

ENRICHMENT AND IDENTIFICATION OF DISULFIDE BONDED PEPTIDES BY
TANDEM MASS SPECTROMETRY

by

POOJA RAJESHBHAI RAMANUJ

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE IN CHEMISTRY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2017

Copyright © by Pooja Rajeshbhai Ramanuj2017

All Rights Reserved



Acknowledgements

I would like to thank my supervising professor, Dr. Saiful M. Chowdhury for his support in the research and my academic career. He has motivated me to work to the best of my potential. I am also thankful to my colleagues and friends for their constant help and encouragement. I also appreciate the contribution and suggestions from my committee members, Dr. Kevin Schug and Dr. Kayunta Johnson-Winters, which have certainly improved the quality of this research. Moreover, I am grateful to the Department of Chemistry and Biochemistry at UT Arlington and Shimadzu Center for Advanced Analytical Chemistry (SCAAC) for mass spectrometric support.

Finally, I am very thankful to my family who has constantly supported and motivated me in my career.

July 28, 2017

Abstract

ENRICHMENT AND IDENTIFICATION OF DISULFIDE BONDED PEPTIDES BY TANDEM MASS SPECTROMETRY

Pooja Rajeshbhai Ramanuj, M.S.

The University of Texas at Arlington, 2017

Supervising Professor: Saiful M. Chowdhury

Proteins are Comprised of different amino acids which can undergo various modifications known as post-translational modifications (PTMs). Amino acid's side chains are the most common sites for the PTMs. Cysteine side chain can undergo several PTMs and one very common one is disulfide bond formation. Disulfide bond aids in maintaining protein tertiary structures, its identification is an inevitable aspect in studying protein functions. Furthermore, disulfide bonds are potential targets to develop new protein therapeutics for life-threatening diseases such as cancer. Literature surveys revealed several methods to identify disulfide bonds, which require either high quality sample or are highly laborious. Moreover, unexpected folding patterns in protein structure limits the use of data software in disulfide identification. In this study, a straightforward and novel method was developed for confident identification of cysteine residues involved in disulfide bond formation. In order to develop a robust strategy for mapping disulfide bonds in proteins, we developed a chemical strategy. First, we reduced the disulfide bonds to two -SH and subsequently reacted with a reagent called 2, 3- dibromomaleimide (DBM). The maleimide reacted with sulfhydryl bond of cysteine amino acid and substituted the disulfide bonds. Cysteine sulfur was converted to sulfoxide by mild oxidation to make this bond low energy tandem mass spectrometry cleavable. Due to cleavage, this produced two different

peptides peaks in the mass spectra so disulfide-bonded peptides can be sequenced as linear modified peptides using tandem mass spectrometry. Furthermore, to enrich the disulfide bonds, we added a biotin group in the dibromomaleimide. Biotin is a small molecule, which has strong interaction with a protein avidin. After reduction of disulfide bonds in peptides, we used a similar dibromomaleimide reagent with a biotin group incorporated in it. After that peptide were enriched using avidin immobilized agarose or magnetic beads. After washing the non-binders, the disulfide-bonded peptide was eluted from the beads and peptide sequence was analyzed using liquid chromatography and tandem mass spectrometry. We tested our approach with standard disulfide containing peptides, somatostatin and a protein lysozyme. Efficient labeling was observed in standard peptides and we have identified several disulfide bond scrambling in lysozyme.

In comparison to conventional approaches, our method has three major advantages: (I) reduced sample complexity (II) precise mapping of disulfide bonds (III) faster and easy MS data analysis. Although, there are certain challenges involving oxidation study using biotin dibromomaleimide (BDBM), this method highly enhances the confidence identification of disulfide bonds. We believe this strategy could be an efficient way to enrich and analyze disulfide bonds in proteins.

Table of Contents

Acknowledgements	iii
Abstract	iv
Table of Contents	vi
List of Figures.....	vii
List of Tables.....	xi
List of Abbreviations.....	xii
Chapter 1 Introduction.....	13
1.1 Background	13
1.2 Conventional method for identification of disulfide bonds.....	18
1.3 Motivation and Objective.....	23
Chapter 2 Materials and methods.....	24
2.1 Materials.....	24
2.2 Sample Preparation	25
2.3 Methods	25
Chapter 3 Results and discussion	29
3.1 Chemical modification and enrichment using biotin maleimide	30
3.2 Bridging characteristics of 2,3-dibromo maleimide and oxidation	34
3.3 Bridging and enrichment characteristics of biotin dibromomaleimide and oxidation ..	47
Chapter 4 Conclusion.....	57
Chapter 5 References	59
Biographical Information	64

List of Figures

Figure 1-1 N-Acetylation mechanism (N-terminal amino group is being modified by acetyl group of acetyl co-enzyme A)	15
Figure 1-2 Types of Post-translational modifications. Three common sites of PTMs (amino acid side chain, disulfide bond and peptide bond) ¹²	16
Figure 1-3 Scheme of profile comparison approach for disulfide bond identification. Chromatogram comparison shows two missing peaks in reduced sample	20
Figure 1-4 Schemes for different alkylation reagents for disulfide mapping.....	20
Figure 1-5 Bioconjugation reaction of thiolate (Sulfhydryl group reacts differently with various bioconjugates reagent. Maleimide reaction uses Michael addition chemistry.) ...	22
Figure 2-1 Preparation of biotin dibromomaleimide. Biotin 1 reacts with thionyl chloride and 2,3-dibromomaleimide to form biotin dibromomaleimide 3.....	24
Figure 2-2 Nomenclature used for inter-peptide fragments data analysis. b and y ion generation from different peptide chain uses different pattern for nomenclature.	28
Figure 3-1 Schematic representation of chemical modification and oxidation of cysteine residue in protein sample	29
Figure 3-2 LC-MS/MS spectra of biotin-maleimide labeled RGDC in ESI-IT-TOF-MS. (A) Full MS (451.19 represents $[M+2H]^{2+}$ of modified RGDC) (B) MS/MS at 451.19 m/z.....	31
Figure 3-3 LC-MS/MS spectra of biotin-maleimide labeled RGDC peptide in ESI-IT-TOF-MS. (A) full MS (451.19 represents $[M+2H]^{2+}$ of modified RGDC) and MS/MS at 451.19 m/z (Myoglobin spiked sample) (B) full MS (451.19 represents $[M+2H]^{2+}$ of modified RGDC) and MS/MS at 451.19 m/z (Enriched sample)	32
Figure 3-4 Identification of somatostatin disulfide bond reduction reaction by TCEP in ESI-IT-TOF-MS. (a) pure somatostatin (b) reduced somatostatin (addition of two hydrogen can be seen from mass difference compare to pure sample.)	33

Figure 3-5 LC–MS/MS spectra of biotin-maleimide labeled somatostatin peptide in ESI-IT-TOF-MS. (A) Full MS (697.98 represents $[M+2H]^{3+}$ of modified somatostatin) (B) MS/MS at 697.98 m/z and 1045.00 m/z (b and y ions from modified peptide backbone under CID MS/MS)..... 34

Figure 3-6 : LC–MS/MS spectra of biotin-maleimide labeled somatostatin peptide in ESI-IT-TOF-MS. (A) Full MS (697.98 represents $[M+2H]^{3+}$ of modified somatostatin in myoglobin digest) and MS/MS at 697.98 m/z and 1045.97 m/z (b and y ions from peptide backbone) (Spiked sample) (B) Full MS and MS/MS at 697.98 m/z and 1045.96 m/z (enriched sample) 36

Figure 3-7 LC–MS/MS spectra of 2,3-dibromomaleimide labeled RGDC peptide in ESI-IT-TOF-MS. (A) Full MS and MS/MS at 623.09 m/z (B) Full MS and MS/MS at 623.09 m/z of RGDC spiked in myoglobin digest sample..... 37

Figure 3-8 LC–MS/MS spectra of 2,3-dibromomaleimide labeled somatostatin peptide in ESI-IT-TOF-MS. (A) Full MS and MS/MS 578.24 m/z (B) Full MS and MS/MS at 578.58 m/z of somatostatin spiked myoglobin digest sample..... 38

Figure 3-9 LC–MS/MS spectra of oxidized somatostatin peptide using 3% H_2O_2 in ESI-IT-TOF-MS. (A) Oxidized somatostatin with mass difference compare to DBM modified peptide (B) MS/MS at 891.84 with characteristic $RS_2O_3H_2$ loss and b and y ions from remaining peptide backbone 40

Figure 3-10 LC–MS/MS spectra of oxidized somatostatin peptide, spiked in myoglobin digest, using 3% H_2O_2 in ESI-IT-TOF-MS. (A) $[M+2H]^{2+}$ and $[M+3H]^{3+}$ of oxidized somatostatin in myoglobin digest (B) MS/MS at 891.84 with characteristic $RS_2O_3H_2$ loss and b and y ions from remaining peptide backbone 41

Figure 3-11 LC–MS/MS spectra of 2,3-dibromomaleimide labeled somatostatin peptide in ESI-IT-TOF-MS. (A) additional proof of unreacted peptide fragment (NFFWK) in full MS (B)

Full MS and MS/MS at 343.46 m/z of 2,3-dibromomaleimide labeled peptide fragments from somatostatin digest.....	42
Figure 3-12 LC–MS/MS spectra of oxidized somatostatin peptide using 3% H ₂ O ₂ in ESI-IT-TOF-MS. (A) 2,3-dibromomaleimide bridged A-G-C-K and T-F-T-S-C fragments from somatostatin digest (B) Oxidized 2,3-dibromomaleimide bridged A-G-C-K and T-F-T-S-C fragments from somatostatin digest and MS/MS at 524.17 with characteristic RSOH loss and b and y ions from remaining peptide backbone. (C) Proof of oxidation study of somatostatin digest in myoglobin digest sample with characteristic RSOH loss and b and y ions from remaining peptide backbone	45
Figure 3-13 LC–MS/MS spectra of 2,3-dibromomaleimide labeled lysozyme digest peptides in ESI-LIT-MS. MS/MS at 733.93 m/z of peptide fragment (CELAAAMK) bonded with (GYSLGNWCAAK)	46
Figure 3-14 LC–MS/MS spectra of 2,3-dibromomaleimide labeled lysozyme digest peptides in ESI-LIT-MS. MS/MS at 1216.54 m/z of peptide fragment NLC76NIPC80SALLSSDITASVNC94AK.....	47
Figure 3-15 LC–MS/MS spectra of biotin dibromomaleimide labeled RGDC peptide in nanoESI-LIT-MS. Full MS with zoomed isotopic peaks (inset) (left). MS/MS at 863.36 m/z (right)	48
Figure 3-16 : LC–MS/MS spectra of biotin dibromomaleimide labeled somatostatin peptide in nano ESI-LIT-MS. (A) Full MS and MS/MS at 987.93 m/z in pure peptide sample (B) Full MS and MS/MS at 987.46 m/z of somatostatin spiked myoglobin digest sample. (C) Full MS and MS/MS at 987.36 m/z of enriched biotin dibromomaleimide labeled somatostatin	50
Figure 3-17 LC–MS/MS spectra of oxidized somatostatin peptide using 3% H ₂ O ₂ in nanoESI-LIT-MS. Full MS of oxidized somatostatin with double and triple charged ions	52

Figure 3-18 LC–MS/MS spectra of biotin dibromomaleimide labeled somatostatin peptides in nanoESI-LIT-MS from somatostatin digest. Full MS of labeled TFTSC fragments (top). MS/MS at 725.73 m/z (bottom) 53

Figure 3-19 LC–MS/MS spectra of biotin dibromomaleimide labeled lysozyme digest peptide in nanoESI-LIT-MS at 1337.52 m/z..... 54

Figure 3-20 LC–MS/MS spectra of biotin dibromomaleimide labeled lysozyme peptides in ESI-LIT-MS from lysozyme digest. Full MS of chemically labeled peptide 1(IVSDGNGMNAWVAWRNRCK) and peptide 2 (CELAAAMK) (top). MS/MS at 1118.00 m/z (bottom) 55

Figure 3-21 LC–MS/MS spectra of biotin dibromomaleimide labeled lysozyme peptides (enriched) in ESI-LIT-MS from lysozyme digest. Full MS of chemically labeled peptide 1(IVSDGNGMNAWVAWRNRCK) and peptide 2 (CELAAAMK) (top). MS/MS at 559.27 m/z (bottom) 56

List of Tables

Table 3-1 Theoretical calculated mass (m/z) for different number of oxygen addition at sulfur atom in thiol group of cysteine modified by dibromomaleimide	39
Table 3-2 Theoretical calculated mass (m/z) for different number of oxygen addition at sulfur atom of cysteine containing peptide fragments of somatostatin digest modified by dibromomaleimide.....	43
Table 3-3 List of possible disulfide bonded lysozyme peptides after trypsin digestion	46
Table 3-4 Theoretical calculated mass (m/z) for different number of oxygen addition at sulfur atom in thiol group of cysteine modified by biotin dibromomaleimide.....	51

List of Abbreviations

BDBM	Biotin dibromomaleimide
BM	Biotin maleimide
CDAP	Dimethylamino-pyridinium tetrafluoroborate
CID	Collision induced dissociation
DBM	2,3-dibromomaleimide
DTT	Dithiothreitol
ECD	Electron-capture dissociation
ESI-IT-TOF	Electrospray ionization-ion trap-time-of-flight
ETD	Electron-transfer dissociation
HPLC	High performance liquid chromatography
LC-MS	Liquid chromatography mass spectrometry
MALDI-TOF	Matrix assisted laser desorption ionization-time of flight
mRNA	Messenger ribonucleic acid
nano LC-ESI-LIT	Nanoelectrospray ionization linear ion trap
NMR	Nuclear magnetic resonance spectroscopy
PBS	Phosphate saline buffer
PTMs	Post -translational modifications
ROS	Reactive oxygen species
TCEP	Tris(2-carboxyethyl)phosphine
UVPD	Ultraviolet Photo dissociation

Chapter 1

Introduction

1.1 Background

A gene is the specific region of DNA, which is made up of various nucleotides and mainly responsible for heredity.^{1,2} A complete set of DNA is known as a genome for an organism.³ DNA further uses this information to form messenger RNA (mRNA) by the process known as transcription. The transfer of genetic information from messenger RNA (mRNA) to proteins is known as translation. A proteome is a group of proteins mainly expressed by an organism or cellular systems. Proteomics is known as the large-scale study of a particular proteome to get information about protein abundance, variations and modifications.^{4,5} It also aids in the clear understanding of the cellular processes.^{6,7} It is known that proteomes have a higher complexity than genomes primarily due to the number of genes involved in the formation of proteins.² A literature review^{2,4,6} reveals that human genomes mainly consist of 20,000 to 25,000 genes, while proteome has more than 1 million proteins. This number helps us to understand the complexity of proteomes as single gene can encode multiple proteins. Another reason behind the complexity of proteome is post-translational modifications (PTMs). Modifications occur after translation has been completed, categorized in post-translation modifications.^{8,9}

Proteins are made up of various amino acids, which can undergo various post-translational modifications (PTMs). Over 200 diverse types of PTMs have been identified until now and new types are continually discovered.^{10,11} Amino acid side chains or peptide linkages are the most common sites of PTMs.⁸ PTMs mainly regulate divergent functions, localization and interaction of proteins. There are mainly three different types of PTMs that control the functionality and structures of proteins. Type 1 modifications are characterized by the addition of small and large molecules at the side chain of amino acids, such as

phosphate group, acetyl group or ubiquitin. Types 2 modifications are characterized by a change in the peptide bonds that can be irreversibly cleaved by many proteases. Type 3 modifications involve changes in disulfide bonds, which include a pair of cysteine residues in the protein structure.¹² There are some common Type 1 modifications, such as phosphorylation, glycosylation, ubiquitination, methylation, acylation and hydroxylation are briefly described below. Post-translational modifications are critical regulators of protein activity and function. Understanding the role of PTMs in disease states is an ongoing effort towards the development of novel biomarkers and therapeutics.

