IDENTIFICATION OF PROTEOLYTIC CLEAVAGE AND ARGININE MODIFICATIONS BY
CHEMICAL LABELING AND MASS SPECTROMETRY

by

MAHESHIKA S.K. WANIGASEKARA

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

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2017

This dissertation is dedicated to my mother and late father who left us on April 02nd 2001, my husband Dhanuka and my son Mayon for their love and support.

Acknowledgements

First and foremost, I owe sincere and earnest thankfulness to my research advisor, Dr. Saiful Chowdhury. My sincere gratitude and profound thanks goes to you for welcoming me into your research group. I am also grateful for your patient guidance and supervision throughout my graduate career.

I express my sincere thanks to my committee members, Dr. Kevin Schug, Dr. Daniel Armstrong and Dr. Jonguyn Heo, for their valuable feedback during my UTA research career.

I would like to express my deepest appreciation to the Department of Chemistry and Biochemistry and all department staff members for their kind assistance from the beginning to the end of my graduate studies at the University of Texas at Arlington. I was also fortunate to have a group of wonderful friends in my research group, who helped me in many ways during my time at UTA.

I am eternally thankful for my husband for his endless love, and support. None of this could be possible without him. Words cannot express how grateful I am for everything he has done for me. He sacrificed immensely for me so I could obtain my Ph.D.

I am very grateful to my brother, Eranda Wanigasekara, who has helped me accomplish many things in my life. Finally, a special thanks to my sister-in-law, Tharanga, and my husband's family for their enormous support and encouragement throughout the years, and especially during my graduate studies.

Abstract

IDENTIFICATION OF PROTEOLYTIC CLEAVAGE AND ARGININE MODIFICATIONS BY
CHEMICAL LABELING AND MASS SPECTROMETRY

Maheshika S.K. Wanigasekara, PhD

The University of Texas at Arlington, 2017

Supervising Professor: Dr. Saiful M. Chowdhury

Post-translational modification of proteins plays a significant role in the regulation of cellular processes. During the translation, process proteins undergo different modifications, which are known as post-translational modifications (PTMs). Due to their low abundances, PTM analysis presents several challenges; therefore, efficient and sensitive PTM detection methods are required. Arginine is an essential amino acid in a cell, which undergoes several kinds of PTMs. Finding these functional arginine residues in a protein is a challenging task. The task was successfully addressed, and the findings are not only reported herein, but an assessment was recently reported covering the new chemical labeling methods developed for identifying functional arginine residues of proteins by comparing two widely used arginine labeling reagents—1,2-cyclohexanedione (CHD) and phenylglyoxal (PG). This dissertation combined the author's and her research team's previous studies with bio-orthogonal chemistry and quantitative mass spectrometry-based proteomics to develop a new approach for the selective enrichment of reactive arginine residues in proteins in complex samples. The primary achievement was the development of a novel arginine-specific, azide-tagged CHD analog, which enables labeling of reactive arginine residues for further studies. For large-scale samples, the workflow from this labeling process is adaptable for gel-based pre-separation. This research established a promising strategy for the effective profiling of reactive arginine residues in large-scale studies. N-terminal modifications of proteins can interpret the functions and stability of proteins, affecting their expression, activation, or degradation. Moreover, selective enrichment

of N-terminal peptides from a complex mixture is a difficult challenge in the proteomics field. Currently we are working with two novel reagents, which have innovative properties to enrich the N-terminal peptides and to generate marker ions from the N-term-labeled peptides during tandem mass spectrometry. These studies will significantly contribute to proteomics and to further research in the bio-analytical mass spectrometry field.

Table of Contents

Chapter 1

List of Illustrations

List of Tables

List of Abbreviations

ADP	Adenosine diphosphate
AGE	Advanced glycation end products
CID	Collision-induced dissociation
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ESI-IT-TOF	Electrospray ionization-ion trap-time-of-flight
HCD	High-energy collision dissociation
HPLC	High performance liquid chromatography
LC/MS	Liquid chromatography mass spectrometry
LC/MS/MS	Liquid chromatography and mass spectrometry in tandem
MALDI-QIT-TOF	Matrix assisted laser desorption ionization-quadrupole ion trap time of flight
MS/MS	Mass spectrometry in tandem
nano LC-ESI-LIT	Nano electrospray ionization linear ion trap
PBS	Phosphate saline buffer
PG	Phenylglyoxal
PTMs	Post -translational modifications
TCEP	Tris(2-carboxyethyl)phosphine

Chapter 1

Introduction

1.1 Thesis Organization

During the past few decades, mass spectrometry-based methods have become an important technique to identify post-translational modification. Determination of post-translational modifications helps to identify the molecular functions of the proteins. After chemical modification followed by trypsin digestion, the peptide mixture is analyzed by liquid chromatography (LC) with a full-scan tandem (MS/MS) mass spectrometry mode (LC/MS/MS). The mass addition can signify a protein modification in the MS analysis. The peptides generated are then fragmented in the tandem mass spectrometer to yield fragments that can be matched to protein amino acid sequences to characterize the specific peptide. This dissertation is focused on identification of proteolytic cleavage and arginine modifications by chemical labeling and mass spectrometry. The dissertation has a total of four chapters. The first chapter describes the basics of the proteomics and the mass spectrometric field. Chapter 2 focuses on the evaluation of chemical labeling methods for identification of functional arginine residues in proteins by mass spectrometry. The continuation work mentioned in chapter 2 is further discussed in chapter 3 with the newly synthesized arginine reactive enrichment reagent. Chapter 4 traces the development of a novel mass spectrometry- cleavable reagent that enables identification of the proteolytic cleavage approach. Chapter 5 presents a general summary of this work and discusses future directions of the projects featured in this dissertation.

1.2 Post-Translational Modifications

Protein biosynthesis occurs in ribosomes. It begins at the DNA level. After the translation, step amino acid residues in the proteins undergo modifications known as post-translational modifications (PTMs). PTMs refer to the covalent processing of proteins. The term denotes not only the changes of the protein chain due to the addition or removal of chemical moiety to the

amino acid residues but also the proteolytic cleavage, the process of breakdown of proteins into smaller polypeptide chains or amino acids as another type of PTM.[1] PTMs play a crucial role in many intracellular processes such as protein-protein interactions, cell signaling, cell division and many cellular processes, which affect protein stability, structural integrity, metabolism etc.[2] There are more than 200 PTMs that have been identified, which may affect the diverse functions of proteins.[3] The analysis of proteins and their PTMs is particularly important for the study of diseases diagnostics and prevention.[4-6] Table 1-1 shows the most common PTMs, the modified amino acid residue, and mass addition.

Table 1-1 Most common PTMs

Type of PTM	Modified Amino Acid Residue	Mass Addition (Da)
Phosphorylation	Tyr(W),Thr(T),Ser(S)	+80
Methylation	Arg(R), Lys(K)	+14
Oxidation	Met(M)	+16,+32
Acetylation	N-terminus	+42
Nitration	Tyr(W)	+45
Deamidation	Asn(N),Gln(Q)	+1
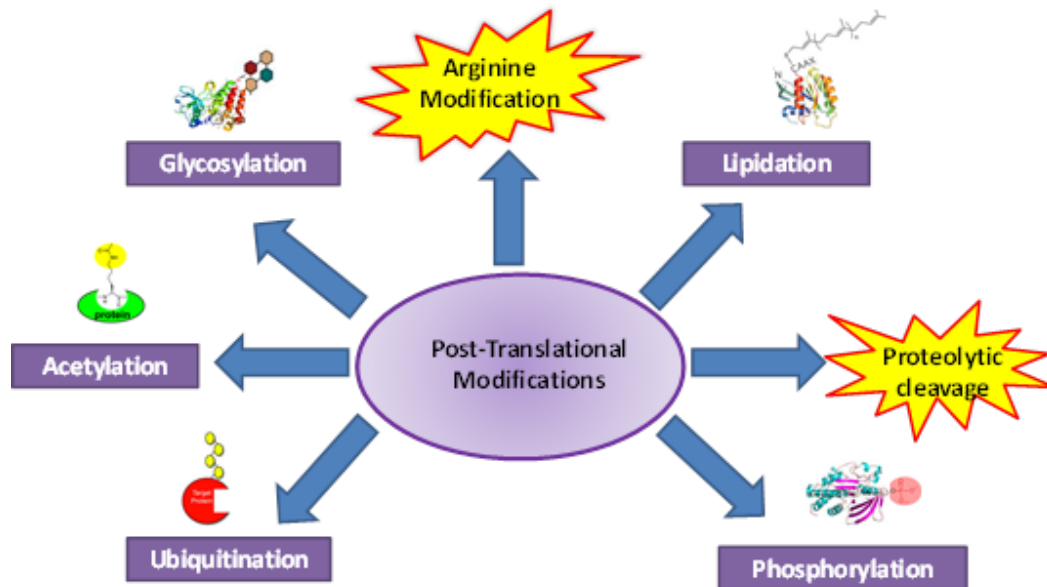
Figure 1-1 Common post-translational modifications (PTMs)

PTMs have been traditionally identified by Edman degradation, Western blotting & amino acid analysis.[7] Over the past several decades technological innovations have positioned mass spectrometry (MS) as an indispensable tool to analyze PTMs.[8] Despite the availability of versatile mass spectrometry-based techniques, PTM analysis is still a challenging task. MS offers many advantages for characterization of PTMs, including (i) high sensitivity, high resolution, and high mass accuracy, (ii) the ability to identify the specific site of modification and (iii) to facilitate detections of PTMs in complex mixtures of proteins, as well as (iv) to quantify the abundance of PTMs. The modification can be defined by the comparison of mass changes. This dissertation is focused on Arginine modification and proteolytic cleavage.

1.3 Proteins

Proteins are macromolecules in the human body. Amino acids are the building blocks of protein, which form when the two amino acids linked through the peptide bonds. Amino acid consists of an amine group (NH_2), carboxylic group (COOH), variable R group as the side chain, and a hydrogen atom attached to the central carbon atom (Figure 1). The simplest amino acid found with hydrogen as the R group is glycine. The R groups are what make the different amino acids with different substituents.
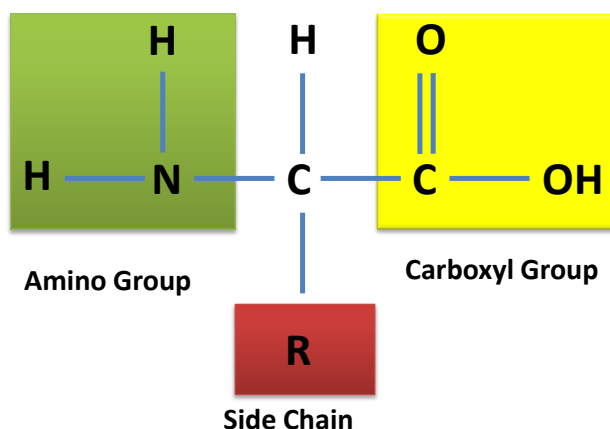
Figure 1-2 Basic structure of an amino acid

There are 20 essential amino acids identified in the human body. Protein structures are classified under four different types. The linear sequence of the amino acids make the polypeptide chain refer to the primary structure. The secondary structure represents the structural conformation of the amino acids. There are two main types of secondary structures, the alpha (α) helix structure and the beta (β) pleated (or folded) sheet, depending on the pattern of hydrogen bonds between the amine hydrogen and carbonyl oxygen atoms in the peptide backbone. The α-helices are right-handed, coiled structures which are stabilized by hydrogen bonds as well β-sheets, which come in two forms—parallel β-sheets where the strands are pointing in an identical direction and in the anti-parallel β-sheets which point in opposite directions. Folding of the secondary structure leads to the 3D structure of the protein known as tertiary structure. Hydrogen bonds, ionic bonds, hydrophobic interactions and disulfide bonds offer a stability which holds the tertiary structure protein together. Protein quaternary structures exist with multiple polypeptide chains known as subunits. In addition, it defines the folding and arrangement of these subunits.

1.4 Arginine

In the early 19th century, few amino acids had been discovered. Asparagine was the first amino acid discovered in 1806 when French scientists, Louis-Nicolas Vauquelin and Pierre Jean Robique, isolated it from asparagus juice. In 1886 Arginine was discovered by a German chemist Ernst Schultze, who isolated it from a lupin seedling extract. The basic structures of arginine contain an amino group at the N-terminal and a carboxylic group at the C-terminal; the side chain consists of a three-carbon aliphatic chain with a guanidine group. In physiological conditions, the guanidine group exists as a protonated ion with a pKa value of about 13.6.
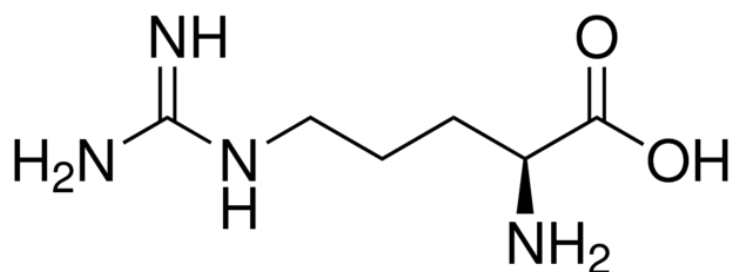
Figure 1-3 Structure of arginine amino acid

1.5 Arginine PTMs

In both prokaryotes and eukaryotes, the arginine residues, which are embedded in the proteins having Post-translational modifications (PTMs) have been considered potential targets for biomarkers. Under current studies researchers have identified four enzymatic PTMs such as methylation, phosphorylation, citrullination and ADP-ribosylation, which are highly regulated during normal cellular functions.[9,10] Two non-enzymatic PTMs known as carbonylation and advanced glycation end products (AGE) which affect human aging causing the loss of activities in metabolites.[11] The most widely studied arginine PTM is methylation. Modifications of arginine residues can affect several cellular processes where the unique structure of arginine with its positively charged guanidine group tends to network with other proteins and nucleic acids by forming hydrogen bonds.[12]

In general, the positional proteomic workflow as an enrichment strategy can be a positive selection or negative selection. In early studies, the protein termini were modified with a tag which enabled a targeted enrichment of the pool of digested peptides. Later, researchers worked with the newly generated termini and followed their depletion, which resulted in enrichment of the target termini.[13-15]

1.6 Proteolytic Cleavage

A crucial problem we are facing today is cellular protein modifications, which change the protein structure and function by driving many intra- and extracellular events. Edman degradation is used for N-terminal sequencing of proteins by serially removing one residue at a time from the N-terminal end of each protein and identifying the N-termini. This method cannot be utilized for sequencing the N-terminally blocked proteins. Sensitivity can be achieved by using radioactive reagents. Nevertheless, the radioactive waste needs careful disposal methods, and these methods are vulnerable due to significant amounts of required material and the inability to identify infrequent PTMs.[16] The key determinants for analyzing N-terminal sequences are found in the evaluation of the potential cleavage of N-Terminus, the evaluation of protein degradation, the ability to identify proteins or peptides and peptide sequencing to determine amino acid sequence changes in selected peptides. Analysis of the N-termini of proteins is essential for the identification of target proteins and for understanding their roles in cell.
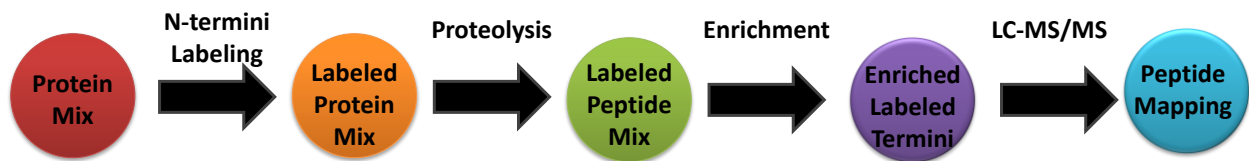
Figure 1-4 – Schematic diagram of positional proteomics workflow

Recently several methods have been reported to isolate the N-terminal peptides. These methods need multiple analyses that require time consuming purification each protein from a complex protein mixture with low sensitivity. During or after the translation of protein synthesis, proteolysis in the body is involved in the complicated process, which regulates intracellular and extracellular signal transduction. This may cause removal of the N-terminal methionine and signal peptide. Removal of signal peptides will generate new N-termini. Several studies have found ways to isolate the N-terminal peptides by MS coupled with gel separation and blotted

proteins. Gevaert et al. developed a peptide isolation method based on diagonal electrophoresis and diagonal chromatography.[17] Recent publications on N-terminal peptide enrichment methods involve disulfide bond reduction and alkylation of proteins followed by the modification of lysine residues to homoarginine via a guanidination reaction. The capture of N-terminal peptides is achieved by tryptic digestion followed by direct or reverse purification of original and newly formed N-terminal peptides by immobilized amine reactive reagents. This research focused on direct and reverse N-terminal purification strategies by the performic acid oxidation of the disulfide bonds as well as reduction and alkylation with dithiothreitol and iodoacetamide. A classical bottom-up workflow was used where the protein was digested with a protease enzyme and resulting peptides were analyzed by tandem mass spectrometry.

1.7 Mass Spectrometry

Mass spectrometry is an extremely useful analytical technique, which plays a major role in the proteomic field. A mass spectrometer can generate multiple ions from the analyte of interest and then sort them according to their specific mass-to-charge ratio (m/z) for detection in proportion to their abundance.[18] In 1912, Joseph John (J.J) Thomson invented first mass spectrometer.[19] Initially the apparatus was used to measure the atomic weight of elements and to record the natural relative abundance of elemental isotopes by physicists. A mass spectrometer is composed of three main components: an ion source, a mass analyzer and a detector.
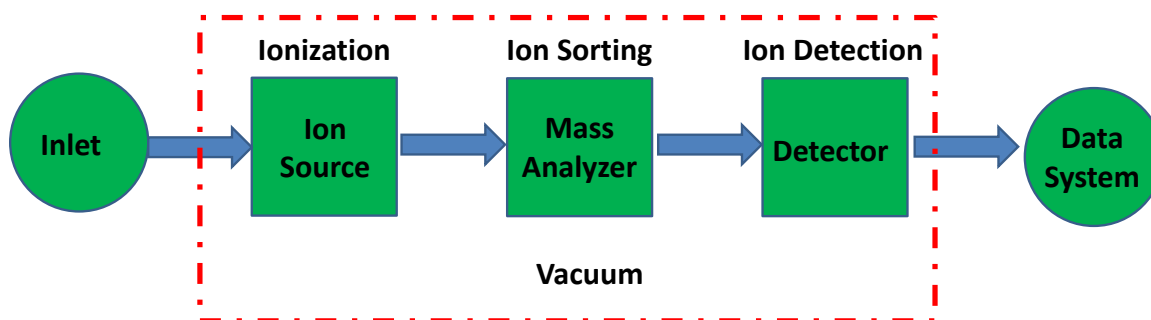
Figure 1-5 Schematic of the main components of a mass spectrometer

7

The system provides a vacuum chamber, where ionization occurs. The gaseous ions then move into and electric field where ion sources change them into charged molecules. The most common ion sources are electrospray ionization (ESI) and matrix assisted laser desorption/ionization (MALDI). In my experiments, I used both the MALDI-QIT-TOF, where ions formed an ion source inside the vacuum, and the LTQ Velos Pro dual-pressure ion trap, which uses an atmospheric pressure ion source. Both ionization techniques provide soft ionization methods, which do not fragment the sample analyte. Therefore, these two techniques are ideal to analyze biological samples such as proteins, lipids, nucleotides etc.[20,21]

1.8 Soft Ionization Methods

Matrix-assisted laser desorption ionization (MALDI) is the most widely employed soft ionization technique in mass spectrometric analysis. In 1985, Franz Hillenkamp, Michael Karas, and their coworkers invented the technique.[22,23] In 2002, Koichi Tanaka received the Nobel Prize in Chemistry for his contribution towards the development of soft ionization methods.[24]

The sample is uniformly mixed with the appropriate matrix. After mixing, the spotted sample is co-crystallized. A focused laser beam irradiation causes the vaporization of matrix which provides nondestructive vaporization of the sample analyte. Matrices are organic compounds having low vapor pressure and low molecular weight with volatile properties. Depending on the type of molecules being analyzed, it is very challenging to choose the best matrix. The matrix can be 2,5-dihydroxybenzoic acid (DHB), α-cyano-4-hydroxycinnamic acid (CHCA) and 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid).

A MALDI ionization source can be paired with different types of mass analyzers. The simplest device is the linear TOF mass analyzer. A TOF mass analyzer consists of ion acceleration, focusing optics, and a flight tube, which will accelerate the set of ions' journey to the detector.

TOF analysis is based on the measured m/z ratios of ions based on time it takes for ions to fly into the analyzer and strike the detector.
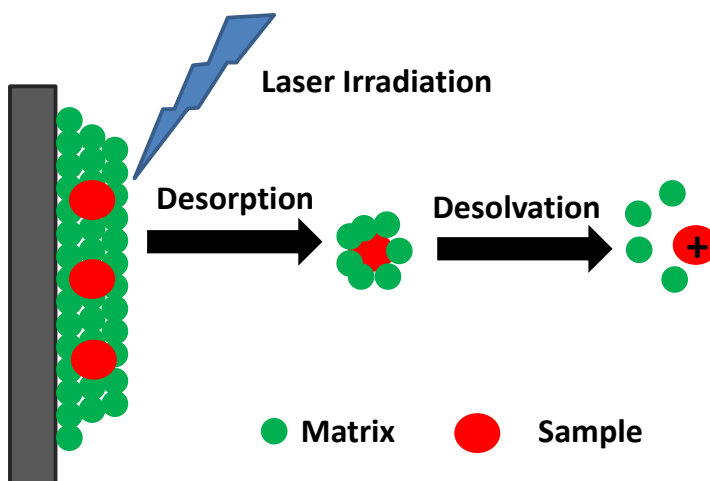
Figure 1-6 Simple flow chart of matrix-assisted laser desorption ionization (MALDI) in mass spectrometry

The figure is adapted from http://nptel.ac.in/courses

Electrospray Ionization (ESI)

ESI is a very powerful soft ionization technique, which will not fragment the analyte of interest where we can analyze the intact molecular ion.[25,26] This can be easily coupled with the high-performance liquid chromatography (LC) system. So, this is known as an ideal technique to analyze different types of biological samples. ESI involves spraying a sample solution through a capillary. Before delivering the ions into the mass spectrometer, with the help of the electrical energy, the ionic species in solution are converted into a gaseous phase,[27,28] as shown in Fig. 1-7. High voltage was applied in between the electrospray tip and the counter electrode. This allows the spraying of charged droplets from the end of the tip with a surface charge of the same polarity to the charge on the tip. Because the same charges repel each, other charged droplets repelled from the electrospray tip towards the counter electrode end up forming a Taylor cone. During this time, the sample droplet shrinks until it reaches the point where the

surface tension can no longer bear the charge at which point Coulombic repulsion occurs and the droplet fission (Coulombic explosion as noted in Fig. 1-7) leads to the formation of smaller droplets (a plasma of ionized atomic particles).
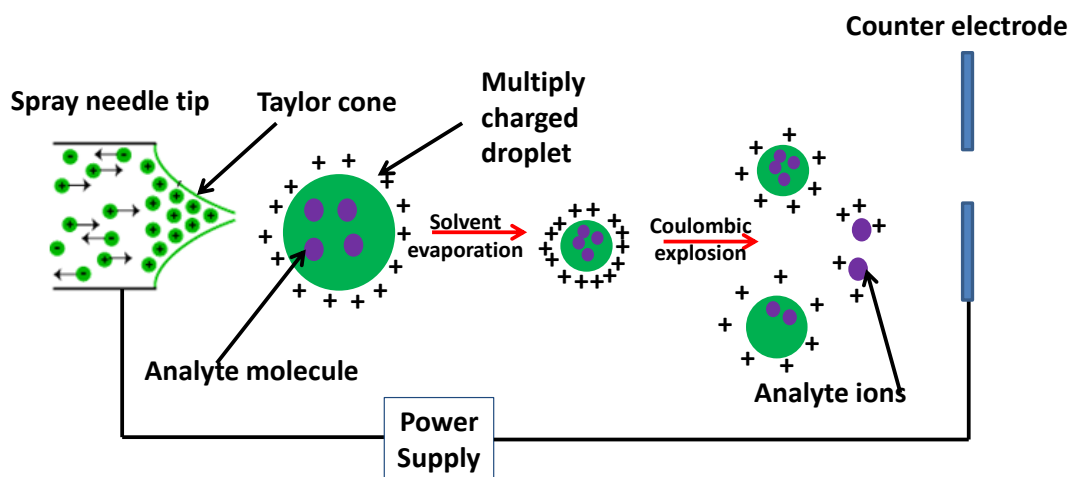
Figure 1-7 Electrospray ionization method in mass spectrometry

This figure is adapted from http://www.bris.ac.uk

These atomic particles can either be singly or multiply charged. Their emitted ions are collected by a skimmer and delivered to the mass analyzer. During this acceleration time the charge droplets are going through pressure gradient and potential gradient in the analyzer region.

1.9 Ion Activation

Tandem mass spectrometry is also known as the MS/MS studies which can be used for structure determination of molecules. In proteomics field peptide sequencing is widely applied. It facilitates characterization of protein structures and the PTMs associated with the proteins. There are several ion activation methods. In CID the ion activation takes place when the selected ions collide with neutral gas in the collision cell.[27,29,30]

10
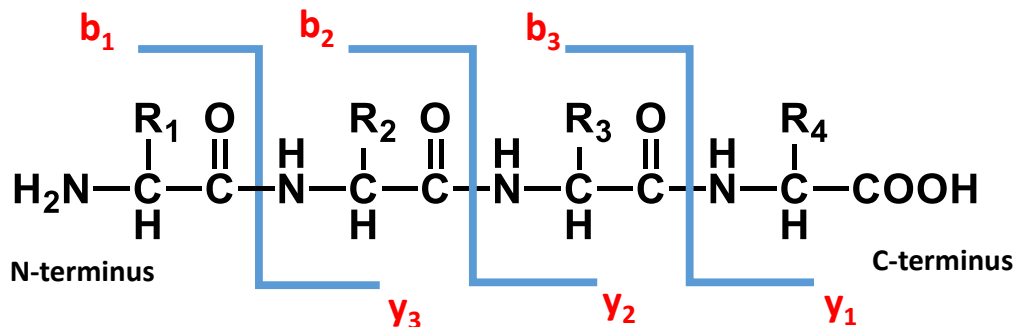
Figure 1-8 Interpretation of b and y ion formation in peptides

Peptide fragmentation nomenclature was introduced by Roepstorff and Fohlman in 1984.[31] Alphabet numbers are used to denote the fragment ions based on the cleavage in the peptide backbone relative to N-terminus or C-terminus. In collision induced dissociation (CID) fragmentation, the C−N bond cleaves and yields by fragment ions. AS depicted in the Figure 1-8 the numerical subscript represents the number of amino acids in the fragment ions. CID fragmentation dominates the other fragmentation techniques. Electron activation methods like electron transfer dissociation (ETD) are responsible for generating c- and z-type fragments by cleavage at the N−Cα bond in the peptide. For large-scale evaluation, multiple activation methods can be performed for the further confirmation.

1.10 Mass Analyzers

A mass analyzer is the heart of the mass spectrometer component that separates the ionized ions and is based on charge-to mass-ratios. Among the many applications of mass spectrometry, the ones that are encountered the most are quadrupole mass analysis, time of flight mass analysis and ion trap mass analysis. In proteomics analysis, common ion-trapping mass analyzers are used. There are two key trapped-ion mass analyzers: dynamic trapping mass analyzers with 3D quadrupole ion traps, and static trapping mass analyzers, i.e., ion

cyclotron resonance mass spectrometers. Consequently, both operate on the same principle of storing ions in the trap and ejecting them via DC and RF electric fields in a series of timely varying manners. Both have their own advantages and limitations. Both can provide high resolution and high sensitivity. The drawback is ion decomposition due to spending more time in the trap, which can also lead to undesirable interactions. Thus, tandem mass spectrometry studies have been established with various types of mass analyzers depending on the type of application.[32-35]

Table 1-2 Comparison of the characteristics of commonly used mass spectrometers in proteomics studies

Instrument	Ion source	Mass accuracy (ppm)	Resolution	Mass range	MS/MS Capability	Dynamic Range
LTQ	ESI	500	2000	50-2000, 200-4000	$MS^{n>13}$	1 E3
Q-q-Q	ESI	1000	1000	10-4000	$MS^{n>13}$	1 E4
Q-q-LIT	ESI	500	2000	5-2800	MS/MS	6 E6
TOF	MALDI	20	20000	No upper limit	$MS^{n>13}$	4 E6
TOF-TOF	MALDI	20	20000	No upper limit	Fragmentation achievable by post-source-decay	1 E4
Q-q-TOF	ESI/MALDI	20	20000	No upper limit	MS/MS	1 E4
FTICR	ESI/MALDI	2	750000	50-2000, 200-4000	$MS^{n>13}$	1 E4
Orbitrap	ESI/MALDI	5	1000000	50-2000, 200-4000	$MS^{n>13}$	1 E3

1.11 Detectors

After the mass separation, the ionized molecules are accelerated toward the detector where they were detected and later converted into a digital output. There are various detectors used in mass spectrometry. The most frequently used detectors are the electron multiplier detector, conversion dynodes, and the Faraday cup collector.

