NOVEL MASS DEFECT LABELED CHEMICAL CROSSLINKERS AND THEIR APPLICATION TO THE 34 kDA – ACTIN BIOLOGICAL SYSTEM

by

LISABETH LEIGH HOFFMAN

(Under the Direction of I. Jonathan Amster)

ABSTRACT

In order to improve the identification of crosslinked species within a complex mixture, mass defect label technology can be employed. The improvement is based upon the addition of a mass defect label into the chemical crosslinker of interest. This derivitization changes the mass defect of the species involved in the crosslinking reaction in a manner that increases the identification specificity. In this study, two mass defect labeled chemical crosslinkers were studied, DiBADMPS and DiBBSIAS. The crosslinkers effectiveness at crosslinking, fragmenting and overall identifying the peptides of interest was tested using two small peptides, Neurotensin and Bradykinin, and a small protein/peptide complex, Ribonuelase S. After initial tests were completed, these crosslinkers were applied to the 34 kDa – Actin biological system in order to identify the regions of interactions between these two proteins. To identify the specific

amino acids of interest in each protein, HPLC-ESI-FTICR-MS/MS was employed. This allowed for the crosslinked peptides to first be identified and then be fragmented within the ICR cell to produce amino acid sequence information. The results of this study identified the major areas of interactions for 34 kDa protein on Actin.

INDEX WORDS: Crosslinker, Tandem Mass Spectrometry, Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, Protein complexes, 34-kDa protein, Actin, EDC, BS^3, N-hydroxysuccinimide esters, Ribonuclease S, Neurotensin, Bradykinin, MS3D Links

NOVEL MASS DEFECT LABELED CHEMICAL CROSSLINKERS AND THEIR APPLICATION TO THE 34 kDA – ACTIN BIOLOGICAL SYSTEM

by

LISABETH LEIGH HOFFMAN

B.S., Rochester Institute of Technology, 2005

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirement of the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2010

NOVEL MASS DEFECT LABELED CHEMICAL CROSSLINKERS AND THEIR

APPLICATION TO THE 34 kDA – ACTIN BIOLOGICAL SYSTEM

by

LISABETH LEIGH HOFFMAN

Major Professor:	I. Jonathan Amster
Committee:	John Stickney
	Lance Wells

Electronic Version Approved:
Maureen Grasso
Dean of the Graduate School
The University of Georgia
July 2010

ACKNOWLEDGMENTS

Firstly, I would like to thank my mom, dad and Brett for all of their encouragement and constant faith that I could do anything that I put my mind to. I would not be where I am today without you guys (ya'll) and I am so thankful that you modeled me into the woman that I am today. I love you both so much. I would also like to thank Dr. Amster for all of his patience with me over the last few years. I understand that I have not always been the model student and I thank him for his commitment, understanding and belief in me. Mike Easterling also deserves a "shout out" for all of his encouragement and for acting as my soundboard when times became difficult. When I first began graduate school I could not have succeeded without the help of Jeremy Wolff and in the end his significant other, Lindsay Renbaum, encouraged me to press on. I would also like to thank my dedicated collaborators, Christoph Borchers, "Jenya" Petrotchenko, Paul Griffin, Marcus Fechheimer and Ruth Furukawa for all of their hard work and persistence in seeing this project through. Thanks to all of you for your help. I truly would not have made it through without your faith, dedication, persistence, and hard work.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SCHEMES

Page

CHAPTER 1:

Introduction/Literature Review

Analytical and Bioanalytical Techniques applied to structural biology questions. Protein-protein interactions are extremely important in the biological world. Biological systems utilize critical interaction networks that range from stable multi-protein complexes all the way to transient interactions.[1] Throughout the years, analytical and bioanalytical techniques have emerged that have allowed the questions of spatial and topological organization of biomolecules, specifically proteins and protein complexes, to be examined. In order to gain a better understanding of the biological world and the functions within it, it is imperative to first understand the structure of these molecules and interactions that take place within them.

X-ray crystallography and NMR spectroscopy have emerged as the leaders in this type of analysis. While these two analytical techniques provide detailed information on the structure of proteins and protein complexes, NMR spectroscopy requires large quantities of pure sample in a specific solvent system, and X-ray studies require the crystallization of the protein or protein complex,[2] thereby limiting their application for the analysis of large multi-subunit biological complexes. Another analytical technique that has been employed for the structural determination

of proteins and protein complexes is electron cryomicroscopy (cryo-EM). This technique offers a number of advantages, specifically for large multi-subunit protein complexes, by providing an estimate of the complexes in their native, interacting states, and giving information related to their cellular functions.[3] However, cryo-EM offers much lower resolution (6-9 Å)[3] than the other analytical techniques mentioned.

A number of bioanalytical techniques have been developed to probe protein-protein interactions and three-dimensional structure as well, such as phage display, yeast two-hybrid screens, protein microarrays, gel filtration, surface plasmon resonance, and fluorescence spectroscopy.[1] Mass spectrometry has also been adapted to answer the questions of low resolution three-dimensional structure of large biomolecules and interacting partners in biological systems. Robinson and coworkers[4] have demonstrated the structural characterization of large intact biological assemblies, such as the intact 70S ribosome, by means of tandem mass spectrometry. Using this method, it is possible to obtain low-resolution details regarding three-dimensional structures, such as subunit stoichiometery and the interaction partners in a protein complex but the regions of contact are not revealed by these measurements. Chemical crosslinking however, is a tool that can be used to elucidate the regions of interaction in protein complexes. Additionally, crosslinking allows interactions that are either transient in nature or dependent on specific physiological conditions to be captured as long-lived covalent complexes,

in which structural information is maintained during sample processing (purification, enrichment and analysis).[1-7]

Mass spectrometry has recently been coupled with studies involving chemical crosslinkers within biomolecules. This analysis technique has the advantages of requiring small amounts of material, does not require high purity of the sample, and is a relatively fast procedure.[8,9] Advances in matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF)[10] and electrospray ionization (ESI)[11] mass spectrometry have pushed these methods to the forefront of cross linker analysis, due to their high sensitivity and ability for the rapid analysis of the complex mixtures obtained after enzymatic digestion of the crosslinking reaction mixtures.[8]

In the last few years, time-of-flight mass spectrometry and tandem mass spectrometry (MS/MS) have been employed to investigate the subunit stoichiometry of large biological complexes and also to gain insight into the interaction regions within the molecules.[4-6] While these techniques have advanced the field of chemical crosslinking, Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) has provided the most significant improvements in mass range, resolving power, mass accuracy, and tandem mass spectrometry (MS/MS) to date, particularly in conjunction with electrospray ionization.[12-14] Tandem mass spectrometry (MS/MS) has also developed over the years and is now used throughout these studies to provide sequence-specific identification of crosslinked peptides and to improve

confidence in crosslink assignments. The most accurate information concerning distance constraints within proteins or protein complexes comes from the precise identification of the sites involved in the crosslinking reaction. In most biological experiments, crosslinked peptides or protein complexes tend to be relatively large and of high mass-to-charge (*m/z*). This often makes extensive fragmentation by MS/MS difficult.[14] As the complexity of the mass spectrum increases, the number of peptides with the same nominal mass but differing amino acid sequences increases. Therefore, the high resolution and mass accuracy that FTICR-MS offers, combined with the ability to comprehensively fragment large peptides by FTICR-MS/MS are indispensible when conducting a crosslinking experiment, and attempting to unambiguously assign the crosslinks.[12,15,16] In addition to providing high mass accuracy, FTICR-MS allows a "gas-phase" purification and subsequent fragmentation of the crosslinked products in the ICR cell.[17] In general FTICR-MS in combination with chemical crosslinking procedures is the most widely used analytical technique for the low-resolution structural determination and the determination of intermolecular relationships between peptides and protein complexes, and thus will be used in the research presented in this dissertation.

Crosslinking approaches involving mass spectrometry. Generally, chemical crosslinking experiments combined with FTICR-MS can be grouped into two specific categories. The first is the analysis of proteolytic digests of crosslinked proteins, otherwise called the 'bottom-up

approach'. The second approach is the analysis of intact crosslinked proteins, referred to as the 'top-down approach'.[17-19] These two strategies are presented in Figure 1.1.[17] Over the years, different variations and adaptations have been made to the general 'bottom-up' and 'top-down' analytical procedures, however the general format has remained the same. In the 'top-down' approach, intact proteins are analyzed directly, mainly by electrospray ionization (ESI)-FTICR mass spectrometry. The overall goal in a 'top-down' analysis procedure is to obtain details of regions of crosslinking reactivity, at the amino acid level.[2,4,6,13,17,18,20] Generally, the crosslinked protein is isolated in the ICR cell and then subjected to one of a number of fragmentation techniques, such as sustained off-resonance irradiation collision-induced dissociation (SORI-CID), infrared multi-photon dissociation (IRMPD), collision-activated decomposition (CAD), or electron capture dissociation (ECD) (Figure 1.2).[17] Exact masses of the intact crosslinked product can then be used to determine the number of molecules incorporated in the crosslinked molecules. Kruppa *et al.* demonstrated that chemical crosslinking followed by a simple, one-step cleanup and mass spectrometric analysis can be used to obtain constraints for a small protein, such as ubiquitin, using the 'top-down approach'.[9] As top down analysis is most applicable to protein complexes under 50 kDa, it is limited in its applicability compared to bottom up analysis.

In a 'bottom-up' analysis, a protein reaction mixture is subjected to enzymatic digestion after a crosslinking procedure has been performed. This is followed by mass spectrometric

analysis of the resulting proteolytic peptides.[2,17-19] A 'bottom-up' procedure can also include one-dimensional gel electrophoresis (SDS-PAGE) and/or matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS analysis to verify that crosslinking has occurred. Protein excision from the 1D-gel or separation by size exclusion chromatography allows the analysis to focus on crosslinked species only, and eliminates unlinked subunits. After separation, it is then necessary to digest the protein enzymatically, which results in a complex mixture of peptides containing both those that have been modified by the crosslinking reaction and those that have remained unmodified. This mixture can be subsequently analyzed by high performance liquid chromatography (HPLC) combined with a mass spectrometer, generally with MS/MS capabilities.[8] Automated computer software can then complete the crosslinking analysis. In 2000, Young *et al.*[21] introduced a 'bottom-up' approached called MS3D, which combined intramolecular chemical crosslinking, proteolysis, and mass spectrometry to identify crosslinks in peptides and then used these identities to define important features of the tertiary structure of a protein.[14,21]

Chemical Crosslinker Design. Chemical crosslinking is the formation of covalent bonds between different molecules (inter-molecular) or parts of a molecule (intra-molecular). The molecules that have been linked by the crosslinking reagent include proteins, peptides, drugs, nucleic acids or even solid particles.[15] When crosslinking intra-molecularly the ultimate goal is

to obtain information on the three dimensional structure of a protein.[23] In contrast, inter-molecular crosslinking elucidates the regions of interaction between protein subunits and provides data about subunit stoichiometry.[24]

Over the years, considerable research effort has been directed toward developing new chemical crosslinkers. There are four major crosslinker designs, homobifunctional, heterobifunctional, zero-length, and trifunctional crosslinkers (Figure 1.3). Homobifunctional crosslinkers are those with two identical functional groups making up their reactive sites.[17,19,25] Heterobifunctional crosslinkers on the other hand contain two different functional groups at their reactive sites, which target different types of side chains. Zero-length crosslinkers are the smallest available and thus can be used to mediate crosslinking between two proteins by creating a bond directly between two amino acid side chains, for example by creating an amide bond between an amine side chain (lysine) and a carboxyl side chain (aspartic or glutamic acid). Trifunctional crosslinkers incorporate two reactive groups for forming a crosslinking, and a third group that can be used for affinity purification of the crosslinker containing species, for example by incorporating a biotin moiety.[18,22,26,27]

Although there are many references to chemical crosslinking scattered throughout the literature, the chemistry behind the overall reactions are very similar, and only utilize a few primary organic reactions. The first type of reaction is crosslinks of amino groups, such as the ε-

amino group of lysine or the protein N-terminus. These are by far the most common types of chemical reactions used. Imidoesters are the most specific acetylating groups, which modify primary amines. N-hydroxysuccinimide (NHS) esters were originally introduced as homobifunctional, highly amine-reactive, crosslinking reagents in the 1970s.[28,29] Recently it has been found that sulfo-NHS esters can be used as an alternative because they are the water-soluble analogue to NHS esters.[18,30] These two reagents react with nucleophiles and liberate a free NHS or sulfo-NHS group, which in turn creates a stable amide or imide bond with a primary or secondary amine. This is the reaction chemistry that will be used throughout the following research. In 1982, Staros demonstrated the utility of sulfo-NHS in solving the hydrolysis problem associated with NHS.[30] This reagent was used in crosslinking experiments with intact human erythrocytes and erythrocyte membranes.[30] The successful use of sulfo-NHS esters in membrane-impermeable crosslinkers eventually led to its use for producing active esters for sensitive reaction schemes in aqueous solutions. In 2005 Bruce *et al.*[31] introduced a novel mass spectrometry identifiable crosslinker, called a protein interaction reporter. This incorporated activated NHS esters for linking lysine side chains (Figure 1.4).

A second type of chemical reaction used, is those used in sulfhydryl-reactive crosslinkers. These reactions become involved in disulfide bonds.[18,32-35] This may disrupt the three-dimensional structure of the molecule, rendering any information obtained less relevant. The

final reaction scheme used in chemical crosslinking strategies, are those reactions that yield photo-reactive crosslinkers. These crosslinkers can be made to react upon irradiation with UV light. The largest number of photo-reactive reagents are those involving nitrene or carbine chemistry, with the photo-labile precursors being azides, diazirines, diazo compounds and benzophenones.[18,36-39] Alley *et al.*[39] demonstrated the use of a novel trifunctional crosslinker, incorporating a thiol-reactive 2-thiopyridine mixed disulfide group, a photo-activatable aryl azide and an affinity probe biotin, to determine the precise location of the protein-protein interactions in solution in the bacteriophage T4 holoenzyme. Trester-Zedlitz *et al.*[27] examined a variety of different trifunctional crosslinking reagents and found that the most promising candidate was that which possessed one amine-reactive group, one photo-reactive site and one biotin moiety (used for affinity-based enrichment). While biotin has the advantage of concentrating crosslinked products, it is a bulky functional group and may prevent the crosslinking reagent from fitting into regions of close contact. It also is thought to reduce the detection sensitivity for MS and MS/MS analysis.

Problems regarding conventional crosslinking experiments and methods of solving them. One of the major challenges of chemical crosslinking experiments, however, is the tremendous sample complexity attributable to the wide variety of crosslinked products that can be created during the crosslinking reaction. After proteolysis, it is possible that the reaction

mixture can contain any of the following types of crosslinked products: (1) peptides that have been modified with a hydrolyzed crosslinker (‘dead-ends), (2) a peptide that has two reactive groups crosslinked within itself (intra-peptide crosslinking product), or (3) crosslinks that occur between two independent peptides (inter-peptide crosslinking products). The first type of product, the hydrolyzed crosslinker, is that which has reacted with a peptide at one of its reactive sites, however the other reactive group has undergone hydrolysis. These products are often called ‘dead end’ crosslinks and offer little information to a researcher.

To circumvent the disadvantage of having a complex mass spectrum, a number of different strategies have been developed that all facilitate identification of crosslinkers containing species, e.g. by employing isotope-labeled crosslinkers,[40,41] isotope-labeled proteins,[42] cleavable crosslinkers,[43] fluorogenic crosslinkers,[19] or crosslinkers creating a characteristic marker ion during MS/MS analysis.[31] Researchers have also attempted to enrich the crosslinker containing species by using trifunctional crosslinkers which contain a biotin moiety[22,27,44,45] and which can be affinity purified with an avidin column. Back *et al.*[46] and Bennett *et al.*[43] have used cleavable crosslinkers and shown mass shifts in the mass spectra before and after crosslinker cleavage. Tang *et al.*[31] reported the use of a novel crosslinker, PIR (Figure 1.4), which releases a low mass diagnostic ion by MS/MS.

In spite of the growing interest in novel crosslinking approaches to profile protein-protein interactions and the commercial availability of a wide array of crosslinkers, no researcher has successfully performed crosslinking on a proteome-wide scale. The limitations as stated above, arise mainly from the intrinsic complexity of crosslinking mixtures. It is often the case that these mixtures contain large amounts of undesirable or unexpected products,[47] in addition to undesirable linkages of crosslinkers such as dead-ends. James Bruce described the data analysis process of crosslinking reactions as looking for a needle in a haystack. The complexity of the crosslinking reaction and digestion mixtures however, is only half of the challenge in a crosslinking approach. The other half of the battle arises in the interpretation of the complex MS/MS spectra of crosslinked peptides.

Mass Defect Label Technology. In order to circumvent some of the issues of the conventional crosslinking experiments, mass defect label technology can be incorporated into the crosslinker design. "Mass defect" is defined as the difference between the exact mass of a compound and its nominal molecular weight calculated using integer atomic masses. Peptides are primarily composed of the elements from the first two rows of the periodic table, most of whose mass defects fall into the range of ± -0.008 amu (C, 0.0 mmu; H, 7.8 mmu; O, -5.1 mmu; N, 3.1 mmu; S, -28.0 mmu (Figure 1.6). In contrast to small molecules, peptides have significant mass defects due to the numerous atoms present; however, the distribution of the peptides mass

defects is generally narrow (Figure 1.5). The peptide molecular weights occupy only one-third of a unit mass at any given nominal mass. Mass defect labeling (MDL), is based on the derivatization of selected amino acids with a reagent that alters their mass defects (Figure 1.5.). Mass defect labeling of the chemical crosslinker allows for rapid, straightforward identification of crosslinked products from unreacted peptides. Crosslinked peptides will have a mass defect label that shifts the molecular weight to a region of the mass scale that is not occupied by unlabeled peptides. In this approach, an element with a large mass defects, such as bromine (-82.0 mmu), are incorporated into the crosslinker to serve as the mass defect label. Hernandez *et al.*[48] recently reported a novel cysteine mass defect label, 2,4-dibromo-(2'-iodo)acetanilide, and its applications to the shotgun proteomic analysis of *Methanococcus maripaludis* using HPLC-MALDI-FTICR mass spectrometry. They demonstrated the effectiveness of this approach for improving the specificity of protein identification. Incorporating two bromine molecules into a chemical crosslinker allows for the direct identification of the crosslinked peptides on the basis of their mass defect.

Incorporating labile bonds within crosslinker. Conventional crosslinkers contain crosslinking sections that are not easily broken during low energy dissociation methods. This poses a challenge to the researchers attempting to determine the amino acids involved in a particular interaction. In this case, MS/MS analysis of a crosslinked peak results in overlapping

series of amino acids, arising from both peptides that were crosslinked. Due to the fact that the peptide bonds are more labile than the crosslinker bonds, fragmentation occurs on each peptide, making spectral analysis difficult. However, incorporating labile bonds within the crosslinker reagent, allows for fragmentation of the crosslinker away from the peptides of interest and then subsequent isolation and fragmentation of the peptides alone, allowing for amino acid sequence information to be obtained.

Electrospray Ionization Mass Spectrometry. To test the effectiveness of this new type of mass spectrometric chemical crosslinker which incorporates mass defect label technology, electrospray ionization (ESI) mass spectrometry will be employed. The development of electrospray for ionizing macromolecules dates to the 1960s, when Malcolm Dole developed a new method to generate small droplets containing single macromolecules, due to a need for a method to analyze high molecular weight biomolecules.[11,49-53] John Fenn then applied this idea and coupled it with a mass spectrometer several decades later,[54] and introduced it to the public in 1988 at a symposium in San Francisco, as the first electrospray ionization source. Today, electrospray ionization is widely used by all types of scientists to obtain detailed information concerning the molecular weights and structures of biological macromolecules and other types of samples. The versatility of this ionization technique allows it to couple with many different

separation techniques and therefore allows for the choice of on-line separation/analysis of analytes.[55,56]

The electrospray process is quite simple. A sample is dissolved in an aqueous / polar organic solvent and pushed through a capillary which is held at a high potential. This high electric field creates an elongated tip of liquid at the exit of the capillary known as a "Taylor cone". At the tip of the cone a mist of highly charged droplets emerges, which subsequently evaporate to produce lone ions that are then analyzed by the mass spectrometer.[57-62] Although many researchers have studied this phenomenon, questions still remain concerning the actual mechanism, which explains how the charged droplets evaporate to ultimately produce gas-phase ions.[53,57,63,64] Malcolm Dole's ion evaporation model and Iribarne and Thomson's charge residue model are the two proposed methods that are widely accepted as possible mechanisms for the production of the gaseous ions.[53,65-67] A representation of the electrospray process can be seen in Figure 1.7.

During the electrospray process, the ions observed by the mass spectrometer are typically ionized by the addition of a proton, or other cation, or the removal of a proton. Therefore, this technique is capable of producing multiply charged ions, which enables the analysis of large biomolecules at relatively low mass-to-charge ratios.[50,62,68,69] This ability also allows for the analysis of analytes by tandem mass spectrometry (MS/MS).[70-74] Another feature of electrospray

ionization is that its sensitivity is enhanced by lowering the flow from μL/min rates to nL/min rates, and a number of researchers have focused on achieving sensitivities of attomole levels using this feature.[55,75-77] This process is commonly referred to as "nanospray", and it is often more tolerant of liquid compositions, allowing researchers more flexibility in the solvent system during analysis.[76,77] In general both "conventional" and "nanospray" electrospray ionization techniques have the following advantages: (1) Soft ionization, which for the most part does not degrade the molecule during the ionization process, (2) multiple charging which allows for the detection of high mass compounds at lower *m/z* ranges, (3) it readily couples to liquid separation techniques for on-line analysis, (4) enables MS/MS analysis of high MW molecules and (5) low chemical background leads to high detection limits. Like all analytical techniques, ESI does have its inherent disadvantages which include; (1) presence of buffers and salts reduces sensitivity, (2) multiple charging introduces a level of complexity within the spectra, and (3) ion suppression can occur, which requires a second level of separation, often on-line.

Tandem Fourier Transform Mass Spectrometry (MS/MS). Tandem mass spectrometry (MS/MS) allows the researcher greater insight into the compound of interest within the original mass spectrum. It is based on the selection of a specific precursor ion from within the many peaks present in the MS, followed by its dissociation or fragmentation further down the instrument to produce smaller ions that are then mass analyzed.[78] MS/MS has become a very

powerful and useful technique for the analysis of large biomolecules, and for the identification of protein sequence information. It also increases the specificity during database searching.[79] Timely advances in this technique have allowed for the development of new applications targeted to the biochemistry and medicinal fields. These include the identification of post-translation modifications and binding sites within large multi-subunit complexes, in addition to sequence information and the identification of unknown proteins.[78-80] A MS/MS experiment can be completed either by space or by time; tandem MS in space involves the coupling of two or more mass analyzers, and tandem MS in time is based on the sequence of events being performed in an ion storage or trap device. In a Fourier transform mass spectrometer, MS/MS can be performed as a series of events in the analyzer cell, due to the efficiency of the magnetic ion trap, which can be manipulated in a number of different ways.

There are several ion dissociation technique for fragmentation of molecules used in an FT-MS/MS experiment including, sustained off-resonance excitation CID (SORI-CID),[81,82] on-resonance CID (RE-CID),[83] infrared multiphoton dissociation (IRMPD),[84,85] blackbody infrared radiative dissociation (BIRD),[86] electron capture dissociation (ECD),[87] and electron detachment dissociation (ECD).[88,89] These techniques all fragment the selected ions in a different fashion, by either low or high energy fragmentation pathways, and generate distinct product ions useful for obtaining detailed structural features of the molecules analyzed.[82,90]

Research Goals: The aim of this research project was to develop a new mass spectrometry identifiable crosslinker that incorporates a mass defect label, which will distinguish crosslinked peptides from simple tryptic fragments by their mass defect. Therefore, the identification of crosslinked species will be possible by a MS measurement rather than MS/MS. This research employed a 'bottom-up' approach and novel crosslinker to facilitate the identification of interaction proteins and their sites of interaction. These crosslinkers were not only designed to distinguish themselves via their mass defect, but also designed to facilitate the release of the attached peptides by low energy collisionally activated decomposition (CAD), allowing further fragmentation by MS^3. This fragmentation will provide sequence information that can be used to identify the protein from which the peptide derived and the location of the peptide in the protein. In order to design such a crosslinker, we positioned a tertiary amine in the center of the crosslinker with activated NHS esters on either side for attachment.

Firstly a novel chemical crosslinker was applied to small peptides neurotensin and bradykinin. While this chemical crosslinking reagent possessed the ability to crosslink the small peptides used in preliminary experiments and dissociated effectively, its diameter was too small to serve as an effective crosslinker, as the spacing between the activated ester groups was only 0.6 nm. The synthesis of the following mass defect labeled crosslinkers can be found in Chapter 2 however, it should be noted that it contains the mass defect label, a tertiary amine and spacing

between reactive groups of approximately 12 Å. In order to test these novel crosslinkers, their effectiveness at crosslinking neurotensin, bradykin and Ribonuclease S was studied before applying them to the complicated biological system of 34 kDa protein and Actin.

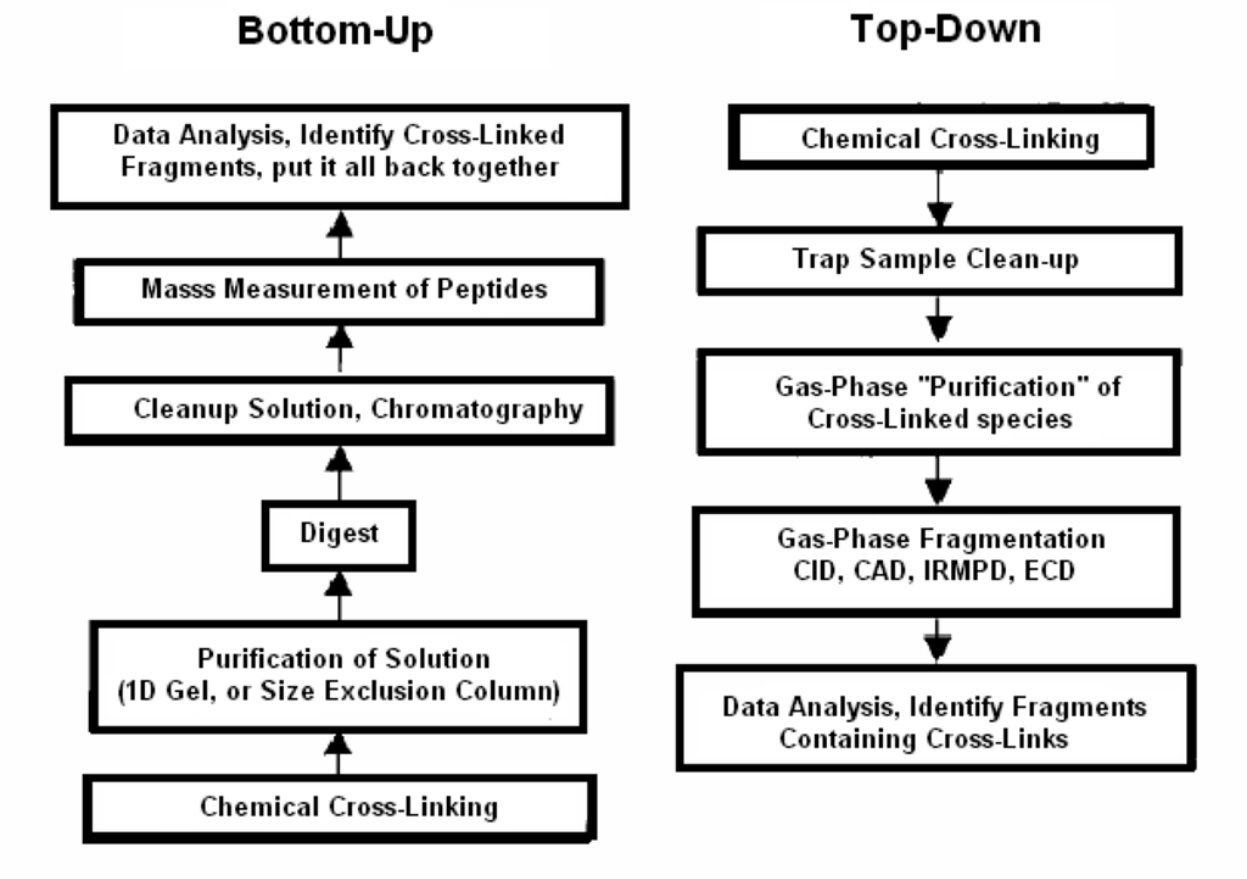


Figure 1.1. Comparison between Bottom-up Approach and Top-Down Approach in a chemical crosslinking study.

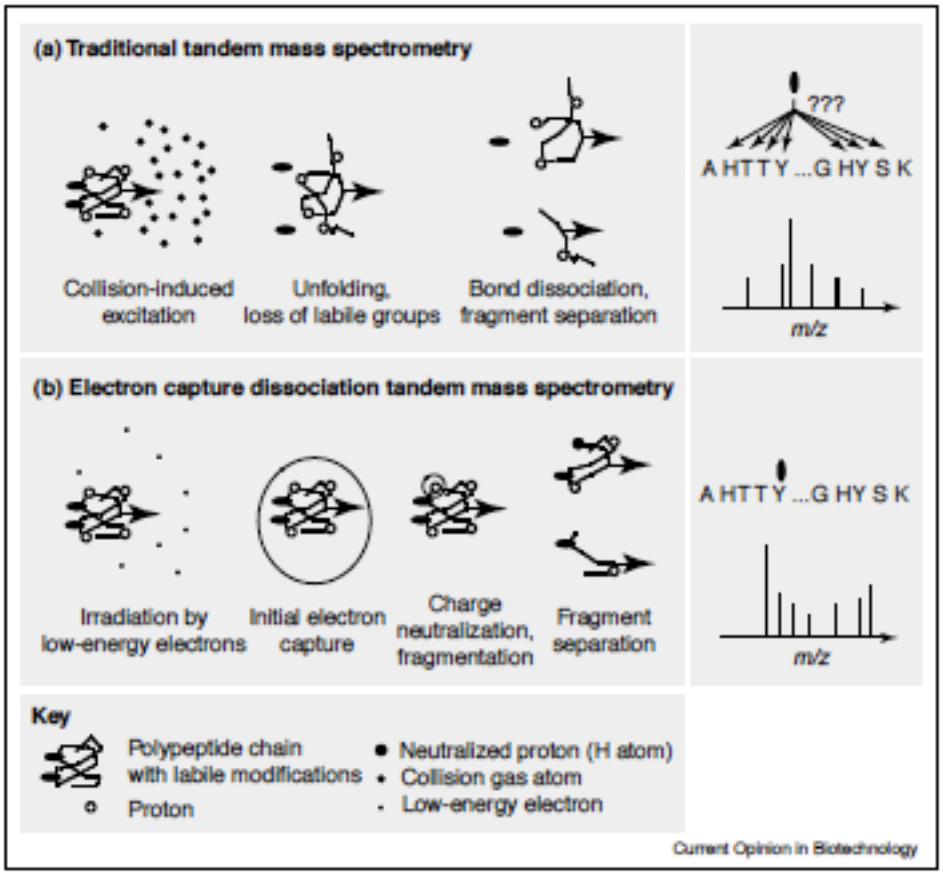


Figure 1.2. Representation of Electron Capture Dissociation.[91]

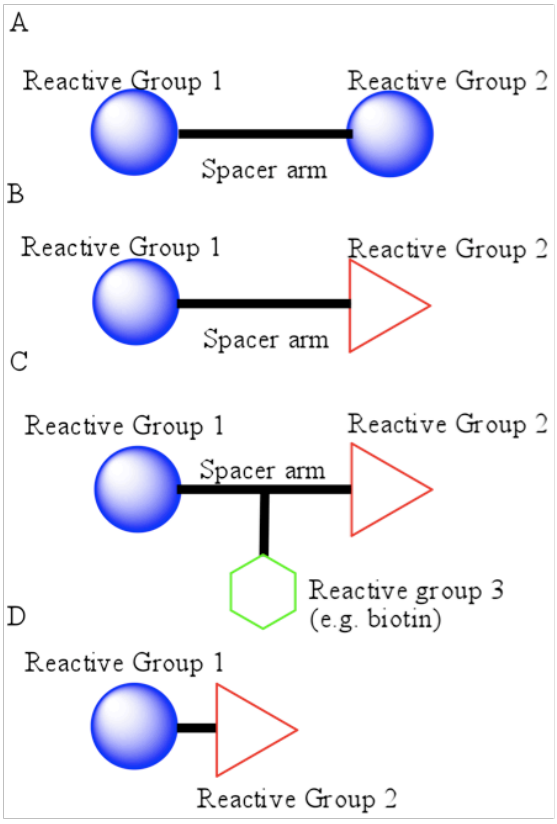


Figure 1.3. Comparison of the different types of chemical crosslinkers that are commercially available.

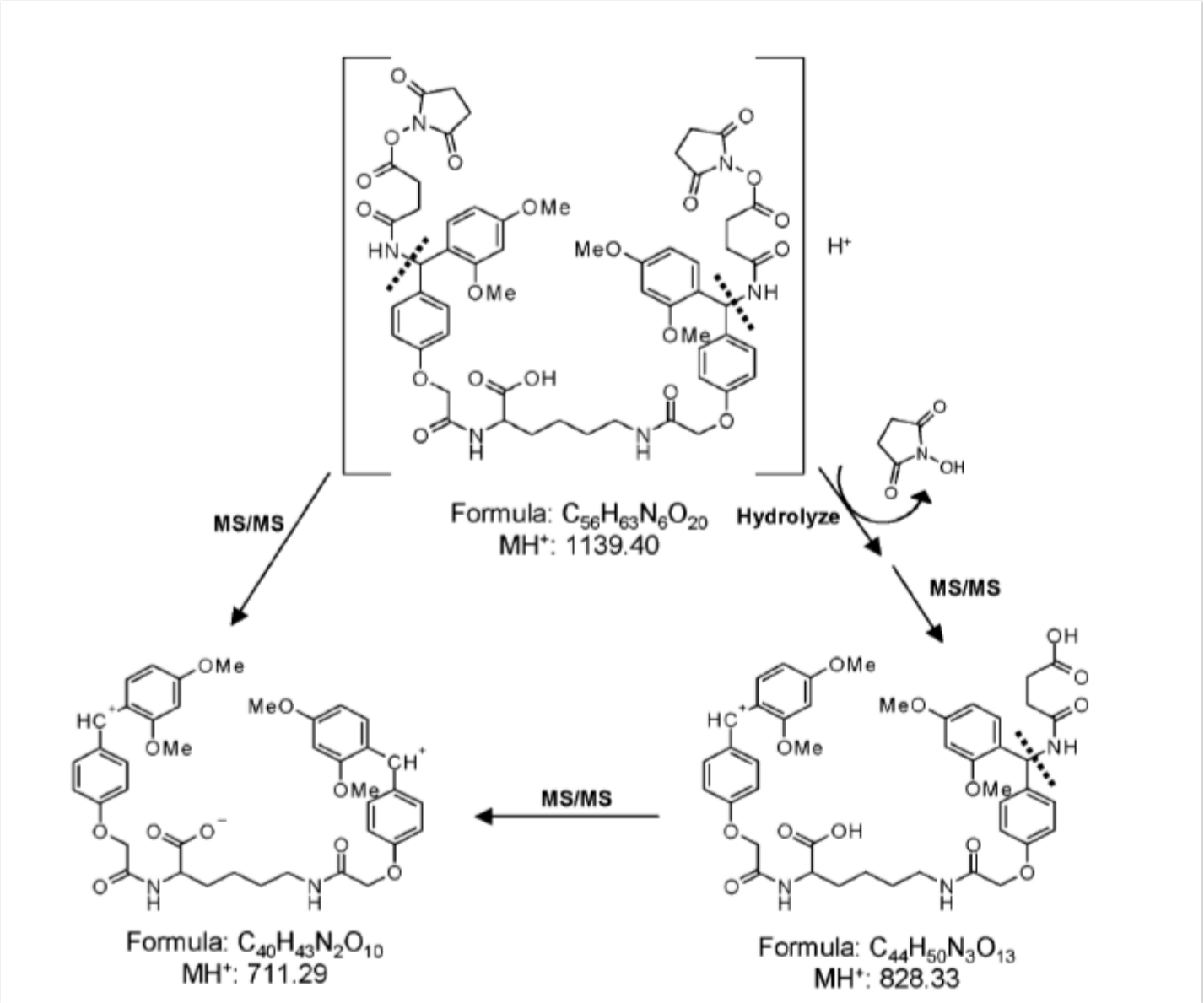


Figure 1.4. Tang *et al.*'s PIR crosslinker.[31]

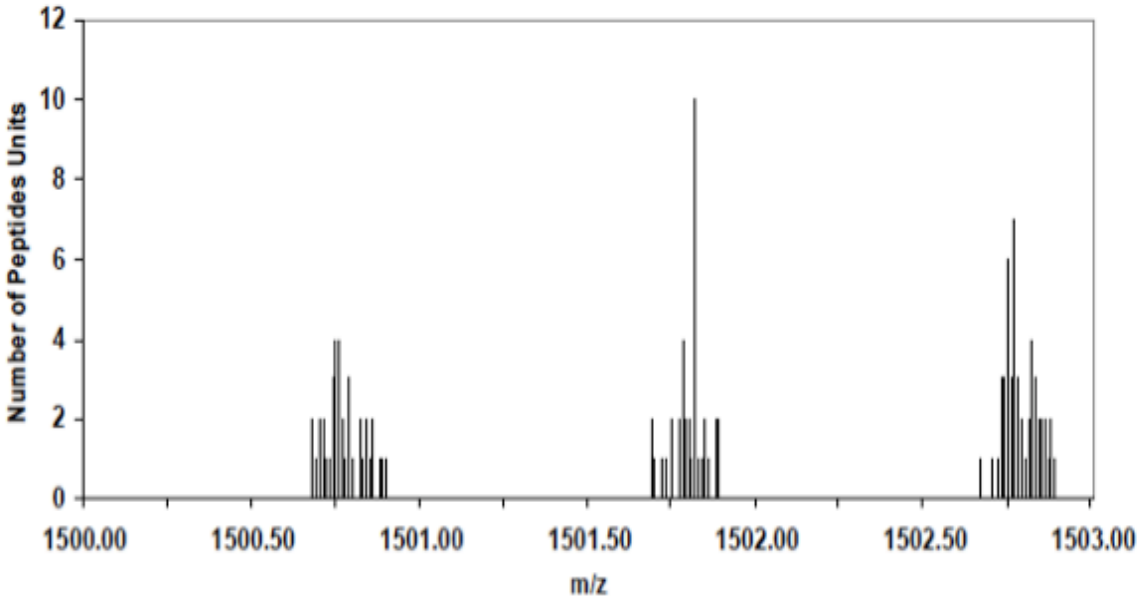


Figure 1.5. Histogram of molecular weight distribution of the predicted tryptic peptides of M. *maripaludis* over the range of 1500 to 1503 Da, illustrating the distribution of mass defect of peptides. The bin size is 0.1 amu. Peptide masses are observed to cluster in approximately one third of the available mass space.

Element	Mass Defect (amu)
^{12}C	0
^{1}H	0.0078
^{16}O	-0.0051
^{15}N	0.0031
^{32}S	-0.0279

Figure 1.6. Conventional mass defects. Mass difference from neuclon value of the most abundant isotope of the elements found in proteins.