- Phosphorylation is a form PTMs where a phosphate group addition occurs at the side chain of serine, tyrosine or threonine residues. Phosphorylation at a specific amino acid side chain represents the nucleophilic behavior of the -OH group of side chains.^{13,14}
- Glycosylation is one of the most complex forms of PTMs among others. It includes addition of carbohydrate molecule to the side chain of amino acids starting from simple monosaccharides to polysaccharides units. They are classified in N-, O- or C- linked glycosylation PTMs based on the nature of attached group to sugar moiety-peptide bonds.¹⁵
- Ubiquitination is the addition of small regulatory protein i.e. ubiquitin (contains 76 amino acid residue). It can be further classified depending on the number of ubiquitin molecules added to the protein. There are two classes of molecules: monoubiquitin and polyubiquitination. A lysine residue and/or the N-terminus are the primary sites of ubiquitin modification.^{16,17}
- *N*-Acetylation is the addition of acetyl group to nitrogen atom of amino acid side chain via reverse or irreversible mechanism as shown in Figure 1-1.^{18,19}

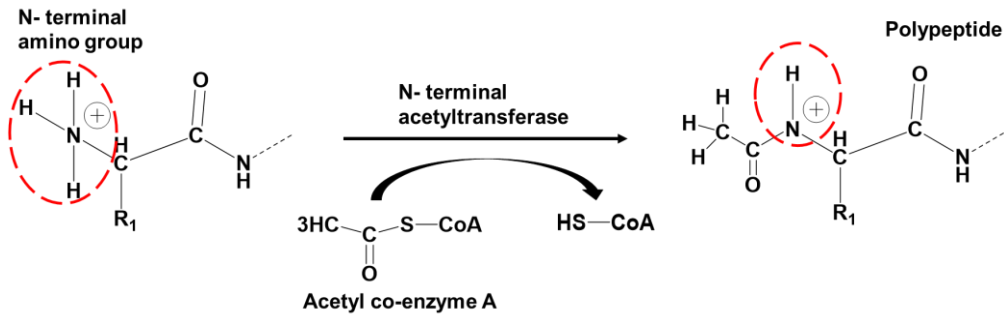
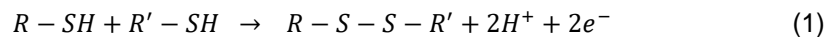


Figure 1-1 N-Acetylation mechanism (N-terminal amino group is being modified by acetyl group of acetyl co-enzyme A)

There are other two types of modifications. They include disulfide bond formation and proteolytic cleavage. The formation of disulfide bonds mainly involves the modification of chemical bond structure via formation of covalent disulfide bonds, while proteolytic cleavage involves the breaking of an amide bond through removal of HOH. The linkage between sulfur atom of two cysteine residues within the same or different polypeptide chain are known as disulfide bonds. The simplest form of disulfide bonds include two cysteine residues as shown in the Figure 1-2. The literature^{12,20} suggests that 50% of the total disulfide bonds in proteins primarily are secreted or function in external environment of endoplasmic reticulum, Golgi apparatus or in endosome. However, some disulfide bonds are found in proteins, which function in the cytoplasm and/or nucleus, although the environment is not favorable for disulfide bond stabilization. Methionine is an amino acid, which also contain sulfur atom but are not capable to form disulfide bonds as they contain an additional methyl group. Reaction (Equation 1) given below shows the general scheme of formation of disulfide bond.



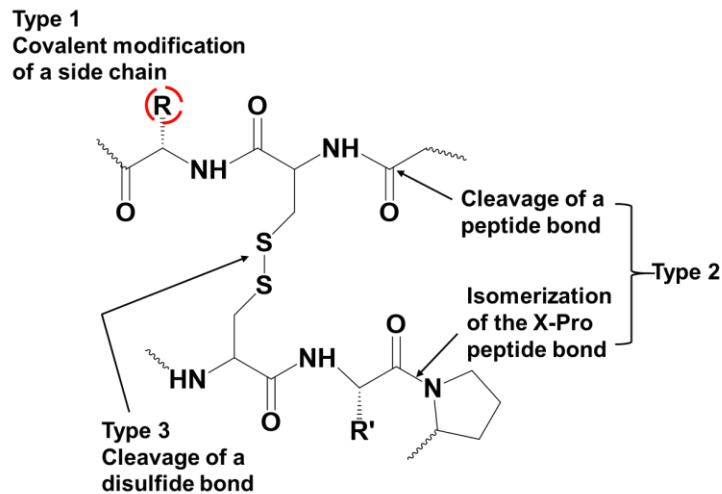


Figure 1-2 Types of Post-translational modifications. Three common sites of PTMs (amino acid side chain, disulfide bond and peptide bond)¹²

Disulfide bonds are known to be relatively stronger, compared to amide bond (between N and C), having dissociation energy between 55-65 kcal/mol. However, they are about 40% weaker compared to carbon-carbon (C-C) and carbon-hydrogen (C-H) bonds in peptides and proteins. Moreover, disulfide bond also plays a significant role in protein stability and the biological functions through maintaining protein modifications and three-dimensional structures.^{21,22} Pace *et al.*²³ and Denton *et al.*²⁴ studied the effects of disulfide bonds on protein stability. Pace *et al.*²³ also reported that the Rnase A structure is 8 times more stable in a folded confirmation compare to the unfolded state. Some researchers have studied the effect of disulfide bonds on antibodies as well. For instance, recombinant of murine IgG CH3 domain disulfide bond stability had been studied by Thies *et al.*²⁵ which showed that tertiary structure was less affected compared to drastic effects on stability.

There are several ways through which disulfide bonds can stabilize the structure of proteins and peptides in their folded forms. Firstly, the peptide sequences come together and form disulfide bonds by arranging it to a folded form. It is also reported that disulfide bonds help to increase the free energy and lower its entropy to form a stable structure.²⁶ The second important folding mechanism is the “hydrophobic effect”. The hydrophobic effect is the phenomenon in which nonpolar surface molecules interact with like molecules rather than polar molecules. The hydrophobic molecules of amino acids remain within the core of protein structure while the periphery comprises of hydrophilic amino acid. This makes water molecule interactions the least favorable and retain peptide bonds as well as maintenance of the stability in secondary structure.²⁷

Disulfide bond containing peptides have various crucial roles in biological systems. Glutathione (cysteinyl peptide) which is found in both oxidized and reduced forms, act as a redox buffer and an antioxidant. Reactive oxygen species (ROS) play a vital role in maintenance of homeostasis and pathophysiological condition of renal diseases. Cellular component damage caused by free radicals and ROS can be prevented or repaired by the different forms of glutathione.^{28,29}

Hormones are also disulfide bonds containing peptides or proteins, which responsible for one’s physical as well mental growth and wellness. Insulin is universally known to have three disulfide bonds containing hormonal peptides. Other examples are oxytocin, vasopressin and somatostatin.³⁰ Growth factors which have the main function of stimulation of cellular growth. Defensins known as disulfide bonds containing proteins to protect our body from various viral, fungal and bacterial infections.³¹

Finally, the last class of compound which is secreted from plant or animals but venomous to human are known as “toxins”. Some of the well-known disulfide bonds containing toxins are snake venom³² and snail venom³³. Disulfide bonds containing

antibodies are responsible for the immune system as well as reproduction and metabolism.

³⁴ In the 2003, the very first disulfide linked drug conjugate was approved for the human use. Hence, identification of disulfide bond connection patterns is an inevitable step to get the detailed information about various functions and biological activity. ³⁵ This concept is the focus of this thesis.

1.2 Conventional method for identification of disulfide bonds

Correct identification of disulfide bonds depends upon several factors. For example, a detailed amino acid sequence of a protein structure is the prime requirement to predict the connectivity of disulfide bonds. Very few analytical techniques available can give the detailed information about amino acid sequences and disulfide bonds connectivity as well. Moreover, intrachain disulfide bond makes protein structure more complicated to investigate. Edman degradations study has been widely used by researchers to evaluate the primary structure of proteins and peptides although it is not able to give any information regarding the disulfide bond pattern. ³⁶ Furthermore, evaluation becomes more complicated when there are more than two cysteines present in the primary sequence. Walewska *et al.* ³⁷ and Klaus *et al.* ³⁸ conducted experiments using nuclear magnetic resonance spectroscopy (NMR) to detect the disulfide bonds in conotoxin. X-rays crystallography is also an effective method for the structure determination of proteins. ³⁹ Both of these techniques offer exceptional identification of disulfide bonds with minimum disulfide scrambling. Although, NMR and x-ray crystallography require high purity, high concentrations, large quantities of sample, tedious sample preparation and grim to crystalize smaller quantity sample make them challenging among researcher and users.

Mass spectrometry has been used as an alternative tool to detect disulfide bonds for many years due to less sample requirements, high speed and sensitivity. Liquid

chromatography mass spectrometry (LC-MS) with database software has gained popularity within the past few years to analyze tremendous data of PTMs.^{40,41,42} Peng *et al.*⁴³ reported the top down approaches for the identification of five disulfide bonds in alpha-amylase from human. Moreover, it is more favorable compared to bottom up approach as it conserves the labile modifications due to less man-made modification and disulfide bond scrambling. However, literature review reveals very limited information about top down approach due to need for costly, highly sensitive instruments.^{43, 44, 45}

As noted earlier, disulfide needs a higher activation energy compare to amide bonds for cleavage, various bottom up methods have been developed. Conventional MS methods utilize various enzymes and alkylating reagents to identify the cysteine connection prior to MS analysis. The very first method to identify disulfide bonds using a bottom up approach is known as profile comparison. Zhang *et al.*⁴⁶ and Zhang and Cockrill⁴⁷ reported disulfide bond connectivity in monoclonal antibody and insulin, respectively. Profile comparison method includes the comparison of reduced and non-reduced enzymatic digest of selected proteins followed by peak identification, fraction collection and mass analysis. Zhang and Cockrill⁴⁷ used Endoprotease Glu-C for digestion and Tris(2-carboxyethyl) phosphine (TCEP) for reduction of protein sample. As a result, reduced and nonreduced peptides of insulin digest is generated. Disulfide bonded peaks collection, from non-reduced sample, were done for further identification using mass spectrometer. General scheme of profile comparison is shown in Figure 1-3.

Profile comparison approach is an upfront way to identify disulfide-associated peptides from High performance liquid chromatography (HPLC) fractions. However, large sample requirement, complicates the experiments when dealing with a high number of disulfide bonds and missed cleaved peptides are some pitfalls associated with this method. Digestion without reduction approach produces very high mass peptides which are beyond

the range of mass analyzer. Hence, universal acceptance is not possible.⁴⁸ In order to overcome this problem some researchers suggested partial reduction followed by alkylation using two different alkylation reagents.^{49,50} Jones *et al.* solved the three among four disulfide bond region pattern for tumor necrosis factor using this approach. Instead of alkylation, cyanylation using dimethylamino-pyridinium tetrafluoroborate (CDAP) and analysis using Matrix assisted laser desorption ionization-time of flight (MALDI-TOF) was reported by some researcher for adjacent cysteine containing samples.^{51,52} Figure 1-4 represents the general steps for both approaches listed above.

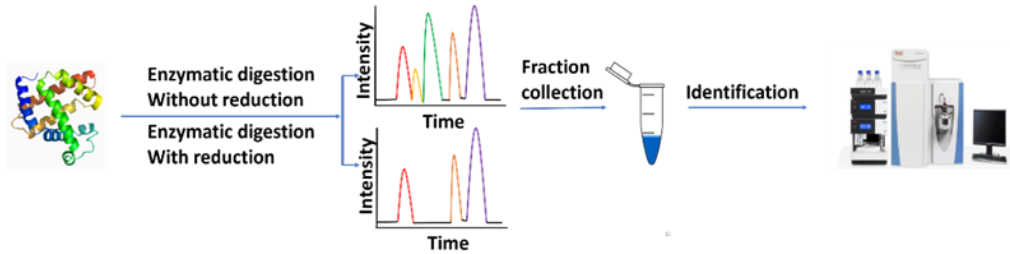


Figure 1-3 Scheme of profile comparison approach for disulfide bond identification.

Chromatogram comparison shows two missing peaks in reduced sample

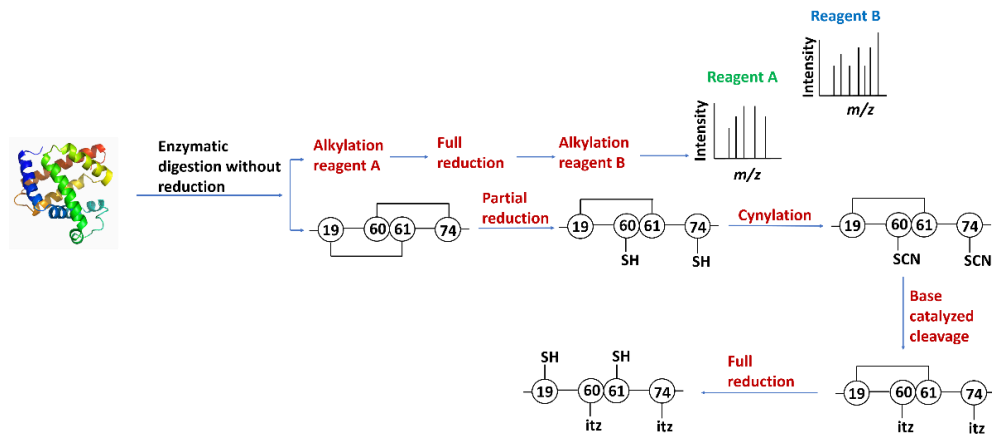


Figure 1-4 Schemes for different alkylation reagents for disulfide mapping.

However, precise reduction condition is the prerequisite for effective reaction. The need of different alkylation reagents for different proteins and peptides make this approach less favorable among researchers. Although mass spectrometric methods possess tremendous advantages, it shows some ambiguity when it comes to disulfide bond identification. This is a major concern when protein sequence contains scrambling patterns. Various gas-phase dissociation techniques such as electron-transfer dissociation (ETD), electron-capture dissociation (ECD), ultraviolet photo dissociation (UVPD) have been developed to simplify the sample preparation steps. Despite of the simplicity, identification of scarce disulfide bond containing proteins remains a challenge.⁵³

Bioconjugation, a method of covalent derivatization of biomolecules, is an alternative and is potentially a more versatile approach to locate disulfide bonds.⁵⁴ Moreover, site-selective modification of any amino acid helps to interrogate detailed in vivo and invitro pathophysiological conditions of any system.⁵⁵ Success of a bioconjugation reaction mainly depends upon the degree of crosslinking between the reaction group of involved amino acid and reagents. Hence, poor selection of site specific modification leads to undesirable product through loss of function of the biomolecules. Recently, many new techniques have been premeditated for site specific bioconjugation that help to retain the biological functions of biomolecules.

Cysteine is one of the least abundant amino acids in nature. Cysteine modification through a thiolate is a widely used bioconjugation method among others available.^{56,57} Polyethyleneglycol (PEG), maleimide and iodoacetamide are commonly used reagents for site specific modification of cysteine residues. Maleimide and iodoacetamide act as strong electrophiles and react with nucleophilic side chains. As shown in Figure 1-5 the maleimide group reacts with cysteine side chain via Michael addition reaction and form succinimidyl thioethers.⁵⁴ However, hydrolysis is a major drawback associated with maleimide

reactions.⁵⁷ Furthermore, some potential limitations of maleimide reagents like nature of reaction and number of attachment positions limit its use. In 2010, the Baker group⁵⁸ published a new class of compounds, known as the bromomaleimide. It was used for the reversible modification of a cysteine residue. However, one point of attachment still causes protein instability. Smith *et al.*⁵⁹ reported that 2,3-dibromomaleimide (DBM) has two points of attachment to maintain protein structure and reversible nature of reaction for biological applications.

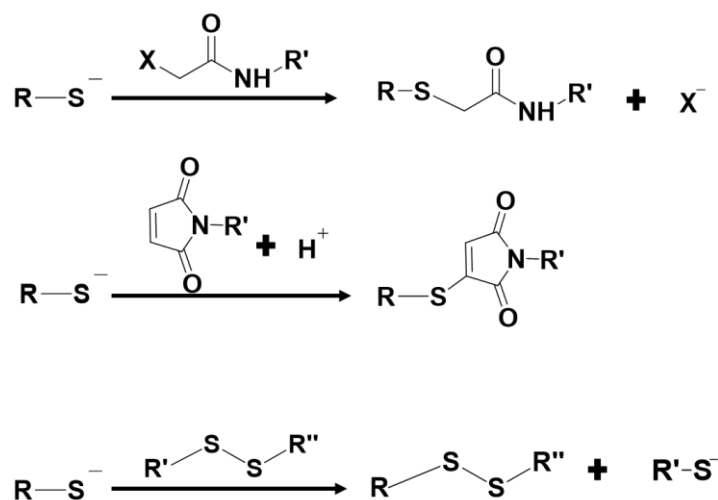


Figure 1-5 Bioconjugation reaction of thiolate (Sulfhydryl group reacts differently with various bioconjugates reagent. Maleimide reaction uses Michael addition chemistry.)

Various affinity tags were used with bioconjugates to increase the effectiveness of precise identification. Biotin can be used as an affinity handle to selectively react with avidin or streptavidin beads. It also helps to detect the targeted proteins or peptides through secondary bioconjugates via immobilized avidin beads or resin or fluorescent avidin tags.