Chapter 2

Evaluation of chemical labeling methods for identifying functional arginine residues of proteins by mass spectrometry

2.1 Abstract

Arginine residues undergo several kinds of post-translational modifications (PTMs). These PTMs are associated with several inflammatory diseases, such as rheumatoid arthritis, atherosclerosis, and diabetes. Mass spectrometric studies of arginine-modified proteins and peptides are very important, not only to identify the reactive arginine residues but also to understand the tandem mass spectrometry behavior of these peptides for assigning the sequences unambiguously. Herein, we utilize tandem mass spectrometry to report the performance of two widely used arginine labeling reagents 1,2 cyclohexanedione (CHD) and phenylglyoxal (PG) with several arginine-containing peptides and proteins. Time course labeling studies were performed to demonstrate the selectivity of the reagents in proteins or protein digests. Structural studies on the proteins were also explored to better understand the reaction sites and position of arginine residues. We found that CHD showed better labeling efficiencies compared to PG. Reactive arginine profiling on a purified albumin protein clearly pointed out the cellular glycation modification site for this protein with high confidence. We believe these detailed mass-spectrometric studies will provide significant input to profile reactive arginine residues in large-scale studies; therefore, targeted proteomics can be performed at the short listed reactive sites for cellular arginine modifications.
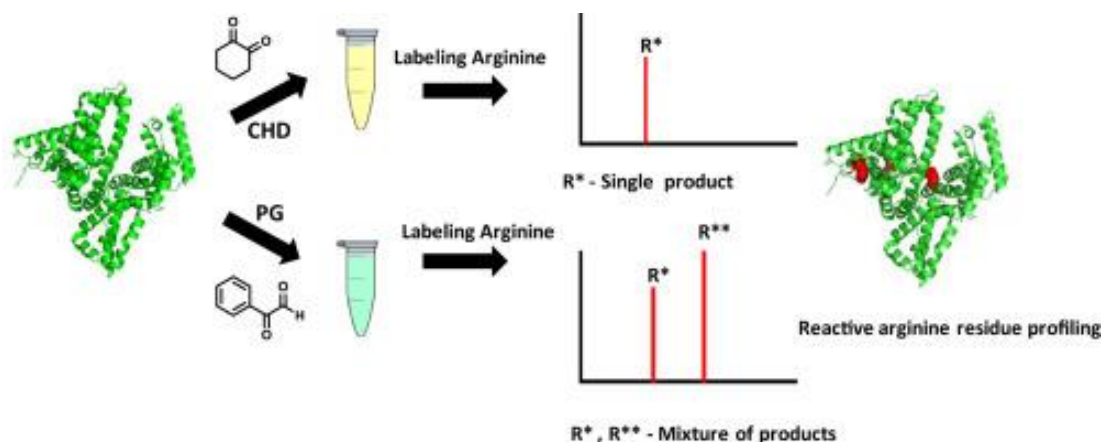
Figure 2-1 Graphical abstract

2.2 Introduction

Post-translational modifications (PTMs) are a series of covalent procedures that process proteins. These procedures result in the cleavage of amino acid sequences as well as an addition or removal of chemical moieties to amino acid residues. Mass spectrometry (MS) is an indispensable tool to characterize PTMs. Arginine is one of the most common natural amino acids, and it plays a major role as a recognition site for other proteins or RNA. These recognition sites are the key to enzyme activities and structures.[36] Changes in protein modification may cause proteins to go from active to inactive states or alter their activities and cause numerous cellular events in disease states. Glycation is a non-enzymatic modification involved with the addition of carbohydrate/reducing sugar moiety to the amine group of the protein. Accumulation of advanced glycation end products (AGE) is an indication of serious complications in the diagnosis and treatment of diabetes.[37,38] AGE products exhibit a wide range of chemical structures, and therefore, carry out different biological activities. Major chemically characterized AGE is evident in the formation of imidazolone, which is a glycation end product formed from 3-deoxyglucosone (3-DG) reacting with the guanidino group of arginine.[39]

14

According to Patthy and Smith, arginine residues of proteins play a significant role in ensuring electrostatic interactions between enzymes and the negatively charged cell surface.[40,41] Several arginine PTMs have been identified recently by proteomic and mass spectrometric methods.[3,42] Various studies have investigated arginine reactivity toward several reagents and a few mechanisms have been proposed for chemical detection of arginine groups.[43,44] Very few studies have been conducted with mass spectrometry.[45-48] Residue-specific chemical modification of proteins is a powerful strategy to identify functional amino acid residues of proteins.[49] Chemical modifications of proteins by arginine selective reagents are beneficial in identification of these functional residues. Phenylglyoxal, 2,3-butanedione, 4-hydroxy-3-nitrophenylglyoxal and camphorquinone (CQ) were found to be the reagents of choice for arginine modifications in proteins and peptides for structural probing with mass spectrometry.[11,50-52] This covalent labeling of amino acids has been a helpful tool to study protein structure and protein-protein interaction by the degree of variation in their reactivity.[53,54] Mainly, this interaction depends on the accessibility of reagents towards amino acids, the reactivity of reagents, and the reactivity of amino acids.[55,56] Friess et al. introduced a computational method to selectively identify arginine residues by considering accessibility parameters,[45] and they did a study on lysozyme, ribonuclease-A, myoglobin, and adenylate kinase. Friess et al.[46] also applied their computational parameter-based accessibility methodology to much larger proteins such as aldolase and albumin.[49] Probing arginine residues in proteins by using MS and a chemical labeling concept was performed by Leitner and Lindner in 2003.[57] They examined the reactivity of arginine residues in model proteins, ubiquitin, cytochrome c, myoglobin, ribonuclease-A and lysozyme, by using 2,3-butanedione and an aryl group boronic acid. Toi et al. showed the initial interest in dicarbonyl compound 1,2-cyclohexanedione and benzyl.[58,59] They showed how under alkaline conditions, benzyl and 1,2-cyclohexanedione undergo condensation reactions. In 1970, Yankeelov et al. found that butanedione is another diketone which can be used as an arginine modification reagent.[60]

Leitner and Lindner presented an alternative covalent labeling technique based on dicarbonyl modification that specifically targets arginine residues in peptides and proteins, but it also includes a novel cyclization step in the method.[44,61] Seeking an alternative to dicarbonyls, Akinsiku et al. introduced the RNA foot printing organic compound, kethoxal, as a reagent to modify guanidinium groups under neutral pH conditions.[62] To map modification sites and characterization of functional arginine residues, we compared two commonly used chemical labeling strategies, which can identify the target sites effectively in the complex mixtures of proteins.

This dissertation presents an explanation of the MS fragmentation pattern of modified peptides with two arginine specific reagents, CHD (1,2-cyclohexanedione) and phenylglyoxal (PG). Our work demonstrated that CHD is a superior labeling reagent compared to PG in peptides and protein labeling studies. We also showed the labeling efficiencies of these reagents in mono- and diarginyl peptides as well as their mass spectrometric fragmentation. Reactive arginine residues were identified in pure protein albumin using CHD at various time points. Intact protein labeling studies pointed out the reactive arginine residues, which showed that effective labeling could be accomplished in previously identified cellular PTM locations. Although several studies have been done on arginine labeling reagents, a comprehensive mass spectrometry study on the labeled peptides is lacking. In this study, we demonstrated a mass-spectrometric study with two widely used arginine selective reagents. This study contributes significantly to targeted proteomics studies for finding cellular arginine reactive residues, which are susceptible to covalent modifications due to disease processes.

2.3 Materials and methods

2.3.1 Chemicals and materials

Bovine serum albumin (BSA), ubiquitin (from bovine erythrocytes), lysozyme (from hen egg white), 1,2-cyclohexanedione(CHD), phenylglyoxal monohydrate (PG), formic acid (FA), acetonitrile (ACN), Iodoacetamide (IAM), ammonium bicarbonate (NH_4HCO_3) and sodium hydroxide (NaOH) pellets were purchased from Sigma Aldrich (St. Louis, MO, USA). Dithiothreitol was obtained from Bio-Rad (Hercules, CA). MALDI matrix 2,5-dihydroxybenzoic acid (DHB) was purchased from ProteoChem (Loves Park, IL USA). Sequencing-grade modified trypsin was used for proteolysis and obtained from Promega (Madison, WI, USA). Samples were desalted by Pierce C18 Tips from Thermo Fisher Scientific (Rockford, IL, USA). Bradykinin, neurotensin (NT) and substance P were purchased from ANASPEC (Fremont, CA). Pierce concentrators, such as 10K MWCO (Molecular Weight Cut-Off) were used to remove excess CHD.

2.3.2 Protein digestion

To the bovine serum albumin (1 mM, 5 µL) solution 1 M DTT was added to make the final concentration of 10 mM; then, the solution was incubated at 56 °C for 45 min. The temperature was lowered to 45 °C for 10 min and then to 25 °C for 10 min. 1 M IAM solution was added to make the final concentration of 55 mM and was incubated at RT in the dark for 30 min. We used trypsin as the enzyme for digestion. Prior to digestion, the sample was diluted with 50 mM NH_4HCO_3; then, protease solution was added to the resulting solution at a ratio of protein: trypsin (100:1) and incubated at 37 °C for 12 h. The solution was acidified with 0.1% formic acid to stop further digestions. Afterward, samples were desalted by using Pierce C18 ZipTips.

2.3.3 Labeling of arginine residues in peptides

The labeling reaction was evaluated using an established method. To evaluate the labeling efficiency of the arginine selective reagent CHD, the reaction was carried out with three model peptides, bradykinin, substance P (C-term amide), and neurotensin in 200 mM NaOH medium. The rate of the reaction of phenylglyoxal (PG) with arginine increased with an increasing pH from 7.5 to 11.5. There are lots of side reactions possible when we increase the pH. PG studies done with different pH and higher pH levels exhibited degradation of products. The model reaction with arginine was much faster in bicarbonate, dimethylamine or trimethylamine buffer than in N-methylmorpholine, borate, phosphate or tris buffer. To minimize the product degradation and to compare CHD and PG reactivities towards arginine residues, we carried out the PG reaction in 0.125 M $KHCO_3$ at pH 8.5.

2.3.4 Chemical modification of intact bovine serum albumin (BSA) and lysozyme

Modification of proteins using 1,2-cyclohexanedione was performed according to the procedure outlined by Toi et al.[56] The reaction mixture contained 1 mM BSA in 200 mM sodium hydroxide (NaOH) solution and was kept at 37 °C under agitation at different protocol time points (30 min, 1 h, 6 h, 12 h, 18 h, 24 h). The guanidino group of arginine reacts with 1, 2-cyclohexanedione (CHD) in alkaline aqueous medium and forms a CHD-arginine covalent product (Fig. 2-2). The condensation of CHD with the guanidino group of arginine results in an imidazolidinone. Excess CHD was removed using a 10K molecular-weight cutoff (MWCO) concentrator. After that, proteins were digested following the digestion procedure mentioned above. The digested covalent product solution was fully dried and then the desalted sample was analyzed with the MALDI-QIT-TOF mass spectrometer.

Modification was performed under agitation at different time points, such as 30 min, 1 h, 6 h, 12 h, 18 h and 24 h. Chemical modification of lysozyme was also performed at 37 °C for different time intervals as established in the protocol time points.
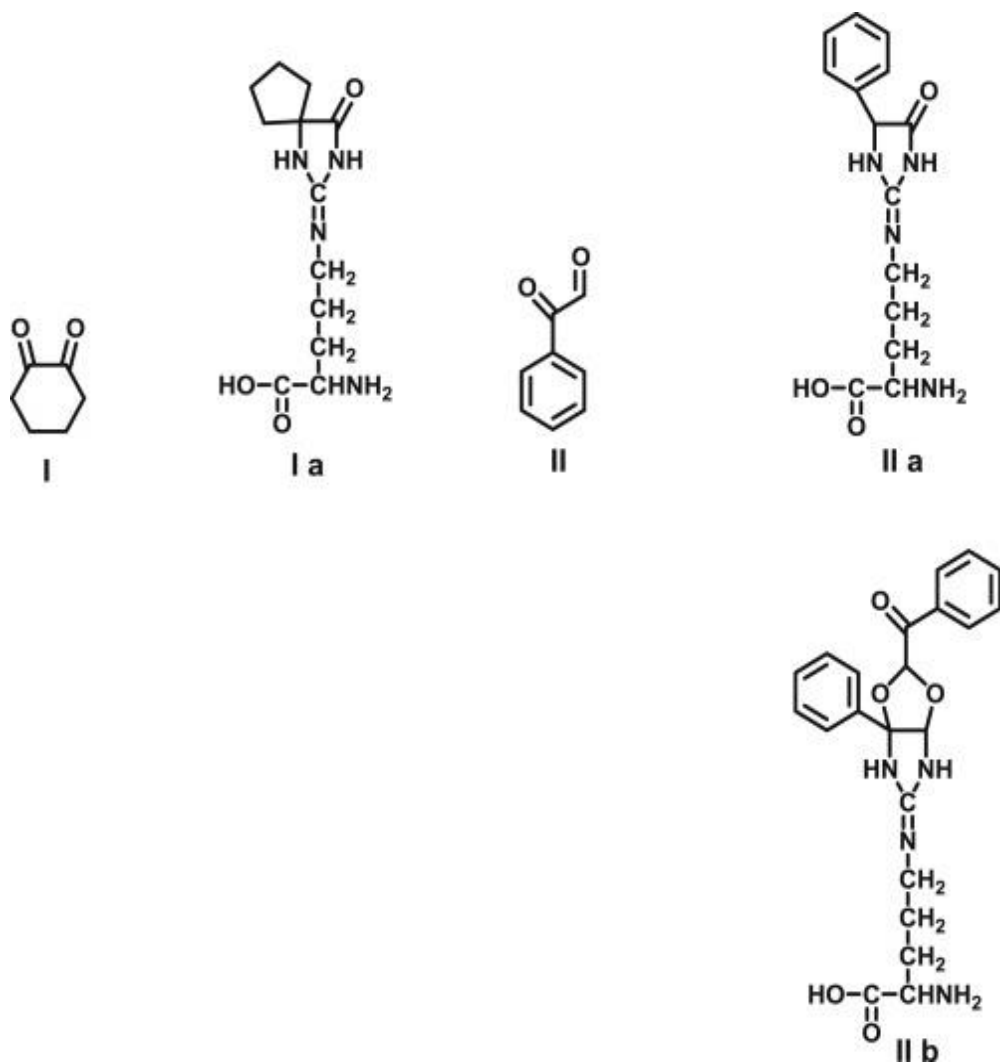
Fig. 2-2 The structures of the modifications resulting in the label peptides with 1,2-cyclohexanedione and phenylglyoxal. I-1,2-cyclohexanedione, II-phenylglyoxal, Ia-Arginine-CHD covalent product, IIa-partial modified arginine-PG covalent product (αPG –single and 2αPG –double), IIb-one full arginine-PG covalent products (addition of two PG products).

2.4 Results and Discussion

2.4.1 Modification by CHD

Bradykinin, neurotensin, and substance P were reacted with 30 M excesses of 1,2-cyclohexanedione in 200 mM NaOH. Under the conditions mentioned in the preceding subsection, an arginine-CHD covalent product was formed. After modification, the product mass for each peptide showed a significant mass shift. The reactivity of CHD towards arginine was revealed by this mass shift. Bradykinin is a peptide with the amino acid sequence of RPPGFSPFR (monoisotopic mass, 1059.5614 Da) containing two arginine residues, located in the C-terminal and the N-terminal end of the peptide. After reaction with CHD, we determined that the modifications of the two arginine residues were 100% complete (Fig. 2-3A). The modified mass for bradykinin (R*PPGFSPFR*) was m/z 1248.6500 $(M + H + 2CHD)^+$. For bradykinin the mass shift was calculated, and the observed m/z was matched with a [peptide mass (1059.5614) + mass addition (188.0838) + H (1.0078) = 1248.6530] mass accuracy of −2.4 ppm (structure in page S2, supplementary data). The added mass corresponds to the addition of two arginine-CHD covalent products. Another peptide used in this study was substance P (C-term amide) which contains only one arginine residue in the N-terminal side of the peptide (RPKPQQFFGLM, monoisotopic mass 1346.7281 Da). Modification of the arginine residue (R*PKPQQFFGLM) gave a modified ion mass of m/z 1441.7913$(M + H + CHD)^+$ (Fig. 2-3B) with a mass accuracy of 9.4 ppm (page S5, supplementary data). As expected, we observed 94.0519 Da mass additions for the single arginine residue in substance P, which matched the calculated mass of a single arginine-CHD covalent product (theoretical, 94.0419 Da). Neurotensin is an N-terminal blocked peptide which was also used for these labeling studies. The amino acid sequence of neurotensin is pyroglu-LYENKPR*R*PYIL (monoisotopic unmodified mass 1671.9097 Da), and it has two arginine residues which are adjacent to each other. Modification with CHD showed a peak at m/z 1861.0154$(M + H + 2CHD)^+$, which proved

that both arginine residues were derivatized by CHD (Fig. 2-3C). For neurotensin, we observed the same mass shift with the addition of two CHD molecules (188.0979 Da, theoretical 188.0838 Da), and the mass accuracy was found to be 7.6 ppm (Table 2-1, supplementary data). All the mass accuracy data for CHD along with the structures of their modified peptides, was provided in the supplementary data. Arginine residues in these peptides were completely derivatized by CHD. The product ions were then used to expound the location of modified peptides by determining the amino acid sequences of the precursor peptides.

To better understand the CHD modified sites, modified peptide MS/MS spectra were presented in Fig. 2-4. Fig. 2-4A showed a MS/MS spectrum for the modified peptide in Fig. 2-4A (top panel at m/z 1248.6500). This peak was assigned to the CHD modified peptide bradykinin with sequence R*PPGFSPFR*, having two modified arginine residues. In this spectrum, we clearly saw the b and y fragment ions of the modified peptide. Fragment ions (b2*) observed at m/z 348.20, m/z 428.23 (b3-NH$_{3*}$) and the b5* ion at m/z 649.50 contained the CHD-modified arginine mass on the side of the N-terminal. The symbol * in the figures designates the peptide fragments with the modified arginine residues. The fragmentation pattern of the peptides as shown in Fig. 3A, provide a clear proof about the CHD mass addition (94.04 Da) to the peptides. Unmodified fragment ion b2 (m/z 254.16) for the peptide bradykinin showed a clear mass difference of 94.04 Da from the modified fragment b2*, which corresponds to the CHD modification at the N-terminal arginine residue (Fig. S8,A). All the y fragment ions y2* (m/z 416.23), y3*(m/z 513.28), y4*(m/z 600.31), y5*(m/z 747.38), y6*(m/z 804.40), y7*(m/z 901.45), y7-H$_2$O*(m/z 883.44) and y8*(m/z 998.51) confirmed another CHD modified arginine mass at the C-terminal end. By comparing with the unmodified-peptide y fragment ions in Fig. S8, A [y5, (m/z 653.34), y7, (m/z 807.41) and y8, (m/z 904.47)], we found the mass difference was 94.04 Da. This MS/MS spectrum confirmed the selective addition of CHD in two different arginines in the bradykinin peptide.
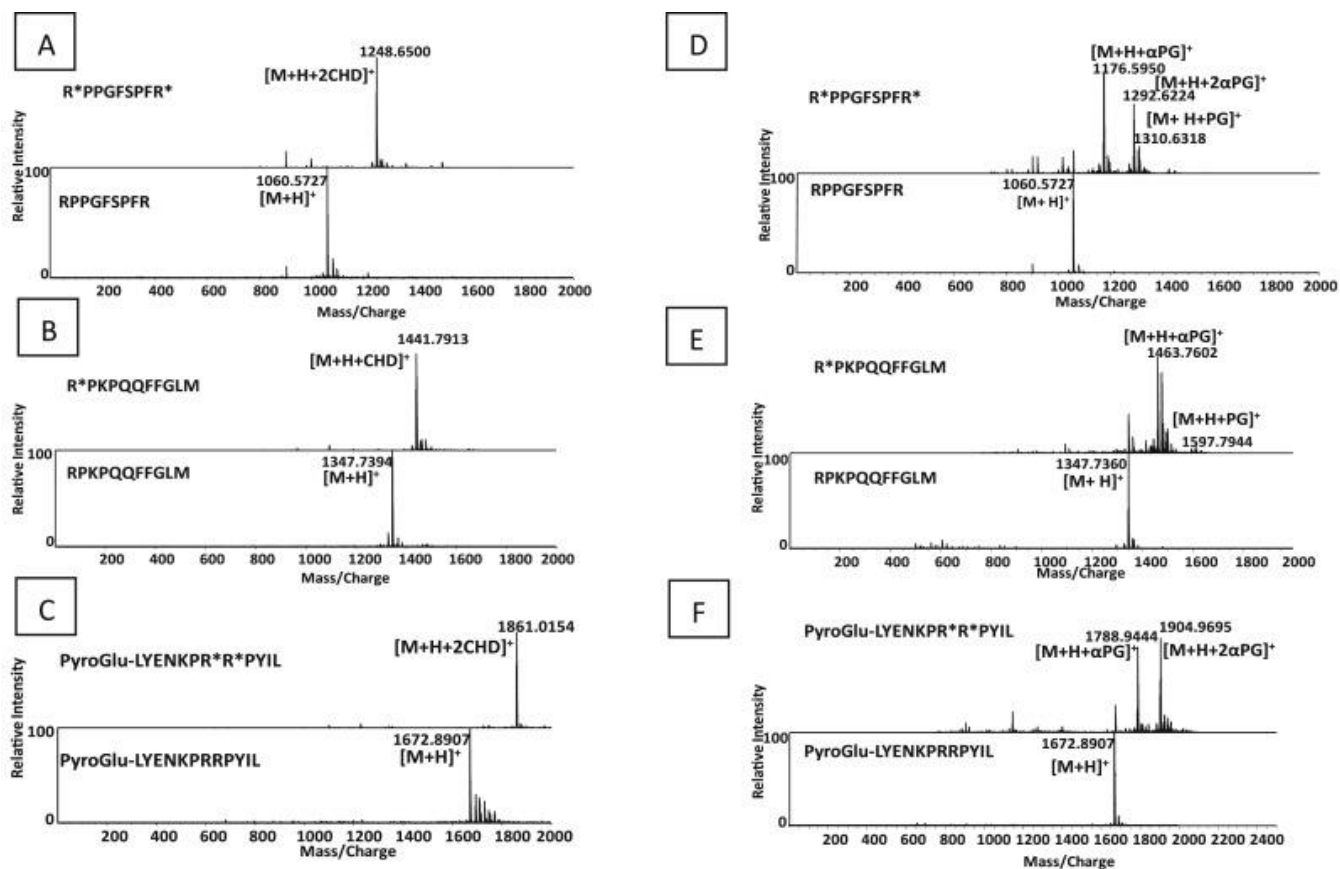
Fig. 2-3 MALDI-QIT-TOF MS spectra of modified peptides with two different arginine-reactive reagents: (A) CHD modified bradykinin at m/z 1248.6500 (M + H + 2CHD)⁺; (B) CHD modified substance P at m/z 1441.7913 (M + H + CHD)⁺; (C) CHD modified neurotensin at m/z 1861.0154 (M + H + 2CHD)⁺; (D) PG modified bradykinin at m/z 1176.5950 [(M + H + αPG)⁺]; 1292.6224 [(M + H+2αPG)⁺)], and 1310.6318 [(M + H + PG)⁺]; (E) PG modified substance P at m/z 1463.7602 [(M + H+αPG)⁺), and 1597.7944 [(M + H + PG)⁺)]; (F) PG modified neurotensin m/z 1788.9444 (M + H + αPG)⁺] and 1904.9695 [(M + H + 2αPG)⁺]. The mass addition for CHD is 94.0519 Da for one CHD, and αPG is 116.0262 Da. (αPG - single partial modification of PG, 2αPG - double partial modification of PG, PG - one full PG modification).
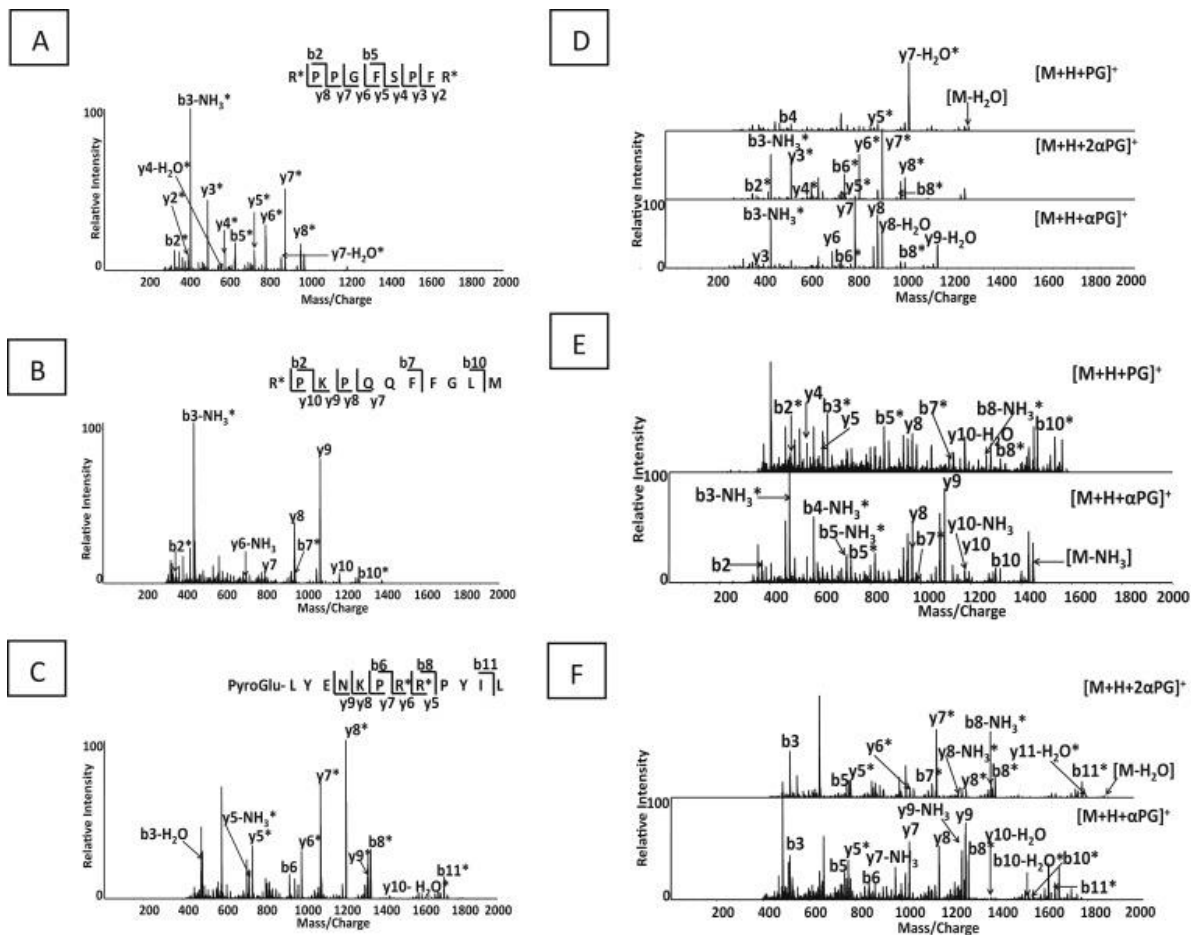
Fig. 2-4. MALDI-QIT-TOF MS/MS spectra of modified peptides with two different arginine reactive reagents: (A) MS/MS spectrum of CHD modified bradykinin at m/z 1248.6500 (M + H + 2CHD)$^+$; (B) MS/MS spectrum of CHD modified substance P at m/z 1441.7913 (M + H + CHD)$^+$; (C) MS/MS spectrum of CHD modified neurotensin at m/z 1861.0154 (M + H + 2CHD)$^+$; (D) MS/MS spectra of different PG modified bradykinin at m/z 1176.5950 [(M + H + αPG)$^+$](bottom), 1292.6224 [(M + H+2αPG)$^+$) (middle)] at 1310.6318 [(M + H + PG)$^+$](top); (E) MS/MS spectra of PG modified substance P at m/z 1463.7602 [(M + H+αPG)$^+$)](bottom) and m/z 1597.7944 [(M + H + PG)$^+$)](top); (F) MS/MS spectra of PG modified neurotensin at m/z 1788.9444 (M + H + αPG)$^+$](bottom) and 1904.9695 [(M + H + 2αPG)$^+$](top). (The symbol * in the figures designates the peptide fragments with the modified arginine residues).