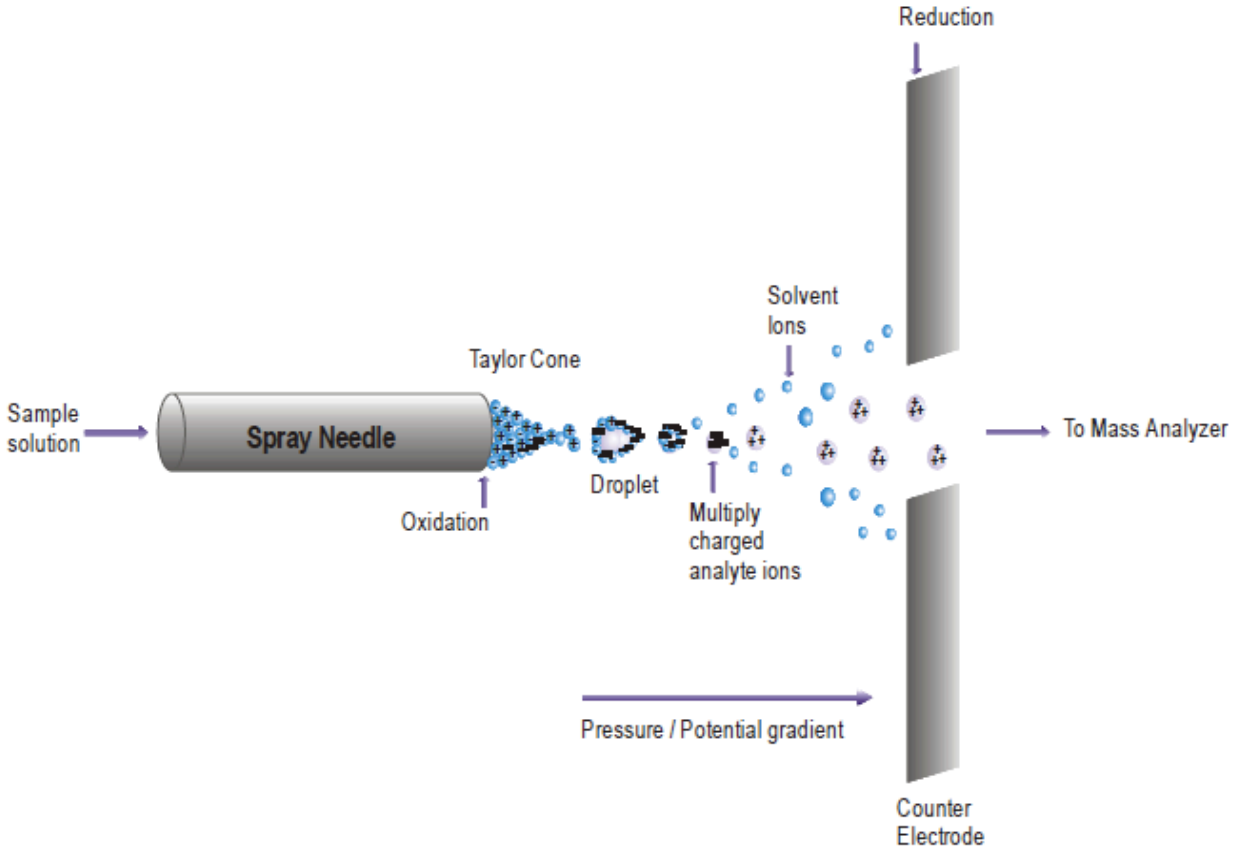


Figure 1.7. Schematic of electrospay process.[11]

References:

1. Trakselis, M. A.; Alley, S. C.; Ishmael, F. T. *Bioconjugate Chem.* **2005**, *16*, 741.

2. Sinz, A. *J. Mass Spectrom.* **2003**, *38*, 1225.

3. Chiu, W.; Baker, M. L.; Jiang, W.; Dougherty, M.; Schmid, M. F. *Structure* **2005**, *13*, 363.

4. Rostom, A. A.; Fucini, P.; Benjamin, D. R.; Juenemann, R.; Nierhaus, K. H.; Hartl, F. U.; Dobson, C. M.; Robinson, C. V. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5185.

5. van Duijn, E.; Simmons, D. A.; van den Heuvel, R. H. H.; Bakkes, P. J.; van Heerikhuizen, H.; Heeren, R. M. A.; Robinson, C. V.; van der Vies, S. M.; Heck, A. J. R. *J. Am. Chem. Soc.* **2006**, *128*, 4694.

6. McKay, A. R.; Ruotolo, B. T.; Ilag, L. L.; Robinson, C. V. *J. Am. Chem. Soc.* **2006**, *128*, 11433.

7. Kluger, R.; Alagic, A. *Bioorg. Chem.* **2004**, *32*, 451..

8. Ihling, C.; Berger, K.; Hofliger, M. M.; Fuhrer, D.; Beck-Sickinger, A. G.; Sinz, A. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1240.

9. Kruppa, G. H.; Schoeniger, J.; Young, M. M. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 155.

10. Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299.

11. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.

12. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1.

13. McLafferty, F. W.; Fridriksson, E. K.; Horn, D. M.; Lewis, M. A.; Zubarev, R. A. *Science* **1999**, *284*, 1289.

14. Novak, P.; Haskins, W. E.; Ayson, M. J.; Jacobsen, R. B.; Schoeniger, J. S.; Leavell, M. D.; Young, M. M.; Kruppa, G. H. *Anal. Chem.* **2005**, *77*, 5101.

15. Back, J. W.; de Jong, L.; Muijsers, A. O.; de Koster, C. G. *J. Mol. Biol.* **2003**, *331*, 303.**331**, 303 (Aug, 2003).

16. Comisaro.Mb; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *25*, 282.

17. Sinz, A. *Anal. Bioanal. Chem.* **2005**, *381*, 44.

18. Sinz, A. *Mass Spectrom. Rev.* **2006**, *25*, 663.

19. Sinz, A.; Wang, K. *Biochemistry* **2001**, *40*, 7903.

20. McLafferty, F. W. *Int. J. Mass Spectrom.* **2001**, *212*, 81.

21. Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5802.

22. Vasilescu, J.; Figeys, D. *Curr. Opin. Biotechnol.* **2006**, *17*, 394.

23. Dihazi, G. H.; Sinz, A. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2005.

24. Schulz, D. M.; Ihling, C.; Clore, G. M.; Sinz, A. *Biochemistry* **2004**, *43*, 4703.

25. Avrameas, S. *Immunochemistry* **1969**, *6*, 43.

26. Fujii, N.; Jacobsen, R. B.; Wood, N. L.; Schoeniger, J. S.; Guy, R. K. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 427.

27. Trester-Zedlitz, M.; Kamada, K.; Burley, S. K.; Fenyo, D.; Chait, B. T.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 2416.

28. Bragg, P. D.; Hou, C. *Arch. Biochem. Biophys.* **1975**, *167*, 311.

29. Lomant, A. J.; Fairbanks, G. *J. Mol. Biol.* **1976**, *104*, 243.

30. Staros, J. V. *Biochemistry* **1982**, *21*, 3950.

31. Tang, X. T.; Munske, G. R.; Siems, W. F.; Bruce, J. E. *Anal. Chem.* **2005**, *77*, 311.

32. Smyth, D. G.; Konigsberg, W.; Blumenfeld, O. O. *Biochem. J.* **1964**, *91*, 589.

33. Gorin, G.; Matic, P. A.; Doughty, G. *Arch. Biochem. Biophys.* **1966**, *115*, 593.

34. Heitz, J. R.; Anderson, C. D.; Anderson, B. M. *Arch. Biochem. Biophys.* **1968**, *127*, 627.

35. Partis, M. D.; Griffiths, D. G.; Roberts, G. C.; Beechey, R. B. *J. Protein Chem.* **1983**, *2*, 263.

36. Brunner, J. *Annu. Rev. Biochem.* **1993**, *62*, 483.

37. Sinz, A.; Kalkhof, S.; Ihling, C. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1921.

38. Suchanek, M.; Radzikowska, A.; Thiele, C. *Nat. Methods* **2005**, *2*, 261.

39. Alley, S. C.; Ishmael, F. T.; Jones, A. D.; Benkovic, S. J. *J. Am. Chem. Soc.* **2000**, *122*, 6126.

40. Collins, C. J.; Schilling, B.; Young, M. L.; Dollinger, G.; Guy, R. K. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4023.

41. Muller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. *Anal. Chem.* **2001**, *73*, 1927.

42. Taverner, T.; Hall, N. E.; O'Hair, R. A. J.; Simpson, R. J. *J. Biol. Chem.* **2002**, *277*, 46487.

43. Bennett, K. L.; Kussmann, M.; Bjork, P.; Godzwon, M.; Mikkelsen, M.; Sorensen, P.; Roepstorff, P. *Protein Sci.* **2000**, *9*, 1503.

44. Schmidt, A.; Kalkhof, S.; Ihling, C.; Cooper, D. M. F.; Sinz, A. *Eur. J. Mass Spectrom.* **2005**, *11*, 525.

45. Hurst, G. B.; Lankford, T. K.; Kennel, S. J. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 832.

46. Back, J. W.; Sanz, M. A.; De Jong, L.; De Koning, L. J.; Nijtmans, L. G. J.; De Koster, C. G.; Grivell, L. A.; Van der Spek, H.; Muijsers, A. O. *Protein Sci.* **2002**, *11*, 2471.

47. Swaim, C. L.; Smith, J. B.; Smith, D. L. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 736.

48. Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J. *Anal. Chem.* **2006**, *78*, 3417.

49. Fenn, J. B. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 524.

50. Cole, R. B. **1997**..

51. Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88*, 4451.

52. Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88*, 4671.

53. Dole, M.; Mack, L. L.; Hines, R. L. *J. Chem. Phys.* **1968**, *49*, 2240.

54. Meng, C. K.; Mann, M.; Fenn, J. B. *Proceedings of the 36th American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics, San Francisco, CA* **1988**.

55. Gaskell, S. *J. Mass Spectrom.* **1997**, *32*, 1378.

56. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Mass Spectrom. Rev.* **1990**, *9*, 37.

57. Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20*, 362.

58. de la Mora, J. F.; Van Berkel, G. J.; Enke, C. G.; Cole, R. B.; Martinez-Sanchez, M.; Fenn, J. B. *J. Mass Spectrom.* **2000**, *35*, 939.

59. Loo, J. A.; Udseth, H. R.; Smith, R. D. *Anal. Biochem.* **1989**, *179*, 404.

60. Ikonomou, M. G.; Blades, A. T.; Kebarle, P. *Anal. Chem.* **1990**, *62*, 957.

61. Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. *Anal. Chem.* **1990**, *62*, 882.

62. Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M.; Udseth, H. R. *Mass Spectrom. Rev.* **1991**, *10*, 359.

63. Enke, C. G. *Anal. Chem.* **1997**, *69*, 4885.

64. Berkel, G. J. V.; Zhou, F.; Aronson, J. T. *Int. J. Mass Spectrom. Ion Processes* **1997**, 55.65. J. V. Iribarne, B. A. Thomson, *Journal of Chemical Physics* **64**, 2287 (1976).

66. Cooks, R. G.; Ouyang, Z.; Takats, Z.; Wiseman, J. M. *Science* **2006**, *311*, 1566.

67. Kebarle, P. *J. Mass Spectrom.* **2000**, *35*, 804.

68. Henry, K. D.; Williams, E. R.; Wang, B. H.; McLafferty, F. W.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 9075.

69. Bruce, J. E.; Anderson, G. A.; Udseth, H. R.; Smith, R. D. *Anal. Chem.* **1998**, *70*, 519.

70. M. A. Mendes, B. M. de Souza, L. D. dos Santos, M. S. Palma, *Rapid Communications in Mass Spectrometry* **18**, 636 (2004).

71. Mendes, M. A.; de Souza, B. M.; dos Santos, L. D.; Santos, K. S.; Palma, M. S. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 2636.

72. Edmonds, C. G.; Loo, J. A.; Loo, R. R. O.; Udseth, H. R.; Barinaga, C. J.; Smith, R. D. *Biochem. Soc. Trans.* **1991**, *19*, 943.

73. Chitta, R. K.; Gross, M. L. *Biophys. J.* **2004**, *86*, 473.

74. Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Science* **1990**, *248*, 201.

75. Valaskovic, G. A.; Kelleher, N. L.; Little, D. P.; Aaserud, D. J.; McLafferty, F. W. *Anal. Chem.* **1995**, *67*, 3802.

76. Jurascheka, R.; Dulcksa, T.; Kara, M. *J. Am. Soc. Mass Spectrom.* **1999**, 300.

77. Wilm, M. S.; Mann, M. *Int. J. Mass Spectrom.* **1994**, *136*, 167.

78. Williams, E. R. *Anal. Chem.* **1998**, *70*, 179A.

79. Roepstorff, P. *Curr. Opin. Biotechnol.* **1997**, *8*, 6.

80. Schwartz, J. C.; Wade, A. P.; Enke, C. G.; Cooks, R. G. *Anal. Chem.* **1990**, *62*, 1809.

81. Senko, M. W.; Speir, J. P.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2801.

82. Herrmann, K. A.; Somogyi, A.; Wysocki, V. H.; Drahos, L.; Vekey, K. *Anal. Chem.* **2005**, *77*, 7626.

83. Cody, R. B.; Freiser, B. S. *Anal. Chem.* **1982**, *54*, 1431.

84. Zhang, J. H.; Schubothe, K.; Li, B. S.; Russell, S.; Lebrilla, C. B. *Anal. Chem.* **2005**, *77*, 208.

85. Little, D. P.; Speir, J. P.; Senko, M. W.; Oconnor, P. B.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2809.

86. Price, W. D.; Schnier, P. D.; Williams, E. R. *Anal. Chem.* **1996**, *68*, 859.

87. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72*, 563.

88. Budnik, B. A.; Haselmann, K. F.; Zubarev, R. A. *Chem. Phys. Lett.* **2001**, *342*, 299.

89. Nielsen, M. L.; Budnik, B. A.; Haselmann, K. F.; Zubarev, R. A. *Int. J. Mass Spectrom.* **2003**, *226*, 181.

90. Domon, B.; Aebersold, R. *Science* **2006**, *312*, 212.

91. Zubarev, R. A. *Curr. Opin. in Biotech.* **2002,** *15,* 12.

CHAPTER 2:

Experimental

This chapter will outline the methodologies and instrumentation that were used to perform the forthcoming experimental procedures. The results of such experiments will be discussed in the following chapters. Specifically this chapter describes the methods for preparing the novel mass defect labeled chemical crosslinkers, preparing Actin and 34 kDa protein for analysis, the crosslinking of specific protein systems, performing proteolytic digestion with and without SDS-PAGE gel-electrophoresis, HPLC separation of peptides, HPLC-ESI-FTMS/MS experiments and the utilization of MS3D[1] data analysis software.

Preparation of novel mass defect labeled crosslinkers. All reagents for the following crosslinker reaction scheme were purchased at Sigma-Aldrich. To prepare the crosslinker, DiBromoMethylAnilineDiPropionateSuccinimide (DiMADPS) was synthesized by mixing acrylic acid with 3,5-dibromo-4-methylaniline, followed by activation with N-hydroxysuccinimide. Briefly acrylic acid (0.65 mmol) was mixed with 0.65 mmol of thiourea and 200 μL of water. The mixture was heated for 2 hours at 99 °C, and 100 μL of 40% (w/w) NaOH was then added and incubated for 1 hour at 99 °C in a thermomixer (Eppendorf). Each

preparation was adjusted to pH ~8 with HCl, and 0.1 mmol of 3,5-dibromo-4-methylaniline (Calbiochem) was added. The solution was mixed until completely dissolved and kept at room temperature for 1 hour. The pH was adjusted to ~8 by addition of NaOH. Bis(2,5-dioxopyrrolidin-1-yl)3,3'-(3,5-dibromo-4-methylphenylazanediyl)diproponate (DiBMADPS) was precipitated by addition of HCl, filtered, washed with 1 mM of HCl, and lyophilized. DiBMADPS was then dissolved in DMSO, and activated with N-hydroxysuccinimide and dicyclohexylcarbodiimide overnight at room temperature. The reaction mixture was then filtered, and DiBMADPS was precipitated by the addition of water, then washed with water, and lyophilized. DiBMADPS was characterized by nano-ESI (QSTAR, Applied Biosystems, Framingham, MA) under the experimental conditions and parameter settings described elsewhere.[2] The synthesis scheme can be seen in Figure 2.1. A 1:1 mixture of DiBMADPS crosslinker to protein in DMSO, was then used for all crosslinking reactions. For storage, 10 mM stock solutions were kept in DMSO at -20 °C and were added directly to a solution of the protein in phosphate buffered saline (PBS) to give a final concentration as described below.

To prepare the second crosslinker, bis(2,5-dioxopyrrolidin-1-yl) 2,2'-(2,4-dibromophenylsulfonylazanediyl)diacetate (DiBBSIAS), 0.1 mmole of dibromobenzenesulfonylchloride was reacted with 0.1 mmole of iminodiacetic acid and 0.3 mmole NaOH in 100 µL of water at 75 ºC for 24 hours. The reaction product was purified using

reversed-phase chromatography using a C_{18}-Sep-Pak (Millipore) and was dried *in vacuo*. 11 mg of dibromobenzenesulfoneiminodiacetic acid was then activated by N-hydroxysuccinimide in the presence of dicyclohexylcarbodiimde, as described before, in dimethylsulfoxide. The final product (DiBBSIAS) was then precipitated from acetone/ethyl ether, dried *in vacuo*, and stored in a 10 mM stock solution at -20 °C until use. Figure 2.2 shows the synthesis scheme.

Crosslinking of Neurotensin, Bradykinin and Ribonuclease S. 1 mg/mL of each protein (Neurotensin, Bradykinin, or Rnase S) was dissolved in 100 mM phosphate buffered saline (PBS) at a pH of 8.0, and added to 0.5 mM DiBMADPS for 30 minutes at 25 °C. The reaction was quenched by addition 50 mM ammonium bicarbonate (ABC), and Rnase S containing species were separated by SDS-PAGE or HPLC. Neurotensin and Bradykinin solutions were simply analyzed by direct infusion ESI-FTICR-MS/MS. For SDS-PAGE experiments, protein bands containing crosslinker were visualized using a UV illuminator with a 362-nm excitation filter. The fluorescent band, corresponding to crosslinked Rnase S was excised and in-gel digestion with porcine trypsin (sequencing grade, Promega) according to the protocol described elsewhere,[3] separated by HPLC, and analyzed by ESI-FTICR-MS/MS as described below.

Preparation of Actin and 34 kDa Proteins. The preparation and purification of actin from rabbit psoas and leg muscles was completed following published procedures.[4,5] The purified G-actin could be maintained in G-Actin buffer (2 mM Tris-HCl, pH 8.0, 0.2 mM $CaCl_2$,

0.2 mM ATP, 0.2 mM DTT, and 0.02% NaN_3) at 4 °C for up to one week if the dialysis buffer is changed daily. F-Actin was produced from G-Actin by polymerization in a high salt buffer composition of 50 mM KCl, 1 mM ATP, and 1 mM $MgCl_2$. The actin was then spun in an ultracentrifuge at 100,000 x g for 2 hours at 4 °C. The F-actin pellet that was produced was then depolymerized in a small volume of G-actin buffer and maintained for up to 1 week in dialysis buffer. 24 hours prior to use in a crosslinking experiment, the actin was exchanged into a buffer composed of 20 mM imidazole pH 7.5, 0.2 mM $CaCl_2$, and 0.2 mM ATP.

34 kDa protein was expressed in the E. coli expression strain BL21(DE3) using the pET-15b vector system (Novagen EMD), and purification of recombinant protein was done as previously described.[6,7] Protein concentrations were determined by the BCA method[8] and protein samples were dialyzed against 2 mM PIPES pH 7.0, 5.0 mM KCl, 0.2 mM DTT, and 0.02% NaN_3, and stored at -80 °C.

Crosslinking of Actin and 34 kDa protein systems. In order to study the effectiveness of the novel mass defect labeled crosslinker, DiBMADPS, comparison studies were also completed using the zero-length crosslinker EDC (1-Ethyl-3[3-dimethylaminopropyl]carbodiimide Hydrochloride) (Pierce Biotechnology, Rockford, IL) and the 11.4 Å amine reactive, homobifunctional, water-soluble crosslinker BS^3 (Bis[sulfosuccinimidyl] suberate) (Pierce Biotechnology, Rockford, IL). 18 µM actin was reacted alone or with 6 µM 34 kDa protein

overnight at 4 °C in 20 mM PIPES pH 7.0, 5.0 mM KCL, 1 mM $MgCl_2$, 1 mM ATP, and 5 mM EDTA. Freshly prepared solutions of 10 mM BS^3 in H_2O, 5 mM EDC/NHS in DMSO, or 10 mM DiBMADPS in DMSO were used for the crosslinking experiments. The final concentration of the crosslinker in the reaction mixture was dependent on the total molar concentration of the protein added. The final crosslinker to protein concentrations were as follows: BS^3 20:1, EDC/NHS 25:1, and DiBMADPS 10:1. All reactions were incubated at room temperature for a total of 60 minutes and quenched using 1 M Tris-HCl at pH 8.0 to produce a final concentration of 10 mM. After quenching the samples, they were prepared for SDS-PAGE analysis by dilution with 4X reducing Laemmli sample buffer and boiled for 5 minutes. Samples were separated by SDS-PAGE9[2] and stained with Coomassie blue. The same procedure was used for both mass defect labeled crosslinkers.

Western blot analysis. This procedure was performed in order to identify the components generated by the chemical crosslinking reaction. Procedural steps were taken from previously published literature.[10] In order to identify the species present, either an anti-34 kDa monoclonal antibody (B2C)[11] or an anti-actin monoclonal antibody (MabGEa; produced by Dr. Rich Meagher, UGA)[12] and a horseradish peroxidase-conjugated goat anti-mouse antibody was used. Visualization of immunoreactive bands was accomplished by chemiluninescence (Pierce Biotechnology, Rockford, IL).

Procedure for in-gel trypsin digestion. The procedure for digestion of the gel bands of interest can be found elsewhere[13] and these procedures were followed with slight modifications. The gel bands were excised using a clean razor blade, cut into 1 mm^3 pieces and placed in an eppendorf tube that was previously cleaned with methanol. The individual pieces were then destained using several aliquots of 50 mM NH_4HCO_3 and 50% methanol. After the pieces were completely destained, the final solution was removed, and the gel pieces were dehydrated in 75% acetonitrile for 20 minutes at room temperature. After removing the acetonitrile solution, the gel plugs were dried at 40 °C. Reduction of the disulfide bonds was completed by inserting the dried gel plugs in a solution of 10 mM DTT, 20 mM NH_4HCO_3 for 45 minutes at 56 °C. This solution was then decanted off the gel plugs and a solution of 100 mM iodoacetamide, 20 mM NH_4HCO_3 was added in order to alkylate the cysteines. This solution was incubated in the dark for 30 minutes at room temperature. The gel plugs were then washed with 50% methanol in 50 mM NH_4HCO_3 for 20 minutes at room temperature. This solution was decanted and the gel pieces were again incubated, this time with 75% acetonitrile for 20 minutes at room temperature. The gel pieces were then dried at 40 °C. Dried gel plugs were then processed in-gel digestion using 100 μL of trypsin (20 μg/mL) in 20 mM NH_4HCO_3 at 37 °C for 12 hours. The trypsin-digested peptides were extracted from the gel plugs by incubation in 50% acetonitrie/0.1% TFA at room temperature for 20 minutes, and then centrifuged for 30 seconds at low speed. The resulting

supernatant was removed and saved in a clean eppendorf tube. A second extraction was performed on the gel plugs, and the supernatants from both extractions were pooled and dried *in vacuo* to < 5 µL. These final peptide solutions were volume adjusted to 10-15 µL in 50% acetonitrile/0.1% TFA and stored at -80 °C.

Mass Spectrometric Analysis. Experiments were performed with a 9.4 T Bruker Apex QhFTMS (Billerica, MA) fitted with an Apollo II dual source, a 25 W CO_2 laser (Synrad model J48-2, Mukilteo, WA) for IRMPD, and an indirectly heated hollow cathode for generating electrons for ECD experiments. Bruker Apex Control was used for data acquisition. Bruker Data Analysis (version 3.4 Build 72) was used for most of the data analysis work. Protein samples were dissolved in a solution of 50% Methanol:50% H_2O:0.1% Formic acid in order to obtain the desired charge states and intensity. Tryptically digested peptides can be ionized by either regular flow ESI or nanospray. For regular-flow ESI, sample solutions were infused at a rate of 120 µL/hour. For nanospray, samples were infused at a rate of 10 µL/hour using a pulled fused silica tip (model #FS360-20-10-D-20, New Objective, Woburn, MA). Solution conditions would vary based on the amount of protein that was present prior to crosslinking. The only samples that were analyzed by direct infusion ESI were those not analyzed by SDS-PAGE (i.e. Neurotensin, Bradykinin and some Rnase S samples).

For infrared multiphoton dissociation (IRMPD) and electron capture dissociation (ECD) experiments, precursor ions were isolated in the external quadrupole and accumulated for 1-10 seconds before injection into the FTMS cell. The isolation/cell fill can be repeated up to 6 times. Typically, a maximum of 6 cell fills is used because the ion intensity reaches a maximum under these conditions. The selection of the precursor ion was further refined by using in-cell isolation, with a coherent harmonic excitation frequency (CHEF) event, which isolates the precursor and its isotopes. The precursor ions are then irradiated with either IR light, or electrons for 0.25-5 seconds. For IRMPD irradiation, the CO_2 laser is set to 70% of its full intensity. For electron irradiation the cathode bias is set to approximately -1 V, the ECD lens is set to approximately - 1 V, and the cathode heater is set to 1.6 A. With the ECD lens set to a potential higher than the cathode potential, it ensures that no electrons enter the analyzer cell. The ECD lens potential is increased by 0.1 V steps until the precursor ion intensity decreases by 75%, or approximately 1 x 10^{-7}. The same procedure is performed for IRMPD experiments, however optimization of the IR pulse length is achieved by increasing it in 0.5 seconds increments to achieve a 75% reduction in precursor intensity. Typically, it takes an IR pulse length of 3 seconds for IRMPD experiments. Ions are excited with an RF frequency chirp that covers the range of *m/z* 200-2500. 24 acquisitions were signal averaged per mass spectrum.

For each mass spectrum, 512K points were acquired at a 2.4 MHz digitization rate, appodized using a sinebell window and padded with one zero fill. Background spectra were acquired by leaving all parameters the same but setting the cathode bias to 0 V or the pulse length to 0 seconds to ensure that no electrons or IR light reached the analyzer cell. Background spectra were not subtracted from the ECD or IRMPD spectra, but were used to ensure that very few of the product ions observed were present during ion accumulation and isolation. External calibration of IRMPD and ECD mass spectra produced mass accuracy of 5 ppm. Internal calibration was performed using confidently assigned tryptic fragment ions as internal calibrants, providing mass accuracy of <1 ppm. Internal calibration was performed using Bruker Data Analysis (DA) software. Peaks were automatically picked using the "SNAP" peak picking feature in Data Analysis with the "S/N threshold" set to 0.001, the "Relative intensity threshold (base peak)" set to≤ 0.001%, and the "Absolute intensity threshold" set to 0.001. This peak-picking feature was used specifically because it is designed for peptide/protein analysis. After internal calibration of the IRMPD and ECD mass spectra, product ions were assigned using accurate mass measurement.

High Performance Liquid Chromatography Analysis (HPLC) Coupled with Tandem Mass Spectrometry. Separation of RNase S and 34 kDa - Actin solutions were analyzed using an Agilent 1100 series HPLC (Santa Clara, CA). Samples were first injected by the autosampler and

loaded on a C18 column (PepMap C18, 300 μm x 15 cm, 3 μm, 100 Å, LC Packing) at a flow rate of 0.4 μL/min with 95/5 (v/v) H_2O (with 0.1% formic acid)/acetonitrile (with 0.1% formic acid), solvent A. Peptides were then separated using the following gradient: 0% B for 0-5 mins, 20-75% B for 5-120 mins, 75-90% B for 120-125 mins, 90-95% B for 125-130 mins, 0% B for 130-140 mins, where solvent B was 0.1% formic acid in Acetonitrile. The column was directly linked to the inlet of the mass spectrometer and therefore the analysis was completed simultaneously. Auto MS/MS was set up to collect compounds from a prefer list compiled by the user and LC/MS/MS mass spectra were acquired using HyStar software (version 3.0 Bruker Daltonics). MS/MS data acquisition was set in automatic mode with active exclusion based on peak intensity and selections of prefer peak lists. Five precursor ions were selected from each MS scan and excluded after two MS/MS scans for 1 minute. Calibration of the instrument prior to any analysis was achieved by injecting a tryptic digest of BSA (Bovine Serum Albumin, Sigma, St. Louis, MO).

MS2Links Software on MS3D Web-portal. The Collaboratory for MS3D, http://www.ms3d.org, was designed by structural biologists working in the area of mass spectrometric based methods for the analysis of tertiary and quaternary macromolecular structures (MS3D). This web-portal was designed to provide collaborators with a shared working environment where data can be uploaded and analyzed. Specially, a series of programs was

designed for the analysis of MS and MS/MS data provided by substrates treated with certain chemical probes or modifications. This software consists of the Links[14] and MS2Links[15,16] programs, which were initially designed for the analysis of data obtained from bottom-up and top-down analysis of crosslinked/covalently labeled proteins and nucleic acids.[17] In this experiment, we used MS3Links and identified the specific crosslinked species formed when either EDC, BS^3, DiBMADPS or DiBBSIAS were crosslinked with Neurotensin, Bradykinin, RNase S, 34 kDA, Actin or 34-Actin complexes. The robust user interface on both the Links and MS2Links software allowed for the direct input of the protein sequences, chemical modifications (including crosslink mass and active site), and peak list. Using this interface, each peak list generated by a HPLC-MS/MS run was subjected to analysis and the peaks containing the modified crosslinker were revealed.

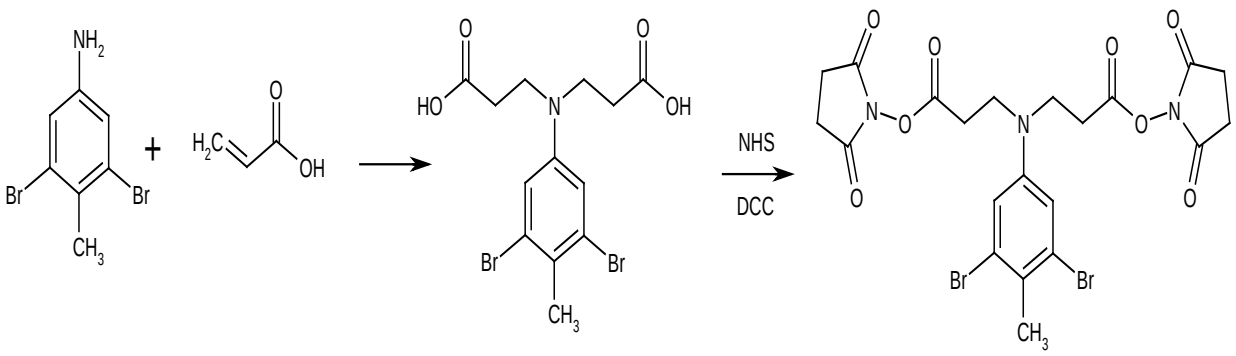


Figure 2.1. Synthesis scheme of DiBMADPS.

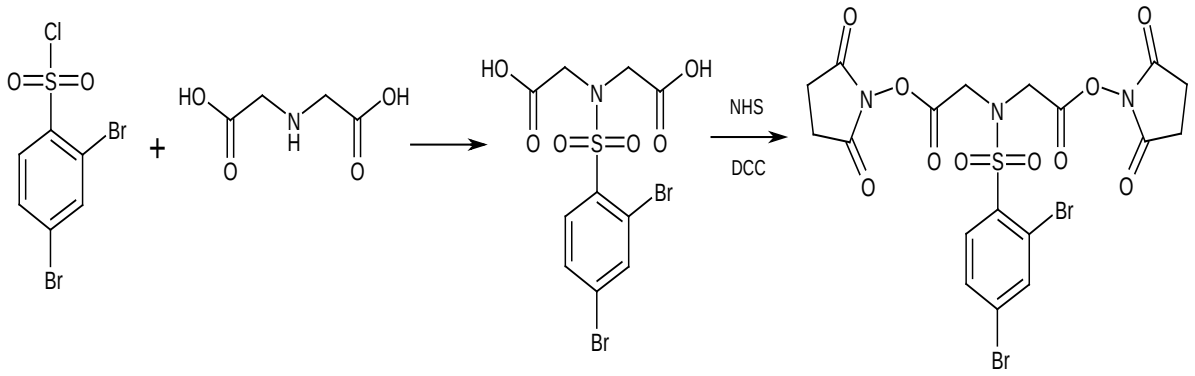


Figure 2.2. Synthesis scheme of DiBBSIAS.

References:

1. Yu, E. T.; Hawkins, A.; Kuntz, I. D.; Rahn, L. A.; Rothfusa, A.; Sale, K.; Young, M. M.; Yang, C. L.; Pancerella, C. M.; Fabris, D. *J. Proteome Res* **1998**, *7*, 4848.
2. Muller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. *Anal. Chem.* **2001**, *73*, 1927.
3. Nobles, K. N.; Guan, Z. Q.; Xiao, K. H.; Oas, T. G.; Lefkowitz, R. J. *J. Biol. Chem.* **2007**, *282*, 21370.
4. Macleanfletcher, S.; Pollard, T. D. *Biochem. Biophys. Res. Commun.* **1980**, *96*, 18.
5. Spudich, J. A.; Watt, S. *J. Biol. Chem.* **1971**, *246*, 4866.
6. Lim, R. W. L.; Fechheimer, M. *Protein Expression Purif.* **1997**, *9*, 182.
7. Griffin, P. A.; Kim, D. H.; Furukawa, R.; Fechheimer, M. *To be submitted* **2010**.
8. Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C. *Anal. Biochem.* **1985**, *150*, 76.
9. Laemmli, U. K. *Nature* **1970**, *227*, 680.
10. Towbin, H.; Staehelin, T.; Gordon, J. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 4350.
11. Furukawa, R.; Kundra, R.; Fechheimer, M. *Mol. Biol. Cell* **1992**, *3*, A38.
12. Kandasamy, M. K.; McKinney, E. C.; Meagher, R. B. *Plant J.* **1999**, *18*, 681.

13. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850.

14. Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5802.

15. Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 834.

16. Kellersberger, K. A.; Yu, E.; Kruppa, G. H.; Young, M. M.; Fabris, D. *Anal. Chem.* **2004**, *76*, 2438.

17. Yu, E. T.; Hawkins, A.; Kuntz, I. D.; Rahn, L. A.; Rothfusa, A.; Sale, K.; Young, M. M.; Yang, C. L.; Pancerella, C. M.; Fabris, D. *J. Proteome Res* **1998**, *7*, 4848.

CHAPTER 3:

A Mass Spectrometry Identifiable Mass-Defect Labeled Chemical Crosslinker for Studies of Protein-Protein Interactions

A novel mass defect labeled chemical crosslinker has been developed to probe protein-protein interactions within large multi-subunit biological complexes. This reagent was designed to utilize both mass defect label technology and incorporate labile bonds that cleave under low energy collisionally induced dissociation. While there are inherent challenges incorporated with chemical crosslinking experiments, this mass defect labeled chemical crosslinker has been designed to circumvent some of these challenges by allowing for quick identification of crosslinked products by means of mass defect label technology, as well as easy fragmentation under collisionally activated dissociation (CAD), to aid identification of the regions of the proteins that are in multi-subunit complexes. The mass defect labeled peptides can be selected in an automated fashion for data-dependent MS/MS experiments. The first stage of MS/MS distinguishes 'dead-end' products from crosslinked products, and provides accurate mass measurement of the linked peptides. MS^3 of the released peptides yields sequence data that can isolate crosslinking sites at the amino acid level. This novel reagent was tested using simple

peptides neurotensin and bradykinin, as well as the more complex ribonuclease S – S peptide (RNase S) system.

Introduction: In the biological world protein-protein interactions play significant roles and there is significant interest in characterizing these molecular interactions to further understand specific biological functions and processes.[1] For the past two decades, researchers have employed a variety of bioanalytical techniques in order to probe these specific sites of interest, including x-ray crystallography, NMR spectroscopy, H-D exchange, cryo-EM, hydroxyl radical footprinting,[2] phage display, yeast two-hybrid screens, protein microarrays, gel filtration, surface Plasmon resonance, and fluorescence spectroscopy.[1] In recent years, mass spectrometry coupled with chemical crosslinking has improved verification of these specific interactions and yielded information regarding the higher order structure of large biomolecules.[3] Fourier-transform ion cyclotron resonance-mass spectrometry (FTICR-MS) provides high-resolution mass spectra, fairly rapidly from a very small amount of sample, without the concern of interfering compounds. Therefore, when coupled with chemical crosslinking approaches, it elucidates low-resolution three-dimensional structure of large biomolecules but more importantly the interactions that take place within them. A second advantage of using this mass spectrometric approach is that interactions that are transient in nature or dependent on specific physiological

conditions can be captured into long-lived covalent complexes, in which structural information is maintained during advanced processing (purification, enrichment and analysis).[1, 4]

In addition to the inherent advantages of using FTICR-MS, including improvements in mass range, resolving power, and mass accuracy, this technique allows for tandem mass spectrometry (MS/MS) experiments to be performed, which are particularly useful for crosslinking experiments.[5] Crosslinked compounds that arise from biological experiments are relatively large and of high mass-to-charge (*m/z*) and therefore fragmentation is preferred in order to gain specific information regarding interactions. MS/MS provides sequence specific identification of crosslinked peptides and also improves the confidence in the crosslink assignments. In addition to the high resolution, mass accuracy, ability to comprehensively fragment large peptides by MS/MS, FTICR-MS also allows for 'gas-phase' purification and subsequent fragmentation in the ICR cell.[6]

Even with the added advantages of FTICR-MS/MS, chemical crosslinking studies involve a number of different challenges, including (1) difficulty in recognizing reacted peptides within complex mass spectra, (2) complexity of the mass spectra due to unreacted, and reacted crosslinked products, (3) difficulty analyzing data when all possible crosslinks have to be considered and (4) sequencing issues when crosslinked peptides undergo dissociation. During a crosslinking experiment, three different types of crosslinks can arise: (1) peptides that have been

modified with a hydrolyzed crosslinker ('dead-end' product), (2) peptides that has two reactive groups crosslinked within itself (intra-peptide product), or (3) crosslinks that occur between two independent peptides (inter-peptide products). In addtion, analysis can be complicated by the presence of non-specifically modified peptides.[7] The only type of chemical crosslink that is analytically useful is the inter-crosslink products. Researchers have creatively attempted to design crosslinker reagents that facilitate the identification of inter-peptide products; e.g., by employing isotope-labeled crosslinkers,[8] isotope-labeled proteins,[9] cleavable crosslinkers,[10] fluorogenic crosslinkers,[11] or crosslinkers creating a characteristic marker ion during MS/MS analysis.[12] Affinity purification has also been utilized in the attempt to enrich the crosslinker containing species. These crosslinkers are often trifunctional crosslinkers that contain biotin,[10-11, 13] and can be used for affinity purification, however introducing the biotin moiety adds to the bulk of the crosslinker and therefore may hinder it from inserting itself into regions of close contact.

Due to the limitations stated above and the intrinsic complexity of the mass spectra that arises during chemical crosslinking studies, no researcher has successfully crosslinked on a proteome wide scale. Bruce likened the data analysis process involved in a crosslinking experiment to searching for a needle in a haystack.[12] The complexity of the crosslinking reaction and digestion mixtures (containing dead-ends, intra-peptide products and inter-peptide products)

only accounts for half of the challenges. The other half, is the direct interpretation of the MS/MS spectra of croslinked peptides, which is a convolution of ion series from two peptides joined by a crosslinker. The aim of the crosslinker design described here is to circumvent both of these intrinsic issues and create a reagent that is both identifiable in a complex mass spectrum, and also cleavable under low energy collisionally induced dissociation, so that the linked native peptides can be identified with high confidence. These features are demonstrated using two small peptides, neurotensin and bradykin, as well as the small protein-peptide complex of RNase S, and incorporating a non-mass defect labeled crosslinker, and two slightly different mass defect labeled chemical crosslinking reagents.

Experimental: *Chemicals.* RNase S, bradykinin and neurotensin were purchased from Sigma (St. Louise, MO) and used without further purification. Sequencing grade modified trypsin was purchased from Promega (Madison, WI). Starting materials for the synthesis of the chemical crosslinkers were purchased from Sigma (St. Louise MO) as well and used without further purification. The initial non-mass defect labeled reagent was a generous gift from Target Discovery (Palo Alto, CA).

Synthesis of Crosslinkers. The first attempt at a mass defect labeled crosslinker, Bis(2,5-dioxopyrrolidin-1-yl)3,3’-(3,5-dibromo-4-methylphenylazanediyl)diproponate (DiBMADPS) was synthesized by mixing acrylic acid (Sigma, St. Louis, MO) with 3,5-dibromo-4-

methylaniline (CalBiochem, Darmtadt, Germany), followed by activation with N-hydroxysuccinimide. Briefly acrylic acid (0.65 mmol) was mixed with 0.65 mmol of thiourea and 200 μL of water. The mixture was heated for 2 hours at 99 °C, and 100 μL of 40% (w/w) NaOH was then added and incubated for 1 hour at 99 °C in a thermomixer (Eppendorf). Each preparation was adjusted to pH ~8 with HCl, and 0.1 mmol of 3,5-dibromo-4-methylaniline was added. The solution was mixed until completely dissolved and kept at room temperature for 1 hour. The pH was adjusted to ~8 by addition of NaOH. DiBMADPS was precipitated by addition of HCl, filtered, washed with 1 mM of HCl, and lyophilized. DiBMADPS was then dissolved in DMSO, and activated with N-hydroxysuccinimide and dicyclohexylcarbodiimide overnight at room temperature. The reaction mixture was then filtered, and DiBMADPS was precipitated by the addition of water, then washed with water, and lyophilized. DiBMADPS was characterized by nano-ESI (QSTAR, Applied Biosystems, Framingham, MA) under the experimental conditions and parameter settings described elsewhere.[2] The synthesis scheme can be seen in Figure 3.1. A 1:1 mixture of DiBMADPS crosslinker to protein in DMSO, was then used for all crosslinking reactions. For storage, 10 mM stock solutions were kept in DMSO at -20 °C and were added directly to a solution of the protein in phosphate buffered saline (PBS) to give a final concentration as described below.