1.3 Motivation and Objective

From literature review, currently available methods have sample complexity on large scale study. Moreover, mass spectrometry data analysis is also challenging. To our knowledge, there is no method and bioconjugates available which can overcome the limitations mentioned above and precisely measure the disulfide bonds in peptides and proteins. The purpose of the work presented here is to develop a chemistry-based approach which can precisely locate and identify the disulfide bonds through tandem mass spectrometry. Additionally, we aim to use a novel maleimide compound and affinity chromatography, to eliminate the unwanted peptide sequence from the sample and enrich cysteinyl peptides for higher confidence identification. The current thesis also addresses higher confidence identification of disulfide bonds using DBM. Chowdhury *et al.*⁶¹ reported neutral fragment loss from oxidized thioether bond under low-energy. Our aim is to use the same cleavable properties of sulfoxide group in a gas phase, after DBM and biotin dibromomaleimide (BDBM) treatment, to get characteristics neutral fragment loss of RSOH from precursor ions, (where R is equal to added mass of maleimide containing bioconjugates), using tandem mass spectrometry (MS/MS). Oxidation properties help to generate different products, which show ideal fragmentation behavior. Oxidation of bioconjugates modified disulfide bonds is a novel confidence identification method described here.

Chapter 2

Materials and methods

2.1 Materials

Biological Materials: Peptide RGDC and somatostatin were purchased from Genscript (New Jersey). Protein lysozyme (from chicken egg white) was purchased from Sigma-Aldrich (St. Louis, MO).

Chemical Materials: 2,3- dibromomaleimide and biotin maleimide (BM) were purchased from Sigma Aldrich and used without further purification. Biotin dibromomaleimide was synthesized from 2,3 dibromomaleimide (explained below). Sequencing-grade modified trypsin was used for protein and peptide digestion and purchased from Promega (Madison, WI). TCEP was acquired from Thermo Scientific. Hydrogen peroxide (28-32% v/v) was purchased from Alfa Aesar. HPLC grade solvents (acetic acid, formic acid) for mobile phase were purchased from Sigma-Aldrich. Water from Milli-Q water system was used for sample and mobile phase preparation when necessary.

Preparation of biotin dibromomaleimide: Biotin dibromomaleimide **3** (MW = 481.16) was synthesized in two steps. First, the carboxylic acid functional group in biotin **1** was reacted with thionyl chloride (SOCl_2) to form its corresponding acid chloride **2**. Then, this freshly prepared acid chloride was coupled with 2,3-dibromomaleimide under basic conditions to afford the pure product (Figure 2-1).

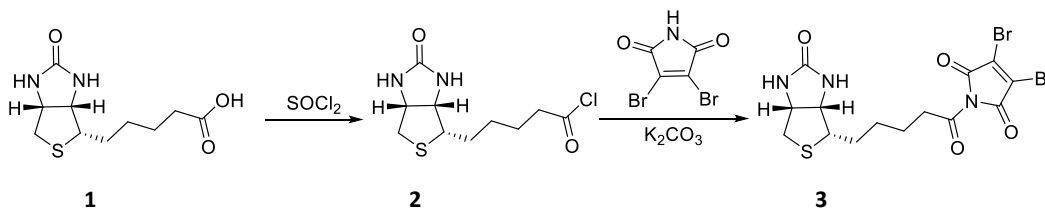


Figure 2-1 Preparation of biotin dibromomaleimide. Biotin **1** reacts with thionyl chloride and 2,3-dibromomaleimide to form biotin dibromomaleimide **3**

2.2 Sample Preparation

Peptide and Protein digestion: Somatostatin (10mM, 10 μ L) and lysozyme (1mM, 5 μ L) solutions were digested using sequential grade trypsin. Lyophilizate trypsin was diluted using HCL to make the finale concentration 0.1 μ g/ μ L. 1:100 protease: protein ratio and 4 hrs incubation time were selected for digestion of peptide. The sample pH (7-8) was maintained during digestion using 50 mM NH_4HCO_3 . Digestion reaction was ceased using 10 μ L of 0.1% formic acid. The sample was vacuum dried followed by reconstitution, using the same solvent. Solution reduction was done using TCEP. The sample was speed vacuumed until dry and desalted using pierce C18 ziptips.

Myoglobin digestion: Myoglobin solution (1 mM, 5 μ L) was taken and trypsin was used for the digestion. Prior to digestion, the sample was diluted with 50 mM NH_4HCO_3 and then trypsin solution was added to the resulting solution at a ratio of protein: trypsin (100:1) and incubated at 37 °C for 12 hrs. Digestion reaction was ceased using 10 μ L of 0.1% formic acid. The solvent was removed, using vacuum and reconstituted using 0.1 % formic acid. Pierce C18 ziptips were used for salt removal.

2.3 Methods

Chemical modification of cysteine residue of peptides using biotin-maleimide: The effectiveness of the BM conjugation reaction toward cysteine was studied using two model peptides. RGDC peptide solution was taken in an Eppendorf. Phosphate saline buffer (PBS) (pH = 7.4, 300 μ L) was added to maintain the pH of the solution. BM solution was added in solution (1:5 peptide: BM) and kept at room temperature for 10 min followed by solvent drying under vacuum and desalted using pierce C18 ziptip for MS analysis.

Somatostatin digestion was carried out as described in section 2.2. Somatostatin solution (10mM, 10 μ L) and somatostatin digest were reduced with 5 equiv. of TCEP for

30 min at room temperature. Phosphate saline buffer (PBS) (pH = 7.4, 300 μ L) was added to maintain the pH of the solution. BM solution was added in solution (1:5 peptide: BM) and kept at room temperature for 60 min followed by solvent drying under vacuum and desalted using pierce C18 zip tip for MS analysis.

Chemical modification of disulfide bonds of peptides and proteins using 2,3-dibromo maleimide and biotin dibromomaleimide: The DBM conjugation reaction was carried out by two model peptides along with lysozyme. Various pH conditions (pH=6.2, 7.5), buffers (sodium phosphate and PBS buffer) and reducing reagents (dithiothreitol (DTT) and TCEP) were tested to select optimum reaction conditions. RGDC peptide solution was taken in an Eppendorf. Phosphate saline buffer (PBS) (pH = 6.2, 300 μ L) was selected and added to maintain the pH of the solution. DBM solution was added in solution (1:5 peptide: DBM) and kept at room temperature for 10 min followed by solvent drying under vacuum and desalted using pierce C18 zip tip for MS analysis.

Somatostatin digestion was carried out as described in section 2.2. Somatostatin solution (10mM, 10 μ L) and somatostatin digest were reduced with 5 equiv. of TCEP for 30 min at room temperature. Phosphate saline buffer (PBS) (pH = 6.2, 300 μ L) was added to maintain the pH of the solution. DBM solution was added in solution (1:5, peptide: DBM) and kept at room temperature for 60 min followed by solvent drying under vacuum and desalted using pierce C18 zip tip for MS analysis.

Lysozyme was digested without reduction as described in section 2.2. The digest was reduced with 5 equiv. of TCEP for 45 min at room temperature. Phosphate saline buffer (PBS) (pH = 6.2, 300 μ L) was added to maintain the pH of the solution. DBM solution was added in solution (1:5 peptide: DBM) and kept at room temperature for 2 hours, followed by solvent drying under vacuum and desalted using pierce C18 zip tip for MS analysis.

Sample preparation for spiking study: Myoglobin digest was prepared as described in section 2.2. Peptides were spiked in myoglobin digest and reaction was done using biotin maleimide and biotin dibromomaleimide as described in earlier sections.

Enrichment reaction using affinity chromatography: Spiked samples were prepared as described in the previous section. 20 μ L of avidin beads were taken followed by washing with 300 μ L PBS (pH =7.4) for 3 times. Samples were incubated with avidin beads from 30 min to 1 hr. Beads were washed with 200 μ L of PBS (pH =7.4), 200 μ L of 25 mM NaCl and 200 μ L of water for 3, 2 and 3 times respectively. Sample elution was done using acetonitrile: water: FA (50:50:0.5%) solution. Sample was dried using vacuum. 0.1% FA was used for reconstitution prior to run in LC-MS.

Mass Spectrometry for BM and DBM modified sample: Mass spectrometric studies were performed by positive ion mode using electrospray ionization-ion trap-time-of-flight (ESI-IT-TOF) from Shimadzu Scientific Instruments. Some samples were also analyzed by Thermo Velos Pro dual-pressure nano electrospray ionization linear ion trap (nano LC-ESI-LIT) mass spectrometer. Products were separated by C18 column (100 \times 2.1 mm, C18, 2.6 μ m, 100 Å) in (ESI-IT-TOF) whereas nano Viper Acclaim PepMap 100 column (30 μ m \times 5 mm, C18, 100 Å) was used with Thermo instrument. A 65-min gradient with 5-100% organic phase and 4-90% organic phase was used with Shimadzu and Thermo instrument respectively. Flow rate 0.2000 mL/min, collision energy 60% and frequency (59.5 KHz) were other parameters for Shimadzu instrument. Flow rate 0.300 μ L/min with 40% collision energy, Collision induced dissociation (CID) activation time 10 ms, isolation with m/z of 1.5 were used for separation and CID experiment of products with Thermo instrument. Data dependent method were used with top 10 intense peak selection for the same experiment.

Oxidation reaction: Chemically modified protein and peptides were taken and reacted with 3% H₂O₂ for 20 to 35 min depending on type of samples. Sample was dried using speed vacuum. C18 ziptip was used prior to MS analysis.

Analysis software and nomenclature: CID can generate various b and y ion fragments during tandem MS/MS as shown in Figure 2-2. Molecular weight calculator was used to get b and y ion list for BM modified peptide. In house software, X-link explorer was used to identify DBM and BDM crosslinked peptides. Peak labeled as [Pept bn] represent intact peptide 1 and b1 fragment of peptide 2 whereas [bn pept] represent intact peptide 2 and b1 fragment of peptide 1. [bn--] and [yn--] show ions from peptide 1 while [--bn] and [--yn] show ions from peptide 2 (see figure below). Stavrox and X-link explorer were used to get modified sequence for lysozyme samples.

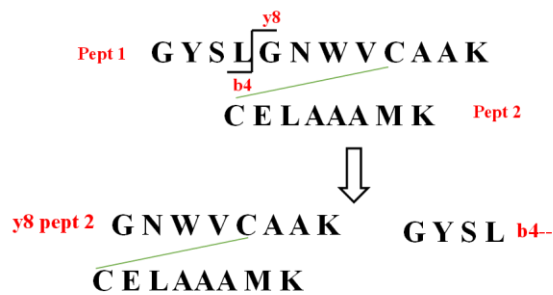


Figure 2-2 Nomenclature used for inter-peptide fragments data analysis. b and y ion generation from different peptide chain uses different pattern for nomenclature.

Chapter 3

Results and discussion

To develop an effective and feasible approach for disulfide bond enrichment and identification, we planned a versatile strategy that is described in Figure 3-1.

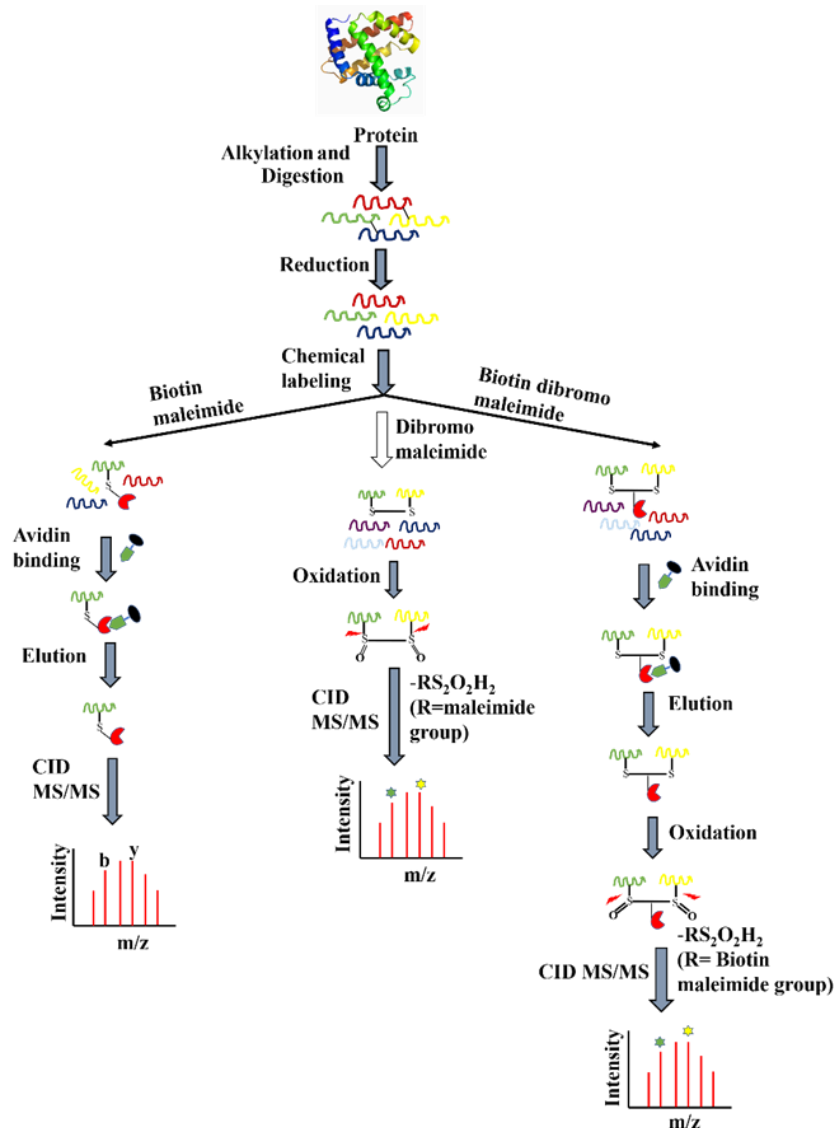


Figure 3-1 Schematic representation of chemical modification and oxidation of cysteine residue in protein sample

First, proteins were alkylated to block the free cysteine residue in its sequence followed by digestion.⁶² After obtaining disulfide bonded peptides, reduction was conducted using TCEP to introduce maleimide containing reagents. Next, three different approaches used for different reagents. In general, biotin containing reagents were used to selectively enrich the modified peptides using avidin-biotin chemistry,⁶³ whereas the dibromo group was used to mimic the disulfide bond property in protein structure. The oxidation and cleavable property of thio-ether bond was used to get signature mass loss (RSOH- depends upon number of oxidation).⁶¹ Oxidation serves two purposes: 1) help to introduce cleavable site in crosslinked peptide and 2) gives detailed information about peptide and cysteine residue involved in disulfide bonds.

3.1 Chemical modification and enrichment using biotin maleimide

The experiment was started with biotin-maleimide reagents. BM was used to check the reactivity of maleimide towards cysteine residue. RGDC (for structure refer Figure 3-2) is a cell binding motif, which can hinder platelet aggregation.⁶⁴ RGDC (monoisotopic mass 450.18) was chemically labeled using biotin-maleimide. The theoretical monoisotopic peak of chemically labeled RGDC peptide was 899.35, with the mass addition of 450.53 Da of biotin-maleimide group. Figure 3-2 (A) shows the intense peaks of biotin-maleimide labeled RGDC peptide at 901.36 [M+H]⁺ and 451.19 [M+2H]²⁺. The peptide peak at 451.19 [M+2H]²⁺ was isolated and fragmented using CID in ESI-IT-TOF-MS. Few b and y ions (from peptide backbone) were observed during tandem MS/MS analysis. Figure 3-2 (B) shows the b and y ions from biotin-maleimide labeled RGDC peptide. MS/MS figure shows very few fragments of b and y ions. As RGDC is a small peptide and some ions are less than 200 Da range, very few b and y ions were expected a priori.

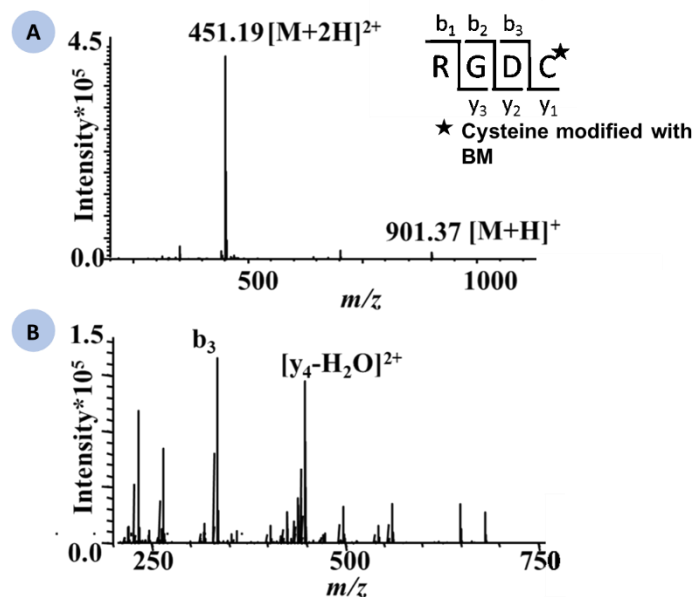


Figure 3-2 LC-MS/MS spectra of biotin-maleimide labeled RGDC in ESI-IT-TOF-MS. (A) Full MS (451.19 represents [M+2H]²⁺ of modified RGDC) (B) MS/MS at 451.19 *m/z*

RGDC peptide reaction with BM was used as a proof of concept for maleimide reactivity towards the sulfhydryl group of cysteine residue. To demonstrate that this peptide can be effectively enriched using avidin-biotin chemistry, affinity chromatography was used to selectively enrich modified peptide from myoglobin digest. Myoglobin digest was used as it does not have any cysteine in its sequence. Selection of 451.19 *m/z* (theoretical monoisotopic mass 450.67) from the MS peak list gives the mass spectra of BM labeled RGDC with intense peaks at 451.19 [M+2H]²⁺ and 901.37 [M+H]⁺. The peptide peak at 451.19 [M+2H]²⁺ was isolated and fragmented using CID in ESI-IT-TOF-MS. Few b and y ions (from peptide backbone) were observed in tandem MS/MS analysis. Figure 3-3(A) shows the b and y ions from biotin-maleimide labeled RGDC peptide in myoglobin spiked sample. It gives further confirmation of maleimide reactivity in complex sample. Figure

3-3(B) shows the successful enrichment of sample using affinity chromatography from myoglobin digest.

