2023

Identification of acidic residues in proteins by selective chemical labeling and tandem mass spectrometry

FNU Ashima

Identification of acidic residues in proteins by selective chemical labeling and tandem mass spectrometry

by

ASHIMA MEHTA

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN CHEMISTRY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2023

Acknowledgments

I would like to extend my gratitude to my supervisor, Dr. Saiful M. Chowdhury for his mentorship in my research and academic career. His unwavering support has been invaluable during my journey. He has always been a constant source of motivation for me. I sincerely appreciate his valuable guidance throughout my journey.

I'd like to acknowledge the contributions of my committee members, Dr. Purnendu Dasgupta and Dr. Joseph Buonomo for their valuable feedback and insights during my project. My heartfelt appreciation goes to Dr. William Cleaver and Dr. Cynthia Griffith for their assistance with my teaching assistant responsibilities during my masters degree. I feel grateful to be a part of the Department of Chemistry and Biochemistry. I am also thankful to Stephanie Henry, Debbie Cooke, and Dorothy Sullivan for their tremendous support. I would like to thank my lab members Adway, Sharel, Akash, and Aurchie for teaching and helping me in the lab. I would like to express my sincere thanks to Dr. Brian Edwards and Dr. Roy McDougald for their guidance in operating and troubleshooting lab instruments. I would like to thank my friends here at UTA and in India for their support.

I am extremely grateful to my parents and family for believing in me and sending me to the United States for higher studies. I would have never achieved any of this without their support. I would also like to thank my younger sister for always supporting me emotionally. I am very thankful to my partner for his endless support.

Abstract

Identification of acidic residues in proteins by selective chemical labeling and tandem mass spectrometry

Ashima Mehta, M.S.

The University of Texas at Arlington, 2023

Supervising professor: Saiful M. Chowdhury

Proteins play important role in carrying out a wide range of cellular activities. They are subjected to various post-translational modifications (PTMs) by the addition of functional groups such as acetyl, phosphoryl, and methyl which leads to changes in their biological functions, localization, activity, and structure. The growth of MS technology has immensely facilitated the identification of PTMs comprehensively. This dissertation focuses on mapping these changes using covalent labeling techniques by tagging carboxylic acid residues, which undergo various kinds of PTMs. The second chapter focuses on the methods designed for Affinity Purification Mass Spectrometry (AP-MS) chemical cross-linking (CXL) of protein complexes.

There are numerous methods available for bioconjugation which target cysteine, lysine, and arginine residues. Although carboxyl residues, which are prevalent in proteins, are essential for preserving the protein's functionality, there are currently few accessible selective labeling procedures. We have described a novel reactive probe that allows for the chemoselective modification of acidic residues in peptides and proteins. We evaluated the reactivity of diphenyldiazomethane (DPDAM) in peptides and proteins in the third chapter. The fragmentation patterns of these labeled peptides have been investigated using tandem mass spectrometry.

Table of Contents

Acknowledgment……iii

Abstract…….....iv

List of Abbreviations ……...……. vii

List of Illustrations……...……...……. viii

Chapter 1 Introduction……..1

1.1 Proteomics…….1

1.2 Protein labeling studies in proteomics……...2

1.3 Chemical strategies for proteomics……3

1.4 Covalent labeling studies in proteomics……5

1.5 Mass spectrometry-based proteomics……5

Chapter 2 Protein-protein interaction network mapping by affinity purification cross-linking mass spectrometry (AP-XL-MS) based proteomics ……...13

2.1 Abstract ……13

2.2 Introduction ……13

2.3 Materials……...…17

2.4 Methods……...….20

2.5 Notes ……27

Chapter 3 Identification of acidic residues in proteins by selective chemical labeling and tandem mass spectrometry…….…….32

3.1 Abstract……32

3.2 Introduction……..32

3.3 Experimental Section……...35

3.3.1 Materials……..……...35

3.3.2 In-solution digestion……..36

3.3.3 Instrumentation…….36

3.3.4 Chemical modification of Neurotensin, Angiotensin I, and Bradykinin with DPDAM….....37

3.3.5 Chemical modification of Bovine Serum Albumin (BSA) ……...…….37

3.4 Results and Discussion…….38

3.4.1 Labeling of peptides by DPDAM.....38

3.4.2 Labeling of proteins by DPDAM.....45

3.5 Conclusion……47

References…….…48

Biographical information……...50

List of Abbreviations

Abbreviations	Name
BSA	Bovine Serum Albumin
CID	Collision Induced Dissociation
HCD	High energy Collisional Dissociation
DTT	Dithiothreitol
HPLC	High performance liquid chromatography
LC-MS	Liquid chromatography mass spectrometry
ESI	Electrospray ionization
MS/MS	Tandem mass spectrometry
PTMs	Post Translational Modifications
MES	Morpholinoethanesulfonic acid monohydrate

List of Illustrations

Figure 1-1 Major applications of proteomics……...……1

Figure 1-2 Peptide sequence fragmentation modeling……...……..6

Figure 1-3 Workflow of bottom-up proteomics……...……8

Figure 2-1 Workflow of AP-CXL-MS……...……..…….16

Figure 3-1 Mechanism of O- alkylation of carboxylic acid using diazo compounds……..……..35

Figure 3-2 ESI-MS and MS/MS spectra of modified peptides with DPDAM at pH 4. (A) Neurotensin labeled with DPDAM at *m/z* 920.55 $[M+2H+DPDAM]^{2+}$ and 614.03 $[M+3H+DPDAM]^{3+}$ (B) MS/MS spectrum of modified neurotensin at *m/z* 614.03 $[M+3H+DPDAM]^{3+}$. *denotes DPDAM-modified sites……...……40

Figure 3-3 ESI-MS and MS/MS spectra of modified peptides with DPDAM at pH 4. (A) Angiotensin I labeled with DPDAM at *m/z* 732.37 $[M+2H+DPDAM]^{2+}$ and 488.60 $[M+3H+DPDAM]^{3+}$. (B) CID-MS/MS spectrum of modified angiotensin I at *m/z* 732.35 $[M+2H+DPDAM]^{2+}$. *denotes DPDAM modified sites.……..……41

Figure 3-4 ESI-MS and MS/MS spectra of modified peptides with DPDAM at pH 2. (A) Neurotensin labeled with DPDAM at *m/z* 920.55 $[M+2H+DPDAM]^{2+}$ and 614.03 $[M+3H+DPDAM]^{3+}$ (B) MS/MS spectrum of modified neurotensin at *m/z* 614.03 $[M+3H+DPDAM]^{3+}$. *denotes DPDAM-modified sites.……..……43

Figure 3-5 ESI-MS and MS/MS spectra of modified peptides with DPDAM. (A) Bradykinin labeled with DPDAM at *m/z* 614.16 $[M+2H+DPDAM]^{2+}$ (B) MS/MS spectrum of modified bradykinin at *m/z* 614.16 $[M+2H+DPDAM]^{2+}$. *denotes DPDAM-modified sites……...….44

Figure 3-6 ESI-MS and MS/MS spectra of modified BSA with DPDAM. (A) BSA labeled with DPDAM at *m/z* 1002.4719 $[M+4H+DPDAM]^{4+}$ (B) MS/MS spectrum of modified peptide at *m/z* 1002.4719 $[M+4H+DPDAM]^{4+}$. *denotes DPDAM-modified sites……...…….46
Figure 3-7 Crystal structure of BSA protein (PDB 3v03) The figure depicts modified sites of BSA protein with DPDAM…….46

Chapter 1 Introduction

1.1 Proteomics

Advances in genomic research, particularly the Human Genome Project, have allowed scientists to sequence and map the entire human genome. Genomics has paved the way for proteomics. Proteins are biomolecules that play a vital role in processes such as cell-to-cell interactions, apoptosis, cell growth, gene expression, transportation, and cell signaling. The set of proteins found in a cell[1] is referred to as proteome. The thorough functional analysis of the proteome is known as proteomics[2]. Proteomics offers important insights into how cells work, including biological pathways, protein-protein interactions[3], post-translational changes, and protein dynamics.

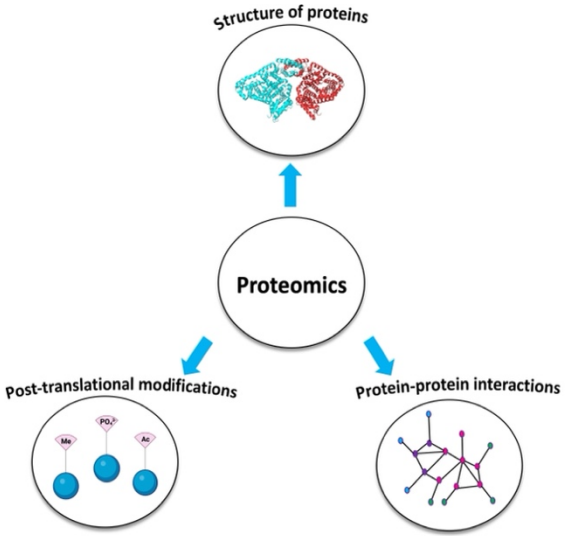


Figure 1-1 Major applications of proteomics

Proteomics can bridge the gap between genomic data and biological processes by examining proteins. As a result, we now have a deeper understanding of how genes are translated into useful proteins and how these proteins affect the general function and traits of cells. It's crucial

to keep in mind that, whereas genomics offers a blueprint of an organism's genetic information, proteomics offers insights into the actual functional molecules in charge of cellular functions, ultimately resulting in a deeper understanding of the biology and complexity of living systems.

1.2 Protein labeling studies in Proteomics

With the availability of complete genome sequences, proteomics has emerged as a field focused on studying the proteins encoded by those genomes. The field of proteomics aims to develop and apply methodologies that accelerate the functional analysis of proteins. Around 30,000 to 40,000 genes could potentially encode 40,000 distinct proteins, and the number of proteins or protein fragments resulting from alternative RNA splicing and post-translational modifications (PTMs) is thought to be substantially higher, potentially reaching up to 2,000,000[4].

A critical component of proteomics is the labeling of proteins with specific labels to enable their identification, quantification, and structural determination[5]. The analysis of post-translational modifications (PTMs) is a crucial part of protein labeling investigations in proteomics. PTMs are chemical modifications that take place after protein synthesis and have a big impact on protein interactions, stability, function, and localization. These PTMs like phosphorylation, acetylation, methylation, and glycosylation can be evaluated by adding certain labels or tags which shed light on their functions in cellular processes, disease mechanisms, and possible therapeutic targets by revealing insights into their dynamics, stoichiometry, and functional effects.

Moreover, protein labeling also facilitates the visualization of proteins in organelles, tissues, and cells. It also makes it possible to analyze dynamic processes, subcellular compartmentalization, and protein localization. The identification of interacting partners is another

application of protein labeling. Protein labeling experiments in proteomics are essential for thorough proteome research. We can quantify proteins, identify particular proteins of interest, observe protein localization and interactions, and examine post-translational modifications by tagging proteins. The knowledge gained from these studies into how proteins are expressed, used, and regulated in biological systems has a wide-ranging impact on biomedical research, drug development, and other domains. The study of protein-protein interactions, protein structures, and PTMs reveals the function of proteins in cells.

1.3 Chemical Strategies for Proteomics

A key step in the development of new therapeutics lies in designing new reagents for labeling proteins. MALDI-MS and ESI-MS are common methods for peptide mass mapping followed by peptide sequencing with tandem MS. Availability of the amino acid residues to the reagent determines how reactive the various functional groups are in a protein. These amino acid residues undergo modifications with specific reagents which can be used to monitor the structure of a protein by utilizing MS. Identifying the specific binding sites of proteins is crucial in the field of drug discovery. These reactive residues can be identified using chemical modifications[6]. Achieving a specific reaction at a particular site in a complex mixture is imperative. Due to the high abundance of lysine and arginine amino acids on the protein periphery, these residues were the focus of the majority of these efforts in the past. Our group has recently evaluated the reactivity of arginine residues which has a strong affinity to dicarbonyl compounds such as 1,2 cyclohexanedione[7]. Along with this compound, an enrichment handle was also introduced to purify the reactive peptides using click chemistry and biotin-avidin chromatography. The utilization of arginine labeling combined with mass spectrometry (MS) is widely employed to

investigate protein structures. Arginine is a basic amino acid and remains protonated under physiological conditions consistently. Those residues which are present on the surface and are reactive are the ones that mainly get modified with these reagents. This study facilitates the investigation of protein surface structure and enables a thorough characterization of their spatial distribution. Additionally, it makes it possible to explore the post-translational modifications that occur on these arginine residues, revealing their functional and regulatory functions. This strategy also enables a confident evaluation of the enzymatic alterations taking place at arginine sites by profiling enzyme activities, facilitating a thorough understanding of protein dynamics and function.

