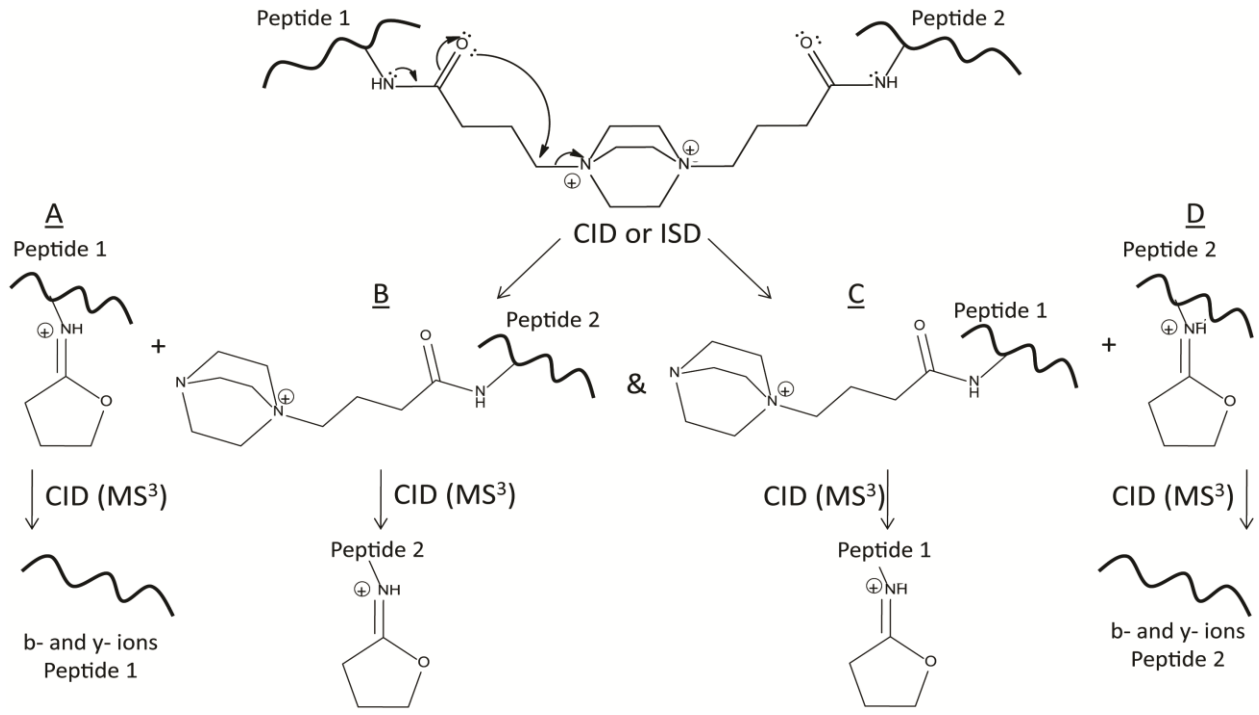


Figure 4.3. Proposed fragmentation mechanism for CID of DC4-crosslinked peptides. Although the crosslinker contained two intrinsic positive charges, peptides modified with DC4 were observed as singly charged ions in MALDI, most likely due to the loss of a proton from a carboxylic acid moiety (not shown). Upon fragmentation of the singly charged precursor by either ISD or CID, the rearrangement produced two singly charged species by mobilizing an existing proton as shown. The difference between A and C, B and D was 112 Da from the loss of the DABCO moiety and can be used as a screen to identify candidate crosslinked peptides.

Tandem mass spectrometry of the dead-end product resulted in a diagnostic ion clearly visible at m/z 199.1. (Figure 4.4) The presence of this ion unambiguously identified this spectrum as being from a modified peptide, although it has been observed occasionally in crosslinked peptides and valid crosslinked peptides may also contain dead-end reaction products. This cannot be used as an absolute exclusion criterion although it was useful. The other major peaks in the MS/MS spectrum resulted from fragmentation at the intrinsic positive charges. Low level b- and y- ions were also observed, however, the same fragments due to cleavage at the

intrinsic positive charges were present in the mass spectrum from in-source decay. CID of the in-source decay product which had lost the DABCO moiety yielded a robust sequence of b- and y- ions. The fragmentation of crosslinked peptides was also evaluated in the ESI-LTQ-Orbitrap-MS via CID and similar fragmentation events were observed as shown in Figure 4.5A and 4.5B.

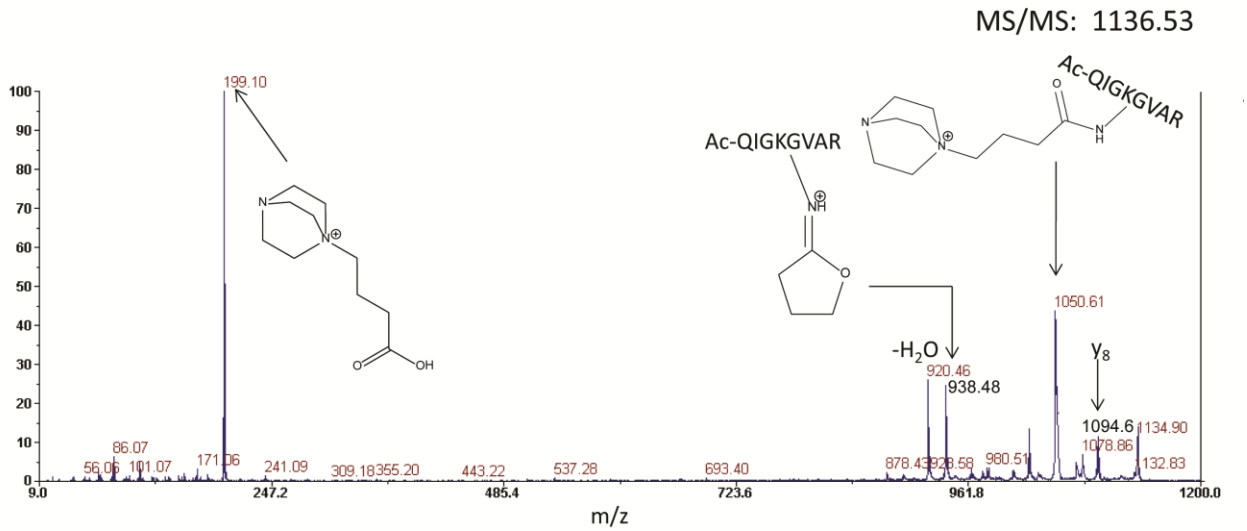


Figure 4.4. CID fragmentation on MALDI-TOF-TOF of a DC4-dead-end reaction product. Similar fragmentation is observed as with the crosslinked peptide but a diagnostic ion at 199.1 is observed, confirming this spectrum as being from a dead-end (type 0) product.

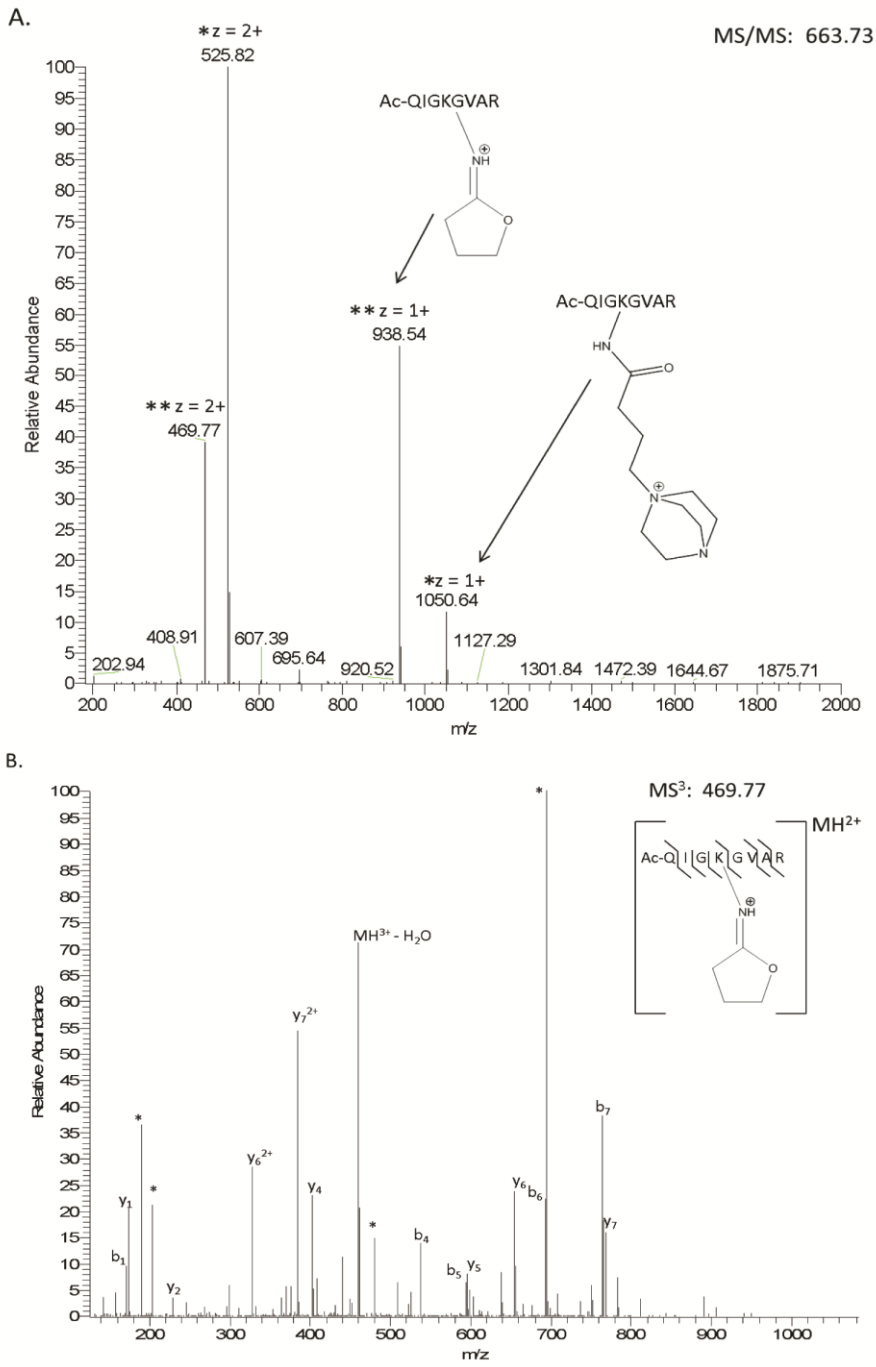
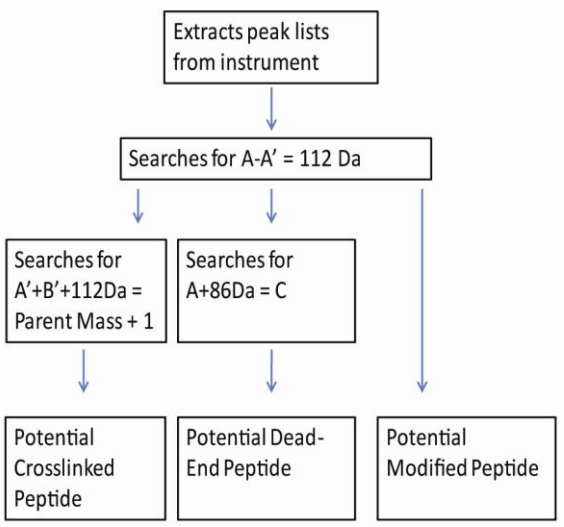


Figure 4.5 CID of a DC4-crosslinked model peptide on an LTQ-Orbitrap. (A) CID on the LTQ-Orbitrap of the crosslinked peptide (Ac-QIGKGVAR) results in the same two major cleavage events observed in MALDI at the intrinsic positive charges. (B) MS^3 of the fragment without an intrinsic positive charge produces a robust series of b- and y-ions for peptide identification.

4.2 Algorithm Development:

The ISD fragmentation of DC4-crosslinked peptides provided a unique series of ions that could be taken advantage of in developing a computational approach to identifying crosslinked peptides. An algorithm was developed to identify modified peptides using a simple filtering schema. (Scheme 4.1) The software application based on this algorithm initially searched MS^2 spectra (or ISD in MS spectra from MALDI) for differences of 112Da between two peaks. Output from this first filter may contain dead-end peptides, crosslinked peptides, or multiply modified peptides. Next, the dead-end peptides were culled out by searching for precursor ions that have an 86Da loss, followed by a 112Da loss for a total loss of 198Da from the precursors. These results were tagged as potential dead-end peptides. The crosslinked peptides were identified by first searching for the pairs of peaks separated by 112Da, then the software determined if the observed peptide pairs sum (while correcting for the DABCO mass of 112Da) to a precursor ion in the spectrum. Those ions that satisfy these criteria were returned to the user as potential crosslinked peptides and were then subjected to MS^3 and detailed de novo analysis.



Scheme 4.1. Filtering system to identify crosslinked peptides

4.3 Application of DC4 to Aldolase:

Aldolase crosslinking by DC4 was optimized in solution by reaction at a range of crosslinker to protein ratios and the products evaluated by SDS gel electrophoresis (Figure 4.6) with the objective of maximizing crosslinks within tetramers while minimizing crosslinks between tetramers (non-specific interactions). Aldolase tetramers were crosslinked at the optimum ratio of DC4 (minimizing non-specific interactions while maximizing crosslinking efficiency) for mass spectrometry analysis. CID of the ISD fragments from DC4-crosslinked peptides generated sufficient backbone fragmentation in the aldolase digest to identify the amino acid sequence of the crosslinked peptides as shown in Figures 4.7 and in Appendix B. The majority of the spectra from modified residues displayed preferential fragmentation, consisting predominantly of b-ions from the N-terminal side of the modified residues and y-ions from the C-terminal side. Not surprisingly, this fragmentation property was somewhat sequence dependent.

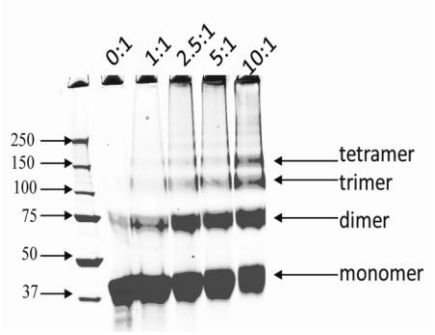


Figure 4.6. SDS-PAGE of aldolase crosslinked with increasing amounts of DC4. The ratio of DC4 to lysine residues was indicated across the top of the gel. Molecular weight standards were indicated in lane 1. Higher ratios of crosslinker to protein resulted in complete loss of the monomer and formation of multimers higher than the tetramer (data not shown).

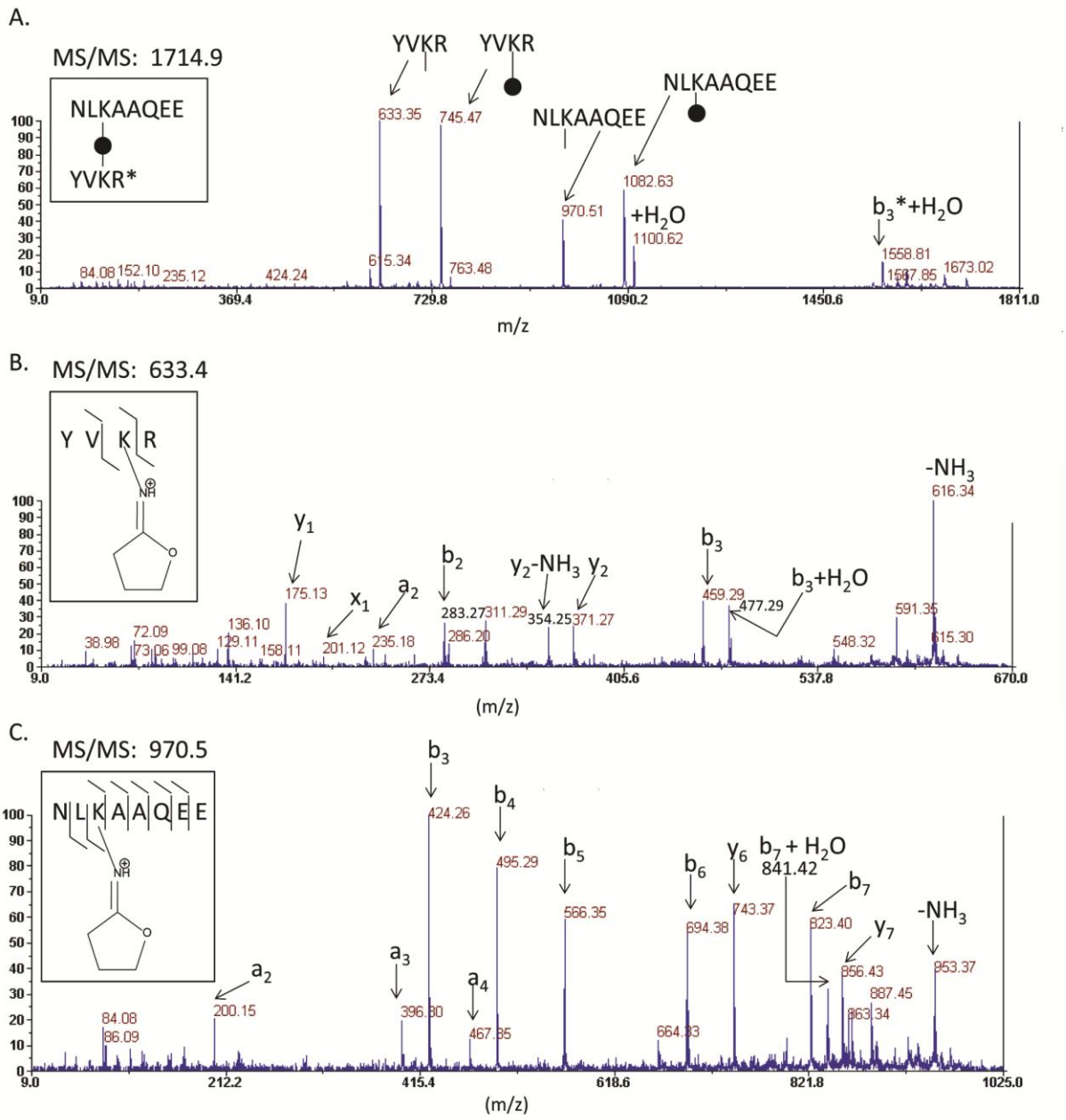


Figure 4.7. MALDI-TOF-TOF spectra of a DC4-crosslinked peptide from aldolase. (A) CID of a DC4-crosslinked peptide produced two sets of peaks where the lower mass was separated from the higher mass peak by 112Da. The black circle represented the central DABCO moiety and the lines on either side represented the carbon spacer arm in the crosslinker. Those same four peaks were observed in the mass spectrum from ISD. CID of the lower mass of each peptide pair (B, C) yielded the b- and y-ions needed to identify the peptide.

Both crosslinked peptides and dead-end peptides were identified from proteolytic digests of DC4-crosslinked aldolase (Tables 4.1 and 4.2, respectively). These modified peptides were evaluated for their consistency with the crystal structure for aldolase[12] and it was found that both the dead-end and crosslinked lysyl residues observed were solvent exposed (Figure 4.8). Two buried lysyl residues were not found to react with DC4, consistent with their solvent inaccessibility as discussed below. Six of the crosslinked peptides identified had inter-Lys distances between 7 and 23 Angstroms. Although the DABCO moiety is rigid, the reactive side arms are quite flexible and capable of coupling to Lys residues-spaced closely together, consistent with the shortest distance observed between two crosslinked Lys residues being 7 Angstroms. The side chain residue distance measurements were based on the rigid crystal structure and some changes in solution phase are likely since aldolase exhibits multiple crystal forms and portions of the molecule exhibit considerable flexibility between crystal structures and in solution.[13-16]

Table 4.1. DC4-crosslinked peptides identified from aldolase on a MALDI-TOF-TOF. Bold residues represent modified residues. The * represents the N-terminus of the protein.

Crosslink Type	Peptide A	Peptide B	Distance (Angstroms)	Precursor Mass	Fragment Masses
2 (Inter)	ILPDGDHDL***K***R (190-200)	Q***KK***E (11-14)	7	2221.2	651.4, 763.3, 875.4, 1278.8, 1346.8, 1458.8
2 (Inter)	***K***VLAAVYK (207-214)	Q***KK***E (11-14)	10	1920.1	959.6, 1071.7, 651.2, 763.4, 875.6
2 (Intra)	YV***K***R (327-330)	NL***K***AAQEE (319-326)	14	1714.9	633.4, 745.5, 970.5, 1082.6
2 (Inter)	ILPDGDHDL***K***R (190-200)	******PHSHPALTPE (1-10)	15	2611.3	1153.6, 1265.7, 1346.7, 1458.8
2 (Intra)	CVL***K***IGE (149-155)	STGSIA***K***R (35-42)	17	1827.9	887.5, 999.6, 941.5, 829.5
2 (Intra)	YV***K***R (327-330)	STGSIA***K***R (35-42)	23	1631.9	999.6, 887.5, 745.5, 633.4
2 (Intra)	STGSIA***K***R (35-42)	NL***K***AAQEE (319-326)	30	1969.0	999.6, 887.4, 1082.5, 970.5
2 (Intra)	STGSIA***K***R (35-42)	******PHSHPALTPE (1-10)	30	2152.1	887.5, 999.6, 1153.5, 1265.7
2 (Intra)	******PHSHPALTPE (1-10)	***K***DGADFAK (139-146)	32	2184.1	1153.5, 1265.6, 1031.5
2 (Inter)	YV***K***R (327-330)	Q***KK***E (11-14)	39	1507.9	875.5, 763.4, 745.5, 633.3

Table 4.2. DC4-dead-ends identified from aldolase on a MALDI-TOF-TOF. Bold residues represent modified residues. The * represents the N-terminus of the protein.

Peptide Sequence	Precursor Mass	Fragment Masses
VD***K***GVVPLAGTNGE (108-121)	1621.8	1535.8, 1423.8
IVAPG***K***GILAADE (22-34)	1519.8	1433.8, 1321.7
NL***K***AQEE (319-325)	1168.6	1082.6, 970.5
KDGADFAK (139-146)	1117.5	1031.5, 919.5
KVLAAVYK (207-214)	1157.6	1071.6, 959.5
CVL***K***IGE (149-155)	1027.5	941.5, 829.4
Q***KK***E (11-14)	961.5	875.5, 763.4, 651.3
STGIAS***K***R (35-42)	1085.6	999.5, 887.5
******PHSHPALTPE (1-10)	1351.6	1265.6, 1153.5
YV***K***R (327-330)	831.5	745.4, 633.4
ILPDGDHDL***K***R (191-201)	1544.8	1458.7, 1346.7

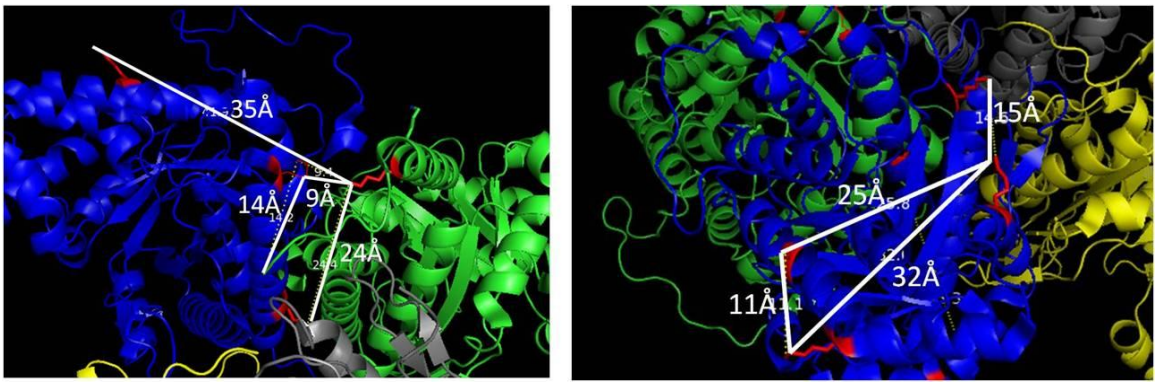


Figure 4.8. Selected aldolase crosslinks mapped onto the crystal structure. Each of subunit of aldolase was differentially colored (blue, grey, green, yellow) and crosslinks were shown in white with the distance next to each crosslink. In the left panel were the intersubunit crosslinks observed and the right panel has a selection of intrasubunit crosslinks.

The other four crosslinked peptides had inter-lysyl residue distances from 30 to 39 Angstroms. Although these distances were longer than we would initially expect based on the length of DC4 alone and a rigid protein structure, they are consistent with flexibility of protein backbones and lysyl side chains in solution. Here, several of the lysyl residues were observed to exhibit crosslinking with more than one other lysyl residue, consistent with earlier claims of flexibility in those regions. This flexibility allows the crosslinker to sample a larger region on the surface of the protein. As an example, Lys 41 was observed crosslinked to both Lys-152 which was 17 Angstroms away as well as Lys-321, 23 Angstroms away but also to Lys 329, 30 Angstroms away and to the N-terminus, also 30 Angstroms away. Deuterium exchange studies[13] in solution indicated very high exchange rates for the short helical region containing Lys-41 and high B-values were observed for this region in the crystal structure[14, 16]. Both of those studies provided further evidence for the flexibility of this region. This high level of flexibility was consistent with the observation for Lys-41 crosslinked to Lys-321 and to the N-terminus, both at 30 Angstroms. Likewise, the four N-terminal residues showed considerable variation between crystal structures indicating this region was flexible and consistent with the

identification of crosslinks between the N-terminus and Lys-199, Lys-139, and Lys-41 with distances of 15, 32, and 30 Angstroms from the N-terminus, respectively. The crosslink exhibiting the longest span between lysine residues was between Lys 13 and Lys 329. The 19 C-terminal residues including Lys 329 have often been found to be too disordered to be crystallized[17, 18] or when they were crystallized, they were found in different conformations depending on their substrate bound state[15, 19, 20]. It has been previously suggested that crosslinker length is not necessarily a precise indicator of inter-residue spacing when flexibility of the peptide backbone and lysyl side chains are factored in[21, 22] and our results were certainly consistent with this assertion.

All the peptides constituting type 2 crosslinked peptides were also observed as type 0 reaction products as shown in Table 4.2. Two peptides (VD***K***GVVPLAGTNGE, IVAPG***K***GILAADE) were only observed as the type 0 products. The closest lysyl residues for Lys-110, and Lys-27 are only 10 and 13 Angstroms, respectively, suggesting that the crosslinked products may not have been suitable for analysis (low ionization efficiency or large fragment size). Two lysyl residues Lys-214 and Lys-146 were observed only as unmodified forms suggesting that they were not as accessible to reagent as those which exhibited modification. Inspection of these residues in the crystal structure (Figure 4.9) revealed that they were much less accessible than other lysyl residues observed to be modified in this study. The two lysine residues were completely buried in the space-filling model, however they were visible in the ribbon model (Figure 4.9).

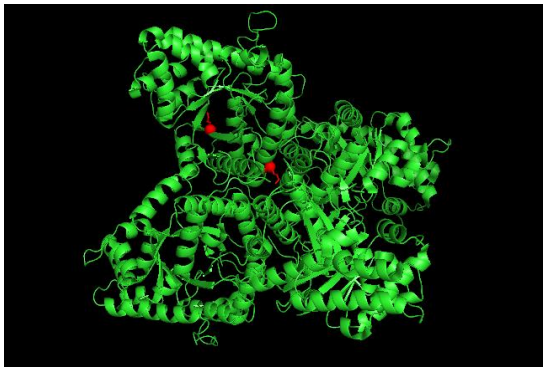


Figure 4.9. **Non-modified residues were buried within the crystal structure**. The two lysine residues that were not modified by DC4 are shown in red and are buried within the aldolase tetramer (green).

4.4 Generation of Antibodies to DC4-crosslinked peptides

As previously discussed in Chapter 1 (section 1.4.4), crosslinked peptides are low in abundance compared to the other peptides in a complex mixture. To increase the likelihood of those peptides being detected, an enrichment strategy is needed. Therefore one of the desired characteristics of a crosslinker was to have a method of enrichment for crosslinked peptides. Affinity purification would be the most specific method of enrichment for crosslinked peptides although to my knowledge, no crosslinkers exist with antibodies specific to crosslinked peptides. Crosslinkers exist containing affinity groups but those groups are often large and hydrophobic, making the crosslinker insoluble in water. Small molecules such as 2,4-dinitrophenol, iTRAQ, 1,3-butadiene, and even aspartic acid have been used as antigens to generate antibodies (www.abnova.com) therefore, it was thought the DABCO ring might be antigenic and an attempt to generate antibodies to DC4-crosslinked proteins was made.

Hemocyanin from keyhole limpet (KLH) is a large antigenic protein that is often conjugated to small molecules for antibody production. To make antibodies against DC4, KLH was crosslinked with DC4 and then sent to a company that specializes in antibody production, Cocalico Biologicals. A rabbit was inoculated with the crosslinked protein according to their

standard protocols. Several boosts were required to get the desired antigenic response which was then evaluated via western blot. The aldolase purchased from Sigma was isolated from rabbit, therefore aldolase was used to test the response of the antibody. Aldolase was reacted with DC4 and with the two tags that were synthesized in chapter 2. Those samples were analyzed via western blot and probed with the DC4 antibody (Figure 4.10). As expected, there was no antigenic response to the unmodified aldolase, but there was a very large response to the DC4 crosslinked aldolase. Discrete bands were distinguished for the monomeric and dimeric forms of aldolase. There was a smear of signal at higher molecular weights indicating the presence of trimer, tetramer, and higher order structures that reacted with the antibody, however, the individual bands were indistinguishable. No antigenic response was observed for either of the tagged aldolase samples. The DABCO tag contained one quaternary amine while the MeDABCO tag contained two quaternary amines. If the antibody were recognizing the DABCO ring with the two quaternary amines, it was expected that the antibody would also co-react with the MeDABCO tag. Because the antibody did not react with either tag, it was unlikely that the antibody reacts with the charged DABCO ring. The antibody could be reacting to two lysine residues crosslinked together or to the dead-end. Further characterization of this antibody is needed to determine its applicability; however this experiment clearly showed that it is possible to generate antibodies to DC4.

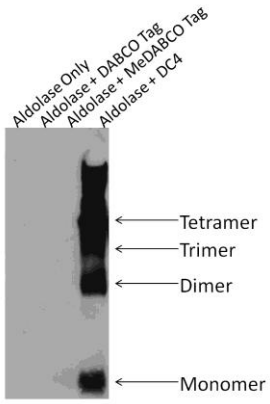


Figure 4.10. Western blot using antibodies generated against DC4. The antibodies did not react with aldolase or with aldolase reacted with either of the synthesized tags. However, there was a strong response to the DC4 crosslinked aldolase.

4.5 Conclusions:

The designed and tested a new crosslinking reagent that overcomes some of the difficulties associated with the use of crosslinking reagents with mass spectrometry. The DC4 crosslinking reagent is stable as a solid over long term storage, highly reactive, highly soluble (1 M solutions), and quite labile to CID. When fragmented by either CID or PSD, DC4-crosslinked peptides fragment characteristically into four rearranged products. Even though the precursor ions are not protonated, the fragmentation products rearrange to generate a mobile proton, and MS^3 results in efficient backbone fragmentation for peptide identification. Database searches can then identify which peptides were crosslinked and based on this characteristic fragmentation behavior, an algorithm was developed to identify which peptides were crosslinked. This crosslinking strategy was used on the model tetrameric protein, aldolase, and the crosslinked peptides identified were consistent with the structure.

4.6 Materials and Methods

Materials. See Section 2.6.

Crosslinking Reaction. In all experiments, 100 mM crosslinker was prepared immediately before use in 250 mM HEPES, pH 7.0. The acetylated standard peptides (10 μg of a 1 μg/μl solution) were reacted overnight with a 5 molar ratio of crosslinker. Fifty micrograms of a freshly prepared 5 μg/μl solution of aldolase was crosslinked using a 25 molar ratio of crosslinker to lysine residues for 30 minutes at room temperature.

SDS-PAGE. See Section 2.6.

Mass Spectrometry. See Section 2.6.

ESI-LTQ-Orbitrap-MS: See Section 2.6.

MALDI-TOF/TOF-MS: See Section 2.6.

Identification of Crosslinked Peptides with a Custom Algorithm and Database Searching Software: A custom database containing all standard proteins used in the Andrews lab was searched using the Mascot search engine[23]. The searches were completed using a peptide mass tolerance of 0.5Da, fragment mass tolerance of 0.6Da, and no enzyme. Variable modifications included the deamidation of glutamine and asparagine residues, oxidation of methionine, pyro-Glu, and the modification of lysine residues or the N-terminus that results from in source decay of the crosslinker (+68Da).

Algorithms written in Java were used to perform three filters of the observed masses to identify potentially modified peptides according to Scheme 1. Peak lists from the MS mode were extracted from the ABI 4800 TOF/TOF. A 0.5Da mass tolerance was allowed for the

searches. The software first searched for differences of 112Da between two peaks. In the second filter, candidate dead-end peptides were identified by searching for a set of three peaks: a precursor mass, a loss of 86Da from the precursor, and then a subsequent loss of 112 Da or a total 198Da loss from the precursor. In the third filter, potential crosslinked peptides were identified by searching the pairs of peaks separated by 112 Da for a set in which the lower of the two masses plus 112 Da added up to a precursor mass observed in the spectrum. Additional MS/MS spectra were obtained based on these results. The Java coding was completed by Bryan Smith.

DC4 Antibodies: Hemocyanin from keyhole limpet (KLH) was purchased from Sigma and prepared in water to 5ug/µl. DC4 was freshly prepared in 100mM HEPES, pH 7.5 at a concentration of 100mM. Equal volumes of KLH and DC4 were combined and allowed to react at room temperature for 30 minutes. The samples were then frozen and sent on ice to Cocalico Biologicals for antibody production using their standard protocols (www.cocalicobiologicasl.com). One rabbit was initially inoculated with 500µg of DC4 crosslinked KLH combined with Complete Freund's Adjuvant. Two and four weeks later, the rabbit was boosted with 250µg of DC4 crosslinked KLH combined with Incomplete Freund's Adjuvant. The rabbit was further boosted each month with 250µg of DC4 crosslinked KLH combined with Incomplete Freund's Adjuvant and a production bleed taken two and four weeks post-injection. To test the reactivity of the sera, aldolase was crosslinked or reacted with the protein tags synthesized in Chapter 3. For each sample, 10µg of aldolase was reacted with 10mM freshly prepared crosslinker or tag at a 50:1 molar ratio (crosslinker to aldolase monomer). Those samples were then analyzed via western blot as described in Chapter 2 except

the primary antibody was UM697 production bleed 1/1/2012 used at a 1:5,000 dilution. All other parameters were the same.

4.7 References

[1] J. W. Back, A. F. Hartog, H. L. Dekker, A. O. Muijsers, L. J. d. Koning, L. d. Jong, *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 222.

[2] E. Soderblom, B. Bobav, J. Cavannagh, M. Goshe, *Rapid Communications in Mass Spectrometry* **2007**, *21*, 3395.

[3] Q. Zhang, E. Crosland, D. Fabris, *Analytica Chimica Acta* **2008**, *627*, 117.

[4] E. V. Petrotchenko, V. K. Olkhovik, C. H. Borchers, *Molecular & Cellular Proteomics* **2005**, *4*, 1167.

[5] S. M. Chowdhury, X. Du, N. Tolic, S. Wu, R. J. Moore, M. U. Mayer, R. D. Smith, J. N. Adkins, *Analytical Chemistry* **2009**, *81*, 5524.

[6] E. V. Petrotchenko, K. Xiao, J. Cable, Y. Chen, N. V. Dokholyan, C. H. Borchers, *Molecular & Cellular Proteomics* **2009**, *8*, 273.

[7] B. Clifford-Nunn, H. D. H. Showalter, P. Andrews, *Journal of the American Society for Mass Spectrometry* **2012**, *23*, 201.

[8] E. Nordhoff, F. Kirpekar, P. Roepstorff, *Mass Spectrometry Reviews* **1996**, *15*, 67.

[9] M. A. Freitas, A. G. Marshall, *International Journal of Mass Spectrometry* **1998**, *182*, 221.

[10] Y. He, J. P. Reilly, *Angewandte Chemie* **2008**, *47*, 2463.

[11] Y. Lu, M. Tanasova, B. Borhan, G. E. Reid, *Analytical Chemistry* **2008**, *80*, 9279.

[12] A. Maurady, A. Zdanov, D. d. Moissac, D. Beaudry, J. Sygusch, *The Journal of Biological Chemistry* **2002**, *277*, 9474.

[13] A. Dalby, Z. Dauter, J. A. Littlechild, *Protein Science* **1999**, *8*, 291.

[14] Z. Zhang, C. B. Post, D. L. Smith, *Biochemistry* **1996**, *35*, 779.

[15] J. Wang, A. J. Morris, D. R. Tolan, L. Pagliaro, *The Journal of Biological Chemistry* **1996**, *271*, 6861.

[16] H. Pan, D. L. Smith, *Biochemistry* **2003**, *42*, 5713.

[17] J. Sygusch, D. Beaudry, M. Allaire, *Biochemistry* **1987**, *84*, 7846.

[18] H. Kim, U. Certa, H. Dobeli, P. Jakob, W. G. J. Hol, *Biochemistry* **1998**, *37*, 4388.

[19] N. Blom, J. Sygusch, *Nature Structural Biology* **1997**, *4*, 36.

[20] G. Hester, O. Brenner-Holzach, F. A. Rossi, D. M. Struck, K. H. Winterhalter, J. D. Smit, K. Piontek, *FEBS Letters* **1991**, *292*, 237.