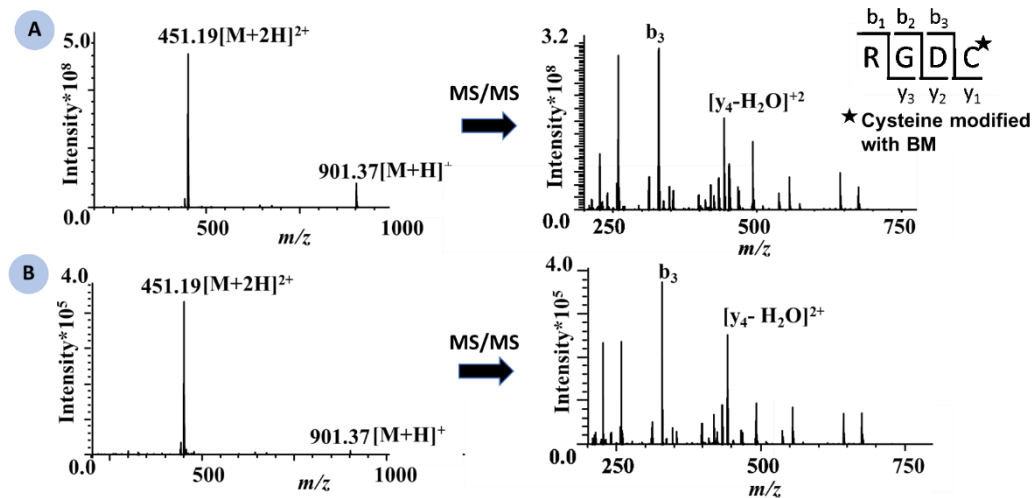


Figure 3-3 LC-MS/MS spectra of biotin-maleimide labeled RGDC peptide in ESI-IT-TOF-MS. (A) full MS (451.19 represents $[M+2H]^{2+}$ of modified RGDC) and MS/MS at 451.19 m/z (Myoglobin spiked sample) (B) full MS (451.19 represents $[M+2H]^{2+}$ of modified RGDC) and MS/MS at 451.19 m/z (Enriched sample)

Somatostatin, a peptide (for structure refer Figure 3-4) that contains 14 amino acids with one disulfide bond was selected for a higher level study. As somatostatin contains one disulfide bond, TCEP was used to break that disulfide bond. We observed that TCEP worked well in breaking disulfide bonds without using any harsh conditions such as incubation at room temperature. Figure 3-4 shows the reactivity of TCEP towards disulfide bonds. The theoretical peak of TCEP reduced somatostatin (monoisotopic mass 1637.73) has a mass of 1639.74 Da. High intensity peaks of 820.37 $[M+2H]^{2+}$ and 547.25 $[M+3H]^{3+}$ in Figure 3-4 (B) further confirmed the addition of hydrogen atoms after the reduction reaction. Next, the reduced somatostatin (monoisotopic mass 1637.73) was chemically labeled using biotin-maleimide. The theoretical monoisotopic peak of

chemically labeled somatostatin was 2088.92, with the mass addition of 450.53 Da of biotin-maleimide group. Figure 3-5 (A) shows the intense peaks of biotin-maleimide labeled somatostatin peptide at 1045.00 $[M+2H]^{2+}$ and 697.98 $[M+3H]^{3+}$. The peptide peaks were isolated and fragmented using CID in ESI-IT-TOF-MS. b and y ions shown in Figure 3-5(B) were in agreement with the expected fragmentation pattern.

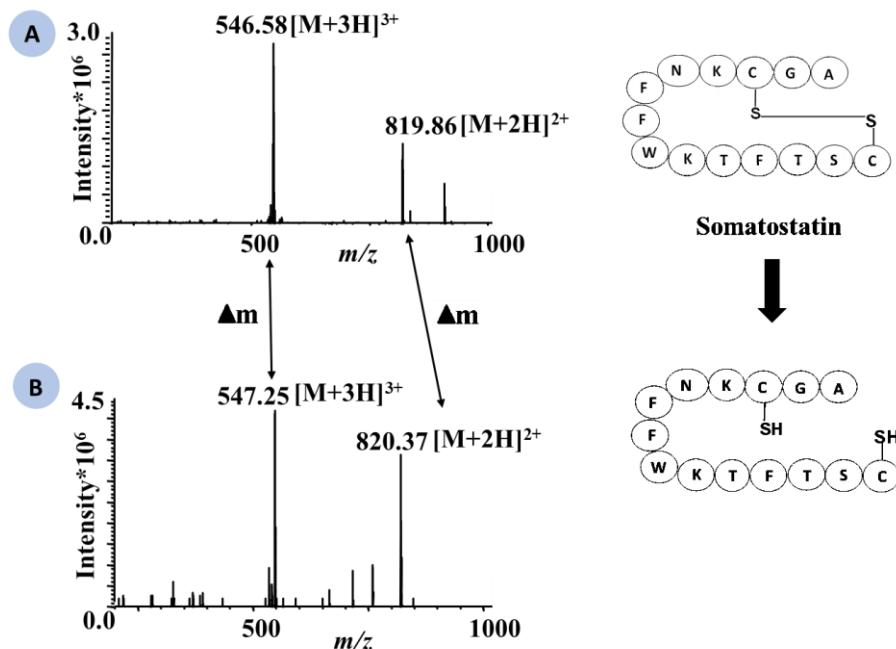


Figure 3-4 Identification of somatostatin disulfide bond reduction reaction by TCEP in ESI-IT-TOF-MS. (a) pure somatostatin (b) reduced somatostatin (addition of two hydrogen can be seen from mass difference compare to pure sample.)

Peptide spiked myoglobin digest and enriched labeled peptides showed the same fragmentation pattern in CID-MS/MS which was observed for a pure sample. We can compare Figure 3-6 (A) and (B) for full MS and MS/MS for further confidence of the reaction. Spiked and enriched samples showed signature mass in full MS scan for $[M+2H]^{2+}$, which match with 1045.46 and 697.30 (calculated mass of $[M+2H]^{2+}$ and

[M+3H]³⁺ modified peptides) respectively. Fragments such as b₈³⁺, y₇, y₈, b₇, b₁₁³⁺ show consistent patterns in CID-MS/MS spectra of pure reaction, spiked as well as enriched sample.

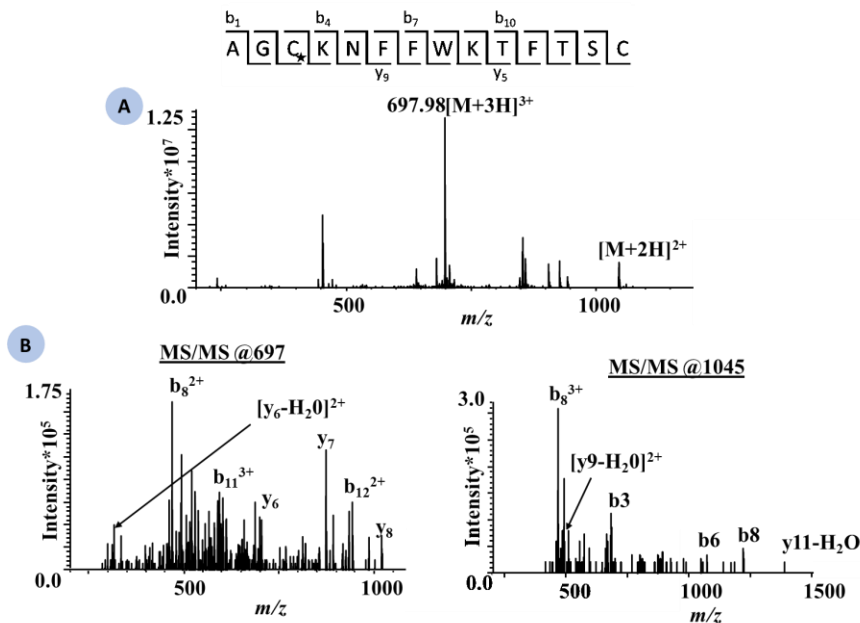


Figure 3-5 LC-MS/MS spectra of biotin-maleimide labeled somatostatin peptide in ESI-IT-TOF-MS. (A) Full MS (697.98 represents [M+2H]³⁺ of modified somatostatin) (B) MS/MS at 697.98 *m/z* and 1045.00 *m/z* (b and y ions from modified peptide backbone under CID MS/MS).

3.2 Bridging characteristics of 2,3-dibromo maleimide and oxidation

After a successful reaction with biotin-maleimide we proceeded to the 2,3-dibromomaleimide reactions. In this venture, RDGC (4 amino acids containing peptide), somatostatin (14 amino acids and one disulfide bond containing peptide) and lysozyme (129 amino acids and four disulfide bonds containing protein) were selected for modification reaction of disulfide bonds using 2,3-dibromomaleimide. Firstly, RGDC

(monoisotopic mass 450.18) was modified using 2,3-dibromomaleimide reagent. We expected to see a sulfhydryl modification of RGDC with mass addition of 174.96 Da. The theoretical monoisotopic peak of chemically labeled RGDC peptide was 623.09, with the mass addition of 174.96 Da from 2,3-dibromomaleimide group. Figure 3-7 shows the intense peaks of 2,3-dibromomaleimide labeled RGDC peptide at 623.08 [M+H]⁺. As bromine has two isotopes, we expected isotopic peaks in full MS spectrum. Figure 3-7(A) (inset) shows isotopic peaks for DBM labeled RDGC at 623.0855 and 625.0846 which further demonstrates cysteine modification using 2,3-dibromomaleimide. Peaks labeled in b₃-NH₃, b₃, RGDC-SH belong to the modified RGDC sequence and indicate that cysteine has been modified using 2,3-dibromomaleimide. Myoglobin digest was spiked with RGDC to validate reactivity in complex sample. Figure 3-7 (B) supplements our hypothesis for 2,3-dibromomaleimide reaction.

After a successful reaction with RGDC our experimental protocol was directed to one small disulfide bond containing peptide, somatostatin was taken to check bridge product formation after disulfide bond reduction and reaction with 2,3-dibromomaleimide. A literature review⁵⁹ reveals that the next generation maleimide has one important property of bridging characteristics that can stabilize the protein structure and decrease the chance of disulfide bonds scrambling. Double and triple charged ion with *m/z* 866.86 and 578.24 were detected, corresponding to the expected *m/z* 867.36 and 578.57 for the somatostatin peptide respectively. The identity of the peptide was slightly challenging to authorize as it forms a bridged product. Lioe *et al.*⁶⁵ noted that disulfide bond cleavage needs higher energy which means it remains intact during CID-MS/MS. As 2,3-dibromomaleimide bridges the two cysteines the way it was bonded by disulfide bonds, product authorization (mentioned above) would be difficult through CID-MS/MS. Although, certain peaks did not match with the intended b and y fragments. Figure 3-8 shows full MS and CID- MS/MS of

2,3-dibromomaleimide modified somatostatin peptide as well as somatostatin spiked in myoglobin digest sample.

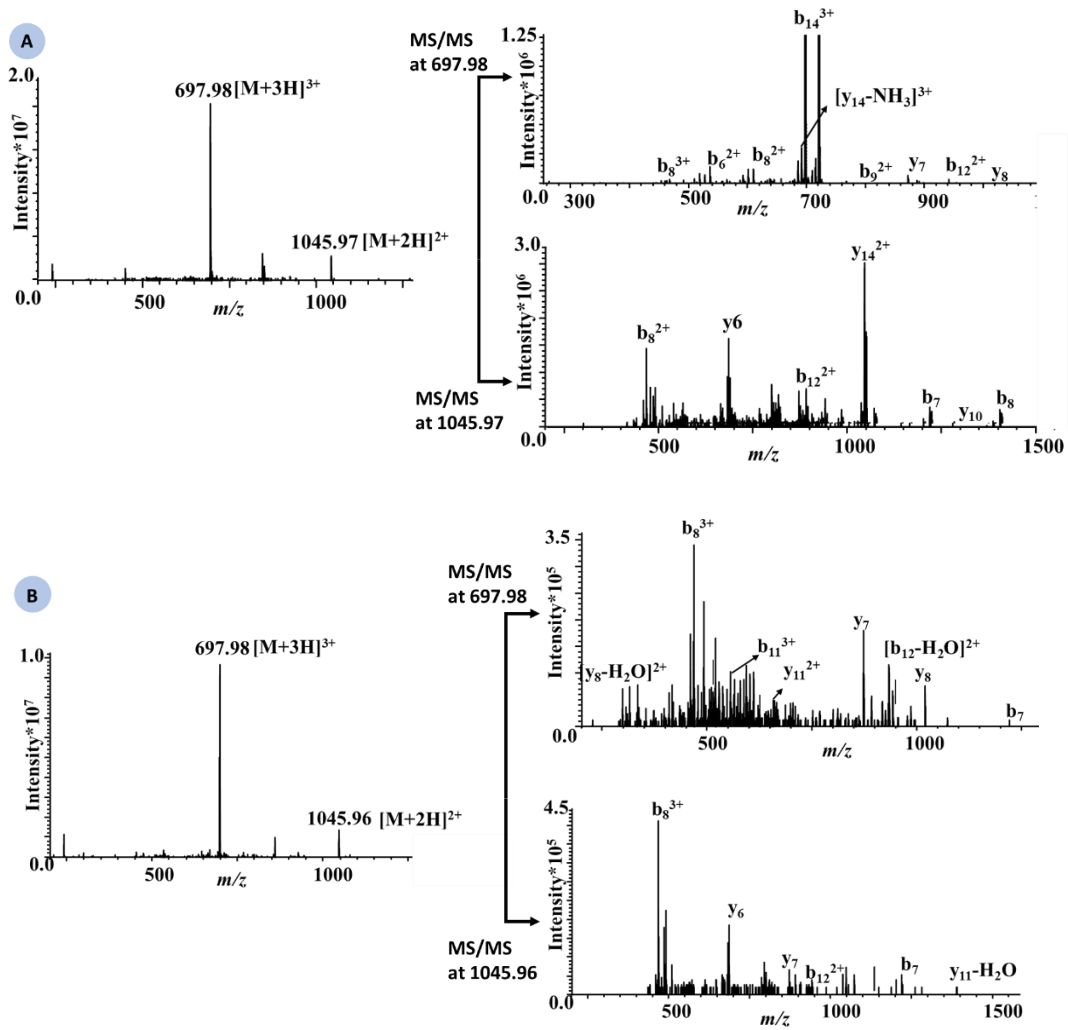


Figure 3-6 : LC-MS/MS spectra of biotin-maleimide labeled somatostatin peptide in ESI-IT-TOF-MS. (A) Full MS (697.98 represents [M+2H]³⁺ of modified somatostatin in myoglobin digest) and MS/MS at 697.98 *m/z* and 1045.97 *m/z* (b and y ions from peptide backbone) (Spiked sample) (B) Full MS and MS/MS at 697.98 *m/z* and 1045.96 *m/z* (enriched sample)

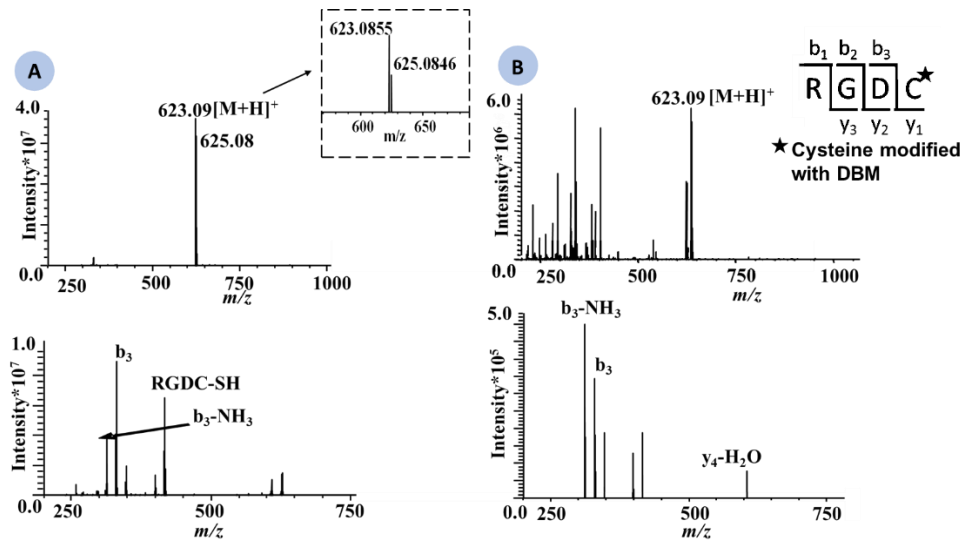


Figure 3-7 LC-MS/MS spectra of 2,3-dibromomaleimide labeled RGDC peptide in ESI-IT-TOF-MS. (A) Full MS and MS/MS at 623.09 m/z (B) Full MS and MS/MS at 623.09 m/z of RGDC spiked in myoglobin digest sample

Once the 2,3-dibromomaleimide reaction with somatostatin was identified, efforts were made to oxidize the thioether bonds in order to get signature mass loss. Oxidation of modified peptide was conducted using 3% H_2O_2 to oxidize sulfur atom of cysteine residue. It can be found from the literature survey that MS/MS fragmentation pattern of mono-oxidized thioether bonds containing peptide produces a loss of characteristic fragment known as RSOH (sulfenic group) from peptides⁶¹. This phenomenon motivates us to design a new strategy of identification and characterization of disulfide bonds using the signature mass loss from the precursor peptides. We started our next step in the protocol of oxidation using 3% H_2O_2 reaction with modified peptide for 15 min. The expected mass from different number of oxidations can be found in Table 3-1 given below. Since, it has two cysteines in the sequence, a maximum of four oxidations are possible.