To prepare the second crosslinker, bis(2,5-dioxopyrrolidin-1-yl) 2,2'-(2,4-dibromophenylsulfonylazanediyl)diacetate (DiBBSIAS), 0.1 mmole of dibromobenzenesulfonylchloride was reacted with 0.1 mmole of iminodiacetic acid and 0.3 mmole NaOH in 100 μL of water at 75 ºC for 24 hours. The reaction product was purified using reversed-phase chromatography using a C_{18}-Sep-Pak (Millipore) and was dried *in vacuo*. 11 mg of dibromobenzenesulfoneiminodiacetic acid was then activated by N-hydroxysuccinimide in the presence of dicyclohexylcarbodiimde, overnight at room temperature in dimethylsulfoxide. The final product, DiBBSIAS, was then precipitated from acetone/ethyl ether, dried *in vacuo*, and stored in a 10 mM stock solution at -20 °C until use. Figure 3.2 shows the synthesis scheme.

Crosslinking Reaction. The crosslinking reagents were resuspended in DMSO at a concentration of 10 mM. RNase S, neurotensin, and bradykinin were dissolved in water to make 1 mM stock solutions. For crosslinking reactions, 10 μL of the stock solutions of Rnase S, bradykinin and neurotensin were diluted to 10 μM in 100 μM phospate saline buffer (100 mM sodium phosphate, 150 mM NAcl, pH 7.2), and the crosslinker was added in 50- and 100- fold molar excess over the peptides and protein, respectively. The reaction mixture was incubated at room temperature (25°C) for 60 minutes. The reaction was then quenched with 1 μL of 1 M Tris, pH 7.5.

SDS-PAGE and In-Gel Digestion. An aliquot of the RNase S reaction mixtures were loaded onto SDS-PAGE for separation using 10% resolving gels made in-house with a 4% stacker. The gels were stained with Coomassie Brilliant Blue (Bio-Rad, Hercules, CA) and imaged with a Densitometer (Molecular Dynamics). The gel bands of interest were excised and destained by incubating with 100 μL of 50 mM ammonium bicarbonate in 50% methanol for 30 minutes at room temperature, or until methanol solution was no longer blue. Gel plugs were then dehydrated by incubating with 100 μL 75% acetonitrile for 20 minutes at room temperature, and then dried at 40 °C for 15-20 minutes. Gel bands were then reduced and alkylated prior to tryptic digestion. 10 mM dithiothreitol (DTT) was prepared in 20 mM ammonium bicarbonate, 40 μL DTT solution was added to each gel plug and allowed to incubate at 37 °C for 1-2 hours. A 100 mM solution of iodoacetamide was prepared in 20 mM ammonium bicarbonate and added to the gel plugs immediately after removing DTT solution, and allowed to incubate at room temperature for 30 minutes in the dark. The plugs were then dehydrated and dried once more as per described above. Tryspin (20 ng/μL) in 20 mM ammonium bicarbone was used for digesting proteins at 37 °C overnight. To extract and dry resulting peptides, 60 μL of 50% acetonitrile/0.1% TFA was added to each plug and allowed to incubate at room temperature for 20 minutes. Extract was removed and set aside. 40 μL of the same solution was then added and allowed to incubate once more, extract removed and combined with earlier solution, and dried.

Liquid Chromatography and Mass Spectrometry. Experiments were performed with a 9.4 T Bruker Apex Ultra QhFTMS (Billerica, MA) fitted with an Apollo II dual source, a 25 W CO_2 laser (Synrad model J48-2, Mukilteo, WA) for IRMPD, and an indirectly heated hollow cathode for generating electrons for ECD experiments. HPLC separation of tryptic peptides was performed with an Agilent 1100 series (Santa Clara, CA) equipped with an autosampler. Neurotensin and bradykinin were directly infused into the mass spectrometer in a solution of 50:50:0.1 (Water:Methanol:Formic Acid) at a flow rate of 120 μ/ hour. Rnase S was directly infused into the mass spectrometer using a static nanospray source in a solution of 50:50:0.1 (Methanol:Water:Formic Acid) at a flow rate of 5 μL/hour.

A second sample of RNase S was injected by the autosampler onto a C18 column (PepMap C18, 300 μm, 3 μm, 100 Å, LC Packing) at a flow rate of 0.4 μL/min with 95/5 (v/v) H_2O (with 0.1% formic acid)(A)/Acetonitrile (with 0.1% formic acid)(B). Peptides were then separated using the following gradient: 0% B for 0-5 minutes, 20-75% B for 5-120 minutes, 75-90% B for 120-125 minutes, 90-95% B for 120-130 minutes, 0% B for 130-140 minutes. The column was directly linked to the inlet of the mass spectrometer and the analysis was performed online. LC/MS/MS spectra were acquired using HyStar software (version 3.0, Bruker Daltonic). MS/MS data acquisition was set in automatic mode with active exclusion based on peak intensity and peak width. Five precursor ions were selected from each MS scan and excluded after two

MS/MS scans. Calibration of the instrument was performed with a tryptic digest of BSA (Bovine Serum Albumin, Sigma, St. Louise MO). Data analysis and processing were performed using Bruker Daltonics DataAnalysis (DA) software (version 3.4 Build 72).

Collisionally activated dissociation (CAD) and electron capture dissociation (ECD) were used as the fragmentation methods in all experiments. Multiply charged precursor ions were selected in the external quadrupole and stored in the external hexapole collision cell for 1-6 seconds where they underwent CAD at voltages from 5-15 V. The directly infused samples were then analyzed by a third MS event in the FT-MS analyzer. The multiply charged precursor ions were again isolated in the analyzer cell and then irradiated for 0.05 s with the ECD cathode set to -1.5 V, the ECD lens was set to between 3 and 5 V and the heater current was set to 1.5 A. Ions were excited with an RF frequency chirp that covered the *m/z* range of 100-2000. Twenty-four acquisitions were co-added for each mass spectrum. 512K points were acquired in 2.4 MHz digitization rate, padded with one zero fill and apodized using a sinebell window.

Results and Discussions*. Crosslinker Design.* The first and second generations of the mass defect label chemical crosslinking reagent (Figures 3.1 and 3.2) are similar in general concept with minor differences in the chemical composition, but significant differences in their performance. This reagent was designed to address the intrinsic challenges of conventional crosslinkers and provide researchers with a means for easily identifying the products of interest

in the complex mass spectrum prior to informative fragmentation. These crosslinkers were designed to have a labile bond that would cleave more readily than bonds within the peptide. As will be shown below, this feature did not perform as expected in the first generation crosslinker (DiBMADPS) (Figure 3.1) however. Possible reasons for the lack of cleavage around the central tertiary amine or at the peptide-crosslinker interface is the extra $–CH_2$ bond on either side of the central nitrogen. Preliminary studies with a reagent that did not contain a mass defect label; shown in Figure 3.3, produced fragmentation readily at both the tertiary amine and the peptide-crosslinker interface. This compound however only contained one $–CH_2$ group between the central nitrogen and the carboxyl group linked to the peptide (Figure 3.3). The first generation crosslinker however had two $–CH_2$ groups in this position, making the actual crosslinking reagent longer, but hindering the desired fragmentation. Fragmentation of the crosslinking moiety from the peptide of interest is important, for it allows easier spectral interpretation, as well as additional stages of MS/MS to be carried out, resulting in peptide bond fragmentation and sequencing of the specific amino acids involved in the interaction.

As can be seen in Figures 3.1 and 3.2, both generations of mass defect labeled chemical crosslinkers contain a dibromophenyl mass defect label moiety. This feature allows the crosslinked peptides to be identified readily. Peptides are primarily composed of the elements from the first two rows of the periodic table, most of whose mass defects fall in the range of ±

0.008 amu (C, 0.0 mmu; H, 7.8 mmu; O, -5.1 mmu; N, 3.1 mmu; S, -28.0 mmu) (Figure 3.4). Therefore, peptide molecular weights tend to occupy only one-third of a unit mass at any given nominal mass, which causes peptide masses to overlap and specificity is reduced, as shown in Figure 1.6. When a mass defect label is incorporated into a peptide, the fractions mass shifts to a value that is not possible with unlabeled peptides. The compound chosen in this approach is bromine, due to its extremely high mass defect (-82.0 mmu). Incorporating two bromine atoms into the chemical crosslinking reagent provides the means identify crosslinked peptides on the basis of their mass defect.

Initial Non-Mass Defect Labeled Crosslinker. To test our initial hypothesis, we investigated the products of the reaction of neurotensin, and bradykinin with the non-mass defect labeled crosslinker (4519-012-IF) from Target Discovery. The peptides were reacted with the crosslinker and Figure 3.5 shows the resulting mass spectrum. It is apparent that in this neurotensin/bradykinin mixture a number of different types of crosslinks occurred. The 5^+, 4^+, 3^+ and 2^+ species of bradykinin crosslinked to neurotensin is observed as well as the 4^+, 3^+, and 2^+ species of bradykinin linked to bradykinin and the 3^+ and 2^+ species of neurotensin crosslinked to a dead end. The most abundant peak in the mass spectrum occurred at *m/z* 722.8760^{4+} and corresponded to the 4^+ species of bradykinin crosslinked to neurotensin and thus was the peak chosen for further analysis. Figure 3.6 shows the results of low energy CAD of this product ion.

It is clear in this spectra, that two types of fragmentation patterns arise during low energy CAD fragmentation. One in which the peptide-crosslinker bond is broken and one in which the bond attached to the tertiary amine is broken, resulting in free bradykinin^{2+} and neurotensin^{2+} as well as an N-acetylated neurotensin molecule. Subsequent MS^3 fragmentation by ECD results in almost complete coverage of the free neurotensin (precursor ion at *m/z* 836.9550^{2+}) (Figure 3.7). The resulting ECD mass spectrum (MS^3) for neurotensin contained fragment peaks, which corresponded to the a peptide backbone. The peak at *m/z* 1673.8604^{1+} represents the charge-reduced species. Peaks *m/z* 801.4857^{1+}, 914.5571^{1+}, 1026.6332^{1+}, 1140.6762^{1+}, 1269.7189^{1+}, 1432.7822^{1+} and 1545.8580^{1+} represent the z ions in the series, while *m/z* 1029.5477^{1+}, 1282.7016^{1+}, 1445.7646^{1+}, and 1558.8490^{1+} represent the c ions.

First Generation MDL Crosslinker. After proof of concept was made with the non-mass defect labeled reagent, the first generation chemical crosslinker (DiBMADPS) (Figure 3.1) was synthesized and reacted with RNase S. This crosslinking reagent contained a tertiary amine and the 2 bromine mass defect labels. Other than the addition of the mass defect label moiety, the only difference between this first generation croslinker (DiBMADPS) and the initial non-mass defect labeled reagent was the number of CH_2 groups between the carboxyl group and the tertiary amine. The initial crosslinker had only one CH_2 group on either side of the tertiary amine, however DiBMADPS had two. Two NHS esters were incorporated to make this

crosslinker reactive with primary amines, however the proposed system could be altered to use reactive groups with different specificity.

Ribonucelase S (RNase S) was chosen as a test compound for crosslinking studies because it is a low molecular weight protein complex formed from S-peptide 1-20 and S-protein 21-124 that are the hydrolysis products from ribonucelase A (RNase A) as a result of cleavage between residues 20 (Ala) and 21 (Ser) by subtilisin. There is known to be a noncovalent interaction between S-peptide and S-protein. This complex retains enzymatic activity similar to that of RNase A. After the crosslinker was prepared, we set out to optimize the reaction conditions in order to facilitate the most successful crosslinking reaction. This was carried out by varying the molar rations of RNase S to the crosslinker at 1:50 and 1:100 and the reaction times were varied between 15, 30, 60 and 120 minutes. Rnase S without crosslinker was also used as a control. Each mixture was then separated by SDS-PAGE followed by Coomassie blue staining. Figure 3.7 illustrates the gel separation of the crosslinking reaction mixture with the varying molar ratios of and reaction times. As can be seen from this gel image, the most ideal reaction conditions occurred at 50-fold molar excess of crosslinker, incubated for 30 minutes. The molecular weights of S-protein, S-peptide and the crosslinker are 11,534, 2,166 and 347 Da respectively and thus the 1:1 Crosslinked species should be located around 14,000 Da. As shown in Figure 3.7, the control (Rnase S only) was approximately 11 kDa. This indicates that only S-

protein appeared on the gel, and this could be due to the fact that S-peptide may be too small to be detected by the gel. When the crosslinker was introduced a distinctive band occurred around mass 14 kDa. Even at the first reaction time interval and a ratio of 50:1, slight crosslinking is observed. As the time interval increases, it is clear that the amount of the higher gel band increases. This is mainly due to an increase in the crosslinking reaction. Since 1:1 linkage of interacting proteins is what we are looking for, a reaction ratio of 50X was selected at an incubation time of 60 minutes. For in-gel trypsin digestion, the gel bands were excised carefully to include as many of the proteins as possible with minimum volume. The tryptic digest was then analyzed by LC-MS/MS.

Analysis of Croslinked Peptides arising from DiBMADPS. The in-gel tryptic digest of the crosslinked band was then analyzed by LC-MS/MS in order to identify the crosslinked species. LC-MS/MS data was acquired in automated fashion. As described in the earlier HPLC section, the eluent from the HPLC C18 column was injected directly into the mass spectrometer for analysis and the 5 most abundant ions in each mass spectrum were selected for MS/MS analysis. Previously calculated potential crosslinked masses were input into a preferred list in the data analysis software. This preferred list was constructed of all possible crosslinkages and used for data dependent acquisition. When an ion present on this list appeared in the mass spectrum, it was selected for fragmentation even if it was not among the five most abundant ions. In this

manner, the crosslinked peptides were selected for analysis even though they were rarely the most abundant ions in the spectra. Once data acquisition was complete we were able to analyze the data to identify which of the peaks chosen for fragmentation were actual crosslinked sites. It was clear that the majority of the peaks picked by the acquisition software were simple tryptic fragments of RNase S and therefore of little use for our analysis, however of the 145 useful compounds identified (Figure 3.9) by the SNAP function in DA, one was an actual inter-peptide crosslinked peak. The peak at *m/z* 1157.04^{2+} has the expected mass defect for a crosslinked product composed of to amino acids 1-7 of S-peptide and residues 12-21 of S-protein (Figure 3.10). The other peaks in the mass spectrum corresponded to either simple tryptic fragments of RNase S, dead-ends or unreacted species.

CAD fragmentation of the m/z 1157.04^{2+} precursor ion yielded fragment ions that corroborated the assignment. As illustrated in Figure 3.11, there are a number of both b and y ions present. We did not observe the fragmentation of the tertiary amine in the crosslinker itself, or the peptide-crosslinker interface however sites of interaction with the crosslinker were able to be determined due to the specific fragmentation pattern. The y ions that are present appear at *m/z* 1722.80^{1+}, 1609.71^{1+}, and 1508.90^{1+}, and the b ions occur at *m/z* 1722.80^{1+}, 1380.80^{1+}, 1309.76^{1+}, 1238.72^{1+}, 1167.69^{1+}, and 1011.36^{1+}. From this data it is clear that the cross-linker attached between amino acid 1 on S-peptide, or the N-terminus, and the lysine residue on S-

protein. This can be concluded from the fragment ion at m/z 1011.36^{2+}, in fact the only amino acids present still attached to the cross-linker are K – E - T on the alpha peptide (S-peptide) and K - D on the beta (S-protein).

Although we did not see the desired fragmentation around the tertiary amine in the cross-linker itself, we were still able to use the information gathered for this experiment to deduce which amino acids are close enough for interaction in the RNase S complex. This conclusion is confirmed by the x-ray structure of RNase S, as is illustrated in Figure 3.12. It is clear from this image that Lys 1 of S-peptide and Lys 17 of S-protein are within the 12.04 Å of each other and therefore can be linked by DiBMADPS. The red line in the figure represents this cross-linking. From the x-ray structure of RNase S, the distance between the N Terminus of S-peptide and Lys 17 of S-protein was estimated to be 14 Å. However, due to the fact that our cross-linker is only 12.04 Å, it can be concluded that this static x-ray structure does not represent the entire range of motion of S-peptide within S-protein, which exemplifies why mass spectrometry needs to be used in conjunction with such static methods of analysis such as x-ray crystallography. It can also be denoted from this cross-link that both Lys 1 and Lys 17 are on the surface of the RNase S complex and therefore readily available for cross-linking. No other cross-links were found at this point.

Second Generation MDL Crosslinker. The first generation mass defect labeled crosslinking reagent, synthesis commenced on the second generation reagent, did not exhibit sufficient lability and so a second generation reagent was prepared as shown in Figure 3.2. It was the aim of this research to design a cleavable mass defect labeled chemical crosslinker that would be easily identifiable within a complex mass spectrum and also cleave at either the peptide-crosslinker interface of the tertiary amine under low energy collisionally activated dissociation conditions. While the initial crosslinker underwent facile cleavage, it did not contain the required mass defect moiety. The first mass defect labeled crosslinker (DiBMADPS) contained the mass defect label moiety, making it stand out within a complex mass spectrum, however cleavage did not occur around either the tertiary amine or the peptide-crosslinker interface. One possible explanation for this is the addition of a CH_2 group between the carboxylic acid and the tertiary amine. This elongated the crosslinker and could have allowed for the fragmentation energy to be too spread out to fragment where predicted.

Analysis of Crosslinked Peptides Using DiBBSIAS. The second generation mass defect labeled crosslinking reagent (DiBBSIAS) likewise contained the mass defect labeled moiety, however it did not contain the extra CH_2 group between the tertiary amine and the carboxylic acid. It also incorporated a sulfone group to increase the water solubility of the compound as a whole. Initially, this compound was just tested with neurotensin. Figure 3.13 shows the MS

spectrum obtained from the mixture of neurotensin with DiBBSIAS. It is clear that there is a crosslinked peak present in the mass spectrum at *m/z* 929.7679 and also unreacted neurotensin 2^+ and 3^+. Upon selection of the crosslinked peak for further MS/MS analysis using CAD, it was found that the crosslinked peak dissociated entirely into charge states of the neurotensin peptide (Figure 3.14). The first fragment ion arises at *m/z* 558.3184^{3+} and corresponds to neurotensin 3^+ and the second at *m/z* 836.9744^{2+} corresponding to neurotensin 2^+. Isolating the neurotensin 2^+ allowed for a third stage of MS to be conducted using ECD and the resulting spectrum can be seen in Figure 3.15. Within this spectrum lies the charge reduced species at *m/z* 1673.9643. Complete sequence coverage is not obtained from this data, however a series of z and c ions can be seen. (z ions occur at *m/z* 1027.6559^{1+}, 1141.7024^{1+}, 1269.7418^{1+}, 1432.8939^{1+}, and 1545.8993^{1+}, and c ions occur at *m/z* 1281.7271^{1+}, 1446.7929^{1+}, and 1558.8825^{1+}). Due to the fact that ECD does not cleave on the N-terminal side of proline, 8 of the possible 10 fragment ions are observed.

LC-MS/MS was then performed on the RNase S reacted compounds. From examiniation of the mass defect products, it was concluded that fractions 4, 5, 6, 14, and 15 are crosslinked peaks (Table 3.1). Fraction 4's MS and MS/MS spectra can be seen in Figure 3.17. The precursor ion, *m/z* 634.2891^{4+}, was selected as the other abundant peaks in the mass spectrum were either tryptic fragments (*m/z* 858.4170) or miscellaneous peaks. Following CAD fragmentation three

distinct peaks arose. The peak at *m/z* 1084.5012^{2+} corresponded to the precursor mass minus the crosslinker mass, thereby indicating that it is either a dead end peptide or an intra-peptide crosslinked peak. The peak at *m/z* 918.4060^{2+} corresponded to the loss of A – A – S – T from the precursor mass and the peak at *m/z* 774.0309^{2+} corresponded to the consecutive loss of S – S – D. This fragmentation patterns yields the conclusion that the fragment selected for analysis at *m/z* 634.2791^{4+} corresponds to an intra-peptide fragment ion between Lys1 and Lys7. Fraction 5 yields similar conclusions (Figure 3.18).

Fraction 15 shown in Figure 3.19 illustrates the fragmentation of the precursor ion at *m/z* 1132.1554^{4+}, and subsequent fragment ions at *m/z* 1350.7008^{1+}, 1236.6552^{1+}, 1084.5012^{2+}, 997.0051^{2+} and 831.3900^{2+}. The peak at *m/z* 1350.7008^{1+} represents the loss of S – S – S – N – Y – C from the precursor mass and subsequently the peak at *m/z* 1236.6552^{1+} represents the loss of N from this series. This corresponds to the loss of amino acids 21 – 27 of S protein. Likewise the peak at *m/z* 1084.5012^{2+} corresponds to the loss of the crosslinker mass from S-peptide, while the peak at *m/z* 1013.4841^{2+} represents the loss of A – A and 831.3900^{2+} represents the loss of S – S – T – S. This sequence corresponds to amino acids 15 – 20 of S-peptide. Therefore it can be concluded that the peak chosen for MS/MS analysis in fraction 15 correlates to a inter-peptide linkage between Lys7 of S-peptide and Lys31 of S-protein.

Fractions 6 and 14 also result from inter-peptide crosslinks between S-peptide and S-protein. The precursor peak at *m/z* 1239.2953^{3+} corresponds to the inter-peptide linkage between Lys7 of S-peptide and Lys37 of S-protein. This is confirmed by the subsequent mass spectrum. It is also seen in Figure 3.20 that the loss of the crosslinker is seen in peak *m/z* 1084.1112^{1+}. The MS/MS spectrum for this fragment only represents the S-peptide portion of the inter-peptide linkage, however utilizing the accurate mass data, we concluded the other portion of the linkage was amino acids 31-41. This conclusion was further confirmed in fraction 14 (Figure 3.21). The precursor ion chosen was *m/z* 971.3666^{4+} which is the 4^{+} equivalent to the precursor in fraction 6. The MS/MS spectrum for this peak however resulted in far more fragmentation. We not only observe the S-peptide fragment ions at *m/z* 1084.0002^{2+}, 1013.0844^{2+} and 831.3900^{2+}, but we also see the fragment ions resulting from fragmentation of S-protein (*m/z* 1220.5582^{1+}, 1117.6635^{1+}, 961.4335^{1+} and 846.5158^{1+}).

This LC-MS/MS data obtained from the crosslinking of RNase S with DiBBSIAS resulted in the following intra-peptide linkage, Lys1 – Lys7 and the following inter-peptide linkages, Lys7 – Lys31 and Lys7 – Lys37. This is depicted in Figure 3.22. Although proof of crosslinker-peptide cleavage is observed during the LC-MS/MS run. The automatically selected collision energy resulted in fragmentation of bonds within the peptide as well as from between the peptides (within the crosslinker). If this procedure is to be adapted to a proteome wide scale,

it is imperative that the novel mass defect labeled crosslinker used, firstly cleaves at the desired position, and also results in specific sequence information. In order to confirm that this crosslinker could provide such cleavage in preference to peptide backbone only, we analyzed the same RNase S sample linked with DiBBSIAS by direct infusion static nano-electrospray ionization, and manually controlled the collision energy.

Static nanospray FTICR-MS of the entire tryptic digest mixture from the reaction between RNase S and DiBBSIAS is shown in Figure 3.23. Most of the peaks represent simple tryptic fragments arising from the proteolysis of RNase S (Denoted by diamonds in the mass spectrum). A peak that contained DiBBSIAS was observed at *m/z* 1132.5051^{4+}. This peak was then isolated in the hexapole and fragmented by CAD with a collision energy of 6 eV (versus 12-18 eV for automated LC-MS/MS) (Figure 3.24). It is apparent that the crosslinker fragments in preference to the peptides and results in two peaks corresponding to intact linked peptides. (*m/z* 1083.5189^{2+} and 997.0299^{2+}). Each of these peaks were then separately isolated in the ICR cell and subjected to ECD. Precursor peak *m/z* 1083.5189^{2+} resulted in a comprehensive series of c and z ions covering amino acids 1 – 20 on S-peptide (Figure 3.25). Precursor peak *m/z* 997.0299^{2+} fragmented into a comprehensive series of c and z ions covering amino acids 21 – 37 (Figure 3.26). This data indicates that the inter-peptide crosslink that arose from the reaction of RNase S and DIBBSIAS occurred between Lys1 or Lys7 on S-peptide and Lys31 on S-peptide.

The reason that the actual amino acid involved on S-peptide could not be determined is due to the clean cleavage of DiBBSIAS from the peptide itself. Since Lys1 is the N-terminal amino acid it can be involved in chemical crosslinking just as Lys7 is. It is more likely however that Lys7 is the amino acid involved in this interaction due to the fact that trypsin cleaves after lysine and arginine and if the crosslinker was positioned at Lys7, it would have been modified enough so that trypsin would not have cleaved there. Therefore, although it could simply be a missed cleavage, it is more likely that this precursor ion represents the linkage between Lys7 and Lys31.

Static Nanospray of Crosslinked RNase S. Although we were unable to achieve complete sequence coverage with the online automated LC-MS/MS run of DiBBSIAS and RNase S, we were able to demonstrate the crosslinkers ability to do so with the static nanospray experiment. This was mainly due to the fact that software needs to be developed to allow the user to control the energy being placed on the automatically chosen precursor ions in the LC-MS/MS run and software also needs to be developed to help select precursor ions by their mass defects within the mass spectrum on the chromatographic time scale. This software is in the process of being written.

It can also be observed from this data that the actual number of crosslinks that occur within a tryptic digest of crosslinking reaction is fairly small. The ions that do represent crosslinked peaks are very small in intensity and abundance and therefore difficult to be

observed in an online analysis. The histograms in Figure 3.27 and 3.28 show this. In order to achieve a more abundant array of inter-peptide crosslinks, some type of enrichment would benefit the procedure by affinity purification or antibody enrichment.

Conclusions. Three different mass defect labeled chemical crosslinkers were examined in this initial study. The first crosslinker did not contain a mass defect label moiety and therefore was just used to determine if the crosslink itself was cleavable under low energy collisionally activated dissociation (CAD). The second crosslinker contained the mass defect label, however it contained an extra $–CH_2$ group between the carboxylic acid and the tertiary amine on each side of the crosslinker. Upon analysis of this crosslinker with neurotensin, bradykinin and RNase S it was concluded that this addition of this extra $–CH_2$ group hindered the desired fragmentation both around the tertiary amine and at the peptide-crosslinker interface. Therefore, a second crosslinker was developed that contained the correct amount of $–CH_2$ groups, the bromine mass defect label moiety and also a sulfone group for added solubility. This crosslinker was again tested with the same three compounds, and performed as expected. CAD of the precursor peaks resulted in fragmentation of the crosslinker, leaving only the peptides that were linked to be analyzed by further MS events. The results obtained from all studies involving the mass defect labeled chemical crosslinker directly correlate to those previously reported by Tane *et al.*.[12] In order to further test this reagent, software will need to be developed that will allow the user the

ability to alter the collision energy within the LC-MS/MS run and also select precursor ions by their mass defects during a chromatographic run.

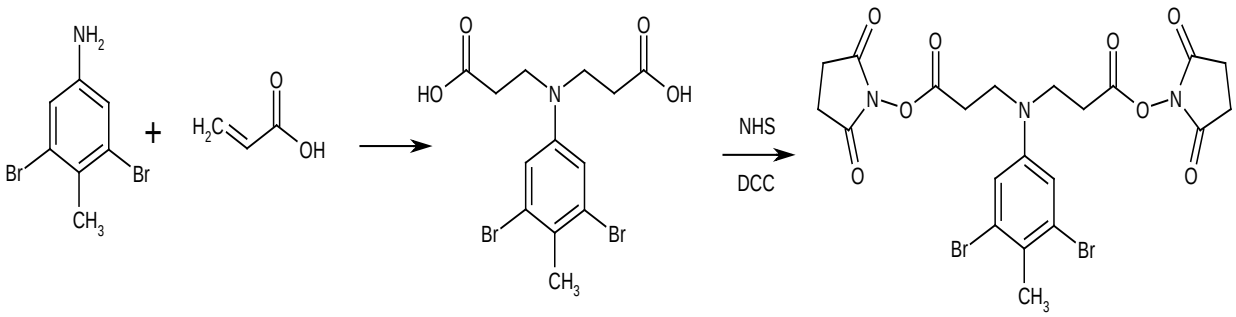


Figure 3.1. Synthesis scheme of DiBMADPS

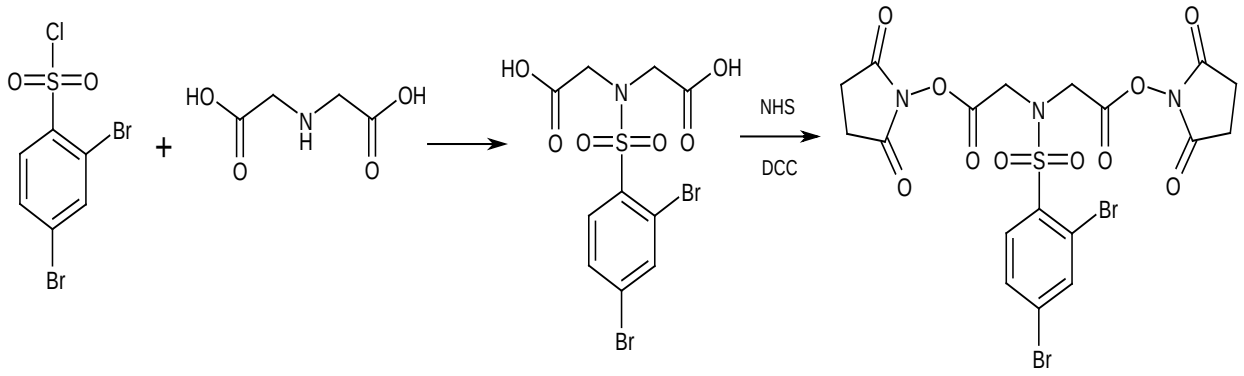


Figure 3.2. Synthesis scheme of DiBBSIAS

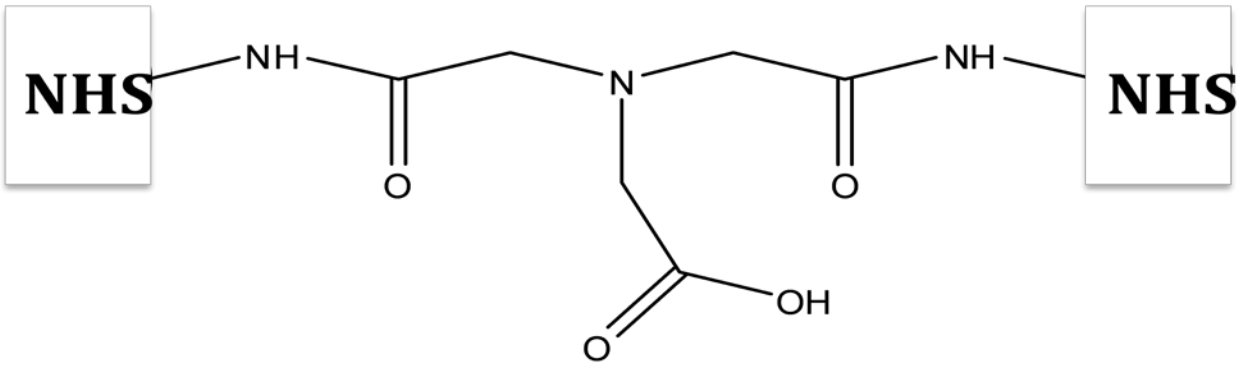


Figure 3.3. 4519-012-1F Crosslinker from Target Discovery used in initial cleavage studies.

Element	Mass Defect (amu)
^{12}C	0
^{1}H	0.0078
^{16}O	-0.0051
^{15}N	0.0031
^{32}S	-0.0279

Figure 3.4. Table showing the common elements and their corresponding mass defects.

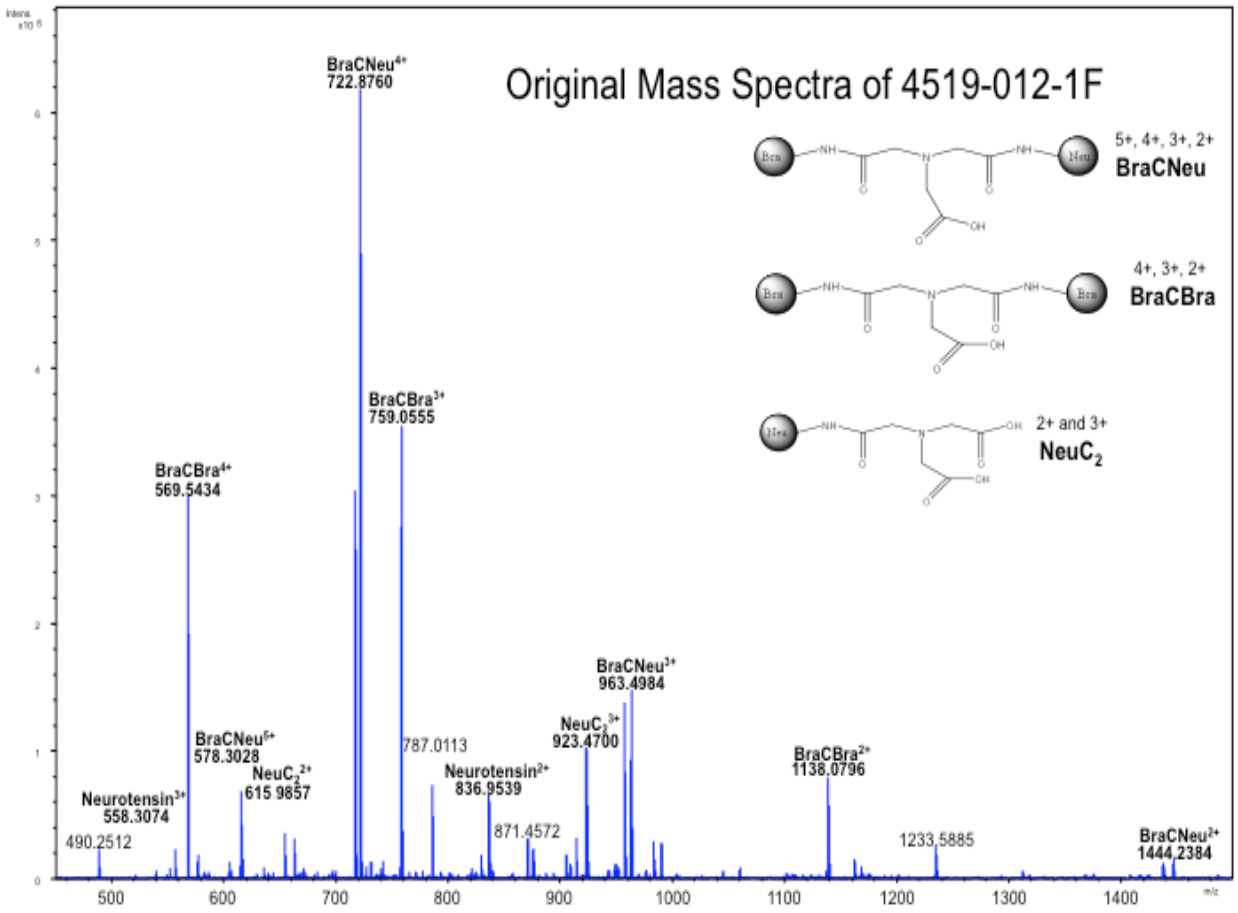


Figure 3.5. MS spectrum of the reaction between bradykinin, neurotensin and the first generation crosslinker, DiBMADPS. This spectrum shows the varying degrees of crosslinking that can occur in a mixture. The peak corresponding to the neurotensin – DiBMADPS – bradykinin^{4+} molecule was isolated for further analysis by CAD.

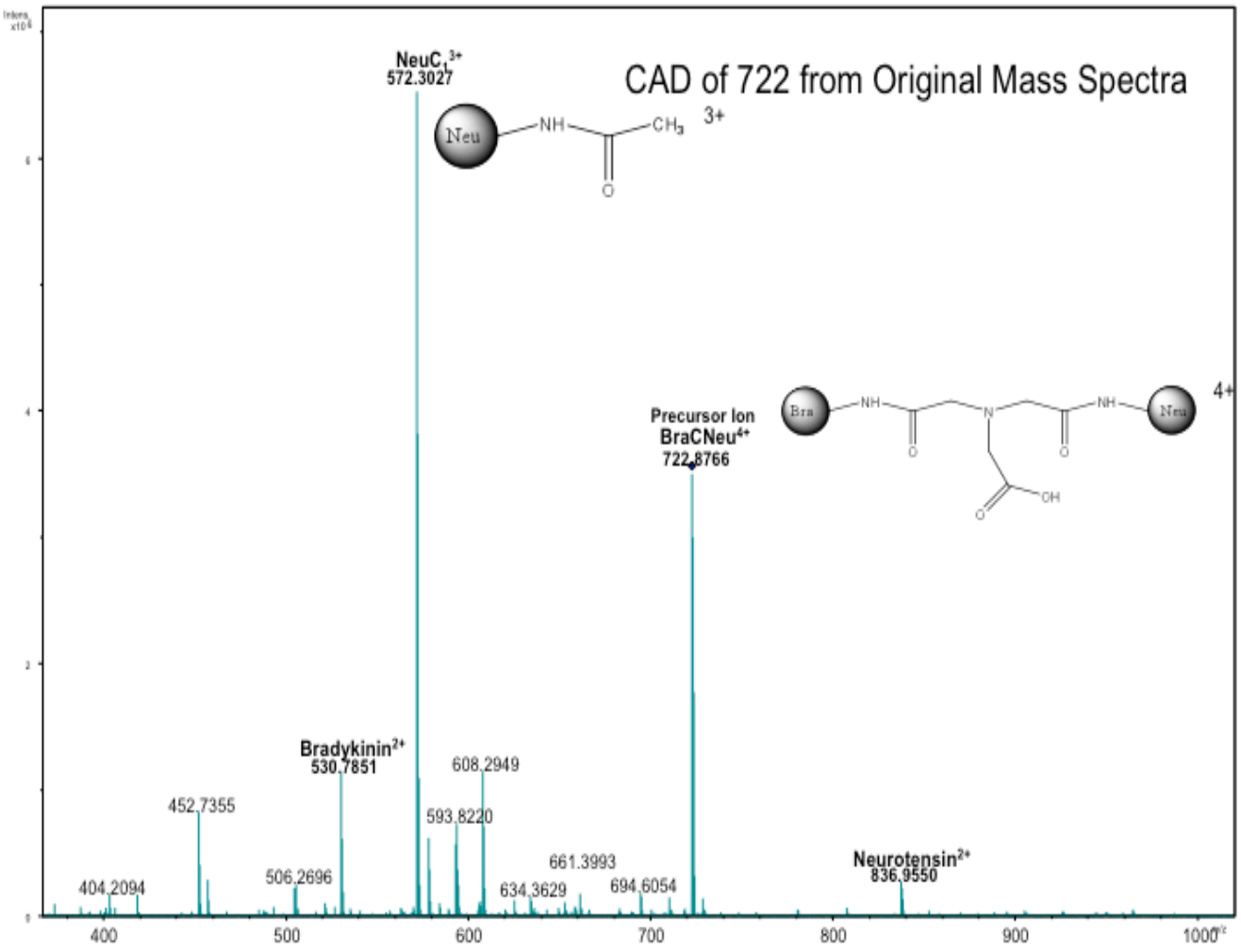


Figure 3.6. MS/MS spectra of bradykinin - neurotensin crosslinked compounds. CAD MS/MS spectrum of triply charged whole molecule (Neurotensin-crosslinker-Bradykinin).