Detailed interpretation of the mass spectrum confirmed the assignment of modification sites in a model peptide substance P (Fig. 2-4B). Observation of fragment b2*, b3-NH$_3$*, b7* and b10* further demonstrated that these fragments contained CHD modified arginine residue mass. The mass difference 94.04 Da between the modified fragment b10* (m/z 1293.71) and the unmodified fragment b10 (m/z 1199.67) indicated the addition of one CHD molecule to the arginine residue (m/z 1293.71- m/z 1199.67 = 94.04 Da). Unmodified peptide MS/MS are also provided in Fig. 2-16B. The MALDI-QIT-TOF MS/MS data obtained for CHD-modified peptide neurotensin, are shown in Fig. 2-4C. Fragment ions b8*, b11*, y5*, y5-NH$_3$*, y6*, y7*, y8*, y9* and y10-H$_2$O* were identified with the addition of modified masses. When we compared these fragments with the unmodified fragment ions of the peptide neurotensin in Fig. 2-16C, it was found that the mass differences between the modified and the unmodified fragments b8, b11, y7, y8 and y9 were 188.08 Da. This data clearly shows the formation of arginine-CHD covalent products at both arginine residues. The MS/MS spectra of unmodified neurotensin were also evaluated (Fig. 2-16, C). It is clear from the b and y ions that both adjacent arginines were selectively modified with CHD. The evidence presented in this study suggests that the addition of 94.04 Da can always be used as a signature for the formation of arginine-CHD covalent product at a single arginine residue and 188.08 Da mass addition for arginine-CHD covalent products at the two arginine residues. All b and y fragment ions were also provided in the supplementary data as an MS Office Excel file.

2.4.2 Arginine modification of peptides by phenylglyoxal (PG)

The same procedure was followed to compare the reactivity of phenylglyoxal (PG), which is an another arginine reactive reagent. The reaction was carried out in 0.125 M KHCO$_3$ buffer. Incubation of 1 mM peptides with 30 mM PG produced various modified products with the model

peptides. After the reaction of PG with bradykinin (Fig. 2D), we saw a mixture of different products at m/z 1176.5950, m/z 1292.6224 and m/z 1310.6318. The m/z 1176.5950 [(M + H+αPG, see Fig. 2-1 for nomenclature)] matched with the mass addition of a single partial PG molecule (m/z 1176.5950 [peptide mass (1059.5614) + H (1.0078) + mass addition (116.0262) = 1176.5954, mass accuracy −0.33 ppm]). Moreover, we observed double partial PG modification [(M + H+2αPG)$^+$ = m/z 1292.6224] and also the full PG molecule after reaction with peptides (m/z 1310.6318[(M + H + PG)$^+$ = peptide mass (1059.5614) + H (1.0078)+ mass addition (250.0630) = 1310.6322, mass accuracy = −0.3 ppm]). We also observed two modified products after substance P reacted with PG (Fig. 2E); a peak at m/z 1463.7602 [(M + H+αPG)$^+$] for addition of single partial PG molecule and another peak at m/z 1597.7944 [(M + H+2 PG)$^+$] of low intensity for the addition of full PG molecules. A peak with addition of water was also observed next to m/z 1463.7602.

For neurotensin we also saw a mixture of product ions at m/z 1788.9444 [(M + H + αPG)$^+$] for the addition of a single partial PG molecule and also for the addition of double partial PG molecules [m/z 1904.9695 [(M + H + 2αPG)$^+$] (Fig. 2-2F). Full nomenclature of αPG (single partial PG modification), 2αPG (single partial PG modification in two different residues) and PG (full double modifications in arginine residues) are mentioned in Fig. 2.1. All the mass accuracy along with the structure of their modified peptides are provided in the supplementary data for PG.

In the MS/MS, spectra of modified peptides with PG (Fig. 2-4D, E, F) were analyzed to confirm the identification of the modified site. In Fig. 3D, the bottom MS/MS spectrum is for the parent ion m/z 1176.5950, which showed a single partial modification of PG at the arginine in the first arginine residue (R*PPGFSPFR). CID fragment ions b3-NH$_3$*, b6*, and b8* contained the partially modified arginine masses. The addition of the single partial PG molecule having a mass of 116.02 Da at the arginine residue can be confirmed by considering the unmodified peptide

fragments and modified peptide fragments. The mass difference between the modified fragment b8* and unmodified peptide fragment b8 showed the addition of a single partial PG modification at the arginine residue (m/z 1002.58–m/z 886.46 = 116.12) (Fig. 2-16, A). Unmodified peptide fragments y3, y5, y6, y7, and y8 were also identified and confirmed the modification on the N-terminal (Fig. 3D, bottom).

Peak at m/z 1292.6224 signifies the double partial PG modification at the two separate arginine residues. This modification was confirmed by the tandem mass spectrometry studies (Fig. 2-4D middle panel). The presence of partial PG modification at the N- terminal arginine was confirmed by the fragment ions b2*, b3-NH$_3$*, b6* and b8*. Additionally, the series of y fragment ions (y3*, y4*, y5*, y6*, y7*, y8*) indicated another PG modification at the arginine on the C-terminal of the peptide. By further analysis of the unmodified fragment ions (Fig. 2-16A), we clearly saw the mass addition of 116.12 Da for the fragment ions of the precursor m/z 1292.6224, which confirmed the addition of two partial PG (2αPG) molecules to both arginine residues.

One full PG modification at a single arginine residue illustrated a peak at m/z 1310.6318 in the MS spectrum in Fig. 2-4D. The existence of y5* and y7-H$_2$O* ion in the MS/MS spectrum (Fig. 3D, top panel) suggested that the PG modified product was formed at the arginine on the C-terminal. Ion fragment b4 defined the unmodified peptide fragment. If we compare Fig. 3D (modified peptide) with Fig. S8A (unmodified peptide), the mass difference between the modified fragment ion y5* and unmodified fragment ion y5 was 250.00 Da (m/z 903.34–m/z 653.34). Thus, it is very clear that mass addition was for the double partial PG modification at the single arginine residue. Fig. 3E provides a detailed interpretation of the MS/MS spectra of modified substance P, which contains one arginine residue. The data clearly shows the single partial PG modification and the full PG modification at the arginine residue. It is evident from the modified peptide presented at Fig. 3E and the unlabeled peptide In Fig. 2-18B that a signature

mass addition of 116.12 Da was present for most of the b series ions, which verifies the existence of a single partial PG modification at the arginine residues.

In Fig. 3F, for the modified neurotensin peptide, the bottom MS/MS spectrum is for the parent ion m/z 1788.9444, which shows a single partial modification of PG at the arginine in the 8th position-pyroglu-LYENKPRR*PYIL. This modification was confirmed by the MS/MS spectrum of the parent ion. A series of b and y fragment ions were observed (b3, b5, b6, b8*, b10*, y10-H_2O*, b11* and y5*, y7*, y7-NH_3* y8*, y9*, y9-NH_3*, y10-H_2O*). Fragmentation of the unmodified peptide gave ion mass for y5 at m/z 661.40 (see the supplementary Fig. 2-16C). Moreover, the presence of y5* ion at m/z 777.52 at the modified peptide in Fig. 3F (bottom panel) indicates that the modification is at the eighth position from the N-terminal of the peptide of the parent ion. The difference between the modified peptide fragment ion y5 and unmodified peptide fragment ion y5 is 116.12, which is the mass addition for a partial PG modification. Other PG modified products observed at m/z 1904.9695 can be explained by considering the CID fragmentation pattern (Fig. 3F top spectrum) of the parent ion. We observed the fragment ions at b3, b5, b7*, b8*, b8-NH3*, b11* and y5*, y6*, y7*, y8*, y8-NH_3* and y11-H_2O*. Identified fragment ions at b7* (m/z 1128.64), b8* (m/z 1400.86), y5* (m/z 777.52) and y6* (m/z 1049.74) indicate that the double partial PG modification took place rather than one full PG modification at a single arginine, which also gave the mass of m/z 1904.9699, after subsequent water loss during the reaction (exact mass 1921.9727 Da [M + H]$^+$−H_2O = 1904.9699). This finding can be used to verify the double partial modification at the two arginine residues in neurotensin. As mentioned before, an MS Office Excel file was provided with the fragment ions.

In contrast to the CHD modification, when we used PG as a modifying reagent and the labeling was not complete in one hour, we observed a mixture of product ions. Our studies always showed that one PG molecule reacted with one arginine residue. Increasing the time to 24

hours showed modification of all residues except arginine (data not shown). Clearly, CHD is more efficient in labeling arginine residues than phenylglyoxal.

2.4.3 Arginine modification of intact bovine serum albumin (BSA) by CHD

To identify the major sites of CHD modification in BSA, we carried out the reaction for the whole protein and the digests of protein. The same procedure was followed for bovine serum albumin fraction V, a serum protein of MW 66432 Da (UniProt P02769). After reaction at the different protocol time intervals (30 min, 1 h, 6 h, 12 h, 18 h, and 24 h), we saw selective labeling of arginine residues in BSA by CHD. According to the mass spectra (Fig. 2-5) of intact BSA modification studies, we identified three tryptic peptides, AWSVARLSQK (unmodified mass m/z 1145.6425), AWSVARLSQKFPK (unmodified mass m/z 1517.8587 with one missed cleavage) and peptide LGEYGFQNALIVRYTRK (unmodified mass m/z 2028.10246), which were modified after reacting with CHD. Both MALDI-QIT-TOF and LC-ESI-IT-TOF experiments were performed to confirm this labeling. In Fig. 2-5, the time course of MALDI-QIT-TOF-MS spectra was shown. In Fig. S1, ESI-MS/MS studies were conducted to reveal characteristics of the selected peptides. Our time course studies pointed out that the modifications were mostly on Arg-241, Arg-433 and Arg-436 residues. Tandem mass spectrometry of peptides AWSVAR*LSQK (m/z 1239.6831) and AWSVAR*LSQKFPK (m/z 1611.5463) confirmed the modification of Arg-241 residues (Fig. 5 and Fig. S1).

The most abundant peak in Fig. 2-5 was at m/z 1239.6831, which related to the modified peptide AWSVAR*LSQK. This peak was assigned to the modification of Arg-241. The MS/MS spectrum (Fig. 2-6A) of this modified peptide generated many fragment peaks. The majority can be assigned to b and y fragments of the peptide. The fragment ions b6*, b7*, b8*, b9*, y7* and y8* contained the modified arginine residue. We observed the same peptide with one missed

cleavage at m/z 1611.5463. The MS/MS spectrum of the parent ion m/z 1611.5463 gave the fragments b6*, b7*, b9* and b10* which contained the modified residue.
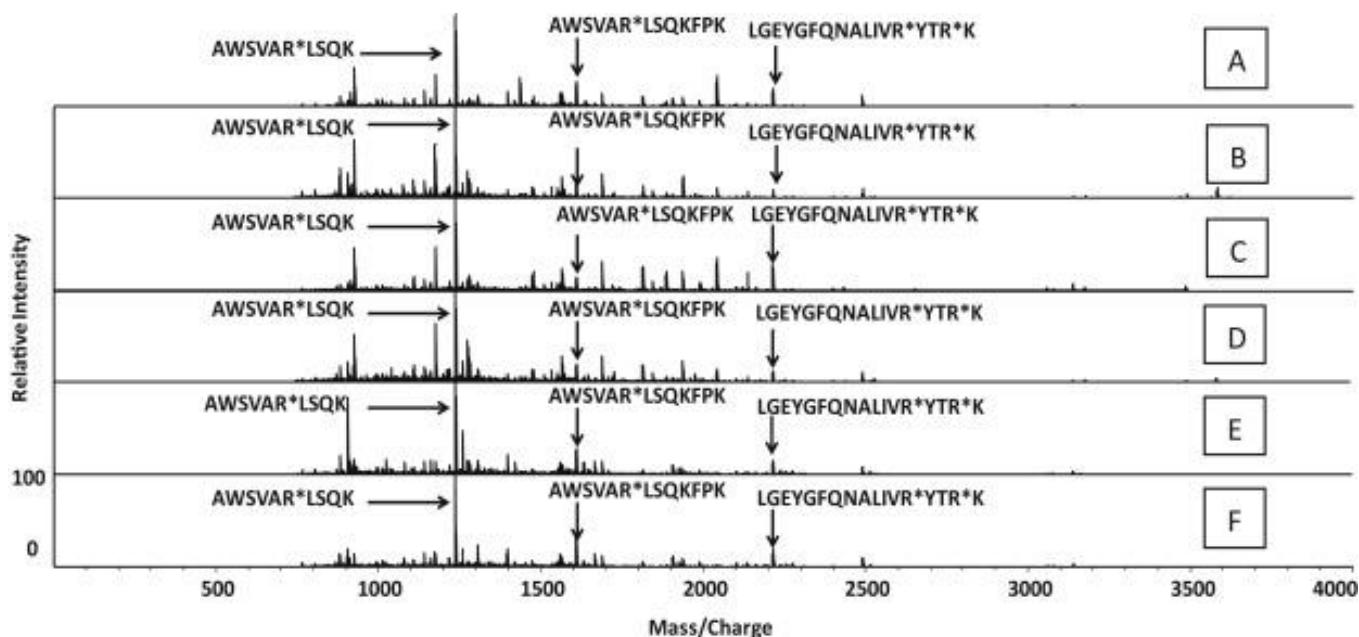
Fig. 2-5. MALDI-QIT-TOF MS spectra of modified bovine serum albumin (intact) with 1,2-cyclohexanedione at different time intervals: (A) 24 h, (B) 18 h, (C) 12 h, (D) 6 h, (E) 1 h, and (F) 30 min.

Both the MS/MS spectra given in Fig. 2-6A & B gave a fragment ion b5, which is an unmodified peptide fragment. As expected the fragmentation data from the two peptides AWSVAR*LSQK and AWSVAR*LSQKFPK demonstrate the mass addition 94.04 Da and verifies the formation of an arginine-CHD covalent product (Fig. 2-6A & B). ESI-LC-MS/MS of this modified peptide (Fig. 2-16-C) also showed that the fragment ions b8*, b9*, y10*, y8* and y6* were modified with CHD. After careful inspection of MS/MS spectra of peptide LGEYGFQNALIVR*YTR*K [m/z 2216.1813 and 740.0601 (3+)], we confirmed the modification of Arg-433 and Arg-436, respectively. Arg-241 is known as a glycation site in human serum albumin.
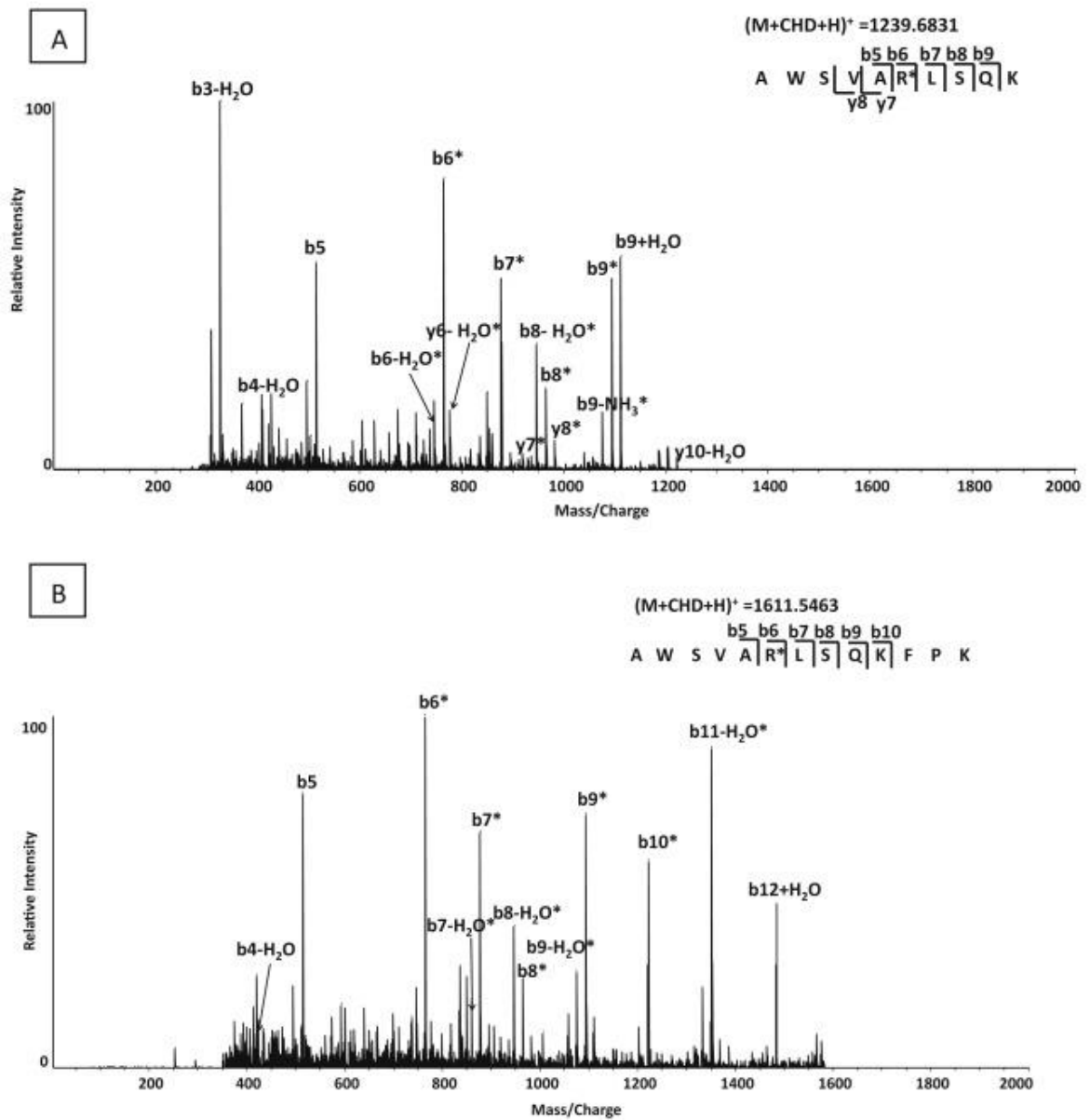
Fig. 2.6. MALDI-QIT-TOF MS/MS spectra of modified bovine serum albumin (BSA) peptides with 1,2-cyclohexanedione (A) MS/MS spectrum of peptide AWSVAR*LSQK at m/z 1239.6831 (B) MS/MS spectrum of peptide AWSVAR*LSQKFPK at m/z 1611.5463 ESI-MS/MS of peptide AWSVAR*LSQK are shown in Fig.S1 at m/z 413.8566(3+). (The symbol * in the figures designates the peptide fragments with modified arginine residues.

30

The work presented by Ahmed et al. in 2002[36] showed the locations of arginine glycated residues in human serum albumin (HSA). They were identified by tryptic peptide mapping by cationic electrospray LC-MS. In their report, they described the sites of modification of HSA by methylglyoxal under physiological conditions. They clearly pointed out the tryptic peptides containing arginine residues converted to MG-H1 residues in human serum albumin modified minimally by methylglyoxal. They reported modification of arginine residue at Arg-218. The peptides which contain the R218 have the amino acid sequence of AWAVAR(218)LSQR. We were able to identify this same peptide AWSVARLSQK with the 1,2-cyclohexanedione modification with BSA. The numbering of BSA, starting with the signal peptide, gives the arginine residue number as R241. But if the numbering starts from the pro-peptide, we get the number R217 because the amino acid sequence of human serum albumin and bovine serum albumin is significantly different. In human serum albumin, the peptide sequence, which contains a glycated arginine, is AWAVAR(218)LSQR, and in bovine serum albumin, the same peptide sequence is ALKAWSVAR(217)LSQK. However, in the study based on a search of reactive arginine residues in BSA, unmodified peptide peaks were also detected, along with modified peptides, (Fig. S9A). The two unmodified peptides YLYEIAR (m/z 927.7405) and RHPYFYAPELLYYANK (m/z 2045.2526) were identified. As shown in Fig. S9B & C, the b and y ion fragments in the CID-MS^2 spectrum provides evidence of the unmodified peptides. The CHD labeling study clearly helped us find reactive residues, which are susceptible to modifications in the disease process. Moreover, a quantitative study on this kind of work can be performed with the isotope labeling method followed by a protease treatment. In that case, a different protease such as Lys-C can be utilized; this avoids the cleavage at the arginine residues.

We also obtained the same reactions in a digest of bovine serum albumin (BSA) control experiments in BSA peptides. They showed more reactivities in arginine residues compared to

the intact protein studies. These experiments were also provided in the supplementary data in detail (page S1 in the supplementary data and Figs. S3 and S4).

2.4.4 Arginine modification in lysozyme

According to Patthy and Smith, all 11 arginine residues in egg white lysozyme were getting modified with CHD.[41] Arg-5 is the most reactive arginine residue, and according to the x-ray diffraction studies, it is included in the α helical structure. It is also found that Arg-5 has relative importance in maintaining an active conformation[63]. Therefore, there is a strong interest in characterizing the immunochemical features of Arg-5. We showed the mass spectrum of the peptide from the reactive sites derived from intact lysozyme labeling and have provided documentation in the supplementary data (Figs. S5, S6). For the whole egg white lysozyme, we found Arg-5 was also modified (VFGR*CELAAAMK, m/z at 1389.7098, unmodified monoisotopic mass 1295.6598 Da, difference 94.05 Da). These modification results are consistent with the results of Patthy and Smith using 1,2-cyclohexanedione in the borate buffer.
[41]Fig. S10-A, displays a MALDI-QIT-TOF mass spectrum of modified hen egg lysozyme with 1,2-cyclohexanedione. Along with the modified peptides, the peaks at m/z 1045.6924 and m/z 1675.8520 correspond to the unlabeled peptides GTDVQAWIR and IVSDGNGMNAWVAWR, respectively. The CID-MS/MS spectra of the unlabeled peptides are shown in Fig. S10,B&C. The b and y fragment ions observed in the CID-MS/MS experiment confirmed the presence of unlabeled arginine residues in the peptides.

2.4.5 Structural analysis of reactive arginine residues

To evaluate the correlation between the modification site and the protein structure, we conducted the analysis of reactive residues (PyMOL version 1.3, 2009–2010) in the published crystal structure of proteins. In Fig. 2-7, we show the structural model of CHD modified proteins.
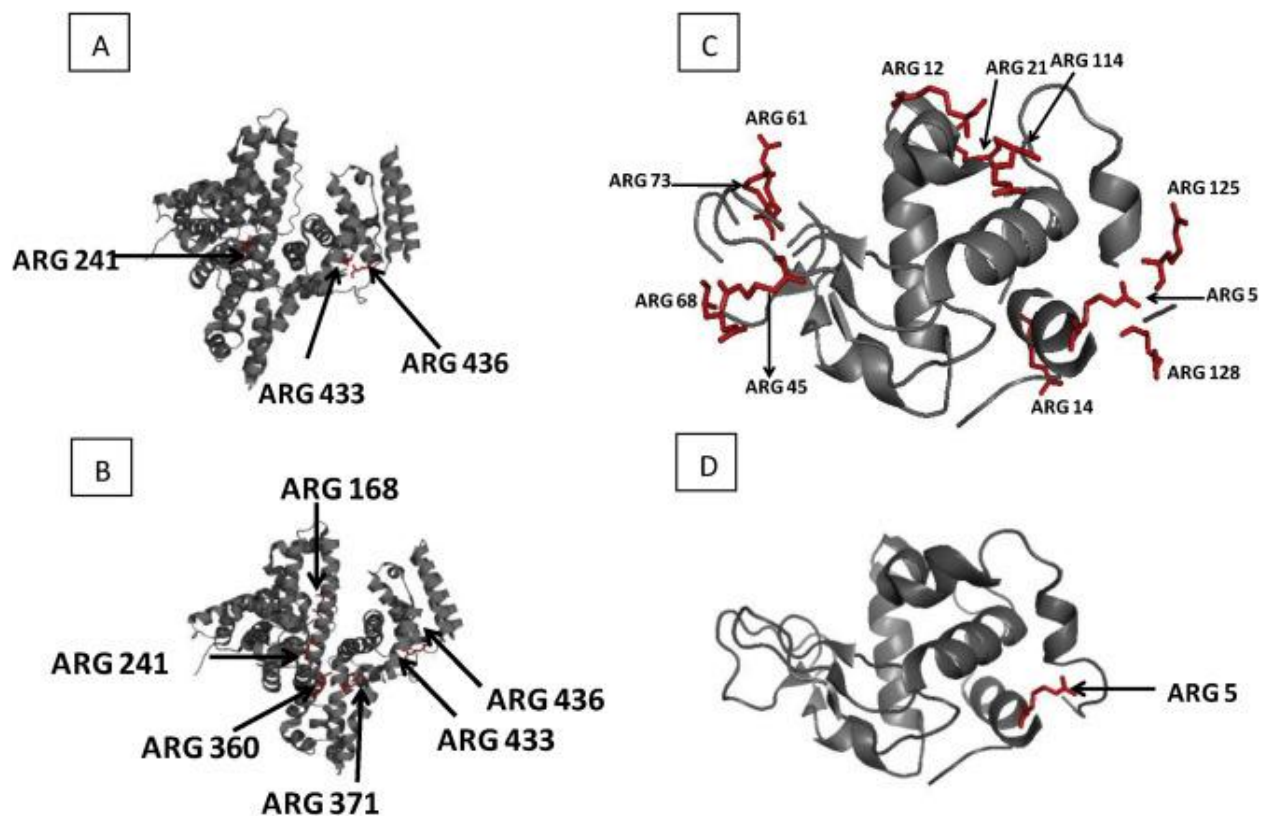
Fig. 2-7. Crystal structure of proteins, BSA (PDB file 4F5S) and lysozyme (PDB file 7LYZ). (A) Modified sites of whole BSA with CHD, (B) modified sites of digested BSA, and (C) all arginine residues in hen egg lysozyme, and (D) most reactive arginine 5 in hen egg lysozyme

In Fig. 2-7A, we show the modified sites of BSA in the intact protein analysis. A few selective residues such as arginine 241, 433 and 436 were labeled quickly while digested protein showed more nonspecific labeling (Fig. 6B). All arginine residues in hen egg lysozyme are shown in Fig. 2-7C and most reactive residues (Arg 5) are shown in Fig. 2-7D. We carefully analyzed the structures and mainly saw the arginine residues in the surface of the protein getting modified. This labeling study will clearly pinpoint solvent accessible arginine and their local environment. If there are strong interactions with other residues, they will not be available for reactions. There

are some concerns about the structural studies using CHD due to buffer and pH. Some of the residues are identified which were also identified on the cellular level with other methods. We believe this structural information will be extremely useful for further exploration of enzyme activities by exploring these residues. These results provide strong evidence that CHD labeling is very selective and efficient in the proteins.

2.5 Conclusions

In this study we showed a clear-cut demonstration that the arginine selective reagent CHD is a better labeling reagent than PG for finding reactive residues in proteins and for analyzing enzyme activities and protein modification sites without ambiguities. The work presented here demonstrates the effectiveness of mass spectrometry analysis of modified peptides. Model peptides were labeled with CHD and PG which contained single and double arginine residues. Detailed MS/MS spectra of modified peptides are shown. We also demonstrated the reactivity of reagents in intact proteins by time course labeling studies. Time course labeling studies clearly pointed out reactive arginine residues in BSA proteins and lysozyme. This study is also useful for selectively identifying the protein glycation sites in a BSA protein. This mass-spectrometric study is beneficial due to its high-confidence identification of the reactive sites, which can be validated in large-scale samples by targeted proteomics using mass spectrometry.