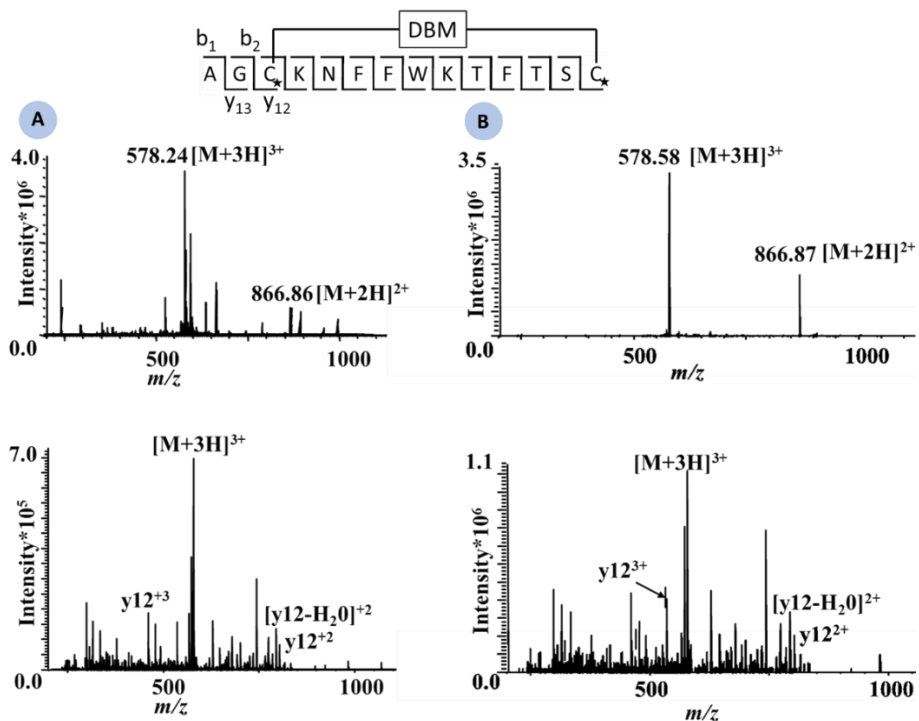


Figure 3-8 LC-MS/MS spectra of 2,3-dibromomaleimide labeled somatostatin peptide in ESI-IT-TOF-MS. (A) Full MS and MS/MS 578.24 m/z (B) Full MS and MS/MS at 578.58 m/z of somatostatin spiked myoglobin digest sample

Formation of oxidation product was confirmed by comparing Figure 3-9 with Table 3-1. It can be seen from the Figure 3-9 (A) that 891.84 $[M+2H]^{2+}$ and 594.89 $[M+3H]^{3+}$ were detected corresponding to 891.35 m/z and 594.57 m/z from the three oxygen addition at cysteine residue ($M+3O$). The peak at 891.84 shows 24.38 Da mass difference whereas the one at 594.89 shows 16.65 Da mass difference further confirming the addition of 3 oxygen atoms, i.e. 3×15.99 (oxygen mass) = 47.97 Da. Thiol-ether bond cleavage was also confirmed by matching peptide mass after $RS_2O_3H_2$ loss, where R is equal to maleimide group, in MS^2 spectra. The peak at m/z 787.37 in MS^2 was formed (Figure 3-9(B)) after losing $RS_2O_3H_2$ from modified somatostatin. It further matches with calculated double

charge m/z 786.88 of peptide after losing $RS_2O_3H_2$. It also confirmed the selectivity of oxidation reaction at the thioether site. This data clearly shows that oxidation at thioether bonds reduces some complexity in disulfide bond mapping. Spiking study was done to validate our method. Myoglobin digest was spiked with somatostatin and reacted with 2,3-dibromomaleimide to get the consistent result with un-spiked sample. Figure 3-10 shows the results from oxidation study in spiked sample. Furthermore, 594.89 m/z and 891.84 m/z detected in sample correspond to the three expected oxidations mentioned in the Table 3-1 and mass after $RS_2O_3H_2$ loss at m/z 786.88 above.

Table 3-1 Theoretical calculated mass (m/z) for different number of oxygen addition at sulfur atom in thiol group of cysteine modified by 2,3-dibromomaleimide

	Somatostatin With DBM	One oxidation	Two oxidations	Three oxidations	Four oxidations
Total mass	1732.78	1748.72	1764.72	1780.71	1796.71
[M+2H]²⁺	867.36	875.36	883.36	891.35	899.35
[M+3H]³⁺	578.57	583.90	589.24	594.57	599.90

As undigested somatostatin reaction was difficult for confident identification in MS/MS, we moved forward with somatostatin digest reaction with 2,3-dibromomaleimide. Tryptic digestion of somatostatin produced three peptide fragments such as N-F-F-W-K, A-G-C-K, and T-F-T-S-C. A-G-C-K (fragment 1) and T-F-T-S-C (fragment 2), that are connected by disulfide bond as they contain cysteine in sequence. Theoretical monoisotopic peak of chemically labeled somatostatin fragments was 1029.38, with the mass addition of 95.06 Da from 2,3-dibromomaleimide group. A triply charged highest intense peak with m/z 343.46 was detected in a somatostatin digest corresponding to the

predicted m/z 344.13 for linked peptide fragments (Figure 3-11(B)). Other peaks from $[M+2H]^{2+}$ and $[M+H]^+$ also agree with expected m/z 515.69 and 1029.38, respectively. Identity of connected fragments was further confirmed by CID-MS/MS. X-link explorer software was used to get the detailed information about b and y ions of connected peptides. Identity of connected peptides was confirmed as it contains $[\text{pept } y_3\text{-H}_2\text{O}]^{2+}$, $[y_3 \text{ pept-H}_2\text{O}]^{2+}$, $\text{pept } y_2$, ions at m/z 381.63, 475.17, 679.22 and 249.12 respectively. All these ions also represented modified and bridged cysteine in A-G-C-K and T-F-T-S-C peptide fragments in somatostatin digest. Third fragment which does not have cysteine can be seen in Figure 3-11(A) without any modification and m/z 741.37 matches exactly with expected m/z 741.86.

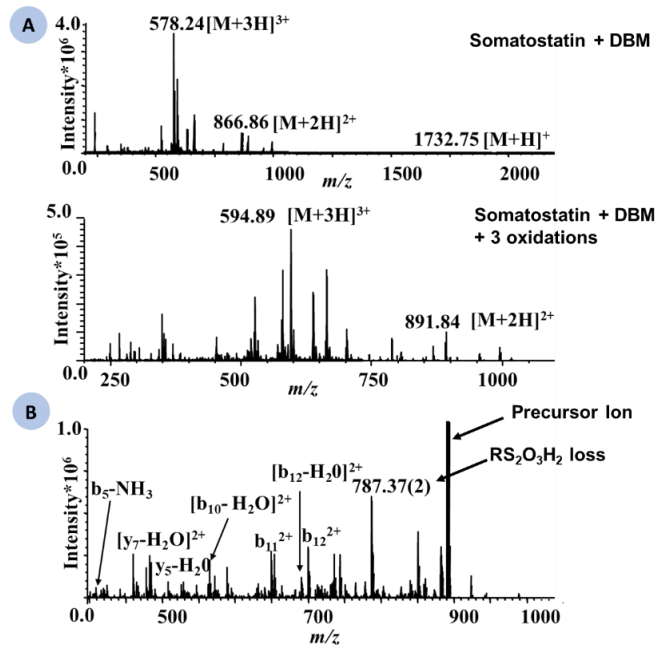


Figure 3-9 LC-MS/MS spectra of oxidized somatostatin peptide using 3% H₂O₂ in ESI-IT-TOF-MS. (A) Oxidized somatostatin with mass difference compare to DBM modified peptide (B) MS/MS at 891.84 with characteristic RS₂O₃H₂ loss and b and y ions from remaining peptide backbone

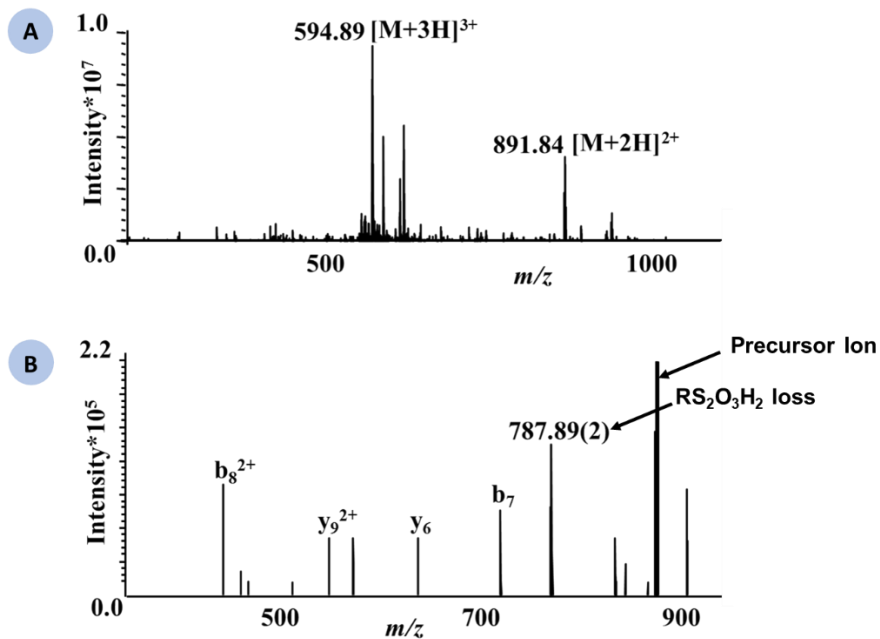


Figure 3-10 LC-MS/MS spectra of oxidized somatostatin peptide, spiked in myoglobin digest, using 3% H_2O_2 in ESI-IT-TOF-MS. (A) $[M+2H]^{2+}$ and $[M+3H]^{3+}$ of oxidized somatostatin in myoglobin digest (B) MS/MS at 891.84 with characteristic $RS_2O_3H_2$ loss and b and y ions from remaining peptide backbone

After productive reaction, we moved to oxidation step in the planned protocol. Again 3% H_2O_2 was used to oxidized sulfur atom of cysteine residue and get a loss of characteristic fragment known as RSOH (sulfenic group) from peptides. Expected mass from different number of oxidations can be found in Table 3-2 given below. Since, it has two cysteines in two peptide sequences, one cysteine in each sequence, a maximum of four oxidations are possible.

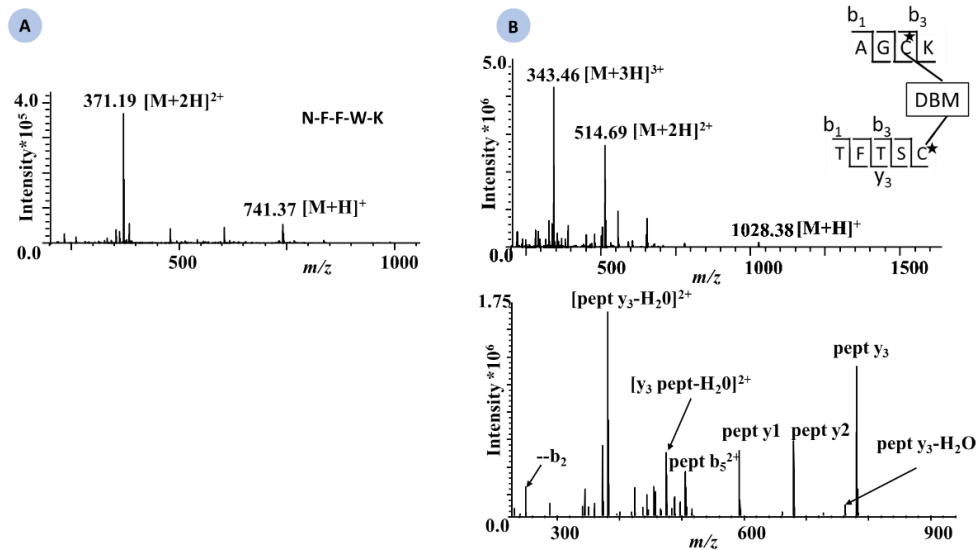


Figure 3-11 LC-MS/MS spectra of 2,3-dibromomaleimide labeled somatostatin peptide in ESI-IT-TOF-MS. (A) additional proof of unreacted peptide fragment (NFFWK) in full MS (B) Full MS and MS/MS at 343.46 m/z of 2,3-dibromomaleimide labeled peptide fragments from somatostatin digest

Formation of oxidation product was confirmed by comparing Figure 3-12 with Table 3-2. It can be seen from the Figure 3-12(B) that 524.17 $[M+2H]^{2+}$ and 349.78 $[M+3H]^{3+}$ were detected corresponding to 523.69 m/z and 349.46 m/z forming mono oxidation (M+O). The peak at 524.17 shows 9.4 Da (+2, charge states) mass difference whereas peak at 349.78 shows 6.3 Da (+3, charge states) mass difference further confirming the addition of one oxygen atom. Thiol-ether bond cleavage was also confirmed by matching peptide mass after RSOH loss, where R is equal to peptide fragment 2 connected with maleimide group, in MS² spectra. The peak at m/z 344.18 in MS² was formed (Figure 3-12(B)) after losing RSOH and matched with expected m/z 344.19. In next step, myoglobin digest was spiked with somatostatin digest and oxidation reaction was performed in the same way it was done for previous sample using 3% H₂O₂ to validate the method in complex sample It shows the

consistent pattern in MS and CID- MS/MS (Figure 3-12(C)) with previous sample (i.e. without spiked study).