The application of cysteine-specific chemical modifications in combination with mass spectrometry (MS) can serve as a highly valuable approach for mapping alteration on protein surfaces. Selectively targeting cysteines enables the identification and characterization of specific modifications occurring at those sites, providing a comprehensive understanding of how these changes impact protein structure and function. Various reagents such as iodoacetic acid and iodoacetamide can be used to map cysteine residues. There is a growing demand to expand the repertoire of amino acids that can be selectively modified.

1.4 Covalent labeling studies

The function of a protein is governed by its higher-order structure, hence approaches that shed light on protein structure are crucial for comprehending protein reactivity. Nuclear Magnetic Resonance (NMR) spectroscopy and X-ray crystallography provide atomic-level resolution and have traditionally been used to provide insights into protein structure. Cryo-electron microscopy (cryo-EM), which was recently developed, is another extremely effective method for analyzing

protein structure. However, in some circumstances, sample quantities, or sample incompatibility prevent these approaches from providing adequate insight into protein's higher-order structure. For protein therapeutic design, their structural information is crucial. When it comes to reaching higher structural resolution, MS-based proteomics technologies have distinct benefits over conventional methods like circular dichroism (CD) spectroscopy or fluorescence spectroscopy including sensitivity, specificity, and speed. MS in conjunction with methods such as covalent labeling[8], crosslinking[9], and noncovalent labeling can unravel information about a protein's three-dimensional structure in solution.

Covalent labeling technique can determine protein's structural information by comparing how differently amino acids respond to a given label. In these investigations, it is implicitly assumed that amino acids that are on the periphery of the proteins are more likely to get modified with the labeling reagent than the embedded ones. The extent of labeling depends on the conformational changes of proteins.

1.5 Mass spectrometry-based proteomics

A comprehensive understanding of proteins within the context of cellular and molecular biology is by no means an easy endeavor, given the diversity and complexity of protein structures, changes, and interactions. For the visualization, identification, and quantitation of proteins gel-electrophoresis, protein microarrays, immunoassays, Edman sequencing, and imaging techniques are just a few of the methods and techniques that have been in use. However, the development of sensitive and specific mass spectrometry (MS) technique is certainly one of the most significant advancements in the field of global, comprehensive proteomics research. For the analysis of biological materials, mass spectrometry is frequently combined with chromatography.[10]

The key components of a mass spectrometer include an ion source, analyzer, and detector. Samples are first introduced in the system through an ion source, and then the ions are isolated depending on their mass-to-charge ratio (m/z) in a mass analyzer followed by their measurement and amplification by the detector. Two soft Ionization techniques that are broadly employed in proteomics include matrix-assisted laser desorption/ionization (MALDI)[11] and electrospray ionization (ESI)[12]. Soft ionization refers to the ionization technique in which the integrity of the entire molecule can be preserved so that the m/z of the native peptides or proteins can be monitored. In MALDI, a laser is irradiated on the matrix of small molecules combined with analyte causing the analyte molecules to enter the gas phase without being fragmented. In the case of ESI, the intense electric field is produced by a high voltage applied at the capillary's outlet and atomizes them into small, charged droplets.[13] After the evaporation of the solvent, there is a high charge density on the droplet's surface eventually the droplet splits into one or several charged ions and enters the gas phase. While ESI produces multiply charged ions in different charge states, MALDI produces predominantly singly charged ions.

Ionized molecules are then transferred to the analyzer to be separated depending on their mass-to-charge ratio. Ion trap (IT)[14], quadrupole (Q), time-of-flight (TOF)[15], ion cyclotron resonance (ICR)[16], and Orbitrap[17] are some of the mass analyzers frequently used in proteomics.

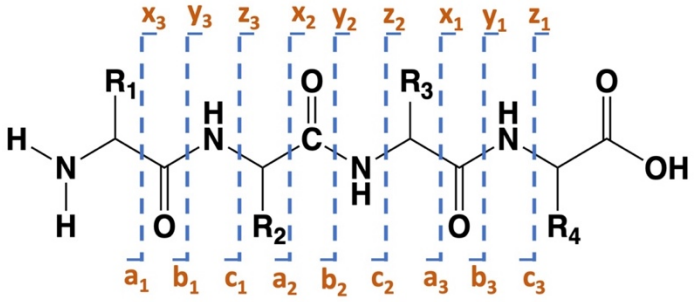


Figure 1-2 Peptide sequence fragmentation modeling

Each analyzer has unique qualities including sensitivity, scan rate, and mass accuracy. In proteomic research, hybrid analyzers are used to improve performance and capabilities. Q-TOF, QQQ, and IT-Orbitrap are a few examples of hybrid mass spectrometers that enable fragmentation to produce product ions tailored to peptide sequences. This fragmentation, known as tandem mass spectrometry, involves several stages of MS acquisition. The precursor ion, corresponding to the intact peptide or protein's m/z in the MS^1 stage, gives rise to product ions (or fragment ions) in the MS^2 stage. IT analyzers have the capability to perform additional fragmentation of the fragments in the MS^2 level, leading to the generation of MS^3 or even MS^n spectra. This capability is valuable as it can provide enhanced qualitative and quantitative information tailored to specific proteomic applications. The fragmentation pattern observed in tandem MS spectra, which is crucial for identifying peptide sequences, is determined by the choice of dissociation methods. Among these methods, collision-induced dissociation (CID) is widely employed. CID involves inducing the dissociation of peptides by subjecting them to collisions with an inert gas within a fragmentation chamber generating *b* and *y*-ions[18]. This collision-based process imparts energy to the peptides, causing them to break apart and generate characteristic fragment ions in the spectrum. Electron Transfer Dissociation (ETD) is another fragmentation technique used in mass spectrometry. It involves the capture of an electron from a reagent anion by positively charged precursor ions, followed by the dissociation of the peptide at N-Cα bonds thereby generating *c* and *z*-ions. ETD differs from CID in that it preserves many labile sites that would be fragmented in CID and is less influenced by peptide sequences. This characteristic makes ETD a valuable tool for studying post-translational modifications (PTMs). By preserving labile modifications, ETD allows for the identification and characterization of PTMs in peptides.[19, 20]

When it comes to protein identification using mass spectrometry (MS) instruments, two main strategies are commonly employed: the bottom-up approach and the top-down approach. These strategies can be used individually or in combination to provide complementary information. The top-down proteomic approach involves the analysis of intact proteins by directly ionizing them and then subjecting them to fragmentation[21]. This approach typically requires a high-resolution mass analyzer to accurately determine protein masses. The top-down approach offers advantages such as comprehensive coverage of protein variants, characterization of protein isoforms, and identification of post-translational modifications (PTMs). However, working with intact proteins presents challenges in terms of solubilization, fractionation, separation, ionization, fragmentation, and bioinformatic analysis.[22]

On the other hand, in bottom-up proteomics proteins are converted into smaller peptide fragments using proteases such as trypsin. These peptides are then analyzed by MS. It allows for easier sample handling, higher sensitivity, and the ability to identify a larger number of peptides and proteins. I have worked on bottom-up proteomics during my research work.

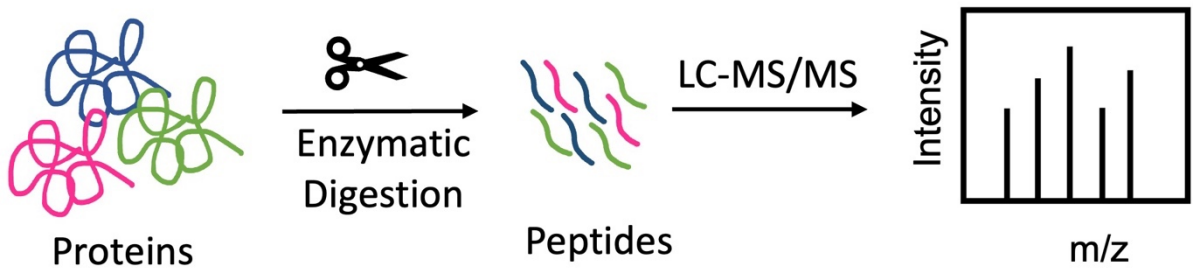


Figure 1-3 Workflow of bottom-up proteomics strategy

The next chapter describes methods employed for mapping the protein-protein interaction network using affinity purification (AP-MS-XL) technology. In this chapter, we have discussed the methods for preparing the samples such as cell culture, cell lysis, protein purification, cell

treatments, in-solution or in-gel digestion, protein separation, liquid chromatography, and mass spectrometry, followed by data analysis of the samples.

References

(1) Pfanner, N.; Warscheid, B.; Wiedemann, N. Mitochondrial proteins: from biogenesis to functional networks. *Nature Reviews Molecular Cell Biology* **2019**, *20* (5), 267-284. DOI: 10.1038/s41580-018-0092-0 (acccessed 2023-08-03T00:41:16).

(2) Dupree, E. J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B. A.; Darie, C. C. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes* **2020**, *8* (3), 14. DOI: 10.3390/proteomes8030014 (acccessed 2023-08-03T00:54:05).

(3) Chakrabarty, J. K.; Naik, A. G.; Fessler, M. B.; Munske, G. R.; Chowdhury, S. M. Differential Tandem Mass Spectrometry-Based Cross-Linker: A New Approach for High Confidence in Identifying Protein Cross-Linking. *Analytical Chemistry* **2016**, *88* (20), 10215-10222. DOI: 10.1021/ACS.ANALCHEM.6B02886/SUPPL_FILE/AC6B02886_SI_003.XLSX.

(4) Mendoza, V. L.; Vachet, R. W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrometry Reviews* **2009**, *28* (5), 785-815. DOI: 10.1002/mas.20203 (acccessed 2023-07-29T07:37:12).

(5) Adam, G. C.; Sorensen, E. J.; Cravatt, B. F. Chemical Strategies for Functional Proteomics. *Molecular & Cellular Proteomics* **2002**, *1* (10), 781-790. DOI: 10.1074/mcp.r200006-mcp200 (acccessed 2023-08-03T05:26:25).

(6) Kjærsgaard, N. L.; Nielsen, T. B.; Gothelf, K. V. Chemical Conjugation to Less Targeted Proteinogenic Amino Acids. *ChemBioChem* **2022**, *23* (19). DOI: 10.1002/cbic.202200245 (acccessed 2023-07-29T07:50:20).

(7) Wanigasekara, M. S. K.; Huang, X.; Chakrabarty, J. K.; Bugarin, A.; Chowdhury, S. M. Arginine-Selective Chemical Labeling Approach for Identification and Enrichment of Reactive Arginine Residues in Proteins. *ACS Omega* **2018**, *3* (10), 14229-14235. DOI: 10.1021/acsomega.8b01729 (acccessed 2022-10-20T15:02:20).

(8) Limpikirati, P.; Liu, T.; Vachet, R. W. Covalent labeling-mass spectrometry with non-specific reagents for studying protein structure and interactions. *Methods* **2018**, *144*, 79-93. DOI: 10.1016/j.ymeth.2018.04.002 (acccessed 2023-07-29T08:09:44).

(9) Chakrabarty, J. K.; Sadananda, S. C.; Bhat, A.; Naik, A. J.; Ostwal, D. V.; Chowdhury, S. M. High confidence identification of cross-linked peptides by an enrichment-based dual cleavable cross-linking technology and data analysis tool cleave-XL. *Journal of the American Society for Mass Spectrometry* **2020**, *31* (2), 173-182. DOI: 10.1021/JASMS.9B00111.

(10) Finehout, E. J.; Lee, K. H. An introduction to mass spectrometry applications in biological research. *Biochemistry and Molecular Biology Education* **2004**, *32* (2), 93-100. DOI: 10.1002/bmb.2004.494032020331 (acccessed 2023-08-08T23:29:14).

(11) 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* **2012**, *26* (10), 1241-1248. DOI: 10.1002/rcm.6220 (acccessed 2023-08-08T23:32:16).

(12) 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*, 1-40. DOI: 10.1155/2012/282574 (acccessed 2023-08-08T23:34:21).

(13) Karas, M.; Bahr, U.; Dülcks, T. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. *Fresenius' Journal of Analytical Chemistry* **2000**, *366* (6-7), 669-676. DOI: 10.1007/s002160051561 (acccessed 2023-08-08T23:35:37).

(14) Douglas, D. J.; Frank, A. J.; Mao, D. Linear ion traps in mass spectrometry. *Mass Spectrometry Reviews* **2005**, *24* (1), 1-29. DOI: 10.1002/mas.20004 (acccessed 2023-08-03T05:46:57).