[21] A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, R. Aebersold, *Molecular & Cellular Proteomics* **2012**, *9*, 1634.

[22] N. S. Green, E. Reisler, K. N. Houk, *Protein Science* **2001**, *10*, 1293.

[23] D. N. Perkins, D. J. C. Pappin, D. M. Creasy, J. S. Cottrell, *Electrophoresis* **1999**, *20*, 3551.

Chapter 5

Chemical Crosslinking Combined With Ion Mobility Mass Spectrometry

One of the major uses of both chemical crosslinking and IM-MS is to generate structural information about protein complexes. Most MS-based chemical crosslinking experiments analyze protein structures by cleaving the proteins into peptides and then analyzing the digest for crosslinked peptides[1-7]. The structures of the complexes can then be pieced together from the sequence of amino acids, the restrictions placed on the structure from crosslinked residues, and any other data from complementary methods[8]. This falls into the category of "bottom-up" methodologies. In IM-MS the entire intact complex is analyzed using more of a "top-down" methodology. Both the size of the entire complex from the collisional cross section (CCS) and the mass of the complex are obtained from a single experiment. Further dissociation experiments can be used to generate information on the components involved of the complex. When the ions are activated via collision-induced dissociation (CID), the weakest interactions break first which usually are the noncovalent interactions that hold the subunits together (hydrophobic, ionic, hydrogen bonding, Van der Waals)[9]. Adding all of the observed fragment ions will provide the mass of the complex and CCS measurements can provide an indication of the subunit sizes. The combined information from an IM-MS experiment and a chemical crosslinking experiment can yield the overall size and shape of the complex, the individual components, and which subunits interact.

5.1 Rationale for Combining Crosslinking and IM-MS

To the author's knowledge, at this time there is only one paper that has been published combining crosslinking and IM-MS[10]. In this paper, the authors analyzed the drift times of dead-end peptides that contained multiple lysyl residues. The drift time varied depending on which lysyl residue was modified and thus the two isomers could be distinguished. Here, crosslinking was completed at the protein level instead of the peptide level and the intact complex was analyzed by IM-MS.

Chemically crosslinking a complex prior to IM-MS has a number of potential benefits. First, chemical crosslinking forms covalent bonds between subunits which should stabilize the complex. This is particularly useful to prevent loss of the weaker interacting partners which could be disrupted during complex purification. The complex potentially may remain folded under harsher ionization conditions as well as during IM-MS analysis. Furthermore, dissociation upon collisional activation should require higher energy. When ions are collisionally activated, alternative unfolding patterns and dissociation events might be observed and when combined with similar data from noncrosslinked structures and modeling, this approach may yield new information about the folding of the complex.

One of the challenges with top-down proteomics is the limited dissociation of subunits that frequently occurs. For example, the L7/L12 stalk complex from the ribosome is selectively dissociated and several studies have been completed analyzing the structure of the stalk in different organisms[11-15]. The mechanism for dissociation of subunits for complexes is still being clarified[16-18], but the current hypothesis is based on the concept that protons are mobilized on the complex surface. Those charges are proposed to be distributed evenly over the surface area of the complex in order to minimize Coulombic repulsion. Occasionally, one subunit will

carry more charge than the others causing the subunit to have increased Coulombic repulsion[19]. When the ions are collisionally activated, the region of the protein with increased charge unfolds to alleviate the stress from the increased Coulombic repulsion. This newly unfolded region then has an increased surface area. The charges redistribute so that the charge distribution over the entire surface area of the complex is more uniform[20]. This results in increased charge on and around the unfolded section which causes further unfolding. This process of unfolding due to Coulombic repulsions continues, leading to dissociation of the highly charged monomeric subunit and the remaining oligomeric complex which is significantly charge reduced[19, 21, 22]. The charge state of the complex is crucial to its ability to dissociate and a minimum amount of charge on the surface area of the protein is required for dissociation of subunits[23]. Therefore, the asymmetric charge distribution after dissociation has detrimental effects on the further fragmentation of the oligomer. For example, in the case of the tetradecamer GroEL (MW = 800 kDa), after two activation events where two monomers were dissociated, the charge of the complex was too small to allow further dissociation[24].

Chemical crosslinking combined with IM-MS may be a solution to the problem of asymmetric charge partitioning. This would increase the utility of mass spectrometric analysis of intact complexes and provide a more systematic approach to top-down analysis of protein complexes. In the case of the ribosome, cleavage of the entire stalk complex often occurs[11, 12, 15, 25] but after chemically crosslinking, the stalk may be stabilized allowing other proteins to dissociate. If those proteins were crosslinked to another protein, a subcomplex could dissociate from the ribosome and the connectivity between new dissociated proteins would be established.

It was hypothesized that protein complexes reacted with DABCO-based crosslinkers or tags might have a more symmetric distribution of charge after fragmentation. The DABCO

crosslinkers and tags have fixed charges in the form of quaternary amines which should allow the labeled protein complex to have a higher charge state. There have been reports that when complexes with higher charge states fragment, the charge distribution is more evenly distributed than in lower charge states[26]. Also, because the DABCO crosslinkers contain fixed charges, the charge on the complex contributed from the quaternary amines should be immobile. These two points suggest that the crosslinked complexes with higher charge states should also have a less asymmetric distribution of charges after fragmentation.

Another challenge for top-down analysis, particularly with single proteins, occurs when the peptide backbone is being fragmented by CID to yield peptides for identification or characterization. It is often observed that peptides from either the N- or C- termini dissociate preferentially and certain regions of the protein rarely dissociate, which leads to low sequence coverage[27]. This is because one cleavage event would result in N- or C-terminal peptides, but internal ions require two cleavage events. Also, the N- and C-termini are generally flexible regions of the protein[28]. If backbone flexibility is indeed a key factor in determining fragmentation sites in intact proteins, then chemical crosslinking might constrain the flexibility of protein domains leading to reductions in the likelihood of backbone fragmentation in those regions. For example the N-terminus could be covalently attached to another section of the protein. This would then both constrain flexibility and require two cleavage events to yield similar fragments. Thus fragmentation of another backbone domain could become more likely. Crosslinking proteins prior to top-down analysis could alter the fragmentation pattern and the peptides observed.

This ability to lock together specific sites on a protein's surface can help retain the solution-phase protein structure in the gas phase. Some proteins have been shown to adapt

conformations in the gas phase that are different from solution phase and crosslinking could prove useful for those proteins[29]. In the absence of bulk solvent, proteins adopt more compact conformations as evidenced by the generally observed decrease in the collisional cross section[30]. One hypothesis for this decrease in CCS is that in solution, the hydrophilic amino acid side chains interact with the water and ions present, resulting in a relatively extended structure. When bulk solvent is removed, those side chains collapse onto the surface of the protein, resulting in a smaller CCS. Chemically crosslinked residues may be held in a rigid conformation and be unable to collapse against the surface of the protein resulting in a larger CCS. This might result in a more native-like state in the gas phase. If these hypotheses are correct and the crosslinkers perform as designed, they could help overcome several of the major difficulties in IM-MS and top-down analyses of protein complexes.

5.2 Analysis of DC4-Crosslinked Proteins

A key feature of the crosslinker DC4 that contributes to many of its unique properties is that it contains two intrinsic positive charges. One of the first questions to answer was whether those positive charges affected the charge state of proteins in the gas phase. DC4 is an NHS-ester and reacts mostly with principally with the primary amines on lysyl residues[1, 31]. In solution at pH 7-8 (physiological pH), lysyl residues are positively charged. When two lysyl residues react with DC4 to form a crosslink, the lysyl residues each lose one positive charge, but the two quaternary amines present in the crosslink retain the same. When one lysyl residue reacts with DC4 to form a dead-end product, the lysyl residue loses its charge, but the presence of the two quaternary amines in DC4 result in a net gain of one positive charge. The proteins were sprayed at pH 7, so the carboxylic acid moiety that resulted from DC4 hydrolysis should

have been deprotonated and thus negatively charged. This would result in nearly complete retention of the net charge at that site however incomplete ionization of the carboxylate could result in conditions which a in net positive charge was expected. If the proteins were ionized under non native acidic conditions, then a large increase in positive charge would be expected.

To determine the effect of the crosslinker on the protein charge distribution, three proteins were crosslinked with DC4 and the charge state distributions of crosslinked and noncrosslinked protein were compared (Figure 5.1) using ESI-IM-MS. The first protein, lysozyme, a monomer[32] exhibited a small but significant increase in the charge state distribution in the crosslinked form with a shift of approximately two charges. The major charge state of lysozyme was observed as +5, with only a low intensity +6 observed. The crosslinked lysozyme exhibited the most intense peaks at the +7 and +6 charge states. The lysozyme charge states may also not have shifted to a much higher charge state because lysozyme only has six lysyl residues[33]. Some of the lysyl residues may have crosslinked (no net increase in charge), and a few could have reacted to form dead-ends which would result in a small charge shift.

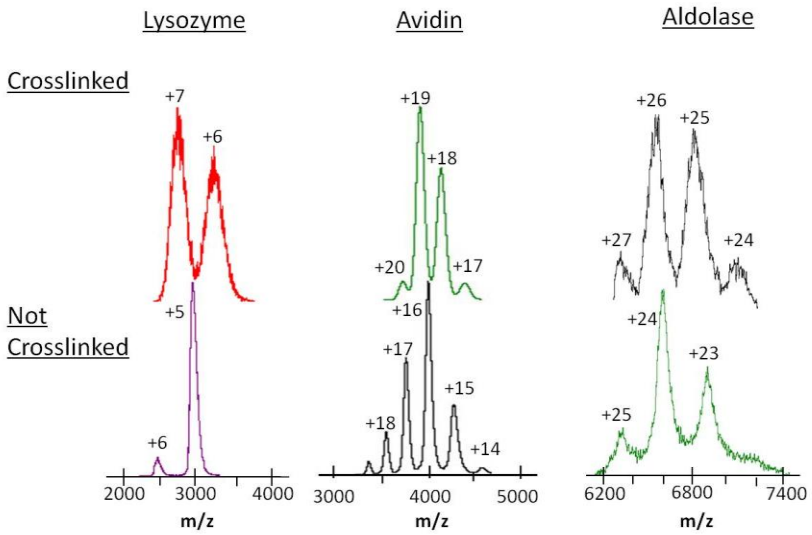


Figure 5.1. Charge state comparisons of three crosslinked proteins. Both crosslinked (top) and not crosslinked (bottom) lysozyme, avidin, and aldolase were analyzed by ESI-IM-MS. The charge states are adjacent to the peak they correspond to.

In order to determine if crosslinking a larger protein complex might result in a larger shift in the charge state distribution avidin (MW 64kD) and aldolase (MW 158kD), both homotetramers[34, 35] were crosslinked. Crosslinked avidin yielded a similar shift in the charge state distribution as observed for lysozyme. The charge states for native avidin were centered around the +16 ion, but the crosslinked avidin spectrum exhibited a charge distribution centered around the +19 state. The avidin monomer[36-38] contains nine lysyl residues and the tetrameric avidin has 40 potential sites for reaction with DC4 (36 lysyl residues and four N-termini). Only a few lysyl residues would react to form crosslinks within each avidin tetramer but it was expected that the majority of lysyl residues that were accessible would be form dead-ends due to the high DC4 concentration used.

Aldolase, a larger tetramer, contains a higher percentage of lysyl residues than avidin and the bottom-up analysis of crosslinked aldolase was presented earlier in this thesis (Section 4.3). Aldolase was chosen as a model protein for the initial characterization of DC4 because it had a high percentage of lysyl residues that were surface exposed[34]. Of the 27 lysyl residues[39] per monomer, ten were observed to react with DC4 and two were shown not to react. The remaining lysyl residues were not observed at all, most likely due to their location in the peptide sequence as previously discussed. Both the C-terminus and the N-terminus of aldolase are known to be flexible regions[40-42], therefore a large number of crosslinks were possible. Because of the increased potential of aldolase to form crosslinks and the larger size of the protein, aldolase may have an increased charge shift.

The three main charge states observed from native aldolase were +23, +24, and +25, but in the spectrum of the crosslinked aldolase, charge states ranged from +24 to +27. Similar to the

lysozyme and avidin results, this also represented a net charge increase of two. Because a large shift in charge distribution wasn't observed for any of the proteins analyzed, either the addition in charge was being balanced by a negative charge, either from negatively charged counter ions or from the formation of internal ion pairs. As indicated above, each dead-end could form an internal ion pair because water reacts with the second end of the crosslinker producing a carboxylic acid. In the gas phase, proteins exist with both positively-charged and negatively-charged residues which adds up to the overall charge on the protein[43].

The molecular weights (Table 5.1) of both crosslinked and noncrosslinked lysozyme were calculated and the observed mass for noncrosslinked lysozyme was consistent with calculated values[33]. Reaction with DC4 will lead to increased protein mass. Crosslinking two lysyl residues leads to a 250 Da mass gain and each individual lysine forming a dead-end product results in a net gain of 268 Da. Reaction of all seven primary amines in lysozyme (six lysyl residues and one N-terminus) with DC4 as dead-end products would lead to a maximum mass gain of 1,876 Da. The formation of crosslinks would only reduce the main gain from this maximum value. The molecular weight of the crosslinked lysozyme was 30% higher than the maximum value, which suggested that other residues reacted with DC4 or that there were a large amount of salt adducts present. NHS esters have been reported to have low but measureable reactivities towards the hydroxyl groups on the side chains of serine, tyrosine, and threonine[31, 44]. These reactivities are much less than primary amines and infrequently observed, therefore it was unlikely that the addition in mass was due to additional reactions with DC4. More likely, the addition in mass was due to salt adducts.

Table 5.1. Observed molecular weights (MW) of native and crosslinked proteins.

	Lysozyme	Avidin	Aldolase
Native MW (Da)	14,700	63,900	158,700
Crosslinked MW (Da)	19,200	74,300	170,700
Total Number of Primary Amines	7	40	112
MW Added if All Primary Amines were Dead-Ends (Da)	1,876	10,720	30,016
MW Added if All Primary Amines were Crosslinked (Da)	750	5,000	14,000
Crosslinked and Native MW Difference (Da)	4,500	10,400	12,000

Crosslinked aldolase and avidin were analyzed by SDS-PAGE (data not shown) and were tetrameric under the denaturing conditions used. SDS-PAGE provides an approximate size, but it is a low resolution technique. The bands representing crosslinked tetramers increased in size by approximately 1-2 kDa as expected if several crosslinks and dead-ends were formed. These results were in contrast to the masses observed by MS (Table 5.1). The mass difference between crosslinked and native avidin observed by MS was much higher than anticipated. Not all of the lysyl residues could have formed dead-ends and avidin must have contained a sufficient number of crosslinks to maintain its tetrameric state as observed via SDS-PAGE. Unlike with the lysozyme and avidin data, the mass of crosslinked aldolase was reasonable.

In addition to the mass shift, an increase in the CCS of each protein was also observed. Lysozyme increased in CCS by 33% (4.5 nm^2) and avidin increased by 12% (4.3 nm^2). There are several possible explanations for this observation. In solution, the side chains from hydrophilic amino acids are usually extended away from the surface of the protein to interact with water and ions in solution and when crosslinked, those side chains would have been crosslinked in their extended state. It has been hypothesized that in the ESI process, as a protein is desolvated and ionized, the side chains flatten against the surface of the protein, resulting in smaller CCS than the in solution structure[30]. After crosslinking, the crosslinked side chains

could have maintained a rigid, more native-like conformation instead of flattening against the surface of the protein, which would result in an increased CCS.

An alternative hypothesis was that the crosslinked proteins increased in size because they were unfolding. Although not always the case, unfolded proteins can have a broad range of drift time distributions and the drift times shift to larger values. With the three crosslinked proteins, the distributions of drift times were narrow but they shifted to a larger time. When proteins are unfolded, the charge states increase[45] because of the increased surface area accessible for protonation. The distribution of charge states also widens due to multiple conformations of unfolded structures[46-49]. The crosslinked proteins did exhibit an increased charge state, but the distribution of charge states remained similar (Figure 5.1). A third hypothesis was that the increased size was due to large amounts of salt adducts. After crosslinking, many quaternary amines and free carboxylates were added onto the surface of the protein. In solution, a large amount of hydrolyzed DC4 was available to electrostatically interact with the protein surface, creating a shell of salt adults around the protein, thus increasing the CCS.

Either of the hypotheses discussed could account for the increased size, increased mass, and small increase in charge state. Further experiments were needed to clarify the processes occurring, however, the increased CCS or apparent surface area could be an explanation for the observed increase in charge states. Proteins of comparable size have similar charge state distributions[50]. When the CCS to charge state ratio was calculated for each of the proteins analyzed, the values were very similar (Table 5.2) regardless of whether the protein was crosslinked. Although ionizable amino acid side chains can be buried within the protein structure[51], most exist on the protein surface[52] resulting in charge also localized on the protein surface[53]. For a given surface area, a limited number of charges can be placed on the protein

such that charge state can be used to predict the protein surface area[46]. Therefore, when the crosslinked proteins increased in CCS, the observation that the charge also increased was expected.

Table 5.2. Charge state, CCS, and primary amine count for three proteins. The CCS of aldolase was provided by Yueyang Zhang. Calibration was not completed on the day crosslinked aldolase was analyzed, however the drift time was longer than for the control aldolase under the same conditions. N.D. stands for no data.

Protein	Lysozyme	Lysozyme + DC4	Avidin	Avidin + DC4	Aldolase	Aldolase + DC4
Average CCS (nm^2)	14.2	17.9	37	41.5	80.4	N.D.
Most Intense Charge State	5	7	16	19	24	26
CCS : Charge State	2.8 : 1	2.5 : 1	2.3 : 1	2.2 : 1	3.4 : 1	N.D.

5.3 Analysis of Tagged Avidin

In the previous IM-MS experiments, three interesting observations were made for the crosslinked proteins. First, the charge state distribution did not increase dramatically, second, the mass of crosslinked proteins increased by more than anticipated, and third, the CCSs increased. One potential hypothesis for the lack of a large increase in positive charge was that the dead-end products form internal ion pairs. If this were true, adding quaternary amines to the protein surface without adding a functional group that could form internal ion pairs would result in a higher charge state distribution. In Chapter 3 of this thesis, two tags were synthesized (Figure 3.1) that contained quaternary amines on the DABCO moiety. These were the DABCO tag which has one quaternary amine and the MeDABCO tag which has two quaternary amines. Avidin was individually reacted with each of the tags and then analyzed by IM-MS (Figure 5.2). The charge state distributions of the DC4-crosslinked avidin were similar to both of the tagged

avidin but for each site that the MeDABCO tag reacted with, a net increase of one charge should have been observed. The similar charge state distributions observed between crosslinked avidin and the two tagged forms of avidin suggested that formation of ion pairs internal to the dead-end products could not explain the charge distribution.

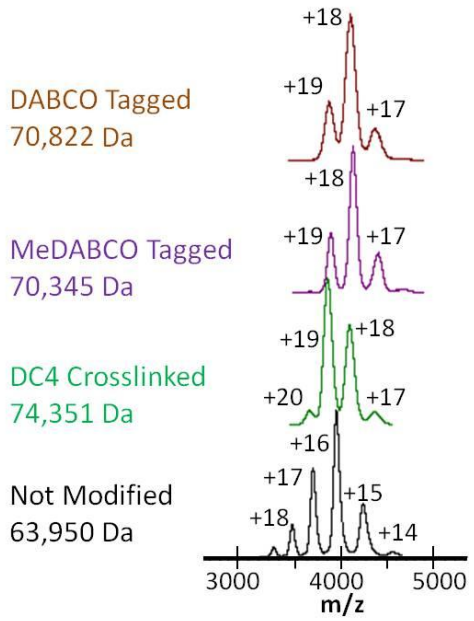


Figure 5.2. Crosslinked and tagged avidin mass spectra. The charge state distributions and molecular weights of avidin were compared using DC4 and two protein tags having one (DABCO tagged) and two (MeDABCO tagged) quaternary amines.

Based on the mass of the crosslinked avidin, it was clear that salts were electrostatically interacting with the DC4 on the surface of the protein. However, if those salts were solely responsible for the increased CCS and charge state, the tagged avidin would exhibit similar charge state distributions, mass shifts, and CCSs. Anions would bind to the surface to reduce the surface charge density added by the tags thus increasing the mass and the CCS. The charge state distributions were similar as previously discussed, but the mass shifts of the tagged avidin were much less than the DC4-crosslinked avidin (Table 5.3). To account for the tagged avidin masses, 35 of the primary amines in the avidin structure would have had to react with the tag but at least

eight of the avidin lysyl residues are buried within the crystal structure[35]. If all of the surface accessible primary amines were modified, salt adducts would still be required to increase the mass and to lower the surface charge density of the complex. The addition in mass was similar between the two tags because tags only differed in mass by 15 Da and the anions balancing the quaternary amines should have been the same. However, the mass difference between the DC4-crosslinked avidin and the tagged avidin was larger than could be explain by the difference in mass between the tags and the crosslinkers. This may have indicated that more salts were electrostatically interacting with the protein surface of DC4-crosslinked avidin than the tagged avidin. The avidin may not have reacted stoichiometrically with the tags and therefore fewer salt adducts were attached to the protein. This would result both in a smaller mass and a smaller CCS. This explanation seems unlikely given that the same conditions were used to tag proteins as were used to crosslink them and the modification of accessible lysyl residues would have been similar. However, avidin treated with DC4 would have both crosslinks and dead-ends. The free carboxylate on the dead-ends could interact with quaternary amines from hydrolyzed DC4 in solution, thus building a shell around the crosslinked avidin. The tagged avidin would not have the added free carboxylates, so there would not be as many salt adducts on the protein. If this hypothesis were true, then the CCS of the tagged avidin would be smaller than the DC4-crosslinked avidin.

Table 5.3. Observed masses and CCSs of tagged and DC4-crosslinked avidin.

	Unmodified	DABCO Tag	MeDABCO Tag	DC4
Observed Mass (Da)	63,914	70,345	70, 822	77,639
Mass if All Primary Amines Reacted (Da)	N/A	71,194	71,794	74,634
Mass if 32 Primary Amines Reacted (Da)	N/A	69,738	70,218	72,490 – Dead-Ends 67,914 – Crosslinks
CCS (nm^2)	34.7 ± .8[a]	36.4 ± .6	38.0 ± .3	38.8± 1.2

[a.] Note that the CCS of avidin on this day was lower than the literature values, however the trend in CCS from the DABCO tag, the MeDABCO tag, and DC4 would have been the same.

The masses of the modified avidin suggested that there were fewer salt adducts and possibly fewer modifications on the tagged avidin than the DC4-crosslinked avidin. If there were fewer salt adducts on the tagged avidin, then the CCS should have been smaller than the DC4-crosslinked avidin. The CCS of the DC4-crosslinked avidin was slightly bigger, but similar to the MeDABCO tagged avidin which was slightly bigger than the DABCO tagged avidin. All three modified avidin samples were bigger than the unmodified avidin (Table 5.3).

At this point, there were several explanations for the data and few concrete conclusions could be made. The masses of the avidin samples showed definitively that there were salt adducts on DC4-crosslinked avidin and that there were fewer salts attached to either of the tagged avidin samples. The charge distributions of the modified avidin shifted equally to slightly higher charge states than modified avidin and the CCSs of the modified avidin complex increased with the size of the tag and crosslinker.

One simple explanation is that the addition of either the tag or the crosslinker causes the protein to increase in CCS. This was unreasonable because the crosslinkers and tags would not

have been protruding away from the surface of the protein. When avidin was desolvated and ionized, the crosslinkers and tags would have collapsed against the surface of the protein[30]. The DABCO moiety is very rigid and would likely have been responsible for any increase in CCS. Both the tags and DC4 each contain the DABCO moiety, so if the CCS increase was due to the increase in size of the tag or crosslinker, the CCS of the tagged and crosslinked avidin should have increased by the same amount. Because that did not occur, the size of the crosslinker and tag could not have been solely responsible for the increased CCS.

The quaternary amines placed on the surface of the protein could be causing avidin to unfold to relieve Columbic repulsions. This would cause the CCS of the tagged avidin to increase from the unmodified avidin, but the crosslinked avidin would be unable to unfold. Instead, the CCS increase of the crosslinked avidin could be due to the crosslinks causing the protein structure to maintain more of rigid structure and a native-like state. Before any further conclusions can be drawn, additional data should be collected. Crosslinking avidin using a crosslinker without quaternary amines would be beneficial and thorough desalting steps need to be done on the DC4 crosslinked and tagged avidin to accurate masses and CCSs.

5.4 Analysis of Crosslinked Avidin Using Multiple Crosslinkers.

In Chapter 3, four crosslinkers were synthesized in this work ranging in length from approximately 13 to 24 Angstroms (Figure 5.3). Crosslinkers with increased length can crosslink lysyl residues over a greater distance. This is beneficial because the more crosslinks that are observed, the more constraints can be put on protein structure. Also, additional complexes are accessible to analysis by crosslinking, particularly large multi-subunit complexes where the subunits are farther apart. In an ideal crosslinking experiment, every residue that

could be crosslinked would be crosslinked, however in reality a mixture of crosslinks, dead-ends, and non modified residues will be observed. Not every lysyl residue will be solvent exposed and accessible to the crosslinker, so buried lysyl residues will remain unmodified. Even surface exposed lysyl residues may not be within crosslinking distance of another lysyl residue, resulting in a dead-end. Residues that are within crosslinking distance may still form a dead-end as opposed to a crosslink because of the sheer excess of water molecules available to react with the crosslinker.

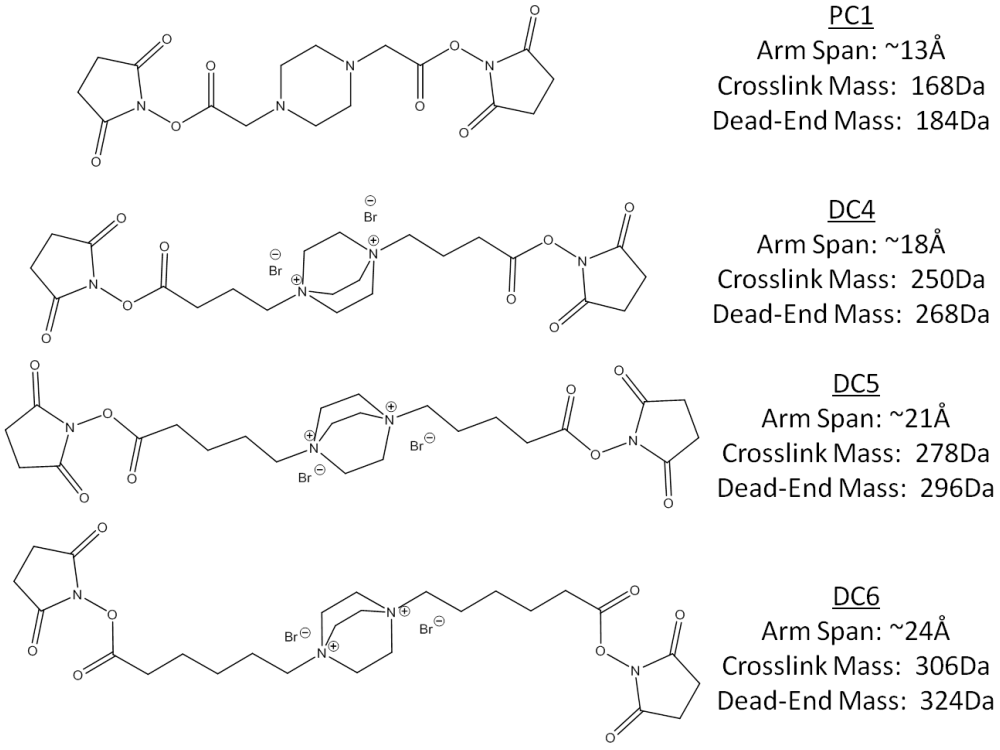


Figure 5.3. Crosslinkers synthesized and utilized in this thesis. The arm span of each crosslinker, and the masses added when the crosslinker reacts with a primary amine to form crosslinks or dead-ends are shown adjacent to the crosslinker structure.

There were several reasons to use varying crosslinker lengths with IM-MS. First, increasing the length of the crosslinker has the potential to yield more crosslinks which should

result in additional stabilization of the complex. Second, if the increased mass from the DC4-crosslinked avidin was due to crosslinker adducts, crosslinking avidin with crosslinkers with higher molecular weight should increase the molecular weight. Similarly, if the increased CCS compared to the tags was due to the increased size of the crosslinker to the tag, DC5- and DC6-crosslinked avidin should have even larger CCSs. Prior to IM-MS, crosslinked avidin was analyzed by SDS-PAGE to verify that the majority was crosslinked (Figure 5.4). When crosslinking was carried out using ratios of 50:1 crosslinker to avidin monomer the majority of avidin existed as a tetramer or higher. For each avidin to appear as a tetramer on a denaturing gel, at least three type two crosslinks per tetramer must have been present. The SDS-PAGE analysis indicated that at least this minimal amount of crosslinking was occurring. Via SDS-PAGE, the sizes of the tetramers crosslinked with the various reagents were in the same region, regardless of which crosslinker was used. The resolution of SDS-PAGE is much less than mass spectrometry, but the monomer bands are shifted up slightly in size by a few kilo Daltons which as expected.

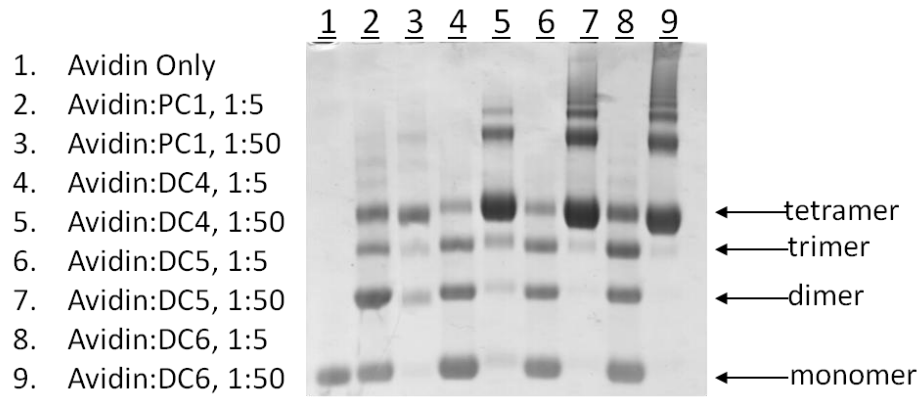


Figure 5.4. SDS-PAGE of the crosslinked avidin analyzed by IM-MS. The crosslinker used and the molar ratio of avidin monomer to crosslinker is shown on the left. Under denaturing conditions, native avidin dissociated into a monomer. However, upon crosslinking with the various crosslinkers, the band representing the tetramer became the major species. Unless otherwise mentioned, the 1:50 ratio of protein to crosslinker was used for IM-MS experiments and the tetramer isolated.

When the crosslinked avidin was analyzed by SDS-PAGE, it was apparent that as the length of the crosslinker increased, so did the amount of crosslinking. This was consistent with the increased length of the crosslinkers being able to access a larger number of lysyl residues. Using a ratio of 50:1 the shortest crosslinker, PC1, crosslinked the majority of avidin to a tetramer. Some monomer, dimer, and trimer remained and there were small amounts of higher order structures (nonbiologically relevant). Under the same conditions, DC4-crosslinked avidin had small amounts of monomer, dimer, and trimer remaining, but had much more bands representing the tetramer and higher order structures. Avidin crosslinked with DC5 contained a small amount of trimer and all other forms were tetrameric or higher structures. DC6-crosslinked avidin followed the same trend. This was not surprising because the longer crosslinkers can span a greater distance and potentially crosslink more residues. This should be a caution on the use of extremely long crosslinkers. As the crosslinker can span longer distances to reach residues in a complex, it could also span longer distances to crosslink nonspecific proteins nearby that were not interacting with the complex.

Avidin crosslinked with this series of crosslinkers was then analyzed by IM-MS. A comparison of the charge distribution in the crosslinked avidin is shown in Figure 5.5. Prior to mass spectrometric analysis, extensive desalting steps were completed, thus the noncrosslinked avidin was slightly unfolded and has a higher charge state distribution than observed for the fully native form. Typically +18 charge states are not observed with the folded avidin. In the PC1 crosslinked avidin, the charge states were not very well resolved, but the charge states were similar to what is typically observed with unmodified avidin. However, PC1 does not have quaternary amines and the stoichiometry of the crosslinking reaction may have lower which

would result in the lower charge state. The spectra for the DC4, DC5, and DC6 crosslinked avidin appeared very similar to each other and exhibited a shift in charge state from a base peak of +16 in the unmodified avidin to a base peak of +19 with each of the DABCO crosslinkers. This was consistent with the previous lysozyme, avidin, and aldolase results. DC4, DC5, and DC6 are very similar in structure and only differ in the arm length.

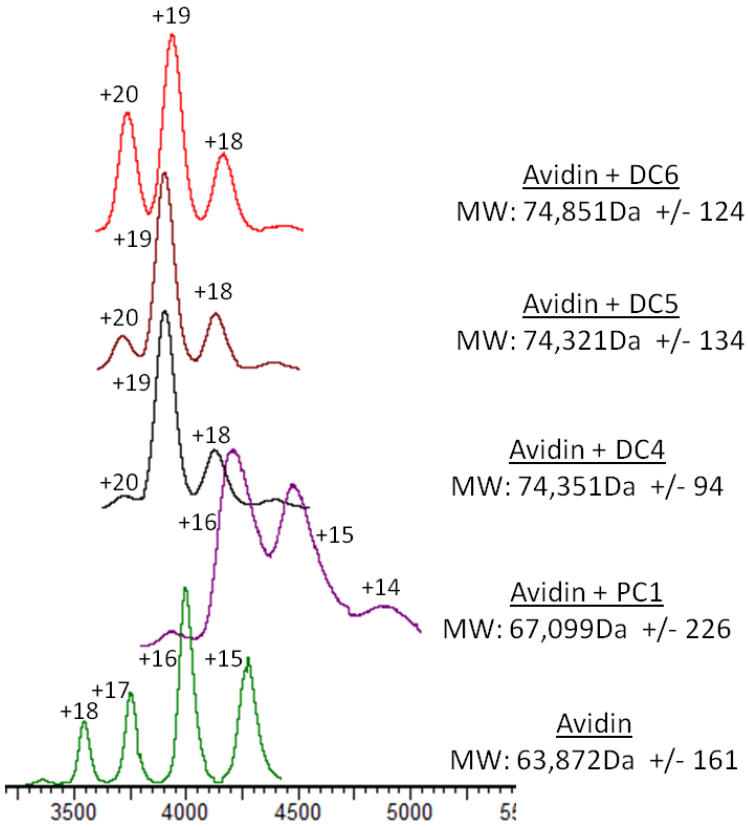


Figure 5.5. **The charge state distributions of crosslinked avidin.** Unmodified avidin does not usually have a +18 charge state, but due to the stringent clean-up methods, this avidin sample became slightly unfolded with a higher charge state.

The increase in mass that was observed when avidin was crosslinked with DC4 was also observed when avidin was crosslinked with the other crosslinkers (Table 5.4). Even though the mass difference between one molecule of DC4 and DC6 was 56 Da, the mass of crosslinked avidin was similar regardless of which DABCO crosslinker was used. These observed masses

were still higher than the expected masses if all lysyl residues reacted and higher than the masses observed by SDS-PAGE. This result was similar to the tagged avidin result because the relative masses of the tags did not appear to change the magnitude of the mass increase on the complex either. However, the mass added when avidin was crosslinked with PC1 was different or low stoichiometry. That increase in mass was within the range of the expected values for an efficient crosslinking experiment. The main difference between PC1 and the DABCO crosslinkers is that each of the DABCO crosslinkers contains two quaternary amines. This is consistent with the increased mass of DC4-, DC5- or DC6-crosslinked avidin being due to the quaternary amines participating in strong electrostatic interactions with ions from solution.

Table 5.4. Masses of avidin complexes using multiple crosslinkers

Crosslinker	Observed Mass (Da)	Mass if All Primary Amines were Crosslinked (Da)	Mass if All Primary Amines were Dead-Ends (Da)
None	63,872	N/A	N/A
PC1	67,099	67,192	71,232
DC4	74,351	68,872	74,592
DC5	74,321	64,432	75,712
DC6	74,851	69,992	76,832

CCSs for noncrosslinked avidin (Table 5.5) were in the 36-37 nm^2 range which were within 3% of the literature values[50]. The addition of the smallest crosslinker, PC1, slightly increased the CCS to approximately 38-40 nm^2. The DABCO crosslinkers further increased the CCS to 41-43 nm^2, but there was not a significant difference between DC4-crosslinked avidin and DC6-crosslinked avidin. As previously discussed, PC1 does not contain quaternary amines and is less likely to form salt adducts as the DABCO crosslinkers. The majority of the increase

in size from 36nm^2 to 38nm^2 of the PC1 crosslinked avidin was likely due to the crosslinking. This size increase suggested that either the protein was unfolded or that the crosslinked lysyl residues were holding the protein in a rigid, but more native conformation. The crosslinker may have placed constraints on the structure by reacting two lysyl residues together so that during desolvation and ionization, the residues may not have been able to collapse against the surface of the protein. If PC1 crosslinked proteins were held in a more rigid structure, then the increase in size of the DABCO crosslinkers was likely due to a combination of the protein being held in a more rigid conformation and the addition of salt adducts.

Table 5.5. Comparison of CCSs of avidin crosslinked with different crosslinkers.