Table 3-2 Theoretical calculated mass (m/z) for different number of oxygen addition at sulfur atom of cysteine containing peptide fragments of somatostatin digest modified by 2,3-dibromomaleimide

	Two fragments With DBM	One oxidation	Two oxidations	Three oxidations	Four oxidations
Total mass	1029.38	1045.38	1061.38	1077.37	1093.37
[M+2H]²⁺	515.69	523.69	531.69	539.69	547.69
[M+3H]³⁺	344.13	349.46	354.79	360.12	365.46

Lysozyme was selected for protein level study. The source of our lysozyme sample was from chicken egg white. It contains eight cysteines in the sequence that are connected by four disulfide bonds with total of 129 amino acids residues. Moreover, it does not contain any free cysteine in sequence. After trypsin digestion, it contains three interchain and one intrachain disulfide bond in sample. There are various possibilities of peptide sequence as trypsin can cut protein sequence after lysine as well as arginine. Most possible peptide sequences after trypsin digestion (without reduction) are listed in Table 3-3. As per our hypothesis, if the very first interchain listed in Table 3-3 is modified by DBM, monoisotopic peak at 1375.69 m/z should be observed. Although, after analyzing the data with Stavrox (open source software tool), some inconsistencies in modified peptides were seen. As seen from Figure 3-13 peptide fragment 1 (CELAAAMK) is bonded with peptide fragment 2 (GYSLGNWCAAK). Theoretical monoisotopic peak of chemically labeled lysozyme fragments was 2197.99, with the mass addition of 95.06 Da from 2,3-dibromomaleimide

group. A triply charged peak with m/z 733.68 was detected in lysozyme digest corresponding to the predicted m/z of 733.66 for linked peptide fragments (Figure 3-13). Second detected peptides were NLC76NIPC80SALLSSDITASVNC94AK. In theory, there is one disulfide bond between C76 and C94, while C80 is attached to C64. In our results C76 and C80 were bonded through 2,3-dibromomaleimide. Theoretical monoisotopic peak of chemically labeled C76-C80 was 2430.11, with the mass addition of 95.06 Da from 2,3-dibromomaleimide group. A doubly charged peak with m/z of 1216.54 was detected in lysozyme digest corresponding to the predicted m/z of 1216.05 for linked peptide fragments. It can be seen from Figure 3-14 that y_{21}^{2+} , y_{22}^{2+} , y_6 and y_7 were present in tandem MS data which further confirms the connection between C76 and C80. These results were not consistent with reported disulfide bond pattern for lysozyme. One probable reason for disulfide bond scrambling is pH condition of reaction. Although our results are not entirely in accordance with our expectation, we did observe C6-C30 disulfide scrambling which is very common with lysozyme protein. Chang and Li⁶⁶ have shown three scrambling patterns where C6-C30 and C115-C127 remained almost consistent while connecting pattern of C64, C76, C80 and C94 were unpredictable. In our findings, C76-C80 was repeatedly observed as described earlier. We expect that the primary reason for the uncertainty is the close proximity of these disulfide bond sites.

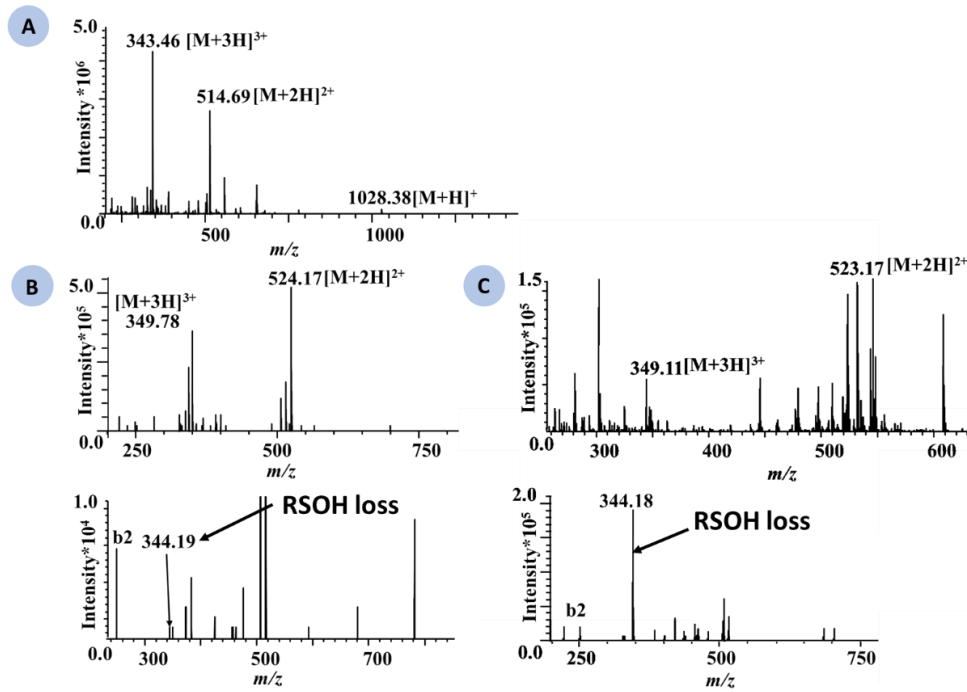


Figure 3-12 LC-MS/MS spectra of oxidized somatostatin peptide using 3% H_2O_2 in ESI-IT-TOF-MS. (A) 2,3-dibromomaleimide bridged A-G-C-K and T-F-T-S-C fragments from somatostatin digest (B) Oxidized 2,3-dibromomaleimide bridged A-G-C-K and T-F-T-S-C fragments from somatostatin digest and MS/MS at 524.17 with characteristic RSOH loss and b and y ions from remaining peptide backbone. (C) Proof of oxidation study of somatostatin digest in myoglobin digest sample with characteristic RSOH loss and b and y ions from remaining peptide backbone

Table 3-3 List of possible disulfide bonded lysozyme peptides after trypsin digestion

Peptide	Structure
Interchain disulfide	C6ELAAAMK GC127RL
Interchain disulfide	GYSLGNWVC30AAK NRC115K
Inter-intrachain disulfide	NLC76NIPC80SALLSSDITASVNC94AK WWC64NDGR

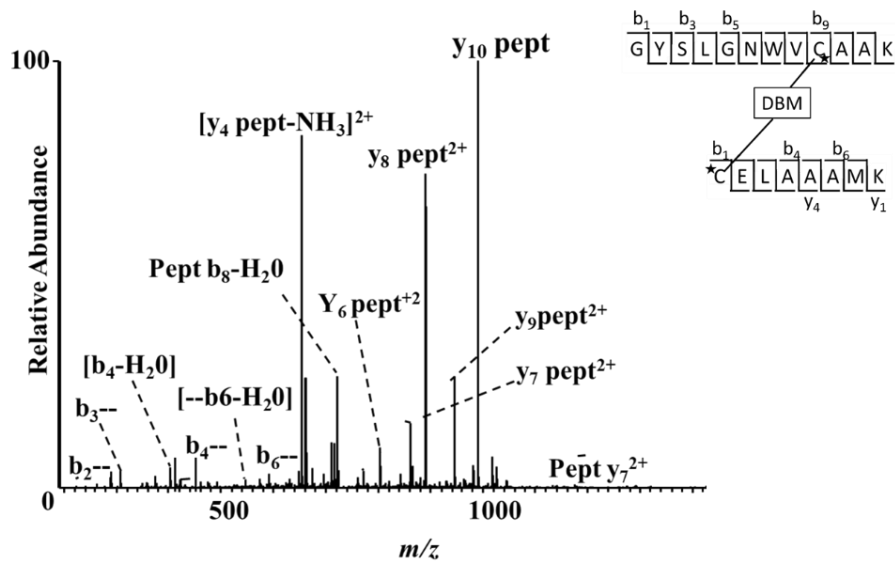


Figure 3-13 LC-MS/MS spectra of 2,3-dibromomaleimide labeled lysozyme digest peptides in ESI-LIT-MS. MS/MS at 733.93 m/z of peptide fragment (CELAAAMK) bonded with (GYSLGNWCAAK)

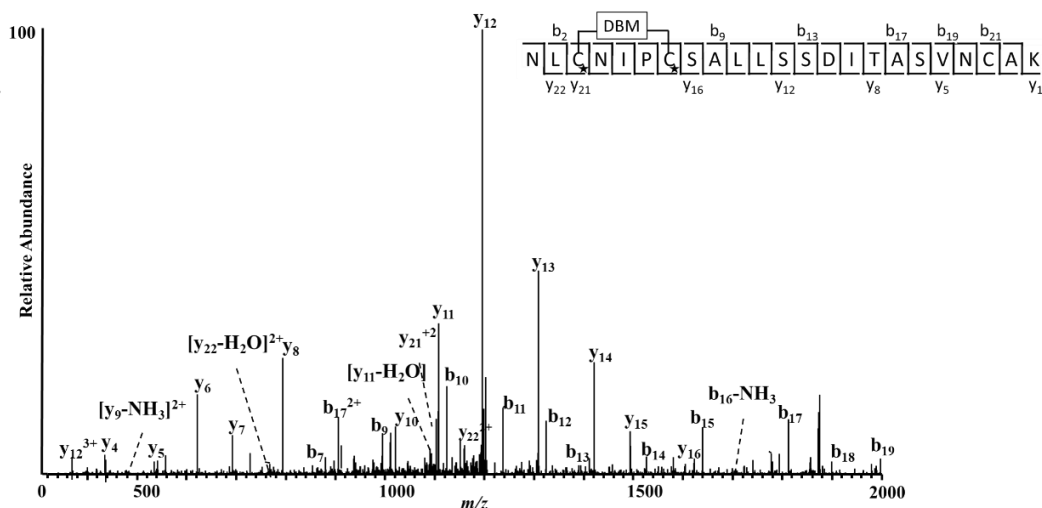


Figure 3-14 LC–MS/MS spectra of 2,3-dibromomaleimide labeled lysozyme digest peptides in ESI-LIT-MS. MS/MS at 1216.54 m/z of peptide fragment
NLC76NIPC80SALLSSDITASVNC94AK

3.3 Bridging and enrichment characteristics of biotin dibromomaleimide and oxidation

BDBM was prepared using the method described in Chapter 2. DBM and BM reaction helped us to select the reaction pH. Firstly, PBS (pH=7.4) was selected for reaction medium but after unfruitful reaction we chose to use PBS (pH=6.2) for all BDBM reaction reactions with successful products. As noted earlier BDM mass is 481.16, although we always saw our products with 16 mass addition more than expected mass. Kowalski *et al.*⁶⁷ reported oxidation of sulfides to sulfoxides in halogen mediated medium. As bromine acts as a leaving group in our reaction, oxidation can be the reason behind 16 mass increment. Experiment was started with small peptide (i.e. RGDC) which contains only one cysteine in sequence (R-G-D-C). RGDC (monoisotopic mass 450.18) was chemically labeled using biotin dibromomaleimide. The theoretical monoisotopic peak of chemically labeled RGDC peptide was 865.16 m/z , with the mass addition of 417.26 Da of biotin dibromomaleimide group. Figure 3-15 shows the intense peaks of biotin-maleimide labeled

RGDC peptide at 863.36 [M+H]⁺. As bromine has two isotopes, doublet peaks can also be seen in zoomed spectra for further confirmation. 861.36 *m/z* and 863.36 *m/z* are the peaks related with bromine isotopes. The peptide peak at 863.36 [M+H]⁺ was isolated and fragmented using CID in nano ESI-LIT-MS. Few b and y ions (from peptide backbone) were observed in tandem MS/MS analysis. Figure 3-15(right) shows the b and y ions from biotin dibromomaleimide labeled RGDC peptide. MS/MS figure shows very few fragments of b and y ions.

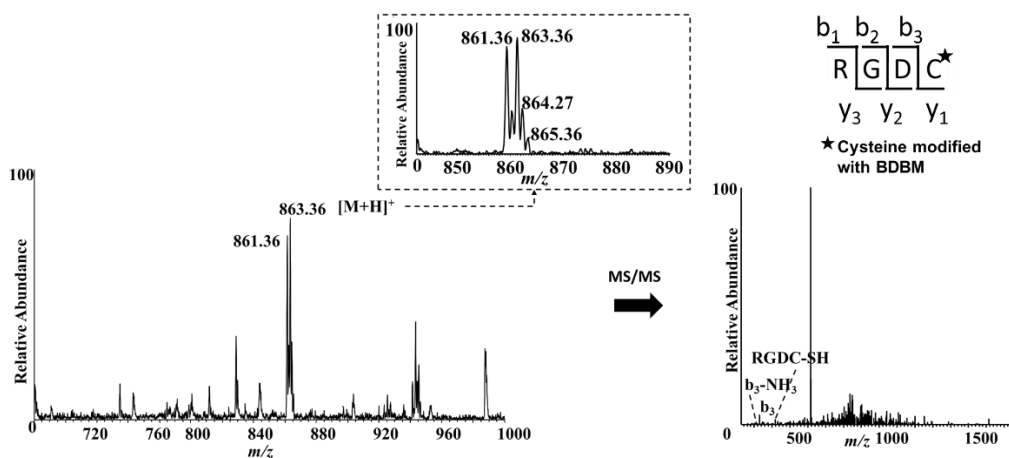


Figure 3-15 LC-MS/MS spectra of biotin dibromomaleimide labeled RGDC peptide in nanoESI-LIT-MS. Full MS with zoomed isotopic peaks (inset) (left). MS/MS at 863.36 *m/z* (right)

After checking reactivity of BDBM, somatostatin was selected for next level study. As somatostatin contains one disulfide bond, it enables us to check bridging property of dibromo group. Experiment was started with somatostatin (monoisotopic mass 1637.73) reduction which gives a linear peptide (monoisotopic mass 1639.74) without any disulfide bond for further reaction. In Figure 3-16(A), an intense peak at *m/z* 987.91 was detected that corresponds to [M+2H]²⁺ of biotin dibromomaleimide modified somatostatin.

658.91 m/z and 1974.73 m/z were also discovered in mass spectra that correspond to $[M+3H]^{3+}$ and $[M+H]^+$ respectively. Furthermore, all peaks agreed very well with the theoretically calculated 988.40 $[M+2H]^{2+}$ and 659.26 $[M+3H]^{3+}$ peaks of biotin dibromomaleimide labeled somatostatin peptide. As mentioned earlier, product authorization using MS/MS spectra of biotin dibromomaleimide bridged product is challenging due to limited b and y fragment formation in CID-MS/MS. Although, there are some b and y ions present in MS/MS spectrum matching with calculated fragments. y_{12} , y_{13} , $[y_{14}-H_2O]^{2+}$, $y_{13}-H_2O$ and y_{14} are some fragments that are identified in MS/MS spectrum to further confirm reaction workability (Figure 3-16(A)).

Enrichment efficiency of biotin group was tested using myoglobin spiked sample. Affinity chromatography was used to selectively enrich biotin dibromomaleimide labeled sample from complex mixture (here myoglobin digest). Figure 3-16(B) shows intense peak of 987.46 m/z $[M+2H]^{2+}$ and 658.69 m/z $[M+3H]^{3+}$ from myoglobin digest spiked with somatostatin sample. Moreover, these peaks correspond to, 988.40 $[M+2H]^{2+}$ and 659.26 $[M+3H]^{3+}$, theoretical peaks of biotin dibromomaleimide labeled somatostatin peptide. There were some difficulties to analyze MS/MS spectrum but we were able to identify certain fragments shown in Figure 3-16(B). Unmodified tryptic digest peaks from myoglobin was also found during data analysis which shows the specificity of biotin dibromomaleimide towards thiol group of cysteine residue. Enrichment reaction was done followed by spiked study. Theoretical monoisotopic peak of chemically labeled somatostatin peptide was 988.40 $[M+2H]^{2+}$ and 659.26 $[M+3H]^{3+}$, with the mass addition of 337.36 Da of biotin dibromomaleimide group. Figure 3-16(C) shows the intense peaks of enriched biotin dibromomaleimide labeled somatostatin peptide at 987.36 $[M+2H]^{+2}$ and 658.55 $[M+3H]^{+3}$.

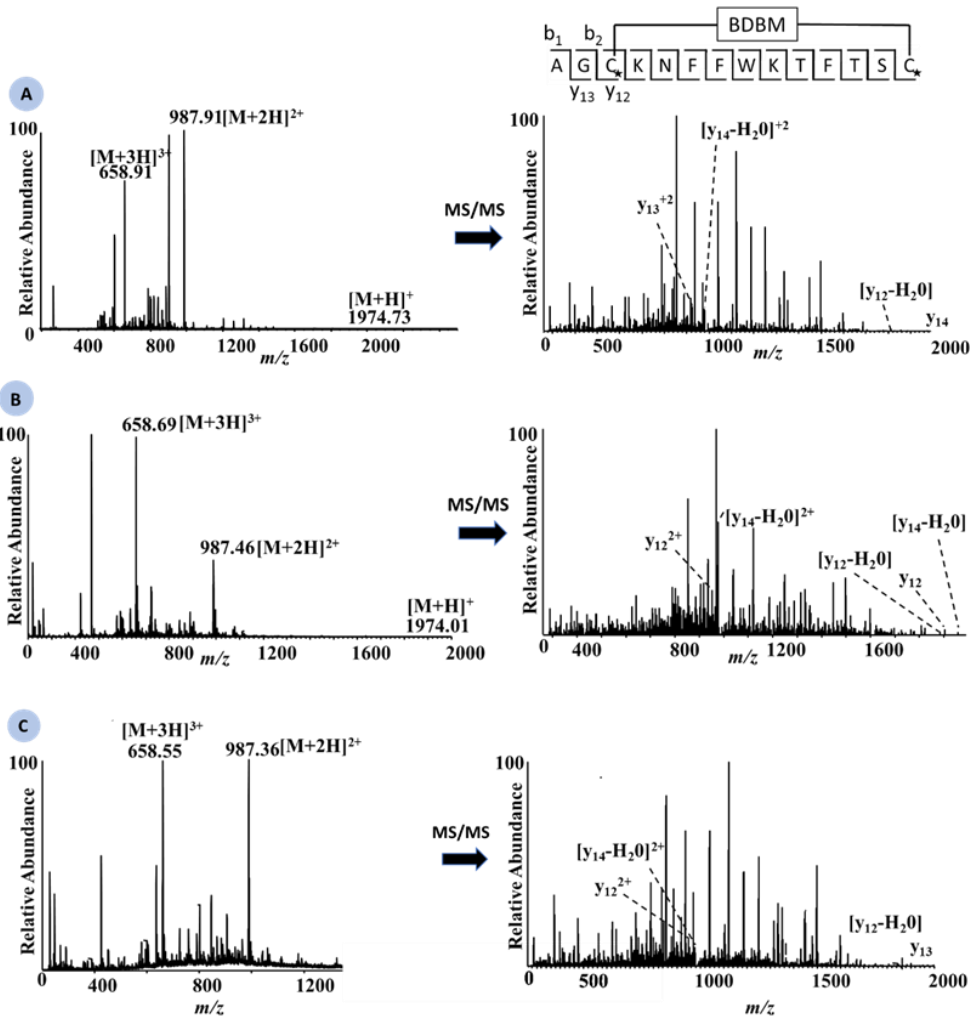


Figure 3-16 : LC-MS/MS spectra of biotin dibromomaleimide labeled somatostatin peptide in nano ESI-LIT-MS. (A) Full MS and MS/MS at 987.93 m/z in pure peptide sample (B) Full MS and MS/MS at 987.46 m/z of somatostatin spiked myoglobin digest sample. (C) Full MS and MS/MS at 987.36 m/z of enriched biotin dibromomaleimide labeled somatostatin

As MS/MS proved difficult for confidence identification of chemically labeled peptide, we moved forward towards the oxidation step in our protocol. 3% H₂O₂ was used to oxidized sulfur atom of cysteine residue and get a loss of characteristic fragment known as RSOH (sulfonic group) from peptides. Expected mass from different number of oxidations can be found in Table 3-4. Since, it has two cysteines in peptide sequence a maximum of four oxidations are possible.