2.6 Acknowledgements

funding from University of Texas proteomics facilities networks, which was awarded to my dissertation supervisor. It is funding like this that makes theses and dissertations possible. My thanks go to all who have made and continue to make research in this area possible in the Department of Chemistry and Biochemistry at the University of Texas at Arlington.

2.8 Supporting information

2.4.6 Arginine Modification in BSA Digests by CHD

To find the reactive arginine residues with CHD in a digest of BSA, we successfully derivatized the BSA peptides with CHD. In the BSA peptides and CHD reactions, we saw where new arginines were modified along with the same arginine residues, which were modified before. More peptides were assigned for the arginine-CHD covalent product compared to the intact BSA protein. It was expected that intact proteins would provide the arginine residues, which are more reactive; whereas, digested peptides mostly label all the arginine residues. For the digests of BSA, Arg-168, Arg-360 and Arg-371 are getting modified in addition to other arginine residues, which were modified in intact BSA (Fig. 2-S8). We have seen how the peptide AWSVAR*LSQK (*m/z* 1238.6831) and AWSVAR*LSQKFPK (*m/z* 1611.5463) were modified, which were the most reactive arginine in intact BSA studies (Arg-241 residue) (see Fig. 2-5). In addition to this, peptide LGEYGFQNALIVR*YTR*K (*m/z* 2216.2021), which was modified at Arg-433, and Arg-436 residues and peptides R*HPEYAVSVLLR*(*m/z* 1627.9174) were seen modified at Arg-360 and Arg-371 (Fig. 2-S3). Modification by CHD on Arg-168 residue was confirmed by a respective mass shift and MS/MS of the modified peptide R*HPYFYAPELLYYANK at *m/z* 2139.0937 (Fig 2-S6). All the MS/MS fragmentation patterns of modified peptides gave clear evidence for the modified sites (Fig. 2-S9). It is also clear from the

studies of intact BSA and digested BSA that reactivity signature is different in intact BSA. Highly

reactive arginine residues were clearly modified in intact BSA at first.
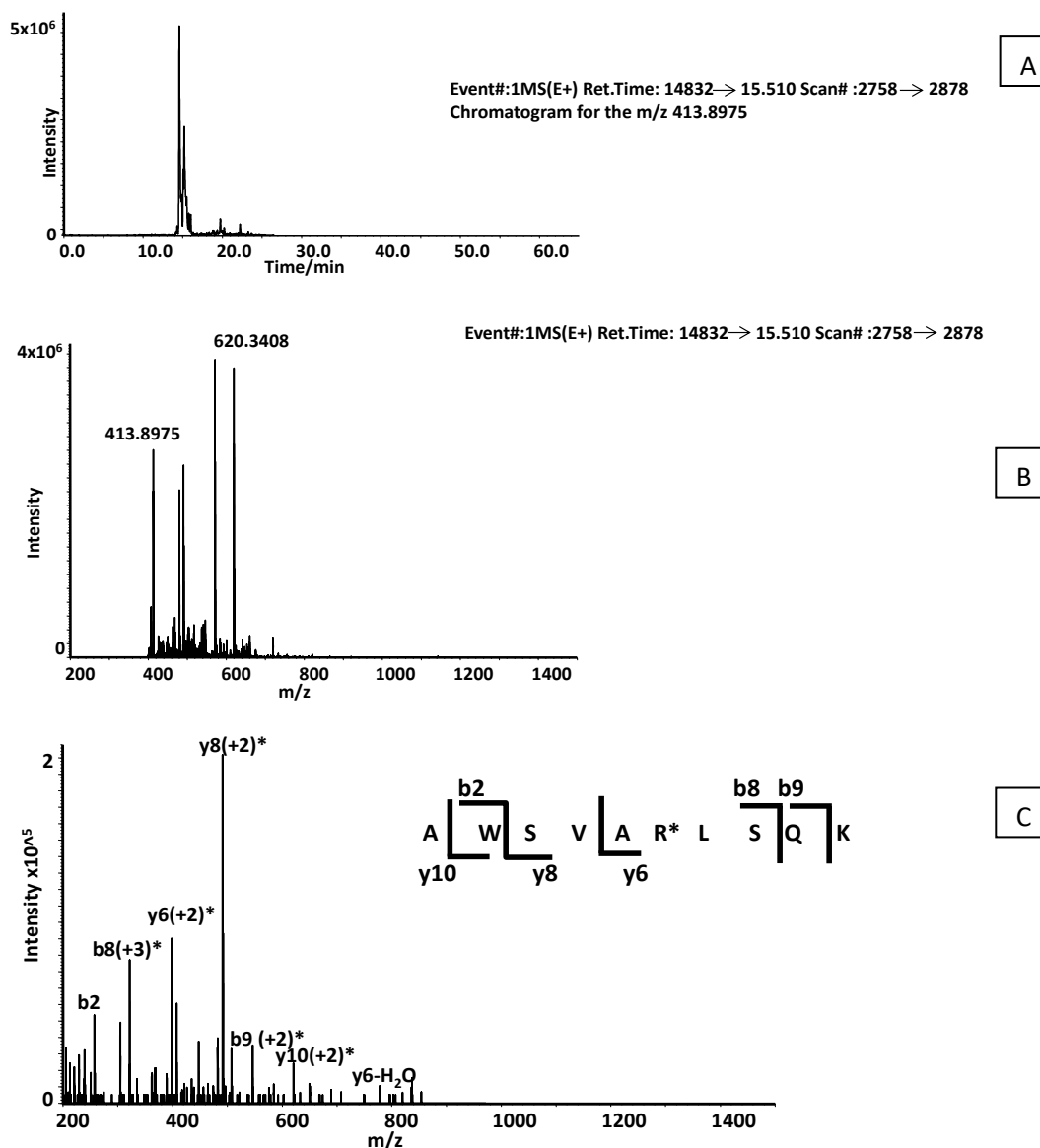
Figure 2-S1. ESI-LC-MS/MS spectra of the modified BSA (intact labeling) with CHD in 1 h time intervals. (A) Retention time of modified BSA peptide AWSVAR*LSQK at m/z 413.8975(+3), (B) mass spectrum of that region, and (C) MS/MS of m/z 413.8975 $(M+3H+CHD^*)^{3+}$, which confirm the modified peptide sequence. (The symbol * in the figures designates the peptide fragments with the modified arginine residues).
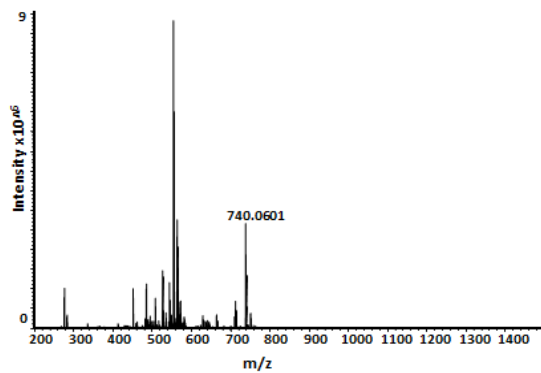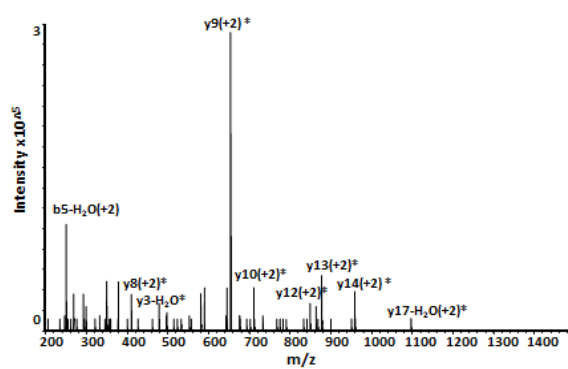
Figure 2-S2. ESI-LC- MS/MS spectra of modified bovine serum albumin (intact labeling) peptides with 1, 2-cyclohexanedione: (A) MS spectrum of LGEYGFQNALIVR*YTR*K at m/z 740.0601(+3) and (B) MS/MS spectrum of LGEYGFQNALIVR*YTR*K at m/z 740.0601 $(M+3H+2CHD)^{3+}$. (The symbol * in the figures designates the peptide fragments with the modified arginine residues).
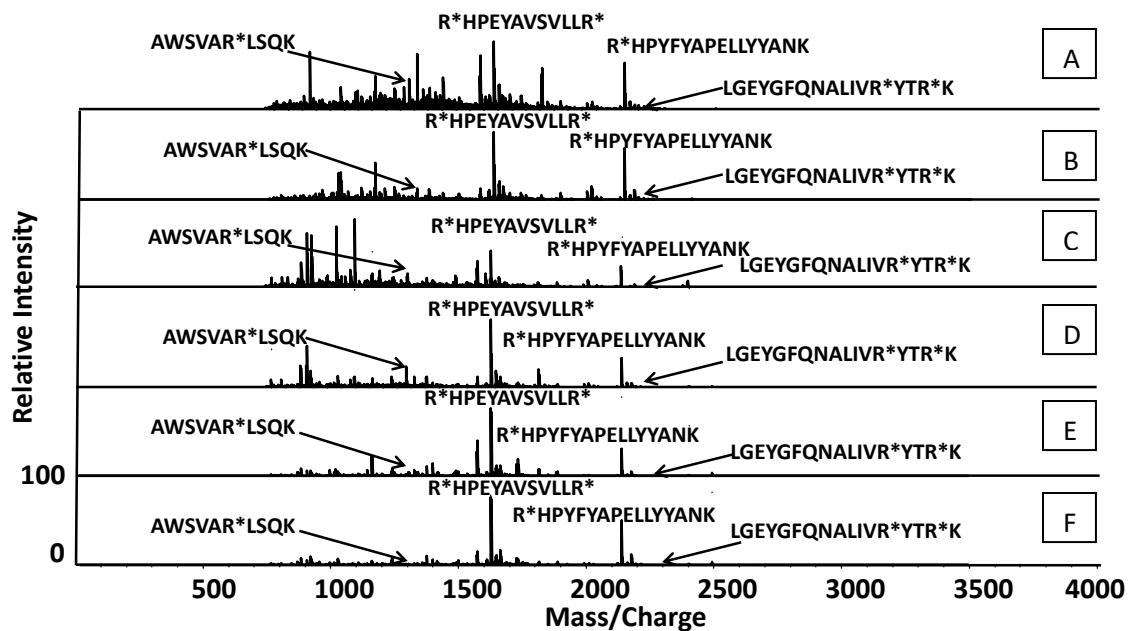
Figure 2-S3. MALDI-QIT-TOF MS spectra of digested bovine serum albumin with 1, 2-cyclohexanedione in different time intervals. (A) 24 hours, (B) 18 hours, (C) 12 hours, (D) 6 hours, (E) 1 hour, and (F) 30 min

37
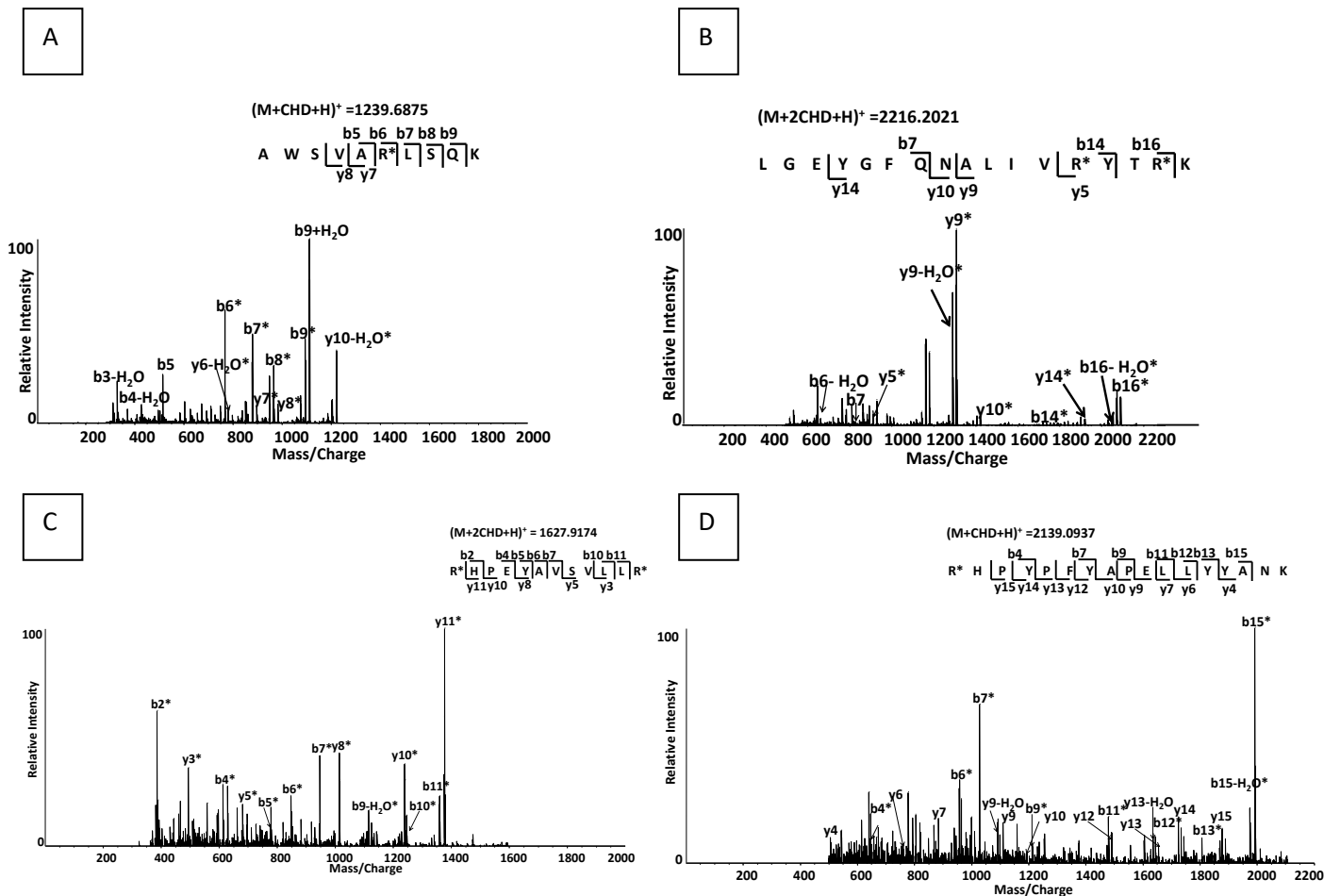
Figure 2-S4. MALDI-QIT-TOF MS/MS spectra of digested bovine serum albumin peptides with 1, 2-cyclohexanedione. (A) MS/MS spectrum of peptide AWSVAR*LSQK at m/z 1239.6875, (B) MS/MS spectrum of LGEYGFQNALIVR*YTR*KI at m/z 2216.2021, (C) MS/MS spectrum of peptide R*HPEYAVSVLLR*(m/z 1627.9174), and (D) MS/MS spectrum of peptide R*HPYFYAPELLYYANK (m/z 2139.0937). (The symbol * in the figures designates the peptide fragments with the modified arginine residues).

Table 2-1 Comparison of the mass accuracy of CHD and PG modified peptides

Peptide	Mass accuracy of Unmodified peptide/ppm	Mass accuracy with CHD/ppm	Mass accuracy with PG-Single partial modification/ppm	Mass accuracy with PG-Double partial modification/ppm	Mass accuracy with PG –Full PG modification/ppm
Bradykinin	3.3	− 2.4	-0.33	0.5	− 0.3
Substance P	2.5	9.3	− 1.3	NA	− 2.8
Neurotensin	16	7.6	0.39	− 0.2	NA
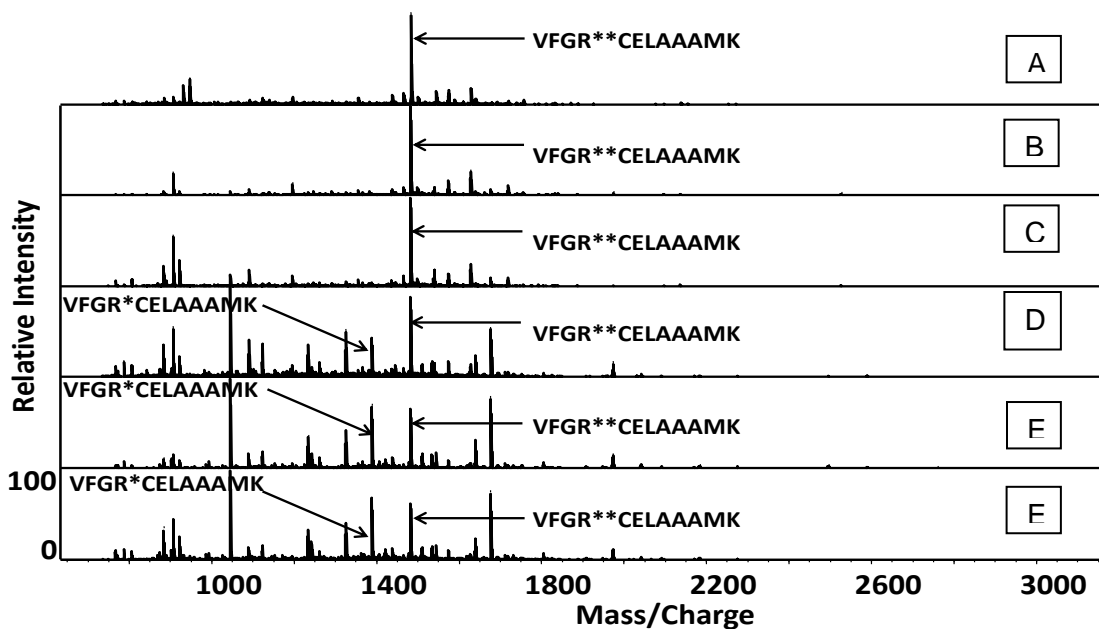
Figure 2-S5. MALDI-QIT-TOF MS spectra of modified hen egg lysozyme with 1,2-cyclohexanedione in different time intervals. (A) 24 hours, (B) 18 hours, (C) 12 hours, (D) 6 hours, (E) 1 hour, and (F) 30 minutes.
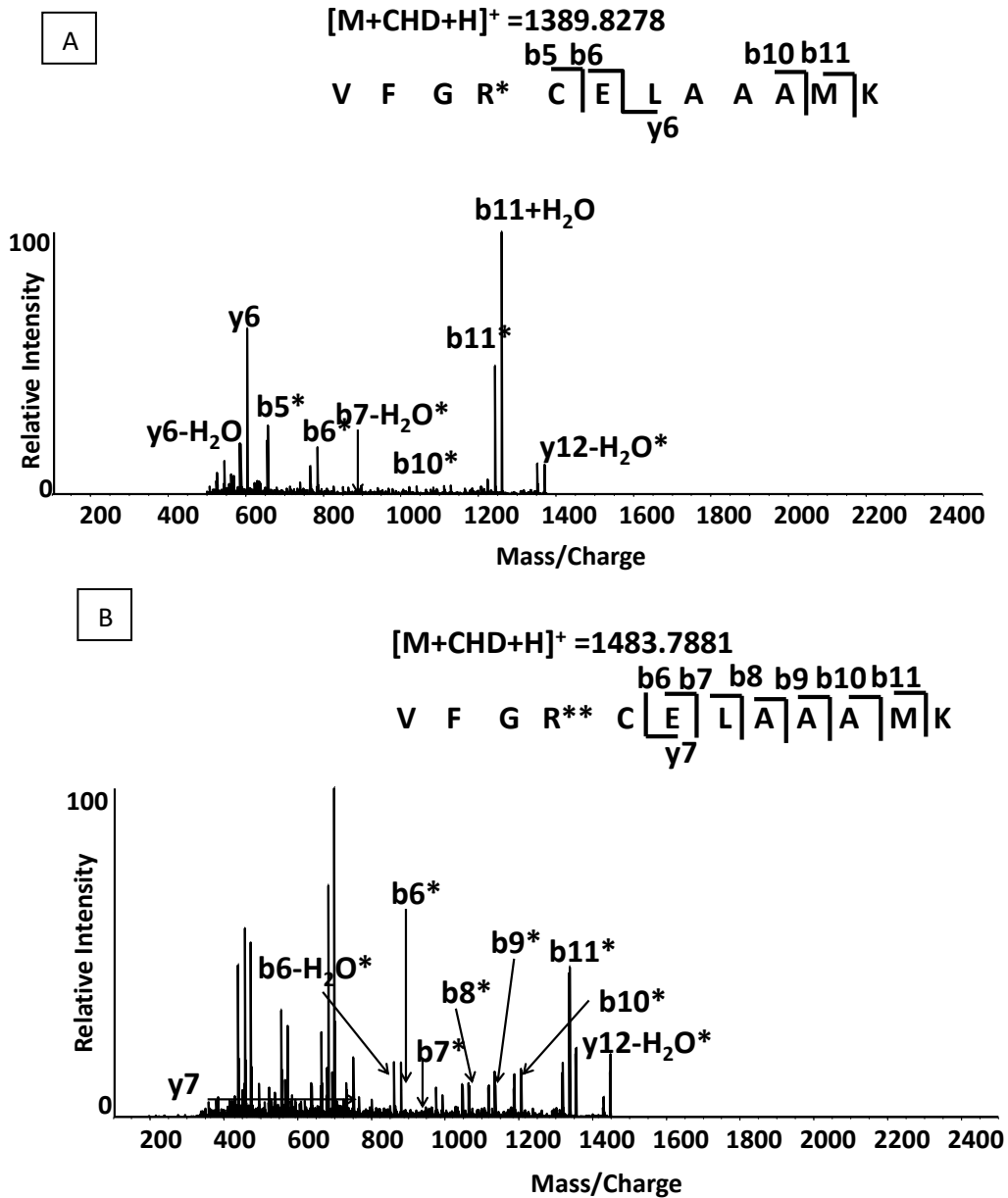
Figure 2-S6. MALDI-QIT-TOF MS/MS spectra of modified hen egg lysozyme peptides with 1,2-cyclohexanedione. (A) MS/MS of VFGR*CELAAAMK at m/z 1389.8278, and (B) MS/MS of VFGR**CELAAAMK at m/z 1483.7881. (The symbol * in the figures designates the peptide fragments with the modified arginine residues).

Table 2-2 Comparison of the relative intensity changes with the reaction time

Kinetic plot for Bovine serm albumin reation with 1,2 Cyclohexanedione					
AWSVAR*LSQK - (M+H)+ = 1239.68					
MALDI-QIT-TOF DATA					
Time (hr)	Relative Intensity				
0.5	8558				
1	20565				
6	22786				
12	29498				
18	30388				
24	54626				
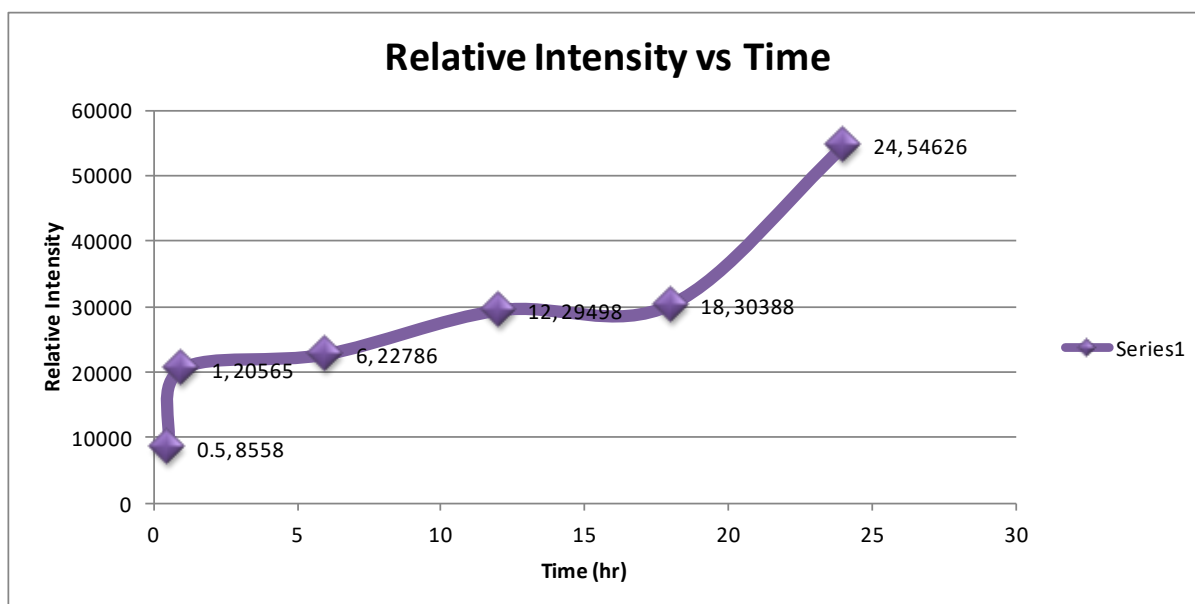
Figure 2-S7. Kinetic plot of the relative peak intensity (normalized) vs. time for the modified peptide AWSVAR*LSQK in MALDI-QIT-TOF.
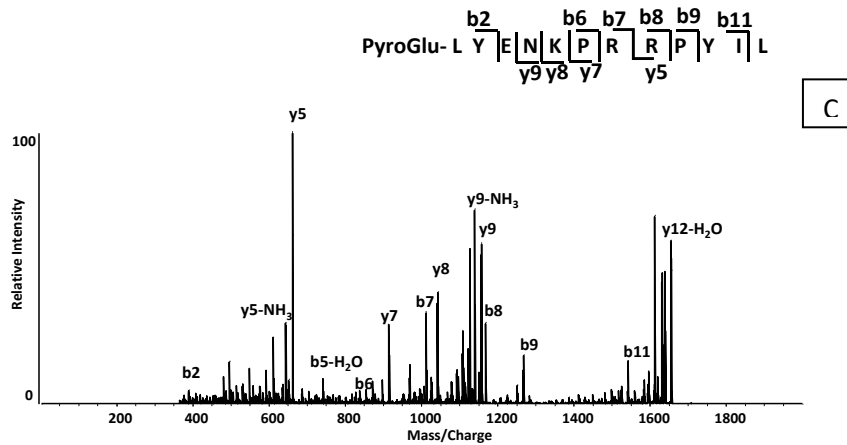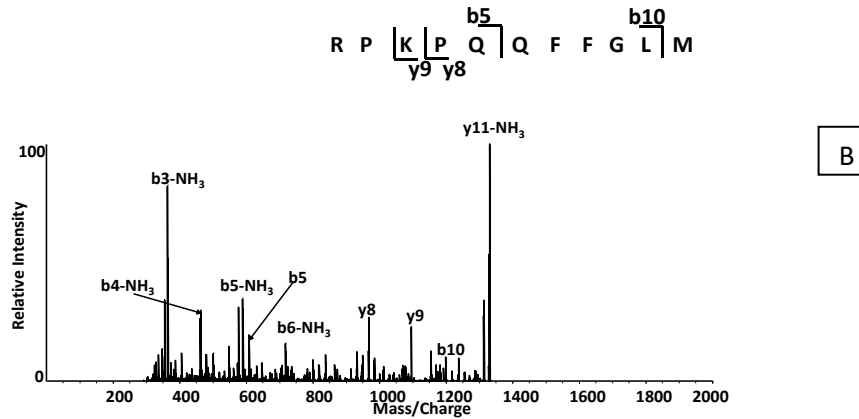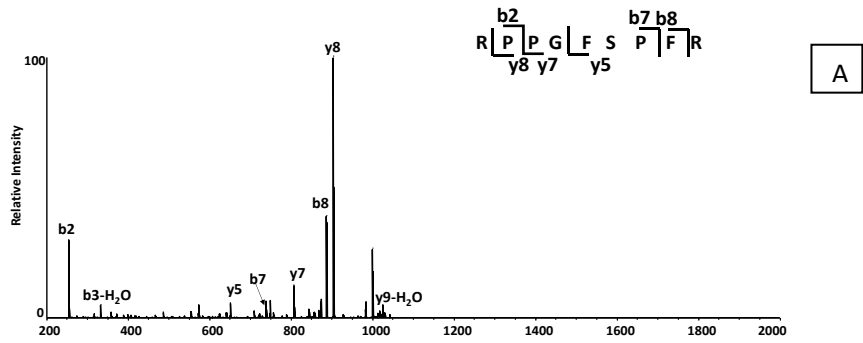
Figure 2-S8. MALDI-QIT-TOF MS/MS spectra of pure peptides. (A) MS/MS spectrum bradykinin at m/z 1060.5727, (B) MS/MS spectrum of substance p at m/z 1347.7394, and (C) MS/MS spectrum of neurotensin at m/z 1672.8907.
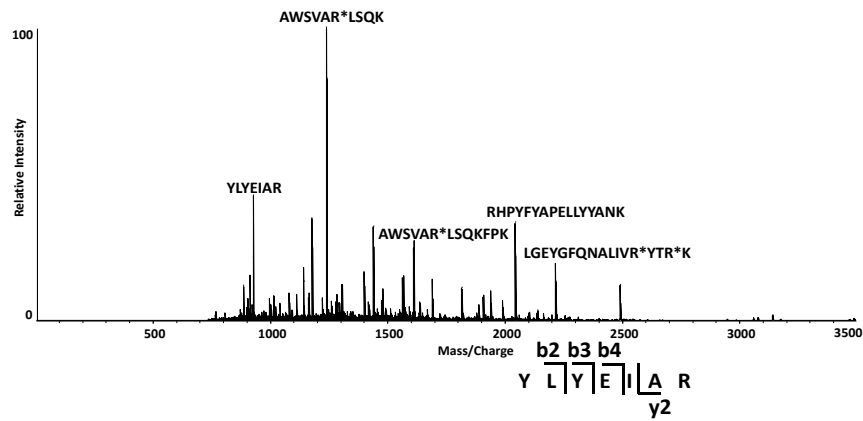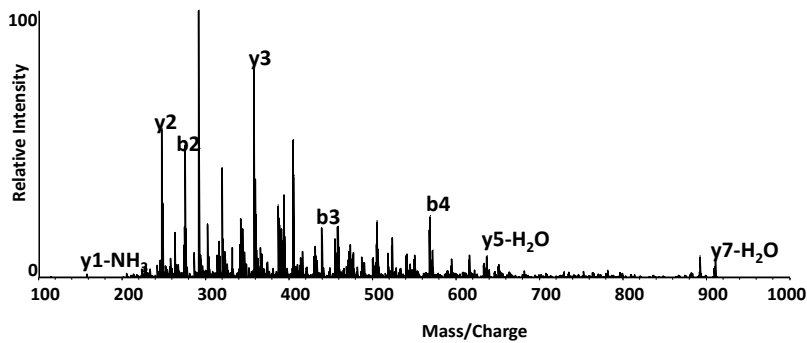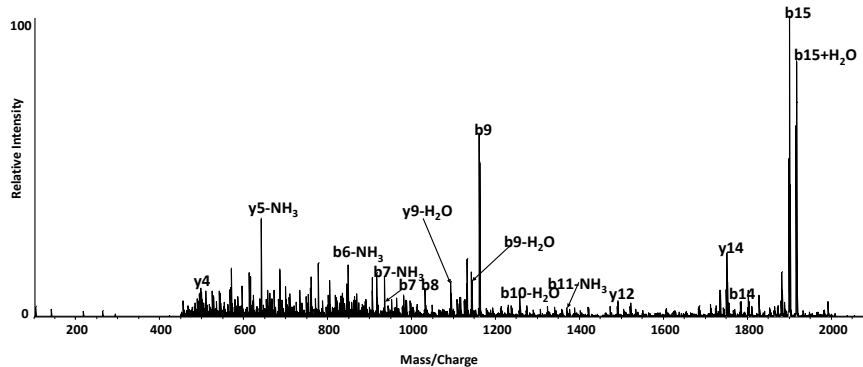
Figure 2-S9. (A) MALDI-QIT-TOF MS spectrum of modified bovine serum albumin (intact) with 1,2-cyclohexanedione, (B) MALDI-QIT-TOF MS/MS spectrum of unlabeled peptide YLYEIAR at m/z 927.7405, and (C) MALDI-QIT-TOF MS/MS spectrum of unlabeled peptide RHPYFYAPELLYYANK at m/z 2045.2526. (The symbol * in the figures designates the peptide fragments with the modified arginine residues).