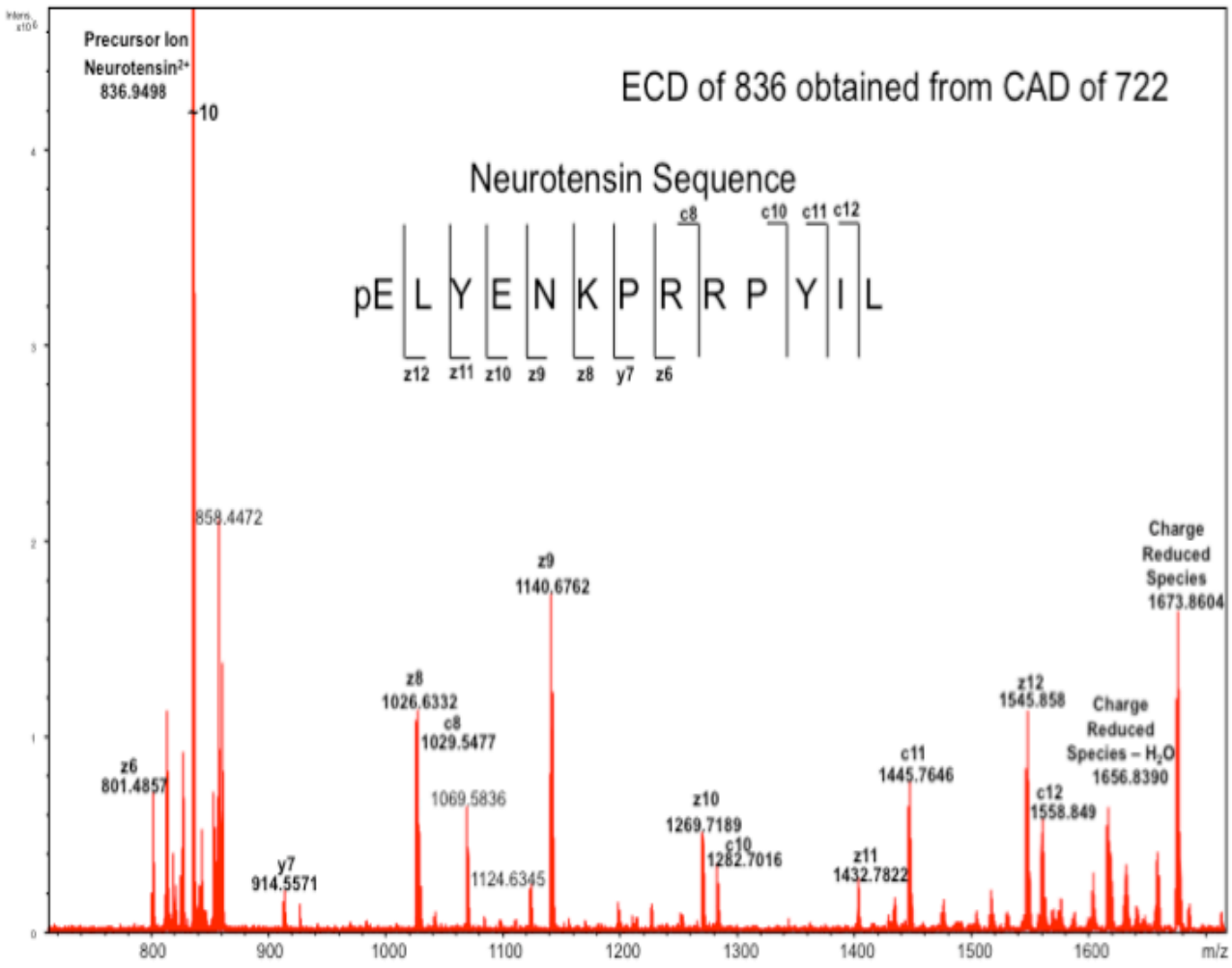


Figure 3.7. MS^3 spectrum of crosslinked bradykinin - neurotensin products. MS^3 mass spectrum obtained isolating CAD products in the ICR-cell and performing ECD on the isolated neurotensin ions.

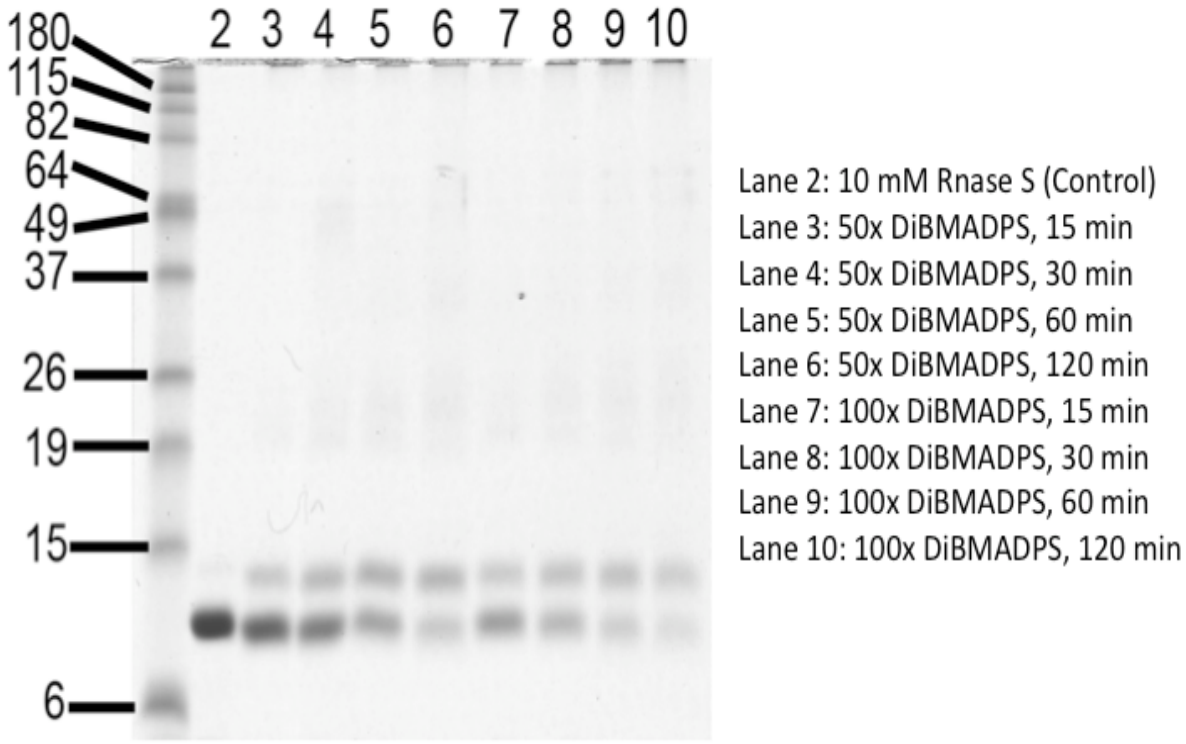


Figure 3.8. Gel of RNase S. SDS-PAGE gel image of 10 μM RNase S reacted with either a 50 fold excess of DiBMADPS or a 100 fold excess of DiBMADPS, reacted for either 15, 30, 60, or 120 minutes.

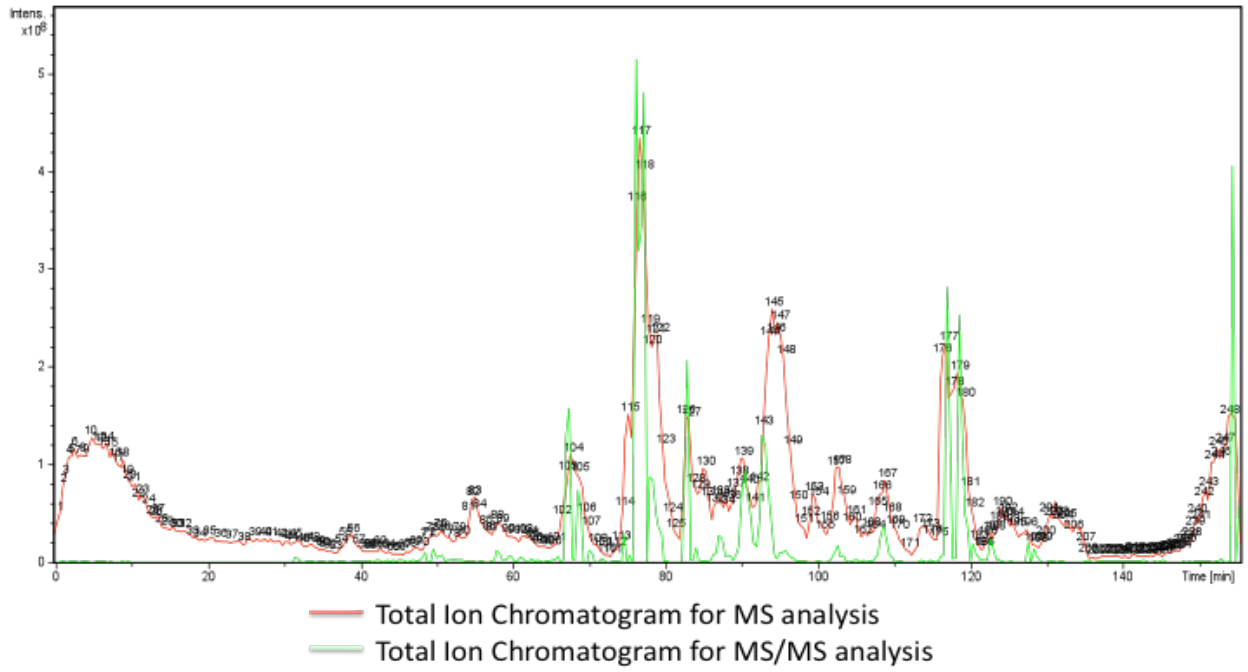


Figure 3.9. Total ion chromatogram for the gel band corresponding to a 50 fold excess of DiBMADPS reacted for 60 minutes. The compounds producing actual fragmentation began at approximately 70 minutes, and ran through 120 minutes.

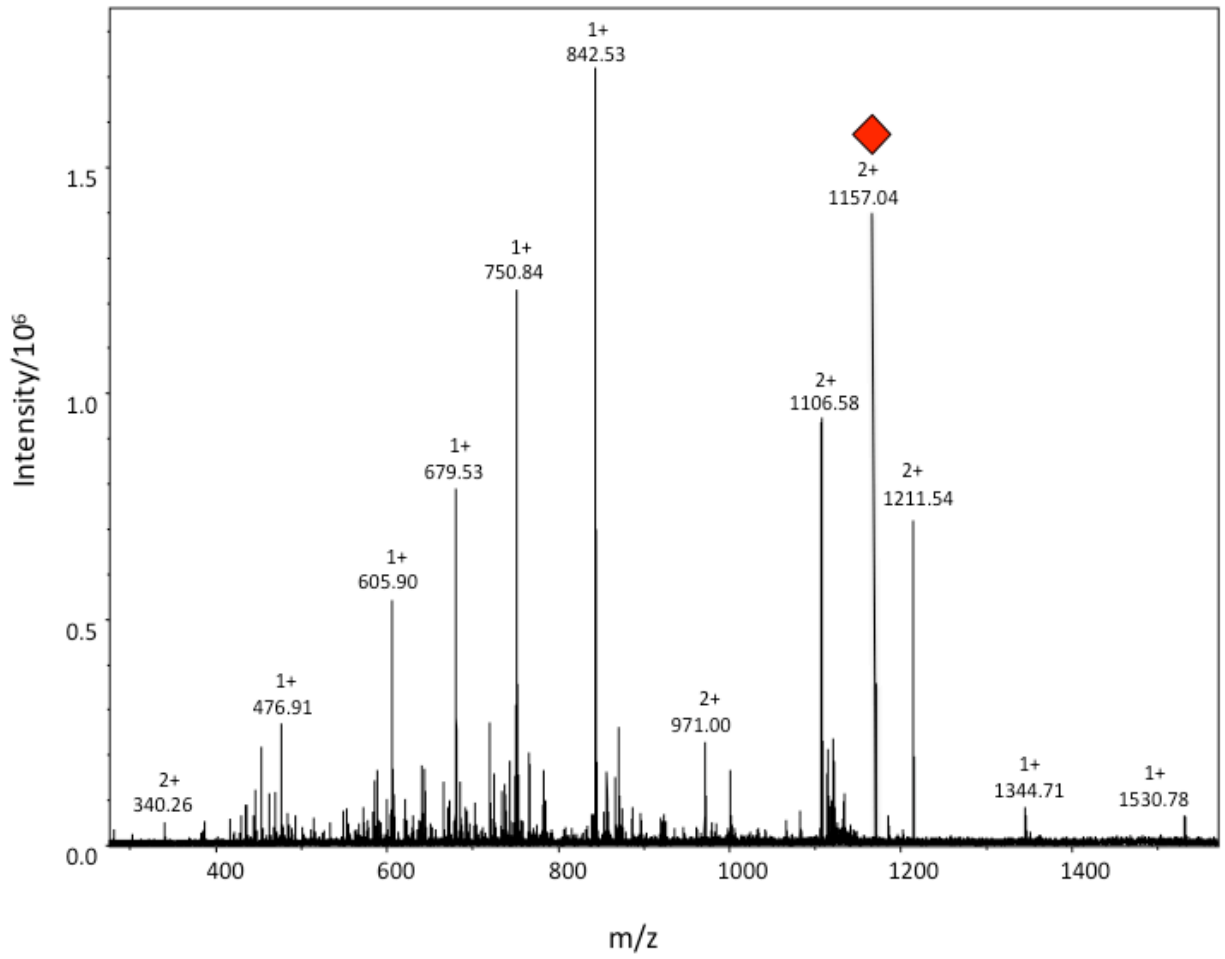


Figure 3.10. LC-MS data for tryptic digest of RNase S after reaction with DiBMADPS crosslinker. Shown is a spectrum of one chromatographic interval, exhibiting a peak corresponding to crosslink between Lys 1 of S-peptide and Lys 17 of S-protein at *m/z* 1157.04^{2+} (marked with a diamond).

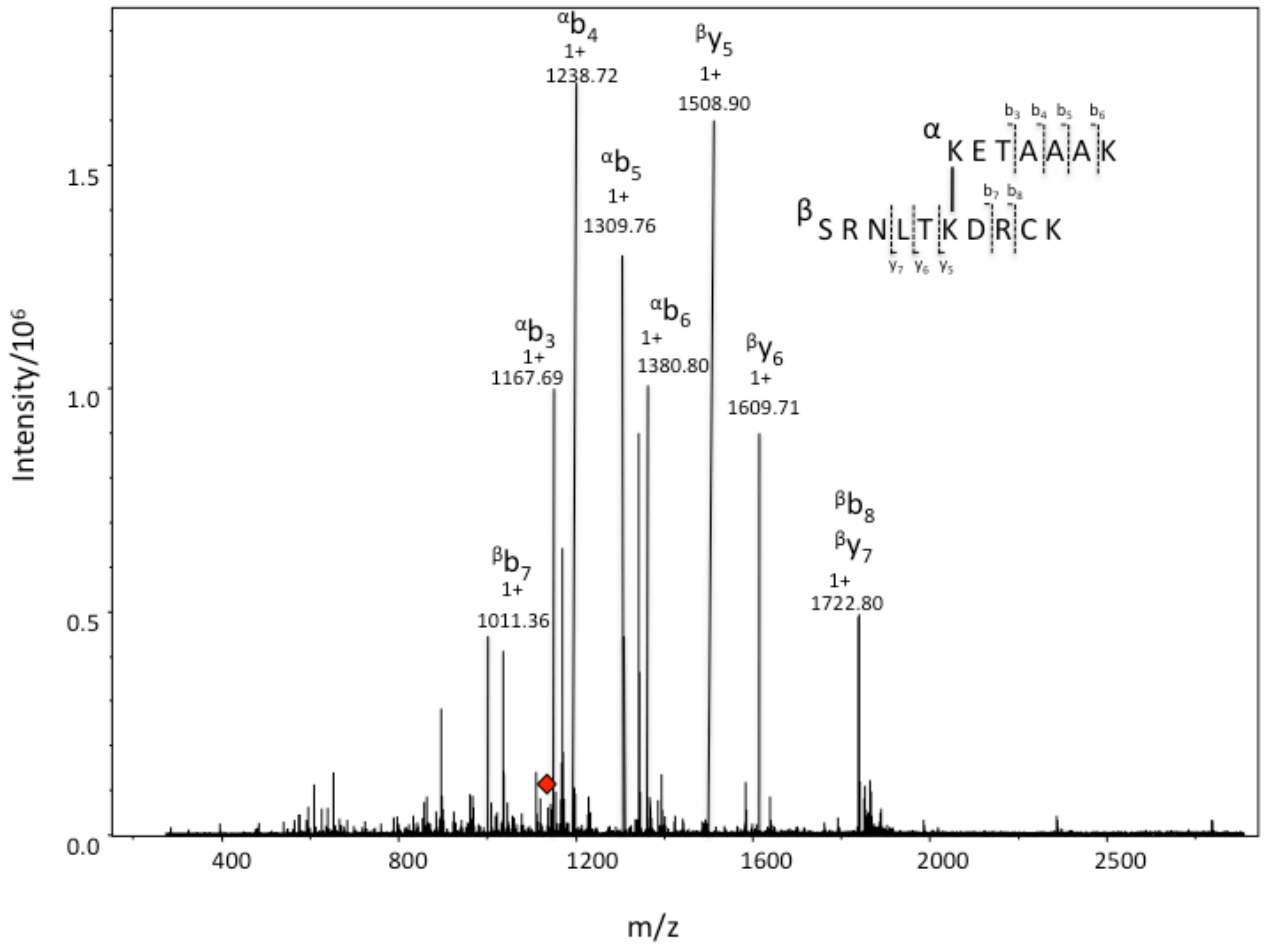


Figure 3.11. MS/MS spectra of 1157.04^{2+}. MS^2 mass spectrum obtained by performing CAD fragmentation of preferred peak, resulting in a series of b and y ions that elucidate the amino acids involved in the cross-linking.

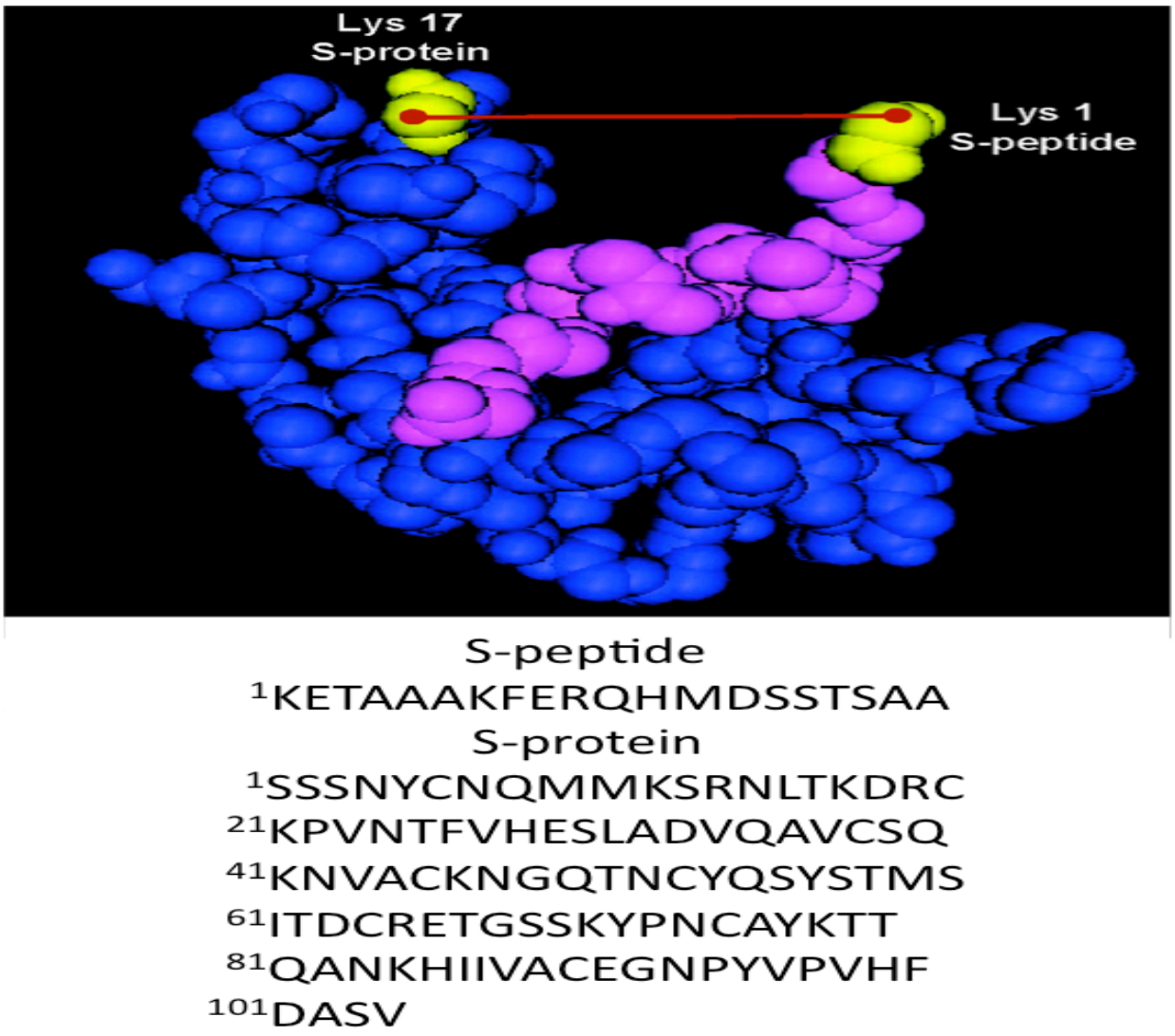


Figure 3.12. X-Ray Crystal Structure representation of Rnase S. X-ray crystal structure and amino acid sequence of RNase S complex. The S-protein region is indicated in blue while S-peptide is indicated in purple. The two amino acids that were found to be involved in the cross-linking reaction are indicated in yellow, and the red line represents the cross-linker.

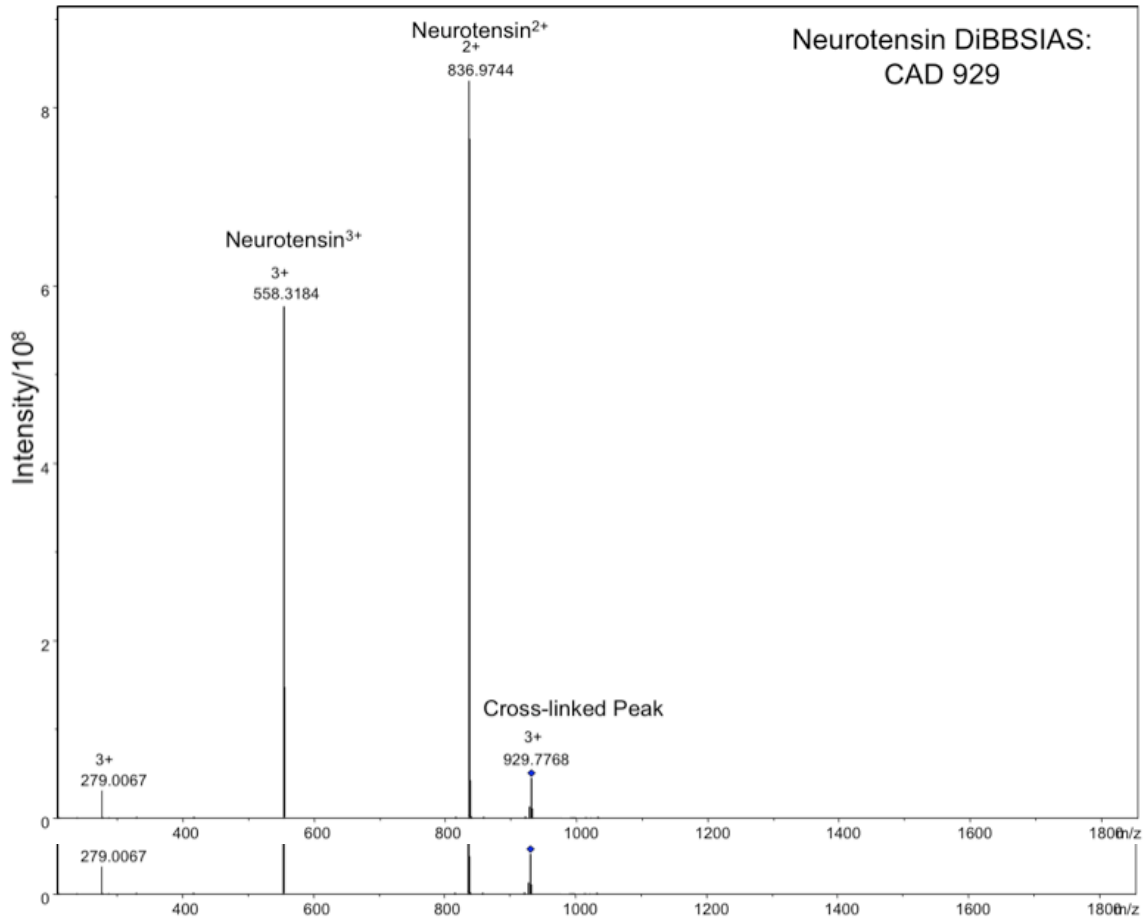


Figure 3.13. MS spectrum of neurotensin crosslinked with the second generation crosslinker, DiBBSIAS. This spectrum shows free neurotensin that was not crosslinked during the reaction as well as the crosslinked peak at *m/z* 994.

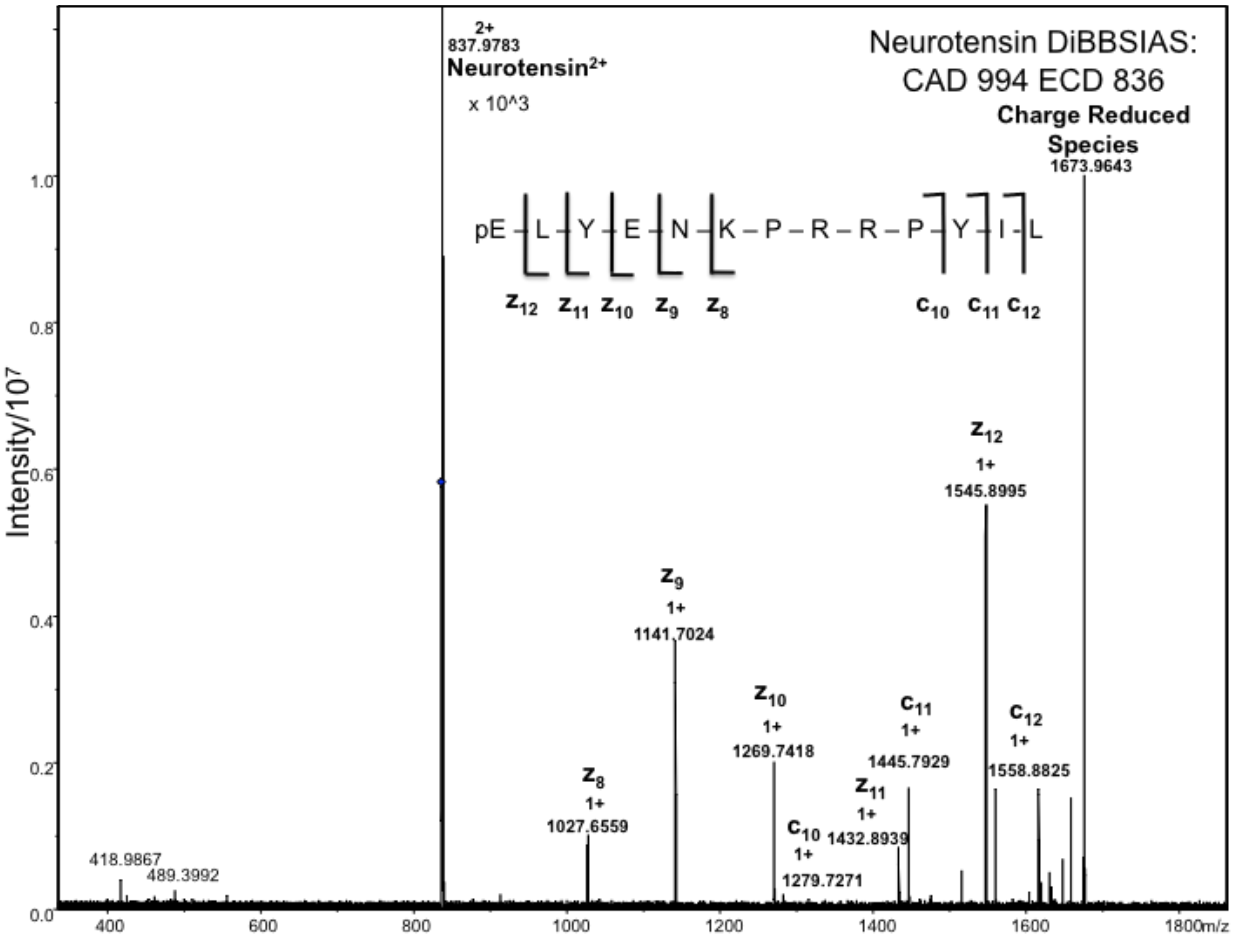


Figure 3.14. MS/MS spectra obtained from the fragmentation of the precursor peak at *m/z* 994, corresponding to the crosslinked neurotensin – DiBBSIAS – neurotensin molecule. The resulting fragment ions demonstrate the ability of DiBBSIAS to cleave off the chemical crosslinking reagent and leave on the peptide of interest to be further analyzed.

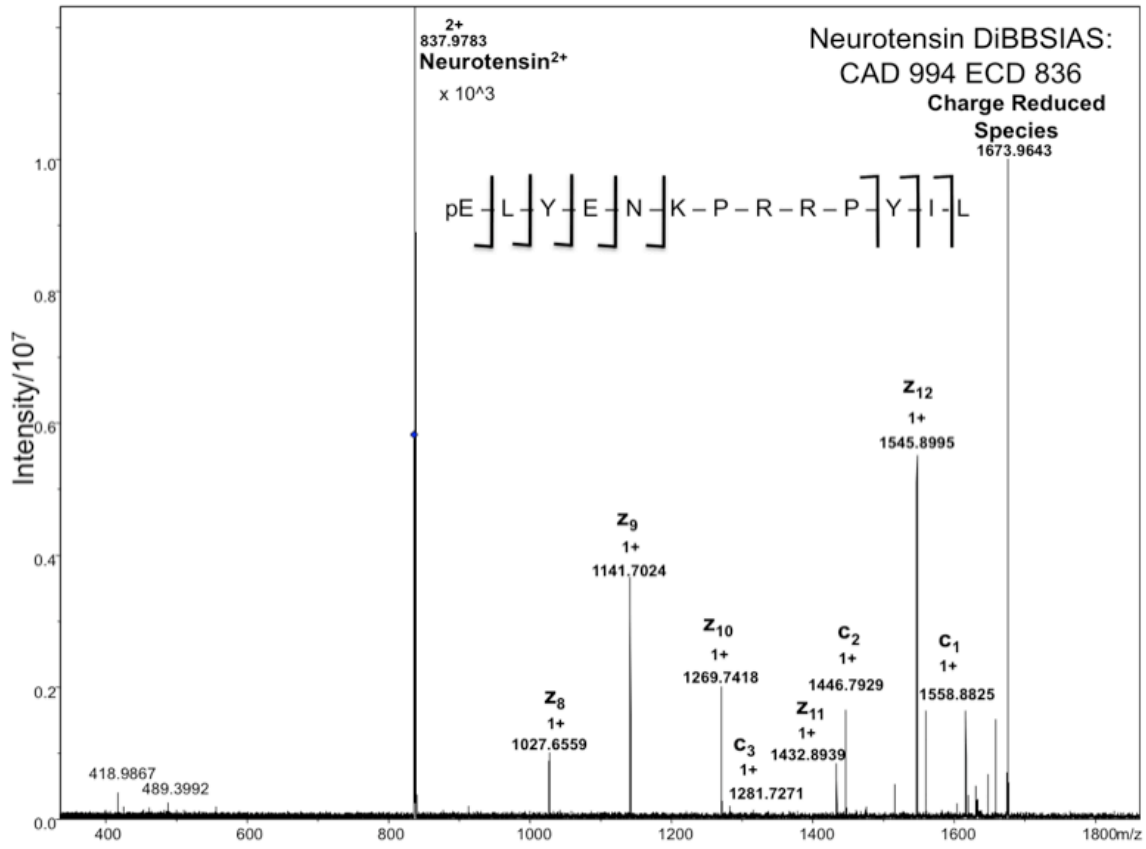


Figure 3.15. MS^3 spectrum that arose from the fragmentation of the secondary fragment, neurotensin2^+ at *m/z* 836. This peaks corresponds to the fragment ion that only contains neurotensin because the crossinker, DiBBSIAS, has been cleaved off by the previous CAD event.

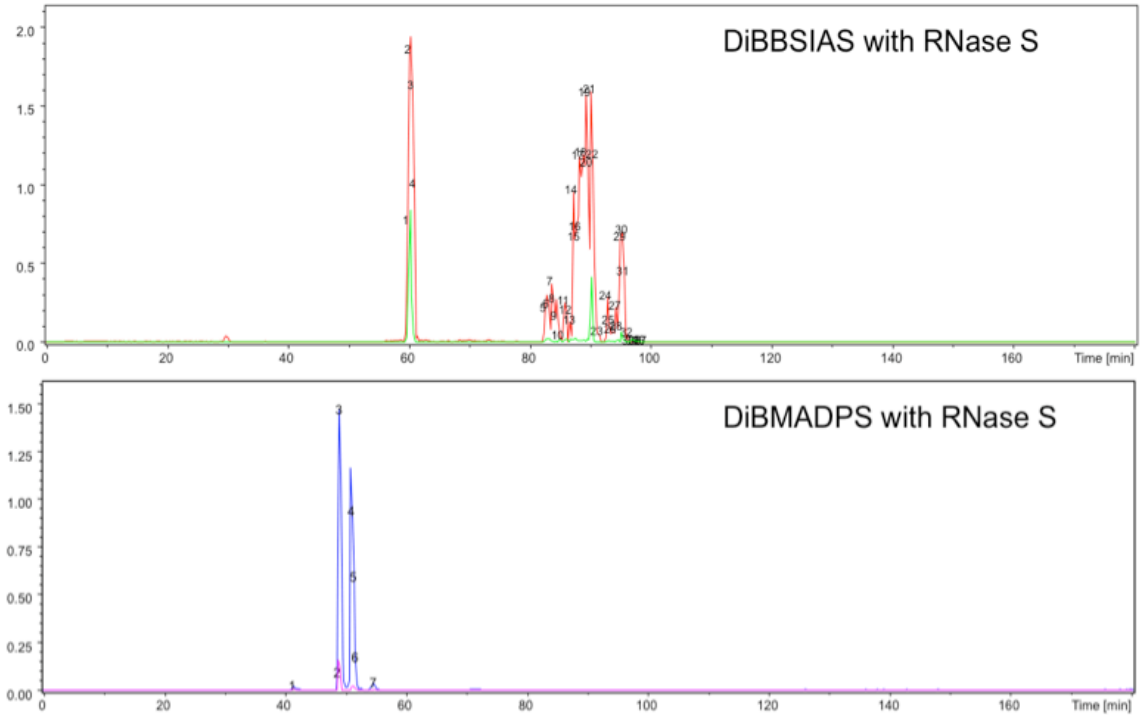


Figure 3.16. Comparison of the two LC-MS/MS Total Ion Chromatograms that arose when the two generations of crosslinkers were applied to the RNase S system. The top TIC represents the second generation crosslinker (DiBBSIAS) and the bottom the first (DiBMADPS).

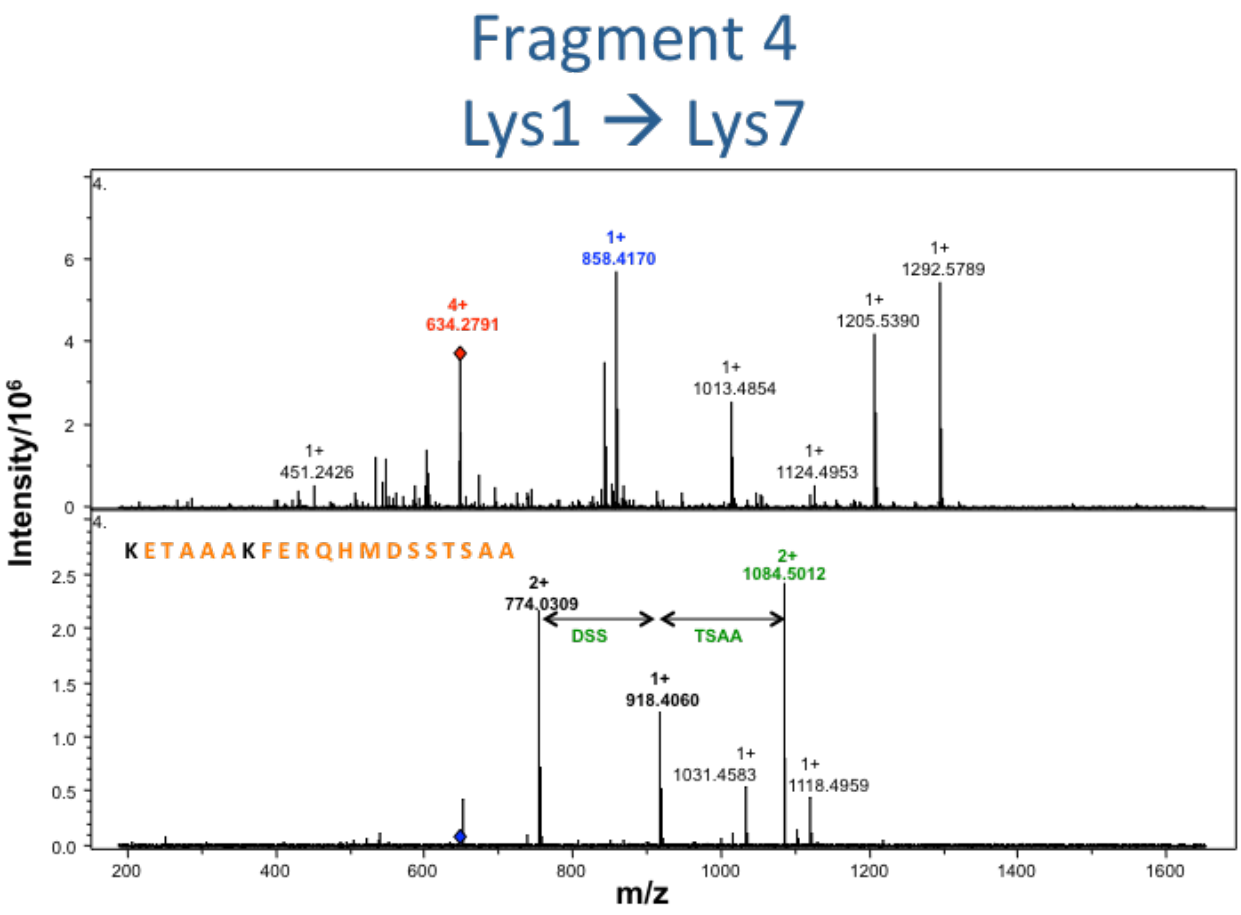


Figure 3.17. Mass spectrum [14] and MS/MS (bottom) spectrum from fraction 4 of the HPLC tryptic digest of RNase S linked with DiBBSIAS crosslinker. It is apparent from the fragmentation that one fractionion resulted from the CAD event on the precursor peak at 634^{4+}. This peak can be seen in green and corresponds to one peptide fractionfrom the RNase S complex. The other peaks in the MS/MS spectrum represent secondary fragmentation that occurred. The blue labeled peak in the MS spectrum represents a tryptic fraction of Rnase S.

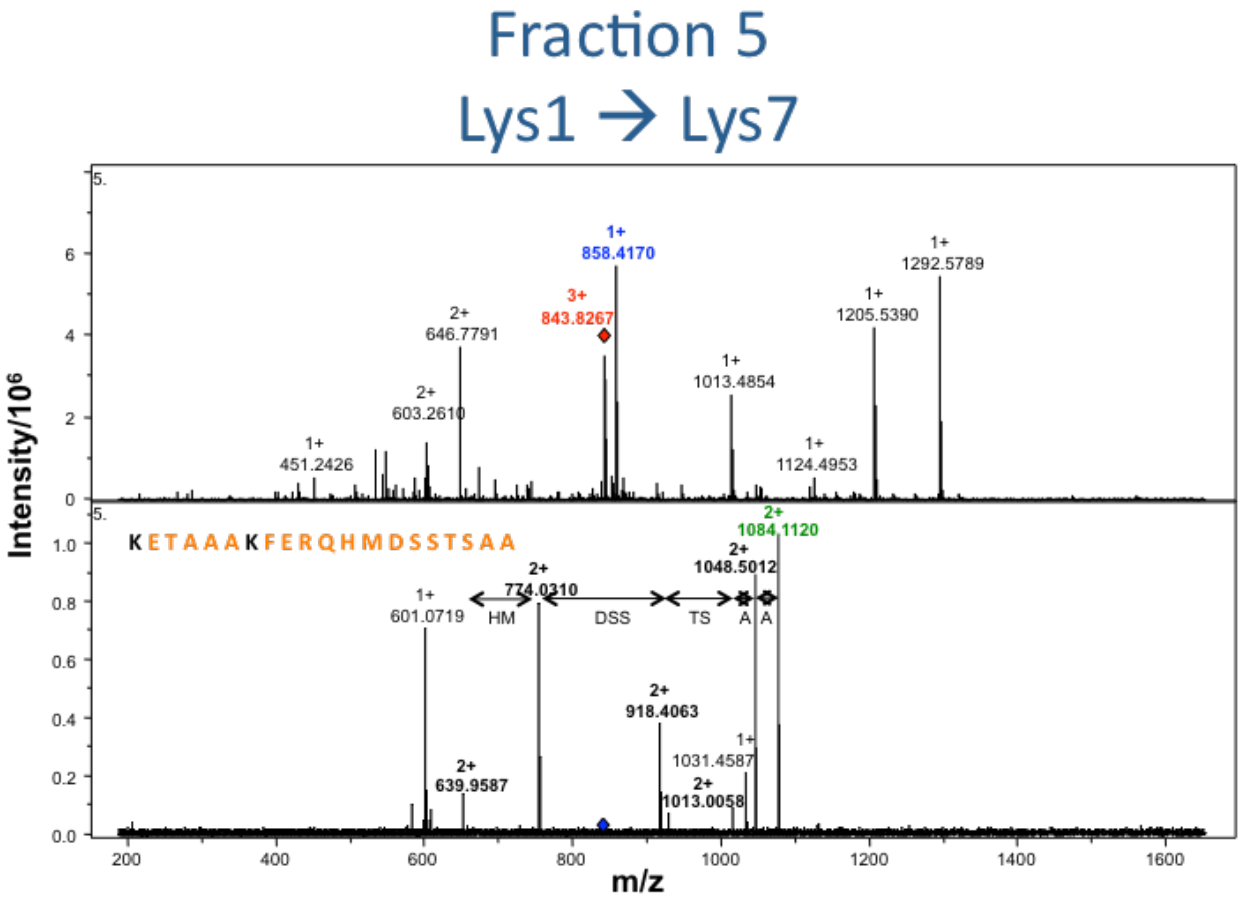


Figure 3.18. Resulting mass spectrum [14] and MS/MS (bottom) spectrum from fraction 5 of the RNase S linked with DiBBSIAS crosslinker. It is apparent from the fragmentation that one product ion resulted from the CAD event on the precursor peak at 843^{3+}. This peak can be seen in green and corresponds to one peptide fragment from the RNase S complex. The other peaks in the MS/MS spectrum represent secondary fragmentation that occurred. The blue labeled peak in the MS spectrum represents a tryptic fragment of Rnase S.