Crosslinker	None	PC1	DC4	DC5	DC6
CCS (nm^2)	37.1 ± 0.4	39.3 ± 0.6	41.5 ± 0.4	42.0 ± 0.5	42.3 ± 0.2

One explanation for the smaller CCS observed for PC1 crosslinked avidin is that it did not have as many salt adducts due to the absence of quaternary amines. Another explanation was that PC1 reacted less stoichiometrically than the other crosslinkers, potentially due to PC1 being the smallest crosslinker and most hydrophobic. Avidin was crosslinked with two different ratios of DC4 and first analyzed by SDS-PAGE (Figure 5.4). The lower ratio of DC4 yielded less crosslinked tetramer, but the higher ratio led to almost complete conversion to the crosslinked tetramer. Via SDS-PAGE small amounts of monomer, dimer, and trimer were observed in the crosslinked avidin which indicated that there were populations of differentially crosslinked avidin. ESI-IM-MS was performed under gentle conditions to maintain the noncovalent interactions, so avidin was observed as tetramer regardless of whether how extensively it was crosslinked. Therefore the mass of the tetramer observed in IM-MS was a composite of a

population of the crosslinked tetramers and noncrosslinked tetramers. At a high ratio of DC4, very few noncrosslinked tetramers were present due to the high stoichiometry of the reaction, but at lower ratios, the stoichiometry of the crosslinking was more incomplete and the tetramers would contain a larger mixture of crosslinked and noncrosslinked subunits. Drift times and charge states are both distributions so on average the charge state and size of the sample crosslinked with a lower level of crosslinker would be smaller and that was what was observed (Table 5.6).

Table 5.6. CCSs of avidin crosslinked with increasing amounts of DC4.

DC4 to Avidin Ratio	None	5:1	50:1
CCS (nm^2)	37.1 ± 0.4	39.8 ± 0.4	41.5 ± 0.3

Under denaturing conditions, the majority of DC4-, DC5-, and DC6-treated avidin were crosslinked in a tetrameric state or higher (Figure 5.4). This implied that the tetramers analyzed in IM-MS were all crosslinked because there was very little avidin not crosslinked into a tetramer or higher. The extent of crosslinking within the tetramer could still vary but each tetramer had the minimum required number of crosslinks to maintain a tetramer under denaturing conditions. The PC1 crosslinked avidin was composed of more noncrosslinked tetramer, thus under IM-MS conditions, the average charge and the average size should have been smaller, as observed (Table 5.4, Figure 5.5). Further clarification of these effects could be obtained by repeating the analysis with only crosslinked tetramer and by analyzing the CCS of crosslinked proteins with increasing amounts of PC1 to avoid the high amounts of salt adducts with on the DABCO crosslinked proteins.

5.5 Collision Induced Unfolding and Dissociation of Crosslinked Avidin

To determine whether crosslinking stabilized the gas phase structures of proteins, the collision induced unfolding (CIU) and collision induced dissociation (CID) was measured by IM-MS. One charge state of tetrameric avidin was isolated and the collision voltage in the trap prior to the ion mobility cell was incrementally increased. The collision voltage is proportional to the energy applied to the ions and as a result, the increased collisional heating causes the ions to first unfold and then to fragment. From the collision voltages, the energy required to unfold or dissociate ions can be calculated. The unmodified avidin tetramer remained folded and compact until a collision voltage of 25V was reached and then the tetramer started to unfold (Figure 5.6). At a collision voltage of 40V, little to no folded tetramer remained. The unfolding was further detailed by monitoring the drift times as a function of collision voltage (Figure 5.7). When proteins unfold, they can unfold into families of structures which have discrete drift times, or they can unfold in a more gradual process and families of structures cannot be distinguished. Unmodified avidin starts as a folded tetramer which unfolds into three, possibly four discrete intermediates before dissociation visible as discrete spots (Figure 5.7). Although the CCS of PC1-crosslinked avidin was larger than unmodified avidin, it remained compact until a collision voltage of 35V was reached. Then the unfolding proceeded similarly to the uncrosslinked avidin with discrete structures. This is illustrated in Figure 5.6 where the increased collision voltage required to unfold PC1-crosslinked avidin and the similar energy ranges are apparent. A shift to higher collision voltages for CID is also observed in Figure 5.7 where the drift time represents a measure of subunit dissociation. The range of collision voltages required for dissociation was larger for the PC1-crosslinked avidin tetramer than for the unmodified avidin reflecting the heterogeneity and incomplete crosslinking described above. The presence of both crosslinked

and uncrosslinked subunits in the sample led to this property because the crosslinked tetramer required more energy to unfold and dissociate, but the noncrosslinked tetramers in the sample did not. A higher level of PC1 crosslinking that contained a higher proportion of tetramer should require a higher trap collision voltage to unfold.

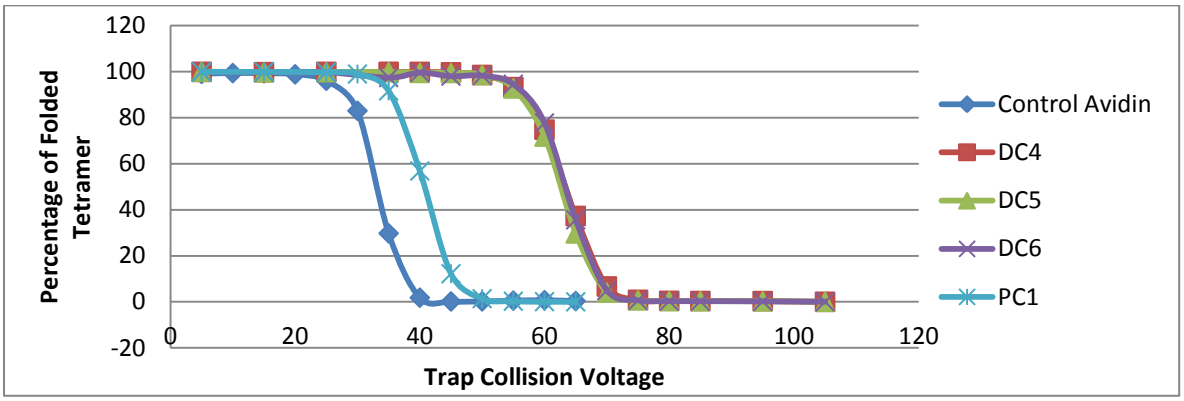


Figure 5.6. A**nalysis of the crosslinked folded avidin tetramer.** Avidin was crosslinked with the four crosslinkers previously synthesized. The +18 charge state of the crosslinked tetramers and the unmodified avidin were isolated and the trap collision voltage was incrementally increased to until the original folded tetramer was unfolded. Then the percentage of folded avidin tetramer was plotted as a function of trap collision voltage.

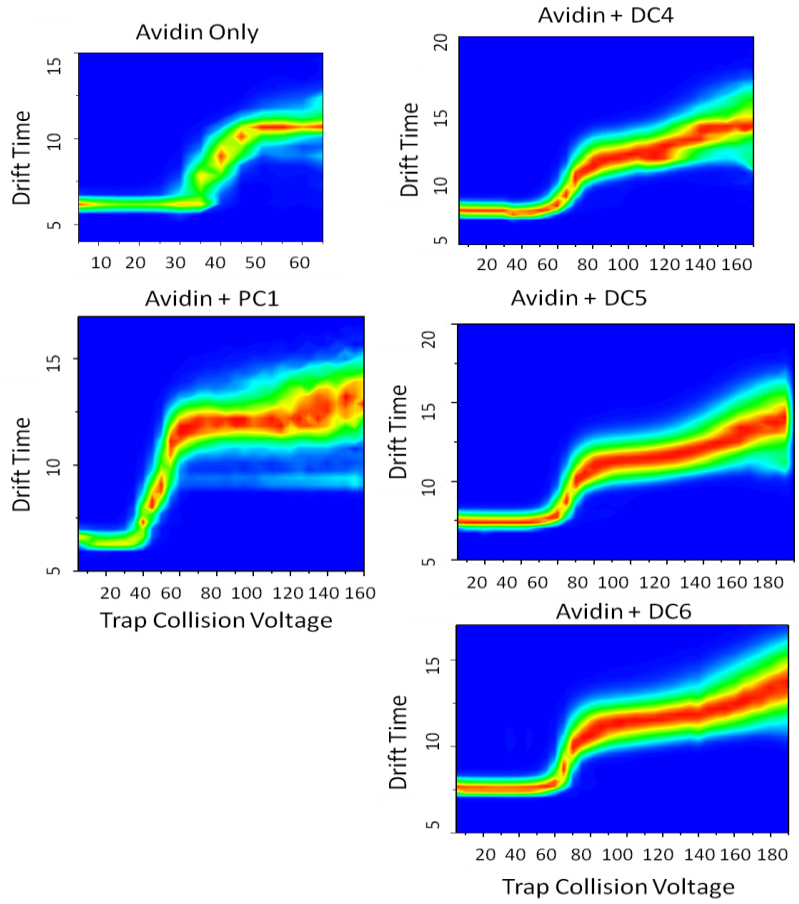


Figure 5.7. Contour plots detailing the collision induced unfolding of crosslinked avidin. These contour plots show the intensity of ions (z axis) as a function of drift times (y axis) and trap collision voltage (x axis). Discrete regions (spots) observed in the top left panel (avidin only), indicated that the presence of distinct unfolding intermediates. The continuous unfolding patterns observed in the right panels result from more heterogeneous components whose unfolding patterns cannot be resolved.

The unfolding profiles of tetrameric DC4-, DC5- or DC6- crosslinked avidin were very similar to each other, but they required more energy to unfold than both the unmodified and PC1 crosslinked avidin (Figure 5.6). The difference between these crosslinked tetramers and the PC1-crosslinked tetramers can be attributed to incomplete crosslinking as described below. Unlike the unmodified avidin and the PC1 crosslinked avidin, the DC4, DC5, and DC6 CIU profiles only had one discrete family of intermediate structures. The drift times were very broad for these species, indicating a broad distribution of unfolded structures. These observations were

consistent with the crosslinks imposing structural constrains on the tetramer which would result in different unfolded structures than observed for the unmodified avidin. The extent of crosslinking will vary from one individual protein molecule to another and this microheterogeneity leads to a heterogeneous distribution of unfolding structures resulting in a broader CIU profile.

A similar CIU analysis was made for avidin crosslinked with two ratios of DC4 described above. The extent of avidin crosslinking was directly proportional to the collision energy required for denaturation. In other words, the highest ratio of DC4 to required the highest collision voltage to unfold the tetramer (Figure 5.8). Avidin crosslinked with a lower ratio of DC4 required more energy to unfold than the unmodified avidin, but less than the highly crosslinked avidin. This was consistent with the lower number off avidin crosslinks expected for the lower ratio of DC4 and supported by SDS-PAGE (Figure 5.4). This result was similar to the PC1 crosslinked avidin which also exhibited incomplete crosslinking and the avidin subunits.

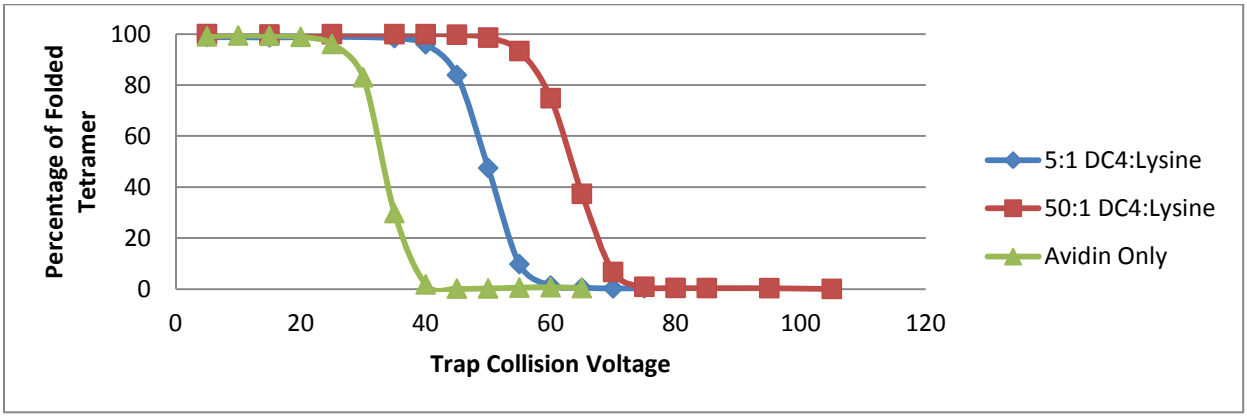


Figure 5.8. A comparison of the trap collision voltage required to unfold differentially crosslinked avidin tetramers. After increasing the amount of DC4 added to avidin, a higher trap collision voltage was needed to unfold the tetramers.

Collisional heating causes protein complexes to unfold and subsequently dissociate into subunits and this dissociation can follow multiple pathways. One possible pathway involves the tetramer dissociating into a monomer and a trimer which further dissociates into the dimer, followed by dissociation into the monomer (or the trimer could directly dissociate into three monomers). Another possible dissociation scheme is dissociation of the tetramer into two dimers which then dissociate into monomers. Lastly, the tetramer could dissociate directly into monomers. To determine the dissociation pathway for avidin, the identities of the dissociated subunits and the percentages of those subunits were used to map a CID profile of unmodified avidin (Figure 5.9). At low collision voltage, the tetramer was stable, but as trap collision voltage increased, the amount of tetramer decreased. Very small amounts of trimer were observed, but the major change was a dramatic increase in the amount of monomer. There was no evidence for any dimer formation. These results were consistent with the avidin tetramer dissociating first into a trimer and a monomer, with the unstable trimer then rapidly dissociating into addition monomers. At the highest trap voltages backbone cleavage occurred and the monomers then dissociated into peptides.

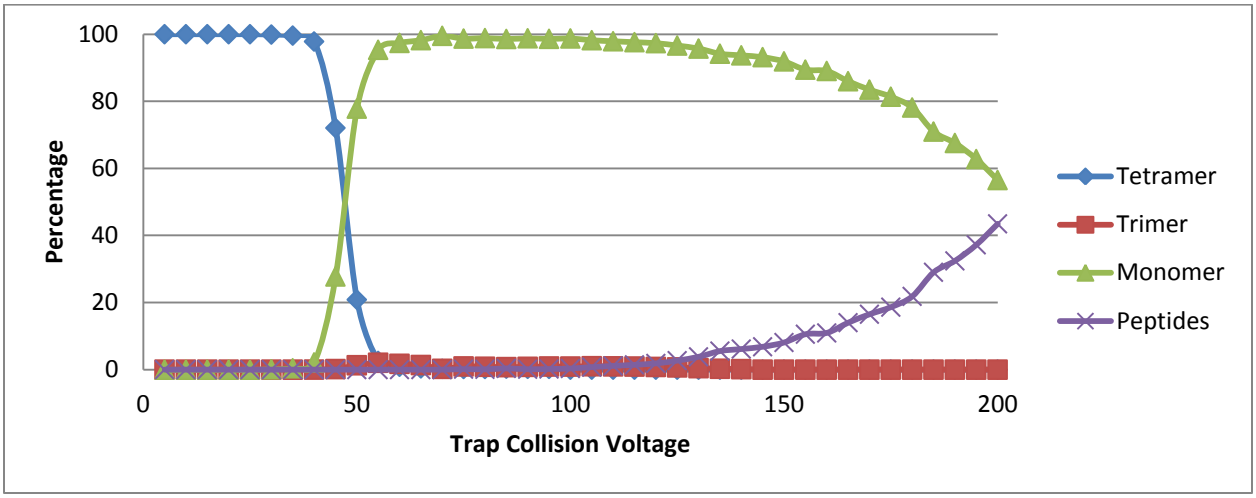


Figure 5.9. CID profile of unmodified avidin. The analysis of dissociation products as a function of trap collision voltage from unmodified avidin revealed that avidin dissociates first into a trimer, then the trimer dissociates into monomers. At higher collision voltages, the monomers dissociate into peptides.

The CID profile for crosslinked avidin (Figure 5.10) was significantly different than unmodified avidin. First, significantly more energy was required to dissociate the DC4-crosslinked avidin compared to the native avidin. At a trap collision voltage of 50 V unmodified avidin was almost completely dissociated but dissociation of the DC4-crosslinked tetramer did not begin until 65 V. The dissociation exhibited a broader energy profile than observed for the unmodified tetramer, consistent with CIU profiles. As the amount of tetramer decreased, low levels of both trimer and dimer appeared. At higher energies monomer appeared and backbone cleavage leading to peptide dissociation followed closely.

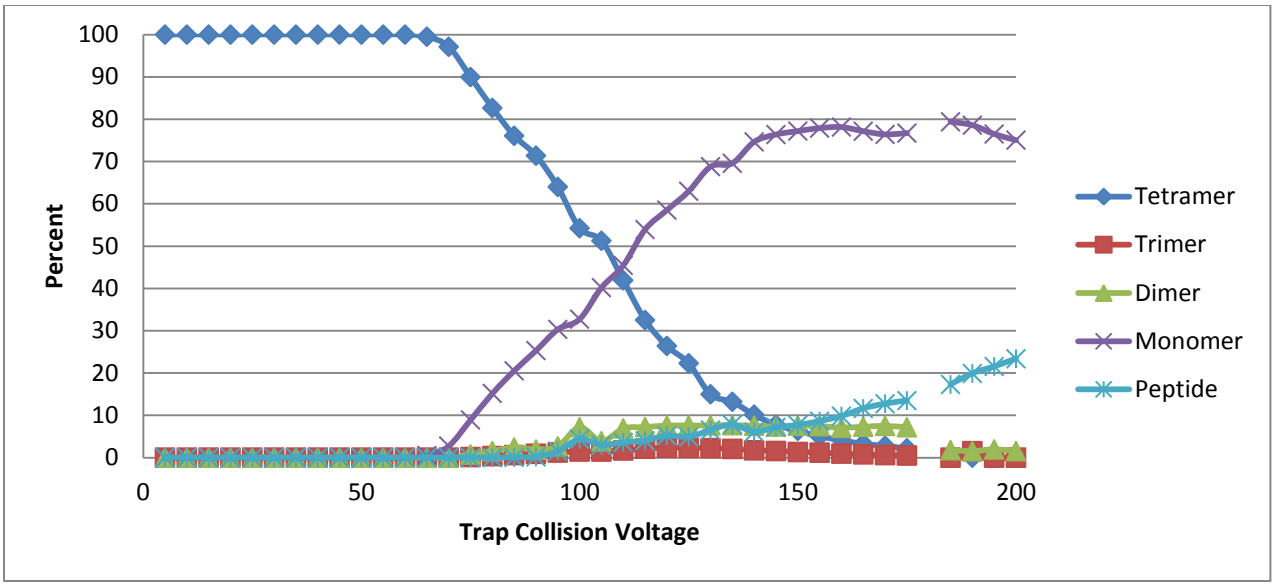


Figure 5.10. CID profile of DC4-crosslinked avidin. Profile of dissociation products as a function of trap collision voltage from DC4-crosslinked avidin.

One of the rationales for combining the DABCO crosslinkers with IM-MS was that the presence of the fixed charges on the crosslinkers might cause a more symmetric distribution of charge after dissociation. To determine if this hypothesis was correct, the charge states of the products from dissociation of tetrameric avidin were compared before and after crosslinking with DC4 (Figure 5.11). When unmodified avidin was dissociated, the most intense charge state of both the resulting trimer and monomer was the +10 charge state. Symmetric portioning would lead to a quarter of the charge being transferred to the monomer, but the typical asymmetric partitioning of charge was observed, as discussed earlier. Dissociation of the DC4-crosslinked avidin resulted in monomer, dimer, and trimer charge envelopes. The monomer had a charge distribution centered near the +8 charge state (lower than the unmodified avidin monomer), while the trimer from DC4-crosslinked avidin was shifted to a slightly higher charge state (+11) than the unmodified trimer. These data suggest that DC4 crosslinking partially alleviates the asymmetric distribution of charges. A minimum amount of charge is required on a complex to

dissociate it, so when asymmetric charge distribution occurs, the dissociation experiments that can be done on the complex are limited. By making the charge distribution more symmetric, a greater amount of charge remains on the larger part of the complex. Further dissociation experiments can be completed, proving additional information about the complex components, structure, and connectivity.

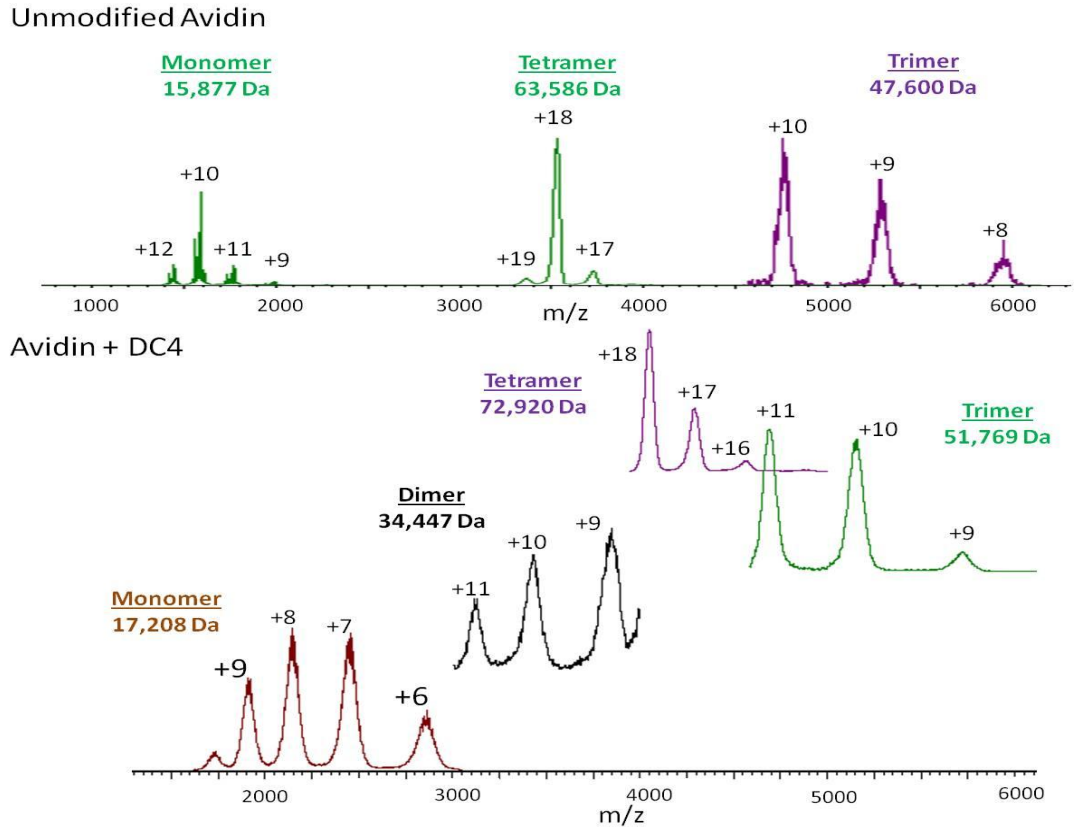


Figure 5.11. Analysis of charge distribution after CID of unmodified and DC4-crosslinked avidin. The unmodified avidin tetramer dissociated into a trimer and monomer and the charge distributions between the two products were approximately equal. The lower set of spectra were the dissociation products from DC4 crosslinked-avidin. The monomer has less charge than the monomer from the noncrosslinked avidin and both the trimer and dimer have higher charge states than the monomer.

Unmodified avidin initially dissociated into a trimer and a monomer with no dimer observed. DC4-crosslinked avidin dissociated to trimer, dimer, and monomer which was

consistent with crosslinker cleavage being the controlling event in dissociation. As the trap collision voltage increased, and thus the energy applied to the ions increased, the level of monomer increased and peptides started appearing. This order suggested that most of the peptide ions observed were generated via dissociation of the monomer. This was consistent with the Chapter 4 data in which the energy required to cleave the crosslinker DC4 was less than the energy required to fragment the peptide bond.

During the CID process salt adducts can also be dissociated from the complex. When this occurs, a mass loss of the complex is observed. The original mass of the DC4-crosslinked avidin tetramer prior to CID was 74 kDa, but after CID, the tetramer mass was 73 kDa indicating a significant amount of salt adducts. Masses were observed that were several hundred Daltons but those masses did not exhibit the isotope distribution of peptides. This indicated that clusters of salt were dissociated from the monomer. One of the most intense ions was found at m/z 199.1 which was observed in Chapter 4 as a fragment from dead-end peptides. Here, the 199.1 may have fragmented off dead-ends on the tetramer or from hydrolyzed crosslinker interacting with the surface of the crosslinked protein. Even after loss of the salt adducts from the tetramer, the masses of the dissociated subunits were larger than expected suggesting the presence of additional tightly bound adducts. The salt adducts may have an effect on charge partitioning and removing them may result in a more symmetric partitioning of charge, however it is also likely that protons were still present which would contribute to charge partitioning.

A second rationale for combining crosslinking with IM-MS was to improve top-down peptide fragmentation of intact proteins. It was hypothesized that constraining the structure by crosslinking might alter the patterns of peptide fragmentation observed. An analysis of the peptides generated by CID of the unmodified avidin tetramer revealed several peptides, primarily

from the C-terminus (Table 5.7). MS/MS was not available for these ions, so they were identified by mass alone, taking the mass accuracy of the instrument into account. A list of theoretical masses for b-, y-, and internal ions from the avidin monomer was generated and compared to the masses observed. Each observed mass had two to three possible peptide sequences. If the listed mass required loss of NH_3, water or carbon monoxide, those possibilities were considered the least likely. Top-down fragmentation frequently occurs at acidic residues and before proline, thus cleavages at those residues were more likely. Masses that required only one cleavage event were considered the most likely because single cleavages are energetically more favorable and top-down fragmentation frequently produces fragments from the termini. When peptides first started appearing at lower trap collision voltages, the peptides observed were peptides which could be assigned to the C-terminus (resulting from one cleavage event). As the trap collision voltage increased, other peptides appeared that could not be assigned to the C-terminus.

Table 5.7. Peptide masses observed after CID of unmodified avidin tetramers. Although suggested peptide sequences were listed, in a theoretical calculation of a top-down avidin experiment, each mass had two to three potential peptides. Residue 152 is the C-terminal residue.

m/z	z	Sequence	Residues
772.9	3	WKATRVGINIFTRLRTQKE	134-152
811.3	3	DWKATRVGINIFTRLRTQKE	133-152
837.8	2	VGINIFTRLRTQKE	139-152
906.3	3	IGDDWKATRVGINIFTRLRTQKE	130-152
989.7	3	SVNDIGDDWKATRVGINIFTRLRTQK-H_2O	126-151
993.3	2	ESTTVFTGQCFIDRNGKE-28	98-115
1057.4	2	DRNGKEVLKTMWLLRSSV	110-127
1138.0	2	FIDRNGKEVLKTMWLLRSS	108-126
1150.5	2	DIGDDWKATRVGINIFTRLR-28	129-148
1158.9	2	WKATRVGINIFTRLRTQKE	134-152
1200.5	2	DDWKATRVGINIFTRLRTQK	132-151
1207.9	2	IGDDWKATRVGINIFTRLRTQ-28	130-150
1216.4	2	DWKATRVGINIFTRLRTQKE	133-152
1236.9	2	SLTGKWTNDLGSNMTIGAVNSRGE-NH_3	29-51
1256.9	2	RNGKEVLKTMWLLRSSVNDIGD	111-132
1265.5	2	VLKTMWLLRSSVNDIGDDWKAT	116-137
1350.5	2	DLGSNMTIGAVNSRGEFTGTYITAVTA-28	37-62
1359.0	2	IGDDWKATRVGINIFTRLRTQKE	130-152

The DC4-crosslinked avidin dissociated into twice as many unique peptides (Table 5.8) as unmodified avidin. The m/z values listed in Table 5.8 were assigned as peptides based on their isotopic distribution, however, the identities of the peptides were not assigned. Each m/z value had at least three potential peptide sequences that would have accounted for the observed mass. This does not take into consideration that the masses could be crosslinked or contain a DC4 fragment or dead-end. Most of the possible peptide matches were not found at the C-terminus so it was not possible to confidently assign the m/z values observed to a specific part of the avidin sequence. However, it was clear that crosslinking intact proteins did change the top-down fragmentation pattern, which resulted in a different series of peptides.

Table 5.8. Peptides masses observed after CID of DC4-crosslinked avidin tetramers

m/z	z	m/z	Z
254.9	2	879.2	2
311.0	2	893.2	2
323.0	2	931.2	3
331.5	2	1204.8	2
335.9	2	1218.3	2
337.0	2	1226.8	2
340.0	2	1244.8	2
343.9	2	1253.8	2
345.5	2	1262.8	2
358.5	2	1266.8	2
367.1	2	1275.8	2
374.9	2	1284.3	2
393.1	2	1309.8	2
396.0	2	1319.3	2
401.5	2	1323.4	2
812.5	3	1331.8	2
818.1	3	1340.4	2
823.2	2	1396.4	2
830.6	2	1427.4	2
849.8	3	855.5	3
855.5	3		

5.6 Conclusions:

In this chapter, it has been demonstrated that chemical crosslinking can be combined with ion mobility mass spectrometry to useful information and the properties of crosslinked protein monomers were characterized by IM-MS. Chemical crosslinking with the DABCO crosslinkers caused the CCS, the mass, and the charge of the complex to increase. Several indications suggested that the presence of significant levels of salt adducts on crosslinked proteins. The initial measurements of crosslinked lysozyme, avidin, and aldolase, the analysis of the tagged avidin, and the results of CID from crosslinked avidin support the hypothesis that salt adducts were at least partially responsible for the increased mass, increased CCS, and charge state

distribution of crosslinked proteins. However, evidence also pointed to an increase in protein rigidity caused by crosslinking, supported by the tagged avidin and by the CIU data.

As anticipated, the crosslinked multimers were more stable and required more energy both to unfold from the native state and to dissociate. However, the data clearly showed that large quantities of salt adducts, most likely hydrolyzed crosslinker, were adhered to the protein surface. Certain ions have been shown to stabilize protein structure[54, 55] by forming ion pairs, so to determine the extent to which crosslinking stabilizes the structure as opposed to the ions, similar experiments must be completed with crosslinked proteins that do not have large quantities of salt adducts. Although extensive desalting was performed on the crosslinked proteins before analysis, it was clear that salt adducts were still present and more effective protein desalting steps are necessary. Long term dialysis, ion exchange purification, or several size exclusion steps might be necessary to remove the salt associated with these complexes. Another possibility is to exchange for salts that bind with less avidity. CIU and CID profiles of the tagged avidin would also provide insight.

It was clear from the CID data that crosslinking changes the protein fragmentation patterns. The percentages of the released subunits differed between crosslinked and unmodified proteins which suggested that the tetramers might be dissociating through a different pathway. This was most likely due to the addition of covalent crosslinks which prevented subunit dissociation. After dissociation, the distribution of charge was also more symmetric for the crosslinked protein complexes with proportionally more charge retained on the larger product. The asymmetric distribution of charge has been a major limitation to CID of intact complexes because it limits the number of dissociations from the complex. By alleviating the asymmetric distribution of charge, additional dissociation experiments can be completed to yield further data

regarding the components and structure of a complex. Only a partial return to charge symmetry was observed and this may be due in part to the presence of salt adducts. Removal of the remaining salt adducts may lead to a further improvement in the symmetrical distribution of charge. The quaternary amine-containing protein tags may also alleviate the asymmetric distribution of charge and CID and CIU data from tagged avidin would allow this to be determined.

A much richer set of peptide fragment data was obtained from the DC4-crosslinked avidin than from the unmodified avidin and the peptides appeared to be from different regions of the avidin structure. With the available data, the DC4-crosslinked peptides could not be identified; however, a bottom-up analysis of the crosslinked avidin, particularly on a high resolution mass spectrometer, should facilitate that analysis.

5.7 Materials and Methods:

Materials. Avidin (chicken egg white), alcohol dehydrogenase (yeast, ADH), pyruvate kinase (rabbit muscle, PK), transthyretin (human, TTR), Concanavalin A (jack bean, ConA), glutamate dehydrogenase (bovine liver, GDH), cytochrome C (equine heart, CYC), aldolase (rabbit muscle), and lysozyme were purchased from Sigma Aldrich. Micro bio-spin columns with bio-gel P6 in SSC buffer were purchased from Biorad and Microcon centrifugal filter devices were purchased from Millipore. All other reagents and supplies were purchased from standard chemical suppliers.

Crosslinking Reaction and Sample Clean-Up. Avidin was dissolved in 100 mM HEPES pH 7.0 to make a 5 μg/μl solution. Solutions of 10 mM of each crosslinker or tag, PC1, DC4, DC5, and DC6, DABCO-Tag, and MeDABCO tag were prepared in 100 mM HEPES, pH 7.0

directly before use. The crosslinker or tag (26 μL) was added to 40 μL of 5 μg/μL avidin in 220 μL 100 mM HEPES and allowed to react for 30 minutes. A dilution of was made of DC4 to 1mM and the same reaction was carried out.

Crosslinked and native samples were then applied to Millipore microcon centrifugal filter devices with a 30,000 Da molecular weight cutoff and used according to the manufacturer's instructions. The samples were concentrated to 10 μl by spinning at 14,000 x g for 12 minutes. The samples were then washed with 500 μL of 1 M ammonium acetate and the volume again reduced to 10 μl by spinning at 14,000 x g for 12 minutes. The sample was eluted from the centrifugation device, diluted to 20 μL, and desalted further with a biorad microspin column according to the manufacturer's instructions. Briefly, the microspin columns had been washed with 1M ammonium acetate and spun for one minute at 1,000 x g four times to exchange the SSC buffer. Then the sample was applied to the column and spun for 4 minutes at 1,000 x g. This procedure was repeated with a microspin column washed with 200 mM ammonium acetate and samples were used without further purification or dilution for ion mobility mass spectrometry. An aliquot of 10 μL was saved for SDS-PAGE analysis. The tagged samples were desalted once on the microspin columns with 200 mM ammonium acetate and used without further purification. Protein solutions for CCS calibration were prepared as either 50 μM or 100 μM solutions in 200 mM ammonium acetate and stored at -80°C. The proteins solutions were desalted immediately before use using the microspin columns equilibrated with 200 mM ammonium acetate and then the protein complexes were diluted to 10 μM. After desalting, 0.5 M m-nitrobenzylalcohol (mNBA) was added to avidin and PC1 crosslinked avidin so the final concentration was 100 mM.

SDS-PAGE. See Section 2.6.

Ion Mobility Mass Spectrometry. A quadrupole ion mobility time-of-flight mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA, USA) was used for all ion mobility experiments. Approximately 10 μL of each sample were injected via nESI source in positive ion mode. The capillary was applied at voltages of 1.7-1.9 kV and the sampling cone was used at 127 V. All instrument settings were optimized to maintain the integrity of the noncovalent complex. An argon gas pressure of 3.3^{-2} mbar and a 50 V trap bias was used for the trap travelling-wave ion guide. The IMS was operated with a wave height at 40V, the IMS wave velocity at 600 m/s, the transfer velocity at 300 m/s, and the nitrogen gas pressure was approximately 3.5 mbar for purposes of achieving IMS separation. The TOF-MS scanned an m/z range of 100-8000 and contained a pressure of $1.6e^{-6}$ mbar.

Collision Induced Unfolding and Dissociation. To obtain the CIU/CID data, ions were selected at the appropriate m/z value which corresponded to the +18 charge state of avidin in the quadrupole mass filter. The +18 charge state was chosen because that charge state was observed in all crosslinked and noncrosslinked samples, and other oligomers were not observed at that m/z value. The trap collision voltage was increased in 5V steps, and then the voltage in the transfer region was increased in 50V steps.

Data Analysis. All mass spectra were analyzed using MassLynx 4.1 software and calibrated externally using a 100 mg/mL solution of cesium iodide. Minimal smoothing and no background correction steps were taken. For CCS measurements, PK, ADH, GDH, and TTR were used as calibrants via a previously published method[50]. CYC was not used because with the current instrument settings, it was highly unfolded.

The relative abundances of each subunit were calculated by selecting the region of the 2D spectrum that that subunit occupied and summing the intensities of ions in that region. That

number was divided by the sum of the intensities of all the ions in the 2D plot. The folded tetramer had a narrow distribution of drift time depending on the charge state. To calculate the percentage of folded tetramer, one charge state was chosen and the intensity of ions in that specific drift time distribution for that charge state was summed. Then it was divided over the sum of all the ions at any drift time in the specific m/z window chosen[55].