Table 3-4 Theoretical calculated mass (*m/z*) for different number of oxygen addition at sulfur atom in thiol group of cysteine modified by biotin dibromomaleimide

	Somatostatin With BDBM	One oxidation	Two oxidations	Three oxidations	Four oxidations
Total mass	1974.79	1990.79	2006.79	2022.78	2038.78
[M+2H]²⁺	988.40	996.40	1004.40	1012.39	1020.39
[M+3H]³⁺	659.26	664.6	669.93	675.26	680.59

Oxidation experiment data is shown in Figure 3-17. Peaks at 1012.55 [M+2H]²⁺ and 676.82 [M+3H]³⁺ in full MS scan gives a confirmation for three oxidation at cysteine residue from Figure 3-17 with Table 3-4. Results from oxidation show some contradiction with our predicted theory. We expected intense peaks, as we established for dibromomaleimide oxidation experiment, of characteristic RS₂O₃H₂ loss in MS/MS spectrum. Characteristic loss is vital in our method to get confirmed identification of disulfide bridging in protein or peptide level. However, MS/MS (not shown) was not very useful, due to lack of signature fragment loss peak, to validate the identity using RS₂O₃H₂ loss. It showed a peak at 784.48 *m/z* that is not very intense as well as 2 Da off from theoretical monoisotopic peak at 786.88 *m/z* which gives another reason to reevaluate the

success of oxidation reaction. However, b any y ions from peptide backbone fragmentation cover majority of the peaks in CID-MS/MS spectrum.

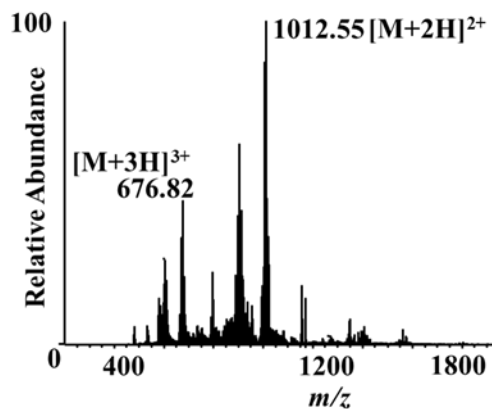


Figure 3-17 LC-MS/MS spectra of oxidized somatostatin peptide using 3% H₂O₂ in nanoESI-LIT-MS. Full MS of oxidized somatostatin with double and triple charged ions

Durand *et al.*⁶⁸ reported that interpeptide bridging gives more information in tandem mass spectrometry due to the generation of more distinct mass fragmentations in MS/MS. Keeping this in mind, somatostatin peptide was digested prior to chemical modification reaction. Two cysteines containing, bridged by disulfide bond, peptides were generated. We expected modification of these two peptides after chemical labelling reaction. However, we got the same peptide sequence bonded together by BDBM. Durand *et al.*⁶⁸ reported disulfide bond scrambling problem in disulfide bond identification reaction. It was also reported that pH condition was the prime reason behind disulfide bond scrambling. Although we have used pH 6.4 which is less than pKa of cysteine, we got unexpected modified peptides. T-F-T-S-C (monoisotopic mass 557.22) was the peptide sequence modified by BDBM with expected peak at 1450.78 *m/z*. Figure 3-18 shows BDBM labeled somatostatin peptide fragment's mass at 1449.55 [M+H]⁺ and 725.73 [M+2H]²⁺. MS/MS data of Figure 3-18 (bottom) [M+2H]²⁺ shows consist results with T-F-T-

S-C modification. $[b_5-H_2O\text{ pept}]^{2+}$, $\text{pept } y_3^{3+}$, $b_5\text{ pept}^{2+}$ and $\text{pept } y_4^{2+}$ are some of the identified fragments. Oxidation reaction was not performed, as enrichment reaction of somatostatin digest peptide was unsuccessful. We need to improve our protocol to selectively enrich disulfide bonded peptides. Incubation time, organic solvent strength, bead concentration are several factors needed to be considered for improvement of the protocol in future.

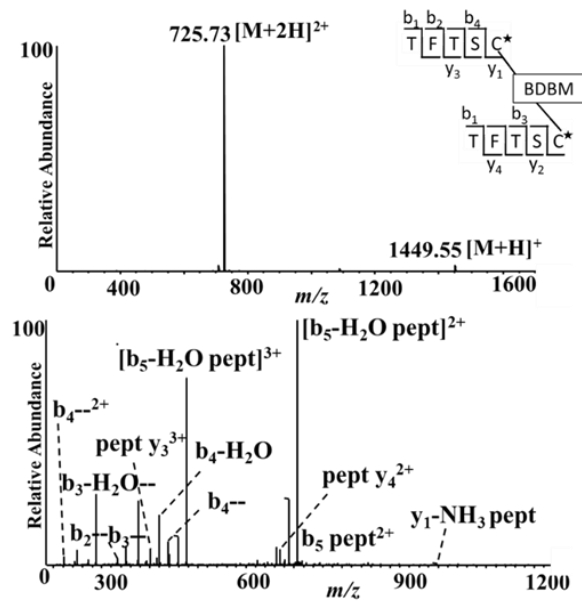


Figure 3-18 LC-MS/MS spectra of biotin dibromomaleimide labeled somatostatin peptides in nanoESI-LIT-MS from somatostatin digest. Full MS of labeled TFTSC fragments (top). MS/MS at 725.73 m/z (bottom)

However, lysozyme reaction with BDBM gave us a successful result before and after enrichment reaction. Chemical modification with BDBM reaction identified two pairs of modified peptides. Lysozyme contains four disulfide bonds that can be seen from Table 3-3. The third disulfide bond listed in the table contains intra as well as inter chain disulfide bonds in sequence. Instead of two modified bonds, we got only one (intrachain) disulfide

modified by BDBM. It also shows consistent result with DBM as C76 and C80 were bonded with BDBM. The theoretical monoisotopic peak of chemically labeled C76-C80 intrachain disulfide (NLC76NIPC80SALLSSDITASVNC94AK) is 2672.43, with the mass addition of 336.37 Da of BDBM group. Figure 3-19 shows MS/MS of 1337.52 m/z which matches with theoretical expected value of 1337.09 m/z . Stavrox was used to identify the modified peptides and X-link explorer was used to get b and y fragments. It must be noted that theoretically C76 and C94 are involved in disulfide bond formation but we saw C76-C80 modification using BDBM.

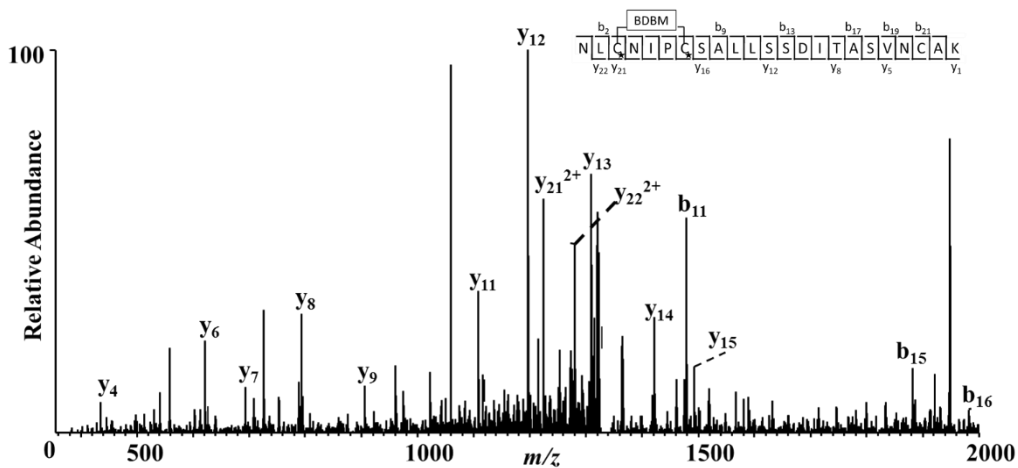


Figure 3-19 LC-MS/MS spectra of biotin dibromomaleimide labeled lysozyme digest peptide in nanoESI-LIT-MS at 1337.52 m/z

Second peptide modified and identified by BDBM in lysozyme digest is listed in Figure 3-20. It (top) shows the intense peaks of modified interchain peptide at 839.09 $[M+4H]^{4+}$, 1118.00 $[M+3H]^{3+}$ and 1676.73 $[M+2H]^{2+}$. All peaks correspond to theoretically calculated $[M+2H]^{2+}$, $[M+3H]^{3+}$, $[M+4H]^{4+}$ at 1675.25, 1117.17, and 838.13 respectively. The peptide peak at 1118.00 $[M+3H]^{3+}$ was isolated and fragmented using CID in nano ESI-LIT-MS. Majority of b and y ions (from peptides backbone) were observed in tandem

MS/MS analysis (Figure 3-20 (bottom)). y_{17} pept, pept y_4^{3+} fragments clearly indicate modified and bridged product after reaction. As listed in Table 3-3 we expected to get CELAAAMK bonded with GCR peptide through BDBM. However, we believe that the major factor behind the modification obtained here is disulfide scrambling. Moreover, it also shows inconsistency with reported lysozyme scrambling pattern, C6-C30.

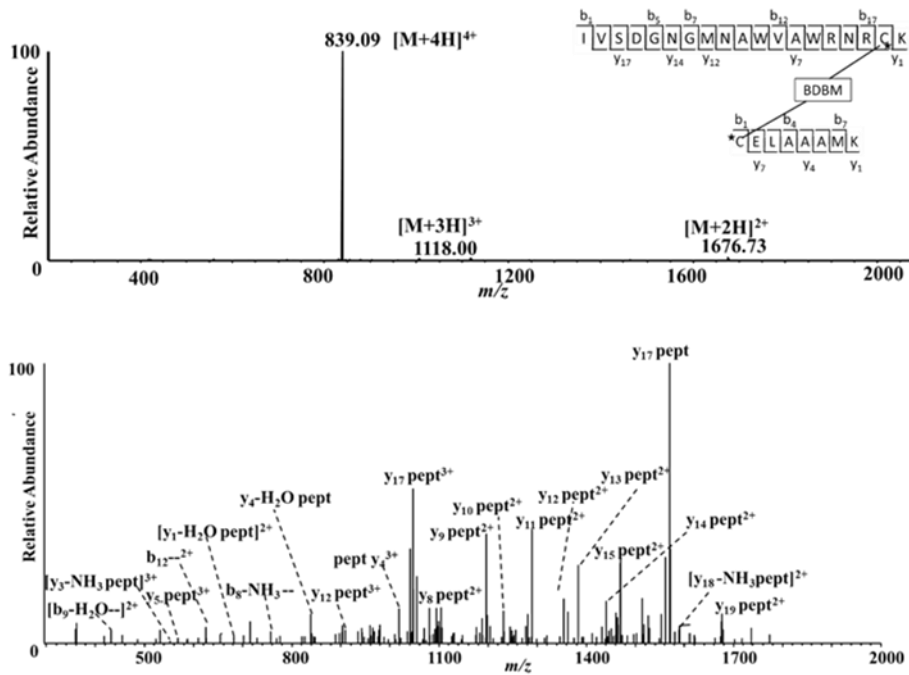


Figure 3-20 LC-MS/MS spectra of biotin dibromomaleimide labeled lysozyme peptides in ESI-LIT-MS from lysozyme digest. Full MS of chemically labeled peptide 1(IVSDGNGMNAWVAVWRNRCK) and peptide 2 (CELAAAMK) (top). MS/MS at 1118.00 m/z (bottom)

In addition, enrichment reaction was done to decrease the complexity and thus increase the confident identification of disulfide bonds. Although results are not so indorsing as only one BDBM bridged peptide was positively enriched form the complex

sample, $[M+6H]^{6+}$ shows the highest intensity peaks with small $[M+3H]^{3+}$ peak at 559.27 and 1118.27 in full mass scan (Figure 3-21(top)) respectively. This mass value resembles peptide 1 (IVSDGNGMNAWVAWRNRCK) and peptide 2 (CELAAMK) modification and conjugation through BDBM. The identity of peptides was further established in CID-MS/MS (Figure 3-21(bottom)). $[y_6 \text{ pept-NH}_3]^{3+}$, $[y_2 \text{ pept-H}_2\text{O}]^{3+}$ and $[b_{7--}]^{+1}$ are few peaks which contain the expected modification that facilitates disulfide bond identification as well as reduce sample complexity.

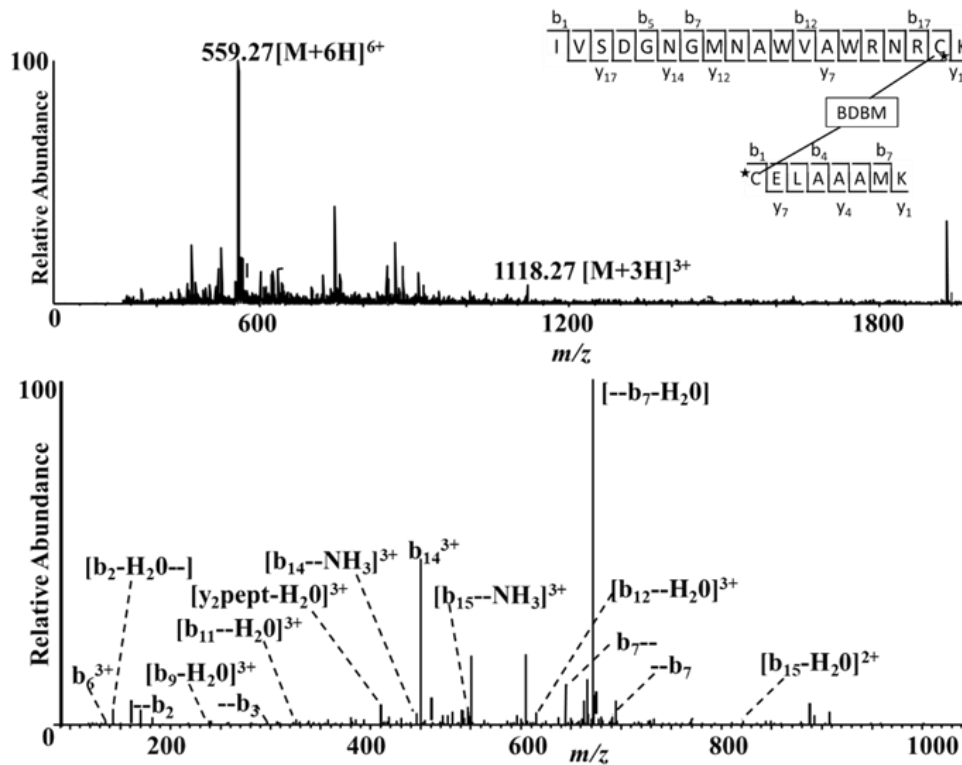


Figure 3-21 LC-MS/MS spectra of biotin dibromomaleimide labeled lysozyme peptides (enriched) in ESI-LIT-MS from lysozyme digest. Full MS of chemically labeled peptide 1 (IVSDGNGMNAWVAWRNRCK) and peptide 2 (CELAAMK) (top). MS/MS at 559.27 m/z (bottom)

Chapter 4

Conclusion

The work presents a novel method for disulfide bond identification using maleimide reagents and sulfur oxidation chemistry. Conventional approaches for disulfide bond identification require specific sample preparation methods and are complicated. Moreover, the results can be ambiguous and data analysis may become extensive. In this study, we demonstrate a generic technique which retains the bridging property of disulfide bonds by using a novel reagent incorporating maleimide group. The biotin maleimide reaction proved the high maleimide reactivity towards thiol group of cysteine residue and effectiveness of biotin group for avidin-biotin affinity purification. Moreover, this thesis, presents an extension of DBM reaction for further confident identification of disulfide bonds using sulfur oxidation chemistry.