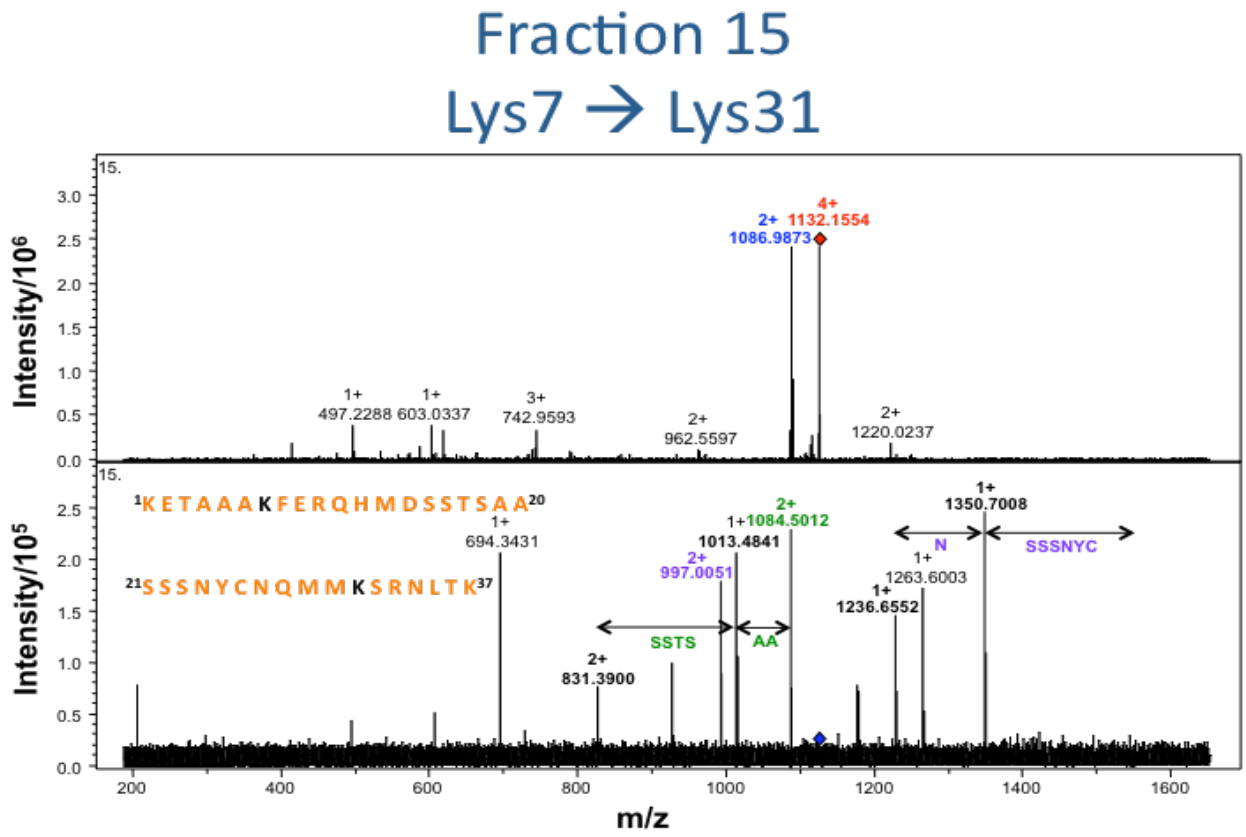


Figure 3.19. Resulting mass spectrum [14] and MS/MS (bottom) spectrum from fraction 15 of the RNase S linked with DiBBSIAS crosslinker. It is apparent from the fragmentation that one product ion resulted from the CAD event on the precursor peak at 1132^{4+}. This peak can be seen in green and corresponds to one peptide fragment from the RNase S complex. The other peaks in the MS/MS spectrum represent secondary fragmentation that occurred. These are labeled in green and purple. The purple fragments represent secondary fragments that arose from a second peptide fragment. The blue labeled peak in the MS spectrum represents a tryptic fragment of Rnase S.

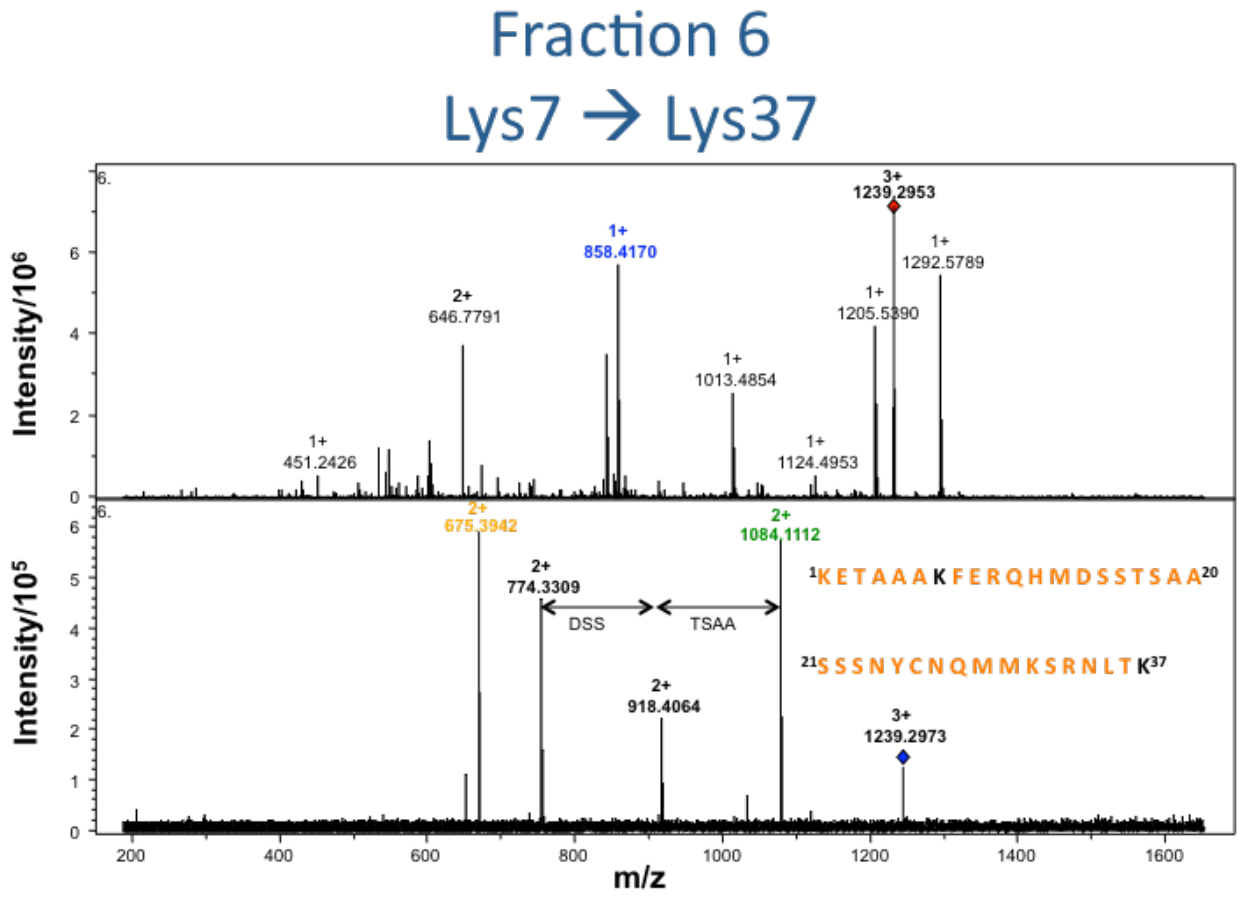


Figure 3.20. Resulting mass spectrum [14] and MS/MS (bottom) spectrum from fraction 6 of the RNase S linked with DiBBSIAS crosslinker. It is apparent from the fragmentation that two product ions resulted from the CAD event on the precursor peak at 1239^{3+}. These two peaks are seen in green and yellow and correspond to two different peptide fragments from the RNase S complex. The other peaks in the MS/MS spectrum represent secondary fragmentation that occurred. The blue labeled peak in the MS spectrum represents a tryptic fragment of Rnase S.

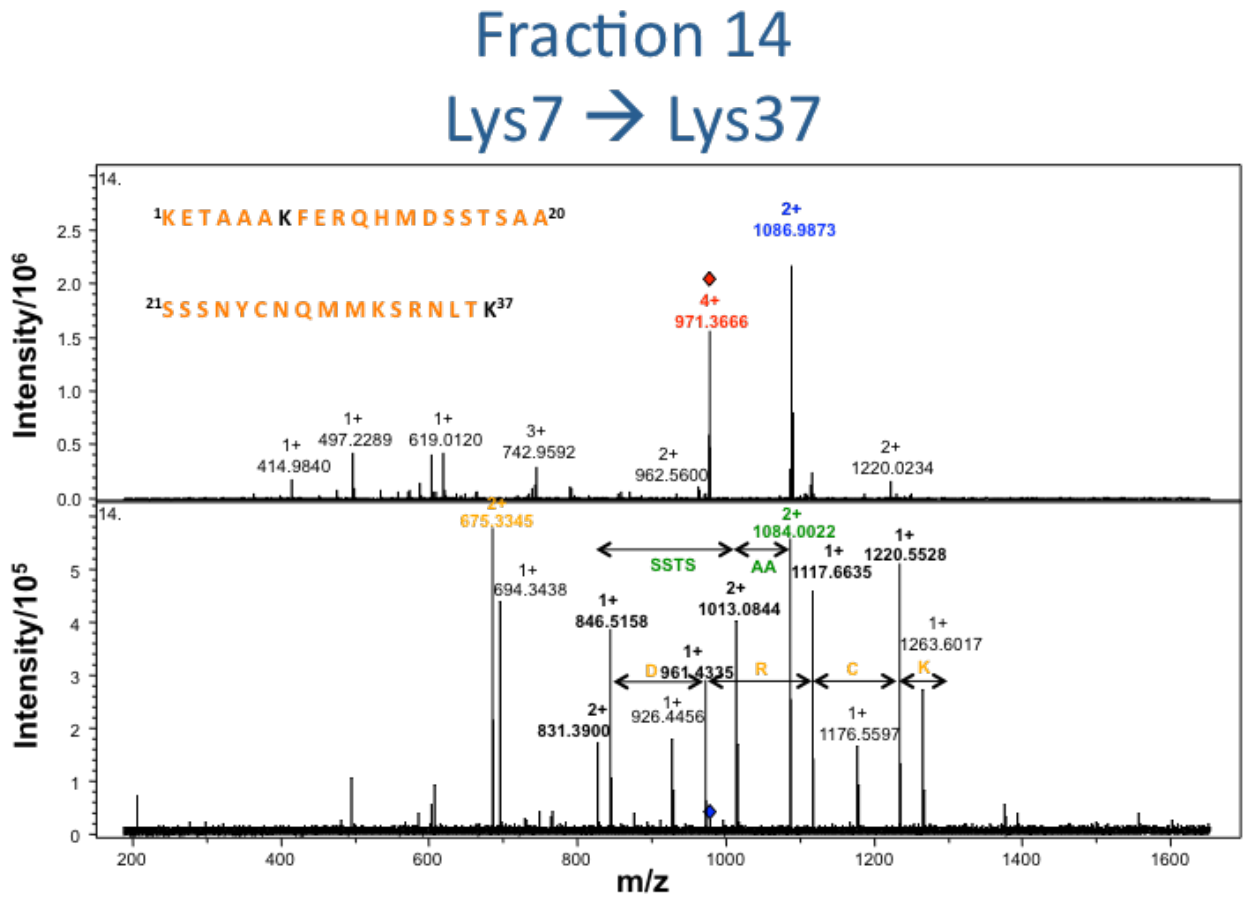


Figure 3.21. Resulting mass spectrum [14] and MS/MS spectrum (bottom) from fraction 14 of the RNase S linked with DiBBSIAS crosslinker. It is apparent from the fragmentation that two product ions resulted from the CAD event on the precursor peak at 971^{4+}. These two peaks are seen in green and yellow and correspond to two different peptide fragments from the RNase S complex. The other peaks in the MS/MS spectrum represent secondary fragmentation that occurred. The blue labeled peak in the MS spectrum represents a tryptic fragment of Rnase S.

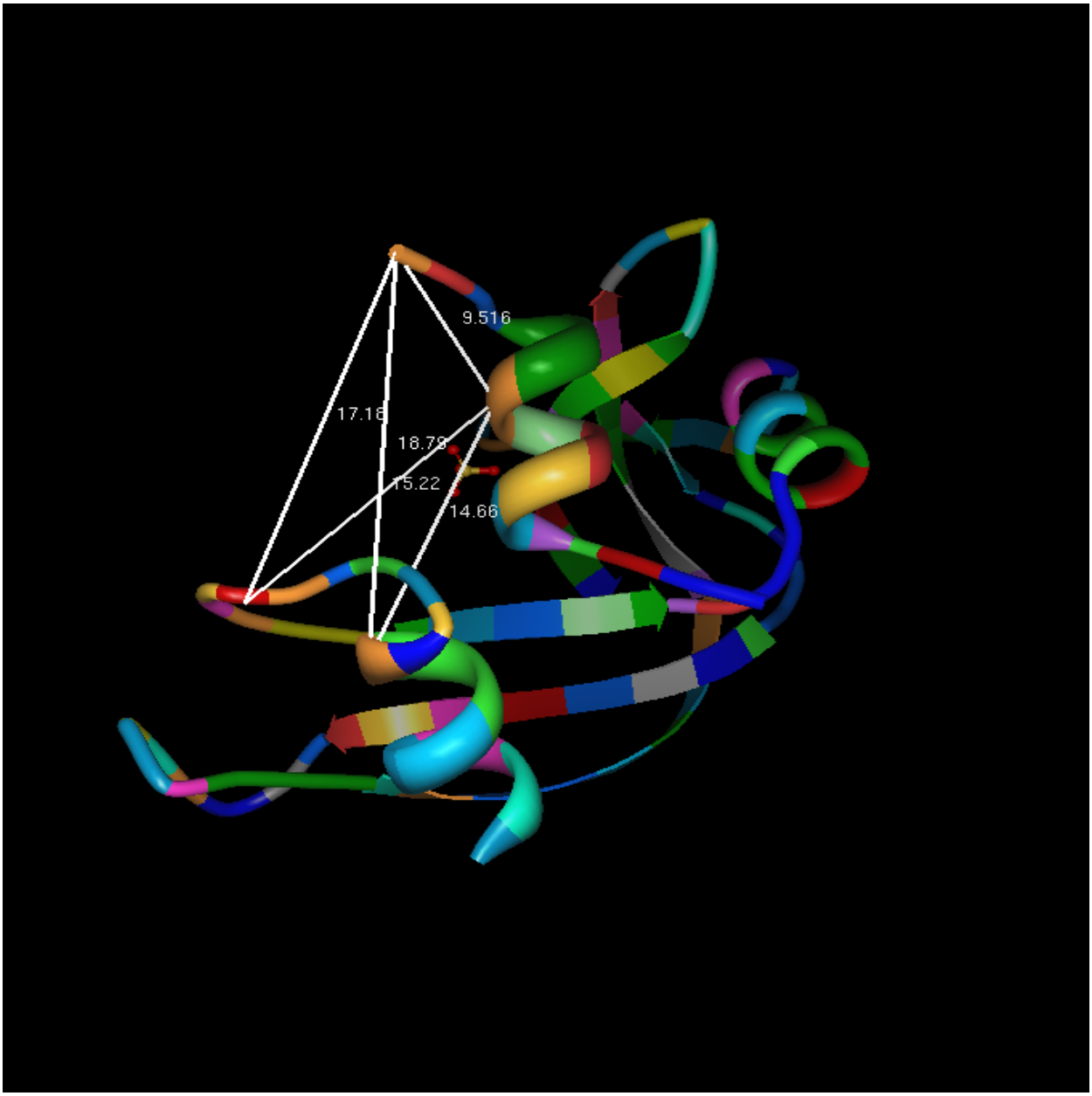


Figure 3.22. Figure of RNase S with all crosslinks found

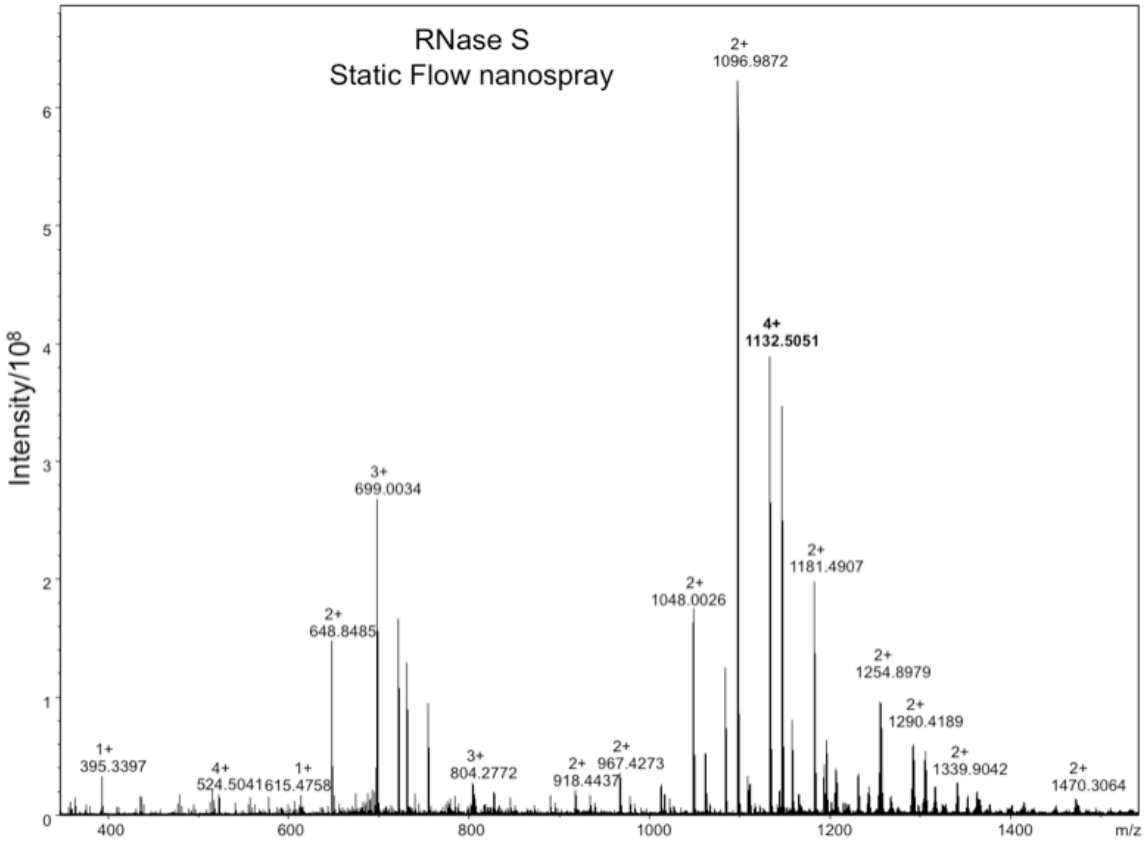


Figure 3.23. Staic nanospray RNase S spectrum

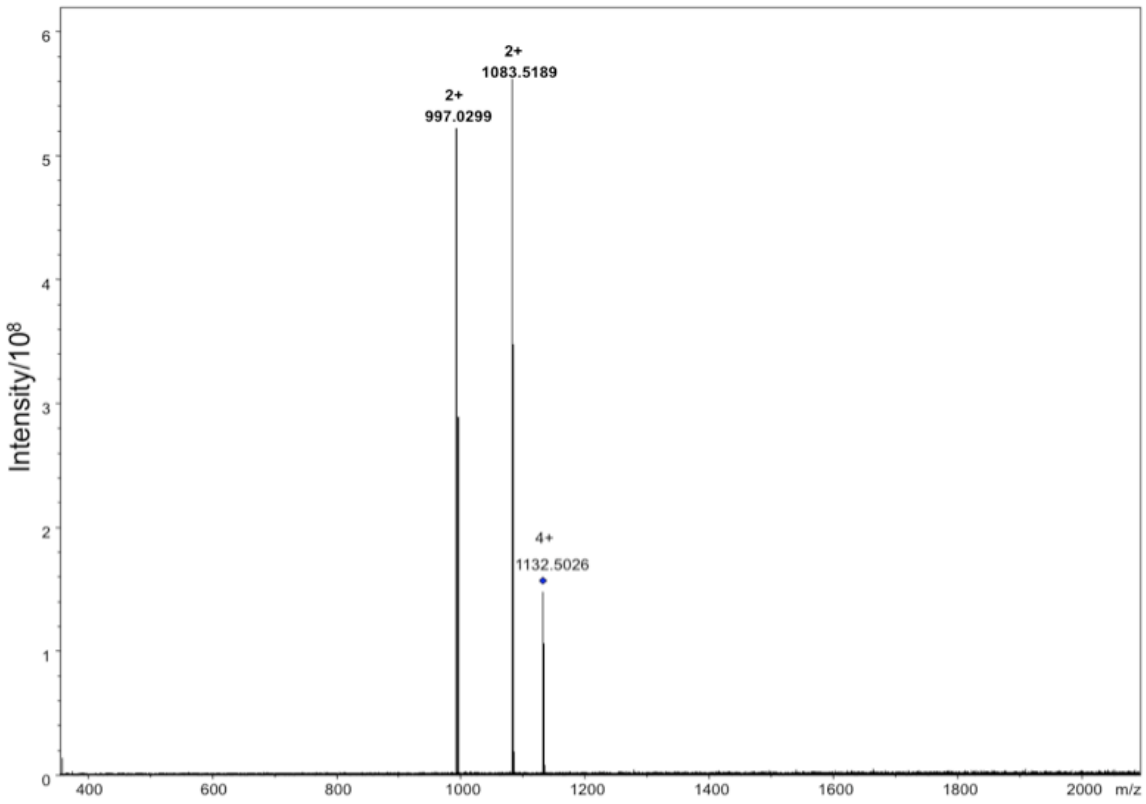


Figure 3.24. CAD spectrum from static nanospray experiment on Ribonuclease S, for *m/z* 1132^{4+}.

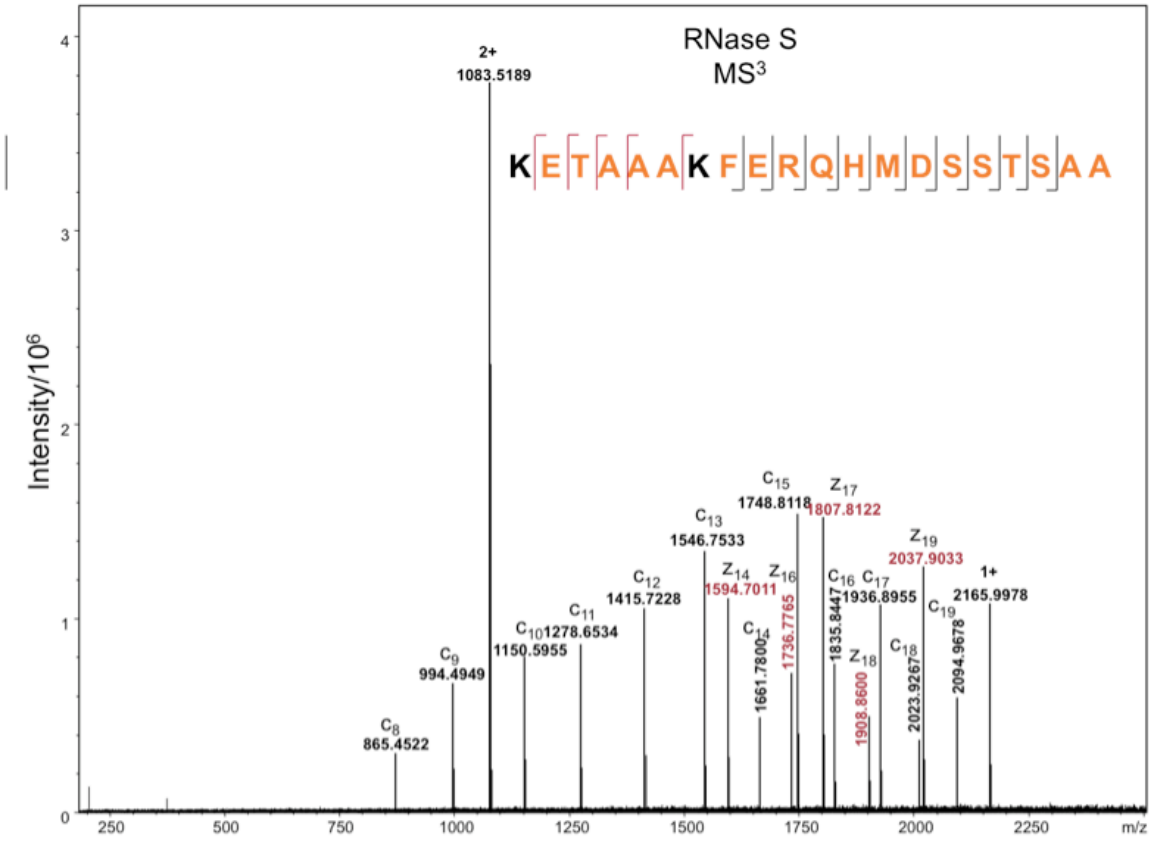


Figure 3.25. ECD spectrum from static nanospray experiment on Ribonuclease S, *m/z* 1083^{2+}.

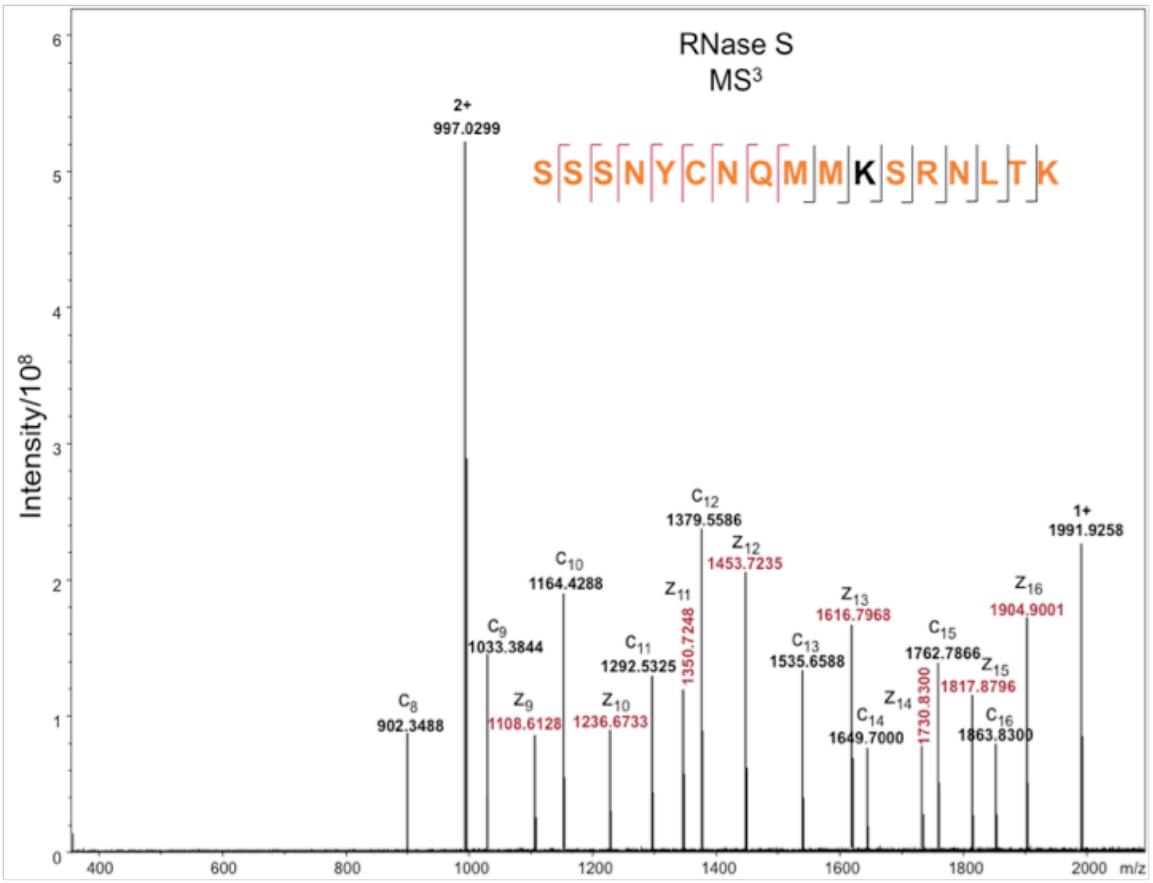
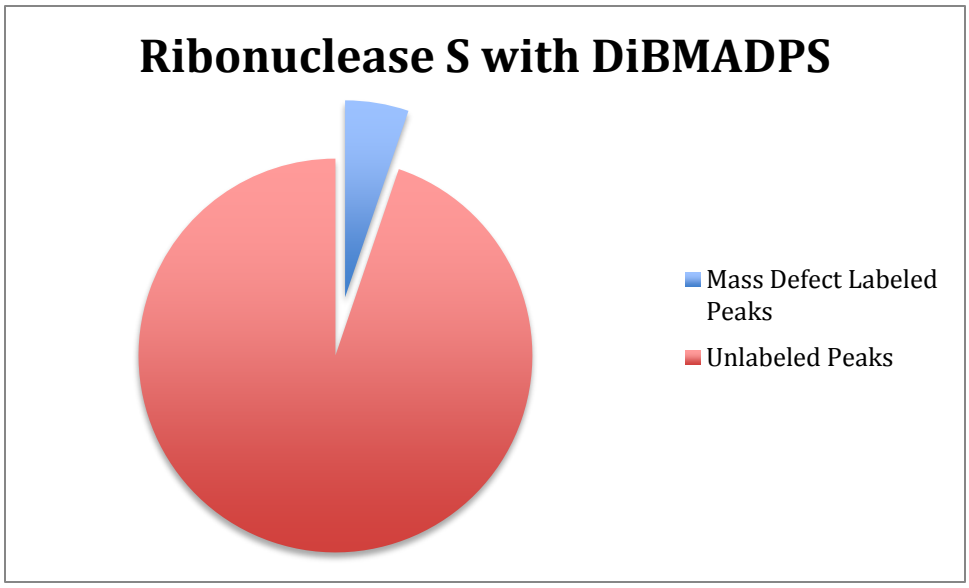


Figure 3.26. ECD spectrum from static nanopsray experiment on Ribonuclease S of *m/z* 997^{2+}.



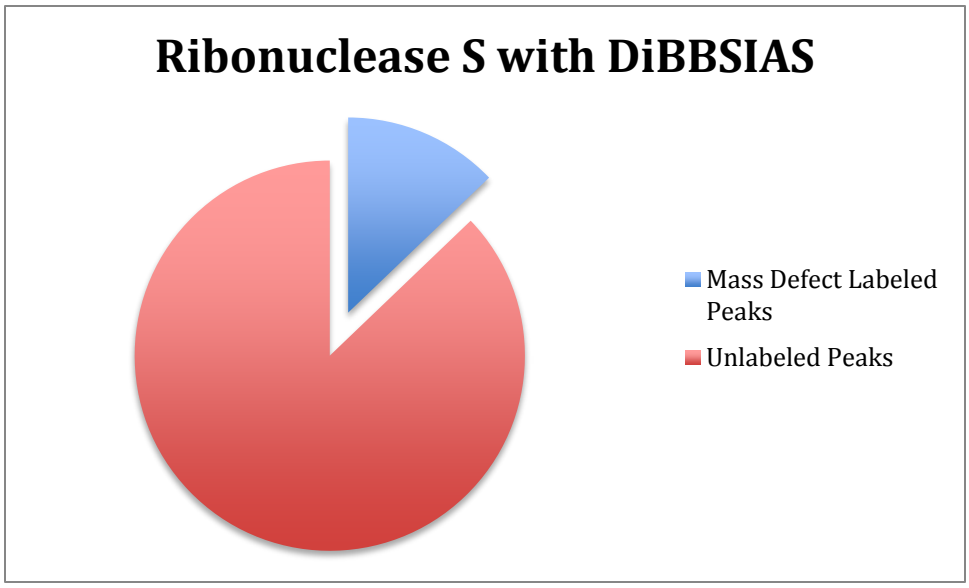


Figure 3.27. Histograms of MDL vs tryptic peaks in HPLC-FTICR-MS/MS experiment of Ribonuclease S reacted with both DiBMADPS and DiBBSIAS (bottom).

References:

1. Trakselis, M. A.; Alley, S. C.; Ishmael, F. T. *Bioconjugate Chem.* **2005**, *16*, 741.

2. Sinz, A. *J. Mass Spectrom.* **2003**, *38*, 1225.

3. Chiu, W.; Baker, M. L.; Jiang, W.; Dougherty, M.; Schmid, M. F. *Structure* **2005**, *13*, 363.

4. Rostom, A. A.; Fucini, P.; Benjamin, D. R.; Juenemann, R.; Nierhaus, K. H.; Hartl, F. U.; Dobson, C. M.; Robinson, C. V. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5185.

5. van Duijn, E.; Simmons, D. A.; van den Heuvel, R. H. H.; Bakkes, P. J.; van Heerikhuizen, H.; Heeren, R. M. A.; Robinson, C. V.; van der Vies, S. M.; Heck, A. J. R. *J. Am. Chem. Soc.* **2006**, *128*, 4694.

6. McKay, A. R.; Ruotolo, B. T.; Ilag, L. L.; Robinson, C. V. *J. Am. Chem. Soc.* **2006**, *128*, 11433.

7. Kluger, R.; Alagic, A. *Bioorg. Chem.* **2004**, *32*, 451.

8. Ihling, C.; Berger, K.; Hofliger, M. M.; Fuhrer, D.; Beck-Sickinger, A. G.; Sinz, A. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1240.

9. Kruppa, G. H.; Schoeniger, J.; Young, M. M. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 155.

10. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.

11. Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299.

12. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1.

13. McLafferty, F. W.; Fridriksson, E. K.; Horn, D. M.; Lewis, M. A.; Zubarev, R. A. *Science* **1999**, *284*, 1289.

14. McLafferty, F. W. *Int. J. Mass Spectrom.* **2001**, *212*, 81.

15. McLafferty, F. W. *Acc. Chem. Res.* **1994**, *27*, 379.

16. Novak, P.; Haskins, W. E.; Ayson, M. J.; Jacobsen, R. B.; Schoeniger, J. S.; Leavell, M. D.; Young, M. M.; Kruppa, G. H. *Anal. Chem.* **2005**, *77*, 5101.

17. Back, J. W.; de Jong, L.; Muijsers, A. O.; de Koster, C. G. *J. Mol. Biol.* **2003**, *331*, 303.

18. Comisaro.Mb; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *25*, 282.

19. Sinz, A. *Mass Spectrom. Rev.* **2006**, *25*, 663.

20. Muller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. *Anal. Chem.* **2001**, *73*, 1927.

21. Collins, C. J.; Schilling, B.; Young, M. L.; Dollinger, G.; Guy, R. K. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4023.

22. Taverner, T.; Hall, N. E.; O'Hair, R. A. J.; Simpson, R. J. *J. Biol. Chem.* **2002**, *277*, 46487.

23. Bennett, K. L.; Kussmann, M.; Bjork, P.; Godzwon, M.; Mikkelsen, M.; Sorensen, P.; Roepstorff, P. *Protein Sci.* **2000**, *9*, 1503.

24. Vasilescu, J.; Figeys, D. *Curr. Opin. Biotechnol.* **2006**, *17*, 394.

25. Tang, X. T.; Munske, G. R.; Siems, W. F.; Bruce, J. E. *Anal. Chem.* **2005**, *77*, 311.

26. Trester-Zedlitz, M.; Kamada, K.; Burley, S. K.; Fenyo, D.; Chait, B. T.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 2416.

27. Sinz, A.; Wang, K. *Biochemistry* **2001**, *40*, 7903.

28. Sinz, A. *Anal. Bioanal. Chem.* **2005**, *381*, 44.

29. Hurst, G. B.; Lankford, T. K.; Kennel, S. J. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 832.

30. Back, J. W.; Sanz, M. A.; De Jong, L.; De Koning, L. J.; Nijtmans, L. G. J.; De Koster, C. G.; Grivell, L. A.; Van der Spek, H.; Muijsers, A. O. *Protein Sci.* **2002**, *11*, 2471.

CHAPTER 4:

Probing the Actin – 34kDa Protein Complex Using A Mass-Defect Labeled Crosslinker

Protein-protein interactions are of extreme importance in the biological world. In order to probe this type of interaction, specific techniques have been developed which allow researchers the ability to freeze transient interactions into covalent bonds, which can be further studied. One such method of studying these interactions is mass spectrometry and chemical crosslinking. There is a wide array of chemical crosslinkers, however for this particular study, a novel mass-defect labeled chemical crosslinker was developed and utilized. This unique crosslinker was designed to utilize mass defect labeling techniques, and incorporates specific bonds capable of cleaving under low energy collisions. The reagent utilizes N-hydroxysuccinimde ester reactive groups to link to secondary amines in the protein system, a labile tertiary amine in the linker portion of the reagent for cleavage, and a dibromophenyl group, which serves as the mass defect label. We employed this specific chemical crosslinker to study the interaction that takes place between Actin and 34 kDa protein, in order to gain insight into the important intermolecular interactions between 34 kDa protein and actin.