5.8 References:

[1] A. Sinz, *Mass Spectrometry Reviews* **2006**, *25*, 663.
[2] A. Sinz, S. Kalkhof, C. H. Ihling, *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 1921.
[3] J. Pettelkau, T. Schroder, C. H. Ihling, B. E. S. Olausson, K. Kolbel, C. Lange, A. Sinz, *Biochemistry* **2012**, *51*, 4932.
[4] Y. Lu, M. Tanasova, B. Borhan, G. E. Reid, *Analytical Chemistry* **2008**, *80*, 9279.
[5] E. V. Petrotchenko, K. Xiao, J. Cable, Y. Chen, N. V. Dokholyan, C. H. Borchers, *Molecular & Cellular Proteomics* **2009**, *8*, 273.
[6] E. V. Petrotchenko, J. J. Serpa, C. H. Borchers, *Molecular & Cellular Proteomics* **2011**, *10*.
[7] E. V. Petrotchenko, V. K. Olkhovik, C. H. Borchers, *Molecular & Cellular Proteomics* **2005**, *4*, 1167.
[8] C. C. Wu, F. Herzog, S. Jennebach, Y. C. Lin, C. Y. Pai, R. Aebersold, P. Cramer, H. T. Chen, *Proc Natl Acad Sci U S A* **2012**, *109*, 19232.
[9] J. L. P. Benesch, J. A. Aquilina, B. T. Ruotolo, F. Sobott, C. V. Robinson, *Chem Biol* **2006**, *13*, 597.
[10] L. F. Santos, A. H. Iglesias, E. J. Pilau, A. F. Gomes, F. C. Gozzo, *J Am Soc Mass Spectrom* **2010**, *21*, 2062.
[11] L. L. Ilag, H. Videler, A. R. McKay, F. Sobott, P. Fucini, K. H. Nierhaus, C. V. Robinson, *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 8192.
[12] C. L. Hanson, H. Videler, C. Santos, J. P. Ballesta, C. V. Robinson, *J Biol Chem* **2004**, *279*, 42750.
[13] H. Videler, L. L. Ilag, A. R. C. McKay, C. L. Hanson, C. V. Robinson, *FEBS Letters* **2005**, *579*, 943.
[14] C. L. Hanson, P. Fucini, L. L. Ilag, K. H. Nierhaus, C. V. Robinson, *J Biol Chem* **2003**, *278*, 1259.
[15] Y. Gordiyenko, S. Deroo, M. Zhou, H. Videler, C. V. Robinson, *J Mol Biol* **2008**, *380*, 404.
[16] J. L. Benesch, B. T. Ruotolo, F. Sobott, J. Wildgoose, A. Gilbert, R. Bateman, C. V. Robinson, *Anal Chem* **2009**, *81*, 1270.
[17] K. Pagel, S. J. Hyung, B. T. Ruotolo, C. V. Robinson, *Anal Chem* **2010**, *82*, 5363.
[18] J. C. Jurchen, D. E. Garcia, E. R. Williams, *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 1373.
[19] S. N. Wanasundara, M. Thachuk, *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 2242.
[20] J. L. Benesch, J. A. Aquilina, B. T. Ruotolo, F. Sobott, C. V. Robinson, *Chem Biol* **2006**, *13*, 597.
[21] J. L. P. Benesch, *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 341.
[22] S. N. Wanasundara, M. Thachuk, *The Journal of Physical Chemistry A* **2009**, *113*, 3814.
[23] J. L. P. Benesch, B. T. Ruotolo, F. Sobott, J. Wildgoose, A. Gilbert, R. Bateman, C. V. Robinson, *Analytical Chemistry* **2008**, *81*, 1270.

[24] E. van Duijn, D. A. Simmons, R. H. H. van den Heuvel, P. J. Bakkes, H. van Heerikhuizen, R. M. A. Heeren, C. V. Robinson, S. M. van der Vies, A. J. R. Heck, *Journal of the American Chemical Society* **2006**, *128*, 4694.
[25] Y. Gordiyenko, H. Videler, M. Zhou, A. R. McKay, P. Fucini, E. Biegel, V. Müller, C. V. Robinson, *Molecular & Cellular Proteomics* **2010**.
[26] J. C. Jurchen, E. R. Williams, *Journal of the American Chemical Society* **2003**, *125*, 2817.
[27] J. A. Loo, C. G. Edmonds, R. D. Smith, *Analytical Chemistry* **1991**, *63*, 2488.
[28] E. Jacob, R. Unger, *Bioinformatics* **2007**, *23*, e225.
[29] B. T. Ruotolo, C. V. Robinson, *Current Opinion in Chemical Biology* **2006**, *10*, 402.
[30] I. Michaelevski, M. Eisenstein, M. Sharon, *Analytical Chemistry* **2010**, *82*, 9484.
[31] S. Kalkhof, A. Sinz, *Analytical Biochemistry* **2008**, *392*, 305.
[32] R. Diamond, *J Mol Biol* **1974**, *82*, 371.
[33] R. E. Canfield, *J Biol Chem* **1963**, *238*, 2698.
[34] J. Sygusch, D. Beaudry, M. Allaire, *Biochemistry* **1987**, *84*, 7846.
[35] L. Pugliese, M. Malcovati, A. Coda, M. Bolognesi, *J Mol Biol* **1994**, *235*, 42.
[36] R. J. DeLange, T. S. Huang, *J Biol Chem* **1971**, *246*, 698.
[37] T. S. Huang, R. J. DeLange, *J Biol Chem* **1971**, *246*, 686.
[38] T.-S. Huang, R. J. DeLange, *Journal of Biological Chemistry* **1971**, *246*, 686.
[39] M. Sajgo, G. Hajos, *Acta Biochim Biophys Acad Sci Hung* **1974**, *9*, 239.
[40] N. Blom, J. Sygusch, *Nature Structural Biology* **1997**, *4*, 36.
[41] Z. Zhang, C. B. Post, D. L. Smith, *Biochemistry* **1996**, *35*, 779.
[42] H. Pan, D. L. Smith, *Biochemistry* **2003**, *42*, 5713.
[43] M. A. Freitas, A. G. Marshall, *International Journal of Mass Spectrometry* **1998**, *182*, 221.
[44] S. Madler, C. Bich, D. Touboul, R. Zenobi, *Journal of Mass Spectrometry* **2009**, *44*, 694.
[45] U. A. Mirza, S. L. Cohen, B. T. Chait, *Analytical Chemistry* **1993**, *65*, 1.
[46] I. A. Kaltashov, A. Mohimen, *Analytical Chemistry* **2005**, *77*, 5370.
[47] A. Dobo, I. A. Kaltashov, *Analytical Chemistry* **2001**, *73*, 4763.
[48] R. Grandori, *Protein Science* **2002**, *11*, 453.
[49] O. O. Sogbein, D. A. Simmons, L. Konermann, *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 312.
[50] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, B. T. Ruotolo, *Analytical Chemistry* **2010**, *82*, 9557.
[51] D. G. Isom, C. A. Castañeda, B. R. Cannon, P. D. Velu, B. García-Moreno E., *Proceedings of the National Academy of Sciences* **2010**, *107*, 16096.
[52] M. F. Perutz, J. C. Kendrew, H. C. Watson, *Journal of Molecular Biology* **1965**, *13*, 669.
[53] J. Fernandez de la Mora, *Analytica Chimica Acta* **2000**, *406*, 93.
[54] L. Han, S.-J. Hyung, J. J. S. Mayers, B. T. Ruotolo, *Journal of the American Chemical Society* **2011**, *133*, 11358.
[55] L. Han, S.-J. Hyung, B. T. Ruotolo, *Faraday Discussions* **2013**.

Chapter 6

Conclusions and Future Directions

Mapping protein-protein interactions and the dynamics of those interactions is crucial to defining physiological states, building effective models for understanding cell function, and to allow more effective targeting of new drugs[1]. Chemical crosslinking combined with mass spectrometry is a method that has been used to put constraints on structures and identify protein interacting partners in systems such as RNA Polymerase I[2], RNA Polymerase II[3], the ribosome[4-6], signal transduction complexes[7], and the membrane proteome of *Shewanella oneidensis* MR-1[8]. Recently, crosslinking studies have been completed *in vivo* to identify protein interacting partners in a more high throughput manner[9-11].

Although the number of publications combining chemical crosslinking and mass spectrometry for studying protein structures and interactions has increased, there are still challenges associated with the method. Many chemical crosslinkers use N-hydroxysuccinimide (NHS) esters as the functional group which reacts with primary amines in lysyl residues and on the N-terminus. At physiological pH, primary amines are protonated and carry a net positive charge. After crosslinking with NHS esters, the primary amine is converted to an amide without a positive charge. There is a relationship between protein structure and charge, therefore, elimination of the positive charge may cause changes in protein structure. Additionally, proteins crosslinked at lysyl residues are more difficult to digest producing large, branched peptides

which can be difficult to ionize and fragment. When the crosslinked peptides fragment well, two individual peptides are simultaneously being fragmented which results in two overlapping series of ions in each spectrum. Therefore, data interpretation is very complex. Lastly, the crosslinked peptides are low in abundance in a complex mixture of unmodified peptides, dead-end peptides, loop-links, and crosslinked peptides. Several efforts have been made to overcome these challenges, mostly by designing new crosslinking reagents with varying levels of success[12-21]. However, there is not a crosslinker that meets all the criteria, so further improvements are still needed.

6.1 Characterization of the Mass Spectrometry Cleavable Crosslinker PC1

One of the main goals of this thesis was to design and synthesize new crosslinkers to surmount the challenges associated with chemical crosslinking, and then to apply those crosslinkers to protein complexes. Initially the mass spectrometry cleavable crosslinker PC1 was designed and synthesized. Then I extensively characterized the fragmentation behavior of model acetylated crosslinked peptides utilizing several of the most common activation methods. Backbone fragmentation was observed using all activation methods and the PC1-crosslinked peptides fragmented within the crosslinking moiety using CID, HCD, and ECD but not IRMPD. During CID on a MALDI-TOF-TOF mass spectrometer, the PC1-crosslinked model peptides exhibited cleavage and rearrangement of the crosslinker moiety, producing marker ions. Those marker ions facilitated the rapid identification of a spectrum as a crosslinked peptide. However, when PC1 was applied to the tetrameric protein aldolase and analyzed on the MALDI-TOF-TOF, the fragmentation pattern changed and backbone fragmentation was no longer observed. The marker ions were still observed but without further backbone fragmentation, the peptide

sequence was not identifiable. This indicated that PC1 would not work for routine analysis of protein complexes on a MALDI-TOF-TOF mass spectrometer.

Similar data was collected for PC1-crosslinked aldolase on an ESI-LTQ-Orbitrap. Crosslinked peptides were not identified, but several highly charged ions that were potentially crosslinked peptides were present in the MS/MS data. These spectra were not identifiable by *de novo* sequencing and the tools to search each ESI spectrum are not available. Therefore, to further process these data, we have started a collaboration with Nuno Bandiera at the University of California – San Diego to generate algorithms capable of analyzing the PC1-crosslinked data generated by electrospray on the Orbitrap. Although PC1 was not a useful crosslinker when analyzing crosslinked complexes on a MALDI-TOF-TOF, it is possible that with a suitable algorithm to analyze PC1-crosslinked data, PC1 may useful for analysis of complexes on the LTQ-Orbitrap.

At this point, the design of PC1 was modified to improve the fragmentation properties of crosslinked peptides. However, there are several questions remaining regarding PC1-crosslinked peptides that could further contribute to the literature and to the design of future crosslinking reagents. In order to generate crosslinked peptides, N-terminally acetylated peptides are often used in initial evaluations of new crosslinkers. With PC1, there was a difference in the fragmentation behavior in a MALDI-TOF-TOF mass spectrometer of PC1-crosslinked peptides when they were acetylated compared to when they were not acetylated. The acetylated peptides were similar to tryptic peptides because each contained a C-terminal arginine residue, but the PC1-crosslinked aldolase was digested with Endoprotease-GluC (GluC). GluC cleaves at acidic residues which would result in a change in basicity of the peptides, and possibly a change in size, either of which may have led to the change in fragmentation behavior. Analyses of the

fragmentation behavior of PC1-crosslinked model peptides with varying C-termini, lengths, and both acetylated and non-acetylated N-termini should yield insights. Furthermore, the fragmentation behavior may have been partially charge state dependent, thus comparing the intensity of fragment ions observed over different charge states could be informative. The data were not conclusive regarding the amount of energy needed to fragment PC1-crosslinked peptides within the crosslinking moiety, however varying the collision energy on the Orbitrap and analyzing the intensity of the fragments may be useful.

Upon fragmentation via CID in the MALDI-TOF-TOF, PC1 crosslinked peptides yield marker ions. Structures and potential mechanisms to account for these marker ions were suggested in Chapter 2, however, they are untested. To verify this mechanism, deuterium labeled PC1 could be synthesized and applied to model peptides. If the proposed mechanism were true and the eight hydrogen atoms in the piperazine ring were deuterium atoms, the marker ion at m/z 111.1 would shift to 118.1. If the four hydrogen atoms between the piperazine ring and the carbonyl were deuterium atoms, the marker ion at m/z 99.1 would shift up by 2 Da and the marker ion at m/z 139.1 should shift up by 3 Da. These observations would be consistent with the suggested mechanism.

6.2 Crosslinkers Containing Quaternary Amines

The fragmentation patterns of PC1 suggested that PC1 was not labile enough for routine CID, therefore, the next goal was to synthesize a more labile crosslinker. In Chapter 3, I synthesized a new series of crosslinkers, DC4, DC5, and DC6, which were predicted to be more labile than PC1 because they each contain two intrinsic positive charges. In Chapter 4, DC4 was extensively characterized and it was observed that the intrinsic positive charges allow

crosslinked peptides to fragment into their component peptides by collision-induced dissociation (CID) or in-source decay (ISD). Initial fragmentation events resulted in cleavage on either side of the positive charges so crosslinked peptides were identified as pairs of ions separated by defined masses. The structures of the component peptides were then robustly determined by MS^3 because their fragmentation products rearrange to generate a mobile proton. On a MALDI-TOF-TOF instrument, ISD occurred readily in DC4-crosslinked peptides but to a lesser degree in DC5- or DC6-crosslinked peptides. Often the ISD products from DC4-crosslinked peptides were of similar or greater intensity than the crosslinked parent. However, the ISD products from the DC5-crosslinked peptides were 10-30 % of the parent intensity and the ISD from the DC6-crosslinked peptides were less than 10 % of the parent intensity. CID of the DC6- and DC5-crosslinked parent ions did result in similar fragmentation patterns but DC4-crosslinked peptides fragmented more efficiently (Figure 6.1). As determined from the synthesis and this comparison of the fragmentation of crosslinked peptides, the shorter crosslinkers are the most labile. A shorter version of the DABCO crosslinkers, DC2, was very reactive so the synthesis and isolation was unsuccessful. However, in all attempts to make DC3, the reactivity of the NHS ester was not the issue but the elimination product was obtained instead of the selective displacement. Because several attempts to synthesize DC3 failed, I am suggesting a slightly different scheme to synthesize a version of DC3 as shown (Scheme 6.1). This synthetic route does not have acidic hydrogen atoms next to the carbonyl, thus preventing the elimination reaction.

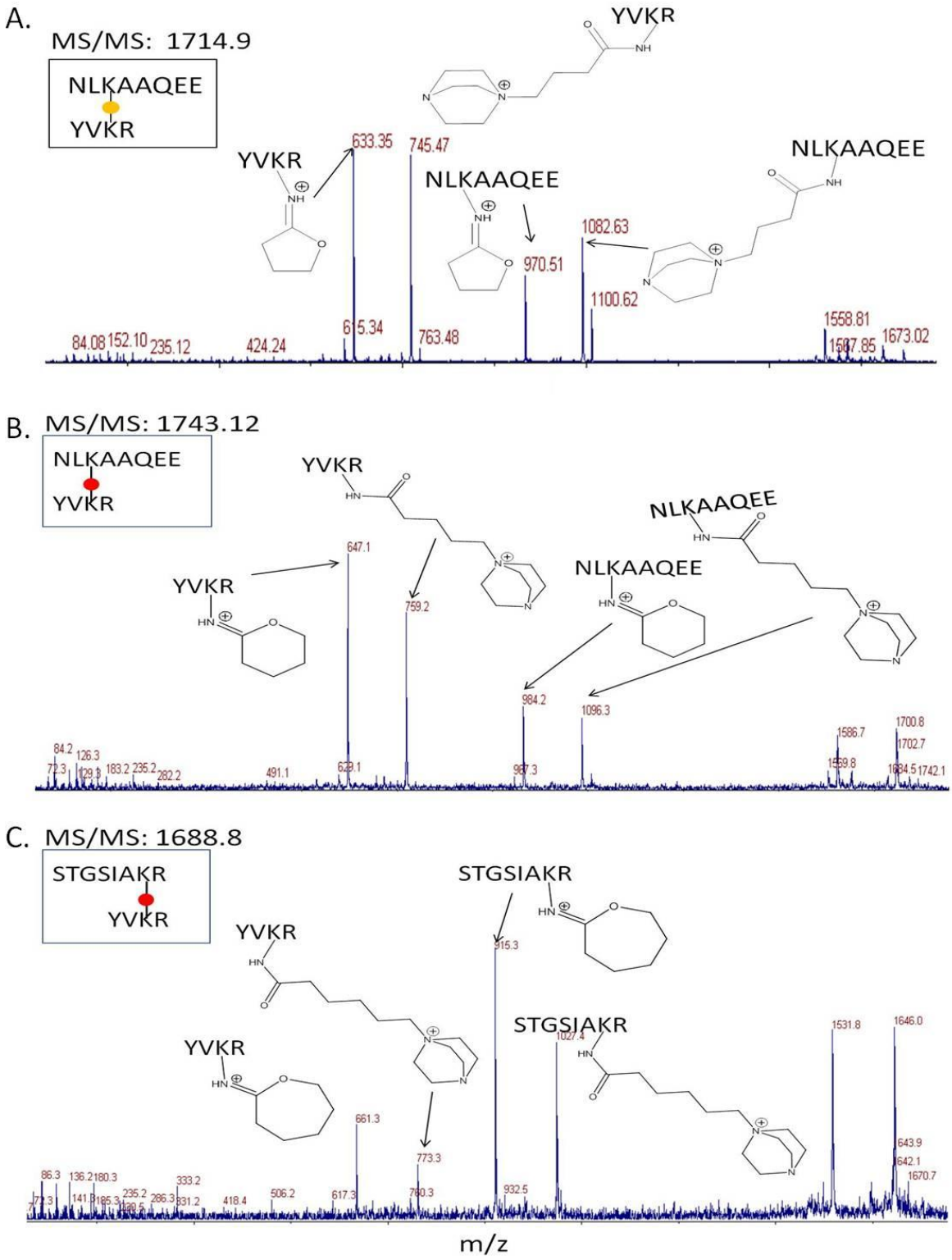
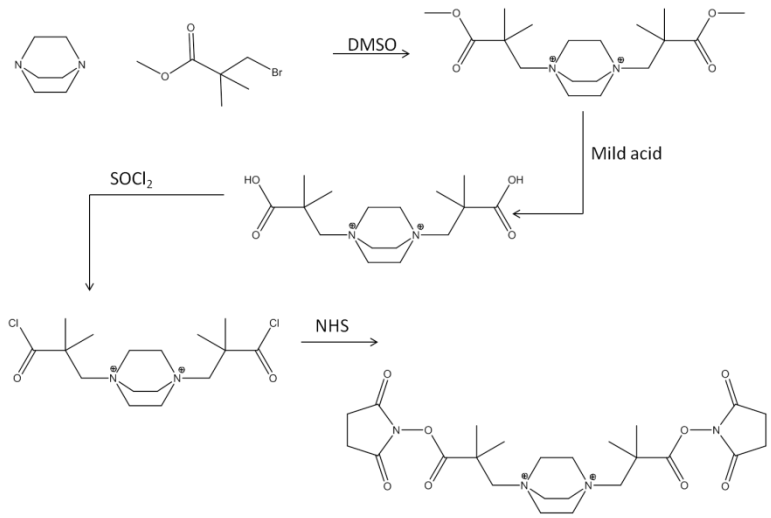


Figure 6.1. CID on a MALDI-TOF-TOF mass spectrometer of crosslinked peptides. DC4-crosslinked peptides fragmented the most efficiently by CID (A) but DC5- (B) and DC6-crosslinked peptides (C) did exhibit similar fragmentation patterns.



Scheme 6.1. **Potential synthetic pathway for DC3.**

The crosslinking reagent DC4 overcomes several of the intrinsic difficulties associated with chemical crosslinking, however, there are still two main problems to be addressed. First, crosslinked peptides are low in abundance compared to non modified peptides and dead-end peptides. We have developed an antibody to the crosslinker DC4 (Section 4.4) and characterization of that antibody has been initiated. The antibody is antigenic to some forms of the crosslinker, but whether it is antigenic to crosslinked peptides or to dead-end peptides has yet to be determined. Further characterization of the antibody is needed. If there are epitopes that recognize crosslinked peptides as opposed to dead-end peptides, those need to be purified. One method of doing this would be to crosslink a large, acetylated peptide with an arginyl residue near the C-terminus and an internal lysyl residue. After crosslinking, trypsin could be used to remove the acetyl group and generate a new primary amine at the N-terminus. Because both crosslinked peptides and dead-end peptides would result, size exclusion chromatography could be used to separate the crosslinked peptide from the dead-end peptide. Next, the separated DC4-

crosslinked peptides could be coupled to a magnetic bead and the sera applied to the magnetic beads. Antibodies that recognize the crosslinked peptides should bind to the magnetic beads, but epitopes that do not recognize the crosslinked peptides should not. Another possibility is to do a reverse purification coupling the dead-end peptides to the magnetic beads. If there are not epitopes that recognize the crosslinked peptides, because the DABCO crosslinkers are antigenic one could use only the crosslinked peptide as an antigen or try to raise monoclonal antibodies for higher specificity. This would be beneficial because the crosslinked peptides could be enriched from the remaining peptides and more crosslinks could be detected.

The second major hurdle that needs to be overcome for this method to become a high throughput technology is the development of further algorithms. The filtering scheme used for DC4 crosslinked peptides with the TOF-TOF is a good first step, however it is very limited and has a high number of false positives. The first filter produces a list of ions within the same spectrum that are separated by 112 Da. These ions are potentially crosslinked. To rule out the false positives, a second tool needs to be developed that analyzes the MS/MS in two steps. First, when two ions are separated by 112 Da, the higher mass ion should fragment into the lower mass. The first filter also produces a parent mass that the two sets of peaks add up to. The second filter should analyze the MS/MS data of the parent and determine if those two sets of peaks are found in that spectrum. If so, that is an authentic crosslinked peptide. If not, the spectrum is most likely a false positive. A similar approach could be used with electrospray data.

With these tools in hand, DC4 crosslinking could become a high-throughput method to analyze protein-protein interactions. Future work in the Andrews lab with these crosslinkers will include analysis of pre-ribosomal complexes. In initial work on *Escherichia coli* cell lysates, I

have demonstrated that it is possible to crosslink cell lysates without grossly perturbing ribosome structure. Ribosomal particles can be separated based on their size and shape by applying cell lysates to sucrose gradients and then centrifuging the gradients at ultrahigh speeds. The small subunit (30S) will migrate slowly into the sucrose gradient, while the large subunit (50S) will migrate further, followed by the intact ribosome (70S), and lastly polysomes (multiple 70S particles on one mRNA). Monitoring the UV absorbance at a wavelength in which RNA absorbs yields a trace known as a polysome profile. When cell lysates were crosslinked and then separated on sucrose gradients by ultracentrifugation, there was not a significant difference in the profiles of DC4-crosslinked lysates to noncrosslinked lysates (Figure 6.2). Furthermore, western blots probed against ribosomal proteins show bands increasing in molecular weight. These data indicate the ribosome was being crosslinked, but without gross changes to the structure.

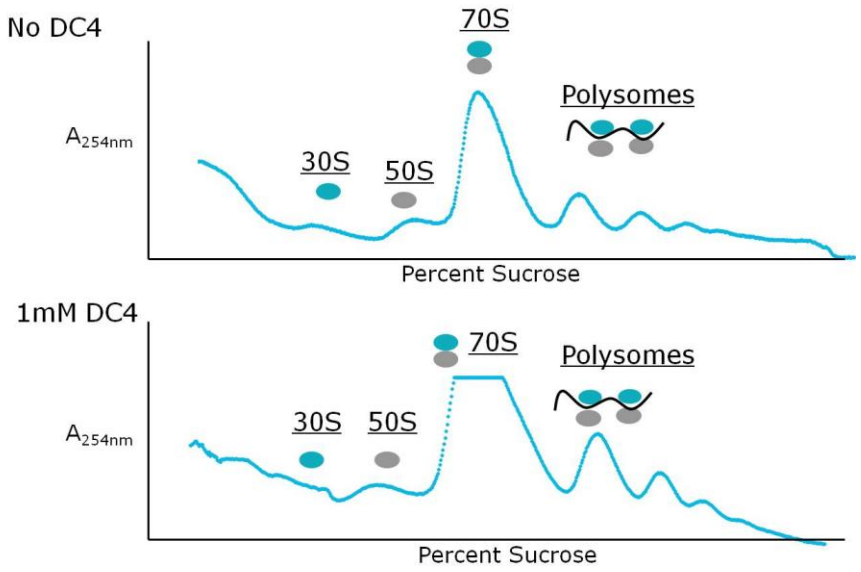


Figure 6.2. Ribosome profiles of DC4-crosslinked cell lysates and unmodified cell lysates. The nonmodified cell lysates (top profile) appeared similar to the DC4-crosslinked cell lysates (bottom profile) when separated via ultracentifugation. DC4 absorbs at similar wavelengths to RNA, therefore the DC44-crosslinked cell lysates appear more intense.

The bands on western blots representing crosslinked ribosome proteins are often found at the top of the well in which indicates that the ribosomal proteins are being crosslinked and that resulting crosslinked product is very large. One potential explanation for this observation was that the ribosomal proteins on the surface of the ribosomal RNA were crosslinked in a manner that prevented the dissociation of those proteins from the ribosome. Ribosomal proteins tend to be basic with a high percentage of lysyl residues thus allowing the formation of many crosslinks between proteins[4, 5, 22]. It was possible that there were enough crosslinks between the ribosomal proteins to trap the RNA to such an extent that even under denaturing conditions, the proteins were unable to be dissociate from the RNA. Another explanation was that the crosslinker was crosslinking the proteins to the RNA. A third possibility was that the crosslinked protein complexes did dissociate from the RNA, but were too large to migrate into the gel. To investigate this third possibility, lysates were crosslinked with increasing amounts of DC4 and then analyzed via agarose gel electrophoresis (Figure 6.3). Without DC4, the ribosomal RNAs (5S, 16S, and 23S) were resolved. As the amount of DC4 was increased, the RNA molecules were unable to migrate into the gel, indicating that either the proteins were not able to dissociate from the RNA or that the RNA itself was being crosslinked.

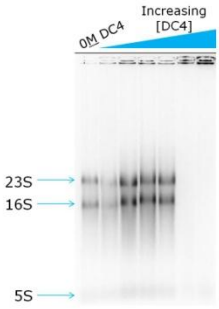


Figure 6.3. Agarose gel electrophoresis of DC4-crosslinked cell lysates. When cell lysates are separated via agarose gel electrophoresis, the ribosomal RNAs can be resolved. When 0M DC4 was added to *E. coli* cell lysates, bands representing the three ribosomal RNAs were visible. However after adding increasing amounts of DC4, the bands got stuck in the wells.

To date, we are not aware of any reports describing the reaction of NHS-esters with RNA. There have been reports that the hydroxyl groups on the side chains of seryl, tyrosyl, and threonyl residues do react with NHS-esters, although to a much lesser extent than primary amines[23, 24]. To determine whether the ribosomal RNA was reacting with DC4, short sequences of RNA were purchased (AAAAG, UUUUC, CCCCA, GGGGA), reacted with DC4, and analyzed by MALDI-TOF-TOF mass spectrometry. The majority of the oligoribonucleotides remained unmodified, but all the RNA sequences did have an observable ion at the calculated m/z for the dead-end reaction. No crosslinks between RNA molecules were observed. CID of the potential dead-end products suggested that the 3' and 5' hydroxyl groups were being modified (Figure 6.4). There was no evidence of nucleobase modification. Similar DNA sequences were purchased and DC4 dead-ends were observed on the DNA as well (data not shown). However, the fragmentation of DC4-modified DNA and RNA was much more efficient than the unmodified oligonucleotides. Furthermore, preferential cleavage occurred within the crosslinking resulting in reduced phosphoric acid loss in DC4-modified RNA spectra, and reduced nucleobase loss in DC4-modified DNA spectra.

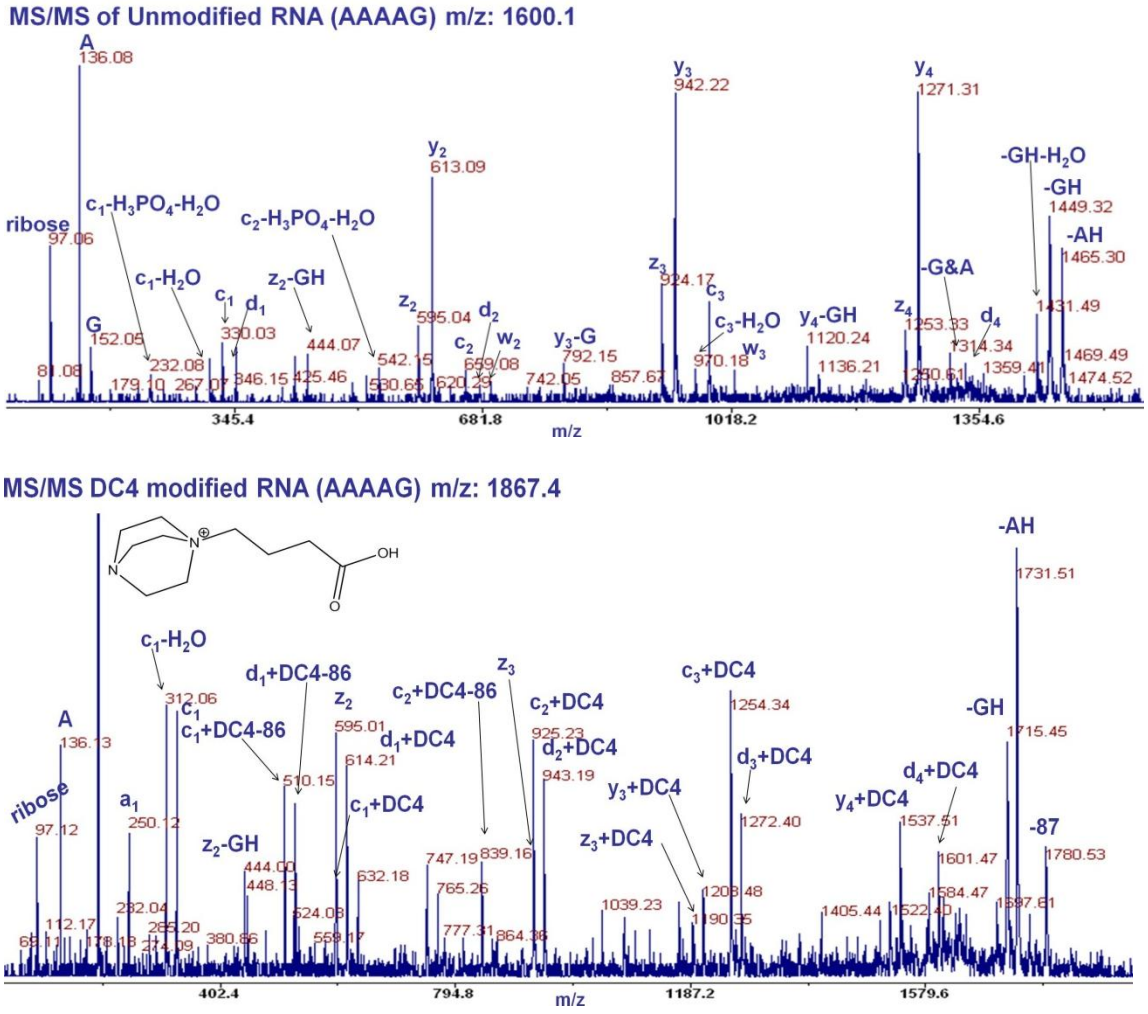


Figure 6.4. CID on a MALDI-TOF-TOF mass spectrometer of a unmodified and DC4-dead-end RNA. The unmodified RNA displayed the expected RNA fragmentation, mostly c, z, d, and w ions and the occasional loss of a phosphate group. The DC4-modified RNA had much more intense ions and preferentially had cleavage in the crosslinker moiety instead of the phosphate group.

The observation that DC4 does react with hydroxyl groups on DNA and RNA is potentially an important breakthrough. First, we have demonstrated feasibility that the ribosomal RNA could be crosslinking with the ribosomal proteins. If this is occurring, the use of DC4 could be used to map protein locations on the ribosome. This would be especially helpful for transiently associated factors to yield insights on their possible functions. Furthermore, other protein-RNA or protein-DNA complexes could be studied with these crosslinkers.

To make this possibility a reality, several key experiments need to be done. First, we have not yet observed a RNA-protein crosslink and are unaware of the fragmentation behavior. To verify that RNA and proteins can crosslink as well as to characterize the fragmentation behavior, highly basic peptides could be synthesized and reacted with the short RNAs discussed above. If ribosomal RNA is crosslinked to ribosomal proteins, then RNA-peptide crosslinks exist in the ribosome data. Those crosslinks may not have been observed because of low abundance or poor digestion and ionization. Therefore digestion protocols need to be developed with multiple RNases or using base hydrolysis to cleave the ribosomal RNA into management fragments and separation procedures such as titanium dioxide enrichment or ion exchange may be necessary to enrich for the RNA fragments. An alternative possibility is to analyze a smaller RNA-protein complex.

Once we have verified that protein-RNA crosslinks can be observed, the computational tools must be improved. The current computational tools to analyze RNA and DNA fragmentation data are scarce and there are not any tools to analyze proteins and RNA crosslinked by DC4. Therefore a key step is to generate RNA and DNA data that can be used as a training set for algorithm development. Then the algorithm would be altered for DC4-modified RNA and DNA. Synthesized oligonucleotide libraries could be used to accomplish this goal.

6.3 Crosslinking Combined with Ion Mobility Mass Spectrometry

The initial results from combining DC4 crosslinking and ion mobility mass spectrometry were very promising, however, further experiments are still needed. From Chapter 5, it was clear that chemically crosslinking a complex results in stabilization, both with unfolding and dissociation. However, there were large quantities of salt adducts present on the complex.

Therefore it was unclear to what extent the stabilization resulted from crosslinking versus the presence of the salt adducts. To further elucidate this, similar dissociation experiments are needed with more efficient desalting. Extensive dialysis against ammonium acetate for long periods of time may be necessary to achieve sufficient desalting. An alternative would be to use a non-cleavable commercial crosslinker such as BS3. The use of BS3 should not result in a large amount of salt electrostatically interacting with the protein, nor should it cleave in the mass spectrometer. Therefore any stabilization effects would be due to the crosslinking. Furthermore, increases in size and mass would be due to the addition of BS3. However, unlike with DC4, crosslinking with BS3 does not maintain the charge on the lysyl residues which could cause perturbations in the protein structure.

DC4 is a mass spectrometry-cleavable crosslinker while BS3 is not. Although DC4-crosslinked avidin required more energy to dissociate, dissociation into subunits was achieved albeit through different pathways. Because BS3 is not a MS-cleavable crosslinker the dissociation of subunits may require additional energy or result in backbone fragmentation prior to monomer dissociation. When peptides were fragmenting from the unmodified avidin, the most likely assignments for those peptides were from the C-terminus. CID of a DC4-crosslinked avidin tetramer produced a much richer set of peptides from locations other than C-terminus, although due to limited mass accuracy and the unavailability of MS^3 their precise identities could not be determined. The identities of those peptide masses are needed to make further conclusions regarding this data. Therefore, the bottom-up analysis of DC4-crosslinked avidin should be completed. When it is known which lysyl residues crosslink to each other, better estimates of the identities of the peptides fragmented by ion mobility can be obtained. An alternative experiment would be to try top-down analysis of DC4-crosslinked avidin on the

Orbitrap. If the same fragment masses resulted, then further MS3 could be done to identify those masses. However, the Orbitrap is not configured to retain protein structure during ionization, so the fragmentation may be different.

In Chapter 4 it was demonstrated with peptides that fragmentation within the DC4 crosslinking moiety is a more facile event than backbone cleavage. However, DC5, and DC6 exhibit less fragmentation within the crosslinker. Therefore, the fragmentation patterns and peptides observed with DC5 and DC6 data may also lend insights into the DC4 crosslinked data. I would expect to see fewer peptides with the DC5- and DC6-crosslinked avidin than with the DC4.