(15) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. An introduction to quadrupole-time-of-flight mass spectrometry. *Journal of Mass Spectrometry* **2001**, *36* (8), 849-865. DOI: 10.1002/jms.207 (acccessed 2023-08-03T05:53:21).

(16) Nikolaev, E. N.; Kostyukevich, Y. I.; Vladimirov, G. N. Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. *Mass Spectrometry Reviews* **2016**, *35* (2), 219-258. DOI: 10.1002/mas.21422 (acccessed 2023-08-03T05:54:57).

(17) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down proteomics: Facts and perspectives. *Biochemical and Biophysical Research Communications* **2014**, *445* (4), 683-693. DOI: 10.1016/j.bbrc.2014.02.041 (acccessed 2023-08-03T00:56:08).

(18) Medzihradszky, K. F.; Chalkley, R. J. Lessons inde novopeptide sequencing by tandem mass spectrometry. *Mass Spectrometry Reviews* **2015**, *34* (1), 43-63. DOI: 10.1002/mas.21406 (acccessed 2023-08-03T22:07:50).

(19) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiaeby electron transfer dissociation (ETD) mass spectrometry. *Proceedings of the National Academy of Sciences* **2007**, *104* (7), 2193-2198. DOI: 10.1073/pnas.0607084104 (acccessed 2023-08-08T23:37:37).

(20) Frese, C. K.; Altelaar, A. F. M.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. *Journal of Proteome Research* **2011**, *10* (5), 2377-2388. DOI: 10.1021/pr1011729 (acccessed 2023-08-08T23:38:55).

(21) Resing, K. A.; Ahn, N. G. Proteomics strategies for protein identification. *FEBS Letters* **2005**, *579* (4), 885-889. DOI: 10.1016/j.febslet.2004.12.001 (acccessed 2023-08-08T23:41:10).

(22) Pandey, A.; Mann, M. Proteomics to study genes and genomes. *Nature* **2000**, *405* (6788), 837-846. DOI: 10.1038/35015709 (acccessed 2023-08-08T23:42:29).

Chapter 2 Protein-protein interaction network mapping by affinity purification cross-linking mass spectrometry (AP-XL-MS) based proteomics

2.1 Abstract:

Protein-protein interactions (PPIs) are the physical interactions formed among proteins. These interactions are primarily functional, i.e., they arise from specific biomolecular events, and each interaction interface serves a specific purpose. A significant number of methods have been developed for protein interactions in the field of proteomics in the last decade. Advanced mass spectrometry technology significantly contributed to the development of these methods. The rapid advancement of groundbreaking MS technology has greatly aided mapping protein interaction from large data sets comprehensively. This chapter describes the Affinity Purification (AP) Mass Spectrometry (MS) based methods combined with chemical cross-linking (CXL) of protein complexes. The chapter includes sample preparation methods involving cell culture, cell treatments with ligands, drugs, and cross-linkers, protein extractions, affinity purification, SDS polyacrylamide gel separation, in solution or in gel digestion, Liquid- chromatography, and mass spectrometry analysis of samples (LC-MS/MS). Application of a cleavable cross-linker, Dual Cleavable Cross-Linking Technology (DUCCT) combining with the Affinity Purification (AP-CXL-MS) method has also been described. Methods for data analysis using unmodified and cross-linked peptide analysis were discussed.

2.2 Introduction

A study of PPIs and protein structures, reveals the protein functions in cells. PPI assays can be broadly categorized as *in vivo* techniques yeast two-hybrid (Y2H)[1] method, *in vitro* methods (Coimmunoprecipitation[2]), and in silico methods (Computational approaches[3]). Integration of protein domains combined with the localization of proteins reveals a plethora of domains existing within subnetworks and their correlation with complexes, thereby deciphering the biological roles of proteins with unknown functionalities. Mass spectrometry is now a major instrument for large-scale protein sequencing. Additionally, tandem mass spectrometry helped utilize a fragment-based marker to explore target proteins and their interactor from large-scale cell signaling experiments. Combining chemical labeling techniques with a traditional biochemical approach is capable of pinpointing protein-protein interactions in the native environment of the cells[4].

The Affinity Purification Mass Spectrometry (AP-MS) approach, a conventional method used for studying protein-protein interactions, utilizes specific binding interactions between the affinity tag or a monoclonal/polyclonal antibody, and macromolecules such as DNA, RNA, lipids, peptides, or proteins (primary target) which act as bait[5]. The inability to distinguish direct interactions from indirect interactions (i.e., between bait and prey proteins) gives rise to high false-positive rates[6,7]. The approach utilizes mild washing conditions to remove non-specific interactions due to the problem of the detachment of bound proteins from the bait. Thus, weak, and transient interactions are potentially not preserved due to the washing steps. Affinity Purification Cross-linking Mass Spectrometry (AP-CXL-MS) overcomes some limitations by covalently linking different proteins with cross-linkers bearing two or more reactive ends separated by the spacer arm[8]. Cross-linkers can be referred to as molecular rulers as they provide the distance between the cross-linked proteins. This linking also helps apply strong denaturing solvent washing steps during sample preparations and preserves weak and transient interactors. Besides, interactors

can be identified by mass spectrometry using non-modified sequences or, more specifically, the specific protein domains by crosslinked peptide analysis.

Different categories of cross-linkers include amine-reactive[9], thiol-reactive[10], carboxyl-reactive[9, 11], MS cleavable[9], and chemical or photocleavable. Cross-linking produces several kinds of peptides after enzymatic digestion, and most are unmodified peptides due to the low reactivity of cross-linkers in an aqueous solvent. Sequencing the inter-crosslinking peptide (two peptides connected with the cross-linkers) is difficult due to the complexity of fragmentation of the two cross-linked peptides during tandem mass spectrometry experiments. Introducing a cleavable group reduces the complexity by cleaving the inter-cross-linked peptides into two parts, and they can be individually sequenced by tandem mass spectrometry. In this chapter, we detailed the identification of interactors by sequencing unmodified peptides. Moreover, we also briefly provided the steps involved in analyzing protein domain using currently available software tools. A novel MS-compatible Dual Cleavable Cross-Linking Technology (DUCCT) has recently been developed by Chowdhury et al[12]. This advanced cross-linking technology effectively contributes to the mapping of PPIs by reducing the complexities encountered during data analysis. DUCCT functions by generating signature ions in the tandem MS/MS by undergoing sequential collision-induced dissociation (CID) and electron transfer dissociation (ETD) cleavage. CID cleavage results from breaking the labile aspartate-proline (DP) bond, while ETD cleavage occurs by breaking an azo bond in DUCCT[13].

The affinity purification MS technique reduces data complexity tremendously. The method also requires the transfection of bait proteins with affinity tags. Due to the poor performance of some antibodies, small molecule affinity tags are popular. Due to the advantage, many reagents came out with affinity tags designed in the structure. Ligands or drugs were also designed with

affinity tags to enrich the target proteins using this approach. Chemical crosslinkers utilized these affinity tags to enrich cross-linked proteins by affinity chromatography. Affinity-tagged ligands or crosslinkers helped reduce data complexities and pinpoint direct or indirect targets in AP-MS methods. Biotin-avidin affinity chromatography is a typical example of affinity tag-based purification techniques. Besides, the length of spacer chain arms in cross-linkers is important to reach a reasonable distance in the protein complexes. Our lab efficiently conducted AP-MS with spacer-chain controlled cross-linking (SPACC)[14]. This AP-SPACC-MS technique connects more

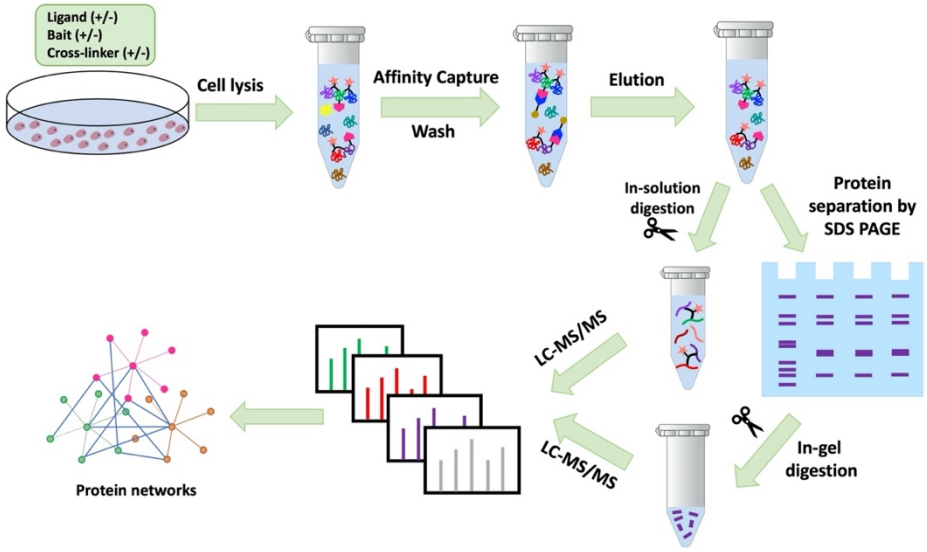


Figure 2-1 Workflow of AP-CXL-MS

proteins with the bait due to the variable spacer arms of the cross-linkers. The chapter also discusses the procedure to conduct the AP-SPACC-MS experiments using DUCCT (18 Å) and a commercial cross-linker called BS3 (11 Å)[13, 15] (See Note 1). A general scheme is provided in Figure 2-1.

Numerous data analysis software like ASAP[16], pLink[3], X-link[17], and Metamorpheus[18] have been used for the analysis of large amounts of data on cross-linked peptides. However, the software's capabilities are limited due to the increase in search space in cross-linking studies. A few software are available to analyze cleavable cross-linked data, e.g., XlinkX[19] and MeroX[20, 21,22]. However, they still suffer from difficulties in analyzing large-scale data. We have developed a user-

friendly software tool Cleave-XL[12] for identifying cross-linked products for DUCCT. This book chapter primarily focuses on identifying unmodified peptides for finding interactions. We will briefly discuss cross-linked peptide analysis using currently availed software tools.

We have developed a user-friendly software tool Cleave-XL for identifying cross-linked products for DUCCT. This book chapter primarily focuses on identifying unmodified peptides for finding interactions. We will briefly discuss cross-linked peptide analysis using currently availed software tools.

2.3 Materials

Cell culture

1. HEK293-MD2-CD14 cells (InvivoGen) were transfected with the HA-tagged human TLR2 gene, InvivoGen, San Diego, CA).
2. Growth medium: Complete cell growth medium Dulbecco's Modified Eagle Medium (DMEM) was prepared by adding 10% fetal bovine serum (FBS), 100U/ml penicillin, 100 µg/ml streptomycin, 50 µg/ml hygromycin, and 10 µg/ml blasticidin to prevent the fungal growth.
3. Radio-immunoprecipitation assay buffer (RIPA) buffer: RIPA lysis and extraction buffer (and protease inhibitor were mixed in 10:1 ratio.
4. Bicinchoninic acid (BCA) Assay: Pierce BCA Protein Assay Kit.

Cell Treatments

1. Lipopolysaccharide (LPS) was purchased from InvivoGen. Prepare the stock solution (5mg/ml) using endotoxin-free water.
2. Simvastatin: Prepare 10 µM simvastatin solution in Dimethyl Sulfoxide (DMSO).

3. 1.5M Tris-HCl buffer (pH 8.0)

Affinity Purification

1. Tris-Buffered Saline (TBS).
2. TBST buffer: 0.05% Tween-20 in TBS buffer.
3. Pierce Monomeric avidin UltraLink Resin. Instead of monomeric avidin, streptavidin can be used.
4. Anti-HA magnetic beads.

SDS Polyacrylamide Gel

1. Resolving gel buffer: 1.5M Tris-HCl (Dissolve 181.7 g Tris-HCl in 1 liter water, and adjust the pH to 8.8, then filter it, and store it at 4 °C). (see Note 2)
2. Stacking gel buffer: 0.5 M Tris-HCl, pH 6.8. (Transfer 60.6 g Tris-HCl in 1 liter water, and adjust the pH with concentrated HCl, then filter it, and store at 4°C.
3. Ammonium persulfate (APS): Prepare 10% APS solution by dissolving 100 mg APS in 1 mL ultra-pure water.
4. 30% Acrylamide/bis-acrylamide solution (28.2:0.8: Store at 4°C.
5. Sodium dodecyl sulfate (SDS): Prepare 10% SDS solution from SDS powder.
6. N, N,N',N'-Tetramethylenediamine (TEMED) was used.
7. Blue pre-stained protein standard with a broad range 11-250 kDa.