Introduction. Hirano bodies are intracytoplasmic structures that are characterized by the presence of paracrystalline arrays of actin filaments. Hirano bodies have been found to be associated with a variety of neurodegenerative diseases, however, most of what is known about Hirano bodies has come from studies using autopsy derived tissue samples. Therefore little is known about the formation of Hirano bodies and their relationship to disease. Recently, a Hirano body cell model system has been developed based on expression of altered forms of the Dictyostelium 34 kDa actin-bundling protein. Intermolecular contacts between actin subunits are extensive, with each actin subunit in contact with four other subunits within the filament. A recent review by dos Remedios et al. catalogued over 160 separate proteins (excluding synonyms and isoforms) that bind to actin.[1] Classification of actin-binding proteins (ABPs) is complicated by the numerous cellular roles in which actin is involved. However, ABPs that are involved in the assembly and disassembly of actin filaments can be classified loosely into 8 major groups (Table 4.1). The specific ABP that this study is interested in is the Filament-crosslinking/bundling proteins. The filament-crosslinking proteins are an important group of proteins because they are responsible for the assembly of actin filaments into organized cellular structures. Crosslinking proteins can crosslink actin filaments into meshworks or into highly ordered arrays called bundles.[2] Proteins that crosslink actin filaments are characterized by the requirement of at least two F-actin binding sites that bind simultaneous to two adjacent

filaments, and as a general rule a large spatial separation between F-actin binding sites is characteristic of proteins that crosslink actin filaments while closely spaced F-actin binding sites are common for bundling proteins.[3] The best characterized family of actin crosslinking proteins is found in the Calponin Homology (CH) domain family of proteins.[4] The smallest and simplest CH protein is fimbrin, which is monomeric and has two actin binding sties along the length of its polypeptide chain. Because of its small size, fimbrin crosslinks actin filaments into bundles and is localized to filopodia and microvilli.[5] Another example of a CH protein is α-actinin which crosslinks and bundles actin filaments in both stress fibers and filopodia.[6,7] Unlike fimbrin, α-actinin is a homodimer and has only one F-actin binding site per polypeptide chain; therefore, the dimerization of α-actinin is necessary for its crosslinking ability. It is important to note that not all crosslinking proteins are CH proteins. For example, APB-34 or 34 kDa actin bundling protein, is an F-actin crosslinking protein found in Dictyostelium.[8,9] Interestingly, even though the 34 kDa protein shows no homology to the CH proteins, it does co-localize with α-actinin in filopodia.[10]

The 34 kDa actin-bundling protein is one of eleven actin crosslinking proteins in Dictyostelium.[2] The 34 kDa protein is dynamic in vivo, and its cellular localization often reflects the dynamic behavior of the actin cytoskeleton. Consequently, it localizes to a number of important actin-dependent structures. For example, it can be found at the leading edge,[11] in

filopodia,[12] at the point of cell-cell contacts,[13] an in phagocytic ups.[14] Clearly, the 34 kDa protein is involved in many diverse functions. Molecular dissection of 34 kDa protein using truncated forms of 34 kDa protein identified three regions important for actin binding. The three regions identified were residues 1-123, 193-254 and 278-295.[15]

Chemical crosslinking coupled with FTICR - tandem mass spectrometry can be used to elucidate low-resolution three-dimensional structures of large biomolecules, as well as the interactions that take place within the molecules, specifically protein-protein interactions fairly rapidly, from a very small amount of sample and without the concern of interfering compounds, therefore it will be employed to elucidate the interaction that takes place between Actin and 34 kDa protein. A specific chemical crosslinker has been designed to facilitate the identification of intermolecular interactions within protein-protein complexes. This reagent employs mass defect label technology to facilitate the identification of peptides that have undergone a reaction from the larger fraction that are unmodified. This crosslinker is designed with labile bonds that release the peptides by low energy CAD, facilitating their identification. The efficiency of CAD is high enough to permit MS^3 sequencing of the released peptides. The goal of this study is to verify the three actin-binding sites identified by Lim *et al.*. Mapping these interaction sites on a model of the actin filament will help reveal the structural basis for stabilization of actin filaments and inhibition of disassembly, which are two elements required for the formation of Hirano

bodies. In order to do this, we utilized two conventional chemical crosslinkers, 1-Ethyl-3-(3-dimethylaminopropyl-carbodimide) hydrochloride (EDC) and bis-sulfosuccinimide suberate (BS^3) in addition to the mass defect labeled reagent (DiBMADPS).

Experimental. *Chemicals.* 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDC) and the supported chemical modification reagent, N-hydroxysulfosuccinimide (Sulfo-NHS) were purchased from Pierce (Rockford, IL). Bis-sulfosuccinide suberate (BS^3) was obtained from ThermoScientific (Waltham, MA). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). The mass defect labeled reagent (DiBMADPS) was synthesized by UVic-GBC Proteomics Centre in Victoria, BC.[16]

Expression and purification of recombinant proteins. 34 kDa protein was expressed in the *E. coli* expression strain BL21(DE3) using the pET-15b vector system (Novagen EMD), and purification of recombinant protein was done as previously described.[17,18] Protein concentrations were determined by the BCA method[19] and protein samples were dialyzed against 2 mM PIPES pH 7.0, 50 mM KCl, 0.2 mM DTT, 0.02% NaN_3 and stored at -80°C.

Actin. Rabbit skeletal muscle actin was prepared from acetone powder [1]. The Sephadex G-150 actin fraction was maintained for up to one week in 2 mM Tris-HCl pH 8.0, 0.2 mM $CaCl_2$, 0.2 mM ATP, 0.2 mM DTT at 4°C with daily buffer changes. G-actin used for chemical

crosslinking was exchanged into 20 mM imidazole pH 7.5, 0.2 mM $CaCl_2$, 0.2 mM ATP, 24 hours prior to use.

Chemical crosslinking reactions. 18 μM actin alone or with 6 μM 34 kDa protein was polymerized at 4°C overnight in 20 mM PIPES pH 7.0, 50 mM KCl, 1 mM $MgCl_2$, 1 mM ATP, and 5 mM EDTA. Freshly prepared stocks of either BS^3 (10 mM in H_2O), EDC/NHS (5 mM each in DMSO), or DiBMADPS (10 mM in DMSO) were used for chemical crosslinking. Final concentrations of each crosslinker were dependent on the total molar concentrations of protein to be crosslinked. The final crosslinker to protein concentrations used were: BS^3 20:1, EDC/NHS 25:1, and DiBMADPS 10:1. All crosslinking reactions were incubated at room temperature for 60 minutes and stopped by quenching with 1 M Tris-HCl pH 8.0 to a final concentration of 10 mM. After quenching, all samples were prepared for SDS-PAGE by dilution with 4X reducing Laemmli sample buffer and boiled for 5 minutes. Samples were subjected to SDS-PAGE [2] and stained with Coomassie blue.

Western blot analysis of crosslinked species. This procedure was performed in order to identify the components generated by the chemical crosslinking reaction. Procedural steps were taken from previously published literature.[10] In order to identify the species present, either an anti-34 kDa monoclonal antibody (B2C)[11] or an anti-actin monoclonal antibody (MabGEa; produced by Dr. Rich Meagher, UGA)[12] and a horseradish peroxidase-conjugated goat anti-

mouse antibody was used. Visualization of immunoreactive bands was accomplished by chemiluninescence (Pierce Biotechnology, Rockford, IL).

Procedure for in-gel trypsin digestion. The procedure for digestion of the gel bands of interest can be found elsewhere[13] and these procedures were followed with slight modifications. The gel bands were excised using a clean razor blade, cut into 1 mm^3 pieces and placed in an eppendorf tube that was previously cleaned with methanol. The individual pieces were then destained using several aliquots of 50 mM NH_4HCO_3 and 50% methanol. After the pieces were completely destained, the final solution was removed, and the gel pieces were dehydrated in 75% acetonitrile for 20 minutes at room temperature. After removing the acetonitrile solution, the gel plugs were dried at 40 °C. Reduction of the disulfide bonds was completed by inserting the dried gel plugs in a solution of 10 mM DTT, 20 mM NH_4HCO_3 for 45 minutes at 56 °C. This solution was then decanted off the gel plugs and a solution of 100 mM iodoacetamide, 20 mM NH_4HCO_3 was added in order to alkylate the cysteines. This solution was incubated in the dark for 30 minutes at room temperature. The gel plugs were then washed with 50% methanol in 50 mM NH_4HCO_3 for 20 minutes at room temperature. This solution was decanted and the gel pieces were again incubated, this time with 75% acetonitrile for 20 minutes at room temperature. The gel pieces were then dried at 40 °C. Dried gel plugs were then processed in-gel digestion using 100 μL of trypsin (20 μg/mL) in 20 mM NH_4HCO_3 at 37 °C for 12 hours. The trypsin-digested

peptides were extracted from the gel plugs by incubation in 50% acetonitrie/0.1% TFA at room temperature for 20 minutes, and then centrifuged for 30 seconds at low speed. The resulting supernatant was removed and saved in a clean eppendorf tube. A second extraction was performed on the gel plugs, and the supernatants from both extractions were pooled and dried *in vacuo* to < 5 µL. These final peptide solutions were volume adjusted to 10-15 µL in 50% acetonitrile/0.1% TFA and stored at -80 °C.

Mass Spectrometric Analysis. Experiments were performed with a 9.4 T Bruker Apex QhFTMS (Billerica, MA) fitted with an Apollo II dual source, a 25 W CO_2 laser (Synrad model J48-2, Mukilteo, WA) for IRMPD, and an indirectly heated hollow cathode for generating electrons for ECD experiments. Bruker Apex Control was used for data acquisition. Bruker Data Analysis (version 3.4 Build 72) was used for most of the data analysis work. Protein samples were dissolved in a solution of 50% Methanol:50% H_2O:0.1% Formic acid in order to obtain the desired charge states and intensity. Tryptically digested peptides can be ionized by either regular flow ESI or nanospray. For regular-flow ESI, sample solutions were infused at a rate of 120 µL/hour. For nanospray, samples were infused at a rate of 10 µL/hour using a pulled fused silica tip (model #FS360-20-10-D-20, New Objective, Woburn, MA). Solution conditions would vary based on the amount of protein that was present prior to crosslinking. The only samples that were analyzed by direct infusion ESI were those not analyzed by SDS-PAGE (i.e. Neurotensin,

Bradykinin and some Rnase S samples). For each mass spectrum, 512K points were acquired at a 2.4 MHz digitization rate, padded with one zero fill, and appodized using a sinebell window. Internal calibration was performed using confidently assigned tryptic fragment ions as internal calibrants, providing mass accuracy of <1 ppm.

Internal calibration was performed using Bruker Data Analysis. Peaks were automatically picked using the "SNAP" peak picking feature in Data Analysis with the "S/N threshold" set to 0.001, the "Relative intensity threshold (base peak)" set to ≤ 0.001%, and the "Absolute intensity threshold" set to 0.001. This peak-picking feature was used specifically because it is designed for peptide/protein analysis.

High Performance Liquid Chromatography Analysis (HPLC) Coupled with Tandem Mass Spectrometry. Separation of 34 kDa - Actin solutions were analyzed using an Agilent 1100 series HPLC (Santa Clara, CA). Samples were first injected by the autosampler and loaded on a C18 column (PepMap C18, 300 μm x 15 cm, 3 μm, 100 Å, LC Packing) at a flow rate of 0.4 μL/min with 95/5 (v/v) H_2O (with 0.1% formic acid)/Acetonitrile (with 0.1% formic acid). Peptides were then separated using the following gradient: O% B for 0-5 mins, 20-75% B for 5-120 mins, 75-90% B for 120-125 mins, 90-95% B for 125-130 mins, 0% B for 130-140 mins. Solvent B was 0.1% formic acid in Acetonitrile. The column was directly linked to the inlet of the mass spectrometer and therefore the analysis was completed simultaneously. Auto MS/MS

was set up to collect compounds from a prefer list compiled by the user and LC/MS/MS mass spectra were acquired using HyStar software (version 3.0 Bruker Daltonics). MS/MS data acquisition was set in automatic mode with active exclusion based on peak intensity and selections of prefer peak lists. Five precursor ions were selected from each MS scan and excluded after two MSMS scans for 1 minute. Calibration of the instrument prior to any analysis was achieved by injecting a tryptic digest of BSA (Bovine Serum Albumin, Sigma, St. Louis, MO).

MS2Links Software on MS3D Web-portal. The Collaboratory for MS3D was designed by structural biologists working in the area of mass spectrometric based methods for the analysis of tertiary and quaternary macromolecular structures (MS3D). This web-portal was designed to provide collaborators with a shared working environment where data can be uploaded and analyzed. Specially, a series of programs was designed for the analysis of MS and MS/MS data provided by substrates treated with certain chemical probes or modifications. This software consists of the Links [3] and MS2Links [4] programs, which were initially designed for the analysis of data obtained from bottom-up and top-down analysis of crosslinked/covalently labeled proteins and nucleic acids [5]. In this experiment we used MS3Links and identified the specific crosslinked species formed when either EDC, BS^3, DiBMADPS or DiBBSIAS were crosslinked with Neurotensin, Bradykinin, RNase S, 34 kDa, Actin or 34-Actin complexes. The robust user

interface (www.ms3d.org) on both the Links and MS2Links software allowed for the direct input of the protein sequences, chemical modifications (including crosslink mass and active site), and peak list. Using this interface, each peak list generated by a HPLC-MS/MS run was subjected to analysis and the peaks containing the modified crosslinker were revealed. Although this software is not 100% accurate in the assignments that it makes, it does allow for a much greater ease in analysis.

Results and Discussion. In most biological experiments, crosslinked peptides or protein complexes tend to be relatively large and of high mass-to-charge (*m/z*). This often makes extensive fragmentation by MS/MS difficult. As the complexity of the mass spectrum increases, the number of peptides with the same nominal mass but differing amino acid sequences increases. Therefore, the high resolution and mass accuracy that FTICR-MS offers, combined with the ability to comprehensively fragment large peptides by tandem mass spectrometry are indispensable when conducting a crosslinking experiment, and attempting to unambiguously assign the cross-links. This combined with mass defect label technology facilitated the identification of specific sites of interaction between 34 kDa protein and Actin in *Dictyostelium.*

Chemical crosslinking of 34 kDa protein and actin. In order to examine the molecular interfaces between 34 kDa protein and actin, we incubated polymerized F-actin and 34 kDa with

three distinct chemical crosslinking agents. Each of these crosslinkers posses' unique qualities, which allow us to obtain different information from each reagent.

The first crosslinker used in this study was the zero-length crosslinker, . 1-Ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDC). This is a commercially available crosslinker that is used to couple carboxyl groups to primary amines. EDC is popular because it provides a neutral linkage without an interfering crosslink, excess reagent and crosslinking byproducts can be washed away using only water, and it is water soluble and very pure. Following the crosslinking reaction, the products were separated using SDS-PAGE (Figure 4.1). Control loads of noncrosslinked actin and 34 kDa protein are shown for comparison. The major products generated with EDC/sulfo-NHS were then estimated to have molecular weights of 191 kDa, 120 kDa, 106 kDa, 87 kDa, and 78 kDa (Figure 4.1). Western blot was then utilized in order to identify the protein components of each new complex using monoclonal antibodies for actin (MAbGEa) and 34 kDa protein (B2C). It was found that the bands at 120 kDa and 87 kDa were immunoreactive for actin only and therefore most likely represented the crosslinked actin trimer and dimmer respectively, as actin has a molecular weight around 60 kDa. The 191 kDa and 78 kDa bands however were found to be immunoreactive for both actin and 34 kDa protein. It was believed that the 78 kDa band was comprised of one actin molecule and one 34 kDa protein and the 191 kDa band a combination of four actin molecules and 1 34 kDa protein,

however this molecular weight could also correspond to three actin molecules and two 34 kDa proteins. It was also found that the 106 kDa band resulted in a mixed immunoreactive results and therefore was not used further.

The second crosslinker studied was the homobifunctiona, water-soluble, non-cleavable and membrane impermeable crosslinker Bis(sulfosuccinimidyl)suberate (BS^3). It contains an amine-reactive N-hydroxysulfosuccinimide (NHS) ester at each end of an 8-carbon spacer arm. NHS esters reactive with primary amines at pH 7-9 to form stable amide bonds, along with the release of the N-hydroxysulfosuccinimide leaving group. Due to the fact that it contains a hydrophilic sulfo moiety it is soluble up to ~10 mM in water and many other commonly used buffer systems. This crosslinker is also very pure and has a spacer length of approximately 11.4 Å. The same reaction conditions were used as described for EDC/sulfo NHS except that the concentration of BS^3 was 600 μM. The products of the BS^3 reaction mixture were again separated by SDS-PAGE (Figure 4.2) and four novel bands were discovered with approximate molecular weights of 153 kDa, 118 kDa, 88 kDa and 78 kDa. Western blot analysis was completed and it was found that the 153 kDa, 118 kDa, and 88 kDa bands were immunoreactive for only actin. These bands thus represent the tetramer, trimer, and dimmer species of actin respectively. Similarly to the EDC experiment, BS^3 produced a band at 78 kDa, which was immunoreactive for 34 kDa protein but gave ambiguous results for actin immunoreactivity.

Over-exposure of blots in the presence of anti-actin antibody always results in varying but slight immunoreactivity but at the cost of making other bands on blots unresolvable (data not shown). Based on this observation and the similarity to the 78 kDa band in the EDC/sulfo-NHS experiments, it is mostly likely the case that the 78 kDa band is composed of one 34 kDa protein and one actin molecule.

The final crosslinker used was the novel mass defect labeled reagent, DiMADPS, developed to be amine reactive with an approximate spacer arm length of 12.04 Å. This crosslinker is designed similarly to the BS^3 crosslinker, however it has an additional feature designed specifically to aid in the identification of crosslinked species in the subsequent mass spectrometry analysis. This reagent contains two bromine molecules that alter the mass defect of the proteolytic crosslinked species significantly, thus separating them from the basic tryptic fragments produced and making identification easier. Mass defect label technology has been employed in order to facilitate the easy identification of a cross-linked species within a complex mass spectrum. A mass defect is the difference between the exact monoisotopic mass of a compound and its nominal molecular weight, or the weight that is based on the nucleon values of the most abundant isotope of each element, e.g., 12 amu for Carbon, 16 amu for Oxygen, ect. Elements from the first two rows of the periodic table are utilized to construct most peptides and proteins, and these elements have an average mass defect range of ±0.08 amu. Therefore for

every 100 amu molecular weight the mass defect will be approximately ±0.05 (a 1 kDa protein has a mass defect of approximately +0.5 amu and a 2 kDa peptide has a mass defect of approximately + 1 amu) [6]. Although peptides have an intrinsic mass defect due to the atoms with which they are assembled, the overall mass defect range for a number of different peptides is generally quite narrow, causing peptide molecular weights at a given nominal mass to occupy only a small area of the entire unit mass. Thus it can be understood that because the peptide molecular weights only occupy about a third of the nominal mass range, there is significant overlap and individual peptide identification becomes difficult. Greater specificity in identifying individual peptides could be accomplished if they were spaced out among the nominal mass range. In order to do this, we added bromine molecules to our cross-linker design. Bromine has a mass defect of – 83 amu and therefore any peptides that contain the cross-linker will be shifted to lower nominal masses and away from the cluster of simple tryptic peptides. Therefore, the DiBMADPS crosslinker was included in the study to increase the probability of identifying the rare 34 kDa-actin crosslinked peptide complexes in the mass spectrum.

Crosslinking of 34 kDa protein and F-actin was performed as described for EDC-sulfo-NHS and BS^3 with a concentration of 600 μM DiBMADPS. As anticipated, the DiBMAPDS crosslinking reaction mixtures separated by SDS-PAGE gave similar results as the reaction mixtures with BS^3. Further, the DiBMADPS reaction mixture contained four major bands of

estimated molecular weights 153 kDa, 118 kDa, 90 kDa, and 78 kDa (Figure 4.3) just as BS^3. However, there was an additional band at 66 kDa that was not observed with either BS^3 or EDC-sulfo-NHS. Just as was found before, the bands at 153 kDa, 118 kDa and 88 kDa were immunoreactive for actin only, and the 78 kDa band was immunoreactive for 34 kDa protein, and not actin (as was stated before using BS^3). Even after overexposure however, the 66 kDa band was immunoreactive for only 34 kDa, leading to the conclusion that it was the crosslinked dimer of 34 kDa protein. This was an exciting result as none of the other crosslinkers used identified this species. Each of the identified crosslinks were then subjected to mass spectrometry analysis for verification of the predicted compositions, and to hopefully identify the amino acids involved in the 34 kDa – Actin interaction zone.

Mass Spectrometry coupled with HPLC. Positively identified bands were excised from the gels, following in-gel tryptic digestion and analyzed using HPLC-MS/MS in order to identify the crosslinked species. Although the HPLC was able to separate the peptides produced into a number of different peaks to be automatically analyzed by the mass spectrometer, the abundance and intensity of the actual crosslinked species was so small that a number of different "tricks" needed to be employed in order to obtain any useful information. In order to identify and analyze the crosslinked peaks, the automated MS/MS program was set to exclude the specific tryptic fragment peaks of both actin and 34 kDa protein, and the S/N ratio was set to 1×10^2, thereby

allowing the MS/MS analysis to be automatically completed on very low abundance peaks. Identification of the peaks of interest in each MS spectra was obtained using a program called MSLinks. This program allows the user to input the specific crosslinker being used as well as the sites of interaction on the peptides. It then compiles a list of the potential crosslinked peaks present in each data set. The user then has the ability to analyze the entire list or pull out the actual crosslinked peaks (as opposed to the dead ends) and determine if the MS/MS spectra corresponds to any of the potential matches.

Analysis of the spectra dataset for actin-34 kDa protein EDC/sulfo-NHS crosslinked peptides yielded the identification of three potential actin-34kDa protein interpeptide complexes with monoisotopic masses $[M+H]^+$ of 2278.0818 amu (Figure 4.4a), 3212.7142 amu (Fig 4.5a), and 2205.0376 amu (Figure 4.6a). Utilizing the information obtained from MSLinks and the MS/MS spectrums, it was determined that each of these peaks identified corresponded to a link between 34 kDa protein and Actin. These peaks correspond to two distinct sites on the 34 kDa – Actin model. Two of the peaks lie within the previously described 34 kDa protein actin binding regions; Glu210 and Glu285. Glu210 occurs within the strong actin binding region and Glu285 is in the C-terminal weak actin-binding region. The second site, Lys184, does not map to any of the three known actin-binding regions, however it does lie close to the strong actin-binding site at residues 193-254. This anomaly could arise due to the fact that the Lys184 strong actin-binding

site was dependent on the proper three-dimensional structural constraints and therefore was not identified in previous studies, due to the fact that they were conducted on truncated forms of the 34 kDa protein.

Analysis of the second crosslinker, BS^3, yielded two potential actin-34 kDa protein interpeptide crosslinks with monoisotipic masses $[M+H]^+$ of 3398.8893 amu (Figure 4.7a) and 3376.8861 amu (Figure 4.8a). The triply charged precursor ion, 1133.6312, was identified using MS/MS as Lys213 on actin linked to Lys17 on 34 kDa protein (Figure 4.7b). The triply charged precursor ion, 1126.2901 however did not result in a viable MS/MS fragmentation spectra, and thus positive identification was not possible. It is however identified as the Lys326 on actin (Figure 4.8b). From the MSLinks data it was determined that both Lys286 and Lys287 could be possible sites of crosslinking. Therefore, although the exact amino acid involved in the crosslink was not determined, the general location of interaction was determined.

Lastly, the mass defect labeled reagent, DiBMADPS, was analyzed by HPLC-MS/MS and MSLinks. Identification of potential actin-34 kDa protein interpeptide crosslinks from DiBMADPS tryptic digests required manual alteration to the search parameters of the MSLinks software. This was completed very easily by simply imputing the mass of the crosslinker as 202.0868 amu. Using this modification to the program, we were able to identify three potential

actin – 34 kDa protein interpeptide crosslinks of monoisotopic masses $[M+H]^+$ of 2554.1354 amu (Figure 4.9a), 2037.8988 amu (Figure 4.10a), and 3488.7271 amu (Figure 4.11a).

Considering the chemistry and length of the commercially available BS^3 crosslinker and the mass defect labeled reagent are very similar the crosslinks identified were also very similar. Due to the fact that these two crosslinkers possess a total length of approximately 12 Å, the actual amino acids involved in the interaction sites were difficult to determine, however they do shed light on the structural organization of 34 kDa protein when it is bound to F-actin. For example, Lys17 crosslinked to actin was found in both BS^3 and DiBMADPS samples (Table 4.2). The simplest interpretation is therefore that the N-terminal portion of 34 kDa protein is in close proximity to F-actin, and this supports the observation that the N-terminus of 34 kDa protein is important in F-actin binding [7].

Additionally, both BS^3 and DiBMADPS illuminated a the crosslink between the C-terminal portion of 34 kDa protein to actin (Lys286/7 and Lys288 respectively). This result is also consistent with previous studies and provides additional support for the EDC/sulfo-NHS crosslinking data which places the C-terminal of 34 kDa protein at a 34 kDa protein – F-actin binding site. The main difference in these two crosslinkers however is that unlike BS^3, the mass defect labeled reagent crosslinked a unique 34 kDa peptide fragment that was crosslinked to actin at Lys204 and this is consistent with the EDC/sulfo-NHS results which place the 34 kDa

protein strong actin-binding site at the 34 kDa protein-F-actin-binding interface. Table 4.2 summarizes the results of all three crosslinkers described above.

34 kDa protein – F-actin binding interface. Following the extensive HPLC-MS/MS analysis of all potential crosslinked fragment ions, the positively identified crosslinks were highlighted in the context of the actin filament structure. This resulted in a patch on one of the faces of F-actin formed by a cluster of three residues that could potentially be a 34 kDa protein binding site. This area is comprised of Lys213 and Lys61 of the lower subunit and Lys326 of the upper subunit, and significantly spans two successive actin subunits along one strand of the filament (Figure 4.12). It has been previously hypothesized that the 34 kDa protein binding site is comprised of two actin monomers and this is what determines the F-actin binding specificity [8]. This 34 kDa protein binding zone is also proximal to well characterized ligand-binding "hot spots" on actin, referred to as the "hydrophobic cleft" and "hydrophobic pocket" [9]. In addition to this 34 kDa protein binding zone, there appears to be a second 34 kDa protein F-actin binding site that is located proximal to the junction between subdomains 3 and 4 in G-actin involving residues Asp179 and Lys191, respectively.

After close examination of these specific locations within the F-actin model structure, it was determined that this second binding site could enable 34 kDa protein to simultaneously contact three adjacent actin subunits at one time (Figure 4.10). Ironically, both Asp179 and

Lys191 have been implicated in molecular interaction studies that determined that they are important for filament stability [10]. Due to the fact that these residues were identified using EDC/sulfo-NHS which is a zero length crosslinker and can only span very short distances, it is reasonable to believe that they are both in direct contact with 34 kDa protein even though the actual amino acids were not determined.

One explanation for this is that the bundling that occurs when 34 kDa protein is introduced into the system induces a conformation change in the filament structure that exposes these residues for 34 kDa protein binding. The F-actin binding protein scruin and cofilin have both been shown to alter F-actin structure is just this manner [11]. Another possible explanation for this phenomenon is that 34 kDa protein is able to bind G-actin which has both Asp179 and Lys191 exposed for crosslinking. However, all of the experimental data to date on 34 kDa protein has demonstrated that it exclusively binds F-actin. Lastly, it is possible that 34 kDa protein may remain "attached" to a newly dissociated actin monomer for enough time to become crosslinked.

The results obtained from the crosslinking and mass spectrometry experiments have identified two putative 34 kDa protein binding sites on F-actin and structural analysis of these sites suggests a possible mechanism by which 34 kDa protein may affect the rate of actin subunit dissociation from an actin filament. Both of the 34 kDa protein F-actin binding sites identified

are located at regions on F-actin that enable 34 kDa protein to interact with more than one actin subunit at one time. Therefore it can be hypothesized that the simultaneous binding of successive actin subunits within the filament by 34 kDa protein may prevent or slow actin subunit dissociation from filament ends and lead to increased filament stability.

These studies have also revealed key residues on both 34 kDa protein and actin that are involved at the 34 kDa – F-actin binding interface. Furthermore, combining previous biochemical data on 34 kDa protein and 34 kDa protein truncations with these studies reveals at least three actin-binding regions. What remains unanswered however, is the mode in which 34 kDa protein uses these multiple actin binding sites to crosslink and bundle actin. F-actin crosslinking proteins must contain at least two distinct regions of binding in order to crosslink two actin filaments [12]. Some proteins do however crosslink F-actin as a monomer and contain at least two separate F-actin-binding sites along the length of their polypeptide chain, while other crosslinking proteins contain only a single F-actin binding site and therefore require dimerization/oligomerization to be able to crosslink actin filaments.

Conclusions. Utilizing the *Dictyostelium* 34 kDa protein system, distinct regions of interaction between 34 kDa and actin were identified using HPLC-MS/MS, multiple commercially available chemical crosslinkers, a mass defect labeled reagent, and sophisticated data analysis software. With the combination of these three chemical crosslinkers, possible

explanations for actin bundling by 34 kDa protein were revealed. Questions still exist as to the actual method by which 34 kDa protein crosslinks actin (either as a monomer or dimer), however the structural information obtained in this study is a huge leap in the positive direction of understanding actin bundling by this particular crosslinking protein.

Table 4.1. Important Diseases associated with the formation of Hirano Bodies. Demonstrates the medicinal use of discovering the key to the formation of Hirano bodies and the bundling of actin fibers by specific proteins such as 34 kDa protein.

Group	Examples	Functional Role	References
Monomer-binding	Profilin	Monomer sequestering, nucleotide exchange	(Schlüter 1997; Paavilainen 2004)
	Thymosin-β4	Monomer sequestering	(Safer 1994; Pollard 2003)
	Srv2/CAP	Recycles ADF/cofilin from G-actin after depolymerization	(Balcer 2003; Paavilainen 2004)
Filament-depolymerizing	ADF/cofilin	Depolymerization of actin filaments/filament turnover	(Bamburg 1999; Ono 2007)
Filament-severing	Gelsolin/ Severin	Severs actin filaments, increases # of (+) ends for polymerization	(Silacci 2004; Ono 2007)
Filament end-binding	CapZ	Binds and caps (+) end of thin filament in muscle	(Weeds 1993; Cooper 2000)
	Tropomodulin	Binds and caps (-) end of thin filament	(Fischer 2003; Kostyukova 2008)
	Capping Protein (CP)	Major (-) end capping-protein in non-muscle cells Major (+) end capping-protein in non-muscle cells	(Wear 2004; Cooper 2008)
De novo filament-nucleation	Arp2/3 complex	Filament side-branching nucleation, crosslinker	(Mullins 1999; Pollard 2003)
	Formin	Nucleation of unbranched, long actin filaments	(Goode 2007)
Motors	Myosin Superfamily	**Currently 23 Classes**	(Foth 2006)
	Myosin I	Single-headed myosin, links filaments to membrane, endocytosis	(Kalhammer 2000)
	Myosin II	Force generation in muscle; non-muscle cells contractile functions	(Sellers 2000)
	Myosin V	(+) end-directed movement of organelles	(Kalhammer 2000)
	Myosin VI	(-) end-directed movement of endocytic vesicles	(Kalhammer 2000; Sellers 2000)
Filament-stabilizing	Tropomyosin	Stabilization/regulates myosin interaction with thin filament	(Gunning 2005; Kostyukova 2008)
Crosslinking/bundling			(Furukawa 1997)
	Fimbrin	Bundles with same polarity/microvilli, filopodia	(Hartwig 1995)
	α-actinin	Crosslinks and bundles/stress fibers, filopodia, muscle Z-line	(Broderick 2005; Sjöblom 2008)
	Spectrin	Links filament networks to plasma membrane	(Hartwig 1995; Broderick 2005)
	Dystrophin	Links cortical actin networks of muscle cells to plasma membrane	(Broderick 2005)
	Filamin	Highly flexible crosslinker/meshworks/leading edge/stress fibers	(Popowicz 2006)
	ABP-34 (34 kDa protein)	Bundles filaments/filopodia/leading edge	(Fechheimer 1987; Fechheimer 1994)

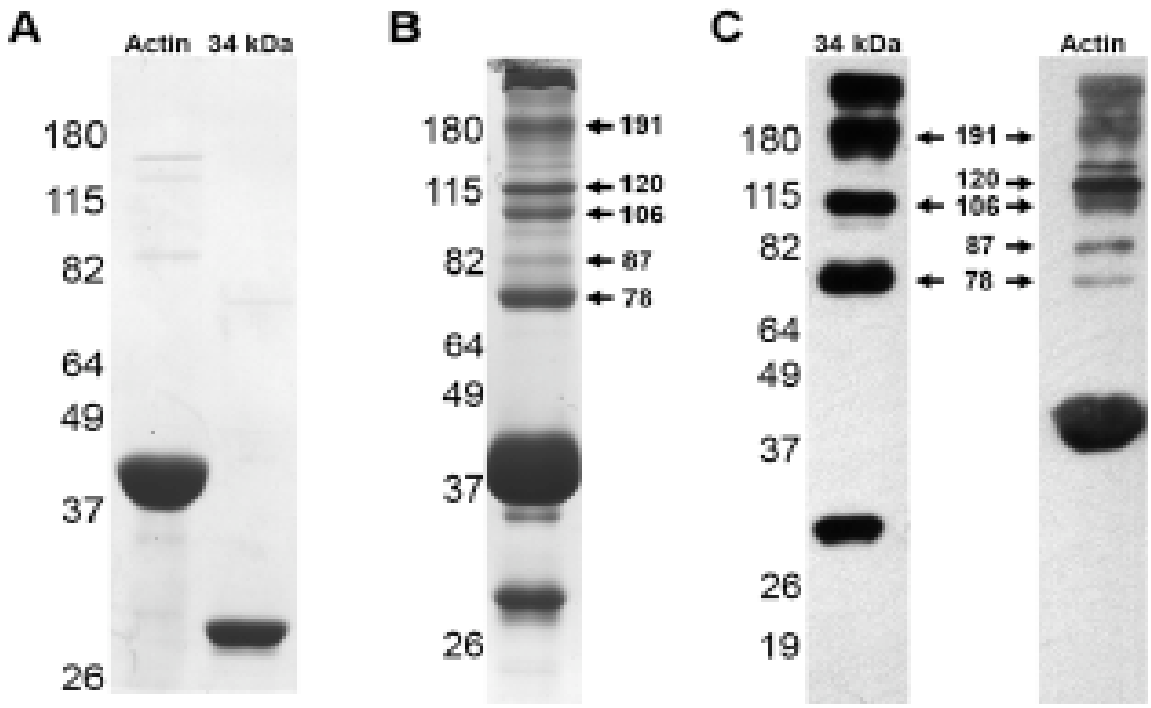


Figure 4.1. EDC/sulfo-NHS reaction products for 34 kDa protein crosslinked with F-actin. (A) SDS-PAGE analysis of control sample loads for actin and 34 kDa protein. (B) SDS-PAGE and (C) Western blot analysis of resulting products generated by EDC/sulfo-NHS chemical crosslinking of 34 kDa protein co-polymerized with F-actin. In (B) the Coomassie-resolvable crosslinked species are indicated with an arrow and corresponding estimated molecular weight (kDa). For (C) the 34 kDa and actin immunoreactivity of the major crosslinked species are indicated by the directed arrow.

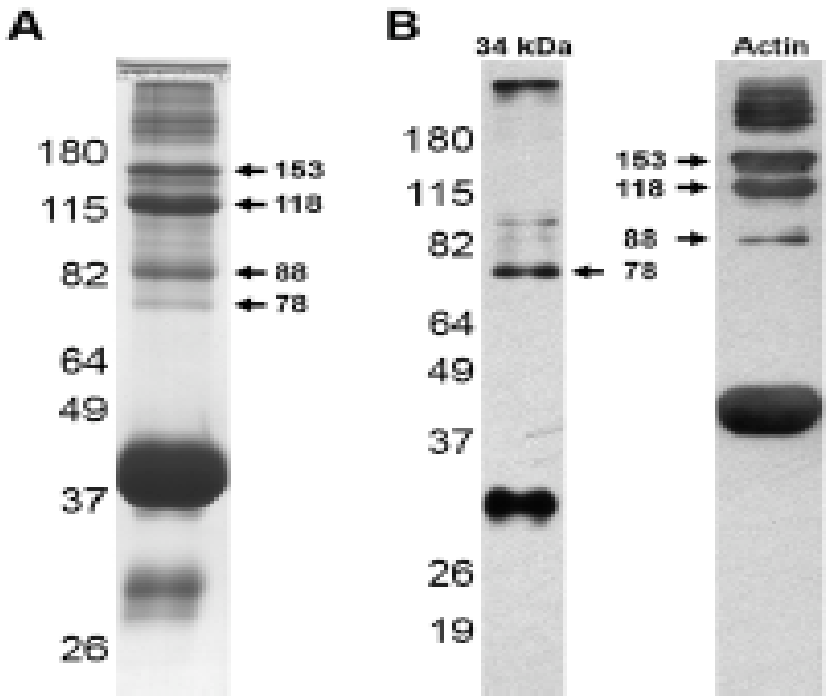


Figure 4.2. BS^3 reaction products for 34 kDa protein crosslinked with F-actin. (A) SDS-PAGE and (B) Western blot analysis of resulting products generated by BS^3 chemical crosslinking of 34 kDa protein co-polymerized with F-actin. In (A) the Coomassie-resolvable crosslinked species are indicated with an arrow and corresponding estimated molecular weight (kDa). For (B) the 34 kDa and actin immunoreactivity of the major crosslinked species are indicated by a directed arrow.

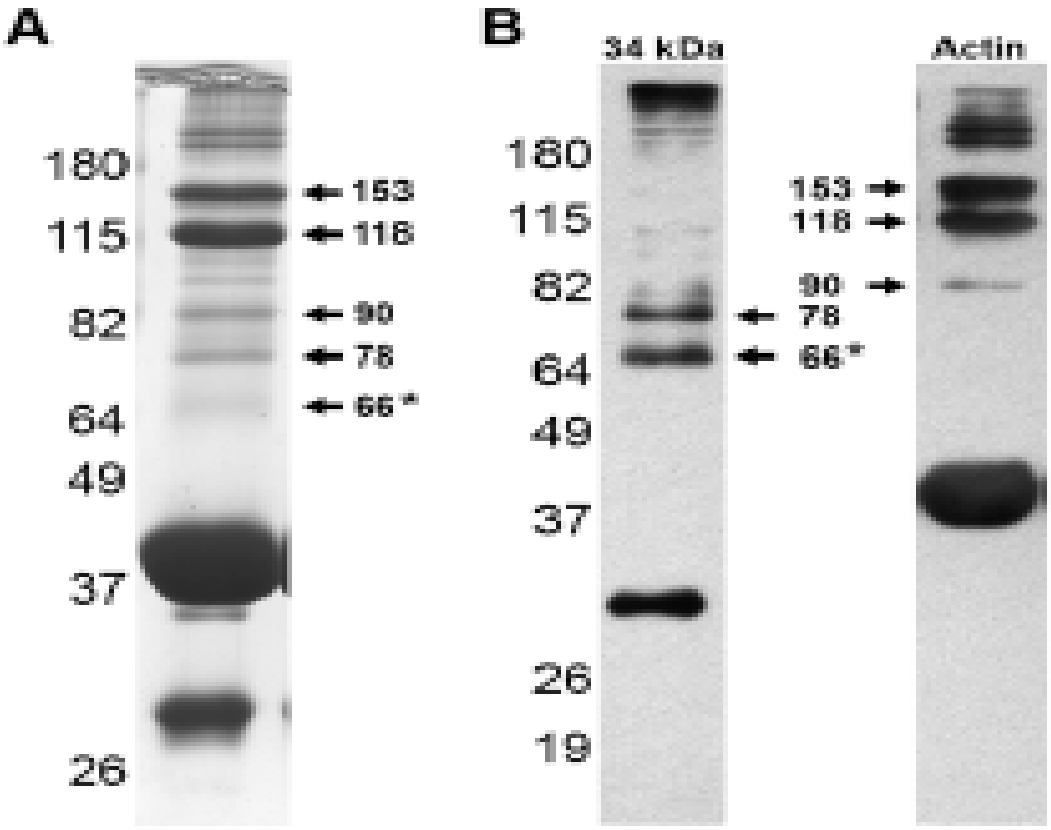


Figure 4.3. DiBMADPS reaction products for 34 kDa protein crosslinked with F-actin. (A) SDS-PAGE and (B) Western blot analysis of resulting products generated by DiBMADPS chemical crosslinking of 34 kDa protein co-polymerized with F-actin. In (A) the Coomassie-resolvable crosslinked species are indicated with an arrow and corresponding estimated molecular weight (kDa). For (B) the 34 kDa and actin immunoreactivity of the major crosslinked species are indicated by a directed arrow. In both (A) and (B) the unique 66 kDa band is highlighted with an asterisk.