6.4 References

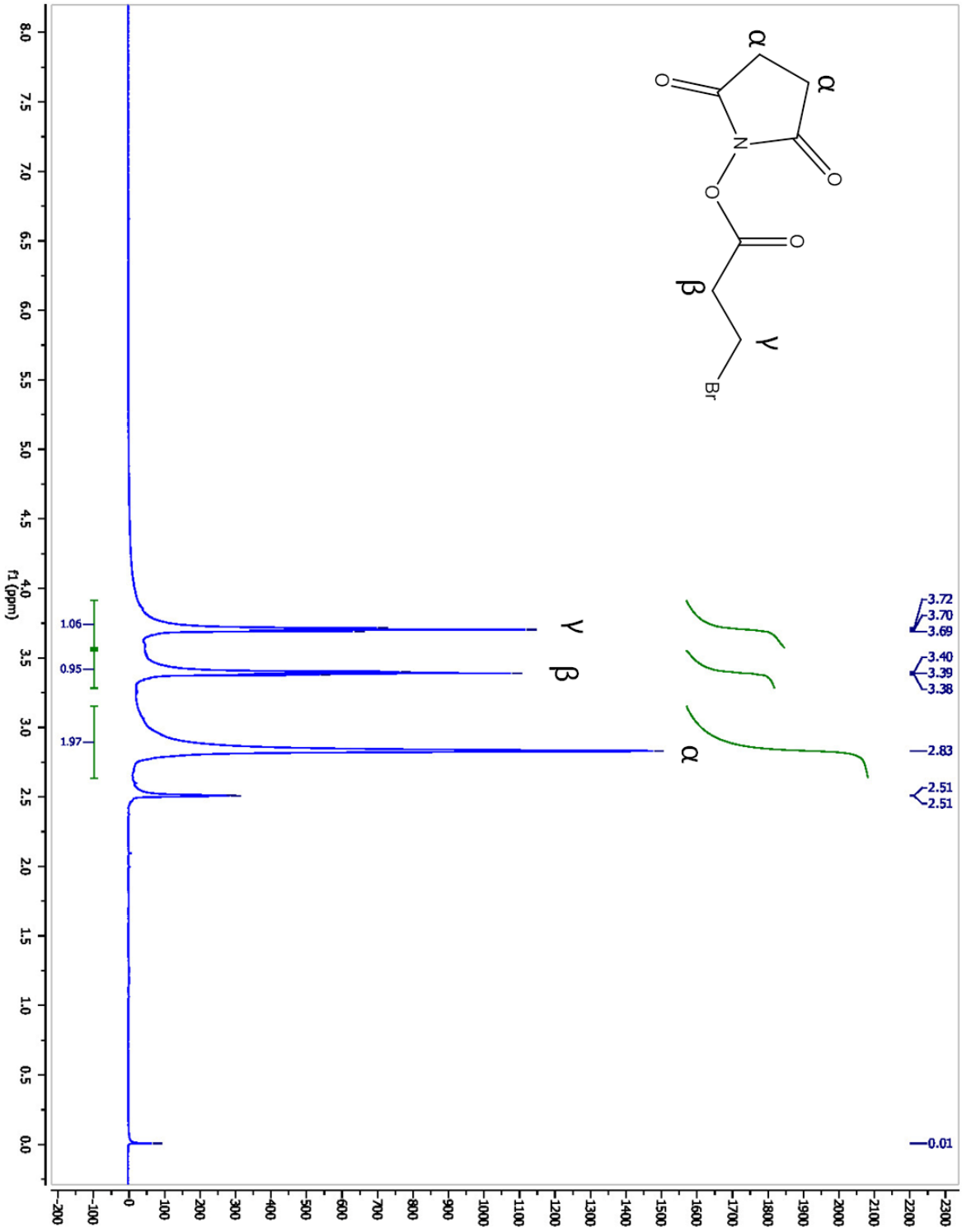
[1] B. Alberts, *Cell* **1998**, *92*, 291.
[2] S. Jennebach, F. Herzog, R. Aebersold, P. Cramer, *Nucleic Acids Research* **2012**, *40*, 5591.
[3] Z. A. Chen, A. Jawhari, L. Fischer, C. Buchen, S. Tahir, T. Kamenski, M. Rasmussen, L. Lariviere, J.-C. Bukowski-Wills, M. Nilges, P. Cramer, J. Rappsilber, *The EMBO Journal* **2010**, *29*, 717.
[4] M. A. Lauber, J. P. Reilly, *Journal of Proteome Research* **2011**, *10*, 3604.
[5] M. A. Lauber, J. Rappsilber, J. P. Reilly, *Molecular & cellular proteomics : MCP* **2012**, *11*, 1965.
[6] E. G. Jaffee, M. A. Lauber, W. E. Running, J. P. Reilly, *Analytical Chemistry* **2012**, *84*, 9355.
[7] J. Freed, J. Smith, P. Li, A. Greene, *Proteomics* **2007**, *7*, 2371.
[8] X. Tang, W. Yi, G. R. Munske, D. P. Adhikari, N. L. Zakharova, J. E. Bruce, *Journal of Proteome Research* **2006**, *6*, 724.
[9] J. E. Bruce, *Proteomics* **2012**, *12*, 1565.
[10] L. Yang, C. Zheng, C. R. Weisbrod, X. Tang, G. R. Munske, M. R. Hoopmann, J. K. Eng, J. E. Bruce, *Journal of Proteome Research* **2011**, *11*, 1027.
[11] C. Zheng, L. Yang, M. R. Hoopmann, J. K. Eng, X. Tang, C. R. Weisbrod, J. E. Bruce, *Molecular & Cellular Proteomics* **2011**, *10*.
[12] J. W. Back, A. F. Hartog, H. L. Dekker, A. O. Muijsers, L. J. d. Koning, L. d. Jong, *Journal of the American Society for Mass Spectrometry* **2001**, *12*, 222.
[13] F. Dreiocker, M. Q. Muller, A. Sinz, M. Schafer, *Journal of Mass Spectrometry* **2009**, *45*, 178.
[14] N. Fujii, R. B. Jacobsen, N. L. Wood, J. S. Schoeniger, R. K. Guy, *Bioorganic & Medicinal Chemistry Letters* **2004**, *14*, 427.
[15] E. V. Petrotchenko, V. K. Olkhovik, C. H. Borchers, *Molecular & Cellular Proteomics* **2005**, *4*, 1167.
[16] E. V. Petrotchenko, J. J. Serpa, C. H. Borchers, *Molecular & Cellular Proteomics* **2011**, *10*.
[17] E. V. Petrotchenko, K. Xiao, J. Cable, Y. Chen, N. V. Dokholyan, C. H. Borchers, *Molecular & Cellular Proteomics* **2009**, *8*, 273.

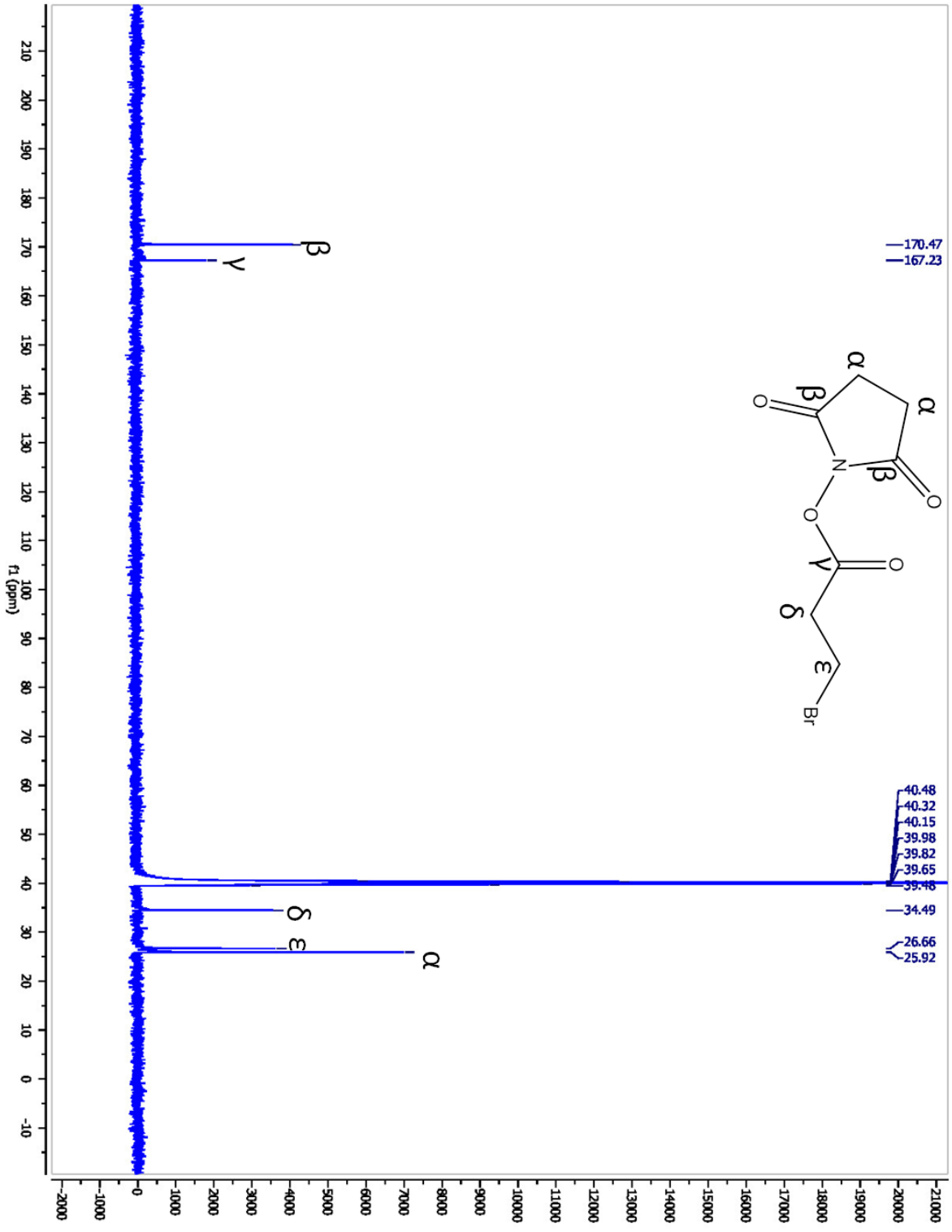
[18] X. Tang, G. R. Munske, W. Siems, J. E. Bruce, *Analytical Chemistry* **2005**, *77*, 311.
[19] A. Sinz, S. Kalkhof, C. H. Ihling, *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 1921.
[20] M. Q. Muller, F. Dreiocker, C. H. Ihling, M. Schafer, A. Sinz, *Analytical Chemistry* **2010**, *45*, 880.
[21] M. A. Lauber, J. P. Reilly, *Analytical Chemistry* **2010**, *82*, 7736.
[22] M. Obwald, B. Greuer, R. Brimacombe, *Nucleic Acids Research* **1990**, *18*, 6755.
[23] S. Kalkhof, A. Sinz, *Analytical Biochemistry* **2008**, *392*, 305.
[24] S. Madler, C. Bich, D. Touboul, R. Zenobi, *Journal of Mass Spectrometry* **2009**, *44*, 694.

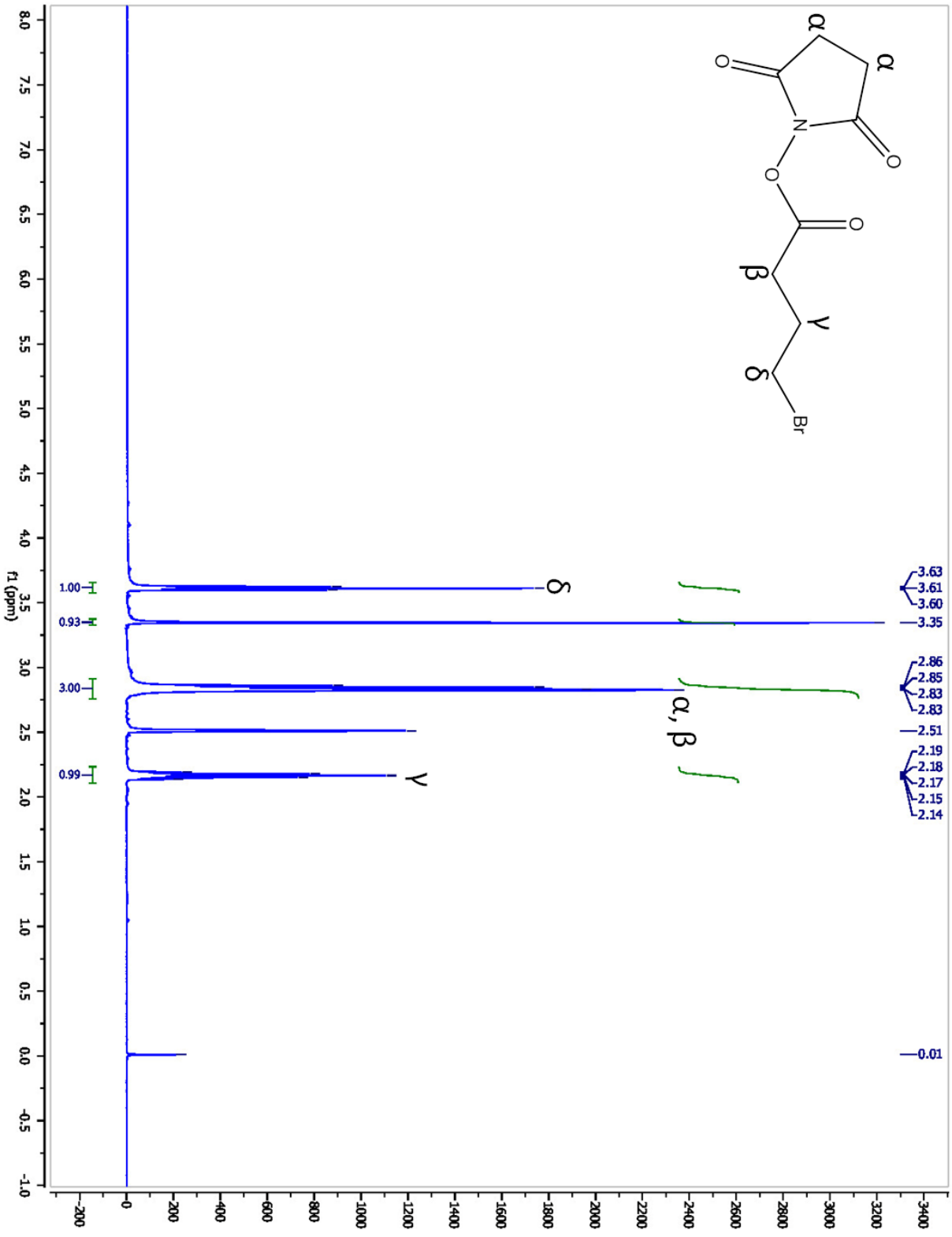
Appendix A

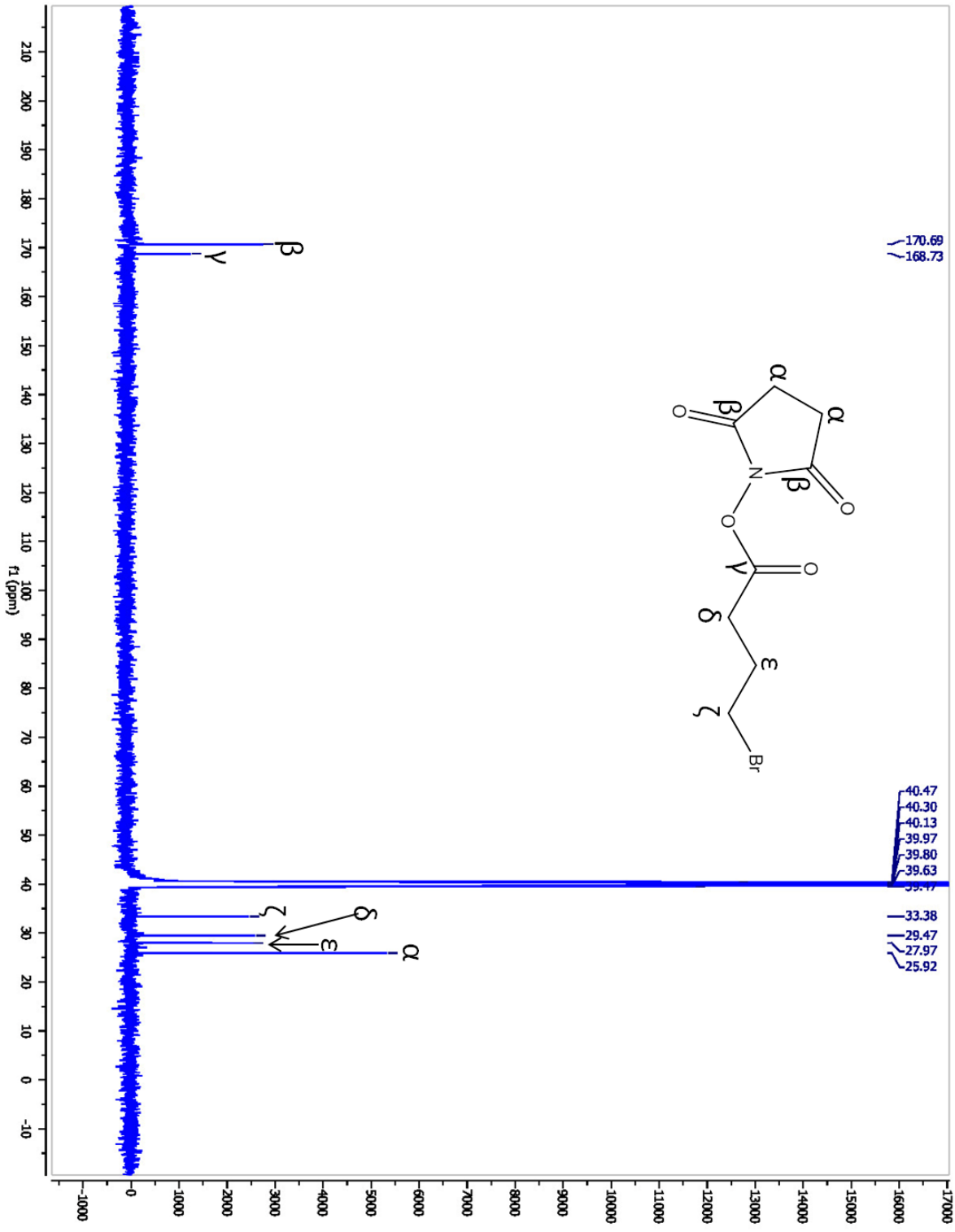
Spectral and Other Characterization Data for Chapter 3

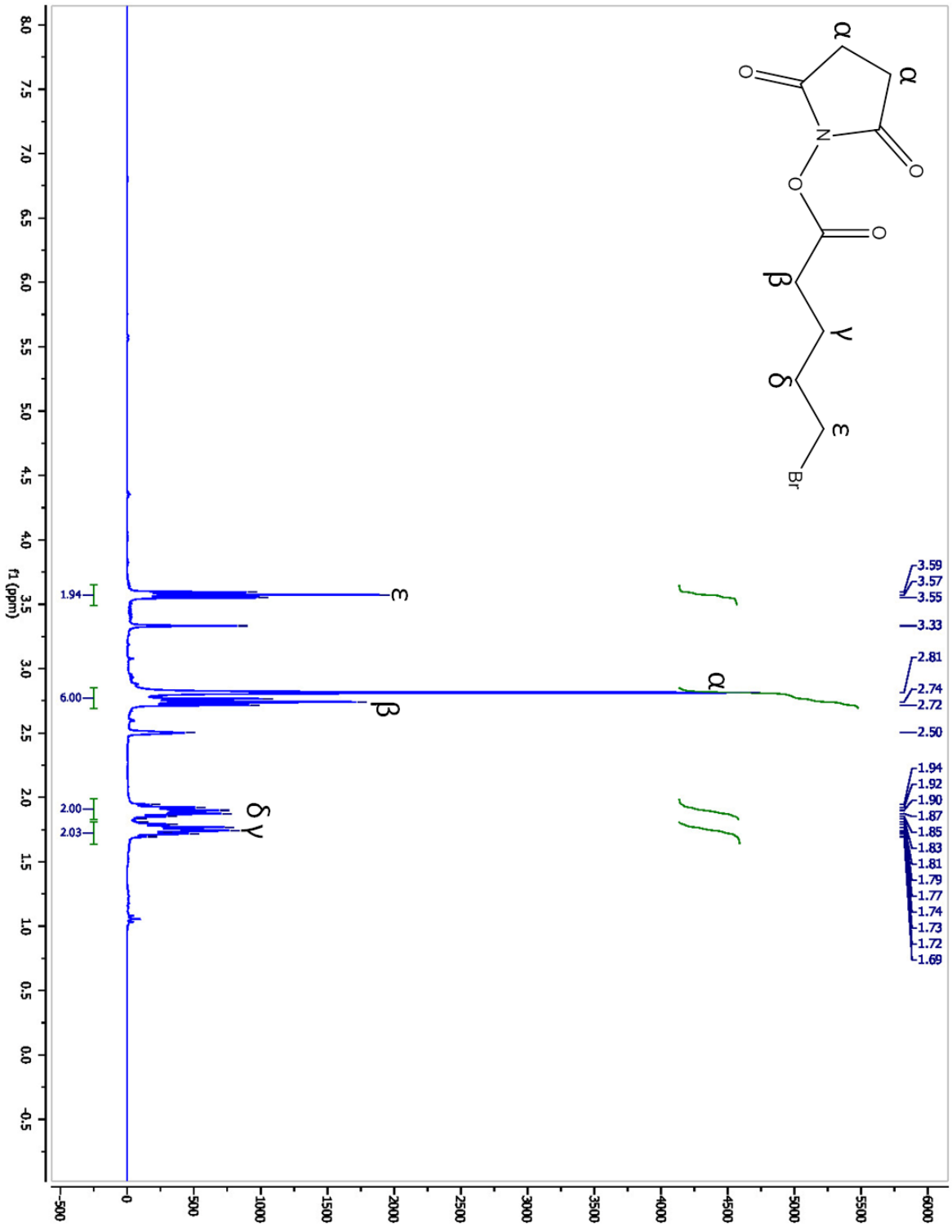
The NMR spectra were recorded on a Bruker instrument at 500 MHz for ^{1}H and 125 MHz for ^{13}C spectra. Chemical shift values are recorded in δ units (ppm). Mass spectra were recorded on a Micromass LCT Time-of-Flight mass spectrometer with an electrospray ionization source.

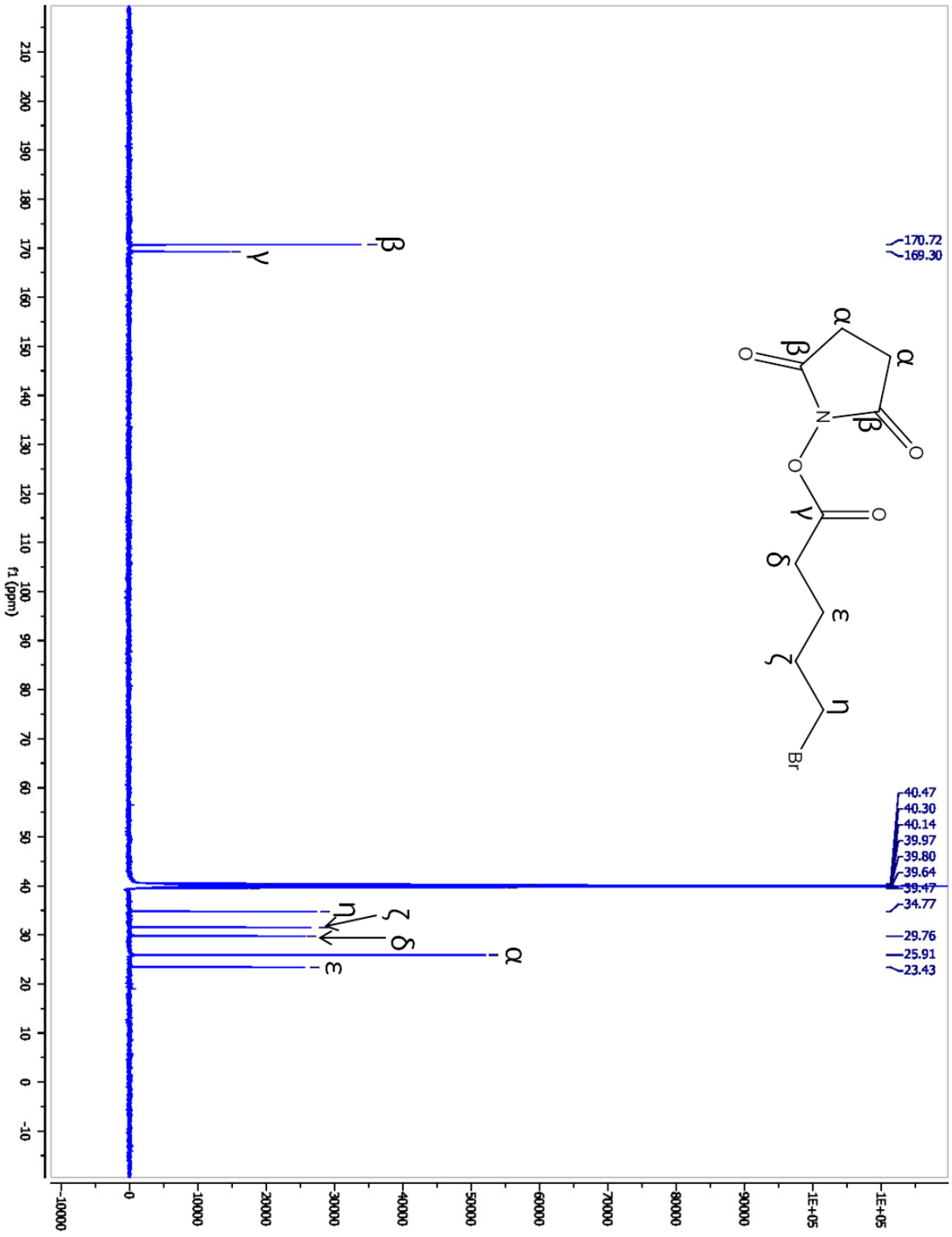

α
α
O
N
O
O
O
β
γ
Br
γ
β
α
1.06
0.95
1.97
3.72
3.70
3.69
3.40
3.39
3.38
2.83
2.51
2.51
0.01
f1 (ppm)

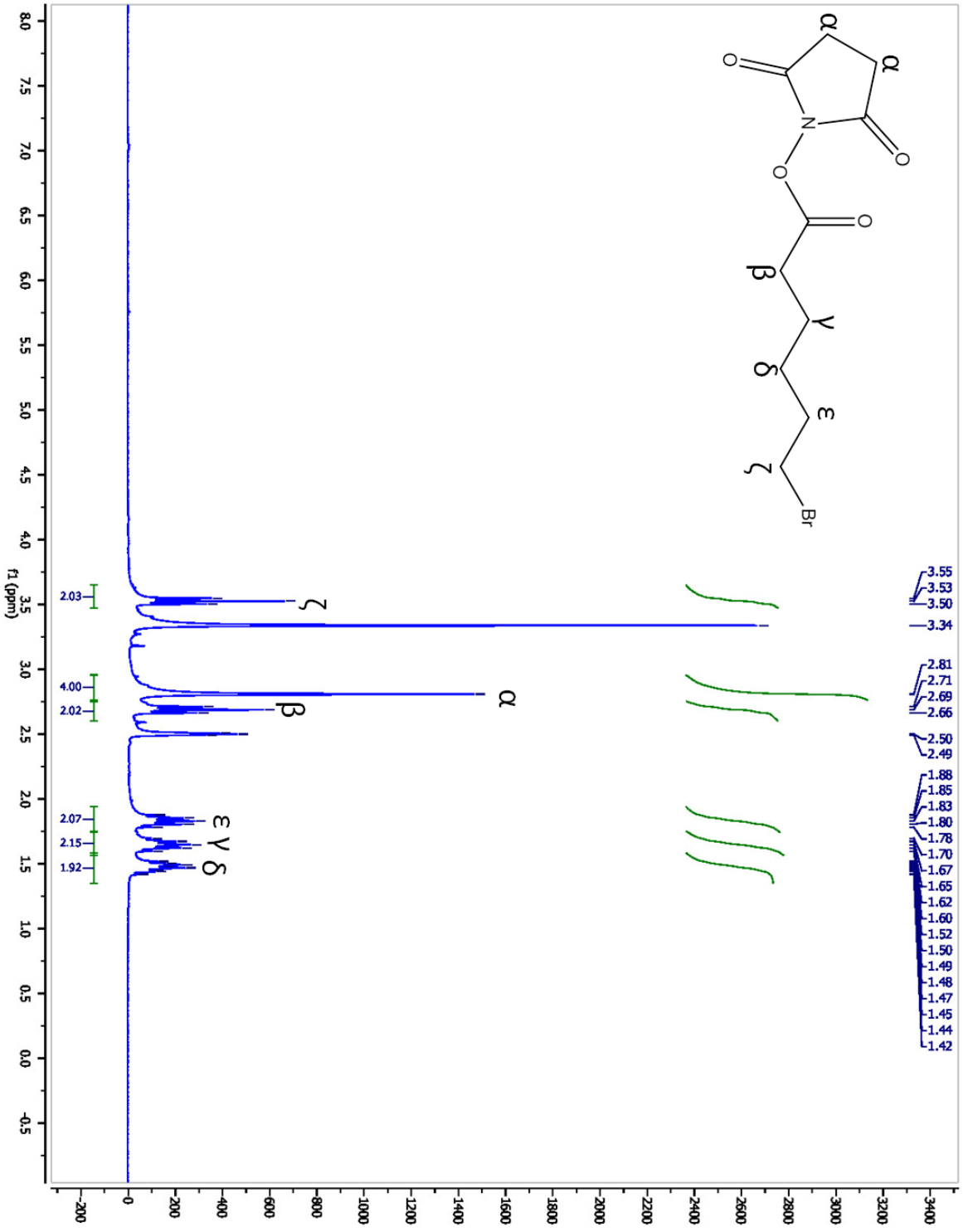

170.47
167.23
40.48
40.32
40.15
39.98
39.82
39.65
34.49
26.66
25.92
β
γ
δ
ε
α
f1 (ppm)

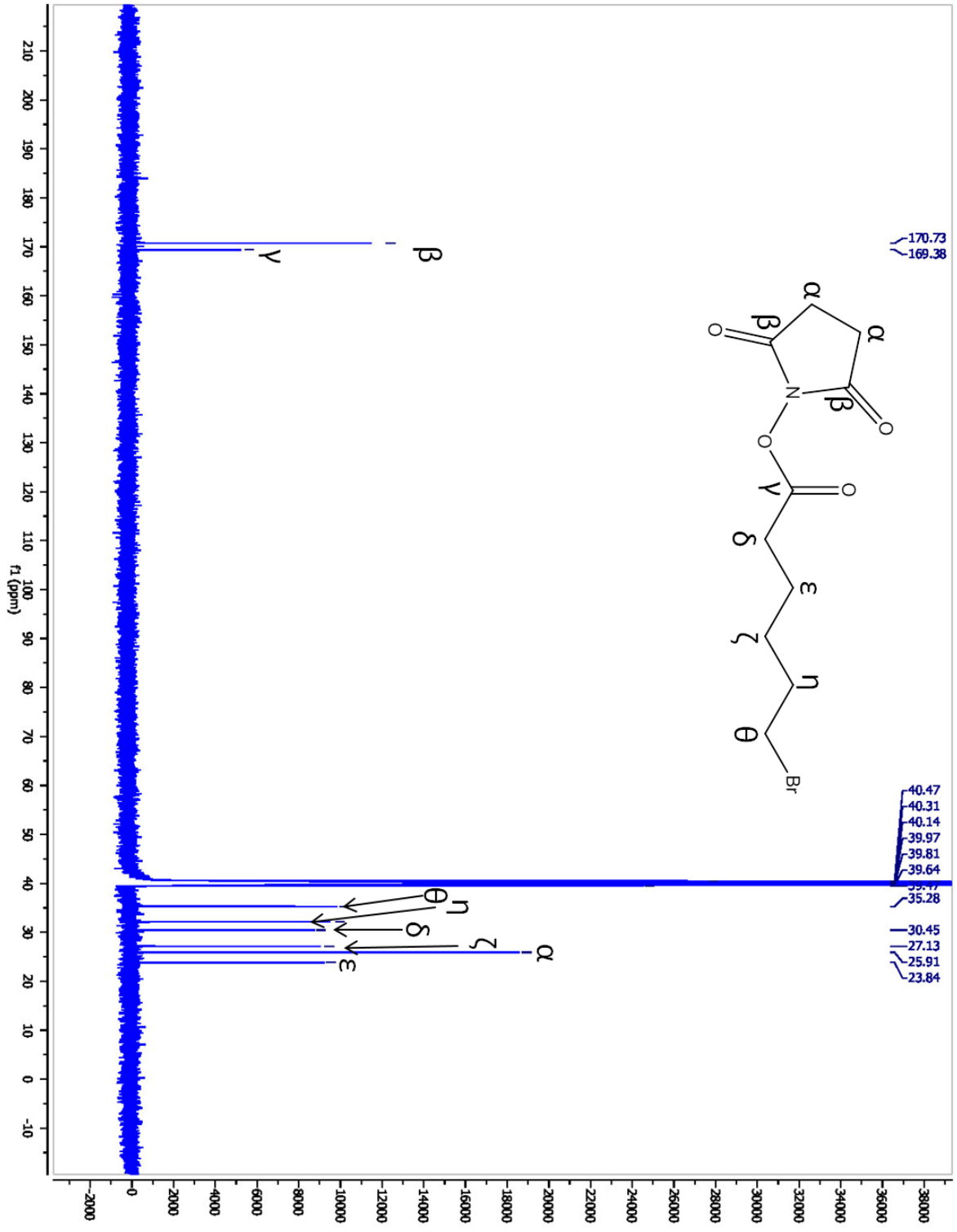

α
α
O
N
O
O
O
β
γ
δ
Br
δ
α, β
γ
1.00
0.93
3.00
0.99
3.63
3.61
3.60
3.35
2.86
2.85
2.83
2.83
2.51
2.19
2.18
2.17
2.15
2.14
0.01
f1 (ppm)

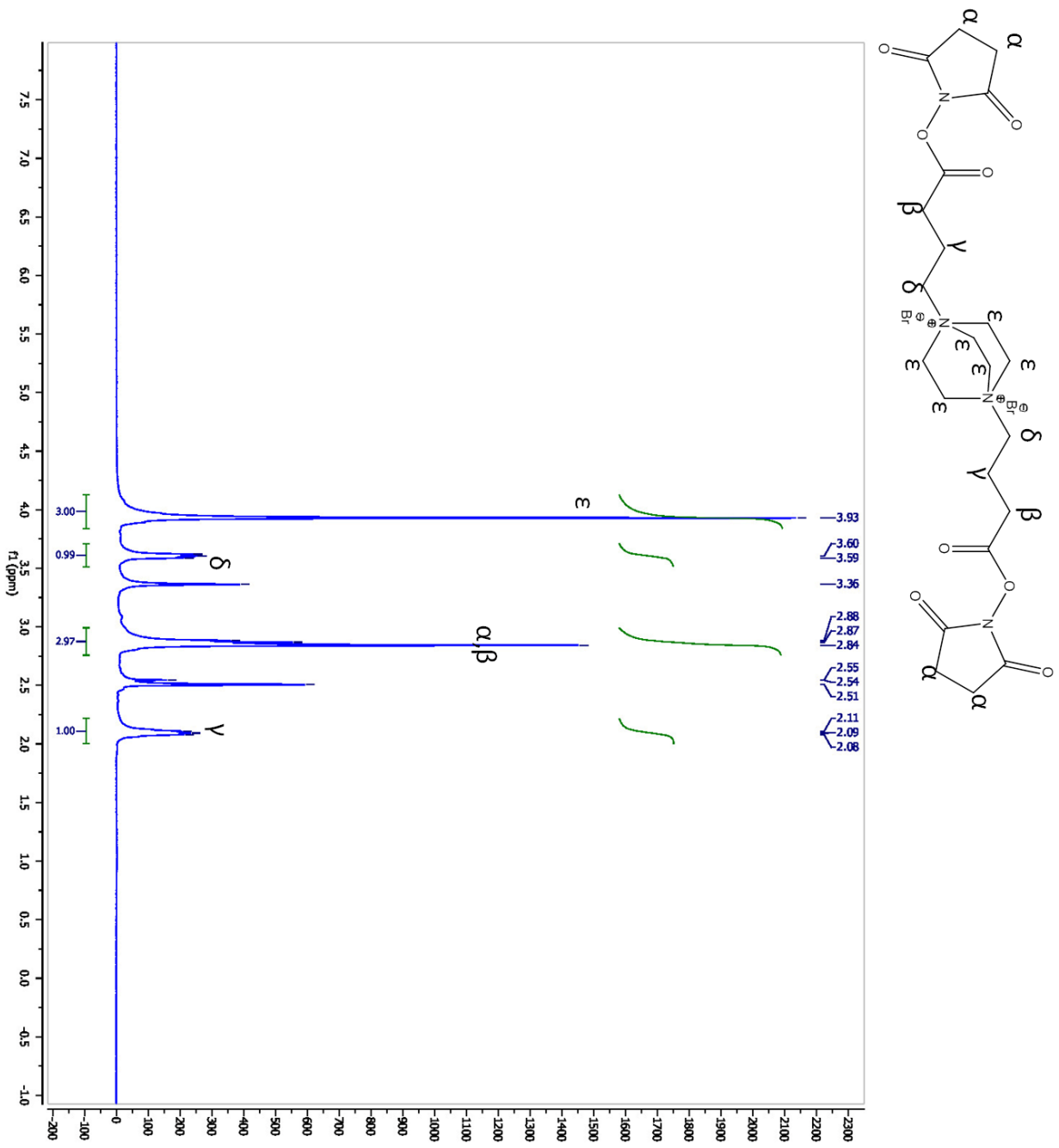

170.69
168.73
40.47
40.30
40.13
39.97
39.80
39.63
33.38
29.47
27.97
25.92
β
γ
ζ
δ
ε
α
f1 (ppm)
Br