Digestion

1. Pierce C18 tips, 100 µL bed.
2. Liquid chromatography-mass spectrometry (LC-MS) grade acetonitrile (ACN), and methanol (MeOH) were obtained (VWR, PA, USA).

3. Ammonium bicarbonate, Iodoacetamide (IAM), chloroform, and formic acid were purchased.
4. Dithiothreitol (DTT) was purchased.
5. Proteolytic enzyme Trypsin is used to digest the proteins.

Sample analysis (Nano ESI-LC-MS/MS)

1. Instrument: LTQ Velos Pro mass spectrometer combined with a Dionex Ultimate 3000 UHPLC system.
2. Acclaim Pepmap C18 column.
3. Solvent system: Mobile Phase A: 0.1% formic acid in water.
4. Mobile Phase B: 0.1% formic acid in 95% acetonitrile, 5% water.

Dual Cleavable Cross-Linking Technology (DUCCT) cross-linker and other cross-linkers

1. Fmoc-Gly-wang resin, Fmoc-Asp(O-2-PhiPr), and Fmoc-Asp-OtBu were purchased from EMD Millipore.
2. Fmoc-Pro-OH, 4-formylbenzoic acid, N-hydroxy succinimide (NHS), N,N-Dicyclohexylcarbodimide (DCC), N,N,N′,N′-Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate, O-(Benzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate (HBTU) were obtained from Sigma Aldrich.
3. Dichloromethane (DCM), Dimethylformamide (DMF), Trifluoroacetic acid (TFA) were acquired from VWR.
4. Solvent systems for Peptide synthesizer PS3, 0.4 M N-methylmorpholine in DMF, 20% piperidine in DMF were purchased from Gyros Protein technologies.
5. 6-Fmoc-hydrazinonicotinic acid was obtained from Advanced automated peptide protein technology.

2.4 Methods

Cell culture

Human embryonic kidney (HEK) 293 cells transfected with affinity-tagged Hemagglutinin (HA) immune receptors/or respective cells with affinity-tagged target or antibody-based affinity purification can be utilized.

HA-tagged HEK293 cell culture

1. For warming the medium, keep the growth medium in a water bath at 37 °C for 15 minutes.
2. Take the cryovial from liquid nitrogen and immediately thaw it at 37° C for 1 minute.
3. The cells were mixed in 5 ml of culture medium into a 15 ml falcon tube.
4. Remove the culture medium after centrifugation at 3000 x g for 5 minutes.
5. Collect the cell pellet and gently resuspend it in 1ml of growth medium.
6. Add 10 mL of pre-made cell growth medium into each 10 cm Petri dish, add the suspended cells drop-wise and incubate at 37° C with 5% CO_2.
7. To perform the subculture, remove the medium, and wash with 1X PBS twice.
8. Trypsinize the cells with Trypsin-EDTA and transfer them into a 1.5 mL Eppendorf tube.
9. Collect the cell pellet by centrifugation at 3000 x g for 3 minutes.
10. Resuspend the cell pellet in 1ml of growth medium.
11. Add 10 mL of growth medium into each petri dish, followed by the addition of suspended cells added based on the experiment, and incubate at 37 °C with 5% CO_2.

Cell treatments

Simvastatin treatment of HA-tagged TLR2-MD2-CD14 cells

1. Allow the cells to attain 70% confluency.

2. Treat the cells with simvastatin to a final concentration of 10 μM and incubate them for 24 hrs followed by stimulation with Lipopolysaccharide (LPS)-biotin.

Lipopolysaccharide (LPS)-biotin treatment of HA-tagged TLR2-MD2-CD14 cells

1. Allow the growth medium to stand in the 37 °C water bath for 15 minutes.
2. Prepare the medium with 1 μg/mL of LPS-biotin (tlrl-3pelps, InvivoGen) and incubate at 37 °C for 4 to 6 hrs.
3. Wash the cells with PBS twice, then scrape the cells.
4. Centrifuge at 3000 x g for 3 min and discard the supernatant.
5. Collect the cell pellet for further use.

Cross-linker

Cross-linker can either be cleavable or non-cleavable. Dual Cleavable Cross-linking Technology (DUCCT~ 18 Å) and commercially available bissulfosuccinimidyl (BS3~ 11 Å) cross-linkers were utilized for this study. Different lengths of spacer arms will provide more coverage for bait to interactors.

1. Prepare 100 mM stock solution of cross-linker in DMSO and treat the cells with cross-linker (DUCCT/BS3) and allow the reaction to proceed for 30 mins.
2. Quench the reaction with 50 mM Tris-HCl buffer (pH 8.0).

Cell Lysis

1. Prepare RIPA Lysis Buffer cocktail by mixing RIPA buffer with protease inhibitor.
2. Wash the cells twice with PBS and transfer them to a falcon tube. Centrifuge at 3000 rpm for 3 minutes. Discard the supernatant. During the second wash, the cells should be transferred to a 1.5 ml Eppendorf tube.
3. Add RIPA cocktail and allow proper dispersion into the cell pellet.

4. Pipette and vortex the cells and incubate on ice for 15 minutes and sonicate for 15 mins.
5. Vortex thoroughly and incubate on ice for 15 minutes.
6. After incubation, centrifuge the cells at 4 °C for 30 minutes at 20,000 x g and collect the supernatant for further analysis.
7. Measure the protein concentration using the BCA method.

Affinity purification

Biotin avidin affinity purification

1. Allow the resin and reagents to attain room temperature.
2. Centrifuge at 500 x g for one minute to remove the storage solution.
3. Wash the monomeric avidin resin three times with 1X phosphate buffer saline (PBS, pH 7.4).
4. Add biotinylated sample to the resin and incubate it at 4°C overnight. Centrifuge for 1 min at 500 x g and discard the supernatant.
5. After that, extensively wash the resin with PBS and ultrapure water to get rid of non-specific interactions.
6. Elute the bound proteins in Laemmli buffer. Centrifuge for 1 min at 500 x g and collect the supernatant.
7. Heat the sample at 95°C for 5 min and load it onto SDS-PAGE gel (See Note 3).

Fusion-tag purification

Any fusion tag can be employed for this study such as Myc, and HA. In this work, we have used HA-tagged proteins.

1. HA-tagged proteins can be purified using anti-HA magnetic beads.

2. Wash the anti-HA magnetic beads with 0.05% Tris Buffered Saline (TBS) and vortex gently.
3. Collect the beads against the side of the tube by placing it into a magnetic stand. Discard the supernatant. Repeat this step twice.
4. Add the HA-tagged protein sample to the beads and allow the reaction to proceed overnight at 4°C with gentle mixing.
5. Place the tube into a magnetic stand and discard the unbound sample.
6. Wash the beads with TBS and ultrapure water twice with gentle mixing.
7. Place the tube into the magnetic stand and discard the supernatant.
8. Elute the beads with Laemmli buffer, heat at 95°C for 5 min followed by centrifugation, and load onto SDS-PAGE (See Note 4).

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE):

1. Prepare separation gel by mixing 3.4 mL of water, 2.5 mL of 1.5M tris buffer, 4 mL 30% Acrylamide, 100 uL 10% SDS, 10 uL TEMED, 50uL 10% ammonium persulfate (APS) in a 50 mL falcon tube. Pour the contents within a 7.25 cm × 10 cm × 1.5 mm gel cassette and overlay a small amount of water on the top of the resolving gel carefully to keep the polymerized gel from drying out.
2. Prepare a stacking gel by mixing 2.65 mL of water, 1.25 mL of 0.5M tris, 1.0 mL of 30% Acrylamide, 50 µL 10% SDS, 5 µL TEMED, 25 µL 10% ammonium persulfate (APS) in a 50 mL falcon tube. Finally, transfer the stacking gel on top of the resolving gel until it overflows and inserts a 10-well gel comb on top of the stacking gel. Allow the gel to polymerize for about 40 minutes.
3. Carefully remove the comb from the gel.

4. Place the glass plate sandwich into the electrophoresis core and keep the assembly in the running tank containing the running buffer.
5. Mix the protein sample with Laemmli buffer and heat at 95°C for 10-20 mins. Load 20 µL of sample into each well using a 200 µL special gel-loading tip.
6. Run the gel with 180 V voltage until the dye reaches the bottom of the gel.
7. Disassemble the running unit and wash the gel with the ultrapure water and use SYPRO ruby protein Gel stain.

Digestion

In-solution Digestion

1. Purified proteins were dissolved in 50 mM ammonium bicarbonate in an Eppendorf tube.
2. Add 5 µL of 5 mM Dithiothreitol (DTT), vortex, and incubate at 56 °C for 45 minutes to reduce disulfide bonds.
3. Add 5 µL, 10 mM Iodoacetamide (IAA) and incubate at 37°C for 30 minutes in the dark (covered in foil) to alkylate cysteines.
4. Bring the volume of the solution up to 150 µL with 50 mM ammonium bicarbonate.
5. Add trypsin solution so that Trypsin: Protein=1:50 w/w and incubate at 37 °C for 12-16 hours in a thermomixer.
6. Allow the sample to cool to room temperature and quench the digestion by acidifying using 1.5 µL of 10% Formic acid.
7. Desalt the sample with C18 columns.
8. Dry the sample in a speed vacuum.
9. Reconstitute the peptides in 0.1% formic acid.

10. Centrifuge the sample at 20,000 X g for 30 minutes before transferring the supernatant to the vial for mass spectrometry analysis.

In-gel Digestion

1. Excise the protein band of interest with a clean scalpel and cut it into smaller pieces.
2. Wash the pieces with double distilled water (DDW) for 10 minutes.
3. Add 500 µL acetonitrile to shrink the gel pieces and vortex well for 15 minutes.
4. Discard the supernatant and dry the gel pieces for 10 minutes.
5. Add 50 µL of 10 mM DTT to the gel pieces. Incubate for 30 minutes at 56°C in thermomixer.
6. Spin down the gel pieces for 2 minutes, allow the tube to stand at room temperature, and carefully remove the supernatant.
7. Add 50 µL of 55 mM IAA and incubate for 30 minutes at room temperature in the dark.
8. Centrifuge the contents for 2 minutes, pull off the supernatant, and wash twice with DDW.
9. Add acetonitrile to shrink the gel pieces followed by removal of acetonitrile by speed vac until dry.
10. Add enough trypsin to cover the dry gel pieces then keep it at 37 °C for 30 minutes.
11. Add 25 µL of 25 mM ammonium bicarbonate and incubate overnight at 37 °C.
12. Centrifuge for 2 minutes and collect the supernatant.
13. To further extract peptides, add 10 µL of 1% trifluoroacetic acid or 1% formic acid solution to gel pieces and incubate for 10 minutes, and add 25 µl of 25 mM ammonium bicarbonate.
14. Spin down for 2 minutes and collect the supernatant.
15. Finally, the sample is ready for LCMS analysis.

Sample analysis (Nano ESI-LC-MS/MS)

LC-MS/MS procedures were detailed using a Thermo velos pro mass spectrometer (See Note 5).

1. For sample analysis, utilize a Dionex Ultimate 3000 nano-UHPLC system coupled with a nano-ESI-linear ion trap (LTQ) Thermo Velos Pro mass spectrometer.
2. For chromatographic separation, a nano viper analytical C18 column (Acclaim Pepmap 150 mm x 75 µm, 3 µm, 100 Å)
3. Employ 90 min binary gradient system to separate peptides. Mobile phase A: 0.1% formic acid in the water, Mobile phase B: 0.1% formic acid in 95% acetonitrile, 5% water.
4. Set the flow rate of nanocolumn to 300 nl/min and injection volume to 5 uL.
5. Perform positive ion mode acquisition with a fixed source voltage 2.20kV, and capillary temperature of 275 °C.
6. Perform full MS scans in normal scan mode from 300 to 2000 m/z mass range.
7. Collect MS/MS spectra in Data Dependent Acquisition (DDA) mode to acquire the MS^2 spectra for the 10 most abundant precursor ions. For data acquisition and processing, Xcalibur software can be utilized.
8. In CID activation, set the collisional energy to 35.0%, activation Q of 0.25, and activation time of 30 ms, along with an isolation width of 1.5 Da.

Data analysis

1. Search the acquired MS raw files through Proteome Discoverer software (PD, version 2.1 ThermoFisher scientific).
2. Download the human protein sequence database from UniProt for the identification of proteins. For all the normal peptides, the search engine node used is Sequest HT.