43
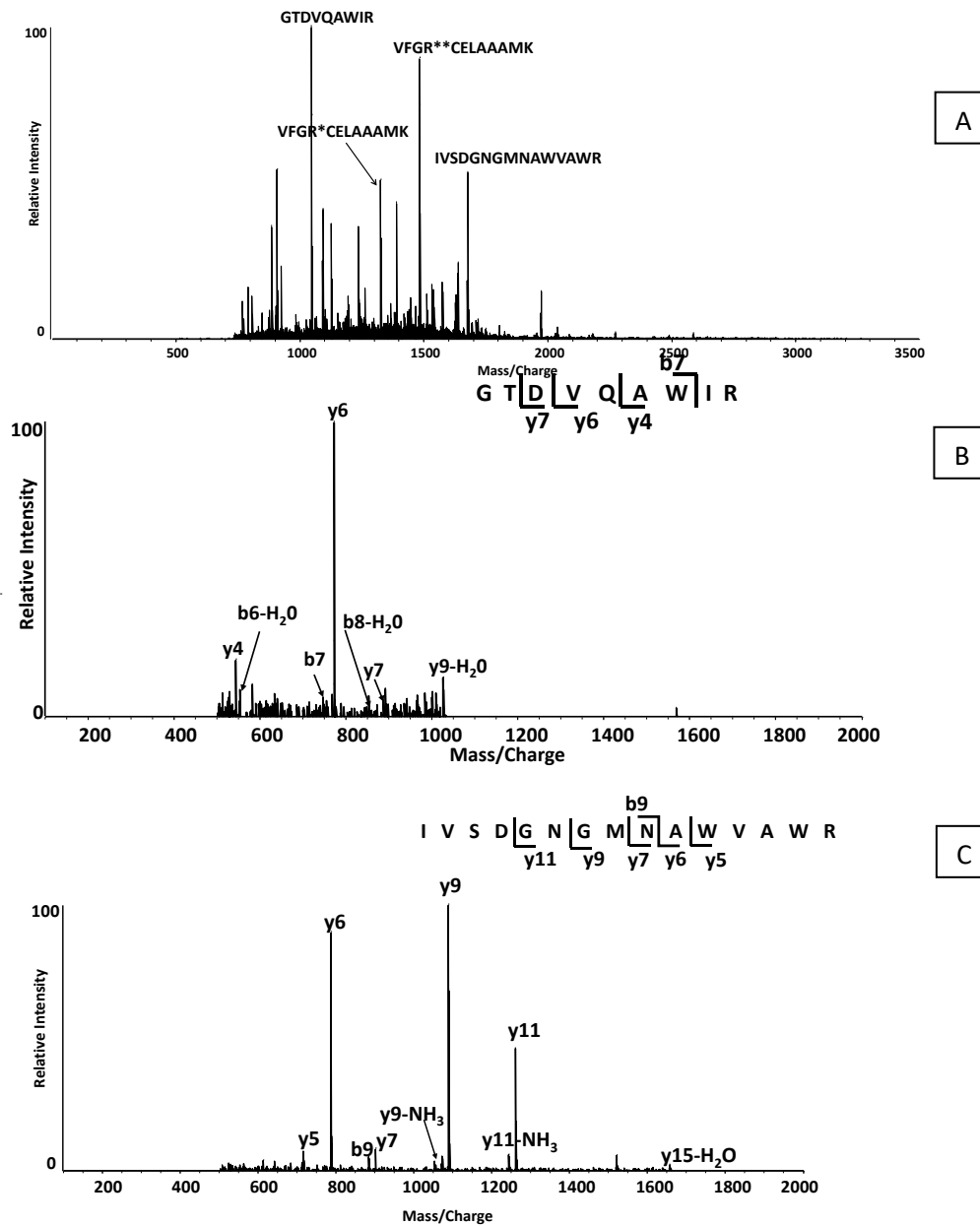
Figure 2-S10. (A) MALDI-QIT-TOF MS spectrum modified hen egg lysozyme (intact) with 1,2-cyclohexanedione (B) MALDI-QIT-TOF MS/MS spectrum of unlabeled peptide GTDVQAWIR at m/z 1045.6924, and (C) MALDI-QIT-TOF MS/MS spectrum of unlabeled peptide IVSDGNGMNAWVAWR at m/z 1675.8520. (The symbol * in the figures designates the peptide fragments with the modified arginine residues).

44

Chapter 3

Enrichment and identification of reactive arginine residues in proteins by bio-orthogonal click chemistry and mass spectrometry

3.1 Abstract

Modification of arginine residues using dicarbonyl compounds is a common method to identify functional arginine residues of proteins or to inactivate enzymes. Arginine also undergoes several kinds of post-translational modifications (PTMs) in these functional residues. Identifying these reactive residues globally is a very challenging task. Several dicarbonyl compounds have been utilized and the most effective ones are phenylglyoxal and cyclohexanedione. However, tracking these reactive arginine residues in a protein or in large-scale protein samples using the labeling approach is very challenging. Thus, we enriched modified peptides to provide reduced sample complexity and confident mass-spectrometric data analysis. To enrich arginine labeled peptides efficiently, we developed a novel arginine selective enrichment reagent. For the first time, we conjugated an azide tag in a widely used dicarbonyl compound cyclohexanedione. This provided us with the ability to enrich modified peptides using a bio-orthogonal click-chemistry and biotin-avidin affinity chromatography. We tested the reagent in several standard peptides and proteins. Three standard peptides, bradykinin, substance P and neurotensin were labeled efficiently with a cyclohexanedione-azide reagent. Enrichment success of the modified peptides was tested by spiking the peptides in a myoglobin digest. A protein, RNase A, was also labeled with the reagent and after click-chemistry and biotin-avidin affinity chromatography; we identified two selective arginine residues. We believe this strategy will be an efficient way to identify functional and reactive arginine residues globally in complex protein mixtures.

3.2 Introduction

Arginine is one of the basic amino acids of proteins and its role in the functionality of protein structures, enzyme activity, and protein interactions is very important. Arginine goes through several kinds of posttranslational modifications during the cellular processes. Among them, glycation is a non-enzymatic modification of proteins, which is the result of the addition of a sugar molecule to a protein by Maillard reactions. In this reaction, nucleophilic amino groups of amino acids react with carbonyl groups of the sugar to form a Schiff-base product. This unstable Schiff base intermediate rearranges to a stable adduct known as an Amadori product, which is a keto-amine compound. This irreversible covalent modification can occur particularly at protein arginine residues by carbonyl compounds resulting in the formation of advanced glycation end-products (AGEs).[38,64,65] AGEs buildup on tissue proteins, which leads to many pathologies including diabetes mellitus, Alzheimer's disease and atherosclerosis.[66-68] AGEs can be formed not only from the reducing sugars but also from dicarbonyl compounds such as glyoxal and methylglyoxal, which are intermediates of several cellular processes. These dicarbonyl compounds are more selective to the guanidine, amino, and thiol group of proteins and nucleic acids.[69] Methylglyoxal modifications have also been observed in arginine residues in recombinant antibody developed for protein therapeutics.[70] Muranova et al. has reported a study on the effects of methylglyoxal modification on the human heat shock protein HspB6 (Hsp20). Their experiment was performed under two different conditions, and it was found that at low MGO, HspB6 molar ratio, the main MGO modified sites were Arg-13, Arg-14, Arg-27, and Arg-102. Muranova et al. could see that all the arginines and lysines residues were modified at high MGO: HspB6 ratio.[71] Methylglyoxal reacts with arginine, and arginyl residues in proteins were reported by Takashi in 1977, by Cheung and Fonda in 1979 and in 1993 by Selwood and Thornalley.[52,72,73]

Brock et al. first addressed the study of site specificity in AGE formation on ribonuclease, and they found that the main site of carboxymethylation was at Lysine-four.[74] In 2004, Cotham et al. reported that the glyoxal-derivative formation takes place mainly at arginine-39 and arginine 85.[42] Identifying the specific binding sites of proteins play a significant role in drug discovery. Chemical modification is a powerful technique for identifying these reactive residues. The reactivity of arginine towards dicarbonyl groups were recently evaluated thoroughly by our research groups.[75] In the cellular process, these dicarbonyl compounds mainly modify arginine residues, which are functional, reactive or surface accessible. Locating these modifications in large-scale samples is very challenging due to sample complexity after digestion. This dissertation study was aimed at the development of an enrichment strategy of the reactive arginine containing peptides. We introduced an azide tag in the widely known arginine reactive reagent, cyclohexanedione (CHD); hence, reactive peptides can be enriched utilizing bio-orthogonal click chemistry. Adding an enrichment tag in dicarbonyl compounds is very challenging due to the difficulty of synthesis. To the best of our knowledge, for the first time, we have developed an arginine selective reagent with an azide functionality. Azides are known to be bio-orthogonal reagents, which do not undergo any side reactions with the functional groups present in the proteins.[76,77] It selectively reacts with alkynes to form a triazole product under the azide-alkyne cycloaddition reactions to click chemistry.[78,79] We performed the chemical labeling of the arginine residues in peptides and proteins using our arginine selective reagent. Finally, chemically modified peptides were then enriched by affinity purification with avidin−biotin coupling. We demonstrated selective labeling of reactive arginine residues in RNase A protein. We believe this chemical approach will help us study reactive or functional arginine's in proteins in large-scale studies so targeted proteomics studies can evaluate PTMs on these residues.

3.3 Experimental section

2.3.1 Materials.

Ribonuclease A from bovine pancreas (RNase A), ubiquitin (from bovine erythrocytes), ammonium bicarbonate (NH_4HCO_3), formic acid (FA), sodium hydroxide (NaOH) pellets, acetonitrile (ACN), and iodoacetamide (IAM) were purchased from Sigma Aldrich (St. Louis,MO, USA). Biotin-PEG4-Alkyne was from Click Chemistry Tools, Scottsdale, AZ. For proteolysis, sequencing-grade modified trypsin was purchased from Promega (Madison, WI, USA). For disulfide bond reduction, dithiothreitol was obtained from Bio-Rad (Hercules, CA). Pierce C18 Zip Tips from Thermo Fisher Scientific (Rockford, IL, USA) were used to desalt the samples. Phosphate buffer saline (PBS) came from VWR (Suwanee GA). Three model peptides bradykinin, neurotensin, and substance P were purchased from ANASPEC (Fremont, CA). To remove excess CHD-Azide, Pierce concentrators with 3K MWCO were utilized. Pierce Ultra Link® monomeric avidin was obtained from Thermo Fisher Scientific (Rockford, IL). All SDS–PAGE supplies were purchased from Bio-Rad (Hercules, CA).

3.3.2 Sample preparation

3.3.2.1 Chemical modification of arginine residues in peptides

 Bradykinin, neurotensin, and substance P were purchased from ANASPEC (Fremont, CA). Recently, we reported an assessment of our chemical labeling method to identify functional/reactive arginine residues of proteins by mass spectrometry by utilizing two widely used arginine reactive reagents. We followed the same labeling protocol mentioned in our recently published article. For each peptide sample (1 mM, 5 µl) we added 5 µl of 30 mM CHD-Azide reagent in 200 mM Sodium hydroxide solution. The reaction was allowed to continue for 2 hours at 37 ˚C under agitation. After the reaction, the samples were dried in a Speedvac at 30 °C for 1 h. The two dried samples were then reconstituted in 0.1% FA. After that samples were

vortexed and desalted using the Thermo Scientific Pierce C18 Zip Tips. Desalted samples were then used for the analysis after diluting1:1 MeOH:H$_2$O with 2% acetic acid.

3.3.2.2 Chemical modification of arginine residues in proteins

Each protein sample was reduced and alkylated using 2 µl of 10 mM dithiothreitol (Bio-rad, CA) and 10 µl of 10 mM iodoacetamide (Sigma-Aldrich, MO), respectively. Each sample was then diluted with 200 mM sodium hydroxide followed by the addition of 5 µl of 30 mM CHD-Azide solution. The reaction was carried out at 37 ˚C for 24 hours. Trypsin digestion was performed overnight at 37 ˚C. Samples were transferred into the pierce concentrators (3 kDa MWCO) for centrifugal ultrafiltration to remove the excess CHD-Azide reagent. Cleared protein samples were digested with trypsin, using 1:100 enzyme-to-protein ratio, for 16 h at 37 °C with frequent mixing. After trypsin digestion the peptide mixture was fully dried and reconstituted in a PBS buffer. At the same time, the azide-containing peptide sample was prepared for the click chemistry reaction by adding 2 µL of 100 mM water-soluble tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) click ligand, 400 µL of 50 mM CuSO$_4$, and 80 µL of 50 mM Tris (2-carboxyethyl) phosphine(TCEP). The enrichment reagent, Biotin-PEG4-Alkyne, was added in a 1:20 molar ratio of peptide and allowed to react for two hours at room temperature in a rotor. The click-labeled sample was then incubated with 30 µL of monomeric avidin beads. Prior to the incubation, the avidin beads were washed with 1x PBS buffer three times. Sample was then incubated with the avidin beads at room temperature for 4 hr. The beads were washed again with the PBS buffer three times, twice with 25 mM NaCl and three times with water. Finally, the captured peptides were eluted from the avidin beads using the elution solution, which contains 50:50 acetonitrile: H$_2$O and 0.4% TFA.

3.3.2.3 The feasibility of click chemistry-based peptide enrichment

To study the feasibility of click chemistry-based peptide enrichment, we performed a spiking study. The CHD-azide modified substance P (1mM 5ul) was spiked into a tryptic digest of myoglobin (1 mM, 5 µl). After the click labeling reaction, the CHD-azide modified peptides were captured by avidin beads following the same protocol mentioned above. Then, the click labeled biotinylated peptide was eluted using the same elution solution containing 50% acetonitrile/50% H_2O and 0.4% TFA.

3.3.2.4 Instrument Set up

The samples were analyzed on the Thermo Scientific Velos Pro dual-pressure linear ion trap mass spectrometer operated in positive ion mode and controlled by Xcalibur™ software (Thermo Fisher Scientific, San Jose, CA, USA). Peptides were loaded onto a Dianex Acclaim[R] PepMap™ C18 column with an inner diameter (i.d.) of 75 µm, particle size of 3 µm, pore size of 100 Å and a length of 15 cm bed for the separation. Reverse phase chromatographic separations of the loaded peptides were performed on the Dianex Ultimate 3000 RSLC nano chromatography (Thermo Fisher Scientific) system across a 65-min run with a flow rate of 300 nl/min. A multi-step gradient was used consisting of 4% B for 3 min, 4 to 30% B in 30 min, 30% to 60% B in 55 min, (A, 0.1% formic acid in water; B, 95%: 5%: 0.1% Acetonitrile: water: formic acid 0.1% formic acid). The mass spectrometer was operated in normal scan modes and a full MS spectrum was obtained. Peptides were identified in the data dependent acquisition (DDA) mode to obtain the tandem mass spectra (MS/MS) for the ten most abundant ions (normalized collision energy of 40%, activation Q of 0.25, and activation time of 10 ms). In DDA analysis, singly charged ions were rejected. To observe the fragmentation behavior of the CHD-Azide modification at the peptide level, a direct infusion analysis was performed for untreated and CHD-Azide-treated peptide samples using ESI-Thermo Velos-MS. The HESI source was

operated in positive ion mode using ESI probe voltage at 2.5 kV, with a capillary temperature or 250 °C, sheath gas = 5, and S-lens RF = 62. Before the injection, the peptide samples were diluted with 1:1 MeOH: H$_2$O: 2% acetic acid and were then injected into the mass spectrometer through the syringe pump where they were ionized and analyzed.

The software Proteome Discoverer (version 1.3, Thermo Scientific, USA) was used to extract peaks from spectra and to match them to the RNase A sequence (UniProt P02769). Trypsin was selected as the cleaving proteases, allowing a maximum of three missed cleavages. Peptide and fragment ion tolerances were set to 5 ppm and 0.7 amu, respectively. Cysteine carbamidomethylation was set as the fix modification (+57.02147). The MS/MS spectra of modified peptides were manually inspected for the confident mapping of the modification sites. Synthesis: Detailed steps of synthesis are provided on pages 9 to 12 in the supplementary data. Briefly, 1,6- dibromohexane was first reacted with p-methoxy benzyl alcohol, under basic conditions, to prepare 1-((6-bromohexyloxy)methyl)-4-methoxybenzene. After this, 4-(6-((4-methoxybenzyl)oxy)hexyl)-2-oxocyclohexyl acetate was synthesized by coupling 2-oxocyclohex-3-en-1-yl acetate with the freshly prepared-((6-bromohexyloxy)methyl)-4-methoxybenzene. This new coupling product was deprotected under oxidative conditions to expose its terminal hydroxyl group, which was tosylated and subsequently substituted with sodium azide to produce the expected alkyl azide intermediate. Finally, an acetyl deprotection was performed followed by a Swern oxidation reaction to produce the CHD-Azide final compound.

3.4 Results and Discussions

3.4.1 Modification of Peptides by CHD-Azide reagent

A chemical labeling scheme and arginine residue enrichment in proteins are shown in Figure 3-1. At first, chemical labeling of arginine residues in peptides/proteins using CHD-azide was

performed. The structure of CHD-azide is shown in the same scheme (see also Figure 3-S1).

After that, proteins were digested, and resultant peptides were coupled using a clickable

biotinylated reagent, biotin-PEG4-alkyne (Figure S1). After click labeling, the CHD-azide labeled

peptides were enriched using biotin-avidin affinity chromatography. LC-MS/MS analysis of

enriched peptides provided the identity and location of the CHD-azide reacted products. ESI-MS

analysis of the modified model peptides is shown in Figure 3-2. The three model peptides we

used in our studies include bradykinin (RPPGFSPFR, monoisotopic mass 1059.5614 Da),

substance P (RPKPQQFFGLM monoisotopic mass 1346.7281 Da) and N-terminal blocked

peptide, neurotensin (pyroglu-LYENKPRRPYIL monoisotopic unmodified mass 1671.9097 Da).

These three model peptides contained two arginine residues at the two terminal ends, a single

arginine at the N-terminal end and two adjacent arginine residues in the middle of the peptide,

respectively. CHD-azide modified bradykinin mass was given in Figure 2A at m/z 749.92

$[M+2H]^{2+}$ and at m/z 500.34 $[M+3H]^{3+}$. Details of mass addition are shown (created with

ChemDraw) in Figure 3-S2A, B. We also compared the fragmentation pattern of unmodified and

modified peptides (see the supplementary data for comparison, Figure S3). The modified

peptide clearly shows complete labeling of the two arginines by this CHD–azide reagent.

Furthermore, the modification with neurotensin yielded two peaks at m/z 1056.60 for the doubly

charged ion and for the triply charged ion at m/z 704.76 (Figure 3-1B). The mass difference

between the unmodified neurotensin mass at m/z 837.45 and m/z 558.66 confirmed dual CHD-

azide incorporation to the two adjacent internal arginine residues in the peptide (Figure S2B and

S3B). The results were further confirmed by tandem mass spectrometry studies after careful

inspection of all MS/MS spectra from modified peptides.

3.4.2 Tandem mass spectrometry studies of the peptides

CHD-azide modification at the two-arginine containing peptide bradykinin R*PPGFSPFR* was

characterized by the CID tandem mass spectrometry spectra (Figure 3-2A). Fragmentation of

the doubly charged molecular ion at m/z 749.92 (Figure 3-2B) and unmodified parent ion at m/z 530.85 are given in the supplementary information Figure S3, respectively. The mass difference 438.14 Da, revealed that the two-arginine residues are modified with the CHD-Azide. For the formation of a single CHD-azide adduct, the mass addition is 219.07 Da. The annotated MS/MS spectrum in Figure 3-2B gave the modified b and y ion series (detailed m/z of fragment ions are provided in an MS Office Excel file). It is notable that the addition of 219.14 Da to modified ion fragments corresponds to the formation of two CHD-azide adducts with the two arginine residues. Figure 1D shows the MS/MS spectrum obtained after reaction between CHD-azide and neurotensin. This data further supported the conclusion that the two arginine residues are modified by CHD-azide. Most of the b and y ion fragments were observed with high abundance (please see the MS Office Excel files for the fragment ion mass). The above deduction was confirmed by the mass difference between the unmodified fragments and the modified fragment ions (see supplementary Figure 3-S3B). The MS^2 data provided confirmed the arginine selectivity and modification sites with high confidence.

3.4.3 The feasibility of click chemistry-based peptide enrichment

Next, we tested click labeling and biotin-avidin affinity enrichment of CHD-azide modified peptides using a commercially available biotin-PEG4-alkyne reagent. Biotin-PEG4-alkyne is not a suitable reagent for mass-spectrometric fragmentation due to the PEG and biotin group, but can be utilized at the protein level for enrichment. Due to the solubility, we decided to use this reagent to provide feasibility of enrichment at the protein level as well as the peptide level. To test the feasibility, a substance P modified peptide was spiked in a mixture of myoglobin digest. After click labeling in this complex mixture, we affinity purify the peptide by avidin. Further analysis of CHD modification followed by click chemistry of substance P peptide indicated that the CHD modified peptide was successfully enriched by biotin-avidin purification. MS/MS was performed on the [M+H]$^{2+}$ ion at m/z 1012.88. The fragmentation behavior of the enriched

peptide is shown in Figure 3-3 (see MS Office Excel files for fragments m/z's). The enriched CHD-azide modified peptide containing the clickable part gave most of the b fragment ions (Figure 3-3). The modified fragment ion series indicated that the addition is 457.58 Da, which corresponds to the BiotinPEG4-Alkyne attached to the CHD-Azide modification site. This together with the fact that unmodified ions b2, b7, b8, b10 and CHD modified fragments b2*, b7*, b8*, b10* as well as the CHD-azide click modified fragments b2#, b7#, b8# and b10# can be used to justify modification at the arginine residue (Figure 3-S4A, B). The mass difference between b2 and b2* is 219.14 Da, which clearly shows the CHD-azide modification at the arginine residue. The mass shift between b2* and b2# fragment ion of the enriched peptide is 457.58 Da, giving a strong evidence of the reaction of biotin-PEG4-alkyne with the azide group. Biotin and PEG functionality sometimes complicates the fragmentation spectra. Our future strategy is also to develop or use a cleavable click reagent to remove these groups before mass-spectrometric analysis. Based on the outcome results noted on the model peptides, we decided to carry out the CHD-azide modification reaction with the RNase A protein to investigate the CHD-azide adduct formation on arginine residues. This model protein system was studied to demonstrate the applicability of this chemical labeling method for locating functional arginine residues in the enzyme.

3.4.4 Modification of RNase A Protein by CHD-Azide compound

To identify the amino acids modified by CHD-azide, the digested peptides of the modified protein samples were analyzed by nano LC-MS/MS. We were able to identify the modified arginine residues in the RNase A protein. Two amino acid sequences were identified with the chemical modification at the arginine residues. Specifically we found two peaks at m/z 469.28 $[M+2H]^{2+}$ and m/z 1003.17 $[M+3H]^{3+}$, which corresponded to a peptide with amino acid sequence SR*NLTK and a large peptide, DR*C!KPVNTFVHESLADVQAVC!SQK with two carbamidomethyl modification at two cysteine residues, respectively. To confirm further the

modification, we investigated the tandem mass spectrometry data of the RNase A protein with the CHD-azide treated MS/MS data. The results in Figure 4A demonstrated the fragmentation of unmodified peptide SRNLTK. The difference between untreated and treated RNase A peptide SRNLTK was 219.14 Da, which suggested that the Arg- 33 was modified by CHD-azide. As presented in Figure 4B, the peptide SR*NLTK confirmed the Arg-33 modification by the CHD-azide compound. The precursor ion at m/z 469.93 generated CID fragment ions corresponding to b2*-NH$_3$, b3*, b4*, b5*, y1, y2, y3, y4 and y5*-H$_2$O. In addition, we saw unmodified peptide SRNLTK at m/z 360.90 give fragment ions b4, b5, y1, y2, y3-H$_2$O, and y6-NH$_3$. To compare the unmodified with the modified peptide SRNLTK, the ion mass at m/z 471.27 (b4) and b4* (CHD-Azide modified) fragment ion mass of the modified peptide SR*NLTK at m/z 690.41 clearly showed the mass addition of 219.14 Da. Closer examination of another fragment ion b5* of the modified peptide at m/z 791.45 with b5 of the unmodified peptide at m/z 572.32 showed the characteristic mass addition of 219.13 Da, suggesting that the CHD-azide modification was at Arg-33. Under identical tandem mass spectrometry conditions, we saw the Arg-39 modification in peptide DR*C*KPVNTFVHESLADVQAVC*SQK, as shown in Figure S5A. As expected, the CID data of the precursor ion peak m/z 1003.17 produced few fragment ions of this large peptide with several modifications of DR*C!KPVNTFVHESLADVQAVC!SQK (*-CHD-Azide, !-carbamidomethylation of cysteine). In this peptide the 219.16 Da increase in the peptide mass between the unmodified peptide and the CHD-azide modified peptide confirmed that the Arg-39 was also modified by the CHD-azide. The tandem mass spectrometry data produced by both sequences were matched against the molecular calculator software (omics.pnl.gov) to confirm the identity of the peptides. The appearance of the most abundant b and y ions of peptide DR#C!KPVNTFVHESLADVQAVC!SQK with the CHD-azide modification (Figure 3-S5B) also confirmed the modification at Arg-39. However, the quality of the tandem mass spectrometry was poor due to the side chain PEG and biotin of the enrichment reagent. As displayed in Figure S5B, some of the b and y ions are absent or they occur at very low intensity m/z values,

since the peptide is having two carbamidomethyl modified cysteine residues along with the CHD-azide-click modified peptides. The difference between the ion fragment b17^{3+} from a click modified peptide at m/z 882.56 (Figure 3-S5B) and CHD-azide modified b17^{3+} at m/z 730.68 (Figure 3S-5A) confirmed the mass addition of 457.58 Da, the MW of the biotinylated click reagent. Our data showed that the Arg-33 and Arg-39 are the most reactive arginine residues present in RNase A. Our findings agree with Patthy and Smith[38,60] on identification of functional arginine residues in ribonuclease A and lysozyme. The data also agree with Brock et al.'s mass spectrometric analysis studies on the detection and identification of arginine modifications on methylglyoxal-modified ribonuclease.[76],[80] These findings were displayed with the crystal structure of RNase A protein with all arginine residues labeled and the selective arginine residues labeled and enriched with our CHD-Azide and bio-orthogonal enrichment strategy (Figure 3-5, A and B). The challenge associated with the CHD type compounds are the high pH of reaction condition. Several researchers have frequently used CHD due to its labeling efficiency compared to phenyl and methylglyoxal. Fortunately, it was determined that even with high pH conditions, the CHD azide and bio-orthogonal enrichment strategy is quite capable of labeling functional arginine residues.