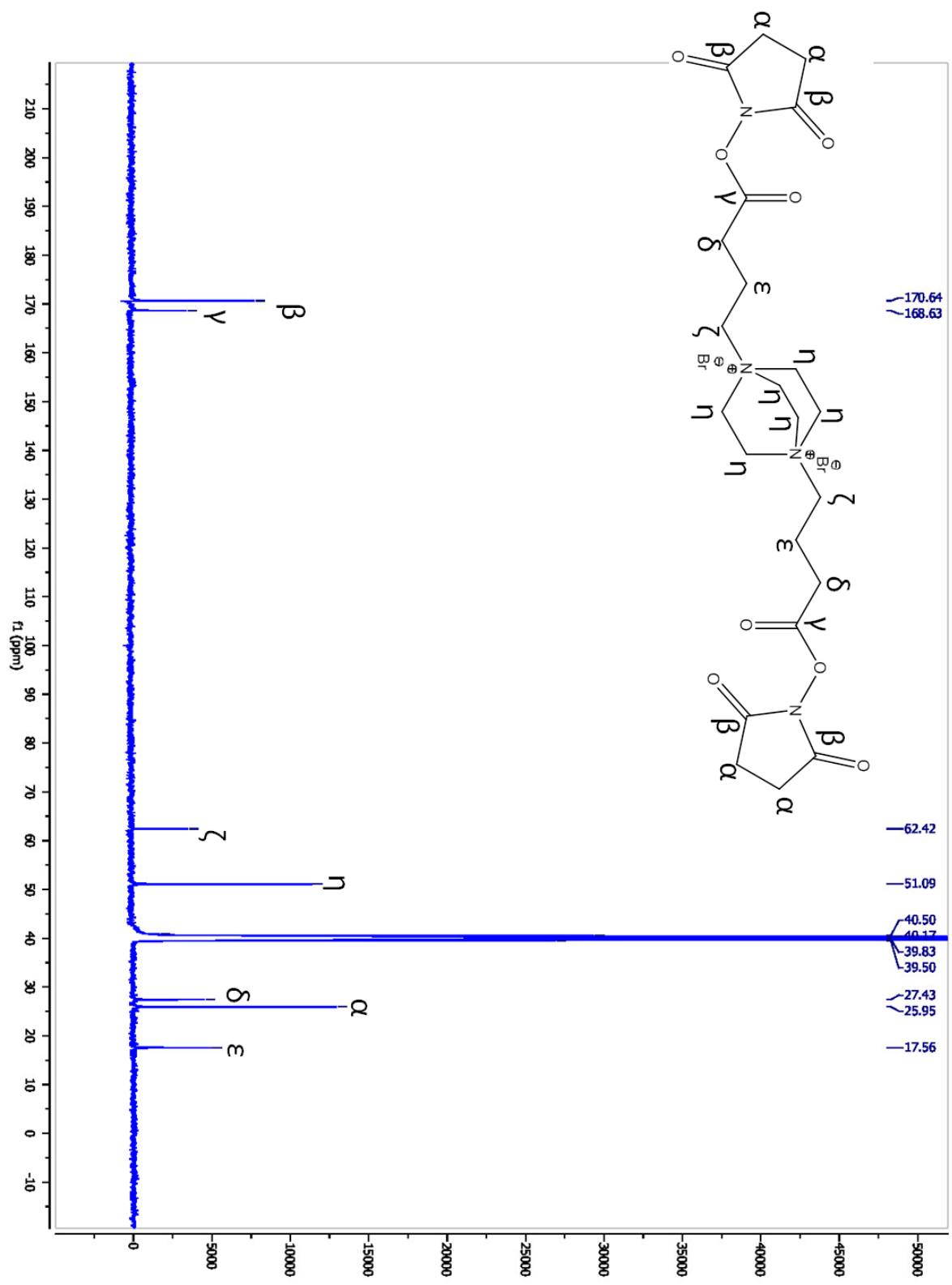

f1 (ppm)
1.94
6.00
2.00
2.03
ε
β
δ γ
α
3.59
3.57
3.55
3.33
2.81
2.74
2.72
2.50
1.94
1.92
1.90
1.87
1.85
1.83
1.81
1.79
1.77
1.74
1.73
1.72
1.69
Br

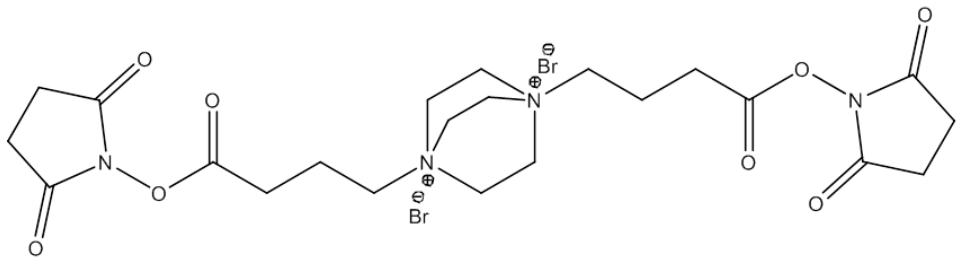

170.72
169.30
40.47
40.30
40.14
39.97
39.80
39.64
39.47
34.77
29.76
25.91
23.43
β
γ
α
β
δ
ε
ζ
η
α
ε
ζ
δ
η
Br
f1 (ppm)

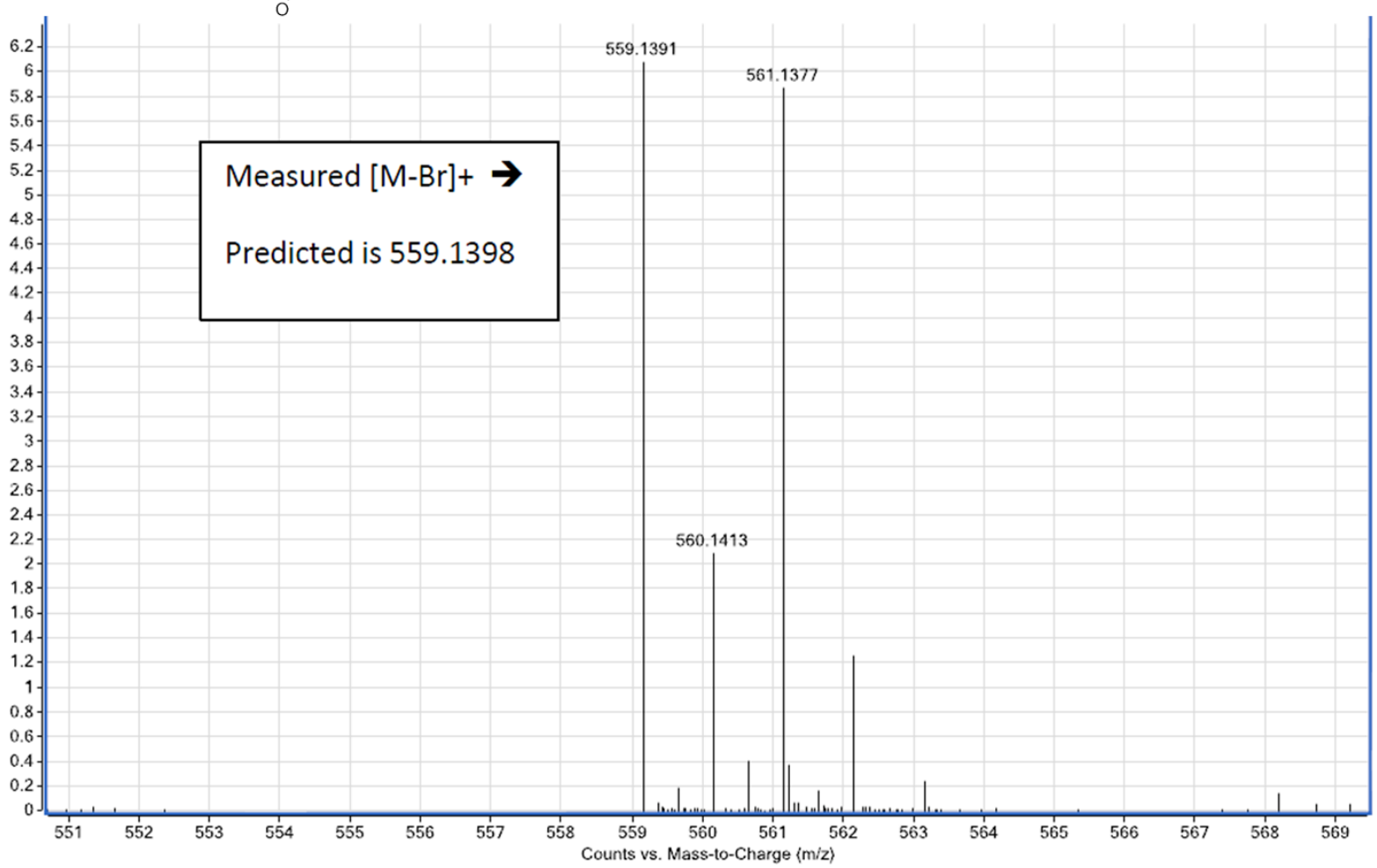

f1 (ppm)
2.03
4.00
2.02
2.07
2.15
1.92
ζ
α
β
ε γ δ
3.55
3.53
3.50
3.34
2.81
2.71
2.69
2.66
2.50
2.49
1.88
1.85
1.83
1.80
1.78
1.70
1.67
1.65
1.62
1.60
1.52
1.50
1.49
1.48
1.47
1.45
1.44
1.42
O
N
O
O
O
O
Br

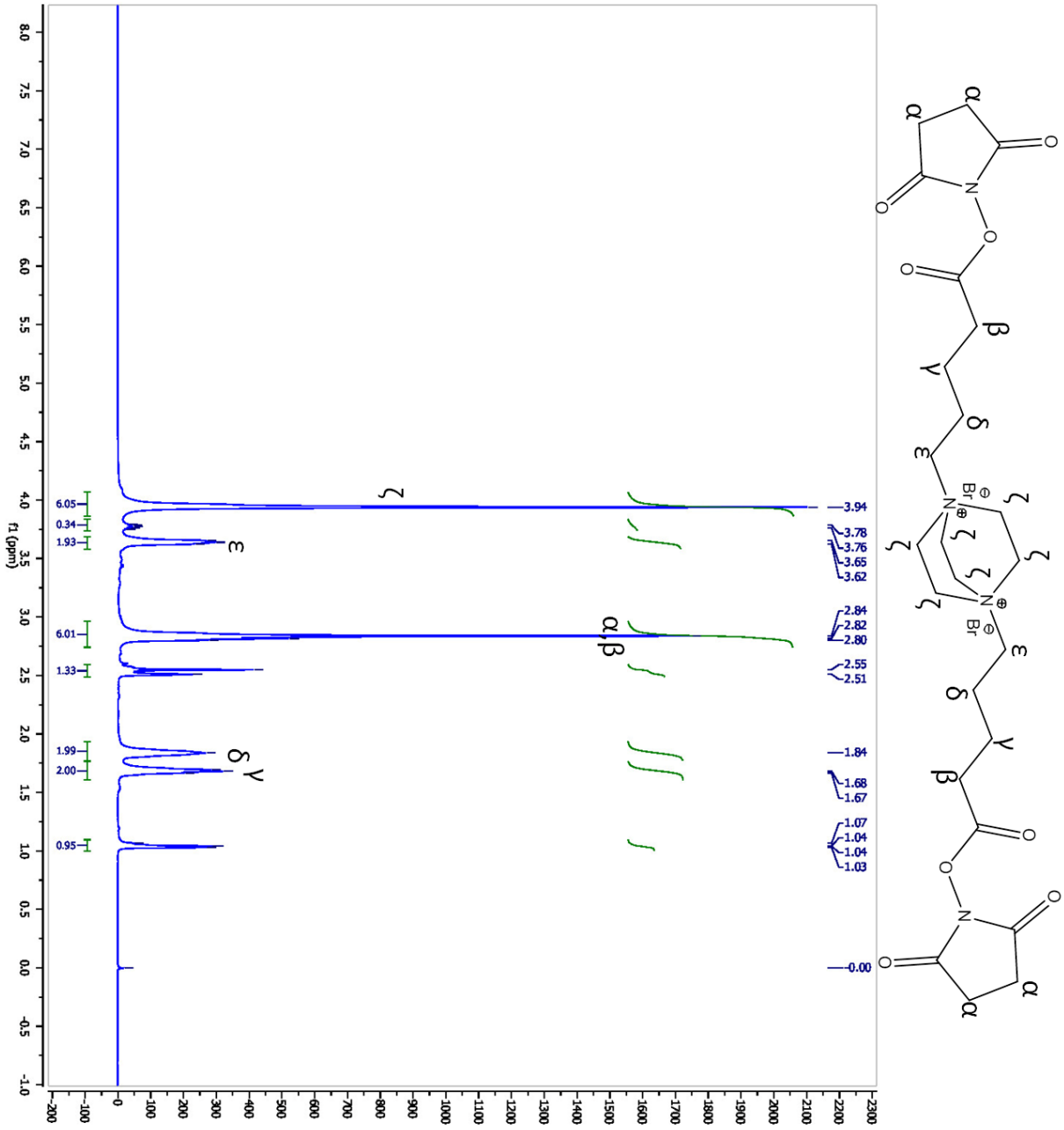

170.73
169.38
40.47
40.31
40.14
39.97
39.81
39.64
35.28
30.45
27.13
25.91
23.84
α
β
γ
δ
ε
ζ
η
θ
Br
f1 (ppm)

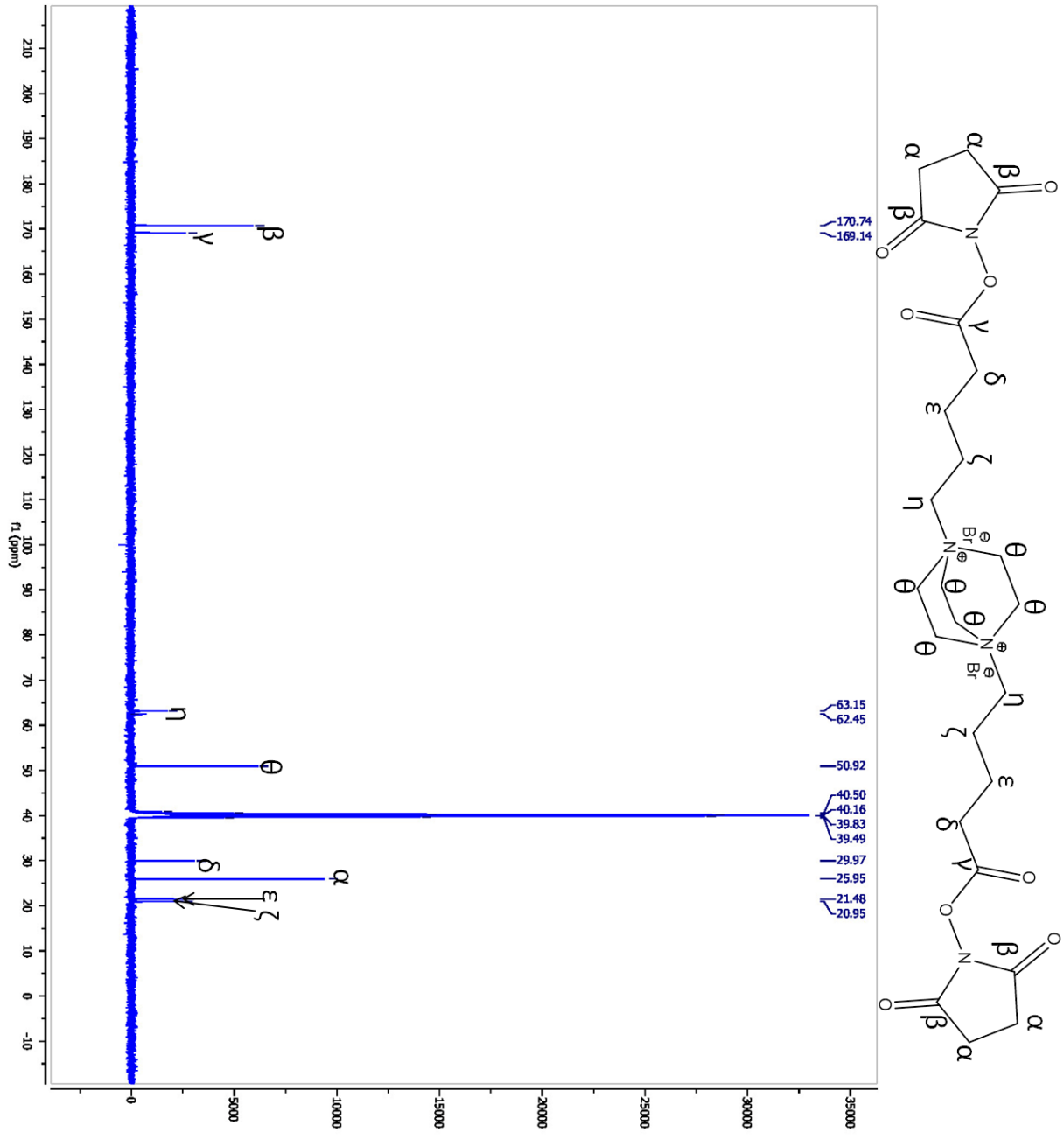

f1 (ppm)
3.00
0.99
2.97
1.00
ε
δ
α,β
γ
3.93
3.60
3.59
3.36
2.88
2.87
2.84
2.55
2.54
2.51
2.11
2.09
2.08

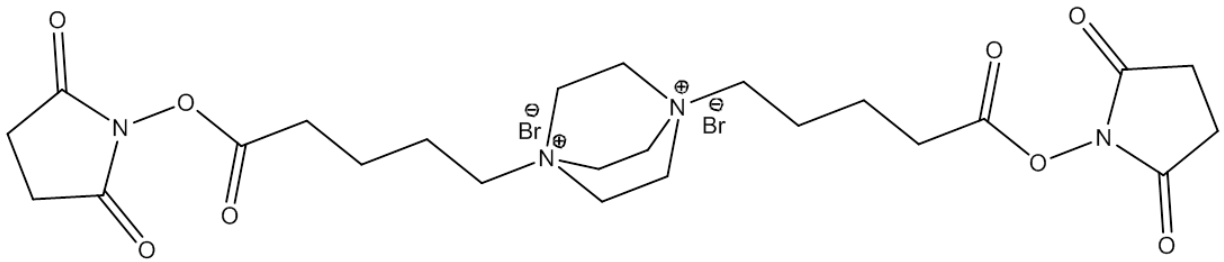

f1 (ppm)
170.64
168.63
62.42
51.09
40.50
40.17
39.83
39.50
27.43
25.95
17.56

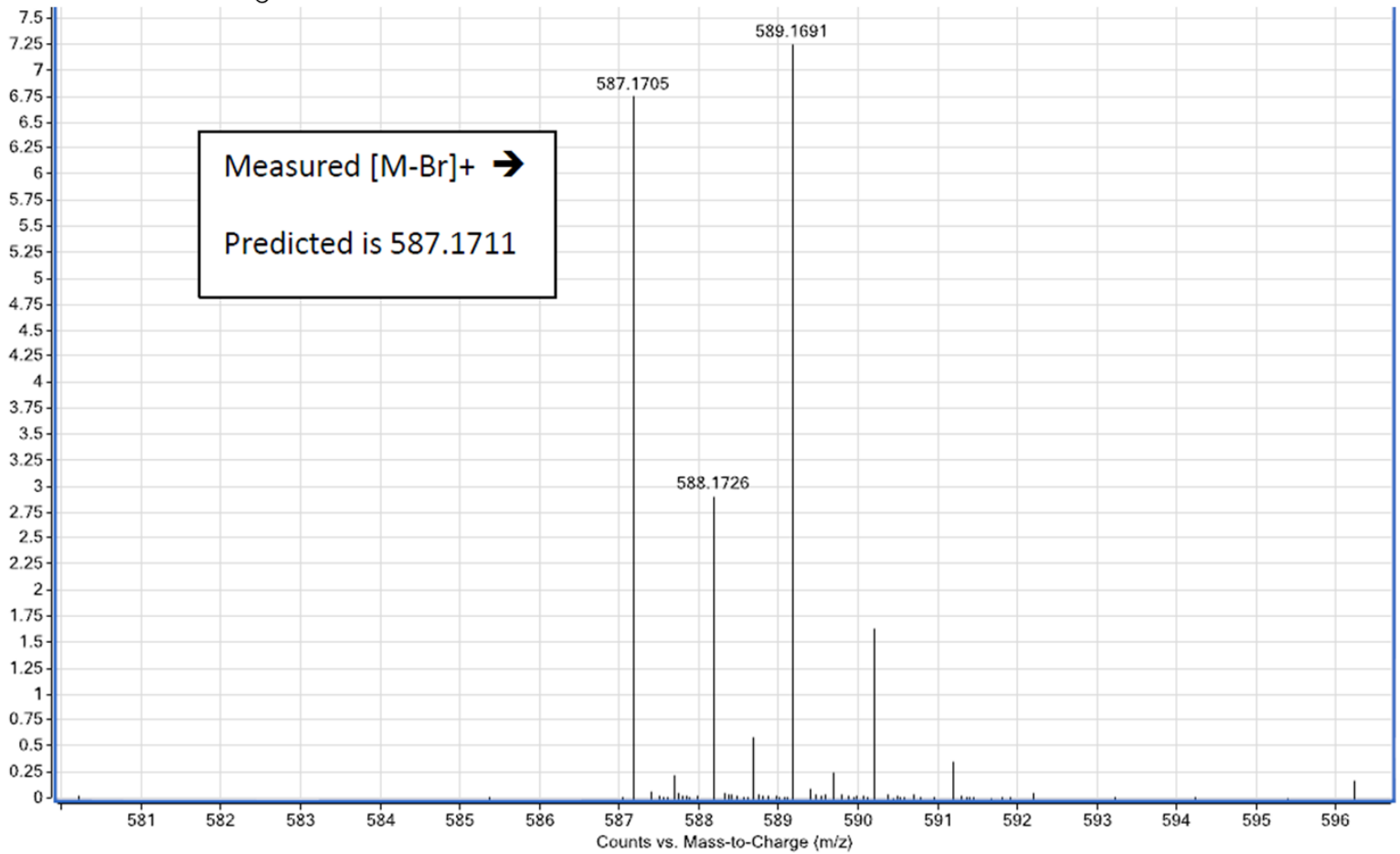

O
O
N
O
O
O
Br
N
N
Br
O
O
O
N
O
O

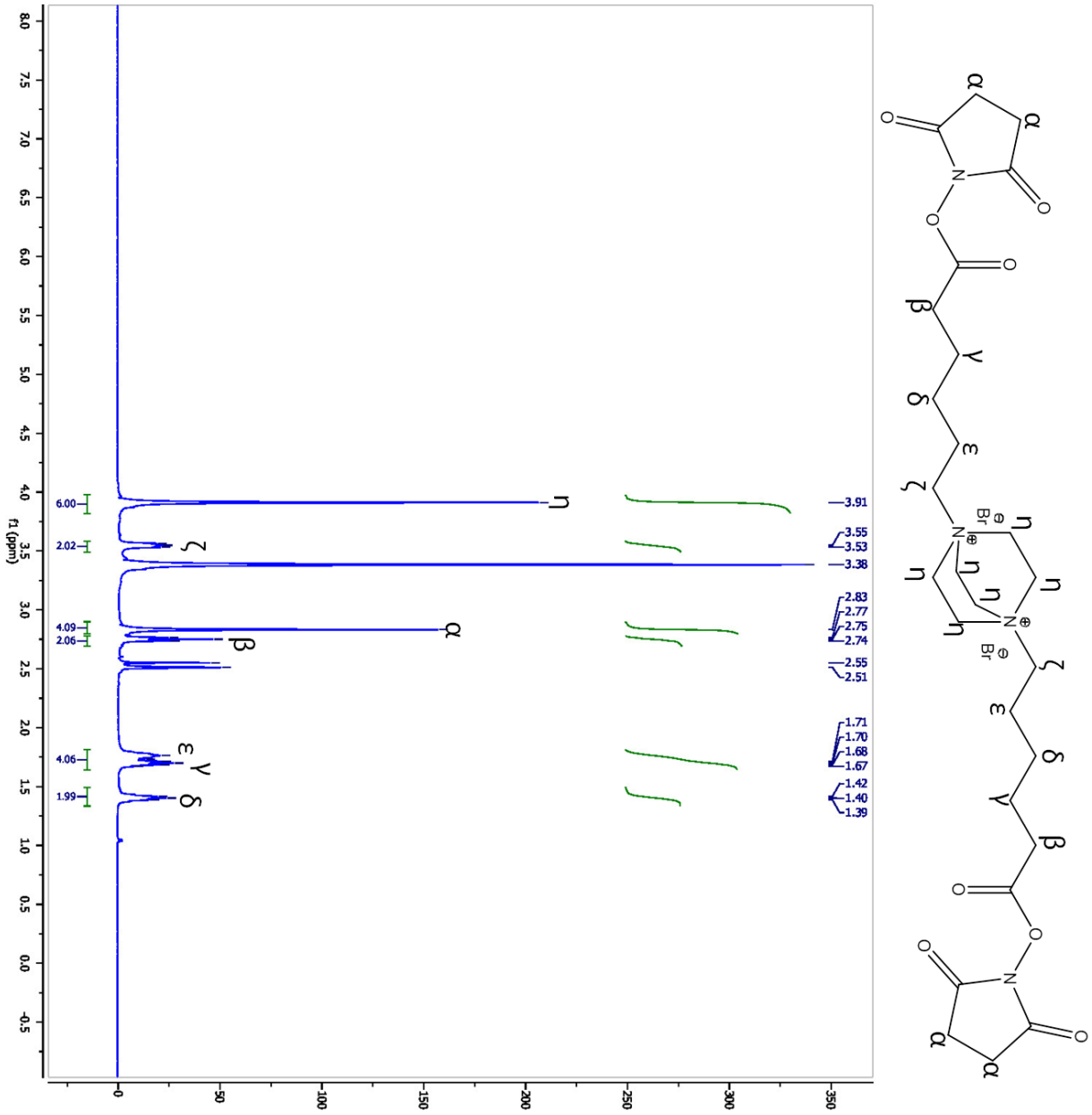

559.1391
561.1377
560.1413
Measured [M-Br]+ ➔
Predicted is 559.1398
Counts vs. Mass-to-Charge (m/z)

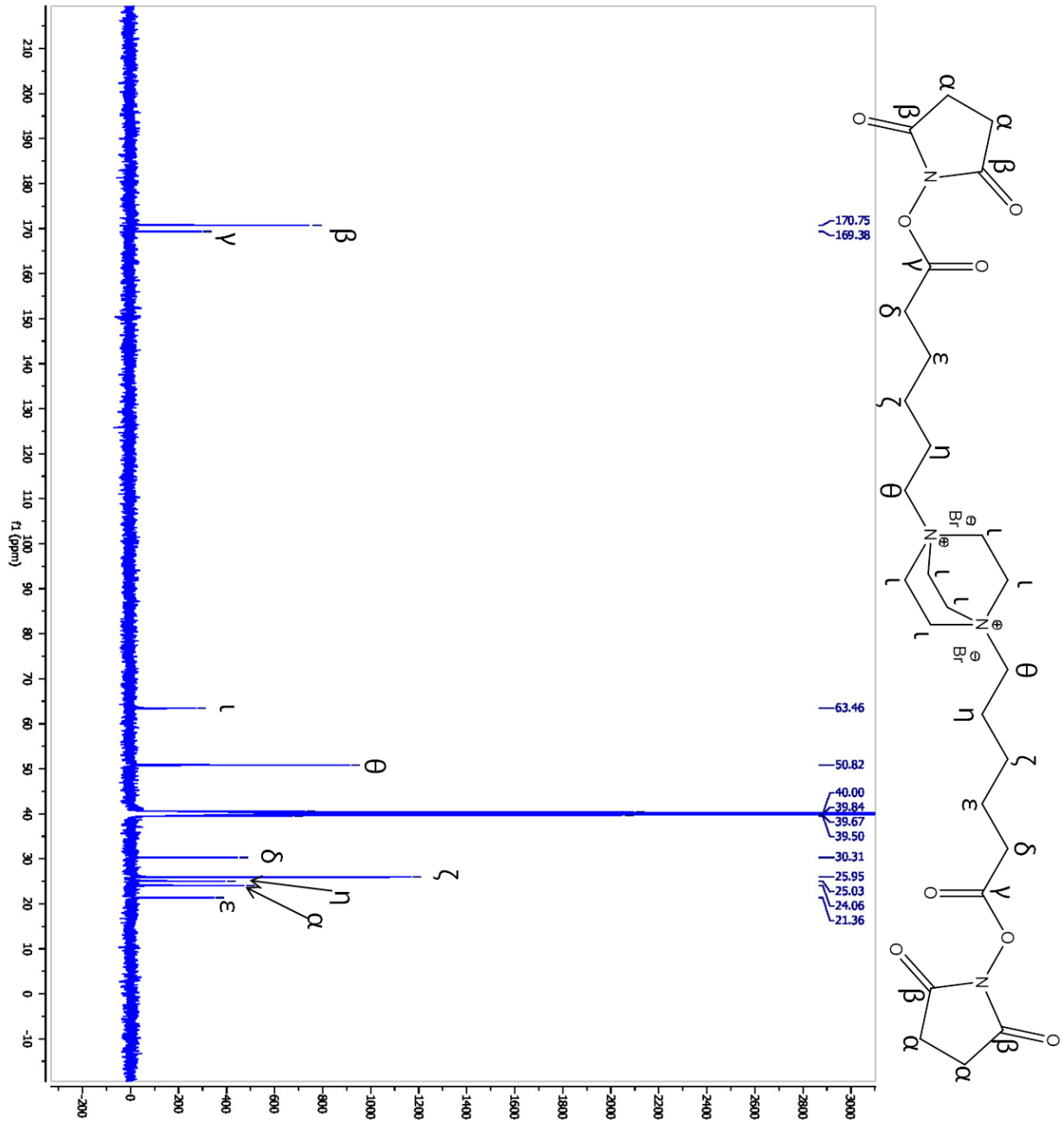

f1 (ppm)
6.05
0.34
1.93
6.01
1.33
1.99
2.00
0.95
3.94
3.78
3.76
3.65
3.62
2.84
2.82
2.80
2.55
2.51
1.84
1.68
1.67
1.07
1.04
1.04
1.03
0.00
ζ
3
α,β
γ,δ
α
β
γ
δ
ε
ζ
Br

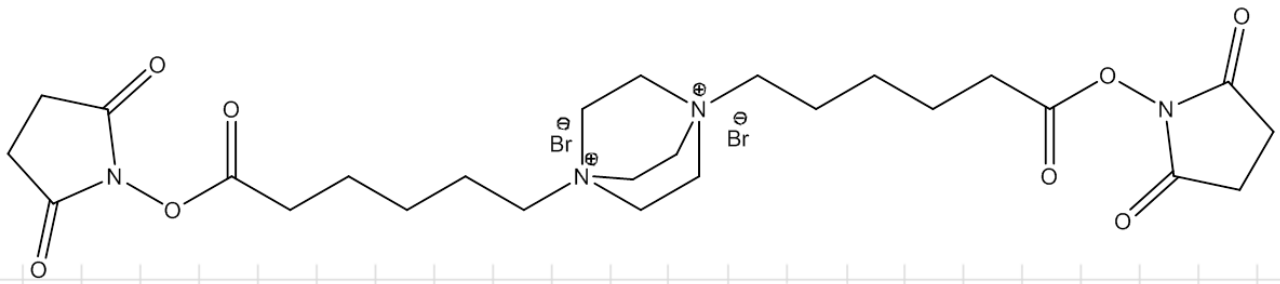

170.74
169.14
63.15
62.45
50.92
40.50
40.16
39.83
39.49
29.97
25.95
21.48
20.95
f1 (ppm)

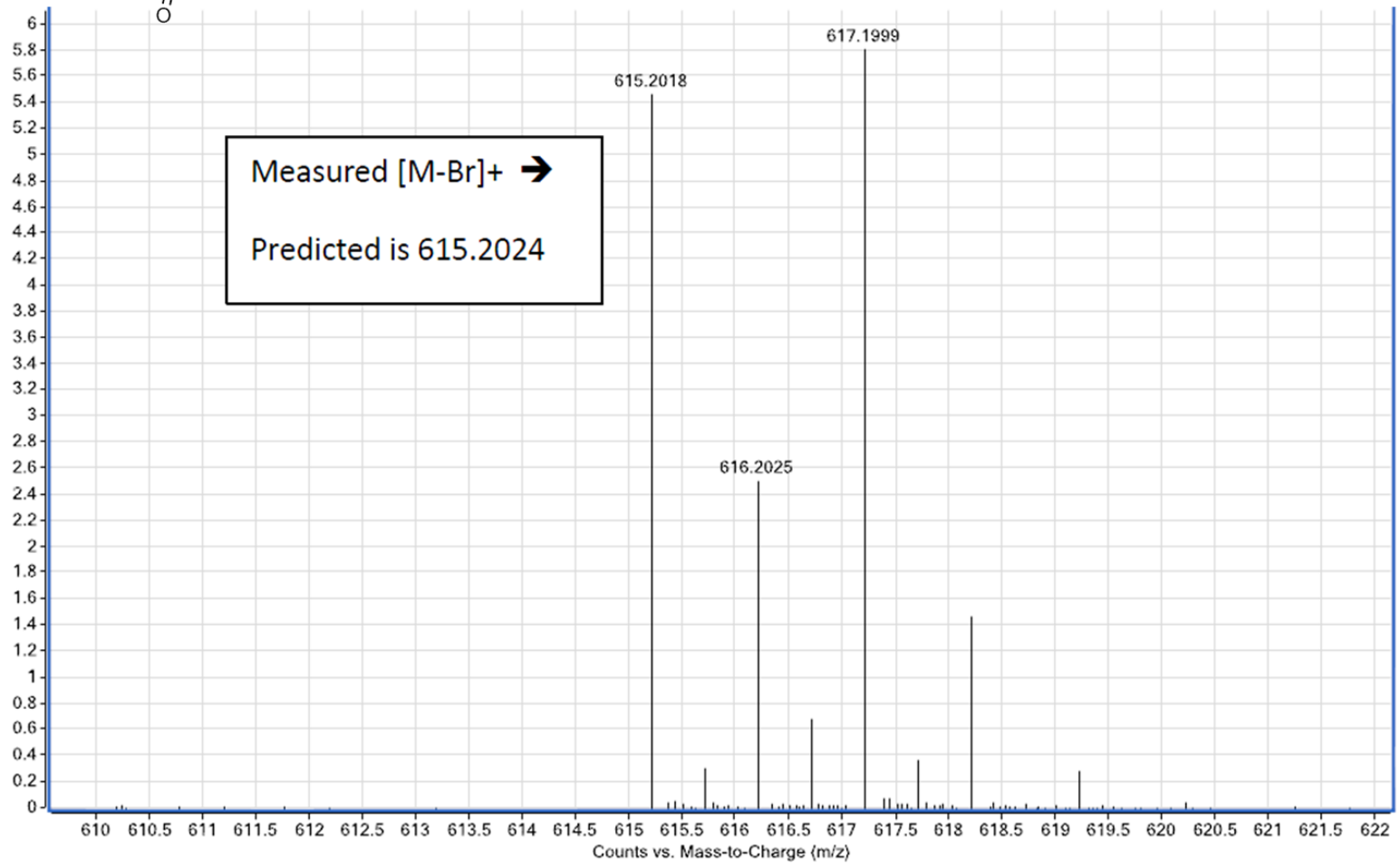

O
N
O
O
O
Br
N
N
Br
O
O
N
O
O

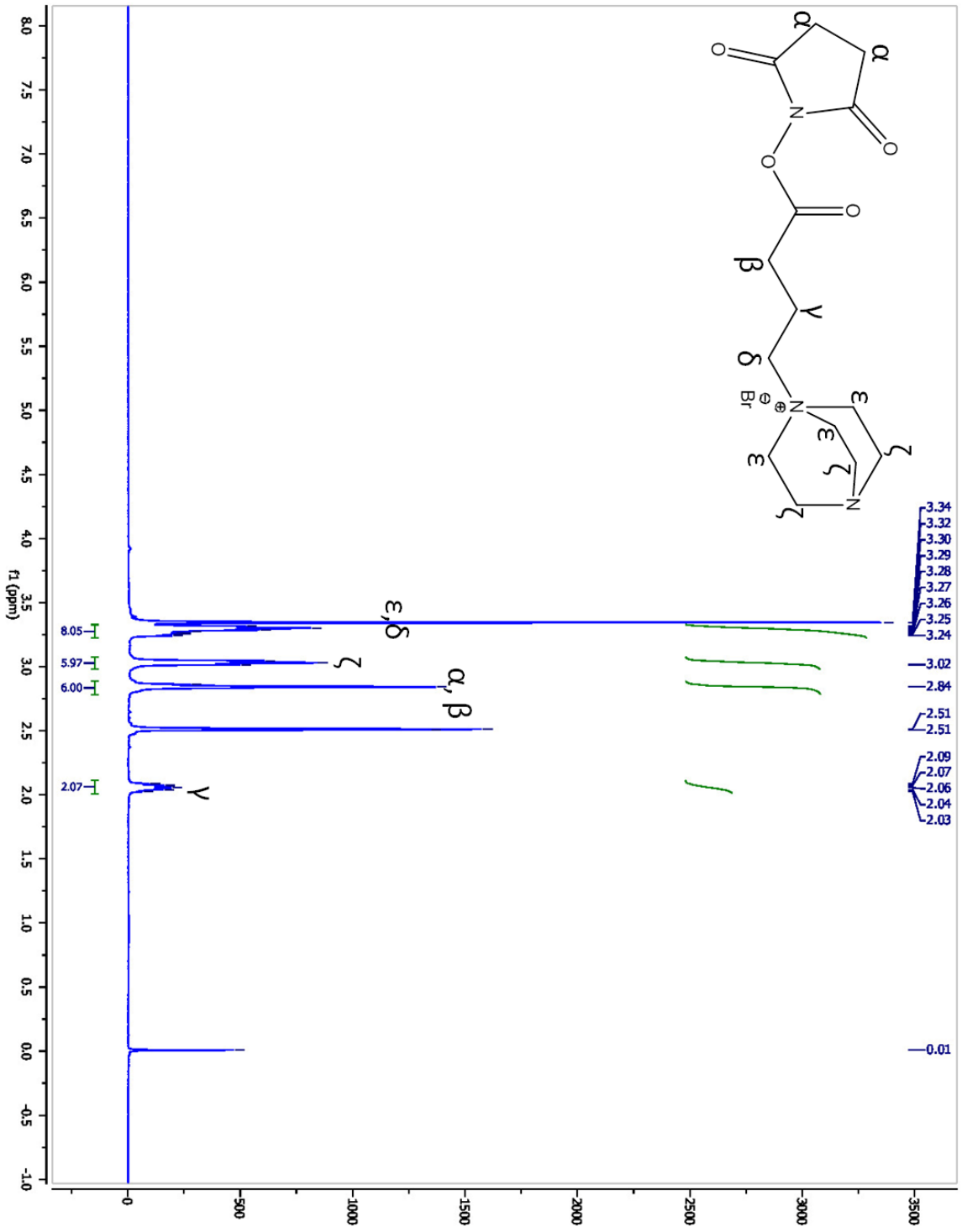

589.1691
587.1705
588.1726
Measured [M-Br]+ ➔
Predicted is 587.1711
Counts vs. Mass-to-Charge (m/z)