3. Create processing and a consensus workflow in the workflow editor and set the parameters such as precursor mass tolerance to 2.5 Da and fragment mass tolerance to 0.6 Da. The proteolytic enzyme used for the study was trypsin with at least two missed cleavages. The precursor mass range was set from 350-3500 Da.
4. Set a static modification parameter, for example, carbamidomethyl/+57.02146 Da for cysteine and methionine oxidation/+15.00492 Da as a dynamic modification.
5. Decoy searches should be applied to the data sets to deliver false positive estimations. Set the false discovery rates (FDRs) for peptide and protein identifications to 1% with at least two peptides per protein.
6. Filter the results to detect the PSMs with the best score using *XCorr* score ≥ 1.5, $\Delta Cn <$ 0.01, and q-values of ≤ 0.05 are used for high confident matches (see Note 6).

2.5 Notes

1. BS3 is available from Thermo Fisher Scientific (Cat # 21580). Cross-linkers with specific reactivities, affinity tags (biotin or clickable), and different spacer arms are available from different vendors.
2. Milli-Q ultrapure water was used for all the experiments.
3. Detailed description of the transfection procedure has been explained in this literature. For in solution digestion, after washing on bead digestion can be performed. In solution digestion protocol is provided above.
4. Mass spectrometry analysis can be performed with other mass spectrometers capable of sequencing large-scale samples.
5. For analyzing cross-linked peptides, a few software are available. The software called XlinkX[17] and MeroX[20,22] can be used in AP-MS for cleavable cross-linkers. For non-

cleavable cross-linkers, pLink[3] software is available online. Limited data analysis can be done with Cleave-XL[12], a software for DUCCT.

References

(1) Ijaz, M.; Ansari, M.-u.-R.; Iqbal, M. Introductory Chapter: Protein-Protein Interactions and Assays. *Protein-Protein Interaction Assays* **2018**. DOI: 10.5772/INTECHOPEN.77337.

(2) MacBeath, G.; Schreiber, S. L. Printing Proteins as Microarrays for High-Throughput Function Determination. *Science* **2000**, *289* (5485), 1760-1763. DOI: 10.1126/SCIENCE.289.5485.1760.

(3) Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M. A. R.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P. I. W.; Daly, M. J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **2007**, *81* (3), 559-575. DOI: 10.1086/519795.

(4) Ewing, R. M.; Chu, P.; Elisma, F.; Li, H.; Taylor, P.; Climie, S.; Mcbroom-Cerajewski, L.; Robinson, M. D.; O'Connor, L.; Li, M.; et al. Large-scale mapping of human protein–protein interactions by mass spectrometry. *Molecular Systems Biology* **2007**, *3* (1), 89. DOI: 10.1038/msb4100134 (acccessed 2022-08-19T05:51:33).

(5) Zacharias, A. O.; Fang, Z.; Rahman, A.; Talukder, A.; Cornelius, S.; Chowdhury, S. M. Affinity and chemical enrichment strategies for mapping low-abundance protein modifications and protein-interaction networks. *Journal of Separation Science* **2021**, *44* (1), 310-322. DOI: 10.1002/JSSC.202000930.

(6) Fang, Z.; Baghdady, Y. Z.; Schug, K. A.; Chowdhury, S. M. Evaluation of Different Stationary Phases in the Separation of Inter-Cross-Linked Peptides. *Journal of proteome research* **2019**, *18* (4), 1916-1925. DOI: 10.1021/ACS.JPROTEOME.9B00114.

(7) Huttlin, E. L.; Ting, L.; Bruckner, R. J.; Gebreab, F.; Gygi, M. P.; Szpyt, J.; Tam, S.; Zarraga, G.; Colby, G.; Baltier, K.; et al. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **2015**, *162* (2), 425-440. DOI: 10.1016/j.cell.2015.06.043 (acccessed 2022-08-19T05:48:50).

(8) Steigenberger, B.; Albanese, P.; Heck, A. J. R.; Scheltema, R. A. To Cleave or Not To Cleave in XL-MS? *Journal of the American Society for Mass Spectrometry* **2020**, *31* (2), 196-206. DOI: 10.1021/jasms.9b00085 (acccessed 2022-08-19T05:54:02).

(9) Piersimoni, L.; Sinz, A. Cross-linking/mass spectrometry at the crossroads. *Analytical and Bioanalytical Chemistry* **2020**, *412* (24), 5981-5987. DOI: 10.1007/S00216-020-02700-X/FIGURES/4.

(10) Iacobucci, C.; Piotrowski, C.; Rehkamp, A.; Ihling, C. H.; Sinz, A. The First MS-Cleavable, Photo-Thiol-Reactive Cross-Linker for Protein Structural Studies. *J. Am. Soc. Mass Spectrom* **2018**, *30*, 139-148. DOI: 10.1007/s13361-018-1952-8.

(11) Sriswasdi, S. ScholarlyCommons ScholarlyCommons Publicly Accessible Penn Dissertations 2013 In-Depth Analysis of Zero-Length Crosslinking for Structural Mass In-Depth Analysis of Zero-Length Crosslinking for Structural Mass Spectrometry Spectrometry.

(12) Chakrabarty, J. K.; Sadananda, S. C.; Bhat, A.; Naik, A. J.; Ostwal, D. V.; Chowdhury, S. M. High confidence identification of cross-linked peptides by an enrichment-based dual cleavable cross-linking technology and data analysis tool cleave-XL. *Journal of the American Society for Mass Spectrometry* **2020**, *31* (2), 173-182. DOI: 10.1021/JASMS.9B00111.

(13) Chakrabarty, J. K.; Naik, A. G.; Fessler, M. B.; Munske, G. R.; Chowdhury, S. M. Differential Tandem Mass Spectrometry-Based Cross-Linker: A New Approach for High Confidence in

Identifying Protein Cross-Linking. *Analytical Chemistry* **2016**, *88* (20), 10215-10222. DOI: 10.1021/ACS.ANALCHEM.6B02886/SUPPL_FILE/AC6B02886_SI_003.XLSX.

(14) Kamal, A. H. M.; Aloor, J. J.; Fessler, M. B.; Chowdhury, S. M. Cross-linking Proteomics Indicates Effects of Simvastatin on the TLR2 Interactome and Reveals ACTR1A as a Novel Regulator of the TLR2 Signal Cascade. *Molecular and Cellular Proteomics* **2019**, *18* (9), 1732-1744. DOI: 10.1074/mcp.RA119.001377.

(15) Chowdhury, S. M.; Du, X.; Tolić, 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. *Analytical chemistry* **2009**, *81* (13), 5524-5532. DOI: 10.1021/AC900853K.

(16) Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrometry Reviews* **2006**, *25* (4), 663-682. DOI: 10.1002/mas.20082 (acccessed 2022-08-19T06:07:23).

(17) Zheng, C.; Weisbrod, C. R.; Chavez, J. D.; Eng, J. K.; Sharma, V.; Wu, X.; Bruce, J. E. XLink-DB: Database and Software Tools for Storing and Visualizing Protein Interaction Topology Data. *Journal of Proteome Research* **2013**, *12* (4), 1989-1995. DOI: 10.1021/pr301162j (acccessed 2022-08-19T06:11:27).

(18) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97* (11), 5802-5806. DOI: 10.1073/PNAS.090099097.

(19) Orsburn, B. C. Proteome Discoverer—A Community Enhanced Data Processing Suite for Protein Informatics. *Proteomes* **2021**, *9* (1), 15. DOI: 10.3390/proteomes9010015 (acccessed 2022-08-19T06:15:08).

(20) Iacobucci, C.; Götze, M.; Ihling, C. H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions. *Nature Protocols 2018 13:12* **2018**, *13* (12), 2864-2889. DOI: 10.1038/s41596-018-0068-8.

(21) Arlt, C.; Götze, M.; Ihling, C. H.; Hage, C.; Schäfer, M.; Sinz, A. Integrated Workflow for Structural Proteomics Studies Based on Cross-Linking/Mass Spectrometry with an MS/MS Cleavable Cross-Linker. *Analytical chemistry* **2016**, *88* (16), 7930-7937. DOI: 10.1021/ACS.ANALCHEM.5B04853.

(22) Götze, M.; Pettelkau, J.; Fritzsche, R.; Ihling, C. H.; Schäfer, M.; Sinz, A. Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. *Journal of the American Society for Mass Spectrometry* **2015**, *26* (1), 83-97. DOI: 10.1007/S13361-014-1001-1.

Chapter 3 Identification of acidic residues in proteins by selective chemical labeling and tandem mass spectrometry

3.1 Abstract

Glutamate and aspartate are the most prevalent excitatory neurotransmitters. Several neurological and psychiatric conditions, such as epilepsy, Alzheimer's disease, and schizophrenia have been linked to abnormal glutamate signaling. Aspartate is the fifth most common residue identified in protein-protein interfaces. Labeling studies will help identify these reactive residues that will be imperative in the development of various prospective therapeutic drugs that target the glutamate (Glu) and aspartate (Asp) systems. While methods such as the activation of carboxylic acids using carbodiimides have been widely used as labeling reagents, however, due to their lack of specificity and formation of numerous side products, they are not considered the best choice. We developed a novel reagent that can specifically label Asp and Glu and its reactivity is examined in several model peptides and proteins. We have demonstrated a proof-of-concept in this study using bradykinin, neurotensin, angiotensin I, and Bovine serum albumin (BSA) protein. We have demonstrated the selective labeling of Asp and Glu amino acids at pH 4 using diphenyldiazomethane (DPDAM). The labeling of the carboxylate group of C-terminal was achieved at lower pH. This method offers a convenient route for the conversion of acids to esters and yields molecular nitrogen as the byproduct. We have employed the tandem mass spectrometry technique to validate the fragmentation characteristics of labeled samples. We believe that this study can be utilized for specifically labeling acidic residues in a protein or a mixture of proteins which will provide invaluable insights into structural biology.

3.2 Introduction

The potential uses for protein bioconjugates include biomedicine, diagnostics, and chemical biology. Proteins are macromolecules with a variety of functional groups, especially nucleophilic groups[1]. Protein conjugation is imperative in building multifunctional constructs in which a desired molecule is coupled with a protein's activity. Cysteine SH and lysine NH groups are by far the two amino acids (AAs) that are targeted the most for the identification and characterization of proteins and their functions[2]. Conjugation to these residues has been well-researched, and it is now common practice in many laboratories. Although more specialized reagents have also been produced, conventional reagents like maleimide and NHS-esters are frequently used. Comprehensive research has also been done on selective conjugation to tyrosine, a few reagents such as 4-phenyl-3*H*-1,2,4-triazoline-3,5(4*H*)-diones (PTAD)[3] -azides and alkynes are also used. Residues with carboxyl-terminating side chains (such as aspartic and glutamic acids) are desirable targets, due to their significant abundance in most proteins. Asp and Glu make up 5.5% and 6.7% of all residues in the most recent edition of the SwissProt database, 2014_04, respectively[4].

Both the sidechains of Aspartic acid and Glutamic acid have a pKa value of about 4. Therefore, at normal pH, they have a negative charge. With regard to solubility and activity, these AAs play a variety of roles in the protein. Because of the restricted carboxylate reactivity in water, Asp- and Glu-selective reactions are a difficult possibility. Also, a carboxylate group is also present at the protein's C-terminus. Except for acylglycerols and amino-acyl tRNAs, esters are hardly present in human cells. This reactivity forms the basis for the clinical efficacy of the numerous prodrugs. Serine and threonine residues can naturally undergo post-translational modifications called acetylation. However, chemical synthesis cannot be used to efficiently produce esters from protein carboxyl[5] or hydroxyl groups. Protein carboxyl groups can be activated using

carbodiimides and other reagents, but the water-soluble solvent and the protein's amino, hydroxyl, and sulfhydryl groups efficiently compete with exogenous alcohols for the resulting activated acyl groups. Since diazo compounds are renowned as flexible and practical substrates for a variety of chemical reactions, diazo chemistry is still very much alive with its long-standing fascination[6]. The usage of diazo compounds is very common in synthetic organic chemistry.

During the most recent worldwide pandemic caused by the SARS-CoV-2 coronavirus, it was found that Omicron (B.1.1.529) and its subvariants have taken over as the predominant circulating variants globally. It has been discovered that the D614G mutation is prevalent in the majority of Omicron subvariants. D614G mutation causes the surface spike protein (S protein) to modify its conformation, which augments its binding affinity and interaction with ACE2. SARS-CoV-2 becomes more contagious as a result. The virus enters the host cell through its interaction with the angiotensin-converting enzyme 2 (ACE2) receptor and its S protein[7]. Given that the S protein of SARS-CoV-2 is directly involved in viral pathogenicity and infectivity, any mutation in this protein is relevant[8]. This particular mutation leads to increased transmission efficiency of the virus[9].