3.5 Conclusions

In this study, we utilized a chemical approach for selective enrichment of the reactive arginine residues in proteins. To the best of our knowledge, for the first time, we developed an arginine selective reagent with enrichment functionality utilizing a cyclic 1,2 dicarbonyl compound. We incorporated an azido group in 1,2 cyclohexanediones so arginine labeled peptides can be enriched with a clickable biotinylated reagent. We tested the chemical labeling methods in several standard peptides with arginine residues in different positions in the sequence. Complete labeling was used and verified as complete in those peptides. We demonstrated affinity enrichment with a commercially available clickable biotinylated reagent. The peptide was

spiked in a protein digest successfully with its affinity purified after click labeling. Furthermore, we tested labeling efficiency of this reagent on RNase A protein to identify and validate the reactive arginine residues. Two functional arginine residues were selectively identified and enriched using this labeling and enrichment approach. We utilized a commercially available clickable biotinylated reagent to evaluate the enrichment of CHD-azide labeled peptides. Due to PEG and biotin functionality in this compound, we have seen poor fragmentation in the click-modified large and small peptides. Nevertheless, we have successfully demonstrated the attachment of CHD-azide and click-adducts after enrichment. We are in the process of employing a cleavable clickable-biotinylated reagent so the modified peptide can be removed from the avidin beads by chemical cleavage. Thus, fragmentation complexity can be reduced from the click-labeled mass spectra. Our approach demonstrated that CHD-azide labeling and the click-based enrichment method is a reliable analytical technique to determine the reactive/functional arginine residues. This study will pinpoint the reactive arginine targets for the biomarker development with high confidence.

3.6 Acknowledgements
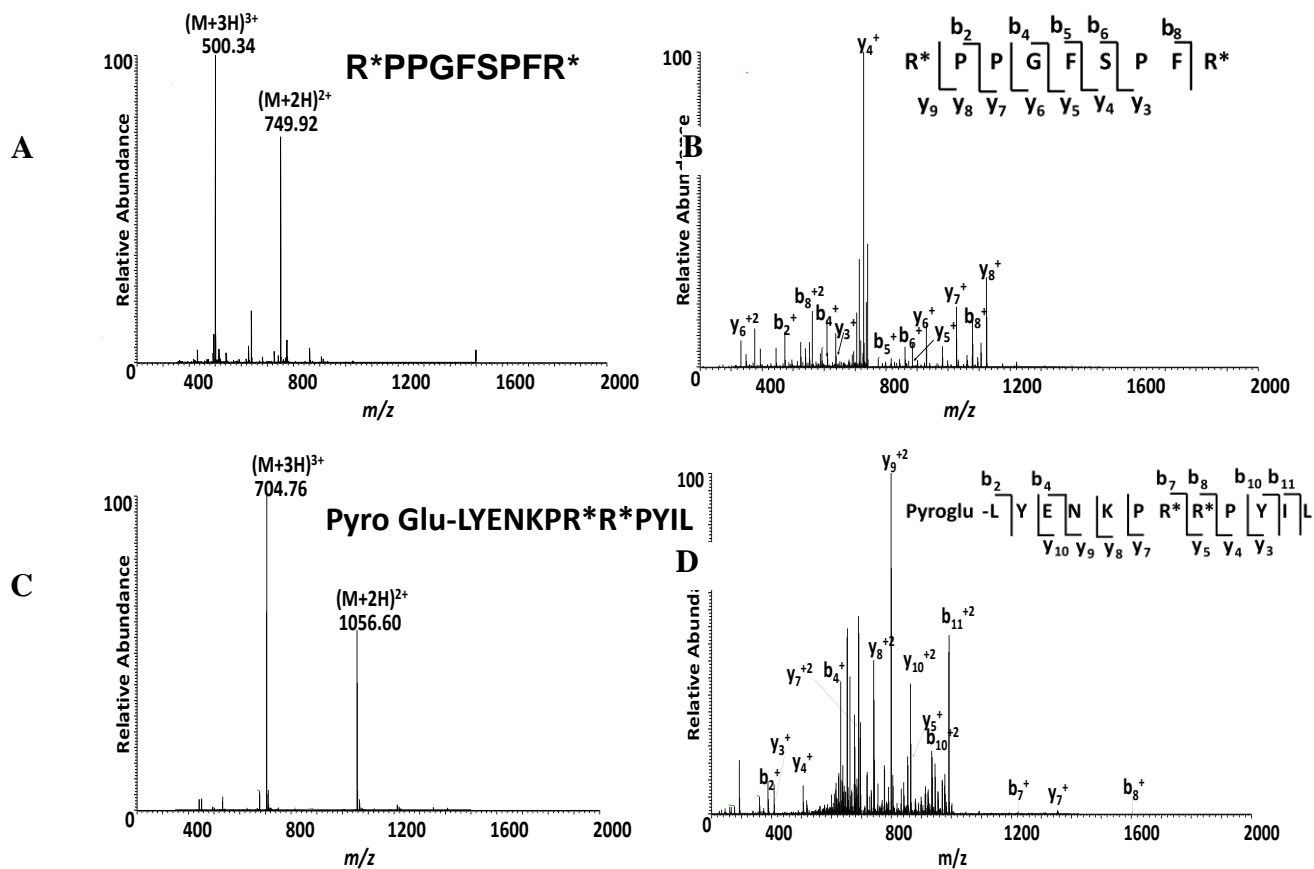
Figure 3-1. ESI-MS and MS/MS spectra of modified peptides with CHD-Azide reagent. (A) CHD-azide modified bradykinin at *m/z* 749.92(M+2H+CHD-Azide)$^{2+}$ and *m/z* 500.34 (M+3H+CHD-Azide)$^{3+}$, (B) MS/MS spectrum of CHD-azide modified bradykinin at *m/z* 749.92 (M+2H+CHD-Azide)$^{2+}$, (C) CHD-azide modified neurotensin at *m/z* 1056.60 (M+2H+CHD-azide)$^{2+}$ and *m/z* 704.76 (M+3H+CHD-Azide)$^{3+}$, and (D) MS/MS spectrum of CHD-Azide modified neurotensin at *m/z* 704.76 (M+3H+CHD-Azide)$^{3+}$. Please see the supplementary data for unmodified spectra for comparison. *denotes CHD-azide modified sites.
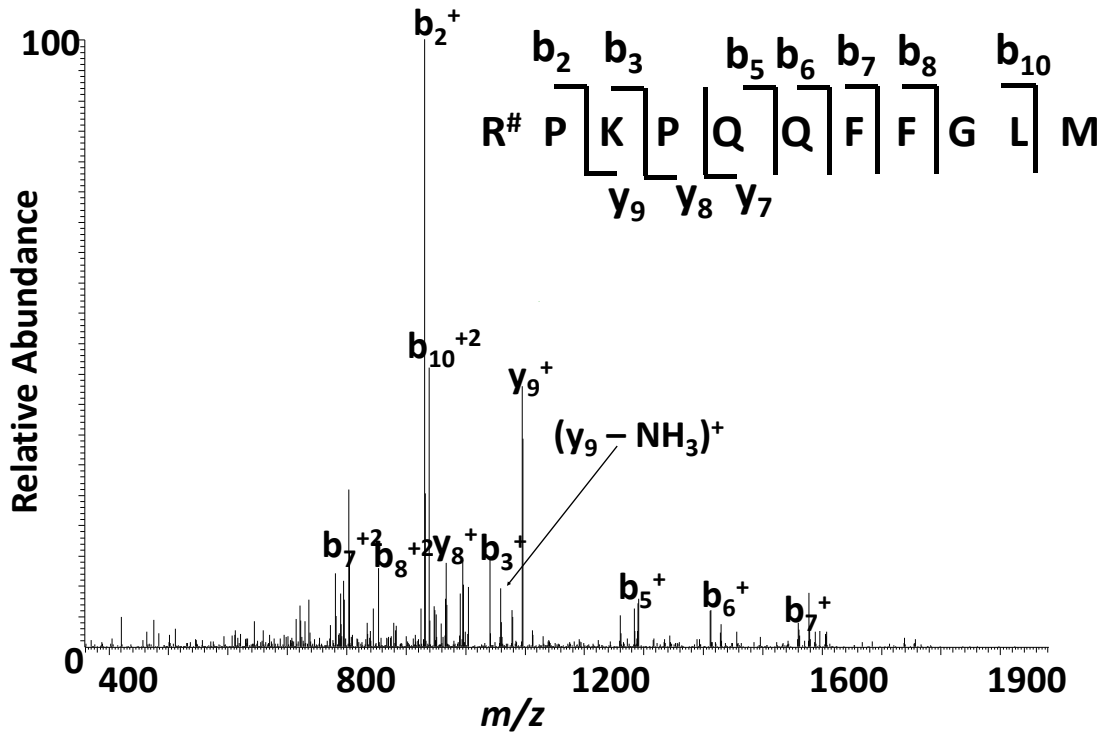
Figure 3-2. MS/MS spectra of spiked CHD-azide modified and click-enriched substance P peptide at m/z 1012.88 $(M+3H+CHD-Az+CLICK)^{3+}$.

A
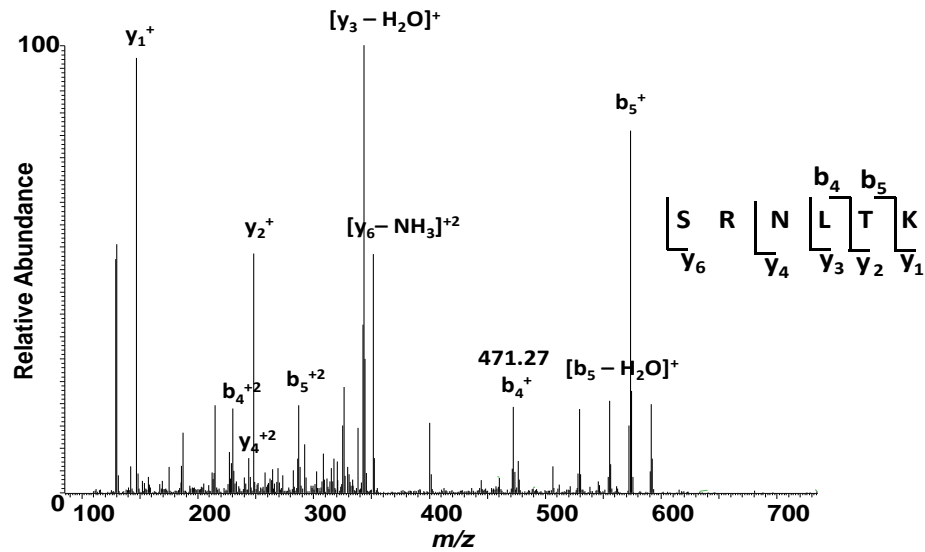
B
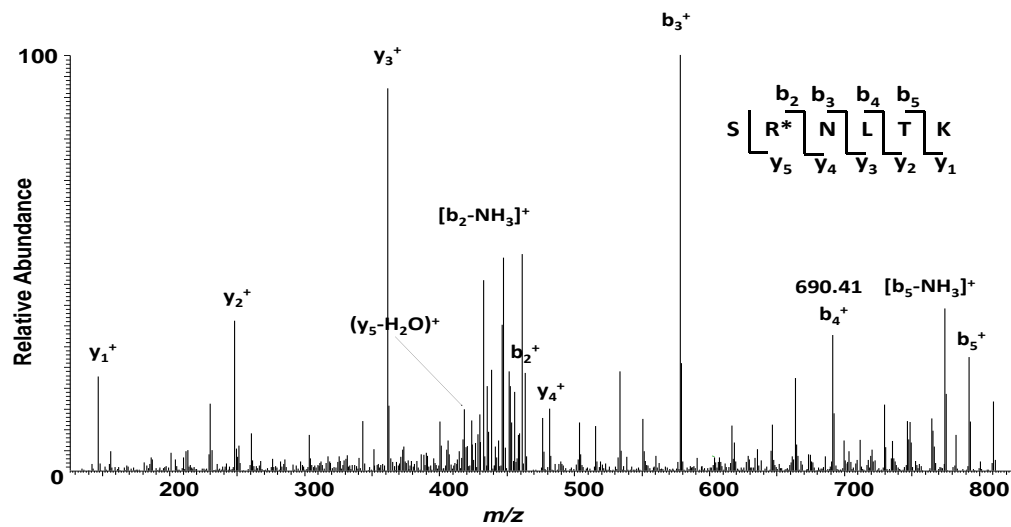
Figure 3-3. MS/MS spectra of unmodified and modified Ribonuclease A peptides (RNase A). (A) MS/MS spectrum of peptide SRNLTK at m/z 359.71$(M+2H)^{2+}$ and (B) MS/MS spectrum of CHD-azide modified peptide SR*NLTK at m/z 469.90 $(M+2H+CHD-Az)^{2+}$. For example, the difference between b4* and b4 corresponds to the addition of 219.14 Da, the CHD-azide modification mass.
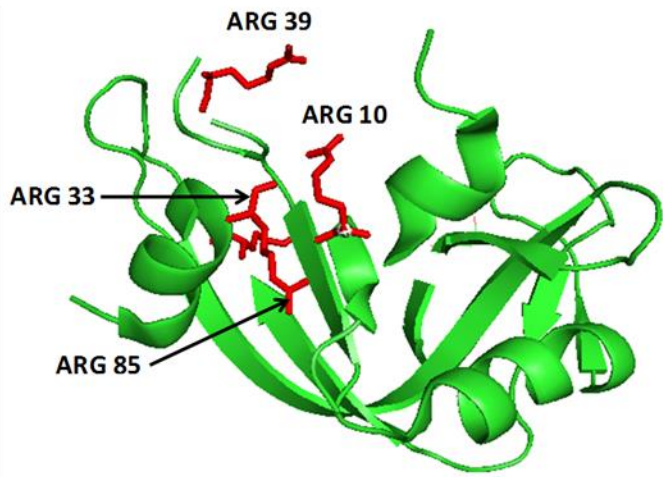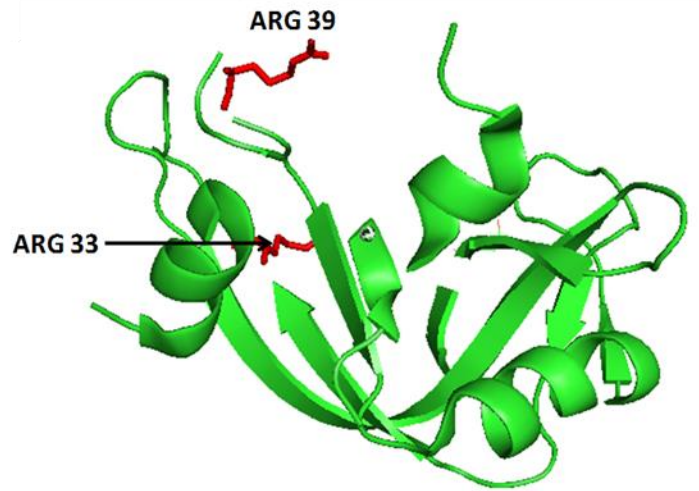
Figure 5A   Figure 5B

Figure 3-4. Crystal structure of RNase A protein (PDB file 4J5Z). (A) All arginine residues in RNase A and (B) modified sites of RNase A with CHD-azide.

A

(M+2H)⁺² 530.85

Unmodified peptide mass [M+H]⁺ = 1060.56
Modified peptide mass [M+H]⁺ = 1498.84
Mass Addition for two arginine residues = 438.28

RPPGFSPFR

(M+3H)⁺³ 354.26

Δm= 146.08

(M+3H)⁺³ 500.34

Δm= 219.07

(M+2H)⁺² 749.92

R*PPGFSPFR*

B

y₈²⁺

R P P G F S P F R
y₉ y₈ y₇ y₆ y₅ y₄ y₃ y₁
b₄ b₅ b₆ b₈

y₁⁺ b₄⁺ y₃⁺ y₇²⁺ (y₉ – H₂O)⁺² y₇⁺ y₈⁺ b₈⁺²
y₄⁺ y₅⁺ b₅⁺ b₆⁺ y₆⁺

y₄⁺

R* P P G F S P F R*
y₉ y₈ y₇ y₆ y₅ y₄ y₃
b₂ b₄ b₅ b₆ b₈

(y₉– H₂O)⁺²
y₆⁺² b₂⁺ b₈⁺² b₄⁺ y₃⁺ b₅⁺ b₆⁺ y₆⁺ y₅⁺ y₇⁺ b₈⁺ y₈⁺

C

(M+2H)⁺² 674.60

Unmodified peptide mass [M+H]⁺ = 1347.73
Modified peptide mass [M+H]⁺ = 1566.87
Mass Addition for single arginine residue = 219.14

RPKPQQFFGLM

(M+H)⁺ 1347.78

Δm= 109.32

(M+3H)⁺³ 522.42

(M+2H)⁺² 783.92

R*PKPQQFFGLM

D

(y₁₁ – NH₃)²⁺ b₁₀²⁺

R P K P Q Q F F G L M
y₁₁ y₁₀ y₉ y₈
b₂ b₃ b₆ b₇ b₈ b₉ b₁₀

b₂⁺ (b₁₀ – NH₃)²⁺ (y₉ – NH₃)⁺ b₉⁺
b₃⁺ b₈⁺² b₆⁺ b₇⁺ b₈⁺ y₉⁺ b₁₀⁺ y₁₀⁺
y₈⁺

b₁₀⁺²

R* P K P Q Q F F G L M
y₁₁ y₁₀ y₉ y₈
b₂ b₄ b₅ b₆ b₇ b₈ b₉ b₁₀

b₂⁺ b₄⁺ (y₁₁ – NH₃)⁺² y₈⁺ y₉⁺ b₈⁺ b₉⁺
b₅⁺ b₆⁺ y₇⁺ y₁₀⁺ b₁₀⁺

62

E

Unmodified peptide mass [M+H]⁺ = 1672.91
Modified peptide mass [M+H]⁺ = 2111.19
Mass Addition for two arginine residues = 438.28

Pyro Glu-LYENKPRRPYIL

Pyro Glu-LYENKPR*R*PYIL

F

Figure 3-S1. MS spectra of unmodified and modified peptides & MS/MS spectra of unmodified and modified peptides.

A

B
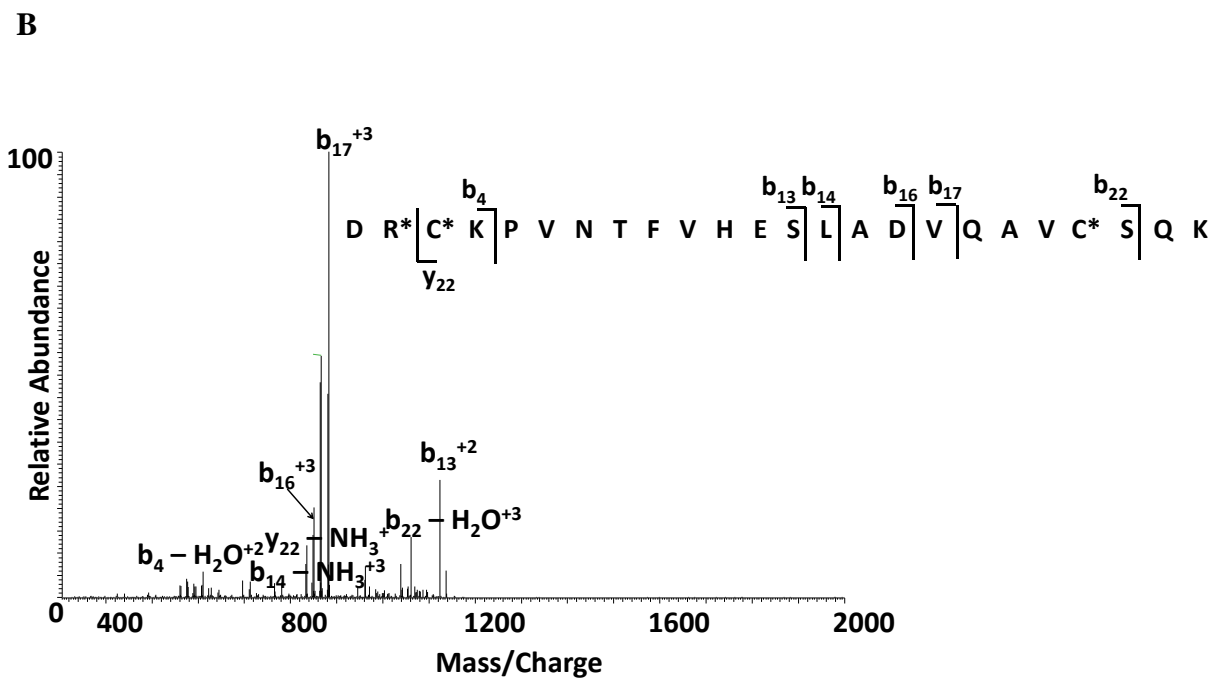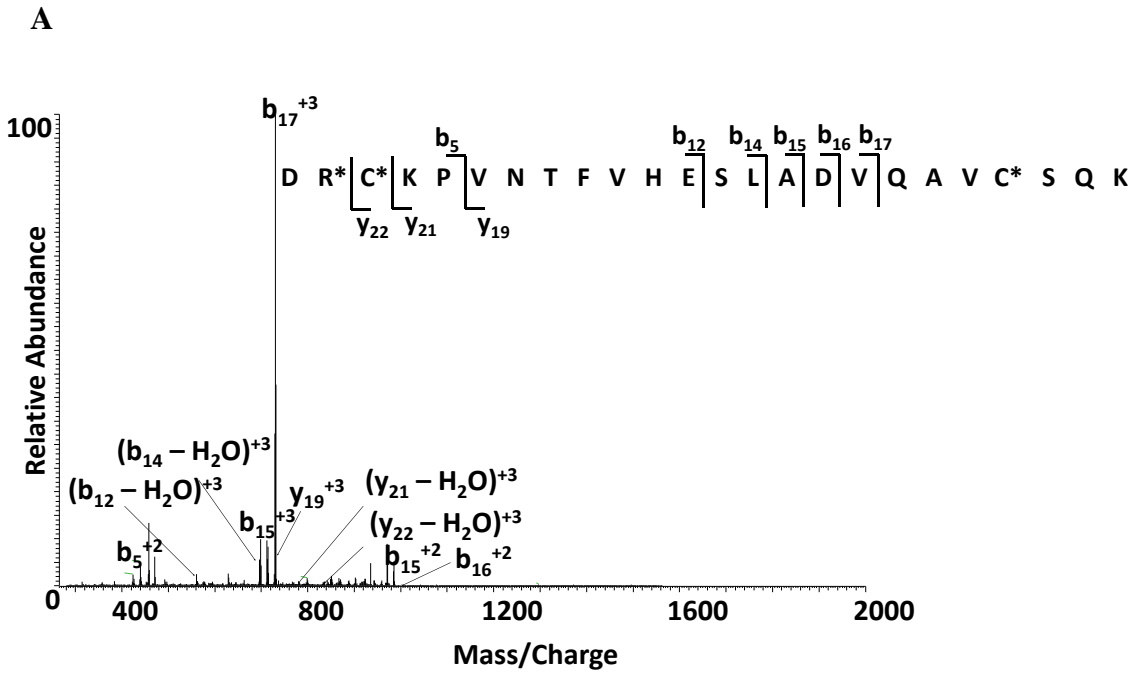
Figure 3-S2. MS/MS spectra of unmodified and modified Ribonuclease A peptides (RNase A).
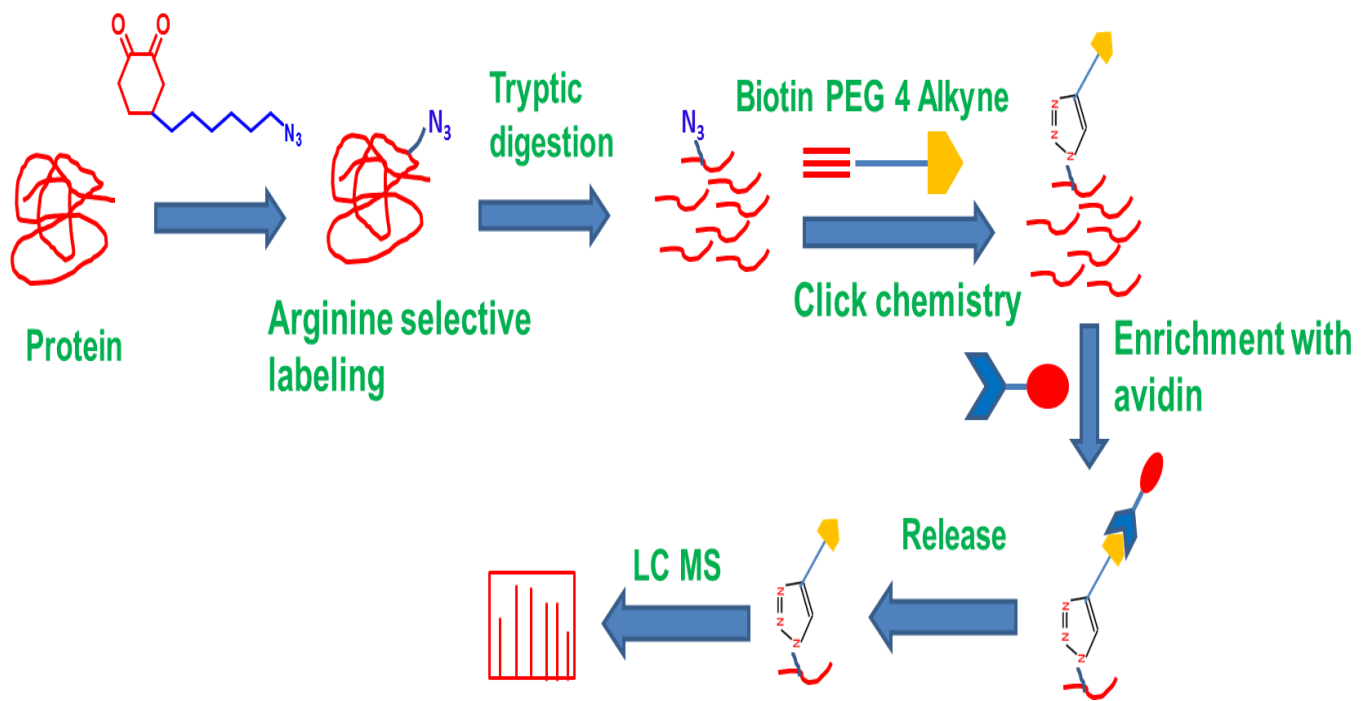
Figure 3-S3. Schematic representation of arginine labeling reaction

Chapter 4

A novel mass spectrometry cleavable reagent that enables identification of proteolytic cleavage

4.1 Abstract

Proteolytic cleavage is an irreversible post-translational modification where the proteases break the peptide bonds in proteins. Recent efforts have focused on identifying the proteolytic cleavage sites, which have become very vital as proteolysis has an all-encompassing ubiquitous involvement in pathogenic processes. One of the approaches is to focus on the analysis of theprotein n-termini. The current literature indicates that a different strategy is required to identify the protein n-termini. Our method employs a solid phase capture and releasing technique, which allows selective isolation of protein n-terminal peptides and identification in mass spectrometry (MS). Not only does the advent of the solid phase reagent provide a more efficient way to achieve enrichment, but it also improves MS analysis of the protein n-terminal. MS is a high-throughput analytical tool for characterization of proteins. Our method allows high throughput profiling of protein n-termini from a complex mixture of proteins. The key to identifying the protein n-terminal is to view and define the signature fragment mass using tandem mass spectrometry data.