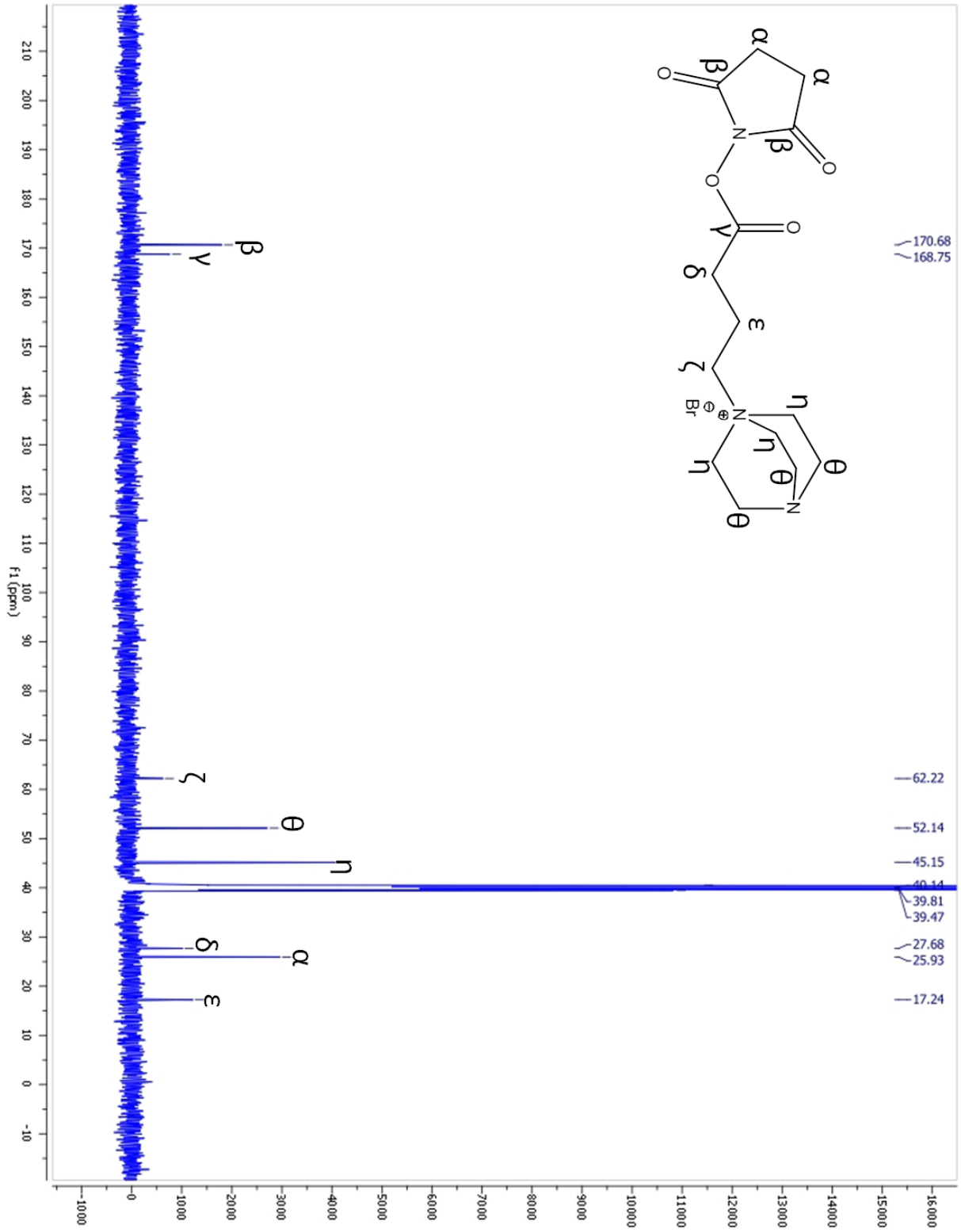

f1 (ppm)
6.00
2.02
4.09
2.06
4.06
1.99
3.91
3.55
3.53
3.38
2.83
2.77
2.75
2.74
2.55
2.51
1.71
1.70
1.68
1.67
1.42
1.40
1.39
η
ζ
α
β
ε
γ
δ
Br

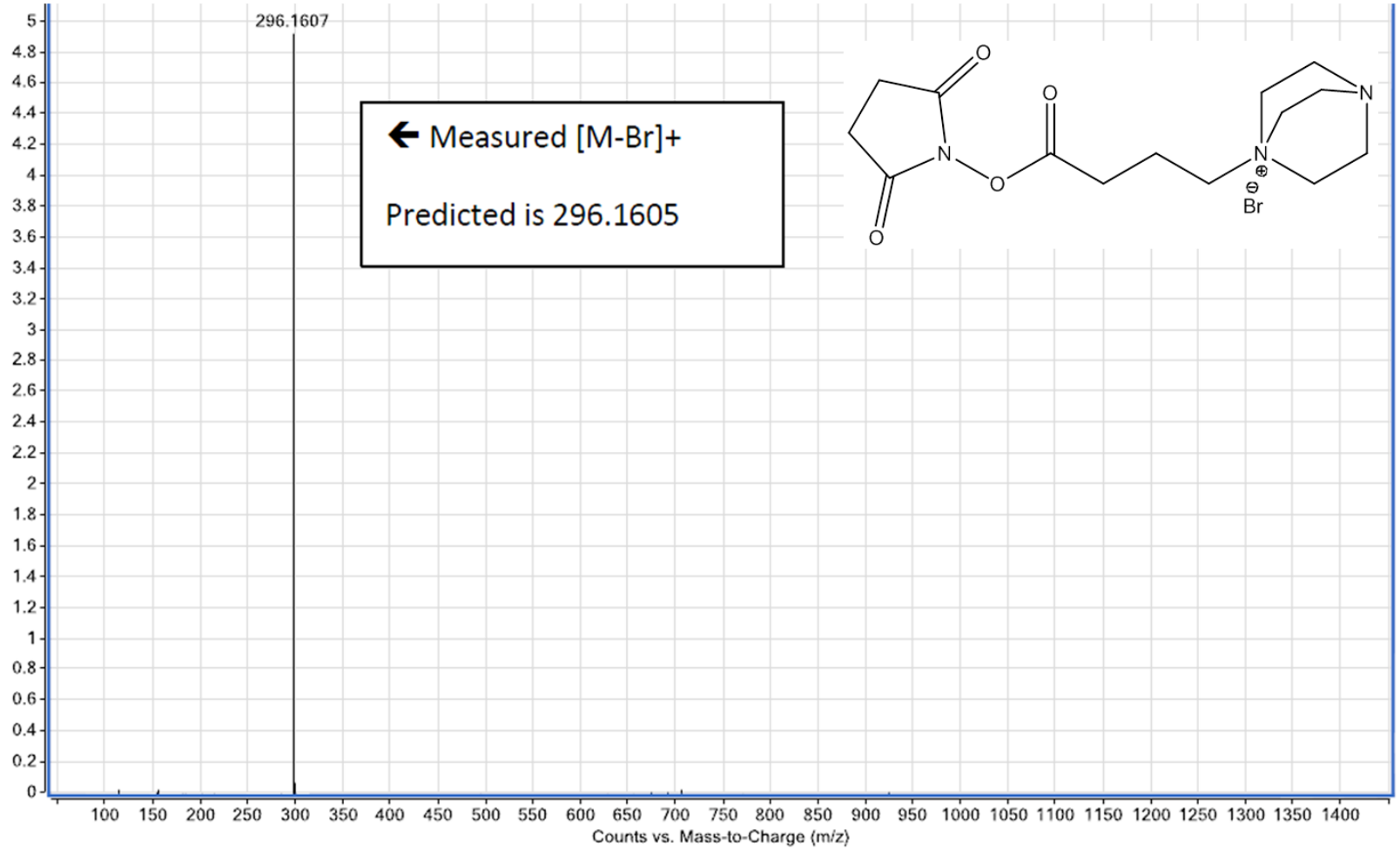

170.75
169.38
63.46
50.82
40.00
39.84
39.67
39.50
30.31
25.95
25.03
24.06
21.36
f1 (ppm)

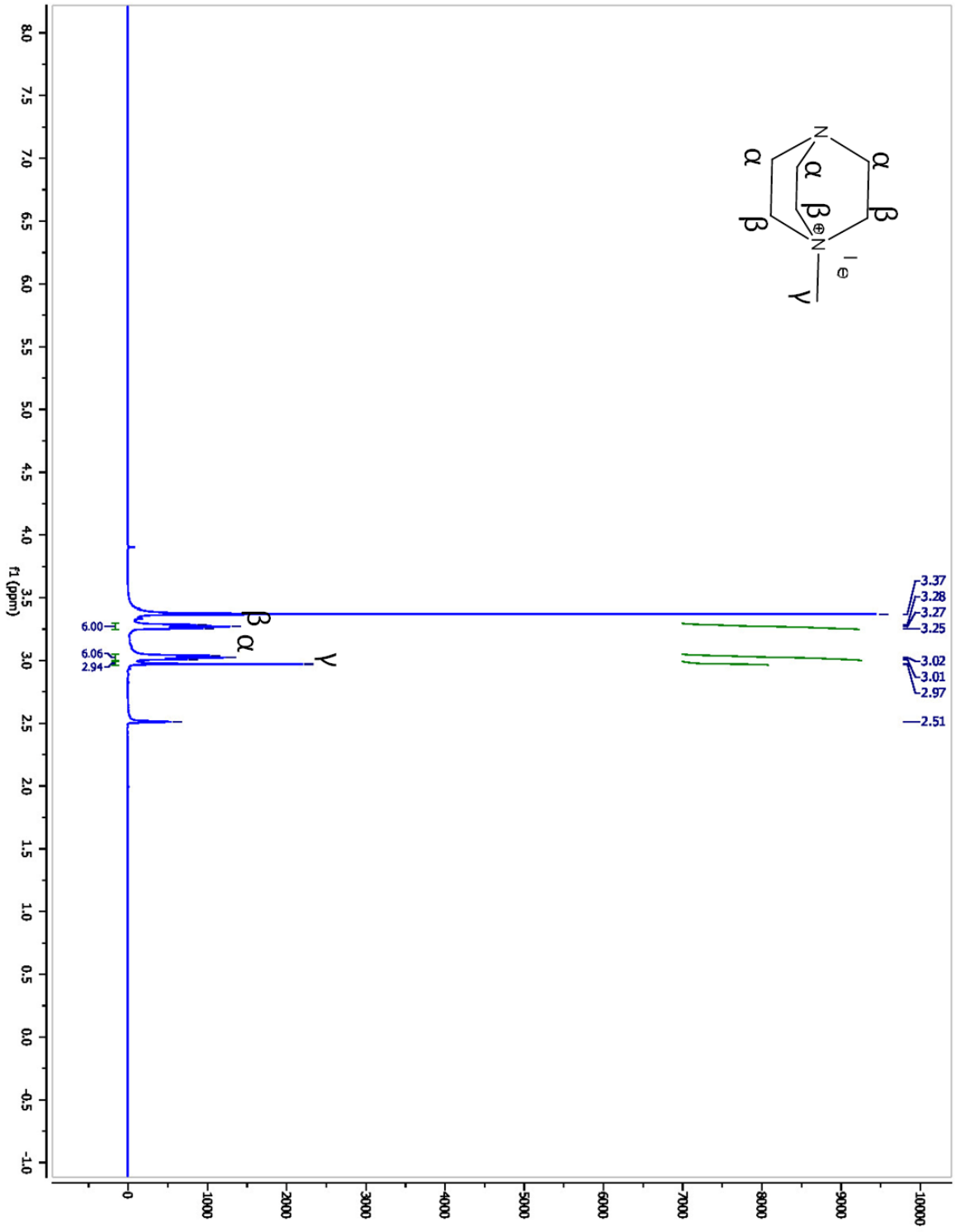

O
O
N
O
O
O
Br
N
Br
O
O
N
O
O

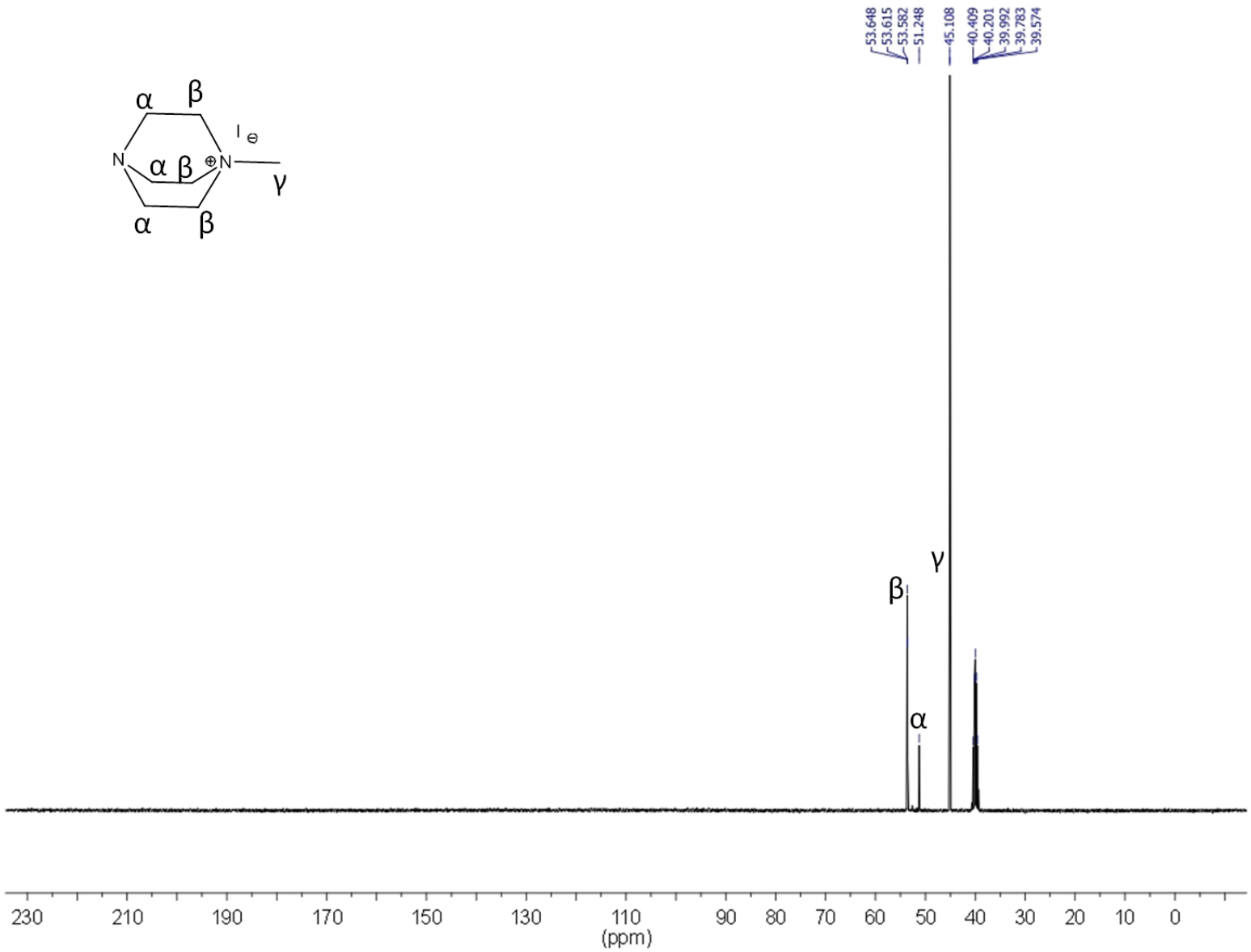

617.1999
615.2018
616.2025
Measured [M-Br]+ ➔
Predicted is 615.2024
Counts vs. Mass-to-Charge (m/z)

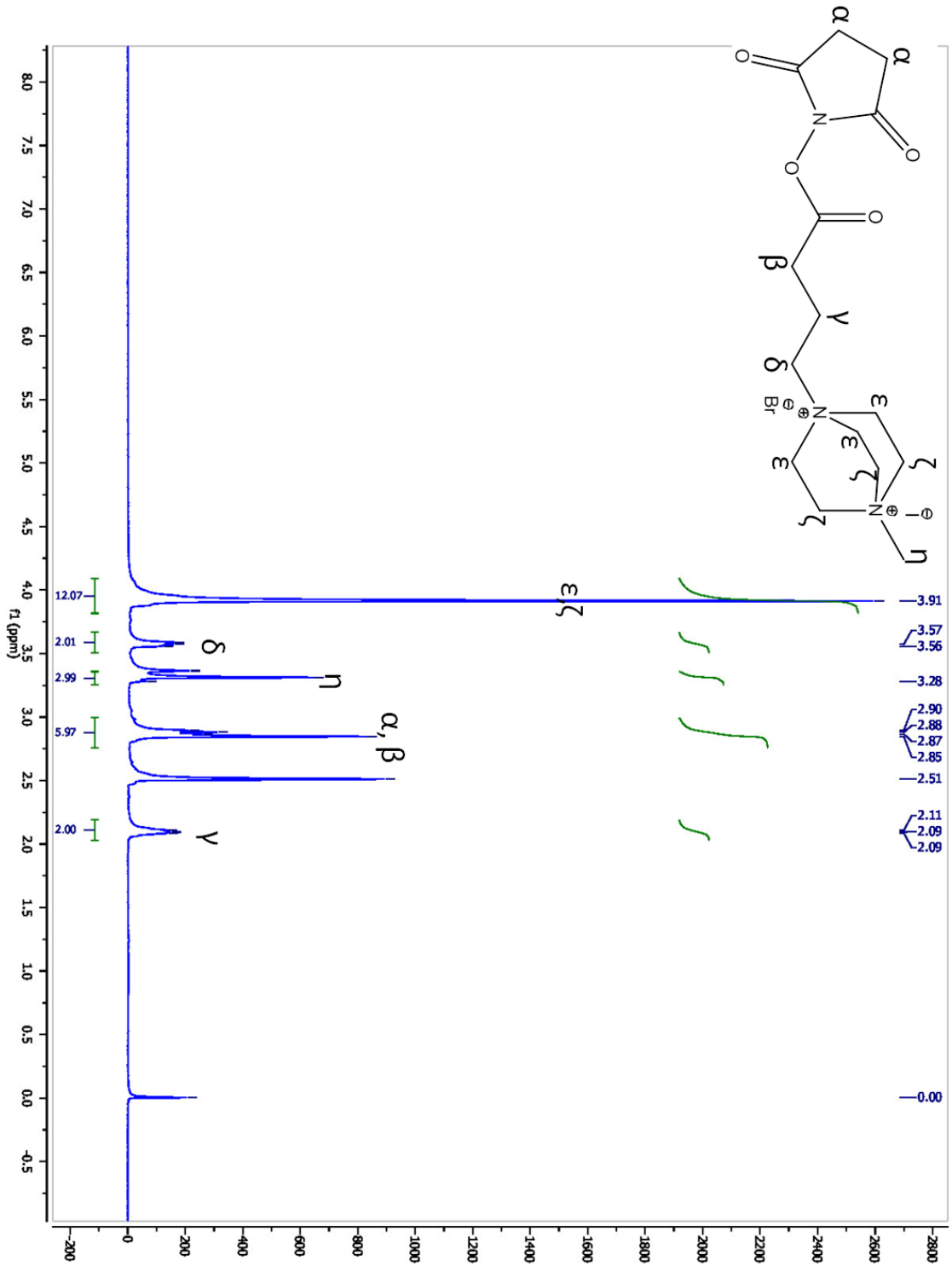

f1 (ppm)
8.05
5.97
6.00
2.07
ε,δ
ζ
α, β
γ
3.34
3.32
3.30
3.29
3.28
3.27
3.26
3.25
3.24
3.02
2.84
2.51
2.51
2.09
2.07
2.06
2.04
2.03
0.01

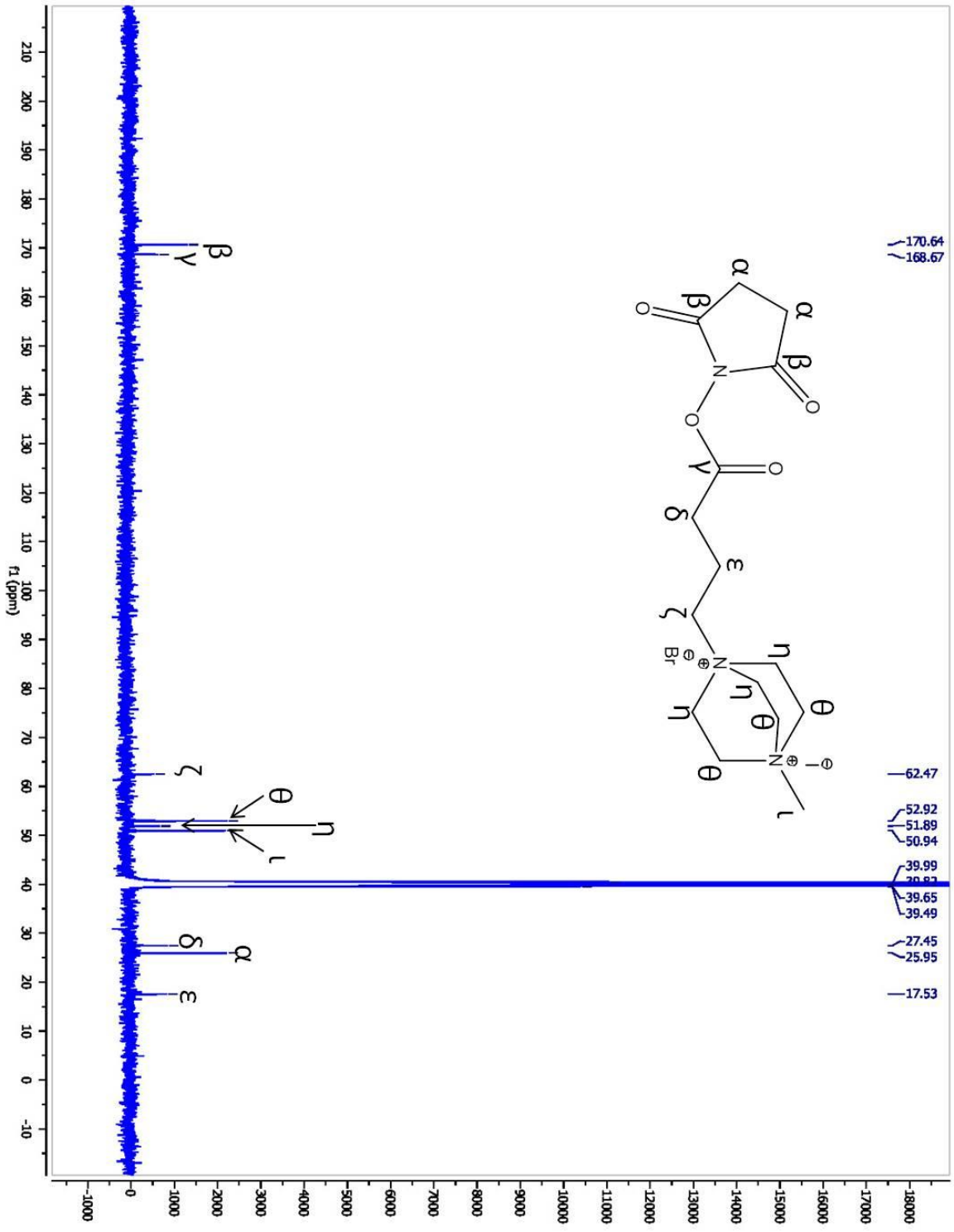

F1 (ppm)
β
γ
ζ
θ
η
δ
α
ε
170.68
168.75
62.22
52.14
45.15
40.14
39.81
39.47
27.68
25.93
17.24

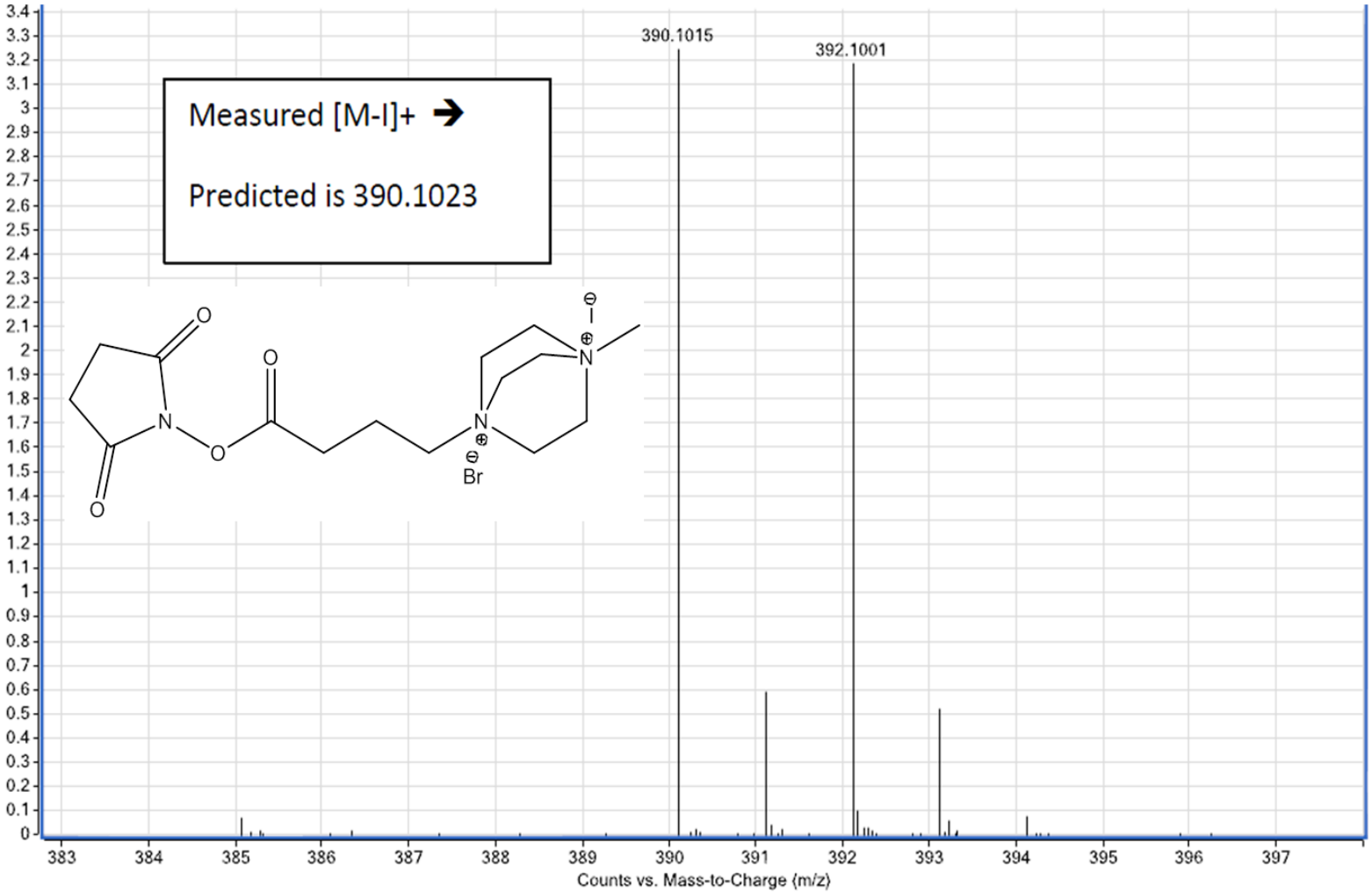

296.1607
Measured [M-Br]+
Predicted is 296.1605
O
O
N
O
O
N
N
Br
Counts vs. Mass-to-Charge (m/z)


f1 (ppm)
6.00
6.06
2.94
3.37
3.28
3.27
3.25
3.02
3.01
2.97
2.51
α
β
γ
I e


α
β
I ⊖
N
α
β
⊕N
γ
α
β
53.648
53.615
53.582
51.248
45.108
40.409
40.201
39.992
39.783
39.574
γ
β
α
230
210
190
170
150
130
110
(ppm)
90
80
70
60
50
40
30
20
10
0


f1 (ppm)
12.07
2.01
2.99
5.97
2.00
3.91
3.57
3.56
3.28
2.90
2.88
2.87
2.85
2.51
2.11
2.09
2.09
0.00
α
α
β
γ
δ
Br
ε
ε
ε
ζ
ζ
ζ
η
η
ζ, ε
δ
η
α, β
γ


170.64
168.67
62.47
52.92
51.89
50.94
39.99
39.65
39.49
27.45
25.95
17.53
f1 (ppm)


Measured [M-I]+ ➔
Predicted is 390.1023
390.1015
392.1001
Br
Counts vs. Mass-to-Charge (m/z)

Appendix B

Additional Annotated DC4 Crosslinked Peptide Spectra

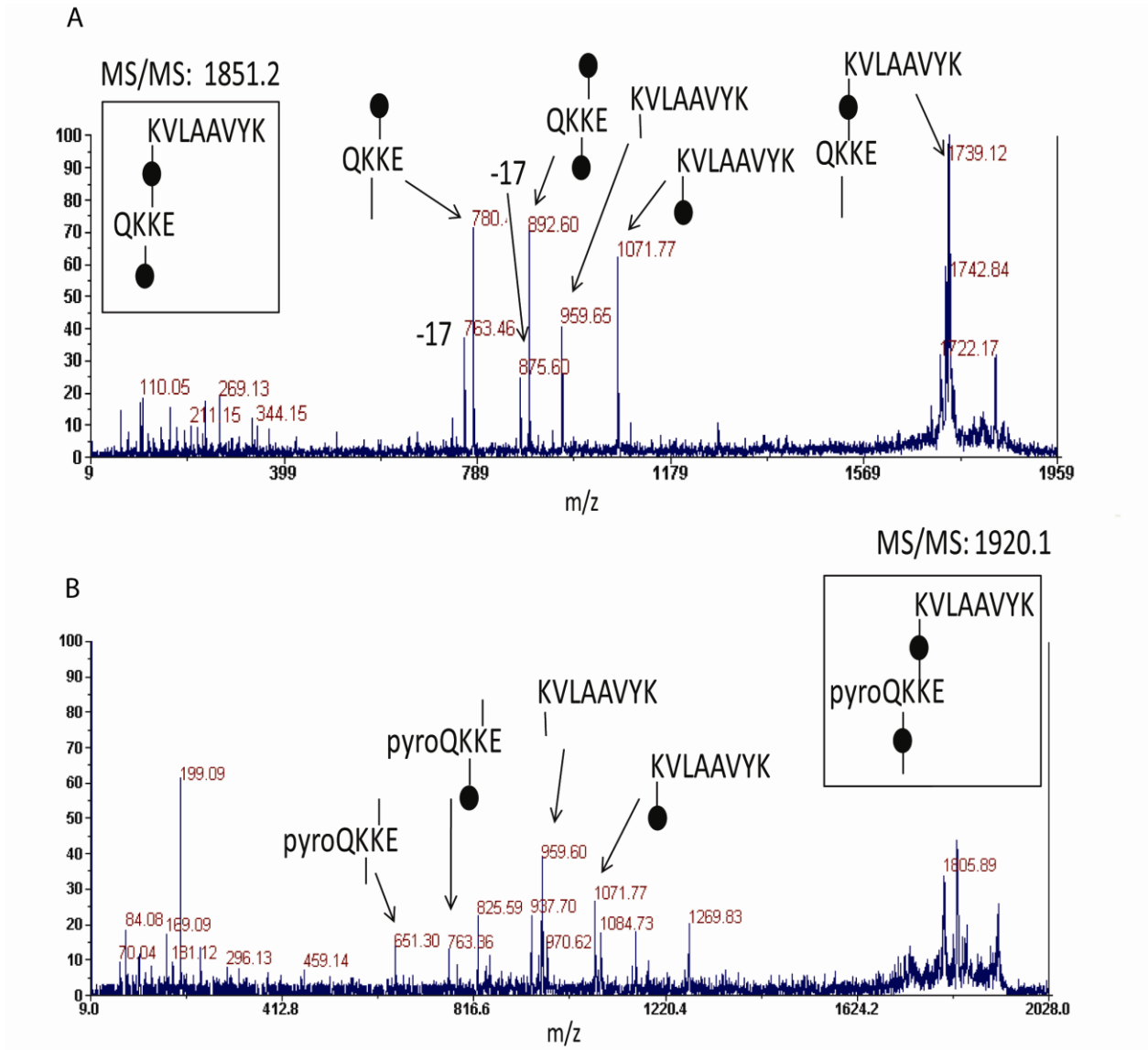


Figure B.1 CID of the DC4-crosslinked peptide at m/z 1851.1. The spectrum was assigned to the crosslinked peptide in shown in the black box.(A) This peptide contained both a crosslink and a dead-end which had already fragmented via ISD. The intact doubly modified peptide was only observed with the n-terminal glutamine cyclized to pyroglutamic acid (B). Similar peaks were observed in the MS spectrum from ISD and CID fragmentation of m/z 651.3 and 959.6 are shown in Figure B.16 and B.15, respectively.

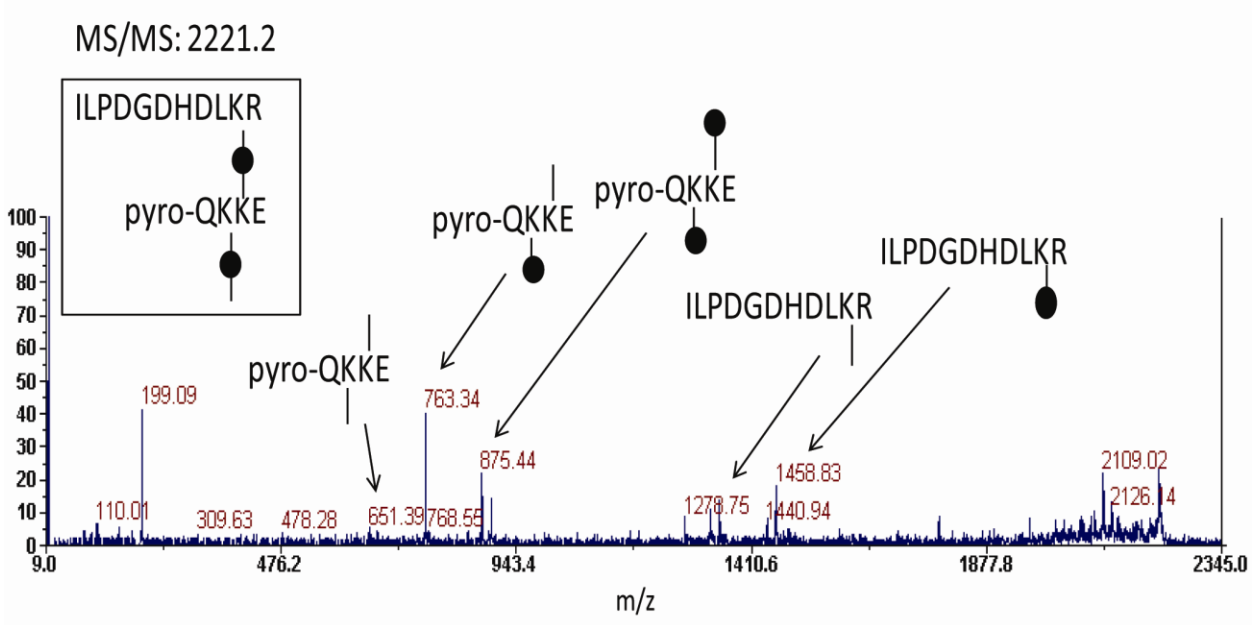


Figure B.2. CID of the crosslinked peptide at m/z 2221.2. This was assigned to the crosslinked peptide in shown in the black box. Similar peaks were observed in the MS spectrum from ISD and CID fragmentation of m/z 651.3 and 1346.5 are shown in Figure B. 16 and B.18, respectively.

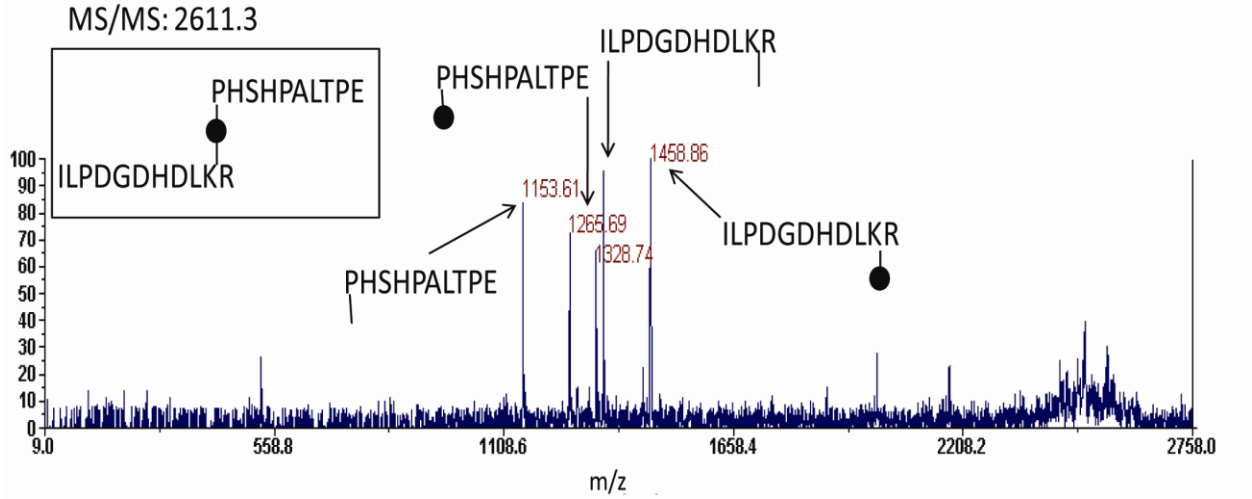


Figure B.3. CID of the crosslinked peptide at m/z 2611.3. This was assigned to the crosslinked peptide in shown in the black box. Similar peaks were observed in the MS spectrum from ISD and CID fragmentation of m/z 1153.5 and 1346.5 are shown in Figure B.12 and B.18, respectively.