We have developed a novel reagent that can specifically modify side chains of aspartic and glutamic acid without labeling the carboxylate group of the C-terminal of the proteins at a particular pH. DPDAM converts carboxylic acids into methyl esters with molecular nitrogen as the by-product. The acidic group of the side chain residues is acidic enough to promote the reaction leading to the formation of diazonium-carboxylate salt. The workflow for chemical labeling of peptides containing acidic residues is shown in Figure 3-1. We investigated the reactions of DPDAM with peptides at different peptide-to-reagent ratios. At peptide level, the labeling was

achieved at 1:20 (peptide: DPDAM) ratio while at the protein level, 200 equivalents of DPDAM were required to achieve the esterification[10].

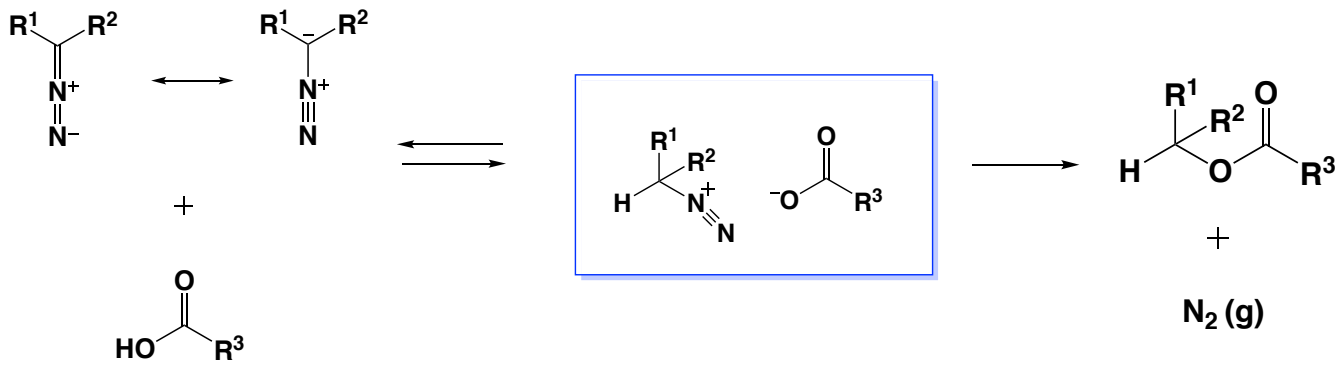


Figure 3-1 Mechanism of O-alkylation of carboxylic acid using diazo compounds

3.3 Experimental section

3.3.1 Materials

Neurotensin (pyr-LYENKPRRPYIL) was obtained from Anaspec, Inc. (San Jose, CA). Angiotensin I (DRVYIHPFHL) and Bovine serum albumin (BSA) were acquired from Sigma-Aldrich (St Louis, MO), and DPDAM was synthesized by Dr. Joseph Buonomo, Assistant professor, University of Texas at Arlington. Acetonitrile, ammonium bicarbonate, and formic acid were purchased from Sigma-Aldrich (St Louis, MO) and C18 desalting tips (100 µL bed) were purchased from Thermo Scientific (IL). For protein digestion, a reducing agent dithiothreitol (Biorad CA), an alkylating agent iodoacetamide (Sigma-Aldrich, MO), and a sequencing-grade modified trypsin protease (Promega, Madison, WI) were purchased. Liquid chromatography-mass spectrometry (LC-MS) grade methanol and acetonitrile were obtained from Fischer Scientific (Pittsburgh, PA). For all the studies, eighteen Milli-Q water was utilized and the water filtration system was obtained from Aries Filterworks (West Berlin, NJ).

3.3.2 In-solution digestion

The disulfide bonds in proteins were reduced using dithiothreitol, 10mM (Biorad, CA), followed by alkylation with 50 mM iodoacetamide (Sigma-Aldrich, MO). The sample was then subjected to trypsin protease (Promega, Madison, WI) for digestion. The ratio of protein-trypsin used was 50:1, and the sample was allowed to incubate overnight at 37°C. 0.1% formic acid was used to quench the digestion process. The sample was dried under vacuum, desalted, and reconstituted in 0.1% formic acid solution in H_2O. The sample was then analyzed using LC-MS/MS.

3.3.3 Instrumentation

Sample analysis was performed on an LTQ Velos Pro mass spectrometer which is coupled with UPLC (Ultimate 3000, Dionex, USA). Labeled peptides were then separated by reversed-phase chromatography, using a nanoviper analytical C18 column (Acclaim Pep Map 100 C18 LC Columns, Thermo Scientific). A binary gradient system was utilized in which the composition of organic and aqueous mobile phases is 95% acetonitrile and 98% water, respectively. The flow rate of nanocolumn was 0.3 µL/min and the injection volume was 5 µL respectively.

The nanoelectrospray ionization (ESI) source was utilized for ionization with spray voltage of 2.0 V and heated capillary temperature of 275°C. LC-MS spectra were acquired from 300 to 2000 *m/z*. Top five most abundant precursor ions were chosen for further MS/MS with the activation energy at 45% with an isolation width of 1.5 Da, activation Q of 0.25, and activation time of 10 ms. The dynamic exclusion time was set for 30 ms. The data was acquired for 90 minutes. For data processing, Xcalibur software was used.

The sample was injected into QExactive HF mass spectrometer coupled to an Ultimate 3000 RSLC-Nano liquid chromatography system. Samples were introduced to a 15-cm long EasySpray

column (Thermo Fisher) with 75 um i.d. and were eluted with a gradient from 0-28% buffer B over a span of 90 min with a flow rate set at 250 nL/min. The composition of buffer A was 2% (v/v) acetonitrile (ACN) and 0.1% formic acid, and buffer B was composed of 80% (v/v) ACN, 10% (v/v) trifluoroethanol, and 0.1% formic acid. The mass spectrometer was operated in positive ion mode while maintaining a source voltage of 2.5 kV and capillary temperature was set to 275°C. MS scans were performed at 120,000 resolution using the Orbitrap mass analyzer. For each full spectrum, up to 20 MS/MS spectra were generated using higher-energy collisional dissociation (HCD). The dynamic exclusion time was set for 20 s.

3.3.4 Chemical modification of Neurotensin, Angiotensin I, and Bradykinin with DPDAM

The standard peptides Neurotensin and angiotensin were treated with DPDAM. Neurotensin sequence contains one acidic residue and modified N-terminus group. Angiotensin I also has one acidic residue. Neurotensin (1mM) was dissolved in MES-NaOH buffer, pH 4. DPDAM (100mM) was dissolved in acetonitrile. The labeling reaction of these peptides with the tag was carried out in 1:20 ratio in MES-NaOH buffer. The reaction was set for 2 hours at 37°C and was subsequently quenched with 10% formic acid. The excess reagent was further removed. The samples were purified using desalting process by utilizing C18 ZipTip, dried by speed vacuum, and reconstituted in 0.1% formic acid solution in H_2O. The labeled samples were then analyzed using a Linear Ion trap (LTQ) Velos Pro mass spectrometer.

3.3.5 Chemical modification of BSA protein with DPDAM

Bovine serum albumin (1mM, 1µL) was treated with DPDAM (100mM) in a 1:200 ratio in 2-Morpholinoethanesulfonic acid monohydrate (MES)-NaOH buffer. The reaction was allowed to proceed for 2 hours at 37°C. The samples were then dried under vacuum and reconstituted in 50 mM ammonium bicarbonate solution in H_2O. The samples were then subjected to in-solution

digestion with trypsin protease overnight. After desalting, the samples were dried by speed vacuum and were reconstituted in 0.1% formic acid solution in H_2O. The samples were further analyzed with QExactive HF mass spectrometer.

3.4 Results and Discussion

3.4.1 Labeling of peptides by DPDAM

Labeling studies can be used to study the dynamics of post-translational modifications in peptides and proteins. A very few studies have used carboxylate-specific chemicals to map protein surfaces due to the challenge of selectively altering carboxylates under native circumstances. Aspartic and glutamic acid residues are abundant on the protein surfaces, and because of their hydrophilicity, these amino acids are less explored. In recent years, protein structure and interactions have been intensively studied using selective covalent labeling and mass spectrometry. This strategy may be enhanced further, which could provide even more knowledge on protein structure in solution. For these studies, it is imperative to develop new labels that are specific to each amino acid. Even though various chemicals are available and have been successfully employed, electrophilic reagents that target nucleophilic functional groups dominate the chemistry utilized to change proteins. Alternative processes that target other amino acid side chains such as aspartate (D) and glutamate (E) with comparable levels of selectivity are required. The reaction conditions for the labels now in use (low pH for EDC, for example) make it difficult to probe protein structure under physiological circumstances. Aspartate and glutamate have been linked to a variety of biological modifications, including acetylation, phosphorylation, and carbonylation, which can lead to neurological and cardiovascular disorders. This has generated significant interest in developing reagents that selectively label acidic residues. We have developed a reagent, DPDAM, that

specifically targets D and E amino acids. These labeled peptides undergo Electrospray ionization (ESI)-MS. We have been successfully able to label three model peptides that include N-terminal blocked peptide neurotensin (pyr-LYENKPRRPYIL, monoisotopic unmodified mass 1671.9097 Da), Angiotensin I (DRVYIHPFHL, monoisotopic mass 1296.6850 Da), and bradykinin (RPPGFSPFR, monoisotopic mass 1059.5614 Da). Neurotensin and Angiotensin I contain one E and D amino acid, respectively. The mechanism of their reaction with DPDAM involves the abstraction of a proton from a carboxylic acid followed by a nucleophilic attack on carbon-bearing diazo group. This leads to the evolution of nitrogen gas, which is a driving force for this reaction. By employing trypsin digestion and mass spectrometry, we were able to identify these esterified residues. Neurotensin and Angiotensin I was labeled with DPDAM at pH 4. In the case of neurotensin, we observed the +2 and +3 charged state of the labeled peptide at *m/z* 920.55 and 620.03, respectively in full MS spectrum. The CID spectrum of $[M+DPDAM+3H]^{3+}$ peptide precursor produced *b* and *y*-ions which assisted in the identification of the peptide sequence shown in Figure 3-2.