4.2 Introduction

In the biosynthetic pathway of protein production, amino acids are assembled in a step-wise fashion starting from the N-terminal to the C-terminal. The protein N-terminal contains all the information about the fate of the mature protein.[8] Early studies on N-termini identification of mature protein was based on Edman degradation.[81] Understanding of protein functions and the physiological process of protein n-termini studies has been a goal of mass spectrometry for

decades. Most strategies are based on bottom up proteomics where the protein of interest enzymatically digested prior to the MS detection.[82-84] Traditional bottom up proteomic strategies were two-fold: The first proteomic strategy was based on direct selection or the positive selection approach and the second strategy was based on counter depletion or the negative selection approach.[85-88] In the direct selection approach for protein analysis, the N-termini was derivatized using different tags after the ε-amino groups on side chain of lysine residues were acetylated or guanidinated and concentrated as n-termini peptides. After comparing the direct selection approach with the general chromatographic separation technique with acetylation, Chen et al. introduced an SCX (strong cation exchange) method to separate acetylated N-termini from original N-termini by the charge difference.[89] When a protein is enzymatically digested with trypsin, the cleavage happens at the C-terminal of the arginine and lysine when there is no proline next to the lysine and arginine. The resulting peptides have a positive charge on N-termini and another positive charge on the C-termini. The net charge is two positive charges. The acetylated peptide has a single charge on the C-terminal. The counter depletion approach was used in the combined fractional diagonal chromatography (COFRADIC) where we can detect both the acetylated and unmodified peptides.[17] In the COFRADIC method, the protein was initially reduced to an alkylate followed by the acetylation. Both the lysine primary amines and n-termini primary amines are blocked. Then the proteins are digested with trypsin and SCX is used to separate the internal peptides. A reverse phase liquid chromatography was followed to fractionate the peptide mixture. Next all the peptides preset in the fraction were derivatized using 2,4,6-trinitrobenzenesulfonic acid (TNBS). The internal peptides become more hydrophobic and shift the elution rate compared to the secondary chromatographic run by the alteration of the TNBS reaction. To enhance the N-terminal peptides separation from the extraneous peptides, TMPP was also used as a labeling reagent. In their studies, they used three different proteases in order to obtain a reliable confirmation of the protein N-termini.[90]

Xu et al. introduced N-CLAP (N-terminal omics by Chemical Labeling of the Alpha Amine of Proteins) strategy where the cleavage sites in proteins can be readily identified by selectively labeling the N-termini amino terminal over the lysine sidechain amino groups.[91] Protein primary amines were modified with phenyl isothiocyanate. After a single Edman degradation step was completed, the neo N-termini were biotinylated using sulfo-NHS-SS-biotin. Finally, the reduction was conducted with TCEP and an analysis of the N-terminal methionine processing is conducted using LC MS and tandem mass spectrometry.

Various potent approaches exist for profiling thespecificity of the proteases.[92-94] Post-translational modifications play a key role in biological processes.[16,95,96] Proteolytic cleavage is a ubiquitous posttranslational modification. MS became an indispensable tool to analyze post-translational modifications. During the proteolytic process the signal peptides can be removed during translocation or N-terminal methionine residues are clipped off during translation.[97-99]

Numerous methods have been reported to isolate N-termini of proteins.[100] In general these approaches involve an enrichment step combined with specific detection steps. A significant advantage of our novel solid phase reagent over the others is that it is comprised of dual properties, which can be enriched and made mass spec cleavable in itself.[101]

Mahrus et al. established a positive selection method to globally identify the proteolytic cleavage sites in proteins.[102] In their method, they used a modified subtiligase enzyme which cleaves the biotinylated peptides of the protein N-termini. For the biotinylation, they used a biotinylated peptide glycolate ester, which contains both a TEV protease cleavage site and a subtiligase cleavage site. Following the enzymatic digestion with trypsin, biotinylated proteins were captured using avidin beads. Recovered peptides with TEV protease were, analyzed with 1D or 2D LC/MS/MS. Limitations of this method includes the specificity of the subtiligase enzyme and the efficiency of ligation. The positive selection method allows high throughput profiling of

unmodified primary amines of the protein N-termini. A major drawback of this type of method is its inability to study the naturally modified n-termini. Over the positive selection technique, negative selection can be used to enrich both unmodified and naturally modified protein N-termini.[103,104]

In recent years, COFRADIC and TAILS methods have been extensively used in the positional proteomics field. Moment et al. has introduced a new technique for the enrichment of protein N-termini through the phosphor tagging approach (PTAG).[105] N-terminal and lysine amino groups are initially completely derivative with formaldehyde at the protein level and after the digestion, the enrichment of phosphor tagged internal peptides were directed using titanium dioxide affinity chromatography. This method supported identification of 753 unique N-terminal peptides in *N. meningitides* in and 928 unique N-terminal peptides in *S. cerevisiae*. Krefeld et al. have developed a negative selection method using the terminal amine isotope labeling of substrates (TAILS) for identification and quantification of N-terminal peptides and proteolytic events.[106] All the selective enrichment techniques prone to the side reactivity and the sample loss.[107] After the digestion, the sample mixture contains lot of internal peptides which cover the N-terminal peptides. In view of the challenges, mass spectrometry-cleavable reagents associated with enrichment methods have been developed.

4.3 Materials and methods

Myoglobin (from equine heart), acetonitrile (ACN), ammonium bicarbonate (NH4HCO3), Iodoacetamide (IAM) and formic acid (FA) were purchased from Sigma Aldrich (St. Louis, MO, USA). Guanidination reaction was performed using isomethylurea hemisulfate, which was obtained from Acros Organics (Morris Plains, NJ, USA). For reduction, dithiothreitol was acquired from Bio-Rad (Hercules, CA) and Iodoacetamide (IAM) for the alkylation from Sigma Aldrich (St. Louis, MO, USA). For proteolysis sequencing-grade modified trypsin obtained from

Promega (Madison, WI, USA) The samples were desalted by C18 ZipTip from Thermo Fisher Scientific (Rockford, IL, USA).Three model peptides bradykinin, neurotensin, and substance P were purchased from ANASPEC (Fremont, CA). To remove the excess solid phase reagent, Pierce concentrators of 3K MWCO were utilized. For the enrichment, Pierce UltraLink® monomeric avidin was purchased from Thermo Scientific (Rockford, IL). Nano-purified water was used in all experiments from the Mili-Q water dispenser.

4.3.1 Spiking Studies

To demonstrate the enrichment strategy, 1mM 5ul of myoglobin was used for the guanidination reaction. The stock solution for the guanidination was prepared by using O-methylisourea by dissolving 50 mg of O-Methylisourea in 51 µL of water. Protein solution was mixed with 30 µL of 7 N NH_4OH and 5 ul of O-methylisourea stock solution. To avoid the incomplete guanidination, a fresh stock solution was prepared prior to each reaction.

The resulting protein mixture was incubated at 65°C for 30 min in thermomixer (Eppendorf). 1 mM 5 µL of bradykinin peptide and the 30mM PFP-RINK-biotin in $NaHCO_3$ was then spiked into the guanidinated protein sample. This mixture was incubated for two hours at 37 °C for the labeling of N-termini for the protein and peptide. After the labelling, the protein sample was digested with trypsin followed by an additional guanidination step to assure all lysine conversion to homoarginine. Protein digestion was performed as follows. Guanidinated and the Bradykinin spiked protein sample was diluted with 50 mM NH_4HCO_3. Then sequence grading trypsin was added to the 1:100 enzymes to substrate ratio respectively. Afterwards the sample was incubated at 37°C overnight. After the reguanidination step, the sample was concentrated using a Speed Vac. A dried peptide sample was reconstituted in PBS buffer and selective enrichment of n-termini modified peptides was performed using traditional affinity chromatography. Initially monomeric avidin beads were washed three times with PBS buffer. Then, the sample was

transferred to the digested peptide sample into the avidin containing Eppendorf tube. The sample was incubated in a rotor at room temperature for one hour. The supernatant was collected into a separate Eppendorf tube. The beads were washed with PBS, NaCl, and water sequentially. Once done with the extensive washing, the beads were incubated with the elution buffer for 1 hr at room temperature. Tagged peptides were then selectively eluted with a higher organic solution from avidin beads. Eluted samples were evaporated to dryness with a SpeedVac rotary evaporator and reconstituted in 0.1% formic acid for MS analysis.
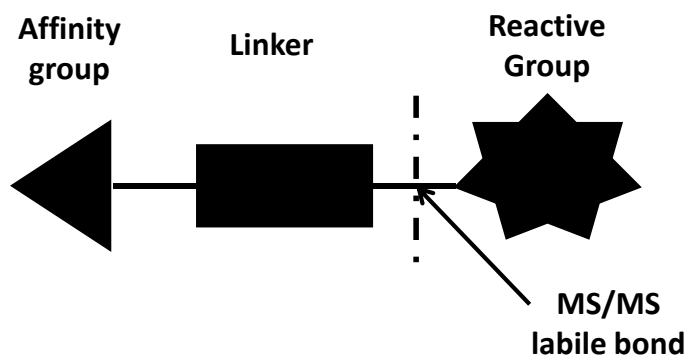
Figure 4-1 Conceptual modular design of the mass spectrometry cleavable reagent

NHS-RINK-BIOTIN

Chemical Formula: C39H42F5N5O9S
Exact Mass: 851.2623

Chemical Formula: C37H46N6O11S
Exact Mass: 782.2945
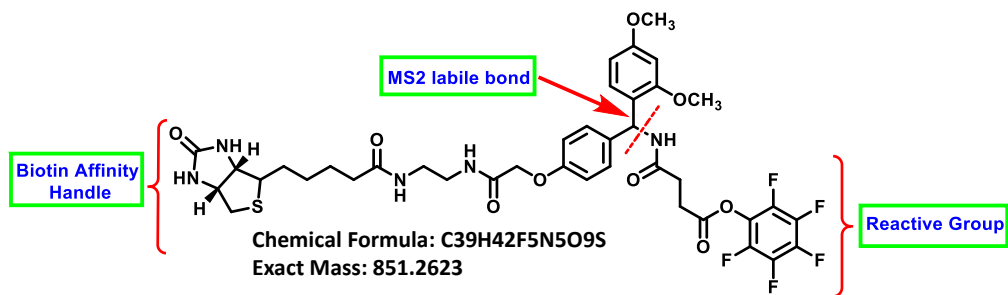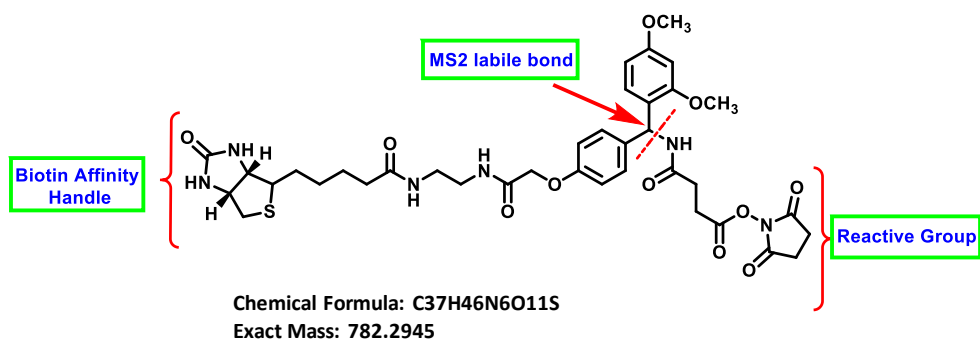
Figure 4-2 Structures of the primary amine reactive CID cleavable enrichment reagents
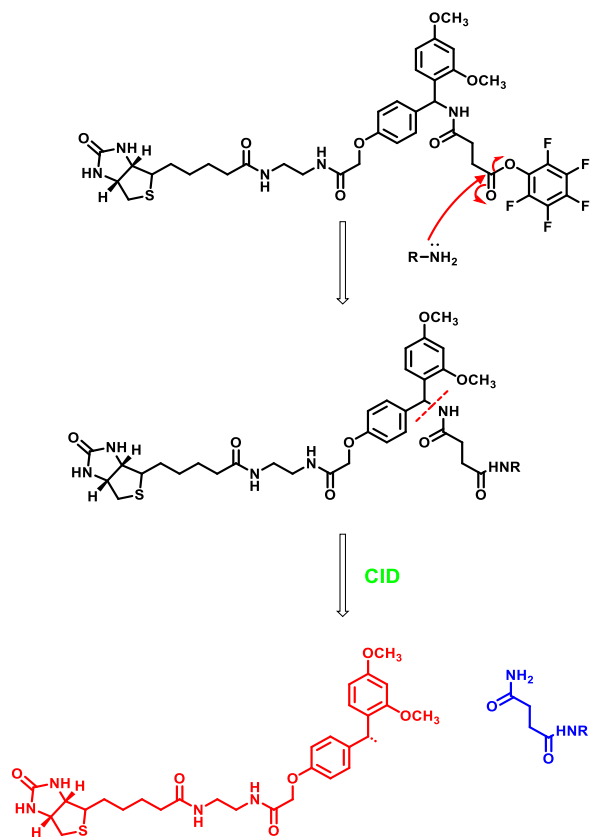
Figure 4-3 Reaction mechanism of reagent with a peptide
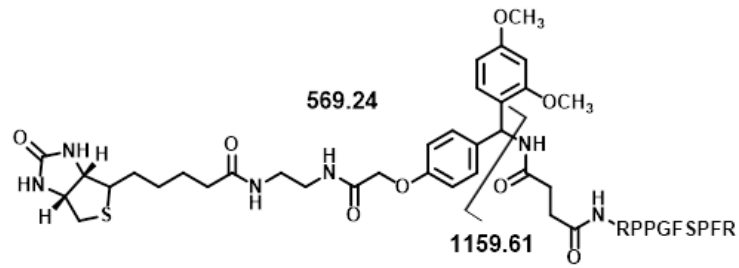
Figure 4-4 Fragmentation of reagent with a peptide
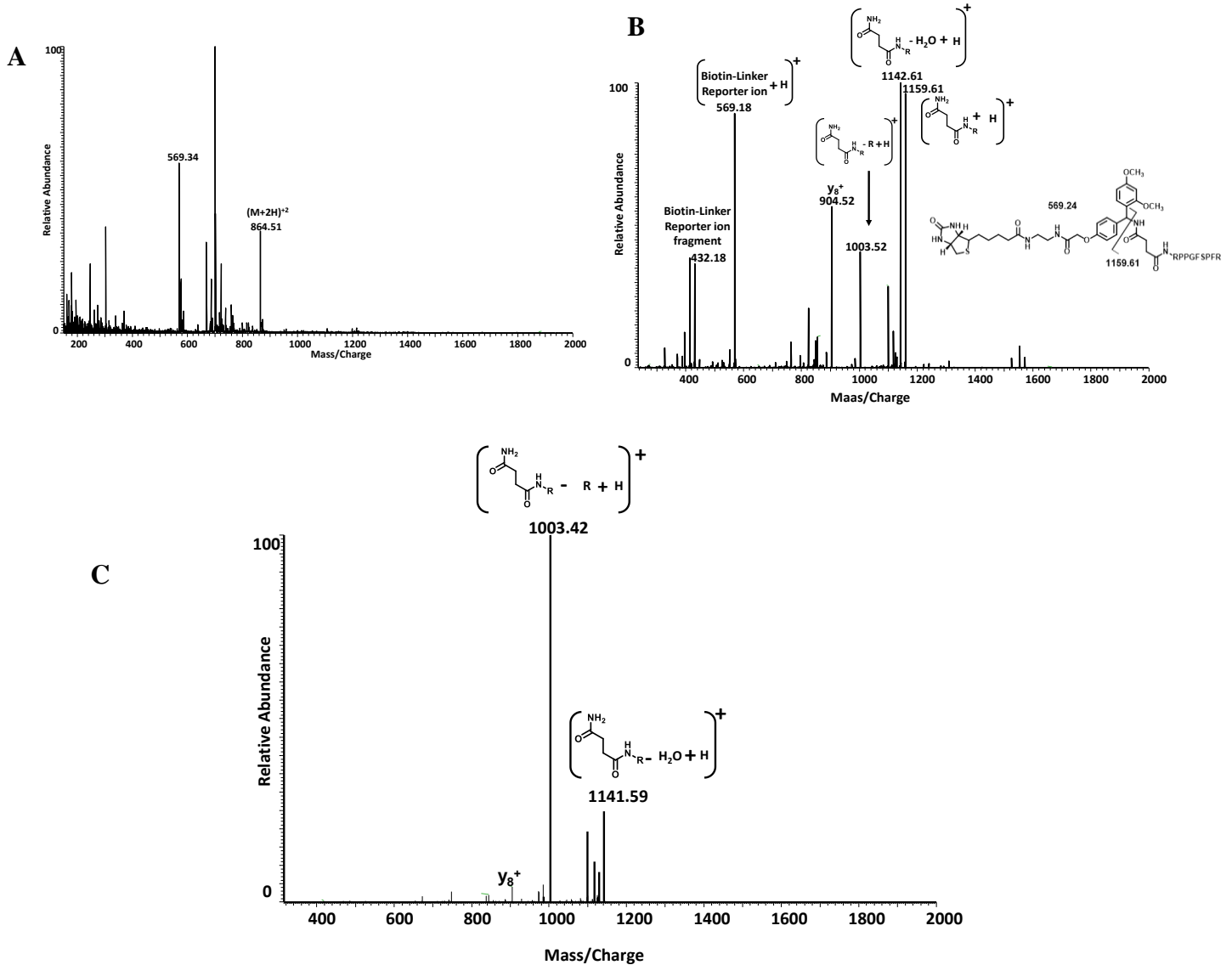
Figure 4-5 (A) Mass spectra of Bradykinin derivatized with PFP-RINK-biotin, (B) MS/MS spectrum of the ion m/z 864.5, and (C) MS3 spectrum of ion m/z 1159.61.

4.4 Results and Discussion

Mass spectrometry analysis of biotinylated peptides was performed using a Nanospray-ESI- MS and MS/MS with a Thermo Scientific Velos Pro dual-pressure linear ion trap mass spectrometer operated in positive ion mode and controlled by Xcalibur software (Thermo Fisher Scientific, San Jose, CA, USA). Using the nano-HPLC system Ultimate 300 coupled to the LTQ Velos Pro mass spectrometer (Thermo Fisher, San Jose, CA), mass spectra were acquired over the range m /z 200-2000. Reverse phase chromatographic separations of the biotinylated peptides were performed with a 65 min run with a flow rate of 300 nl/min. A multi-step gradient was used containing 4% B for 3 min, 4 to 30% B in 30 min, 30% to 60% B in 55 min (A, 0.1% formic acid in water; B, 95%: 5%: 0.1% Acetonitrile: water: formic acid 0.1% formic acid) on a Dianex Acclaim® Acclaim PepMap™ C18 column (Thermo Fisher Scientific) with an inner diameter (i.d.) of 75 μm, particle size of 3 μm, pore size of 100 Å and a length of 15 cm bed for the peptide separation. The column temperature was set to 35 °C. Automatic gain control (AGC) was 1*E6, normalized collision energy (NCE) was 35%, and a data dependent scan was used from MS/MS top 10 with S/N 5.

As described during the LC MS analysis observation of ion at m/z 864.51 indicates the potential existence of the precursor ion having the doubly charged PFP-RINK-biotin modified Bradykinin. The doubly charged precursor ion was isolated for the CID fragmentation. The MS/MS spectrum shown in Figure 4-3C shows the efficient cleavage of the CID cleavable bond. It generates the reporter ion mass at m/z 569.18. As indicated in Figure 4-3b the ions generated at m/z 432.18, 904.58, 1003.52, 1142.61, 1159.61, which represent biotin linker reporter ion fragment, biotin linker reporter ion, y8$^+$ of Bradykinin peptide, linker with one arginine loss in the peptide, linker with a loss of a single water molecule, peptide with the group of the linker respectively. This MS/MS data further confirmed the Bradykinin modification with the PFP Rink biotin. As depicted in Figure 4-3C the MS^3 spectrum of the further fragmented 1159.61 ion supported the

assignment of the amino acid sequence for the confident identification of the bradykinin peptide. The observation of a series of b and y ions is lacking due to the high intense peak at m/z 1003.42, which is the arginine loss from the peptide attached to the group of the linker. Ion at m/z 1141.59 is formed when a single water molecule is lost from the peptide with the group of linkers.

The MS^2 and the MS^3 spectra clearly show that our novel enrichment reagent is reacting with the N-termini of the peptide.

Traditionally biotin-avidin interaction is used for specific enrichment of molecules of interest from complex biological mixtures. Introducing a biotin group into the structure of the enrichment reagent provides an affinity tag that can be used for specific pulldown studies. Currently we are working on the HCD fragmentation studies of the peptide,

4.5 Conclusions

Based on the excellent performance of our solid phase reagent we believe that this enrichment strategy will improve the mass spectrometric identification of the n-termini of proteins. In addition, this approach significantly reduces the sample complexity compared to the traditional approaches and allows elucidating the proteolytic cleavage pathways.

Chapter 5

Summary and Future Work

Through the development of our method we were able to identify the reactive arginine residues in protein bovine serum albumin. For further analysis of the modified peptide, we performed an enrichment study with a newly synthesized arginine reactive reagent. This eliminated the sample complexity. However, the biotin alkyne reagent used in the enrichment strategy gave potential issues on the sample cleanup process and on the LC MS separation. Therefore, new methods are needed for faster analysis steps and for improved enrichment strategies. To obtain a better quality data, we are currently working on two different biotin alkyne reagents. Diol Biotin-Alkyne is one azide reactive reagent we used in our new studies. The reagent consists of a biotin group and a vicinal diol where the cleavage can be done using sodium periodate. Another reagent we used is an alkyne activated photo cleavable compound containing a biotin moiety linked to a terminal alkyne group. The captured azide modified peptides were then released by exposing to UV light. These reagents are capable of effectively pulling down the click-functionalized peptides from a complex mixture. Click Chemistry is a powerful reliable selective reaction. The concept was introduced by K. B. Sharpless in 2001. Copper-catalyzed azide-alkyne cycloaddition (CuAAC) is catalyzed by copper followed by the triazole formation when the alkyne and azide reacts together. Partial work on these two reagents has been done. We believe future work in this aspect will improve our click-based enrichment method. We studied the chemical labeling of arginine residues and found the functional arginine residues. It is critical to find out the chemically modified arginine residues in a complex sample mixture. Moreover, after the enrichment studies, successful cellular level studies with a complex matrix are planned in the future. The amount of methylglyoxal accumulation is high in diabetes patients andl ends up as a more advanced glycation end product (AGEs) formation. This increases the severity of the complications of diabetes. Our future goal is to identify the CHD adduct and to

76

utilize our targets to design a quantitative approach for this modification using multiple reaction monitoring (MRM) mass spectrometry. We plan to determing whether these modified arginine residues are more prone to form AGE adducts through profiling, and we will design a quantitative experiment using MRM.
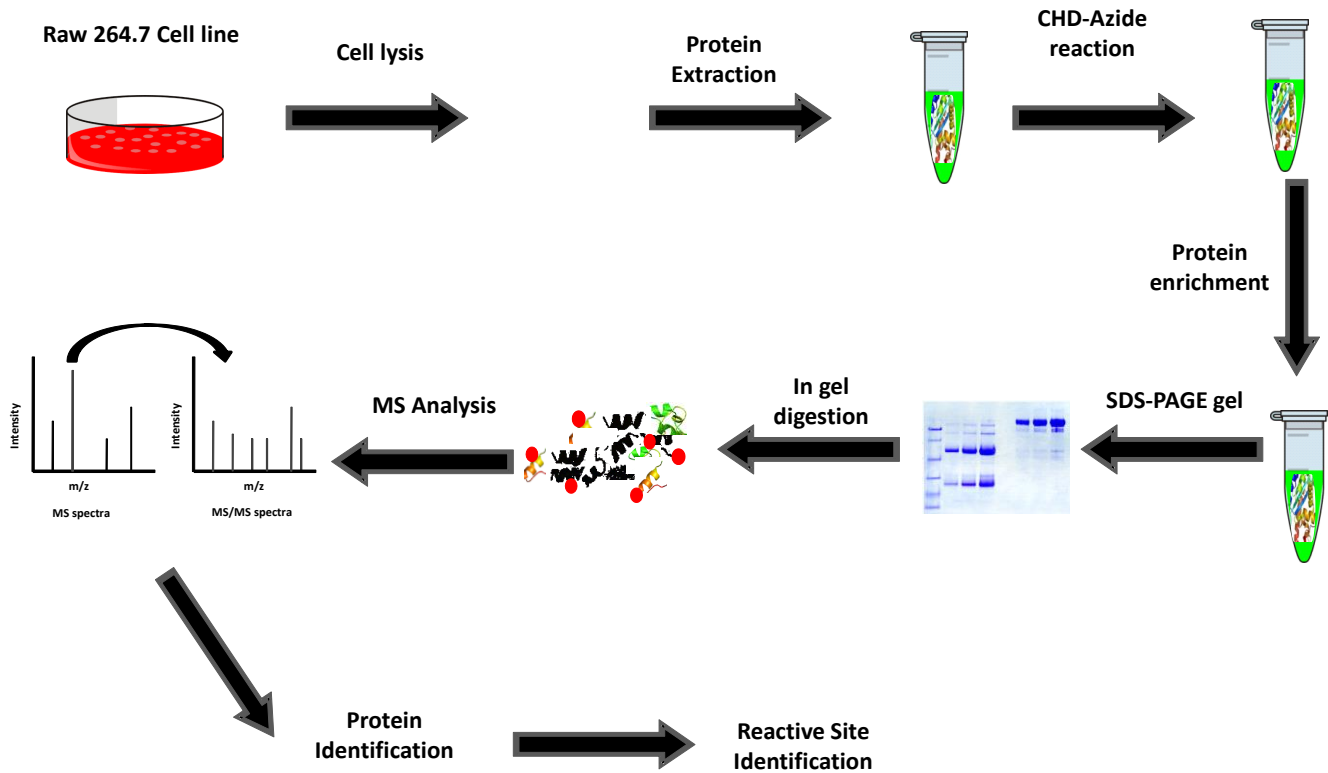
Figure 5-1 General scheme for enrichment of CHD-azide modified peptides using chemical derivatization method in RAW 264.7 macrophage cell
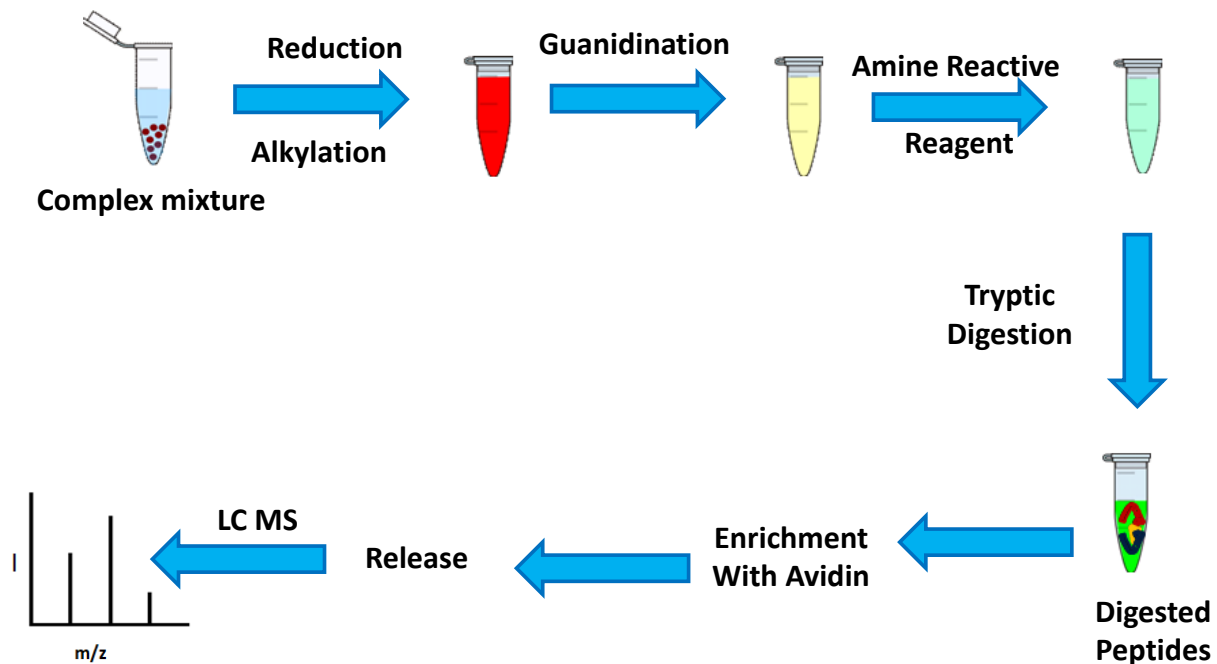
Figure 5-2 General scheme for enrichment of N-termini modified peptides using a complex mixture of samples

Our study on the proteolytic cleavage site identification highlights a new method to analyze the protein N-termini. This method facilitates the enrichment of only the N-termini peptide. By comparing to the general strategies, we envisage our method will provide a versatile analysis and further confirmation with the tandem mass spectrometry analysis. After we determine the utility of the novel capture reagent, we will demonstrate the proposed enrichment strategy in large-scale samples. Capture and release methods will be demonstrated for the N-terminal peptide enrichment in macrophages, and these raw macrophage cells will be stimulated with lipo-polysaccharide ligand prior to the modification. Hence, we will show the applicability of our method in real biological samples.