BM results show great effectiveness of maleimide reagents towards modification and enrichment of cysteine residue involved in disulfide bond formation. Our study was validated using two small peptides: RGDC and somatostatin. DBM is highly effective in overcoming the disadvantages associated with biotin maleimide reagent. In comparison with commonly available methods, our DBM method offers the following advantages (1) Characteristic RSOH loss minimizes complexity of data analysis (2) Retention of protein structure and stability by mimicking disulfide bond properties. However, it has some challenges arising from the presence of unwanted peptides that need to be addressed such as: (1) interference due to lack of enrichment tag (2) low efficiency of oxidation reaction. BDBM was used and showed motivating results to overcome the difficulties listed above.

However, enrichment and oxidation reaction with BDBM did not seem to be consistent in our findings. Future work in this aspect will be focused on improving the current enrichment protocol. Firstly, the avidin bead volume will be modified along with

organic to aqueous solvent ratio in elution step. Secondly, the precursor ion mass deviation and signal to noise ratio in the data analysis software needs to be tuned to identify more peptides in protein enrichment sample. As mentioned in the results, oxidation of somatostatin digest needs further study while lysozyme oxidation also needs to be done following expected enrichment. Moreover, once the method has been rigorously tested, more complex level study with large protein molecules and palmitoylation on cellular level study will be possible using this novel approach.

Chapter 5

References

1. NIH. A Brief Guide to Genomics, 2015. <https://www.genome.gov/18016863/a-brief-guide-to-genomics/> (accessed June 6, 2017).
2. Collins, F. S.; Lander, E. S.; Rogers, J.; Waterston, R. H. Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931-945.
3. Slack,. In *Genes-A Very Short Introduction*; Oxford University Press: Oxford, 2014.
4. NIH. What is Cancer Proteomics? <https://proteomics.cancer.gov/whatisproteomics> (accessed June 2017, 06).
5. Anderson, L.; Anderson , N.. Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis* **1998**, *19*, 1853-1861.
6. Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198-207.
7. Mann, M.; Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **2003**, *21*, 255 - 261.
8. Walsh, C. T. Proteien translation and posttranslational modification. In *Posttranslational Modification of Proteins: Expanding Nature's Inventory*; Roberts and Company;; Englewood, Colorado, 2006; pp 1-3.
9. Ayoubi, T. A. Y.; Van De, W. J. M. Regulation of gene expression by alternative promoters. *FASEB J.* **1996**, *10* (4), 453-460.
10. Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *WIREs Syst Biol Med.* **2012**, *4* (6), 565-583.
11. Scientific, T. f. Overview of Post-Translational Modifications (PTMs). <https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-post-translational-modification.html> (accessed June 2017, 07).
12. Hogg, P. J. Targeting allosteric disulphide bonds in cancer. *Nat. Rev. Cancer* **2013**, *13*, 425-431.
13. Khoury, G. A.; Baliban, R. C.; Christodoulos , F. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **2011**, *1* (90), 1-5.
14. Bischoffa, R.; Schlüterb, H. Amino acids: Chemistry, functionality and selected non-enzymatic post-translational modifications. *J. Proteomics* **2012**, *75* (8), 2275–2296.
15. Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, Jr., G. J. Protein Posttranslational Modifications: The Chemistryof Proteome Diversifications. *Angew. Chem. Int. Ed.* **2005**, *44*, 7342 – 7372.
16. Swatek, K. N.; Komander, D. Ubiquitin modifications. *Cell Res.* **2016**, *26*, 399–422.
17. Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **2009**, *458*, 422-429.
18. Arnesen , T. Towards a Functional Understanding of Protein N-Terminal Acetylation. *PLOS Biol.* **2011** , *9* (5), 1-5.

19. Starheim, K. K.; Gevaert, K.; Arnesen, T. Protein N-terminal acetyltransferases: when the start matters. *Trends Biochem. Sci.* **2012**, *37* (4), 152-161.
20. Bullied, N. J.; Ellgaard, L. Multiple ways to make disulfides. *Trends Biochem Sci.* **2011**, *36* (9), 485-492.
21. Mcauley, A.; Jacob, J.; Kolvenbach, C.; Westland, K.; Lee, H. J.; Brych, S.; Rehder, D.; Kleemann, G.; Brems, D.; Matsumura, M. Contributions of a disulfide bond to the structure, stability and dimerization of human IgG1 antibody CH3 domain. *Protein Sci.* **2008**, *17*, 95-106.
22. Durand, K. L.; Ma, X.; Plummer, C. E.; Xia, Y. Intra-molecular Reactions between Cysteine Sulfinyl Radical and a Disulfide Bond within Peptide Ions. *Int. J. Mass Spectrom.* **2015**, *378*, 246-254.
23. Nick Pace, C.; Grimsley, G.; Thomson, J.; Barnett, B. Conformational Stability and Activity of Ribonuclease T1 with Zero, One and Two Intact Disulfide Bonds. *J. Bio. Chem.* **1988**, *263* (24), 11820-11825.
24. Denton, M. E.; Scheraga, H. A. Spectroscopic, Immunochemical and Thermodynamic Properties of Carboxymethyl (Cys6, Cys127) - hen egg white lysozyme. *J. Protein Chem.* **1991**, *10* (2), 213-232.
25. Thies, M.; Talamo, F.; Mayer, M.; Bell, S.; Ruoppolo, M.; Marino, G.; Buchner, J. Folding and Oxidation of the Antibody Domain CH3. *J. Mol. Biol.* **2002**, *319*, 1267-1277.
26. Betz, S. F. Disulfide bonds and the stability of globular proteins. *Protein Sci.* **1993**, *2*, 1551-1558.
27. Tsai, C. J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Sci.* **1997**, *6*, 53-64.
28. Winther, J. T.; Jakob, U. Redox Control: A black hole for oxidized glutathione. *Nat. Chem. Biol.* **2013**, *9* (2), 69-70.
29. Ceballos-Picot, I.; Witko-Sarsat, V.; Merad-Boudia, M.; Nguyen, A. T.; Thevenin, M.; Jaudon, M. C.; Zingraff, J.; Verger, C.; Jungers, P.; Descamps-Latscha, B. Glutathione Antioxidant System as a Marker of Oxidative Stress in Chronic Renal Failure. *Free Radic. Biol. Med.* **1996**, *21* (6), 845-853.
30. Chang, S. G.; Choi, K. D.; Jang, S. H.; Shin, H. C. Role of Disulfide Bonds in the Structure and Activity of Human Insulin. *Mol. Cell.* **2003**, *16* (3), 323-330.
31. Ganz, T. Defensin: Antimicrobial Peptides of Innate Immunity. *Nat. Rev. Immunol.* **2003**, *3*, 710-720.
32. Huang, S. Y.; Chen, F. S.; Chen, H. C.; Huang, H. W.; Wu, W. G.; Sung, W. C. Global Disulfide Bond Profiling for Crude Snake Venom Using Dimethyl Labeling Coupled with Mass Spectrometry and Radar Algorithm. *Anal. Chem.* **2014**, *86*, 8742-8750.
33. Safavi Hemami, H.; Qing, L.; Jackson, R.; Song, A. S.; Boomsma, W.; Bandopadhyay, P. K.; Gruber, C. W.; Purcell, A.; Yandell, M.; Olivera, B. M.; Ellgard, L. Rapid expansion of the protein disulfide isomerase gene family facilitates the folding of venom peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113* (12), 3227-3232.
34. Schumacher, F. F.; Nunes, J.; Maruani, A.; Chudasama, V.; Smith, M.; Chester, K.; Baker, J.; Caddick, S. Next generation maleimides enable the controlled assembly of antibody–drug conjugates via native disulfide bond bridging. *Org. Bio. Mol. Chem.* **2014**, *12* (37), 7261-7269.

35. Saito, G.; Swanson, J. A.; Lee, K.-D. Drug delivery strategy utilizing conjugation via reversible disulfide linkages: role and site of cellular reducing activities. *Adv. Drug Delivery Rev.* **2003**, *55*, 199-215.
36. Peterson, J. D.; Nehrlich, S.; Oyer, P. E.; Steiner, D. F. Determination of the amino acid sequence of the monkey, sheep and dog proinsulin C-Peptides by a semi-micro Edman Degradation Procedure. *J. Biol. Chem.* **1972**, *247* (15), 4866-4871.
37. Walewska, A.; Skalicky, J. J.; Davis, D.; Zhang, M.-M.; Estuardo, L.-V.; Watkins, M.; Han, T. S.; Yoshikami, D.; Olivera, B.; Bulaj, G. NMR-Based Mapping of Disulfide Bridges in Cysteine-Rich Peptides: Applications to the u-Conotoxin SxIIIa. *J. Am. Chem. Soc.* **2008**, *130* (43), 14280-14286.
38. Klaus, W.; Broger, C.; Gerber, P.; Senn, H. Determination of the Disulfide Bonding Pattern in Proteins by Local and Global Analysis of Nuclear Magnetic Resonance Data. *J. Mol. Biol.* **1993**, *232*, 897-906.
39. Drenth, J. *Principles of Protein X-ray Crystallography*, 3rd ed.; Springer-Verlag New York, 2007.
40. Gorman, J. J.; Wallis, T.; Pitt, J. Protein Disulfide Bond Determination by Mass Spectrometry. *Mass Spectrom Rev.* **2002**, *21*, 183-216.
41. Barbirz, S.; Jakob, U.; Glocker, M. Mass Spectrometry Unreveals Disulfide Bond Formation as the Mechanism that Activates a Molecular Chaperone. *J. Biol. Chem.* **2000**, *275* (25), 18759-18766.
42. Yang, H.; Liu, N.; Qiu, X.; Liu, S. A new method for analysis of Disulfide-Containing Proteins by Matrix-Assisted Laser Desorption Ionization (MALDI) Mass Spectrometry. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 2284-2293.
43. Peng, Y.; Chen, X.; Sato, T.; Rankin, S.; Tsuji, R.; Ge, Y. Purification and High-Resolution Top-Down Mass Spectrometric Characterization of Human Salivary alpha-Amylase. *Anal. Chem.* **2012**, *84*, 3339-3346.
44. Zhang, Y.; Cui, W.; Zhang, H.; Dewald, H.; Chen, H. Electrochemistry-Assisted Top-Down Characterization of Disulfide Containing Proteins. *Anal. Chem.* **2012**, *84*, 3838-3842.
45. Zhang, H.; Ge, Y. Comprehensive Analysis of Protein Modification by Top-down Mass Spectrometry. *Circ. Cardiovasc. Genet.* **2011**, *4* (6), 711-731.
46. Zhang, W.; Marzilli, L. A.; Rouse, J.; Czupryn, M. Complete disulfide bond assignment of a recombinant immunoglobulin G4 monoclonal antibody. *Anal. Biochem.* **2002**, *311*, 1-9.
47. Zhang, B.; Cockrill, S. Methodology for Determining Disulfide Linkage Patterns of Closely Spaced Cysteine Residues. *Anal. Chem.* **2009**, *81*, 7314-7320.
48. Tsai, P. L.; Chen, S.-F.; Huang, S. Y. Mass spectrometry-based strategies for protein disulfide bond identification. *Rev. Anal. Chem.* **2013**, *32* (4), 257-268.
49. Foley, S. F.; Sun, Y.; Zheng, T.; Wen, D. Picomole-level mapping of protein disulfides by mass spectrometry following partial reduction and alkylation. *Anal. Biochem.* **2008**, *377*, 95-104.
50. Jones, M. D.; Hunt, J.; Liu, J. L.; Patterson, S.; Kohno, T.; Lu, H. Determination of Tumor Necrosis Factor Binding Protein Disulfide Structure: Deviation of the Fourth Domain Structure from TNFR/NGFR Family Cysteine-Rich Signature. *Biochemistry* **1997**, *36*, 14914-14923.

51. Wu, W.; Huang, W.; Qi, J.; Chou, Y.-T.; Torng, E.; Watson, J. T. Signature Sets, Minimal Fragment Sets for Identifying Protein Disulfide Structures with Cyanylation-Based Mass Mapping Methodology. *J. Proteome Res.* **2004**, *3*, 770-777.
52. Yang, Y.; Wu, J.; Watson, T. Disulfide Mass Mapping in Proteins Containing Adjacent Cysteines is Possible with Cyanylation/Cleavage Methodology. *J. Am. Chem. Soc.* **1998**, *120*, 5834-5835.
53. Yu, X.; Khani, A.; Ye, X.; Petruzziello, F.; Gao, H.; Zhang, X.; Rainer, G. High-Efficiency recognition and identification of disulfide bonded peptides in rat neuro peptidome using targeted electron transfer dissociation tandem mass spectrometry. *Anal. Chem.* **2015**, *87*, 11646-11651.
54. Kalia, J.; Raines, R. Advances in Bioconjugation. *Curr. Org. Chem.* **2010**, *14* (2), 138-147.
55. Hull, E.; Livanos, M.; Miranda, E.; Smith, M.; Chester, K.; Baker, J. Homogeneous Bisspecific by Disulfide Bridging. *Bioconjugate Chem.* **2014**, *25*, 1395-1401.
56. Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. Chemical modification of proteins at Cysteine: Opportunities in chemistry and biology. *Chem. Asian J.* **2009**, *4* (5), 630-640.
57. Gunnoo, S. B.; Madder, A. Chemical modification through Cysteine. *ChemBioChem* **2016**, *17* (7), 529-553.
58. Tedaldi, L.; Smith, M.; Nathani, R.; Baker, J. Bromomaleimides: New Reagents for Selective and Reversible Modification of Cysteine. *Chem. Commun.* **2009**, 6583-6585.
59. Smith, M.; Schumacher, F.; Ryan, C.; Tedaldi, L.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. Protein modification, bioconjugation and disulfide bridging using Bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132*, 1960-1965.
60. Glick, D. *Methods of Biochemical Analysis*; John Wiley & Sons: New York, 2009.
61. Chowdhury, S. M.; Munske, G. R.; Ronald, R. C.; Bruce, J. E. Evaluation of low energy CID and ECD fragmentation behavior of mono-oxidized thio-ether bonds in peptides. *J. Am. Soc. Mass. Spectrom.* **2007**, *18*, 493-509.
62. Winther, J.; Thorpe, C. Quantification of Thiols and Disulfides. *Biochim. Biophys. Acta* **2014**, *1840* (2), 838-846.
63. Bayer, E.; Wilchek, M. Application of Avidin-Biotin technology to affinity-based separations. *J. Chromatogr., A* **1999**, *510* (27), 3-11.
64. Cao, L.; Wang, L.; Fan, L.; Xiao, W.; Lin, B.; Xu, Y.; Liang, J.; Cao, B. RGDC peptide-induced biomimetic calcium phosphate coating formed on AZ31 magnesium alloy. *Materials* **2017**, *10* (358), 1-15.
65. Lioe, H.; O'Hair, R. A Novel Salt Bridge Mechanism Highlights the Need for Nonmobile Proton Conditions to Promote Disulfide Bond Cleavage in Protonated Peptides Under Low-Energy Collisional Activation. *J. Am. Soc. Mass. Spectrom.* **2007**, *18*, 1109-1123.
66. Chang, J.-Y.; Li, L. The unfolding mechanism and disulfide structures of denatured lysozyme. *FEBS Lett.* **2002**, *511*, 73-78.
67. Kowalski, P.; Mitka, K.; Ossowska, K.; Kolarska, Z. Oxidation of sulfides to sulfoxides, part 1: oxidation using halogen derivatives. *Tetrahedron* **2005**, *61* (8), 1933-1953.

68. Durnad, K.; Tan, L.; Stinson, C.; Love-Nkansah, C.; Ma, X.; Xia, Y. Assigning peptide disulfide linkage pattern among regio-isomers via methoxy addition to disulfide and tandem mass spectrometry. *J. Am. Soc. Mass. Spectrom.* **2017**, *28*, 1099-1108.

Biographical Information

Pooja Rajeshbhai Ramanuj was born and brought up in Rajkot, India. She earned a Bachelor's and Master's degree in Pharmacy from Saurashtra University in India. She later joined The University of Texas at Arlington, where she decided to pursue a graduate degree in Analytical Chemistry under the supervision of Dr. Saiful M. Chowdhury. Her research interests include, identification of post-translational modifications, protein-protein interactions and method development using mass spectrometry.