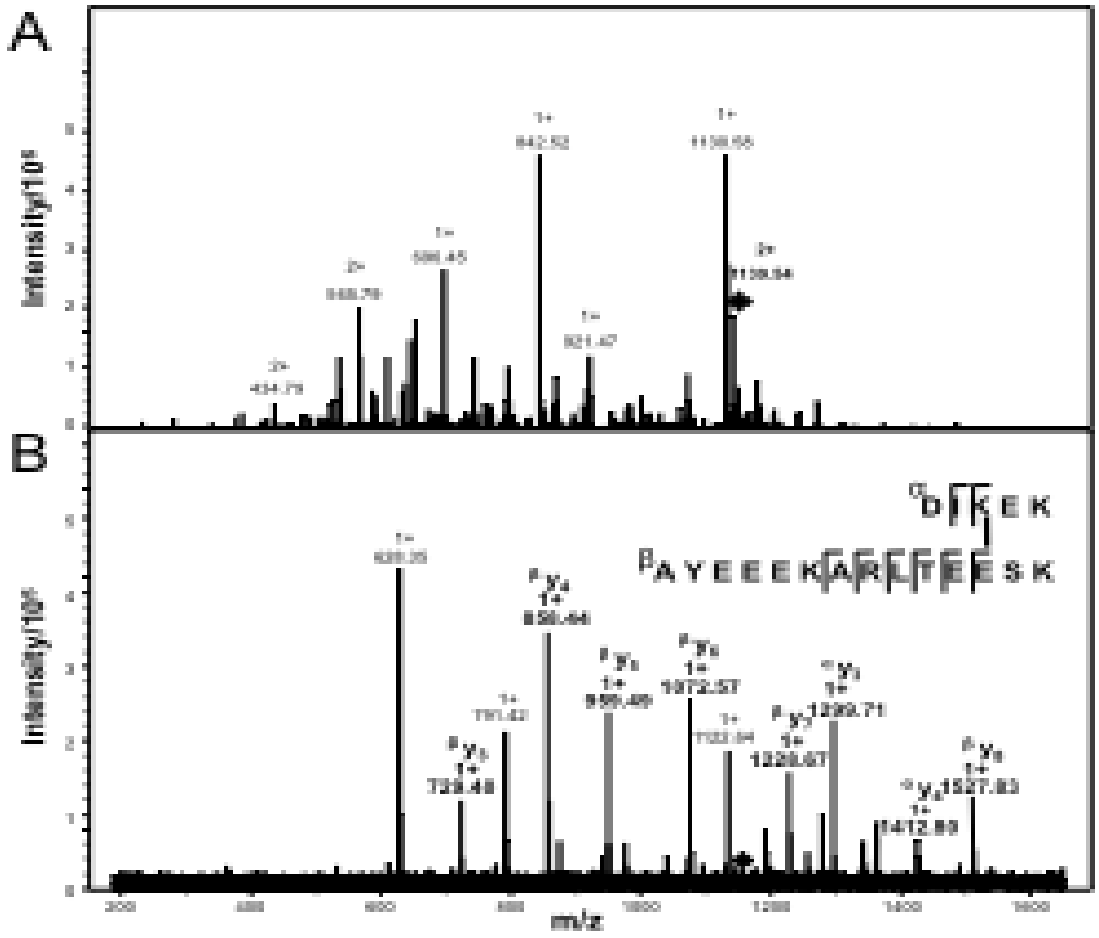


Figure 4.4. Mass spectra set for EDC/Sulfo-NHS monoisotopic mass at 2278.0818. (A) ESI mass spectrum for EDC/sulfo-NHS crosslinked interpeptide species of mass 2278.0818. (B) MS/MS spectrum of precursor ion at m/z 1139.54^{2+}, which is labeled with a diamond. Peptide (α) represents the peptide fragment arising from F-actin and peptide (β) represents the peptide fragment arising from 34 kDa protein. The sequence of each peptide as well as the identified b and y ions are shown, corresponding to the m/z indicated in the spectra. The results shown that Lys213 of actin is crosslinked to Glu210 of 34 kDa protein.

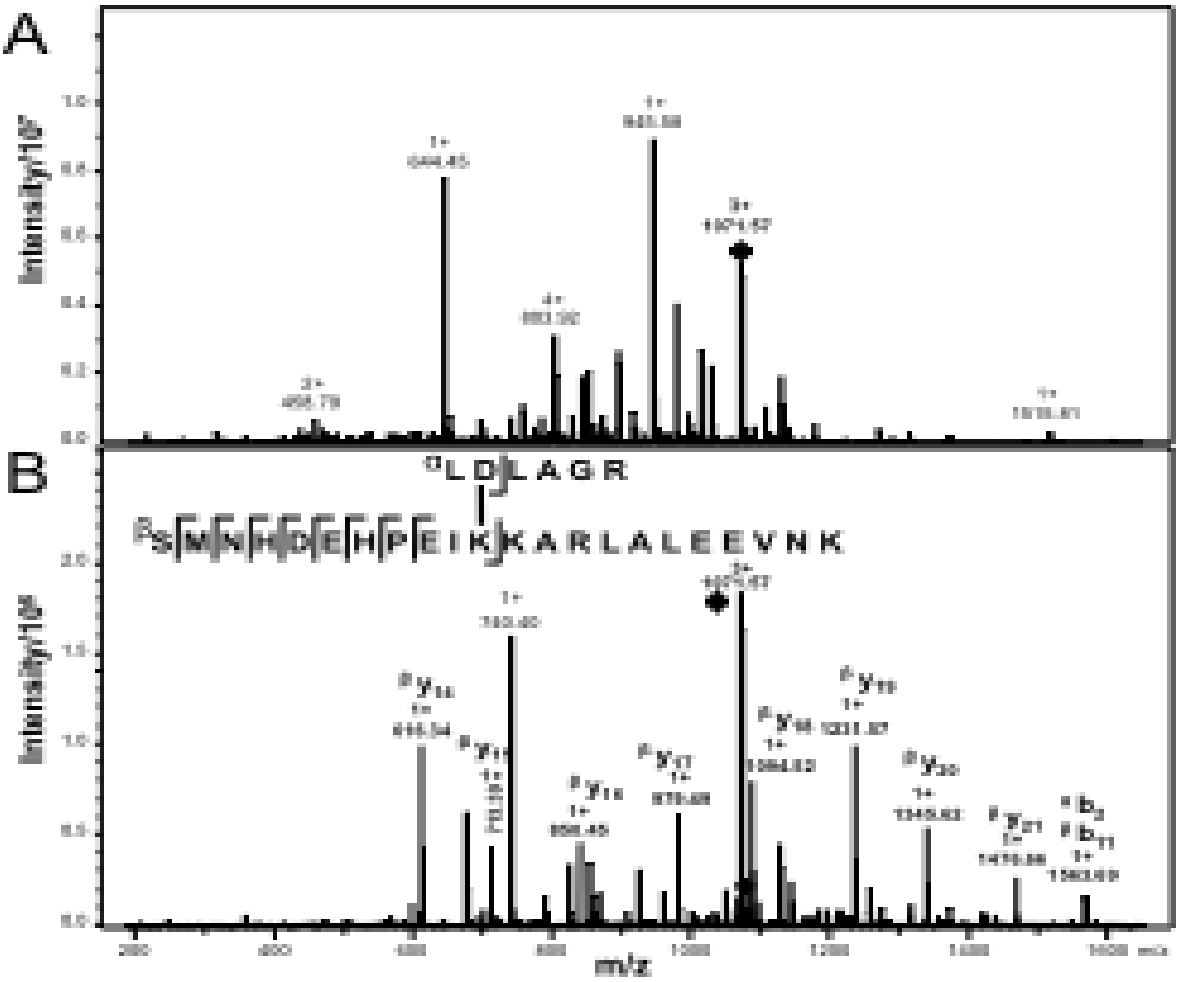


Figure 4.5. Mass spectra set for EDC/Sulfo-NHS monoisotopic mass at 3212.7142. (A) ESI mass spectrum for EDC/sulfo-NHS crosslinked interpeptide species of monoisotopic mass 3212.7142. (B) MS/MS spectrum of precursor ion at m/z 1071.57^{3+}, which is labeled with a diamond. Peptide (α) represents the peptide fragment arising from F-actin and peptide (β) represents the peptide fragment arising from 34 kDa protein. The sequence of each peptide as well as the identified b and y ions are shown, corresponding to the m/z's indicated in the spectra. The results illustrate that Asp179 of actin is crosslinked to Lys184 or 34 kDa protein

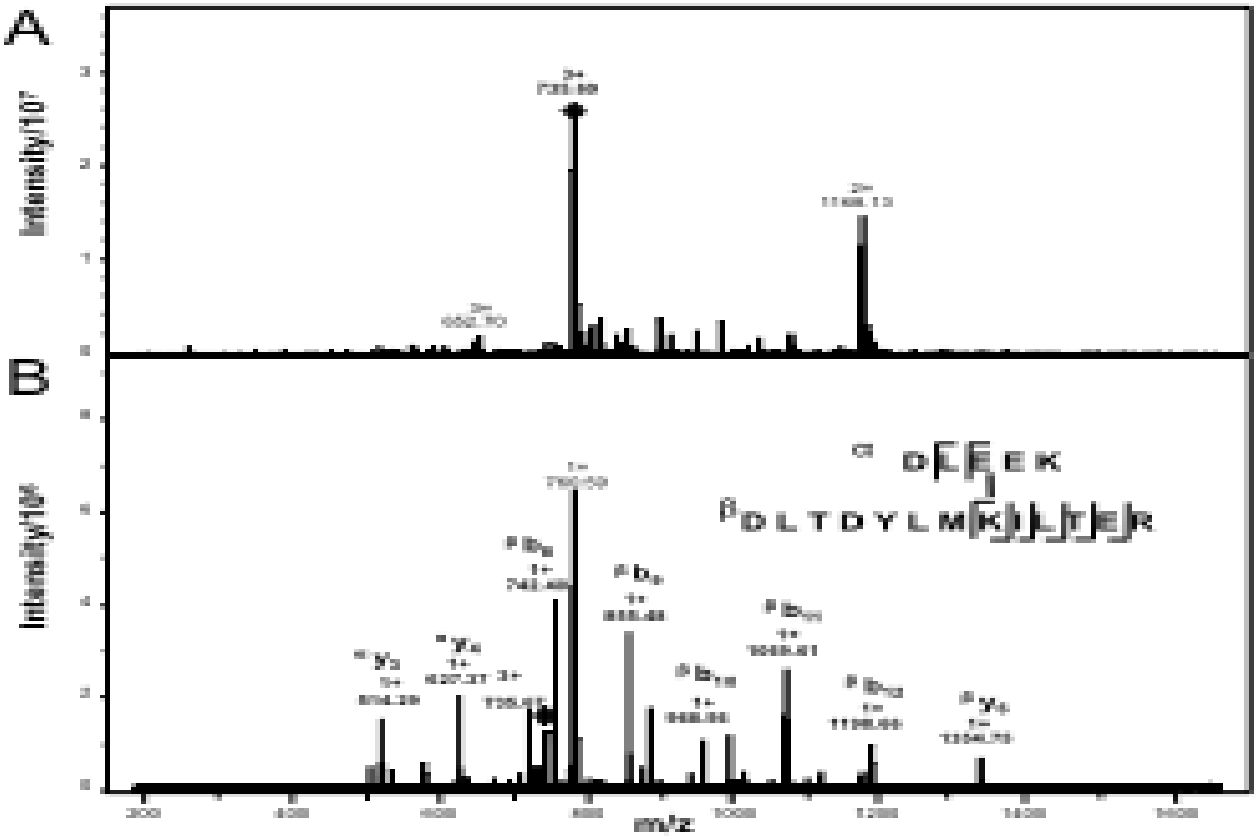


Figure 4.6. Mass spectra set for EDC/Sulfo-NHS monoisotopic mass at 2205.0376. (A) ESI mass spectrum for EDC/sulfo-NHS crosslinked interpeptide species of monoisotopic mass 2205.0376. (B) MS/MS spectra of precursor ions at m/z 735.69^{3+}, which is labeled with a diamond. Peptide (α) represents the peptide fragment arising from 34 kDa protein and (β) represents the peptide fragment arising from F-actin. The sequence of each peptide as well as the identified b and y ions are shown, corresponding to the masses indicated in the spectra. The results show that Lys191 of actin is crosslinked to Glu285 of 34 kDa protein.

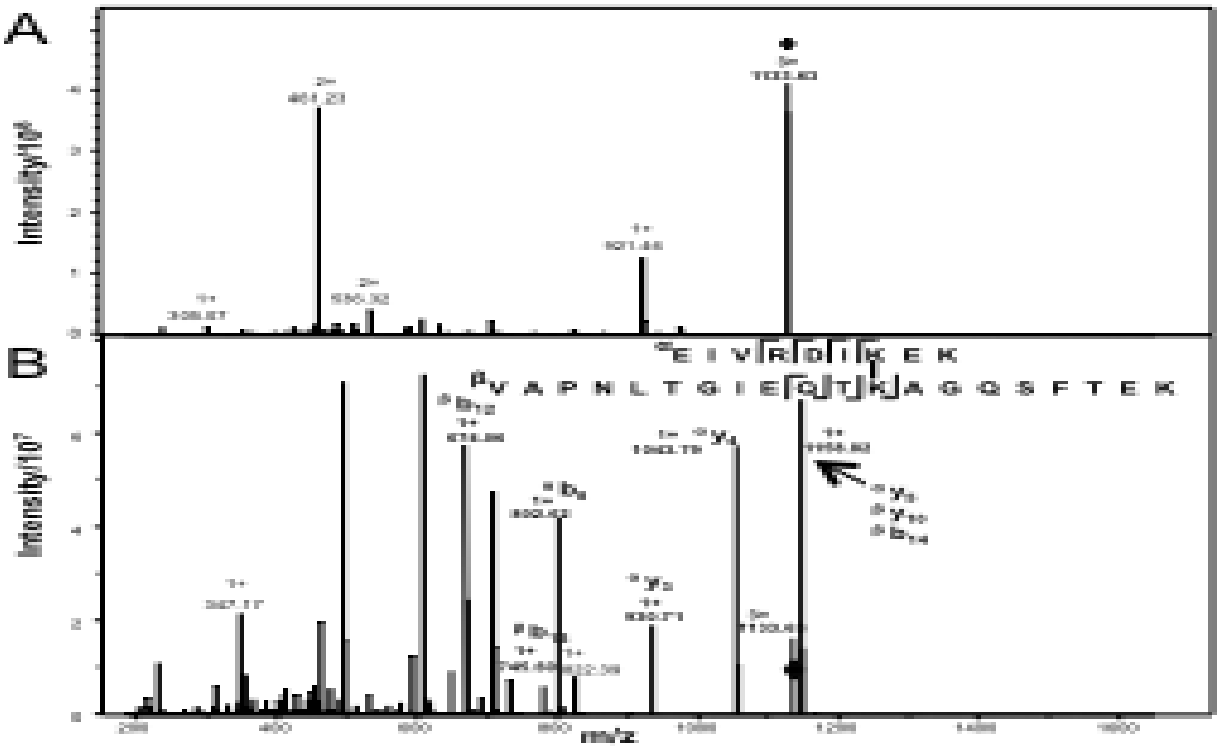


Figure 4.7. Mass spectra set for BS^3 monoisotopic mass at 3398.8893. (A) ESI mass spectrum for BS^3 crosslinked interpeptide species of monoisotopic mass 3398.8893. (B) MS/MS spectra of ions at m/z 1133.63^{3+}, which is labeled with a diamond. Peptide (α) represents the peptide fragment arising from F-actin and peptide (β) represents the peptide fragment arising from 34 kDa protein. The sequence of each peptide and the identified b and y ions are shown, corresponding to the masses indicated in the spectra. The results show that Lys214 of actin is crosslinked to Lys17 of 34 kDa protein.

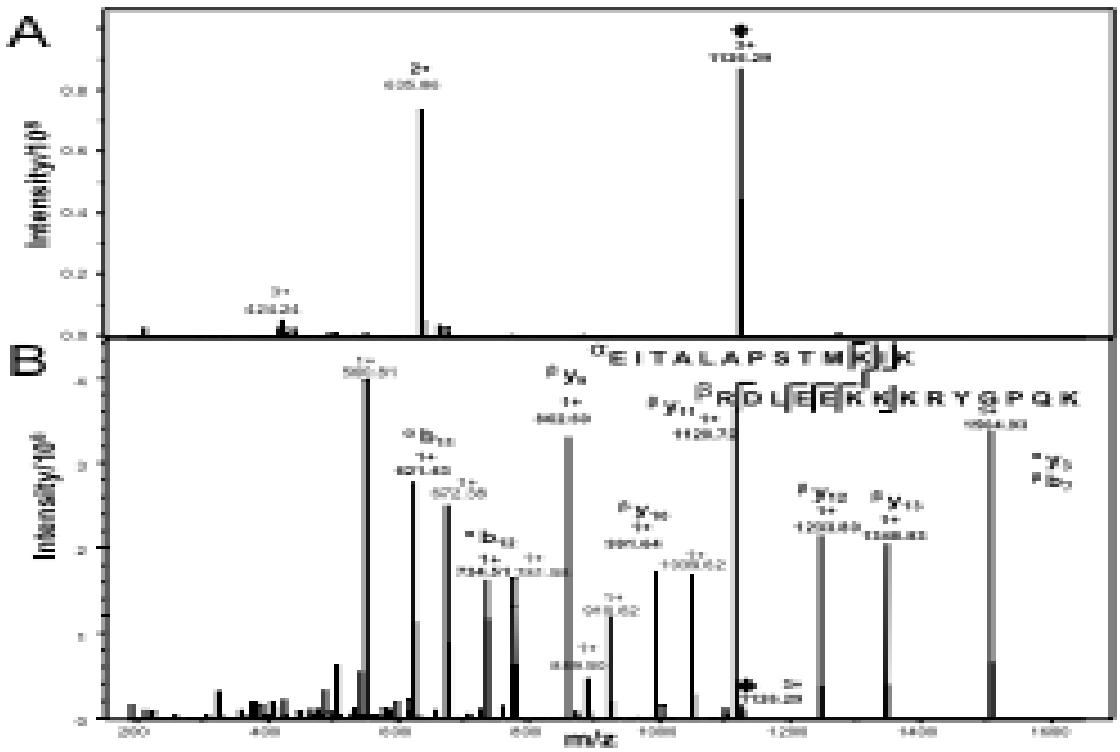
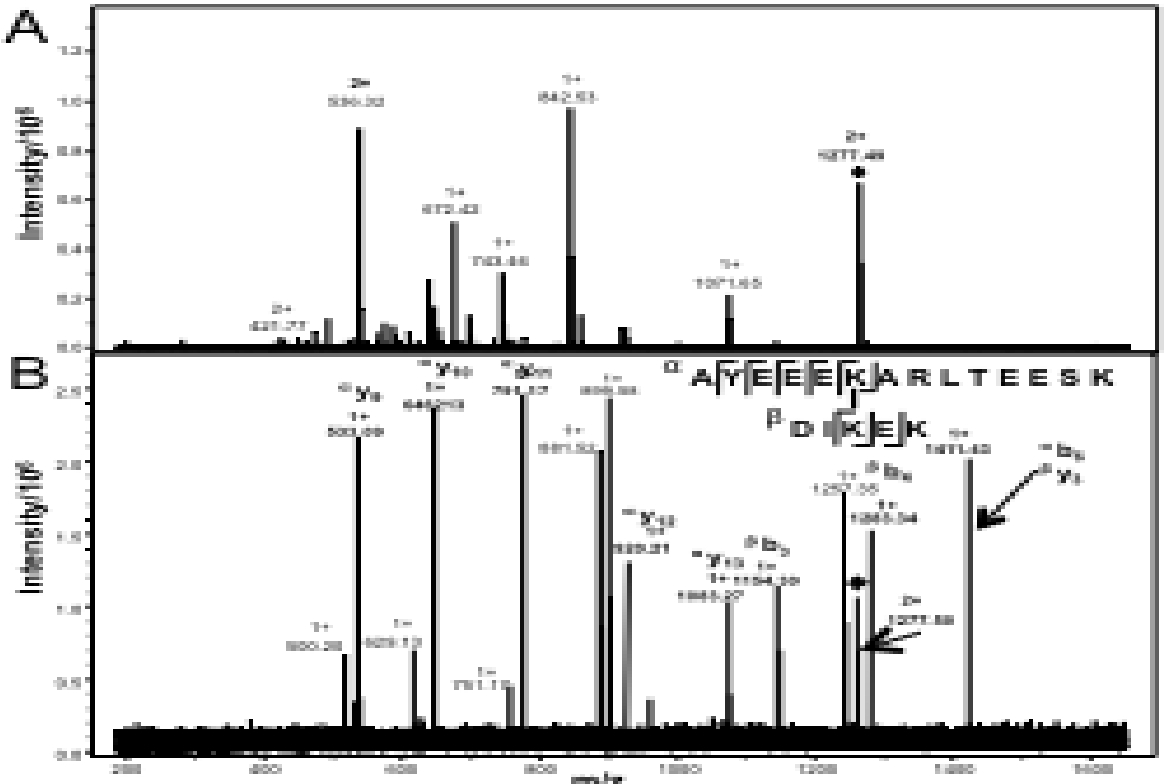


Figure 4.8. Mass spectra set for BS^3 monoisotopic mass 3376.8661. (A) ESI mass spectra for DiBMADPS crosslinked interpeptide species of monoisotopic mass 2554.1354. (B) MS/MS spectra of ions at m/z 1277.49^{2+}, which is labeled with a diamond. Peptide (α) represents the peptide fragment arising from 34 kDa protein and peptide (β) represents the peptide fragment arising from actin. The sequence of each peptide and the identified b and y ions are shown corresponding to the masses indicated in the spectra. The results show that Lys213 of actin is crosslinked to Lys204 of 34 kDa protein.



4.9. Mass spectra set for DiBMADPS monoisotopic mass 2554.1354. (A) ESI mass spectra for DiBMADPS crosslinked interpeptide species of monoisotopic mass 2554.1354. (B) MS/MS spectra of ions at m/z 1277.49^{2+}, which is labeled with a diamond. Peptide (α) represents the peptide fragment arising from 34 kDa protein and peptide (β) represents the peptide fragment arising from actin. The sequence of each peptide and the identified b and y ions are shown corresponding to the masses indicated in the spectra. The results show that Lys213 of actin is crosslinked to Lys204 of 34 kDa protein.

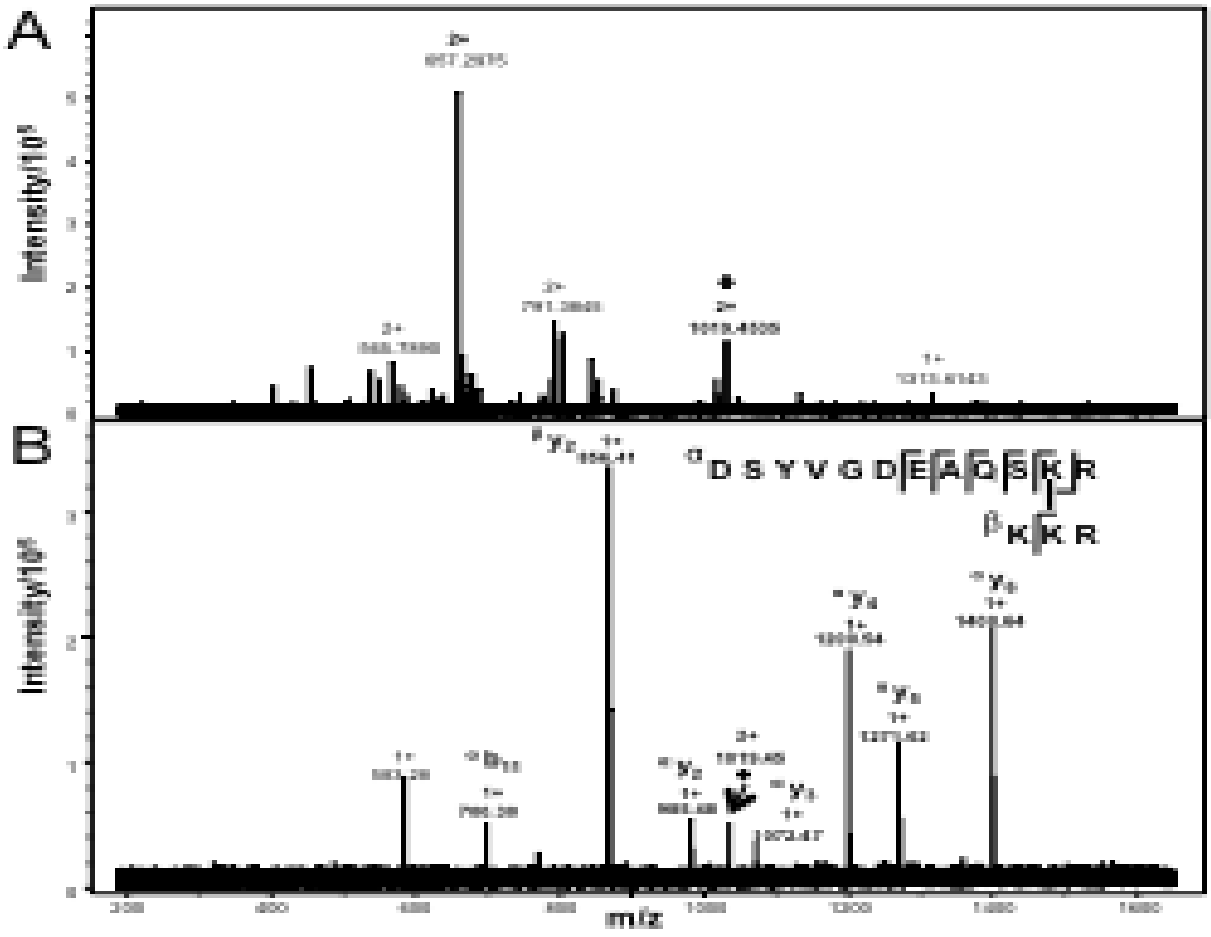


Figure 4.10. Mass spectra set for DiBMADPS monoisotopic mass 2037.8988. (A) ESI mass spectra for DiBMADPS crosslinked interpeptide species of monoisotopic mass 2037.8988. (B) MS/MS spectra of ions at m/z 1019.45^{2+}, which is labeled with a diamond. Peptide (α) represents the peptide fragment arising from actin and peptide (β) represents the peptide fragment arising from 34 kDa protein. The sequence of each peptide and the identified b and y ions are shown, corresponding to the masses indicated in the spectra. The results show that Lys61 of actin is crosslinked to Lys288 of 34 kDa protein

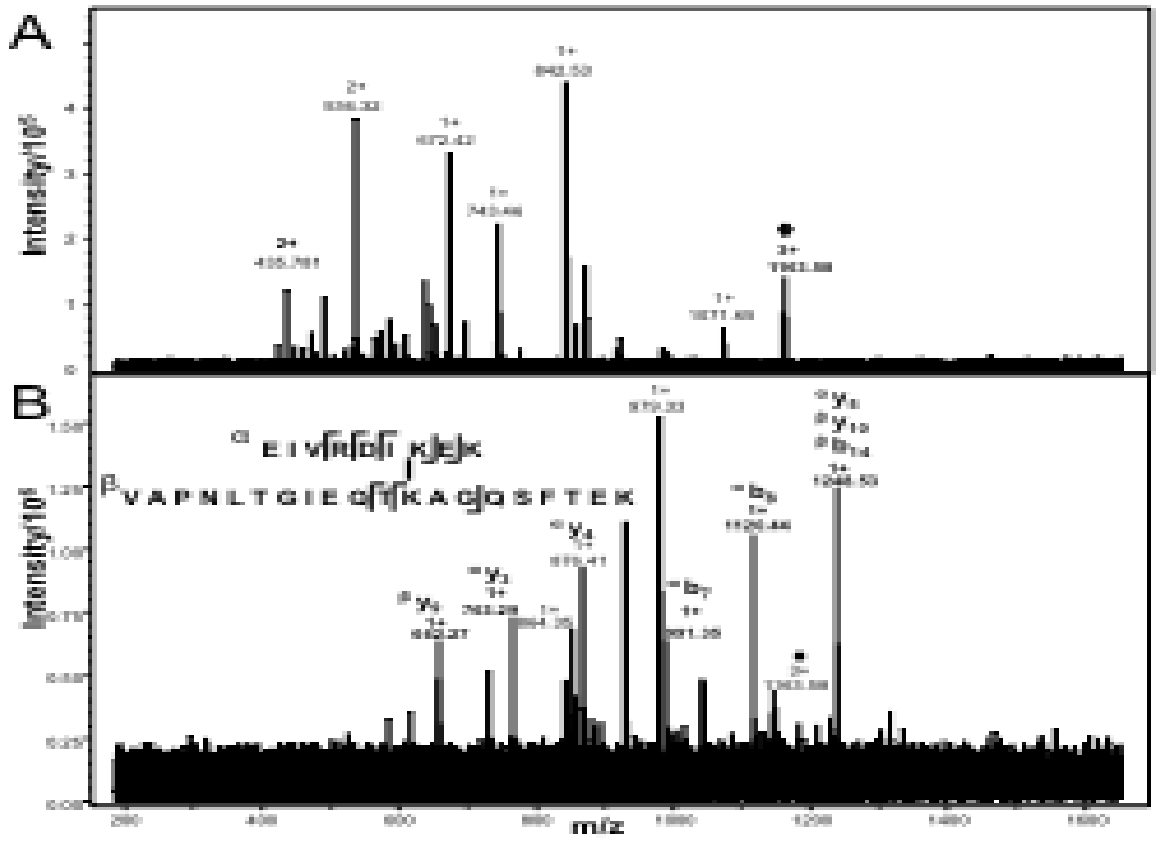


Figure 4.11. Mass spectra set for DiBMADPS monoisotopic mass 3488.7271. (A) ESI mass spectra for DiBMADPS crosslinked interpeptide species of monoisotopic mass 3488.7271. (B) MS/MS spectra of precursor ions of m/z 1163.58^{3+}, which is labeled with a diamond. Peptide (α) represents the peptide fragment arising from actin and peptide (β) represents the peptide fragment arising from 34 kDa protein. The sequence of each peptide and the identified b and y ions are shown, corresponding to the masses indicated in the spectra. The results show that Lys213 of actin is crosslinked to Lys17 of 34 kDa protein.

Table 4.2. Summary of crosslinked species and crosslinked residues as identified by chemical crosslinking studies.

$[M+H]^+$ observed	$[M+H]^+$ theoretical	Error (ppm)	Interpeptide species		Crosslinked residues		Crosslinker
			Actin	34 kDa	Actin	34 kDa	
2278.0818	2278.2394	0.1516	211-215	199-212	Lys213	Glu210	EDC/NHS
3212.7142	3212.6724	- 0.0418	179-183	174-195	Asp179	Lys184	EDC/NHS
2205.0376	2205.1113	0.0737	184-196	282-286	Lys191	Glu285	EDC/NHS
3398.8893	3398.9936	0.1043	207-215	6-25	Lys213	Lys17	BS[3]
3376.8861	3376.8912	0.0015	316-328	281-294	Lys326	Lys286/7	BS[3]
2554.1354	2554.1645	0.0286	211-215	199-212	Lys213	Lys204	DiBMADPS
2037.8988	2037.8993	0.0005	51-62	287-289	Lys61	Lys288	DiBMADPS
3488.7271	3488.7640	- 0.0631	207-215	6-25	Lys213	Lys17	DiBMADPS

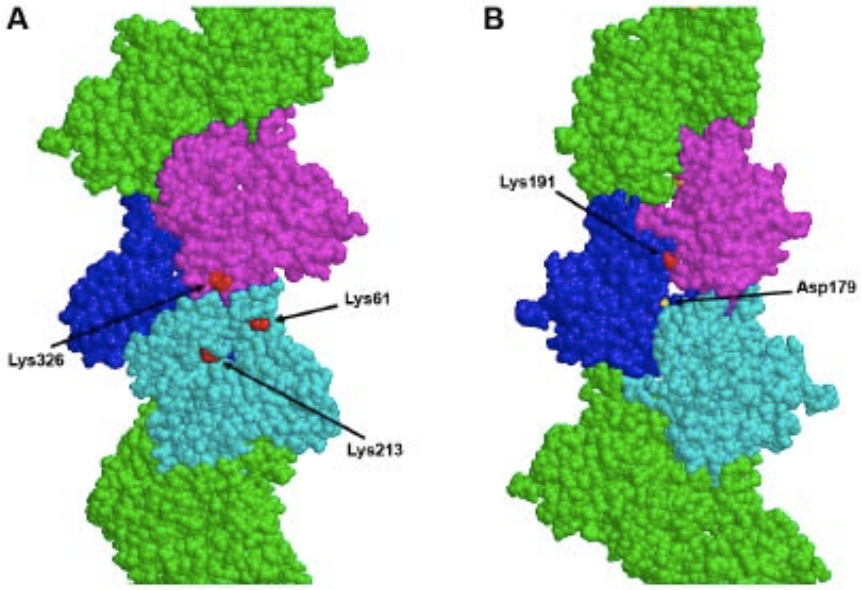


Figure 4.12. Space-filling F-actin models highlightin putative 34 kDa binding sites. The filament models are orientated with the pointed end on the top and the barbed end on the bottom. Lysine residues are colored red and aspartate is colored yellow. Three actin subunits within the actin filament are colored magenta, blue and cyan. (A) The 34 kDa binding patch on F-actin is comprised of Lys326, Lys61 and Lys213 and spans two successive actin subunits. (B) The second 34 kDa protein binding site is comprised of Lys181 and Asp179 on the same actin subunit (cyan). Note that Lys191 and Asp179 are located at sites of contact with different actin subunits. Thus 34 kDa protein binding at this location on F-actin likely involves simultaneous contact on all three actin subunits.

References:

1. Dos Remedios, C. G.; Chhabra, D.; Kekic, M.; Dedova, I. V.; Tsubakihara, M.; Berry, D. A.; Nosworthy, N. J. *Physiological Reviews* **2003**, *83*, 433.

2. Furukawa, R.; Fechheimer, M. In *International Review of Cytology - a Survey of Cell Biology, Vol 175*; Academic Press Inc: San Diego, 1997; Vol. 175, p 29.

3. Puius, Y. A.; Mahoney, N. M.; Almo, S. C. *Curr. Opin. Cell Biol.* **1998**, *10*, 23.

4. Gimona, M.; Djinovic-Carugo, K.; Kranewitter, W. J.; Winder, S. J. *FEBS Lett.* **2002**, *513*, 98.

5. Hartwig, J. H. *Protein Profile* **1995**, *2*, 703.

6. Broderick, M. J., and Winder, S. J. *Adv. Protein Chem.* **2005**, 203.

7. Sjoblom, B.; Salmazo, A.; Djinovic-Carugo, K. *Cell. Mol. Life Sci.* **2008**, *65*, 2688.

8. Fechheimer, M.; Taylor, D. L. *J. Biol. Chem.* **1984**, *259*, 4514.

9. Fechheimer, M. *J. Cell Biol.* **1987**, *104*, 1539.

10. Fechheimer, M.; Ingalls, H. M.; Furukawa, R.; Luna, E. J. *J. Cell Sci.* **1994**, *107*, 2393.

11. Furukawa, R.; Kundra, R.; Fechheimer, M. *Mol. Biol. Cell* **1992**, *3*, A38.

12. Okazaki, K.; Yumura, S. *Eur. J. Cell Biol.* **1995**, *66*, 75.

13. Lim, R. W. L.; Furukawa, R.; Eagle, S.; Cartwright, R. C.; Fechheimer, M. *Biochemistry* **1999**, *38*, 800.

14. Hoffman, L.; Griffin, P. A.; Furukawa, R.; Fechheimer, M.; Petrotchenko, E.; Borchers, C.; Amster, I. J. *To Be Submitted* **2010**.

15. Lim, R. W. L.; Fechheimer, M. *Protein Expression Purif.* **1997**, *9*, 182.

16. Griffin, P. A.; Kim, D. H.; Furukawa, R.; Fechheimer, M. *To be submitted* **2010**.

17. Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C. *Anal. Biochem.* **1985**, *150*, 76.

18. Spudich, J. A.; Watt, S. *J. Biol. Chem.* **1971**, *246*, 4866.

19. Macleanfletcher, S.; Pollard, T. D. *Biochem. Biophys. Res. Commun.* **1980**, *96*, 18.

20. Laemmli, U. K. *Nature* **1970**, *227*, 680.

21. Towbin, H.; Staehelin, T.; Gordon, J. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 4350.

22. Kandasamy, M. K.; McKinney, E. C.; Meagher, R. B. *Plant J.* **1999**, *18*, 681.

23. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850.

24. Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5802.

25. Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 834.

26. Kellersberger, K. A.; Yu, E.; Kruppa, G. H.; Young, M. M.; Fabris, D. *Anal. Chem.* **2004**, *76*, 2438.

27. Yu, E. T.; Hawkins, A.; Kuntz, I. D.; Rahn, L. A.; Rothfusa, A.; Sale, K.; Young, M. M.; Yang, C. L.; Pancerella, C. M.; Fabris, D. *J. Proteome Res* **1998**, *7*, 4848.

28. Hernandez, H.; Niehauser, S.; Boltz, S. A.; Gawandi, V.; Phillips, R. S.; Amster, I. J. *Anal. Chem.* **2006**, *78*, 3417.

29. Lim, R. W. L.; Furukawa, R.; Fechheimer, M. *Biochemistry* **1999**, *38*, 16323.

30. McGough, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 166.

31. Dominguez, R. *Trends Biochem. Sci.* **2004**, *29*, 572.

32. Holmes, K. C.; Popp, D.; Gebhard, W.; Kabsch, W. *Nature* **1990**, *347*, 44.

33. Lorenz, M.; Popp, D.; Holmes, K. C. *J. Mol. Biol.* **1993**, *234*, 826.

34. Holmes, K. C.; Angert, I.; Kull, F. J.; Jahn, W.; Schroder, R. R. *Nature* **2003**, *425*, 423.

35. Oda, T.; Iwasa, M.; Aihara, T.; Maeda, Y.; Narita, A. *Nature* **2009**, *457*, 441.

36. M. F. Schmid, J. M. Agris, J. Jakana, P. Matsudaira, W. Chiu, *Journal of Cell Biology* **124**, 341 (Feb, 1994).

37. Schmid, M. F.; Sherman, M. B.; Matsudaira, P.; Chiu, W. *Nature* **2004**, *431*, 104.

38. McGough, A.; Pope, B.; Chiu, W.; Weeds, A. *J. Cell Biol.* **1997**, *138*, 771.

39. Matsudaira, P. *Trends Biochem. Sci.* **1991**, *16*, 87

CHAPTER 5:

MS^3 Confirms the Composition of In Situ Generated 'Internal Elimination' Production Ions that Arise from Synthetic Neoglycolipid Derivatives

In situ 'internal elimination' has recently been proposed to produce a C-glycoside product by fragmentation of synthetic amphiphilic neoglycolipid cholesteryl derivatives using low-energy collisonally activated decomposition (CAD) in tandem mass spectrometry (MS/MS) experiments. Here we present MS^3 data confirming the composition of this novel C-glycoside product ion, using CAD and infrared multiphoton dissociation (IRMPD) to fragment six different synthetic neoglycolipids, ranging in size and composition. These neoglycolipids contain similar cholesterol moieties, however differ in both sugar composition (N-acetyl-D-glycosamine, N-acetyllactosamine, or O-acetylated derivatives of these carbohydrate moieties), and poly(ethoxy) linker lengths. These variables help establish the conditions required for formation of the C-glycoside product ions. In addition to the $[\text{C-glycoside} + H]^+$ product ions, we also identified a number of common fingerprint fragment ions for this family of neoglycolipids.

Introduction. Synthetic neoglycolipids are of great current interest for their ability to stabilize liposomes that are used for delivery of therapeutic agents.[1,2] It has been found that those

liposomes which are stabilized with neoglycolipids exhibit higher gene transfer efficiency than those stabilized using poly(ethylene glycol) (PEG) groups.[3] Recently, published work on synthetic neoglycolipids has produced evidence for the *in situ* formation, both at the electrospray interface and in the collision cell, of a C-glycoside species during low-energy collisionally activated decomposition (CAD) in tandem mass spectrometry (MS/MS) measurements.[4,5] The formation of this product ion has been described to be the result of an 'internal elimination' reaction, forming a C-glycosylation product from a gas-phase ion-molecule reaction. The facile formation of a C-glycoside is surprising, because in the condensed phase, C-glycosylation is much more difficult to achieve synthetically than O- or N-glycosylation. The therapeutic values of C-glycoside analogues have been established as anticancer vaccines,[6] treatments for melanoma metastases,[7] and protection against malaria.[8] The phenomenon of "*in situ*" C-glycosylation can potentially provide insight into new synthetic routes for these compounds.[4]

The formation of the C-glycoside product ion is also of interest because of the rarity of elimination of such large groups in mass spectrometry. A few publications have described the elimination of internal monosaccharide residues from oligosaccharides during MS/MS. McNeil first described a glycosyl-residue elimination that occurred in per-O-alkylated oligosaccharide-alditols during chemical ionization-mass spectrometry in 1983.[9] Kovacik et al. examined a protonated tetrasaccharide under high energy CID conditions and found evidence of the

elimination of the internal galactose residue.[10] It was proposed from this study that a simple mechanism corresponding to a charge-remote rearrangement and elimination of the 1 → 2 substituted galactose was responsible for this 'internal elimination'.[10] However, shortly after this, Brüll et al. demonstrated that internal monosaccharide residue loss was not limited to those ions containing 1 → 2 substituted residues, but also occurred in trisaccharides containing 1 → 6 linked monosaccharide residues.[11] The occurrence of internal monosaccharide residue loss for electrospray generated ions of sialyl-Lewis-type oligosaccharides under low-energy collision induced dissociation (CID) conditions was also reported.[12] Warrack et al. also observed internal monosaccharide losses in their collisionally activated dissociation mass spectra of anthracycline aminodisaccharides, in which product ions resulted from the loss of an internal fucose residue.[13] Ma et al.[14] and Tadano-Aritomi et al.[15] published low energy CID data which confirmed an interesting 'internal elimination' production ion, resulting in the loss of a monosaccharide.