References

(1) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angewandte Chemie* **2005**, *44*, 7342-7372.

(2) Deribe, Y. L.; Pawson, T.; Dikic, I. Post-translational modifications in signal integration. *Nature structural & molecular biology* **2010**, *17*, 666-672.

(3) Mann, M.; Jensen, O. N. Proteomic analysis of post-translational modifications. *Nature biotechnology* **2003**, *21*, 255-261.

(4)Wang, Y. C.; Peterson, S. E.; Loring, J. F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell research* **2014**, *24*, 143-160.

(5) Basle, E.; Joubert, N.; Pucheault, M. Protein chemical modification on endogenous amino acids. *Chemistry & biology* **2010**, *17*, 213-227.

(6) Anderson, N. L.; Anderson, N. G. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **1998**, *19*, 1853-1861.

(7) Seo, J.; Lee, K. J. Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *Journal of biochemistry and molecular biology* **2004**, *37*, 35-44.

(8) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198-207.

(9)Cuthbert, G. L.; Daujat, S.; Snowden, A. W.; Erdjument-Bromage, H.; Hagiwara, T.; Yamada, M.; Schneider, R.; Gregory, P. D.; Tempst, P.; Bannister, A. J.; Kouzarides, T. Histone deimination antagonizes arginine methylation. *Cell* **2004**, *118*, 545-553.

(10) Kearney, P. L.; Bhatia, M.; Jones, N. G.; Yuan, L.; Glascock, M. C.; Catchings, K. L.; Yamada, M.; Thompson, P. R. Kinetic characterization of protein arginine deiminase 4: a transcriptional corepressor implicated in the onset and progression of rheumatoid arthritis. *Biochemistry* **2005**, *44*, 10570-10582.

(11) Slade, D. J.; Subramanian, V.; Fuhrmann, J.; Thompson, P. R. Chemical and biological methods to detect post-translational modifications of arginine. *Biopolymers* **2014**, *101*, 133-143.

(12) Thompson, P. R.; Fast, W. Histone citrullination by protein arginine deiminase: is arginine methylation a green light or a roadblock? *Acs Chem Biol* **2006**, *1*, 433-441.

(13) Gawron, D.; Ndah, E.; Gevaert, K.; Van Damme, P. Positional proteomics reveals differences in N-terminal proteoform stability. *Molecular systems biology* **2016**, *12*, 858.

(14) Prudova, A.; Serrano, K.; Eckhard, U.; Fortelny, N.; Devine, D. V.; Overall, C. M. TAILS N-terminomics of human platelets reveals pervasive metalloproteinase-dependent proteolytic processing in storage. *Blood* **2014**, *124*, e49-60.

(15) Kim, J. S.; Dai, Z.; Aryal, U. K.; Moore, R. J.; Camp, D. G., 2nd; Baker, S. E.; Smith, R. D.; Qian, W. J. Resin-assisted enrichment of N-terminal peptides for characterizing proteolytic processing. *Anal Chem* **2013**, *85*, 6826-6832.

(16) McDonald, L.; Robertson, D. H.; Hurst, J. L.; Beynon, R. J. Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. *Nature methods* **2005**, *2*, 955-957.

(17) Gevaert, K.; Goethals, M.; Martens, L.; Van Damme, J.; Staes, A.; Thomas, G. R.; Vandekerckhove, J. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nature biotechnology* **2003**, *21*, 566-569.

(18) Finehout, E. J.; Lee, K. H. An introduction to mass spectrometry applications in biological research. *Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology* **2004**, *32*, 93-100.

(19) Downard, K. M. 1912: a Titanic year for mass spectrometry. *Journal of mass spectrometry : JMS* **2012**, *47*, 1034-1039.

(20) Aminlashgari, N.; Hakkarainen, M. Surface assisted laser desorption ionization-mass spectrometry (SALDI-MS) for analysis of polyester degradation products. *J Am Soc Mass Spectrom* **2012**, *23*, 1071-1076.

(21) Liebisch, G.; Scherer, M. Quantification of bioactive sphingo- and glycerophospholipid species by electrospray ionization tandem mass spectrometry in blood. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2012**, *883-884*, 141-146.

(22) Fagerquist, C. K.; Sultan, O. A new calibrant for matrix-assisted laser desorption/ionization time-of-flight-time-of-flight post-source decay tandem mass spectrometry of non-digested proteins for top-down proteomic analysis. *Rapid communications in mass spectrometry : RCM* **2012**, *26*, 1241-1248.

(23) Karas, M.; Bachmann, D.; Hillenkamp, F. Influence of the Wavelength in High-Irradiance Ultraviolet-Laser Desorption Mass-Spectrometry of Organic-Molecules. *Anal Chem* **1985**, *57*, 2935-2939.

(24) Karas, M.; Bahr, U.; Dulcks, T. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius' journal of analytical chemistry* **2000**, *366*, 669-676.

(25) Konermann, L.; Vahidi, S.; Sowole, M. A. Mass spectrometry methods for studying structure and dynamics of biological macromolecules. *Anal Chem* **2014**, *86*, 213-232.

(26) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the mechanism of electrospray ionization. *Anal Chem* **2013**, *85*, 2-9.

(27) Banerjee, S.; Mazumdar, S. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International journal of analytical chemistry* **2012**, *2012*, 282574.

(28) Salih, B.; Masselon, C.; Zenobi, R. Matrix-assisted laser desorption/ionization mass spectrometry of noncovalent protein-transition metal ion complexes. *Journal of mass spectrometry : JMS* **1998**, *33*, 994-1002.

(29) Sleno, L.; Volmer, D. A. Ion activation methods for tandem mass spectrometry. *Journal of mass spectrometry : JMS* **2004**, *39*, 1091-1112.

(30) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *Journal of mass spectrometry : JMS* **2000**, *35*, 1399-1406.

(31) Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry* **1984**, *11*, 601.

(32) Bantscheff, M.; Kuster, B. Quantitative mass spectrometry in proteomics. *Analytical and bioanalytical chemistry* **2012**, *404*, 937-938.

(33) Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry* **2012**, *404*, 939-965.

(34) Pellati, F.; Orlandini, G.; Pinetti, D.; Benvenuti, S. HPLC-DAD and HPLC-ESI-MS/MS methods for metabolite profiling of propolis extracts. *Journal of pharmaceutical and biomedical analysis* **2011**, *55*, 934-948.

(35) Siebenberg, S.; Kaysser, L.; Wemakor, E.; Heide, L.; Gust, B.; Kammerer, B. Identification and structural elucidation of new caprazamycins from Streptomyces sp. MK730-62F2 by liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid communications in mass spectrometry : RCM* **2011**, *25*, 495-502.

(36) Morris, S. M., Jr. Arginine: beyond protein. *The American journal of clinical nutrition* **2006**, *83*, 508S-512S.

(37) Ahmed, N.; Thornalley, P. J.; Dawczynski, J.; Franke, S.; Strobel, J.; Stein, G.; Haik, G. M. Methylglyoxal-derived hydroimidazolone advanced glycation end-products of human lens proteins. *Investigative ophthalmology & visual science* **2003**, *44*, 5287-5292.

(38) Lopez-Clavijo, A. F.; Duque-Daza, C. A.; Canelon, I. R.; Barrow, M. P.; Kilgour, D.; Rabbani, N.; Thornalley, P. J.; O'Connor, P. B. Study of an Unusual Advanced Glycation End-Product (AGE) Derived from Glyoxal Using Mass Spectrometry. *J Am Soc Mass Spectr* **2014**, *25*, 673-683.

(39) Ahmed, N.; Argirov, O. K.; Minhas, H. S.; Cordeiro, C. A.; Thornalley, P. J. Assay of advanced glycation endproducts (AGEs): surveying AGEs by chromatographic assay with derivatization by 6-aminoquinolyl-N-hydroxysuccinimidyl-carbamate and application to Nepsilon-carboxymethyl-lysine- and Nepsilon-(1-carboxyethyl)lysine-modified albumin. *The Biochemical journal* **2002**, *364*, 1-14.

(40) Lazo, J. S.; Aslan, D. C.; Southwick, E. C.; Cooley, K. A.; Ducruet, A. P.; Joo, B.; Vogt, A.; Wipf, P. Discovery and biological evaluation of a new family of potent inhibitors of the dual specificity protein phosphatase Cdc25. *Journal of medicinal chemistry* **2001**, *44*, 4042-4049.

(41) Patthy, L.; Smith, E. L. Identification of Functional Arginine Residues in Ribonuclease-a and Lysozyme. *Journal of Biological Chemistry* **1975**, *250*, 565-569.

(42) Cotham, W. E.; Metz, T. O.; Ferguson, P. L.; Brock, J. W.; Hinton, D. J.; Thorpe, S. R.; Baynes, J. W.; Ames, J. M. Proteomic analysis of arginine adducts on glyoxal-modified ribonuclease. *Molecular & cellular proteomics : MCP* **2004**, *3*, 1145-1153.

(43) Taniguchi, K.; Kuyama, H.; Kajihara, S.; Tanaka, K. MALDI mass spectrometry-based sequence analysis of arginine-containing glycopeptides: improved fragmentation of glycan and peptide chains by modifying arginine residue. *Journal of mass spectrometry : JMS* **2013**, *48*, 951-960.

(44) Leitner, A.; Lindner, W. Functional probing of arginine residues in proteins using mass spectrometry and an arginine-specific covalent tagging concept. *Anal Chem* **2005**, *77*, 4481-4488.

(45) Deyl, Z.; Miksik, I. Post-translational non-enzymatic modification of proteins .1. Chromatography of marker adducts with special emphasis to glycation reactions. *J Chromatogr B* **1997**, *699*, 287-309.
(46) Julian, R. R.; Beauchamp, J. L. The unusually high proton affinity of aza-18-crown-6 ether: implications for the molecular recognition of lysine in peptides by lariat crown ethers. *J Am Soc Mass Spectrom* **2002**, *13*, 493-498.

(47 Krause, E.; Wenschuh, H.; Jungblut, P. R. The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. *Anal Chem* **1999**, *71*, 4160-4165.

(48) Krell, T.; Pitt, A. R.; Coggins, J. R. The Use of Electrospray Mass-Spectrometry to Identify an Essential Arginine Residue in Type-Ii Dehydroquinases. *Febs Lett* **1995**, *360,* 93-96.

(49) Friess, S. D.; Daniel, J. M.; Zenobi, R. Probing the surface accessibility of proteins with noncovalent receptors and MALDI mass spectrometry. *Phys Chem Chem Phys* **2004**, *6*, 2664-2675.

(50) Gauthier, M. A.; Klok, H. A. Arginine-specific modification of proteins with polyethylene glycol. *Biomacromolecules* **2011**, *12*, 482-493.

(51) Pande, C. S.; Pelzig, M.; Glass, J. D. Camphorquinone-10-Sulfonic Acid and Derivatives - Convenient Reagents for Reversible Modification of Arginine Residues. *P Natl Acad Sci-Biol* **1980**, *77*, 895-899.
(52) Takahashi, K. The reactions of phenylglyoxal and related reagents with amino acids. *Journal of biochemistry* **1977**, *81*, 395-402.

(53) Mendoza, V. L.; Vachet, R. W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass spectrometry reviews* **2009**, *28*, 785-815.

(54) Takagi, T.; Naito, Y.; Oya-Ito, T.; Yoshikawa, T. The role of methylglyoxal-modified proteins in gastric ulcer healing. *Current medicinal chemistry* **2012**, *19*, 137-144.

(55) Lo, T. W. C.; Selwood, T.; Thornalley, P. J. The reaction of methylglyoxal with aminoguanidine under physiological conditions and prevention of methylglyoxal binding to plasma proteins. *Biochemical Pharmacology* **1994**, *48*, 1865-1870.

(56) Suckau, D.; Mak, M.; Przybylski, M. Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping. *Proceedings of the National Academy of Sciences* **1992**, *89*, 5630-5634.

(57) Leitner, A.; Lindner, W. Probing of arginine residues in peptides and proteins using selective tagging and electrospray ionization mass spectrometry. *Journal of mass spectrometry : JMS* **2003**, *38*, 891-899.
(58) Itano, H. A.; Gottlieb, A. J. Blocking of Tryptic Cleavage of Arginyl Bonds by the Chemical Modification of the Guanido Group with Benzil. *Biochemical and biophysical research communications* **1963**, *12*, 405-408.

(59) Toi, K.; Bynum, E.; Norris, E.; Itano, H. A. Studies on the chemical modification of arginine. I. The reaction of 1,2-cyclohexanedione with arginine and arginyl residues of proteins. *J Biol Chem* **1967**, *242*, 1036-1043.

(60) Yankeelov, J. A., Jr. Modification of arginine in proteins by oligomers of 2,3-butanedione. *Biochemistry* **1970**, *9*, 2433-2439.

(61) Leitner, A.; Lindner, W. Chemistry meets proteomics: the use of chemical tagging reactions for MS-based proteomics. *Proteomics* **2006**, *6*, 5418-5434.

(62) Akinsiku, O. T.; Yu, E. T.; Fabris, D. Mass spectrometric investigation of protein alkylation by the RNA footprinting probe kethoxal. *Journal of mass spectrometry : JMS* **2005**, *40*, 1372-1381.

(63) Patthy, L.; Smith, E. L. Identification of functional arginine residues in ribonuclease A and lysozyme. *J Biol Chem* **1975**, *250*, 565-569.

(64) Bilova, T.; Paudel, G.; Shilyaev, N.; Schmidt, R.; Brauch, D.; Tarakhovskaya, E.; Milrud, S.; Smolikova, G.; Tissier, A.; Vogt, T.; Sinz, A.; Brandt, W.; Birkemeyer, C.; Wessjohann, L. A.; Frolov, A. Global proteomic analysis of advanced glycation end products in the Arabidopsis proteome provides evidence for age-related glycation hot spots. *J Biol Chem* **2017**, *292*, 15758-15776.

(65) Saraiva, M. A.; Borges, C. M.; Florencio, M. H. Mass spectrometric studies of the reaction of a blocked arginine with diketonic alpha-dicarbonyls. *Amino Acids* **2016**, *48*, 873-885.

(66) Makino, H.; Shikata, K.; Kushiro, M.; Hironaka, K.; Yamasaki, Y.; Sugimoto, H.; Ota, Z.; Araki, N.; Horiuchi, S. Roles of advanced glycation end-products in the progression of diabetic nephropathy. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* **1996**, *11 Suppl 5*, 76-80.

(67) Smith, M. A.; Richey, P. L.; Taneda, S.; Kutty, R. K.; Sayre, L. M.; Monnier, V. M.; Perry, G. Advanced Maillard reaction end products, free radicals, and protein oxidation in Alzheimer's disease. *Annals of the New York Academy of Sciences* **1994**, *738*, 447-454.

(68) Yamagishi, S.; Nakamura, K.; Matsui, T. Regulation of advanced glycation end product (AGE)-receptor (RAGE) system by PPAR-gamma agonists and its implication in cardiovascular disease. *Pharmacological research* **2009**, *60*, 174-178.

(69) Rabbani, N.; Xue, M.; Thornalley, P. J. Methylglyoxal-induced dicarbonyl stress in aging and disease: first steps towards glyoxalase 1-based treatments. *Clinical science* **2016**, *130*, 1677-1696.

(70) Chumsae, C.; Gifford, K.; Lian, W.; Liu, H.; Radziejewski, C. H.; Zhou, Z. S. Arginine modifications by methylglyoxal: discovery in a recombinant monoclonal antibody and contribution to acidic species. *Anal Chem* **2013**, *85*, 11401-11409.

(71) Muranova, L. K.; Perfilov, M. M.; Serebryakova, M. V.; Gusev, N. B. Effect of methylglyoxal modification on the structure and properties of human small heat shock protein HspB6 (Hsp20). *Cell stress & chaperones* **2016**, *21*, 617-629.

(72) Cheung, S. T.; Fonda, M. L. Reaction of phenylglyoxal with arginine. The effect of buffers and pH. *Biochemical and biophysical research communications* **1979**, *90*, 940-947.

(73) Selwood, T.; Thornalley, P. J. Binding of methylglyoxal to albumin and formation of fluorescent adducts. Inhibition by arginine, N-alpha-acetylarginine and aminoguanidine. *Biochemical Society transactions* **1993**, *21*, 170S.

(74) Brock, J. W.; Cotham, W. E.; Thorpe, S. R.; Baynes, J. W.; Ames, J. M. Detection and identification of arginine modifications on methylglyoxal-modified ribonuclease by mass spectrometric analysis. *Journal of mass spectrometry : JMS* **2007**, *42*, 89-100.

(75) Wanigasekara, M. S.; Chowdhury, S. M. Evaluation of chemical labeling methods for identifying functional arginine residues of proteins by mass spectrometry. *Analytica chimica acta* **2016**, *935*, 197-206.

(76) Chowdhury, S. M.; Du, X. X.; Tolic, N.; Wu, S.; Moore, R. J.; Mayer, M. U.; Smith, R. D.; Adkins, J. N. Identification of Cross-Linked Peptides after Click-Based Enrichment Using Sequential Collision-Induced Dissociation and Electron Transfer Dissociation Tandem Mass Spectrometry. *Anal Chem* **2009**, *81*, 5524-5532.

(77) Smits, A. H.; Borrmann, A.; Roosjen, M.; van Hest, J. C. M.; Vermeulen, M. Click-MS: Tagless Protein Enrichment Using Bioorthogonal Chemistry for Quantitative Proteomics. *Acs Chem Biol* **2016**, *11*, 3245-3250.

(78) Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. Labeling live cells by copper-catalyzed alkyne--azide click chemistry. *Bioconjugate chemistry* **2010**, *21*, 1912-1916.

(79) Speers, A. E.; Cravatt, B. F. Activity-Based Protein Profiling (ABPP) and Click Chemistry (CC)-ABPP by MudPIT Mass Spectrometry. *Current protocols in chemical biology* **2009**, *1*, 29-41.

(80) Brock, J. W.; Hinton, D. J.; Cotham, W. E.; Metz, T. O.; Thorpe, S. R.; Baynes, J. W.; Ames, J. M. Proteomic analysis of the site specificity of glycation and carboxymethylation of ribonuclease. *Journal of proteome research* **2003**, *2*, 506-513.

(81) Bradshaw, R. A.; Brickey, W. W.; Walker, K. W. N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families. *Trends in biochemical sciences* **1998**, *23*, 263-267.

(82) Chowdhury, S. M.; Munske, G. R.; Siems, W. F.; Bruce, J. E. A new maleimide-bound acid-cleavable solid-support reagent for profiling phosphorylation. *Rapid communications in mass spectrometry : RCM* **2005**, *19*, 899-909.

(83) Galan, J. A.; Guo, M.; Sanchez, E. E.; Cantu, E.; Rodriguez-Acosta, A.; Perez, J. C.; Tao, W. A. Quantitative analysis of snake venoms using soluble polymer-based isotope labeling. *Molecular & cellular proteomics : MCP* **2008**, *7*, 785-799.

(84) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nature biotechnology* **2002**, *20*, 512-515.

(85) Azad, G. K.; Tomar, R. S. Proteolytic clipping of histone tails: the emerging role of histone proteases in regulation of various biological processes. *Molecular biology reports* **2014**, *41*, 2717-2730.

(86) Biniossek, M. L.; Niemer, M.; Maksimchuk, K.; Mayer, B.; Fuchs, J.; Huesgen, P. F.; McCafferty, D. G.; Turk, B.; Fritz, G.; Mayer, J.; Haecker, G.; Mach, L.; Schilling, O. Identification of Protease Specificity by Combining Proteome-Derived Peptide Libraries and Quantitative Proteomics. *Molecular & cellular proteomics : MCP* **2016**, *15*, 2515-2524.

(87) Brown, M. S.; Ye, J.; Rawson, R. B.; Goldstein, J. L. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* **2000**, *100*, 391-398.

(88) Klingler, D.; Hardt, M. Profiling protease activities by dynamic proteomics workflows. *Proteomics* **2012**, *12*, 587-596.

(89) Chen, T.; Muratore, T. L.; Schaner-Tooley, C. E.; Shabanowitz, J.; Hunt, D. F.; Macara, I. G. N-terminal alpha-methylation of RCC1 is necessary for stable chromatin association and normal mitosis. *Nature cell biology* **2007**, *9*, 596-603.

(90) Deng, J.; Zhang, G.; Huang, F. K.; Neubert, T. A. Identification of protein N-termini using TMPP or dimethyl labeling and mass spectrometry. *Methods in molecular biology* **2015**, *1295*, 249-258.

(91) Xu, G.; Shin, S. B.; Jaffrey, S. R. Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 19310-19315.

(92) Bond, J. S.; Butler, P. E. Intracellular proteases. *Annual review of biochemistry* **1987**, *56*, 333-364.
(93) Duffy, M. J. Do proteases play a role in cancer invasion and metastasis? *European journal of cancer & clinical oncology* **1987**, *23*, 583-589.

(94) Neurath, H.; Walsh, K. A. Role of proteolytic enzymes in biological regulation (a review). *Proceedings of the National Academy of Sciences of the United States of America* **1976**, *73*, 3825-3832.

(95) Zhang, X.; Ye, J.; Engholm-Keller, K.; Hojrup, P. A proteome-scale study on in vivo protein Nalpha-acetylation using an optimized method. *Proteomics* **2011**, *11*, 81-93.

(96) Beardsley, R. L.; Reilly, J. P. Optimization of guanidination procedures for MALDI mass mapping. *Anal Chem* **2002**, *74*, 1884-1890.

(97) Olivero, S.; Novakovitch, G.; Ladaique, P.; Sainty, D.; Viret, F.; Faucher, C.; Stoppa, A. M.; Blaise, D.; Chabannon, C. T-cell depletion of allogeneic bone marrow grafts: enrichment of mononuclear cells using the COBE Spectra cell processor, followed by immunoselection of CD34(+) cells. *Cytotherapy* **1999**, *1*, 469-477.

(98) Rocchi, P.; Boudouresque, F.; Zamora, A. J.; Muracciole, X.; Lechevallier, E.; Martin, P. M.; Ouafik, L. Expression of adrenomedullin and peptide amidation activity in human prostate cancer and in human prostate cancer cell lines. *Cancer research* **2001**, *61*, 1196-1206.

(99) Schilling, O.; Barre, O.; Huesgen, P. F.; Overall, C. M. Proteome-wide analysis of protein carboxy termini: C terminomics. *Nature methods* **2010**, *7*, 508-511.

(100) Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. Improved matrix-assisted laser desorption/ionization mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine-containing peptides. *Rapid communications in mass spectrometry : RCM* **2000**, *14*, 2070-2073.

(101) Thinon, E.; Serwa, R. A.; Broncel, M.; Brannigan, J. A.; Brassat, U.; Wright, M. H.; Heal, W. P.; Wilkinson, A. J.; Mann, D. J.; Tate, E. W. Global profiling of co- and post-translationally N-myristoylated proteomes in human cells. *Nature communications* **2014**, *5*, 4919.

(102) Mahrus, S.; Trinidad, J. C.; Barkan, D. T.; Sali, A.; Burlingame, A. L.; Wells, J. A. Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* **2008**, *134*, 866-876.

(103) Verhelst, S. H. L. Intramembrane proteases as drug targets. *The FEBS journal* **2017**, *284*, 1489-1502.

(104) Yamasaki, K.; Di Nardo, A.; Bardan, A.; Murakami, M.; Ohtake, T.; Coda, A.; Dorschner, R. A.; Bonnart, C.; Descargues, P.; Hovnanian, A.; Morhenn, V. B.; Gallo, R. L. Increased serine protease activity and cathelicidin promotes skin inflammation in rosacea. *Nature medicine* **2007**, *13*, 975-980.

(105) Mommen, G. P.; van de Waterbeemd, B.; Meiring, H. D.; Kersten, G.; Heck, A. J.; de Jong, A. P. Unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phospho tagging (PTAG) and TiO(2)-based depletion. *Molecular & cellular proteomics : MCP* **2012**, *11*, 832-842.

(106) Kleifeld, O.; Doucet, A.; auf dem Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R. K.; Starr, A. E.; Foster, L. J.; Kizhakkedathu, J. N.; Overall, C. M. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nature biotechnology* **2010**, *28*, 281-288.

(107) Ye, J. Roles of regulated intramembrane proteolysis in virus infection and antiviral immunity. *Biochimica et biophysica acta* **2013**, *1828*, 2926-2932.

Oct 28, 2017

This Agreement between Maheshika Wanigasekara ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4200841099824
License date	Oct 02, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	Analytica Chimica Acta
Licensed Content Title	Evaluation of chemical labeling methods for identifying functional arginine residues of proteins by mass spectrometry
Licensed Content Author	Maheshika S.K. Wanigasekara,Saiful M. Chowdhury
Licensed Content Date	Sep 7, 2016
Licensed Content Volume	935
Licensed Content Pages	10
Start Page	197
End Page	206
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Identification of proteolytic cleavage and arginine modifications by chemical labeling and mass spectrometry
Expected completion date	Dec 2017
Estimated size (number of pages)	200
Requestor Location	Maheshika Wanigasekara 1020 W Abram Street Apt 155 Arlington, TX 76013 United States

Vita

Maheshika S.K. Wanigasekara

The author was born in Kandy Sri Lanka. She obtained her Bachelor of Science degree in chemistry from the College of Sciences-Institute of Chemistry Sri Lanka in May 2010. During her undergraduate career she started working with Dr. R.M.G Rajapaksha in the Department of Chemistry, faculty of science, University of Peradeniya, researching hydroxyapatite coating on stainless steel implants. While she was working as a chemistry instructor in a private school, she also assisted Dr. Anoja Wanigasekara in her research work in the Department of Basic Veterinary Science as part of the Faculty of Veterinary Medicine & Animal Science, University of Peradeniya. Maheshika was a visiting scholar at the time. In the fall of 2012, she joined the Department of Chemistry and Biochemistry at the University of Texas at Arlington and joined Dr. Saiful Chowdhury's group in January 2013. Her research has been based on identification of proteolytic cleavage and arginine post-translational modification by chemical labeling and mass spectrometry. She received her Ph.D. degree in Analytical Chemistry from the University of Texas at Arlington in December 2017. In the summer of 2017, she worked for Vertex Pharmaceuticals in Boston as an analytical development summer intern. This confirmed her career plan of working in or with the pharmaceutical industry.