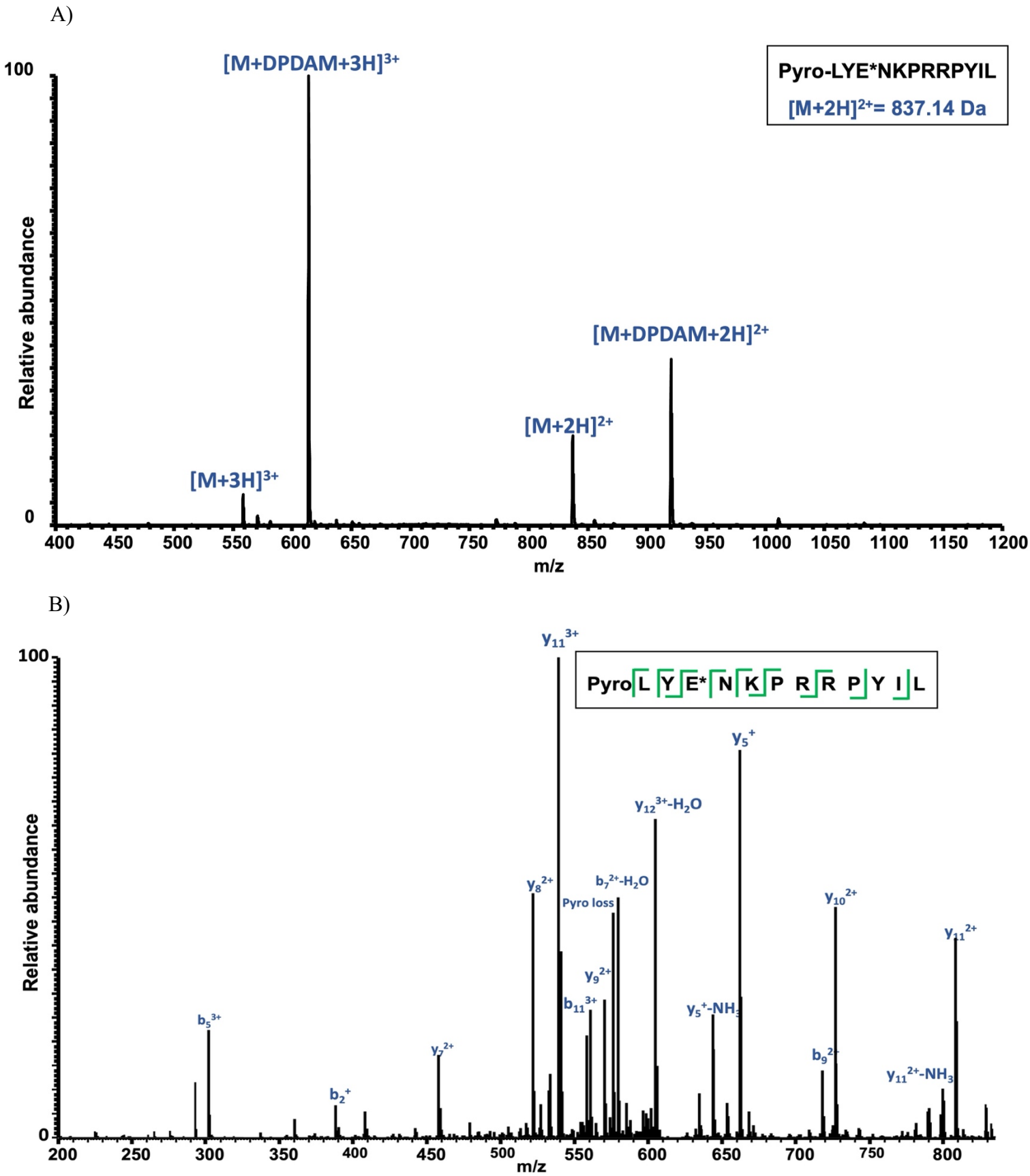
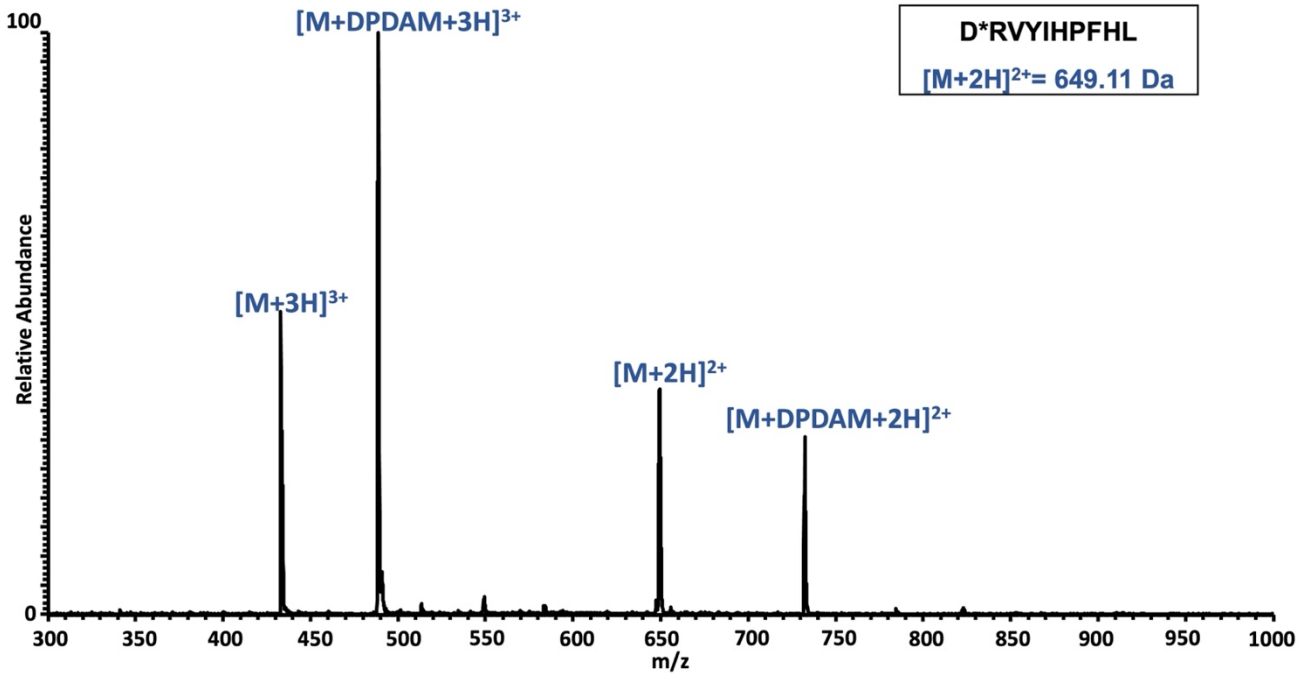


Figure 3-2 ESI-MS and MS/MS spectra of modified peptides with DPDAM at pH 4. (A) Neurotensin labeled with DPDAM at *m/z* 920.55 $[M+2H+DPDAM]^{2+}$ and 614.03 $[M+3H+DPDAM]^{3+}$ (B) MS/MS spectrum of modified neurotensin at *m/z* 614.03 $[M+3H+DPDAM]^{3+}$. *denotes DPDAM-modified sites

A)



B)

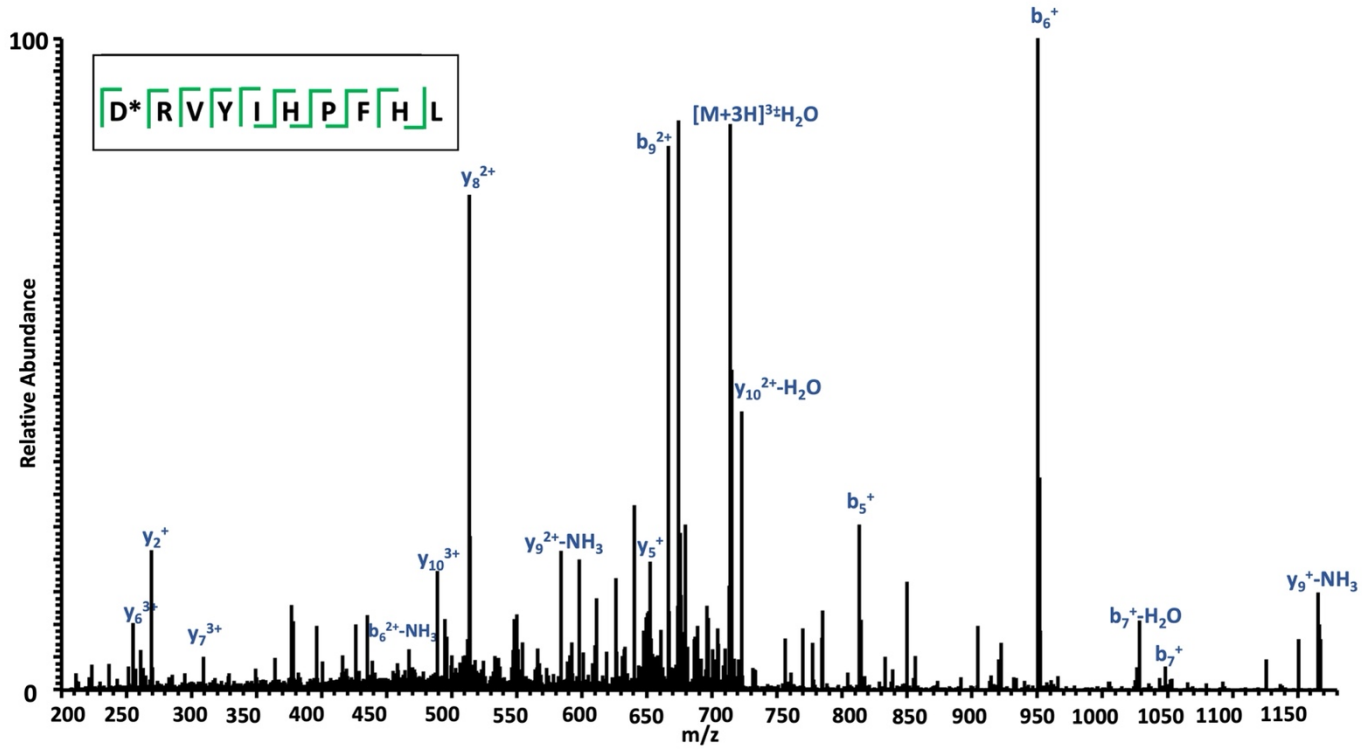


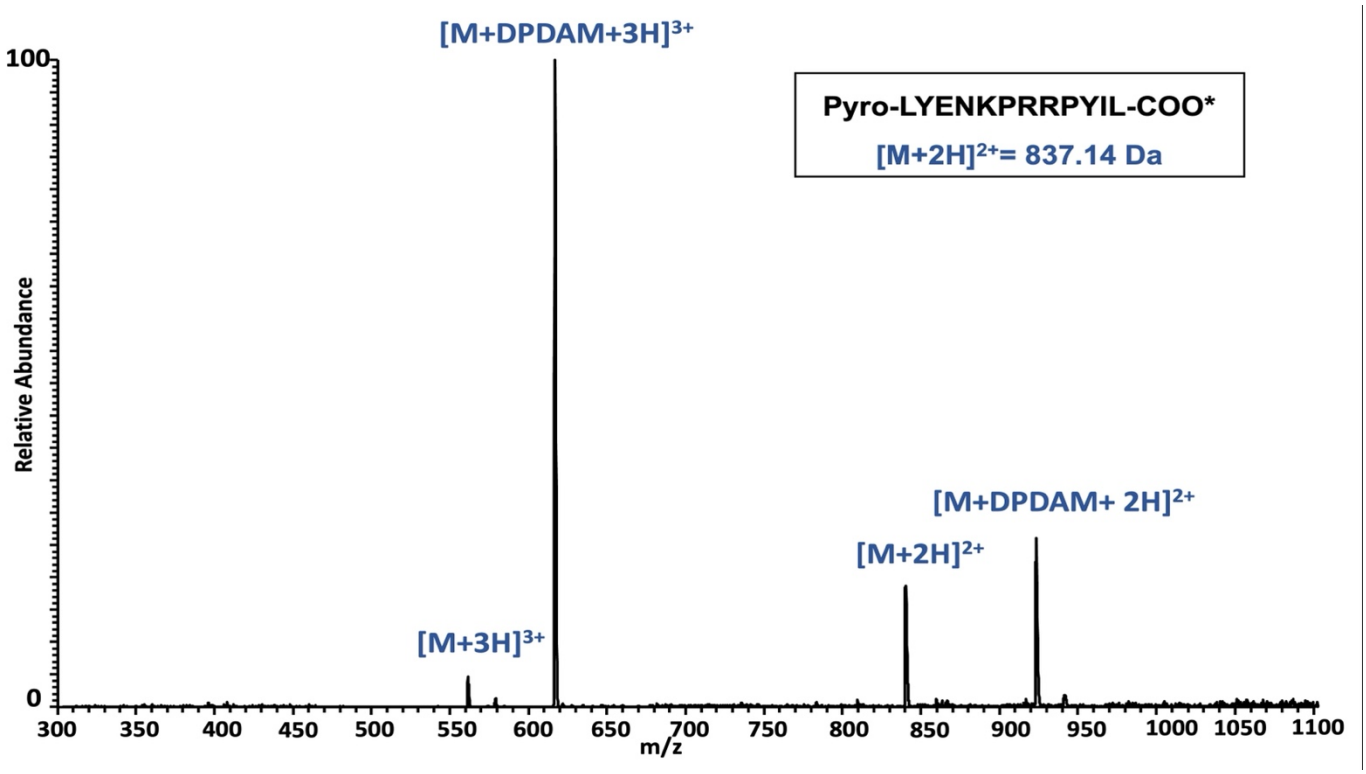
Figure 3-3 ESI-MS and MS/MS spectra of modified peptides with DPDAM at pH 4. (A) Angiotensin I labeled with DPDAM at *m/z* 732.37 $[M+2H+DPDAM]^{2+}$ and 488.60 $[M+3H+DPDAM]^{3+}$. (B) CID-MS/MS spectrum of modified angiotensin I at *m/z* 732.35 $[M+2H+DPDAM]^{2+}$. *denotes DPDAM modified sites

Similarly, the DPDAM was allowed to react with angiotensin I, and labeled peptides were identified at *m/z*'s of 732.37 ($[M+DPDAM+2H]^{2+}$) and 488.60 ($[M+DPDAM+3H]^{3+}$), respectively. CID-MS/MS spectrum of $[M+2H]^{2+}$ facilitated the identification of the peptide sequence shown in Figure 3-3.

Interestingly, it was found that the DPDAM only labeled D and E amino acids at pH 4. However, when the reaction was allowed to proceed with neurotensin at pH 2, it was found that only the C-terminal of the peptide got labeled with DPDAM, not the D/E residue as shown in Figure 3-4. Figure 3 (A) shows the full MS spectrum of neurotensin labeled with DPDAM and 920.95 and 614.03 represent +2 and +3 charged states, respectively. DPDAM modification at neurotensin was identified by the CID-MS/MS spectra. The fragmentations of the triply charged molecular ion at *m/z* 614.03 confirmed the peptide sequence as well.