Recently, the elimination of a polyethylene glycol (PEG) oligomer in PEG linked neoglycolipds has been proposed as the source of an unusual product ion that arises during MS/MS experiments[4,5]. In the present study, Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) is used to examine the products of this elimination reaction. FTICR-MS offers extraordinarily high mass resolving power and mass accuracy, as well as tandem mass spectrometry (MS^n) capabilities and thus it has become one of the key tools used for protein and

peptide characterization, accurate mass measurement experiments, as well as molecular composition and overall structure determination and verification for biological macromolecules.[16] Structurally significant fragments can be produced in tandem mass spectrometry experiments using various excitation methods, such as collisionally activated dissociation (CAD),[17] infrared multiphoton dissociation (IRMPD)[16,18-20] and electron aided dissociation (ECD, ETD, EDD).[18,21-23] Here, we use CAD combined with IRMPD, in a MS^3 experiment, to confirm the composition of the *in situ* 'internal elimination' fragment ion that has been observed by ESI of neoglycolipids previously.[4,5]

Experimental. *Synthesis of Neoglycolipids.* The synthesis of neoglycolipids **1**-**4**[24-26] and **5**-**6**[25,27] was performed as reported earlier by El-Aneed *et al.*.[4] In these particular amphiphilic neoglycolipid cholesteryl derivativies, the cholesterol and carbohydrate moieties were attached either directly, as was the case of the simple O-glycoside, or by means of a polyethoxy variable spacer unit. The carbohydrate portion was chosen to be either N-acetyl-D-glucosamine (GlcNAc) (**1**-**3**), N-acetyllactosamine (LacNAc) (**5**), the O-acetylated derivatives of the GlcNAc carbohydrate residue (**4**), or the O-acetylated derivative of the LacNAc carbohydrate residue (**6**). The molecular structure for each of these compounds is shown.

Mass Spectrometry Analysis. Experiments were performed on a 9.4 T Bruker Apex IV QeFTMS (Billerica, MA) fitted with an Apollo II dual source, and a Synrad 75W maximum

power continuous wavelength CO_2 laser (model J48-2) for infrared multiphoton dissociation (IRMPD) experiments. Solutions of each synthetic neoglycolipid were made at a concentration of 5 μM in 50:50:0.1 chloroform:H_2O:formic acid (Sigma, St. Louis, MO) and ionized by electrospray using a flow rate of 2 μL/min. All neoglycolipids were analyzed in both positive and negative ion modes.

For MS^2 experiments, CAD was used for ion dissociation. Precursor ions were mass selected in the external quadrupole and stored for 2 to 4 s before fragmentation, using argon as a collision gas, followed by their injection into the ICR analyzer cell. For MS^3 experiments, CAD was performed as outlined above for the first stage of MS/MS. The product ion of interest was then isolated in the analyzer cell with a coherent harmonic excitation frequency (CHEF)[28] waveform before dissociation by IRMPD, in the ICR cell. In order to perform the IRMPD MS^3 experiment, the laser was set to between 50-80% of its maximum power, with an irradiation time of 0.5-1.5 seconds. Product ions were then excited and detected using standard protocols.

Results and Discussions. *Tandem mass spectrometry and MS^3 measurements.* The six different neoglycolipid derivatives examined in this study, Structures **1-6**, consist of a monosaccharide or disaccharide residue attached to a cholesterol moiety via a poly(ethoxy) linker, or in the case of compounds **2** and **4**, by a simple glycosidic bond. The ESI-MS spectrum of compound **1** is shown in Figure 1a. The protonated molecular ion, $[M+H]^+$, and the sodium

cationized molecular ion, $[M+Na]^+$ both appear in the mass spectrum, although their relative intensities depend on sample handling. It was observed that the abundance of the protonated molecule could be increased, relative to its sodiated counterpart, by washing the syringe and the fused silica capillary that was used to deliver the sample to the electrospray capillary with concentrated formic acid before injection of the samples. This procedure was utilized for all the experiments reported here to facilitate the acquisition of the MS^3 data for the protonated molecular ion. The protonated molecule of **1** was dissociated by CAD to yield the mass spectrum shown in Figure 1b. A number of product ions are observed, and their proposed formation is outlined in Scheme 1. The major product ion arises due to cleavage of the glycosidic bond between the sugar and the linker region at m/z 563, $[\text{-O-}(CH_2O)_4\text{-cholesterol} + H]^+$. In contrast to the more rigid bonds comprising sugar residues, glycosidic bonds are flexible and labile, and produce the majority of the high intensity peaks in the mass spectra of carbohydrates.[29] Accordingly, the most abundant ion in the MS/MS spectrum of neoglycolipid **1** results from glycosidic bond cleavage. Bond cleavage is also observed to occur between the cholesterol and the linker region, resulting in the $[\text{GalNAc-}(CH_2O)_4 + H]^+$ product ion at m/z 398. Cleavage of the bond between the linker and the cholesterol with retention of positive charge on the cholesterol moiety produces $[\text{cholestene} + H]^+$ at m/z 369.

The second most abundant product ion in the CAD mass spectrum of **1** appears at m/z 572, and has been assigned as the result of the 'internal elimination' of the poly(ethoxy) linker that resides between the cholesterol and sugar moieties in the original compound, with the composition of $[\text{C-glycoside} + \text{H}]^+$.[4] The structure of the proposed C-glycoside ion at m/z 572 was examined in a MS^3 experiment by using IRMPD. This produced two fragment ions, seen in Figure 1c, which correspond to $[\text{cholestene} + \text{H}]^+$ at m/z 369 and $[\text{GalNAc} + \text{H}]^+$ at m/z 204. These data confirm the composition of this ion, and suggest that m/z 572 is indeed the $[\text{C-glycoside} + \text{H}]^+$ ion, which dissociates as shown in Scheme 2. These data support the work of El-Aneed *et al.*, who proposed the *in situ* generation of a C-glycoside by 'internal elimination' of the linker region of the neoglycolipid.[4]

Effect of linker length on $[\text{C-glycoside} + \text{H}]^+$ *in situ formation.* The conditions necessary for the production of the $[\text{C-glycoside} + \text{H}]^+$ product ion were studied. Of the six compounds examined, two compounds, **2** and **4**, contained a single oxygen molecule as the linker between the sugar and cholesterol moieties, in contrast to the other compounds, which contain poly(ethoxy) linkers, $-(C_2H_4O)_nO-$, where n=3 or 4. In order to test the effect of the length of the linker region on the formation of the 'internal elimination' product, the CAD mass spectra of the protonated molecules of compounds **1**-**3** were compared. The CAD mass spectra obtained for **2** and **3** were very similar to those obtained for **1**. As **2** does not have a poly(ethoxy) linker, the

peaks observed for **1** that contain the linker are not observed for **2**. However, the [cholestene + H]$^+$ at m/z 369 and the [C-glycoside + H]$^+$ at m/z 572 do occur (supplemental data). **3** produces fragment ions at m/z 369, [cholestene + H]$^+$, and m/z 354, [GalNAc-$(CH_2O)_3$ + H]$^+$, as well as the [C-glycoside + H]$^+$ at m/z 572 (Figure 3a). It should be noted that for **3,** the 'internal elimination' product ion was the most intense fragment ion in the CAD spectrum, and this was found to be concentration dependent, as has been reported previously.[4] An increase in the concentration of the solution, causes an increase in the intensity of the [C-glycoside + H]$^+$ fragment ion. As a result of these data, it was concluded that the length of the poly(ethoxy) linker region, does not effect the *in situ* formation of the novel 'internal elimination' product ion. The [C-glycoside +H]$^+$ ion is formed both when the linker is a single oxygen atom or a much longer poly(ethoxy) chain. As mentioned above, the [C-glycoside + H]$^+$ product ions were isolated for each compound and an MS^3 experiment was performed to verify their composition (supplemental data).

Effect of saccharide length and acetylation on in situ formation of [C-glycoside + H]$^+$. The effect of saccharide length on the formation of the C-glycoside ion was examined by performing CAD on compound **5**, for which the structure differs from compound **3** only by the extension of the sugar to a dissacharide. The MS/MS spectrum of the protonated molecule of **5**, Figure 2a, exhibits [C-glycoside + H]$^+$ at m/z 734 as the third most abundant fragment ion. Thus

the elimination reaction does not require a monosaccharide moiety. The most abundant peaks from CAD of **5** result from glycosidic bond cleavage to yield a disaccharide "b" ion (Domon and Costello nomenclature [Domon, 1988 #49]) at m/z 366 and loss of the nonreducing end saccharide residue to yield the "y" ion at m/z 722, [GalNAc-$(CH_2O)_3$-cholesterol + H]$^+$. Loss of the cholesterol moiety yielded a fragment at m/z 516, [Gal-GalNAc-$(CH_2O)_3$ + H]$^+$.

The effect of acetylation of the saccharide on the formation of the C-glycoside product ion was examined by CAD studies of compounds **4** and **6**, which are identical to compounds **2** and **5** respectively, except for per-O-acetylation of the sugar moiety. CAD of the protonated molecule of compound ***6*** produced the disaccharide fragment at m/z 618, (Gal-GalNAc + H)$^+$, the disaccharide and linker fragment at m/z 768, [Gal-GalNAc-$(CH_2O)_3O$ + H]$^+$, and [Gal-GalNAc-$(CH_2O)_3$ - H]$^+$ at m/z 750, shown in Figure 2b. Two other product ions resulted from glycosidic cleavage between the sugar residues, the "z" ion, [GalNAc-$(CH_2O)_3O$-cholesterol + H]$^+$, at m/z 788, and its related "y" ion at m/z 806.[29] Significantly, no 'internal elimination' product ion arose for this acetylated compound. Similar results are observed for the pair of compounds, **2** and **4** (data not shown). CAD of the protonated molecule of compound **2** yields a C-glycoside product, while the CAD mass spectrum of its counterpart with a per-O-acetylated saccharide, compound **4**, does not exhibit this elimination product.

The observation that the per-O-acetylated compounds do not form the $[C\text{-glycoside} + H]^+$ product ion, although similar those of El-Aneed et al.,[4] varies somewhat from those described by McNeil, who demonstrated that a gas-phase glycosyl-residue elimination can occur in per-O-alkylated tetrasaccharide-alditols, partially O-ethylated trisaccharided-alditols and per-O-methylated methyl glycosides of oligosaccharides.[9] Although no mechanism was purposed for the glycosyl-residue elimination, it proved that simply hydrogen or hydroxyl groups do not need to be present in order for an *in situ* 'internal elimination' reaction to occur. Warrack et al. also found that peracetylated derivatives of anthracycline aminodisaccharides underwent an 'internal rearrangement' leading to the loss of an internal monosaccharide, proving that only one hydroxyl group is necessary for an internal rearrangement to occur.[13] This finding is also interesting due to the contradiction that it makes with conventional carbohydrate chemistry. It was believed that per-O-acetylated neoglycolipids would form a more reactive 1,2-cyclic oxonium ion, thus increasing the rate of reaction between it and the neutral cholesterol molecule.[4] It is now believed however, that these per-O-acetylated sugar oxonium ions are extremely reactive and thus can eliminate molecules of acetic acid and ketene to stabilize the molecule before it has time to react with the cholesterol, as is explained by El-Aneed et al..[4] Future explanations for this phenomenon may result when gas-phase reactions are further explored.

Effect of sodium ion adduction on in situ formation of [C-glycoside + H]$^+$. As mentioned above, Tadano-Aritomi et al. observed an unexpected internal residue loss from collision induced dissociation (CID) of a protonated molecule of glyceroplasmalopsychosine,[15] and Brull et al. and Kovacik et al. demonstrated, using CAD, that an internal residue loss occurs on 1→2 substituted oligosaccharides, as well as per-O-methylated trisaccharides having 1→6 linkages.[10,11] In each of these studies however, it was found that sodium-cationized glyceroplasmalopsychosine and oligosaccharides do not undergo internal residue losses during MS/MS experiments.[11] Previous research on neoglycolipids also suggests that protonation is necessary to trigger the 'internal elimination' ion-molecule reaction.[4,5] These finding were confirmed here by comparing the MS/MS behavior of protonated and sodium-adducted neoglycolipid ions. Figure 3 compares the CAD mass spectrum of the protonated molecular ion of compound **3**, Figure 3a, with the CAD spectrum of its sodium-adducted molecular ion, Figure 3b. The C-glycoside product is the most abundant ion in the MS/MS spectrum of the protonated ion, but is completely absent in the CAD mass spectrum of the sodium-adducted ion. The only fragment ions observed by CAD of the sodium-adduct are [GalNAc-$(CH_2O)_3$ + H]$^+$ and [-O-$(CH_2O)_3$-cholesterol + H]$^+$. These data, along with the fact that negative ion spectra did not produce the C-glycoside product ion (data not shown), confirm the hypothesis that a proton is necessary for initiation of the ion-molecule reaction that produces the C-glycoside gas-phase 'internal elimination' product ion.

Conclusions. By means of MS^3, these data lend further support to the composition of a novel *in situ* generated 'internal elimination' product ion, which is believed to occur through a gas-phase ion-molecule reaction upon CAD fragmentation of synthetic amphiphilic neoglycolipid cholesterol derivatives.[4] This C-glycoside product ion has been proposed to arise due to the elimination of a poly(ethoxy) linker region, producing a product ion consisting of a sugar and a cholesterol molecule. The present study provides the first example of MS^3 to examine the *in situ* C-glycosidation product which results during ESI-MS/MS of O-linked amphiphilic neoglycolipids, and shows that the length of the linker region or the sugar moiety does not affect the ability of these compounds to undergo this unusual elimination reaction.

In addition to confirming the composition of this novel product ion, the present study confirms that protonation is required for initiation of the proposed gas-phase ion-molecule reaction, which produces the C-glycoside product ion. This finding corroborates the work of El-Aneed et al., Tadano-Aritomi et al., and Brull et al. who all postulated that a proton is the driving reagent in unique gas-phase 'internal elimination' reactions.[4,11,15] It was also confirmed in this study that acetylation blocks the ion-molecule 'internal elimination' reaction. The high abundance of C-glycoside products from a variety of neoglycolipids by low energy CAD should encourage the exploration of alternative routes to the synthesis of these important targets.

Acknowledgments. The authors gratefully acknowledge financial support from the National Institute of Health grant RR19767.

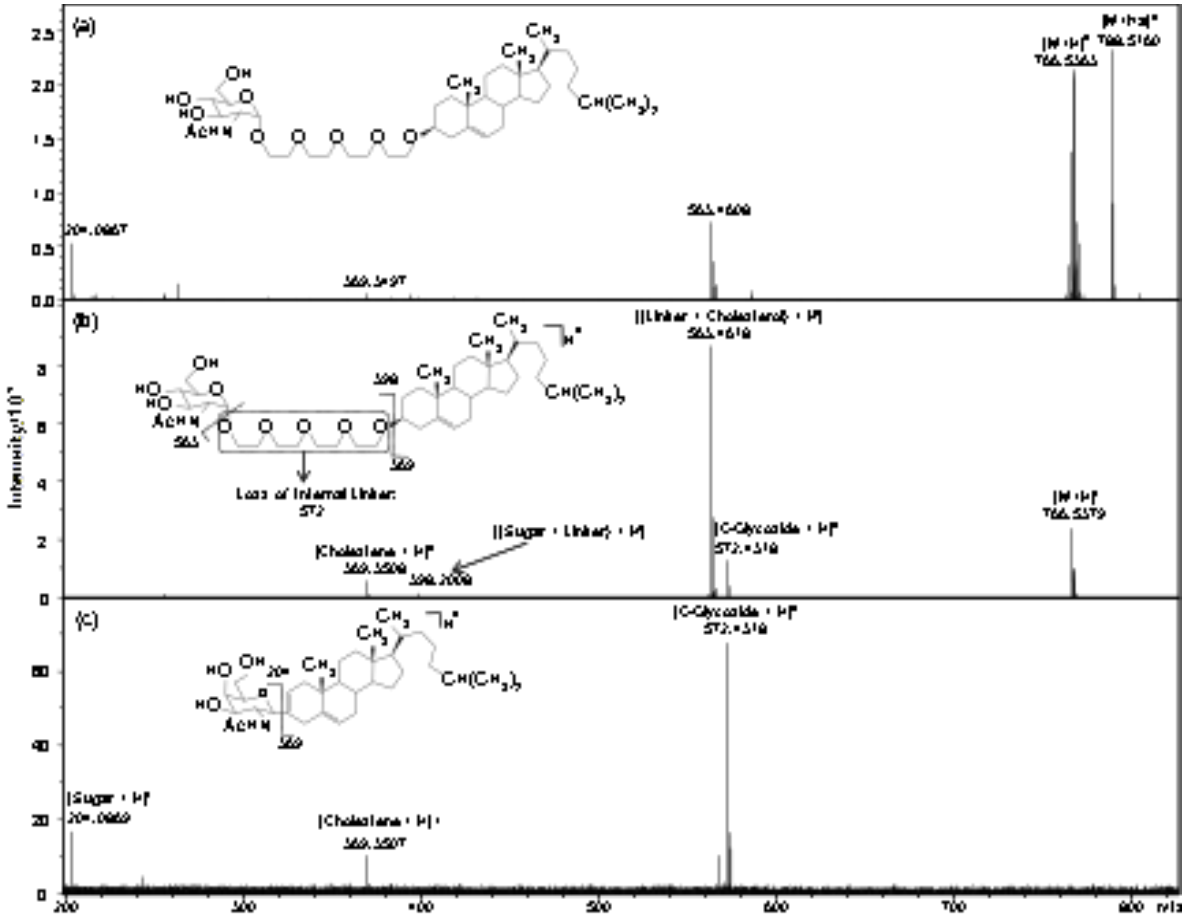
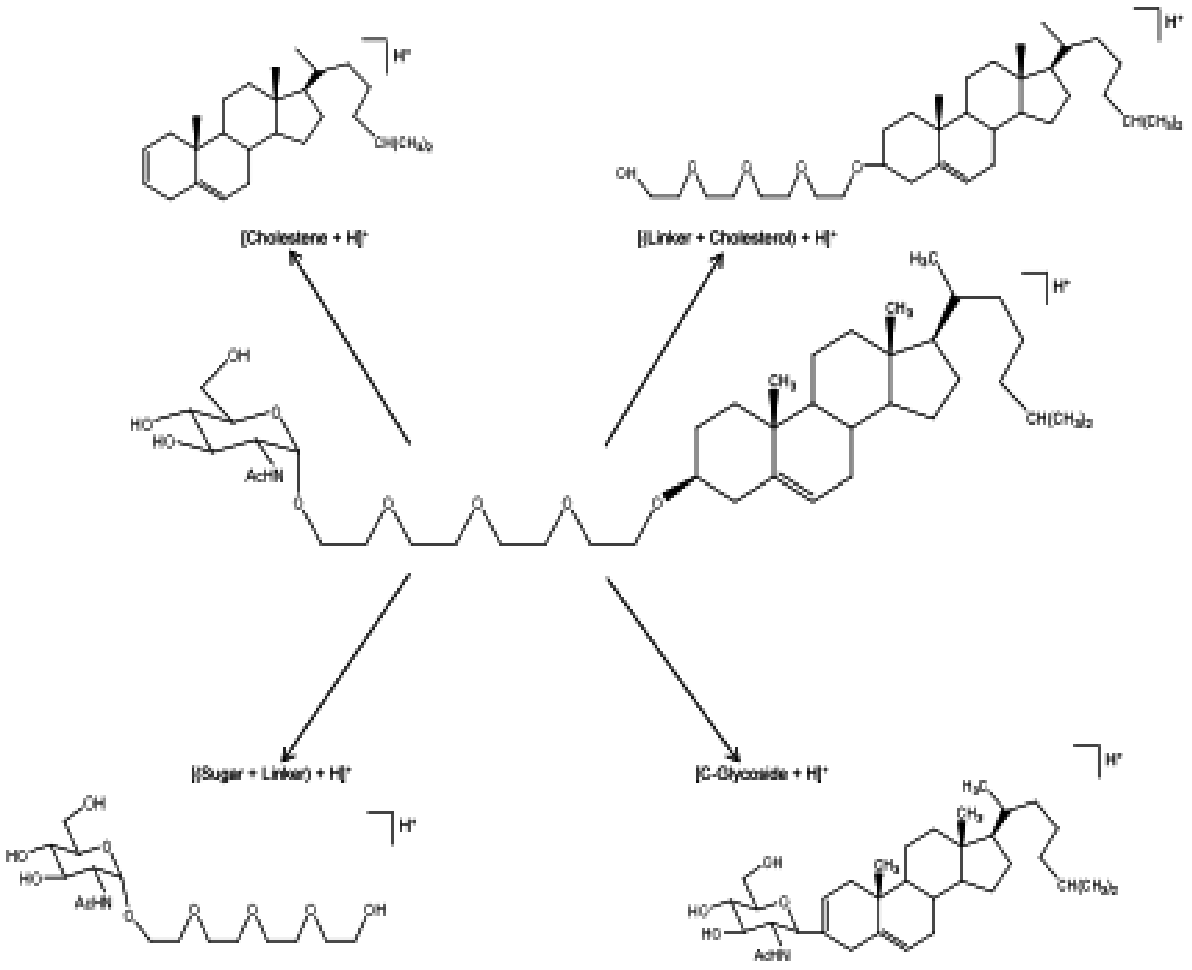
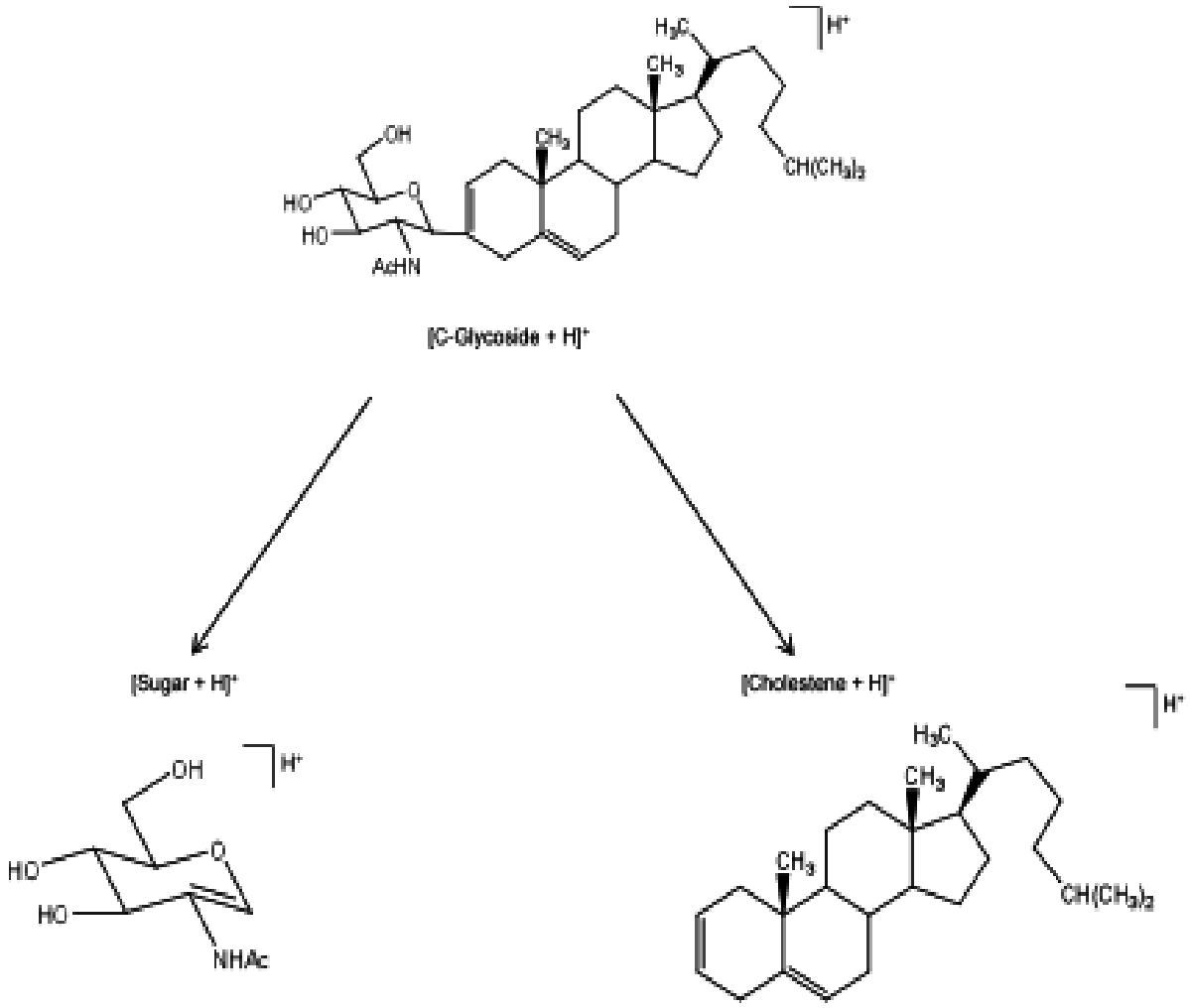


Figure 5.1. (a) ESI FTICR/MS of synthetic neoglycolipid **1**, (b) low-energy CAD spectra of the precursor $[M+H]^+$ at m/z 766, protonated **1**, and (c) MS^3 spectra using low-energy CAD to dissociate $[M+H]^+$, followed by IRMPD to dissociate the $[C\text{-glycoside} + H]^+$ product ion at m/z 572.

Scheme 5.1. Scheme representing possible fragmentation pattern resulting in [C-Glycoside + H]$^{+}$, 'internal elimination' production ion.



Scheme 5.2. Scheme representing possible fragmentation pattern resulting in [C-Glycoside + H]$^+$, 'internal elimination' production ion.



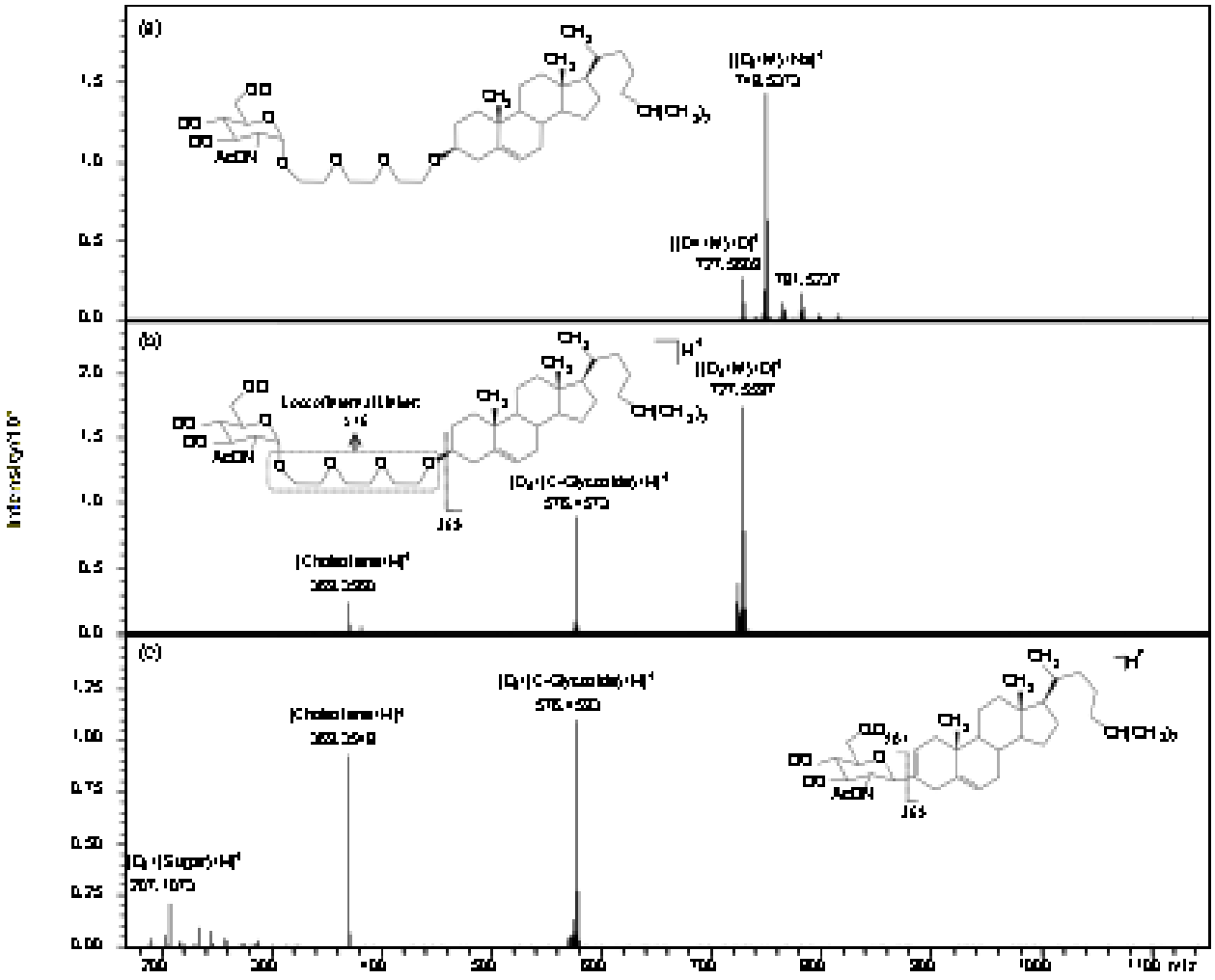


Figure 5.2.(a) FTICR MS/MS using CAD of **5**, non-acetylated neoglycolipid, and (b) CAD of **6**, a per-O-acetylated neoglycolipid.

Figure 5.3.

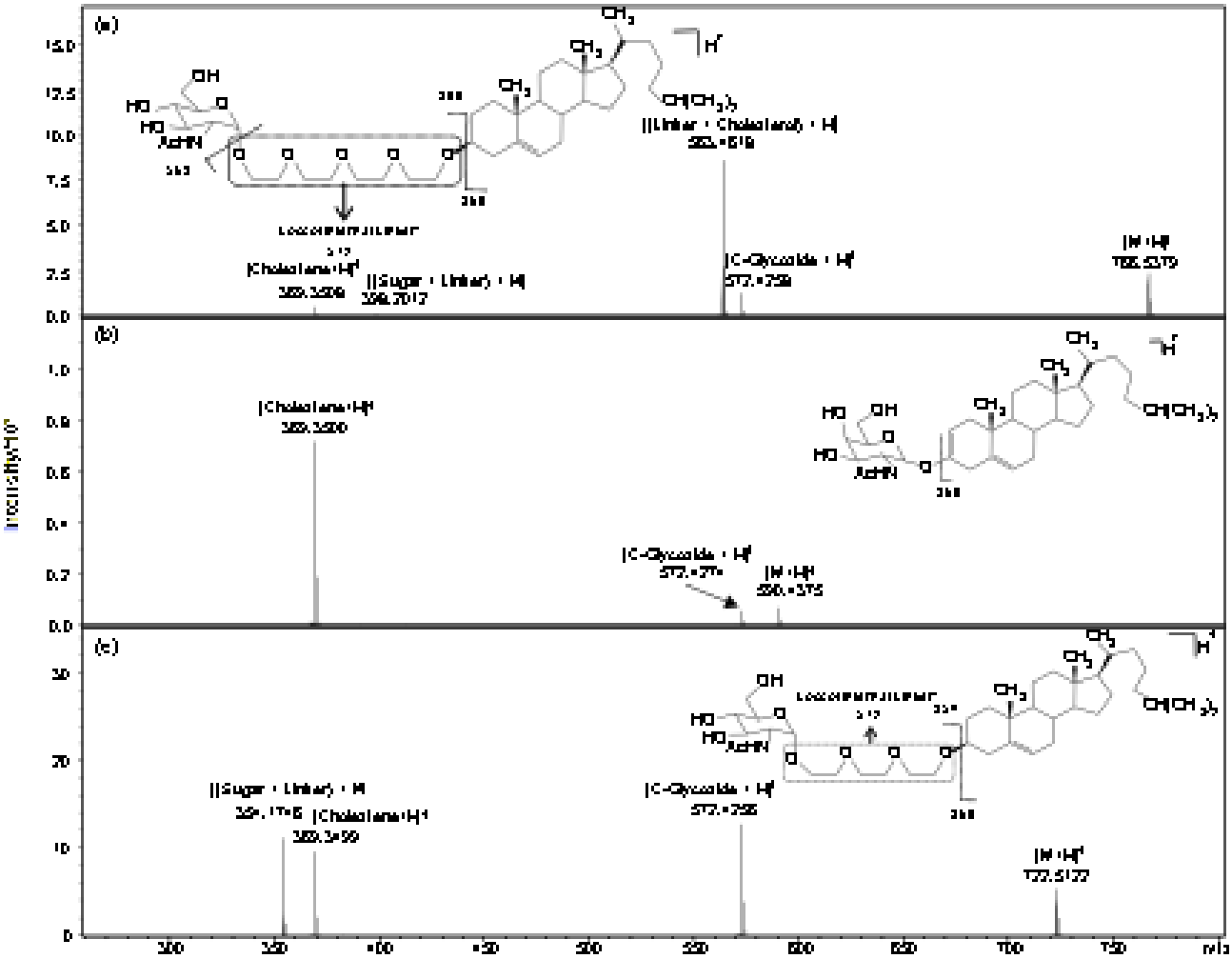


Figure 5.3. (a) FTICR-MS/MS using CAD of the protonated molecular ion of **3** at m/z 722, and

(b) FTICR-MS/MS using CAD of m/z 744, the sodium cationized molecular ion of **3**.

References:

1. Xu, Z.; Jayaseharan, J.; Marchant, R. E. Synthesis and characterization of oligomaltose-grafted lipids with application to liposomes. *J. Colloid Interface Sci.* **2002**, *252*, 57-65.

2. Duffels, A.; Green, L. G.; Ley, S. V.; Miller, A. D. Synthesis of high-mannose type neoglycolipids: Active targeting of liposomes to macrophages in gene therapy. *Chemistry-a European Journal* **2000**, *6*, 1416-1430.

3. Perouzel, E.; Jorgensen, M. R.; Keller, M.; Miller, A. D. Synthesis and formulation of neoglycolipids for the functionalization of liposomes and lipoplexes. *Bioconjugate Chem.* **2003**, *14*, 884-898.

4. El-Aneed, A.; Banoub, J.; Koen-Alonso, M.; Boullanger, P.; Lafont, D. Establishment of mass spectrometric fingerprints of novel synthetic cholesteryl neoglycolipids: The presence of a unique C-glycoside species during electrospray ionization and during collision-induced dissociation tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 294-310.

5. Banoub, J.; Boullanger, P.; Lafont, D.; Cohen, A.; El Aneed, A.; Rowlands, E. In situ formation of c-glycosides during electrospray ionization tandem mass spectrometry

of a series of synthetic amphiphilic cholesteryl polyethoxy neoglycolipids containing N-acetyl-D-glucosamine. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 565-570.

6. Peri, F.; Cipolla, L.; Rescigno, M.; La Ferla, B.; Nicotra, F. Synthesis and biological evaluation of an anticancer vaccine containing the C-glycoside analogue of the Tn epitope. *Bioconjugate Chem.* **2001**, *12*, 325-328.

7. Schmieg, J.; Yang, G. L.; Franck, R. W.; Tsuji, M. Superior protection against malaria and melanoma metastases by a C-glycoside analogue of the natural killer T cell ligand alpha-galactosylceramide. *J. Exp. Med.* **2003**, *198*, 1631-1641.

8. Kuberan, B.; Sikkander, S. A.; Tomiyama, H.; Linhardt, R. J. Synthesis of a C-glycoside analogue of sTn: An HIV- and tumor-associated antigen. *Angew. Chem., Int. Ed.* **2003**, *42*, 2073-2075.

9. McNeil, M. Elimination of Internal Glycosyl Residues During Chemical Ionization-Mass Spectrometry of Per-O-Alkylated Oligosaccharide-Alditols. *Carbohydr. Res.* **1983**, *123*, 31-40.

10. Kovacik, V.; Hirsch, J.; Kovac, P.; Heerma, W.; Thomasoates, J.; Haverkamp, J. Oligosaccharide Characterization Using Collision-Induced Dissociation Fast-Atom-Bombardment Mass-Spectrometry - Evidence for Internal Monosaccharide Residue Loss. *J. Mass Spectrom.* **1995**, *30*, 949-958.

11. Brull, L. P.; Heerma, W.; Thomas-Oates, J.; Haverkamp, J.; Kovacik, V.; Kovac, P. Loss of internal 1 --> 6 substituted monosaccharide residues from underivatized and per-O-methylated trisaccharides. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 43-49.

12. Ernst, B.; Muller, D. R.; Richter, W. J. False sugar sequence ions in electrospray tandem mass spectrometry of underivatized sialyl-Lewis-type oligosaccharides. *Int. J. Mass Spectrom. Ion Processes* **1997**, *160*, 283-290.

13. Warrack, B. M.; Hail, M. E.; Triolo, A.; Animati, F.; Seraglia, R.; Traldi, P. Observation of Internal Monosaccharide Losses in the Collisionally Activated Dissociation Mass Spectra of Anthracycline Aminodisaccharides. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 710-715.

14. Brull, L. P.; Kovacik, V.; Thomas-Oates, J. E.; Heerma, W.; Haverkamp, J. Sodium-cationized oligosaccharides do not appear to undergo 'internal residue loss' rearrangement processes on tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1520-1532.

15. Tadano-Aritomi, K.; Hikita, T.; Kubota, M.; Kasama, T.; Toma, K.; Hakomori, S.; Ishizuka, I. Internal residue loss produced by rearrangement of a novel cationic glycosphingolipid, glyceroplasmalopsychosine, in collision-induced dissociation. *J. Mass Spectrom.* **2003**, *38*, 715-722.

16. Shi, S. D.-H.; Hendrickson, C. L.; Marshall, A. G.; Siegel, M. M.; Kong, F.; Carter, G. T. Structural validation of saccharomicins by high resolution and high mass accuracy fourier transform-ion cyclotron resonance-mass spectrometry and infrared multiphoton dissociation tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1285-1290.

17. McLuckey, S. A. Principles of Collisional Activation in Analytical Mass-Spectrometry. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 599-614.

18. Zubarev, R. A. Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* **2004**, *15*, 12-16.

19. Tsybin, Y. O.; Witt, M.; Baykut, G.; Kjeldsen, F.; Hakansson, P. Combined infrared multiphoton dissociation and electron capture dissociation with a hollow electron beam in Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1759-1768.

20. Little, D. P.; Speir, J. P.; Senko, M. W.; Oconnor, P. B.; McLafferty, F. W. Infrared Multiphoton Dissociation of Large Multiply-Charged Ions for Biomolecule Sequencing. *Anal. Chem.* **1994**, *66*, 2809-2815.

21. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* **1998**, *120*, 3265-3266.

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

23. Wolff, J. J.; Amster, I. J.; Chi, L.; Linhardt, R. J. Electron Detachment Dissociation of Glycosaminoglycan Tetrasaccharides. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 234-244.

24. Boullanger, P.; Chevalier, Y.; Croizier, M. C.; Lafont, D.; Sancho, M. R. Synthesis and Surface-Active Properties of Some Alkyl 2-Amino-2-Deoxy-Beta-D-Glucopyranosides. *Carbohydr. Res.* **1995**, *278*, 91-101.

25. Lafont, D.; Boullanger, P. Syntheses of L-glucosamine donors for 1,2-trans-glycosylation reactions. *Tetrahedron-Asymmetry* **2006**, *17*, 3368-3379.

26. Lafont, D.; Boullanger, P.; Chierici, S.; Gelhausen, M.; Roux, B. Cholesteryl oligoethyleneglycols as D-glucosamine anchors into phospholipid bilayers. *New J. Chem.* **1996**, *20*, 1093-1101.

27. Lafont, D.; Boullanger, P.; Carvalho, F.; Vottero, P. A convenient access to beta-glycosides of N-acetyllactosamine. *Carbohydr. Res.* **1997**, *297*, 117-126.

28. Heck, A. J. R.; Dekoning, L. J.; Pinkse, F. A.; Nibbering, N. M. M. Mass-Specific Selection of Ions in Fourier-Transform Ion-Cyclotron Resonance Mass-Spectrometry - Unintentional Off-Resonance Cyclotron Excitation of Selected Ions. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 406-414.

29. Domon, B.; Costello, C. E. A Systematic Nomenclature for Carbohydrate Fragmentations in FAB-MS MS Spectra of Glycoconjugates. *Glycoconjugate J.* **1988**, *5*, 397-409.

CHAPTER 6:

Conclusions

It is apparent from the above studies that chemical crosslinking combined with FTICR-MS/MS can provide insight into the protein protein interactions that take place within biological complexes, such as 34 kDa protein and Actin. Applying mass defect label technology along with labile bonds within the chemical crosslinker, allows the researchers to identify the compounds of interest with greater ease, as well as fragment the crosslinker completely away from the peptides of interest, without fragmenting the peptide bonds. Subsequent MS/MS fragmentation then allows for the amino acids involved in the interaction to be elucidated. Although continued work needs to be focused on development of software for both data acquisition and data analysis, it was proven during this study that the novel mass defect labeled chemical crosslinker, DiBBSIAS, was easily identifiable within the mass spectra and was also labile enough to cleave under low energy dissociation methods, yielding ultimate amino acid sequence information. Using this crosslinker, we were able to identify specific areas of interaction between 34 kDa protein and Actin, thereby aiding in the research of Hirano bodies and hopefully the cure for diseases such as Alzheimer's in the future.