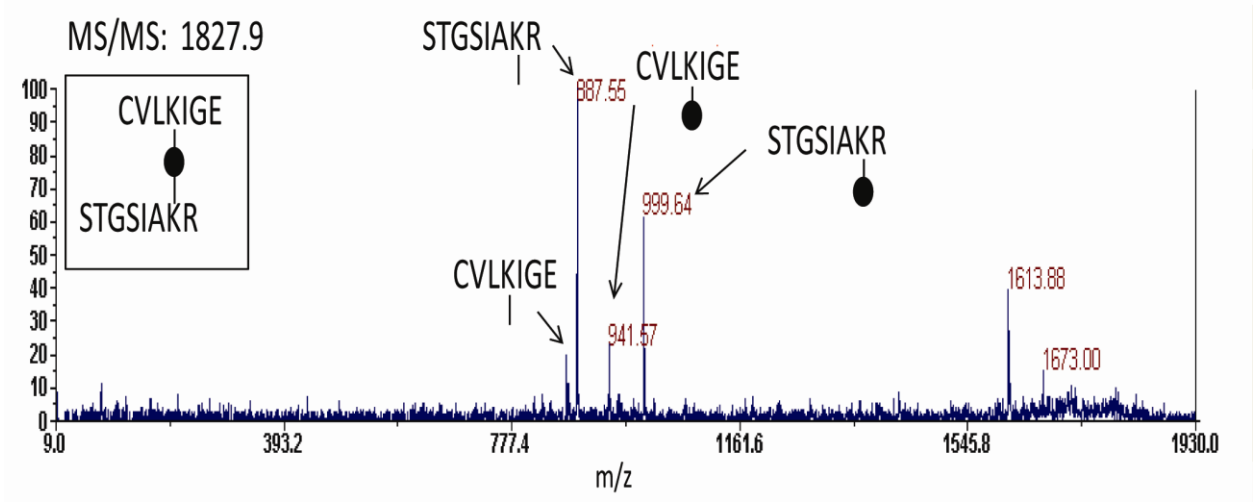


Figure B.4. CID of the crosslinked peptide at m/z 1827.9. This was assigned to the crosslinked peptide in shown in the black box. Similar peaks were observed in the MS spectrum from ISD and CID fragmentation of m/z 829.4 and 887.5 are shown in Figure B.11 and Figure B.17, respectively.

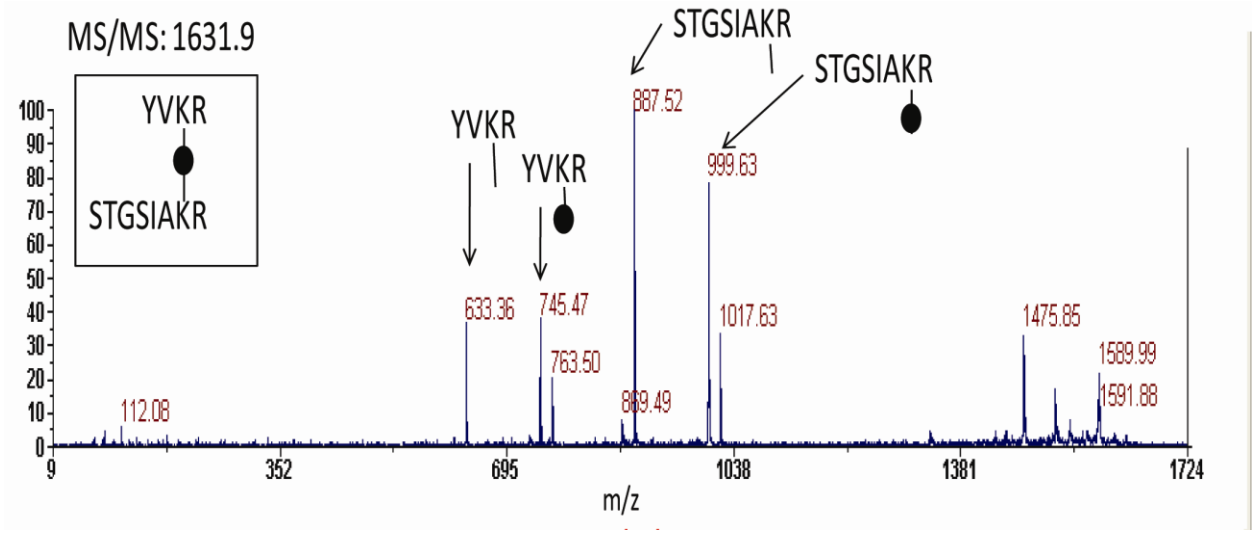


Figure B.5. CID of the crosslinked peptide at m/z 1631.9. This mass was assigned to the crosslinked peptide in shown in the black box. Similar peaks were observed in the MS spectrum from ISD and CID fragmentation of m/z 633.5 and 887.5 are shown in Figure B.14 and Figure B.17, respectively.

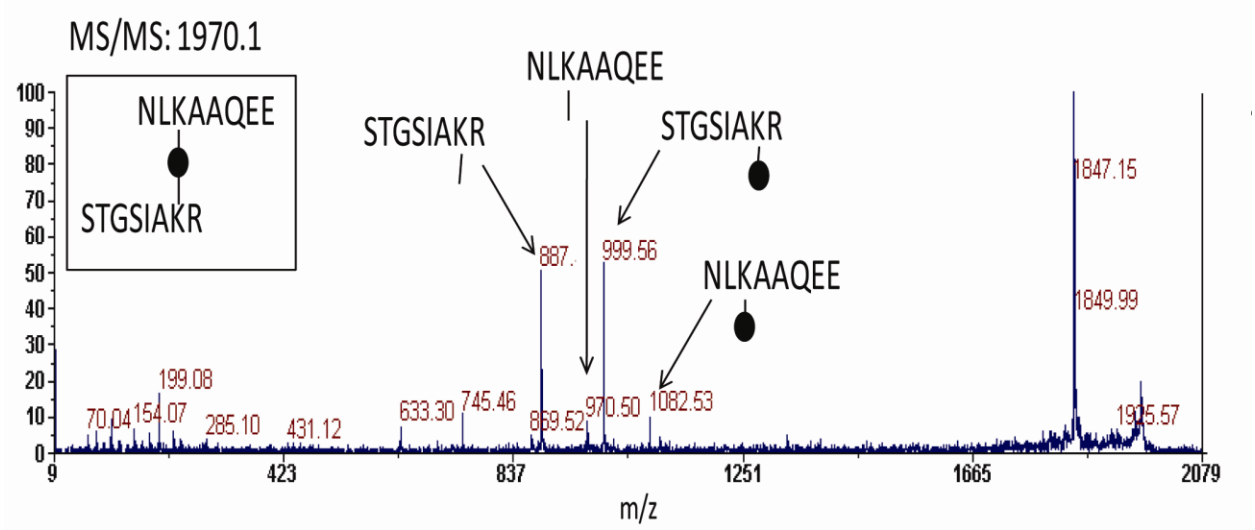


Figure B.6. CID of the crosslinked peptide at m/z 1970.1. This ion was assigned to the crosslinked peptide in shown in the black box. Similar peaks were observed in the MS spectrum from ISD and CID fragmentation of m/z 970.5 and 887.5 are shown in Figure B.10 and B.17, respectively.

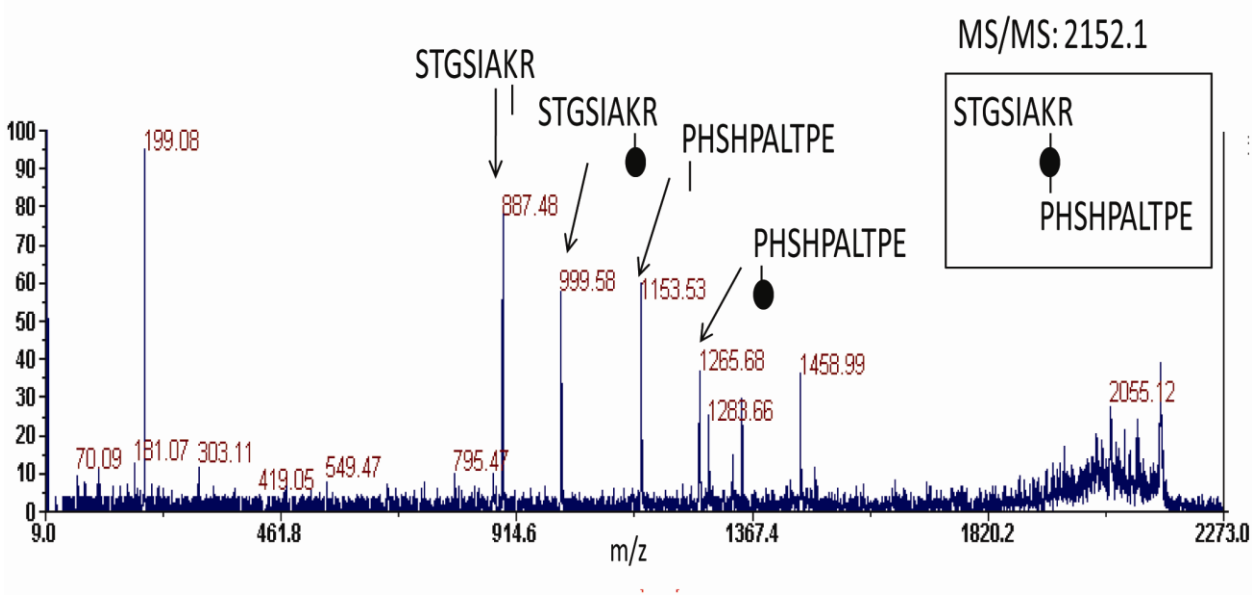


Figure B.7. CID of the crosslinked peptide at m/z 2152.1. This ion was assigned to the crosslinked peptide in shown in the black box. Similar peaks were observed in the MS spectrum from ISD and CID fragmentation of m/z 887.5 and 1153.5 are shown in Figure B.17 and B.12, respectively.

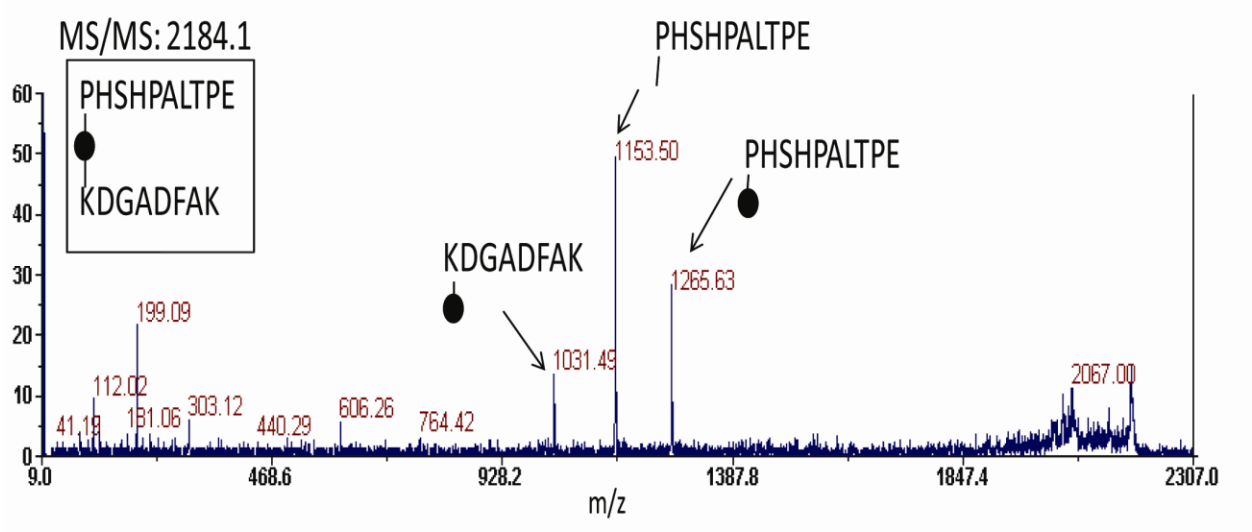


Figure B.8. CID of the crosslinked peptide at m/z 2184.1. This ion was determined to the crosslinked peptide in shown in the black box. Similar peaks were observed in the MS spectrum from ISD and CID fragmentation of m/z 919.5 and 1153.5 are shown in Figure B.13, and B.12, respectively.

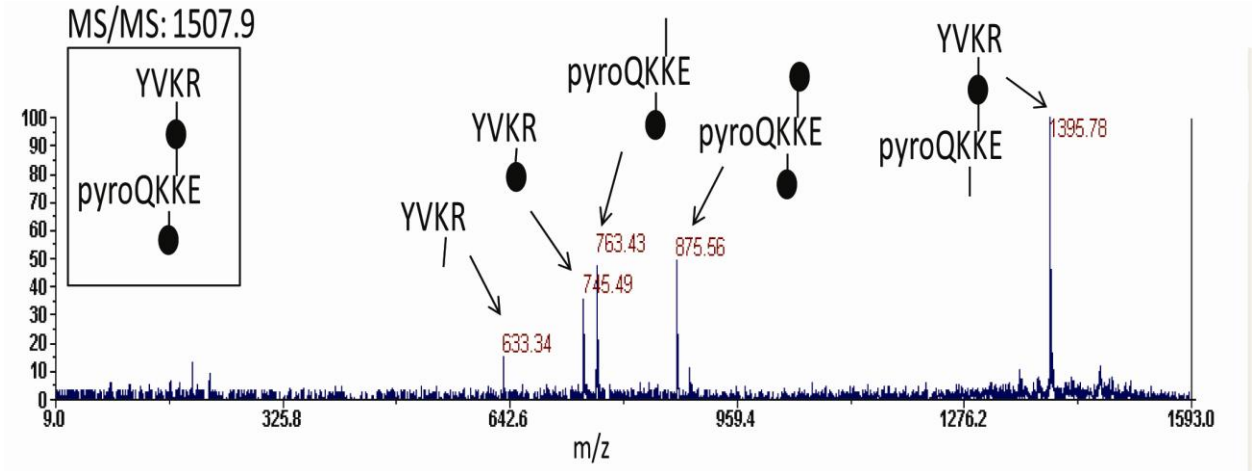


Figure B.9. CID of the crosslinked peptide at m/z 1507.9. This ion was determined to the crosslinked peptide in shown in the black box. Similar peaks were observed in the MS spectrum from ISD and CID fragmentation of m/z 633.5 and 651.3 are shown in Figure B.14 and B.16, respectively.

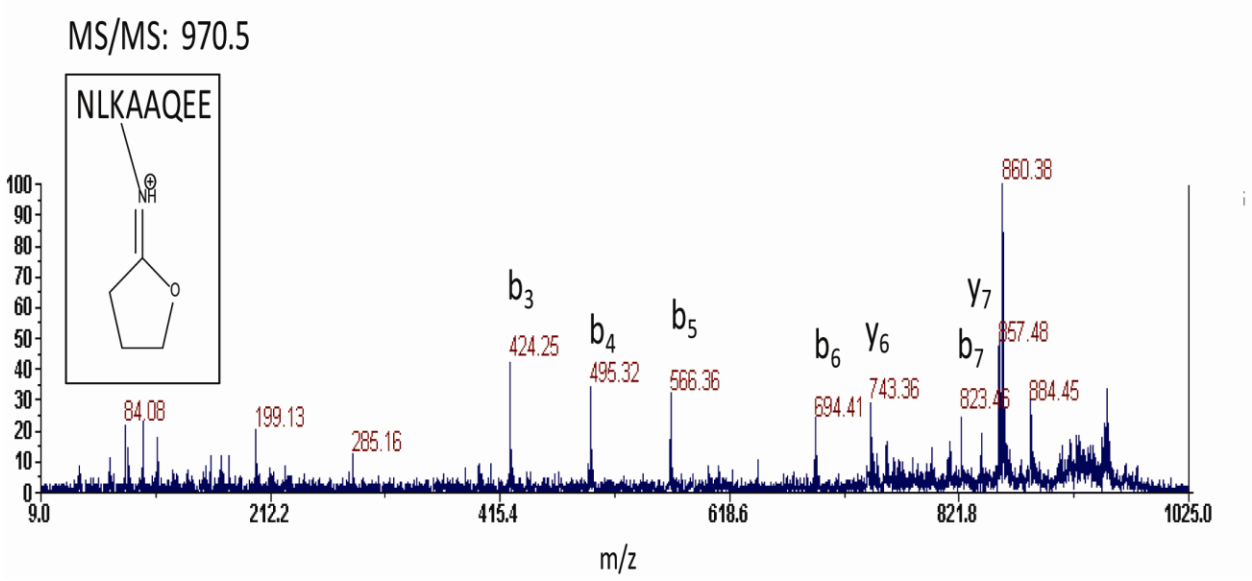


Figure B.10. CID fragmentation of the ISD product at m/z 970.5. This spectrum was identified to be the peptide NLKAAQEE with the lysine modified as shown.

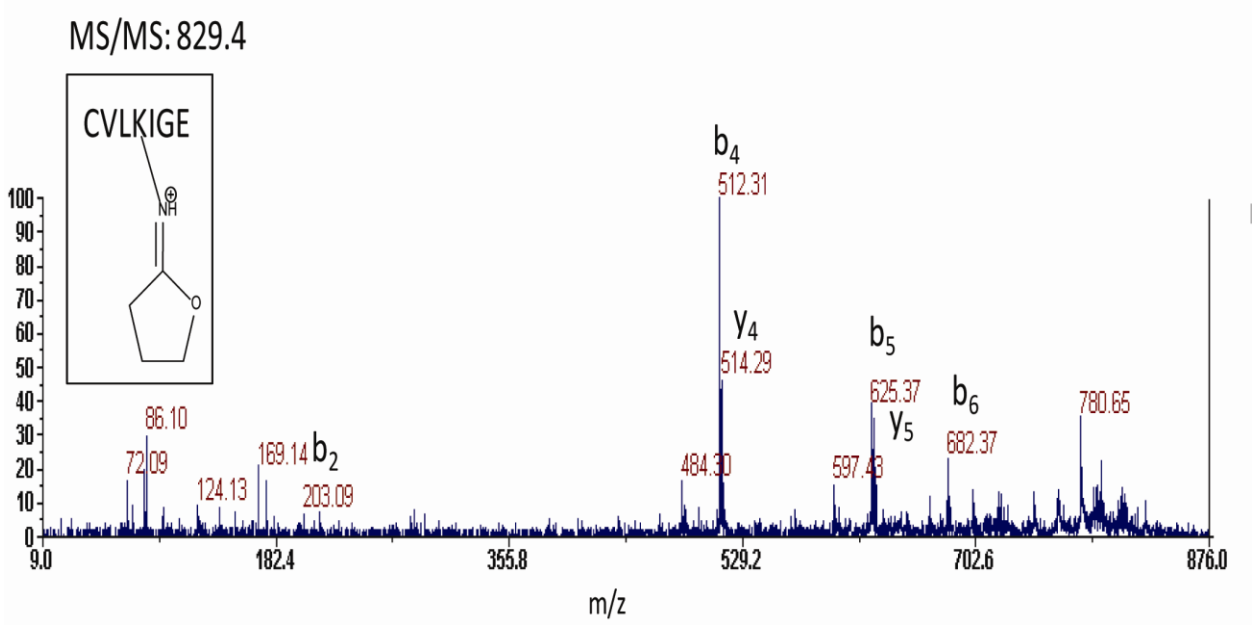


Figure B.11. CID fragmentation of the ISD product at m/z 829.4. This spectrum was identified to be the peptide CVLKIGE with the lysine modified as shown.

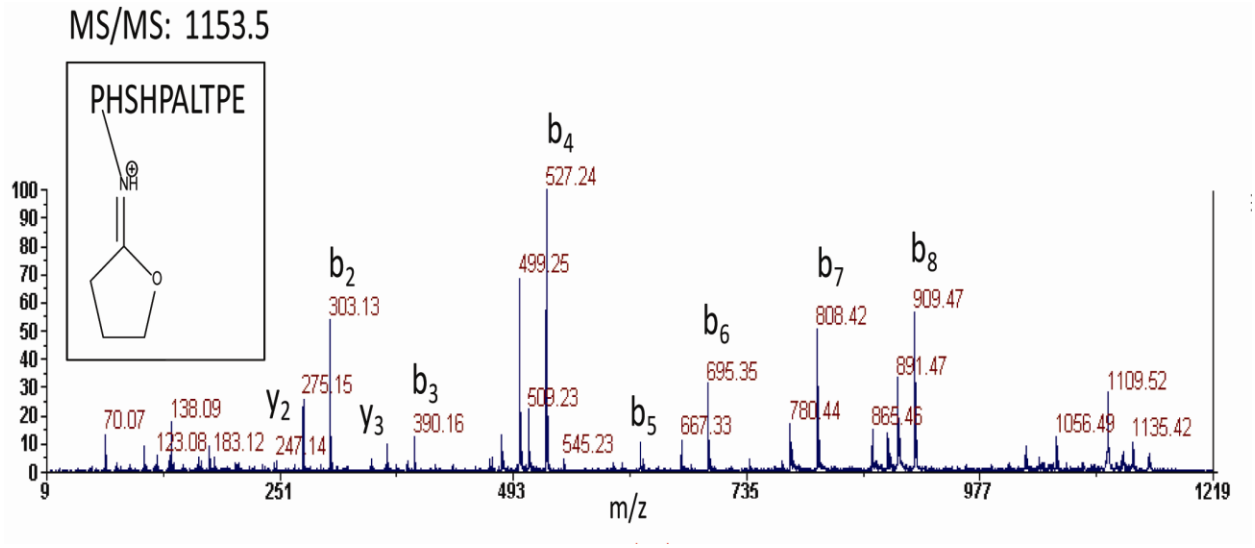


Figure B.12. CID fragmentation of the ISD product at m/z 1153.5. This spectrum was identified to be the peptide PHSHPALTPE with the n-terminus modified as shown.

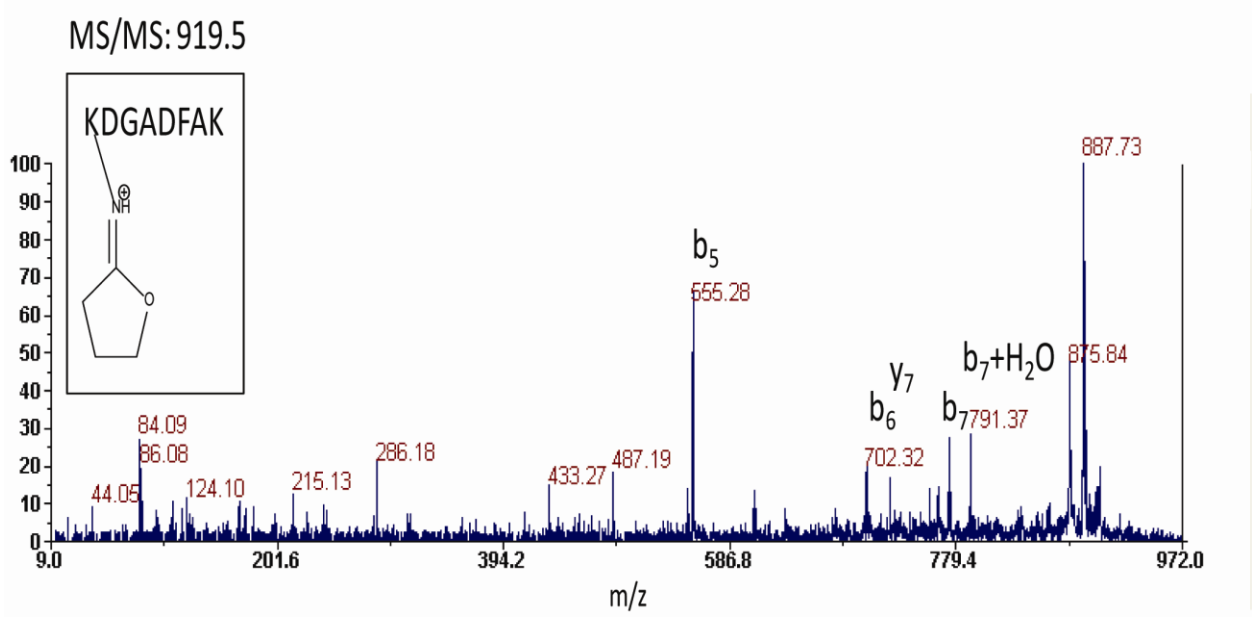


Figure B.13. CID fragmentation of the ISD product at m/z 919.5. This spectrum was identified to be the peptide KDGADFAK with the only the first lysine modified as shown.

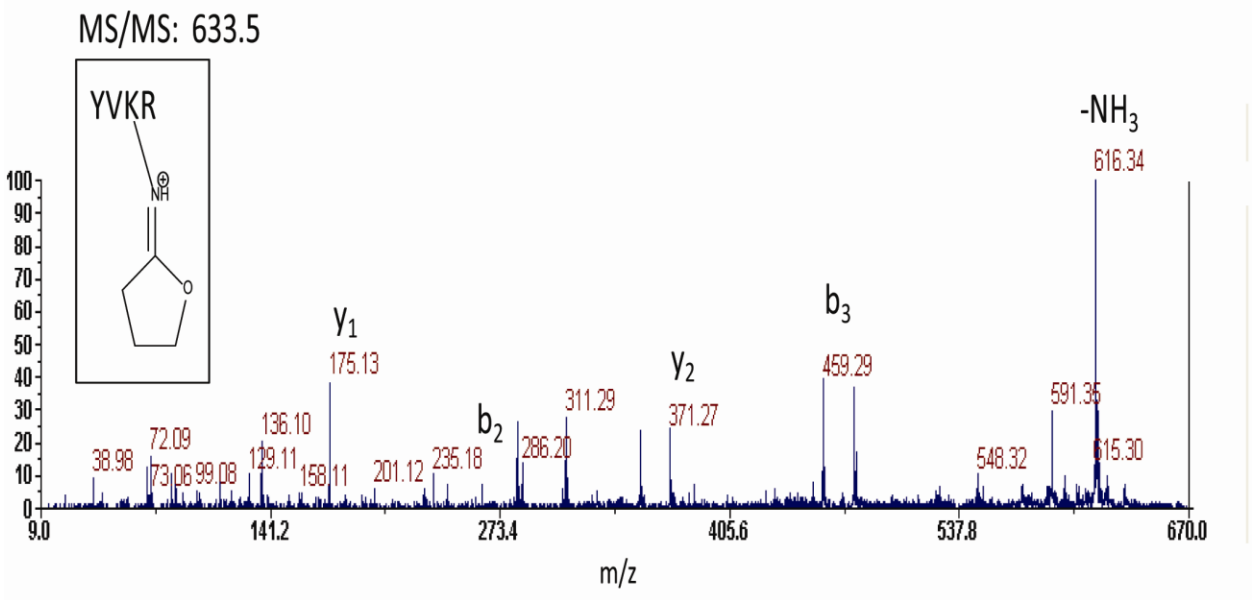


Figure B.14. CID fragmentation of the ISD product at m/z 633.5. This spectrum was identified to be the peptide YVKR with the lysine modified as shown.

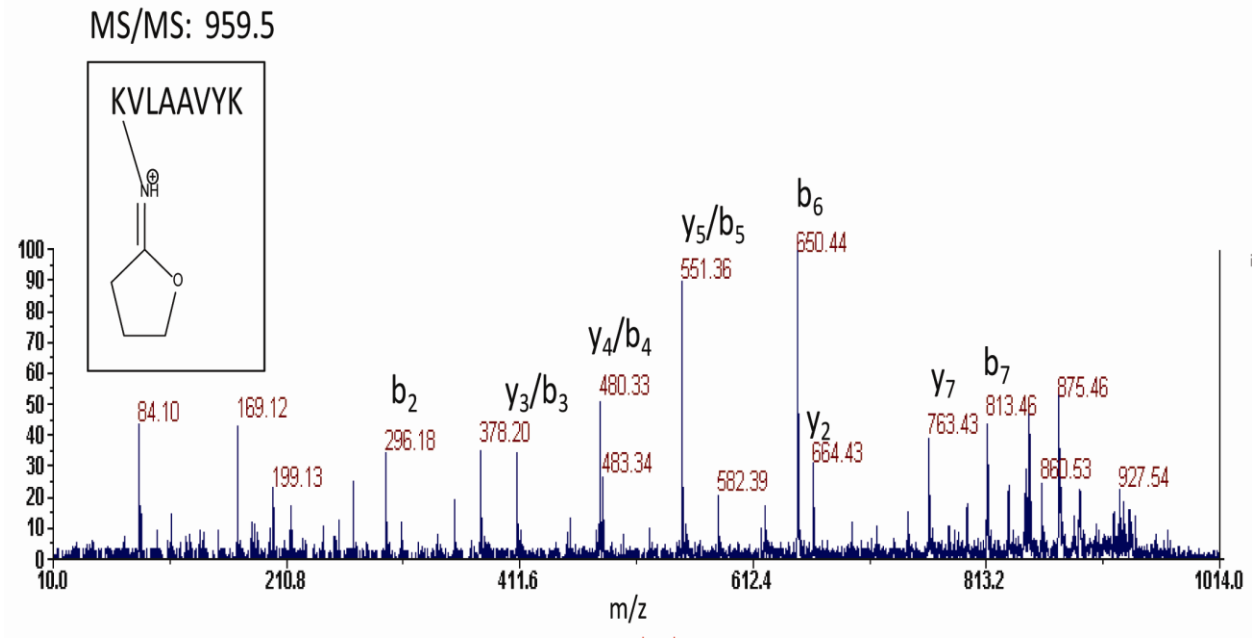


Figure B.15. CID fragmentation of the ISD product at m/z 959.5. This spectrum was identified to be the peptide KVLAAVYK with the first lysine modified as shown.

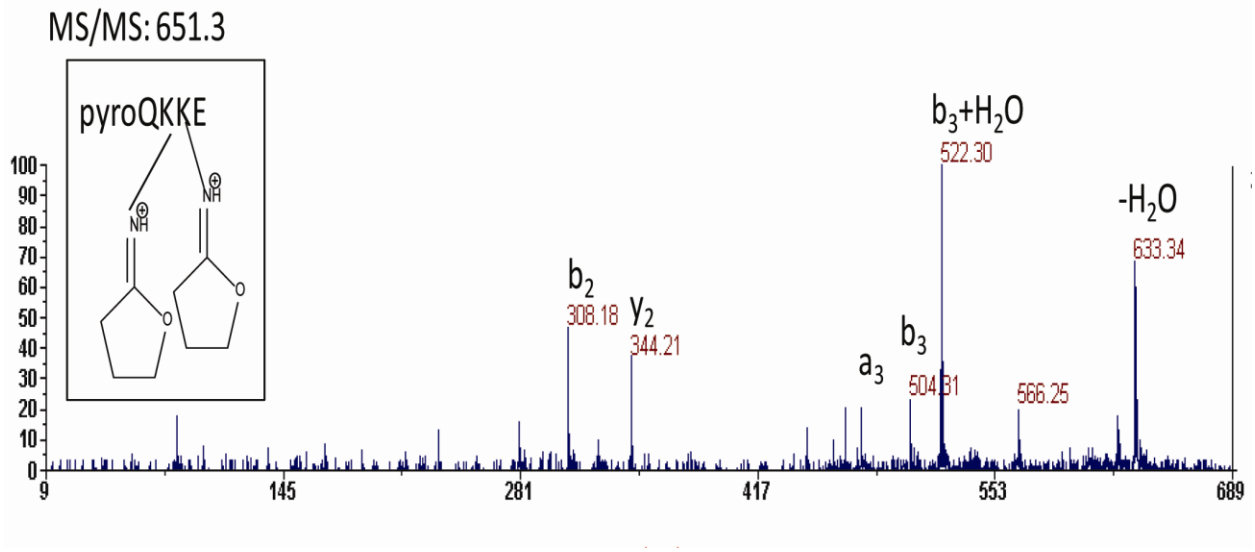


Figure B.16. CID fragmentation of the ISD product at m/z 651.3. This spectrum was identified to be the peptide QKKE with the lysines modified as shown and the glutamate pyrolyzed to glutamic acid. Although this species has two modified lysines, it is still singly charged due to the loss of a proton, most likely from the C-terminus or the glutamic acid side chain.

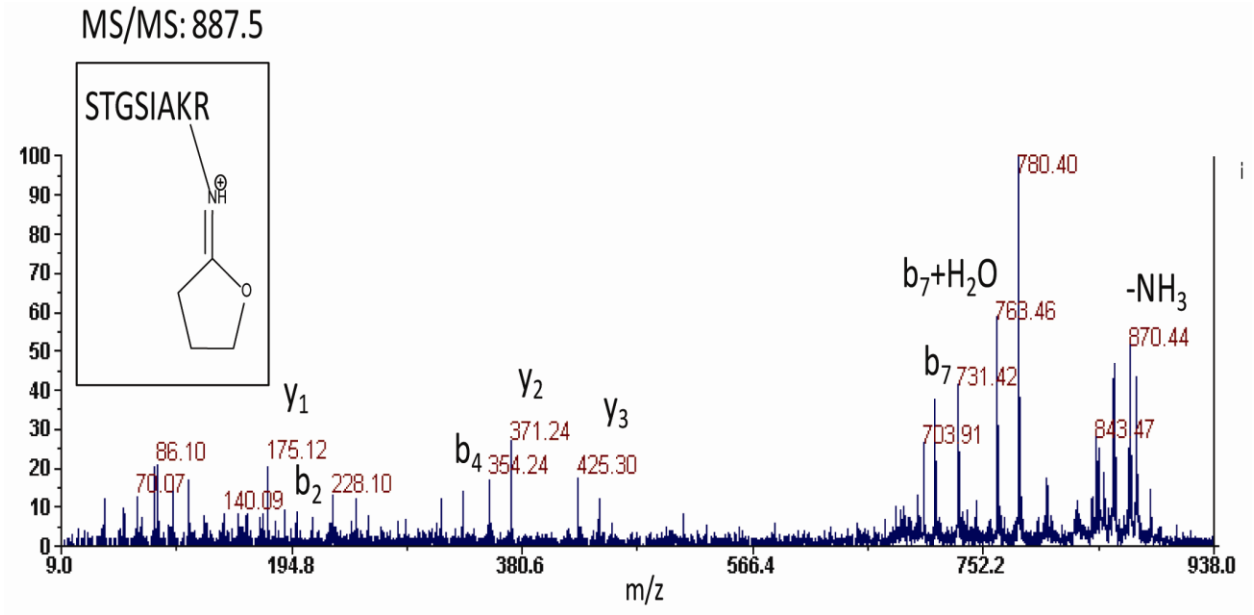


Figure B.17. CID fragmentation of the ISD product at m/z 887.5. This spectrum was identified to be the peptide STGSIAKR with the lysine modified as shown.

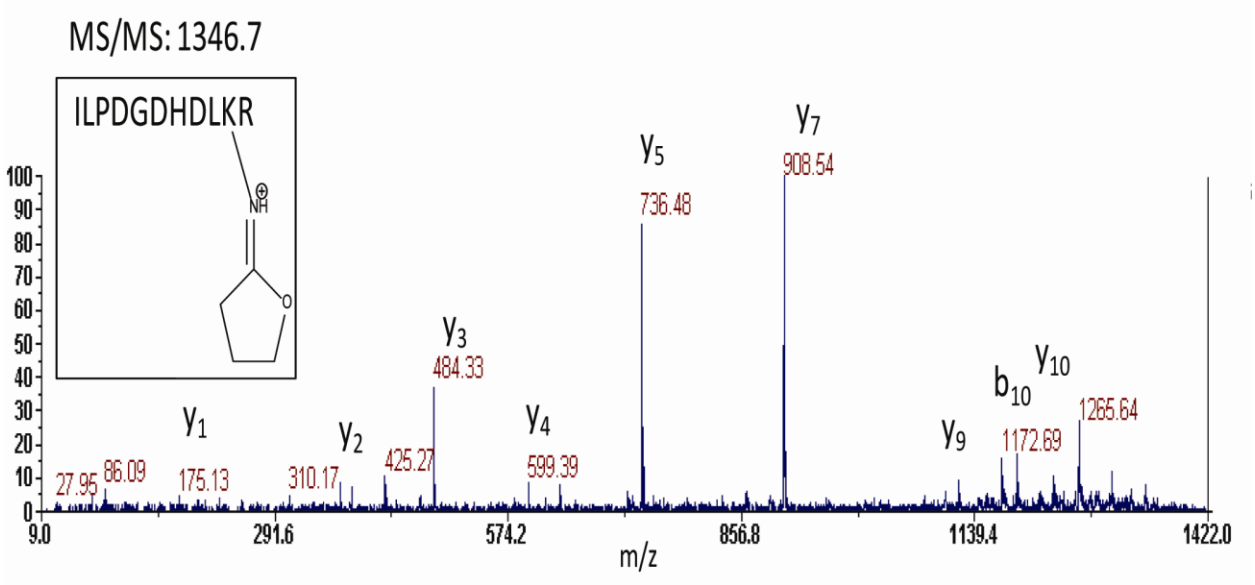


Figure B.18. CID fragmentation of the ISD product at m/z 1346.7. This was identified to be the peptide ILPDGDHDLKR with the lysine modified as shown.