We also performed a control reaction in which bradykinin, which neither has E nor D, was allowed to react with DPDAM and it was found that the C-terminal of the peptide got labeled at pH 2 and 4. The full MS spectrum of the labeled peptide is shown in Figure 3-5, 613.78 represents the +3 charged state of the peptide labeled with DPDAM.

A)



B)

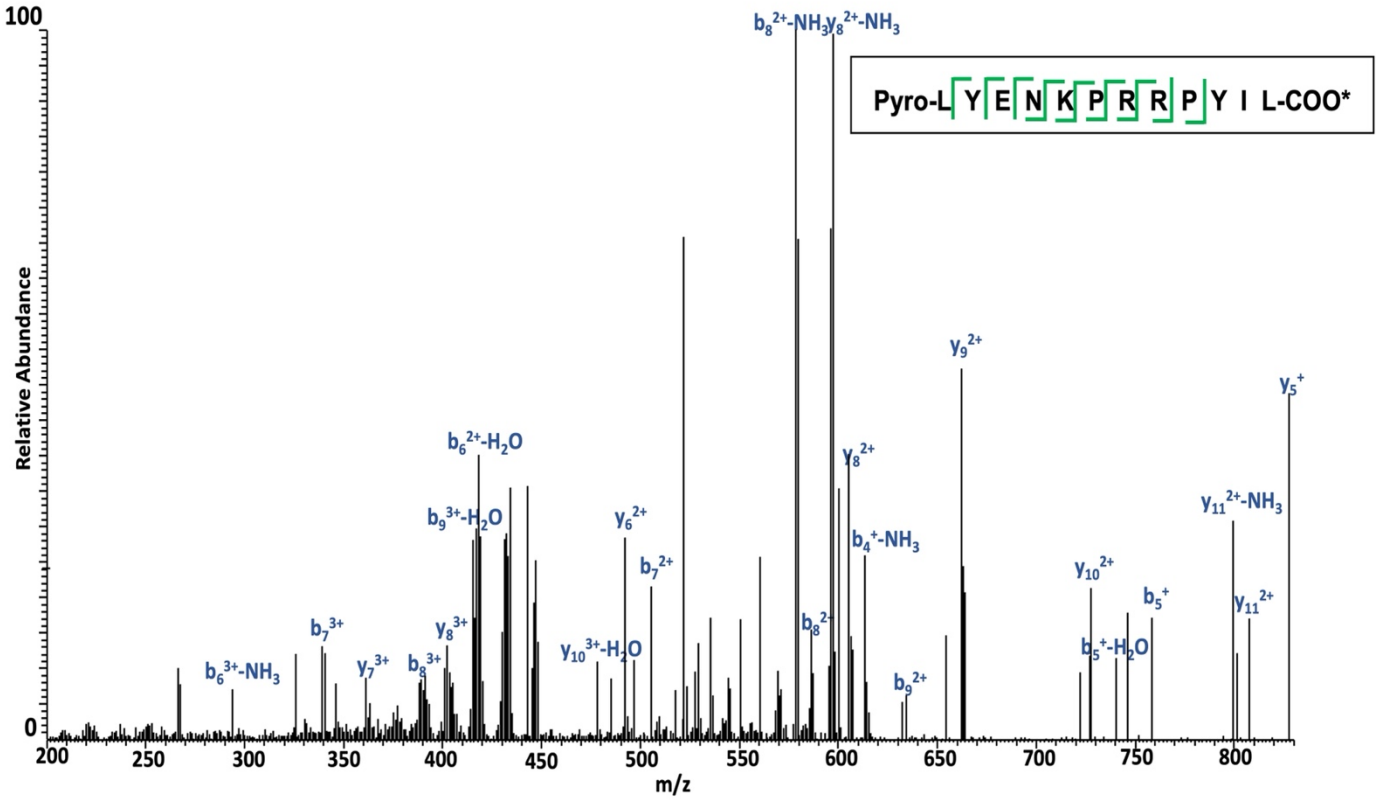
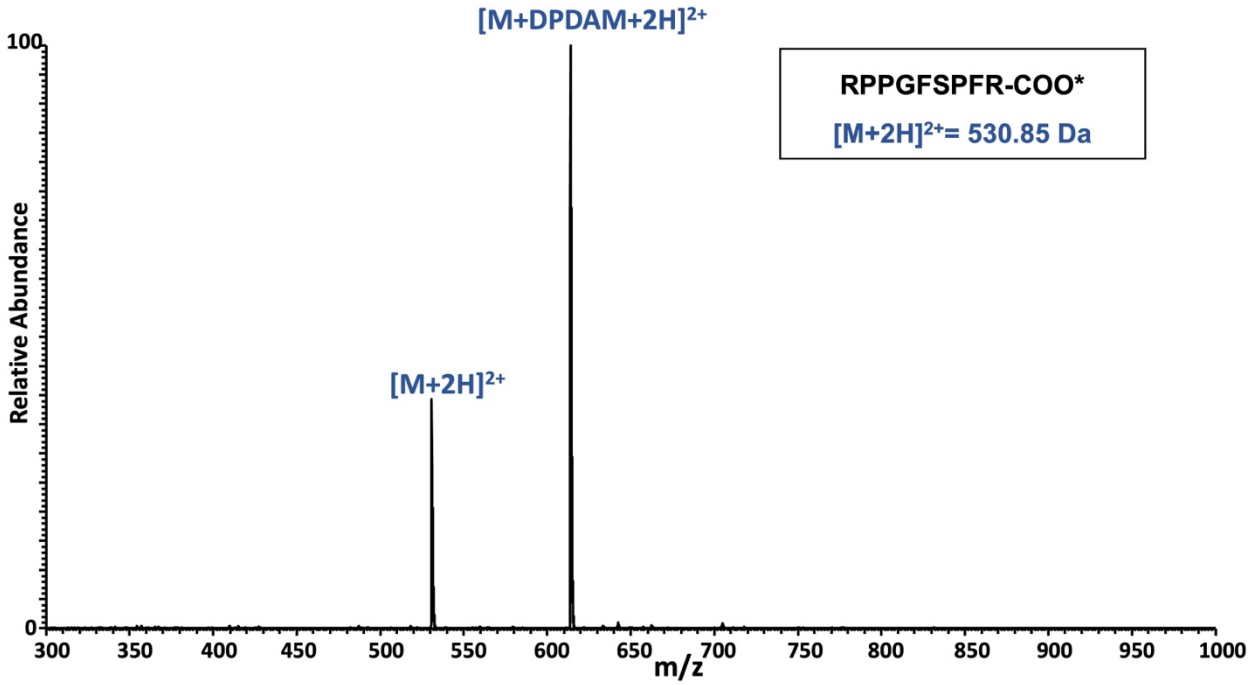


Figure 3-4 ESI-MS and MS/MS spectra of modified peptides with DPDAM at pH 2. (A) Neurotensin labeled with DPDAM at *m/z* 920.55 $[M+2H+DPDAM]^{2+}$ and 614.03 $[M+3H+DPDAM]^{3+}$ (B) MS/MS spectrum of modified neurotensin at *m/z* 614.03 $[M+3H+DPDAM]^{3+}$. *denotes DPDAM-modified sites.

A)



B)

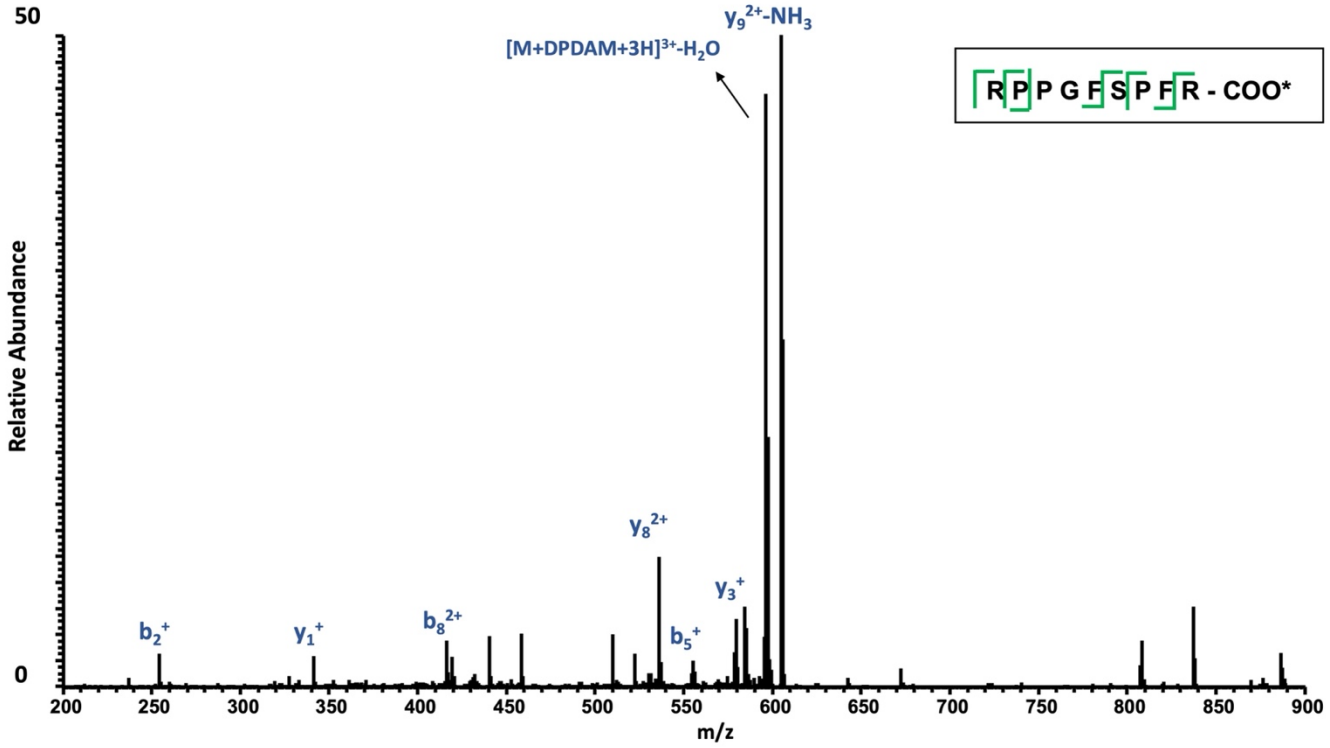


Figure 3-5 ESI-MS and MS/MS spectra of modified peptides with DPDAM. (A) Bradykinin labeled with DPDAM at *m/z* 614.16 $[M+2H+DPDAM]^{2+}$ (B) MS/MS spectrum of modified bradykinin at *m/z* 614.16 $[M+2H+DPDAM]^{2+}$. *denotes DPDAM-modified sites

3.4.2 Modification of proteins by DPDAM

After the successful labeling of peptides with DPDAM, we then increased the sample complexity. The labeling reaction of DPDAM was performed with a standard protein, bovine serum albumin (BSA). After in-solution digestion, the labeled peptides were identified in HCD-MS/MS. We were able to identify modified D and E residues in BSA protein. We found the peak at *m/z* 1002.4719 $[M+DPDAM+4H]^{4+}$ which corresponded to a peptide with amino acid sequence SHC!IAEVEKDAIPENLPPLTAD*FAED*KD*VC!K (*-DPDAM modified,!-cysteine -SH blocked by carbamidomethylation).

To confirm the labeled sites, we employed HCD-MS/MS of labeled protein. MS^2 of the labeled peptide is shown in Figure 3-6. The MS/MS spectra of the labeled precursor ion at *m/z* 1002.4719 $[M+DPDAM+4H]^{4+}$ selected for HCD-MS/MS gave good coverage of corresponding *b* and *y*-ions shown in the spectrum. Asp307, Asp311, and Asp313 residues were labeled by DPDAM. It has already been reported that an epitope of intact BSA found in a peptide with residues 299–338 is likely recognized by the anti-BSA AB6 mAb. Close observation of X-ray crystal structure of BSA revealed that these residues are exposed to the surface and therefore are more susceptible to DPDAM conjugation as shown in Figure 3-7. D and E residues which are embedded were less likely to get labeled. Loss of phenyl cation from DPDAM was also observed in MS^2 spectra. These efforts will open new opportunities for studying post-translationally modifications such as nitration, sulfation, oxidation, on Asp and Glu residues.

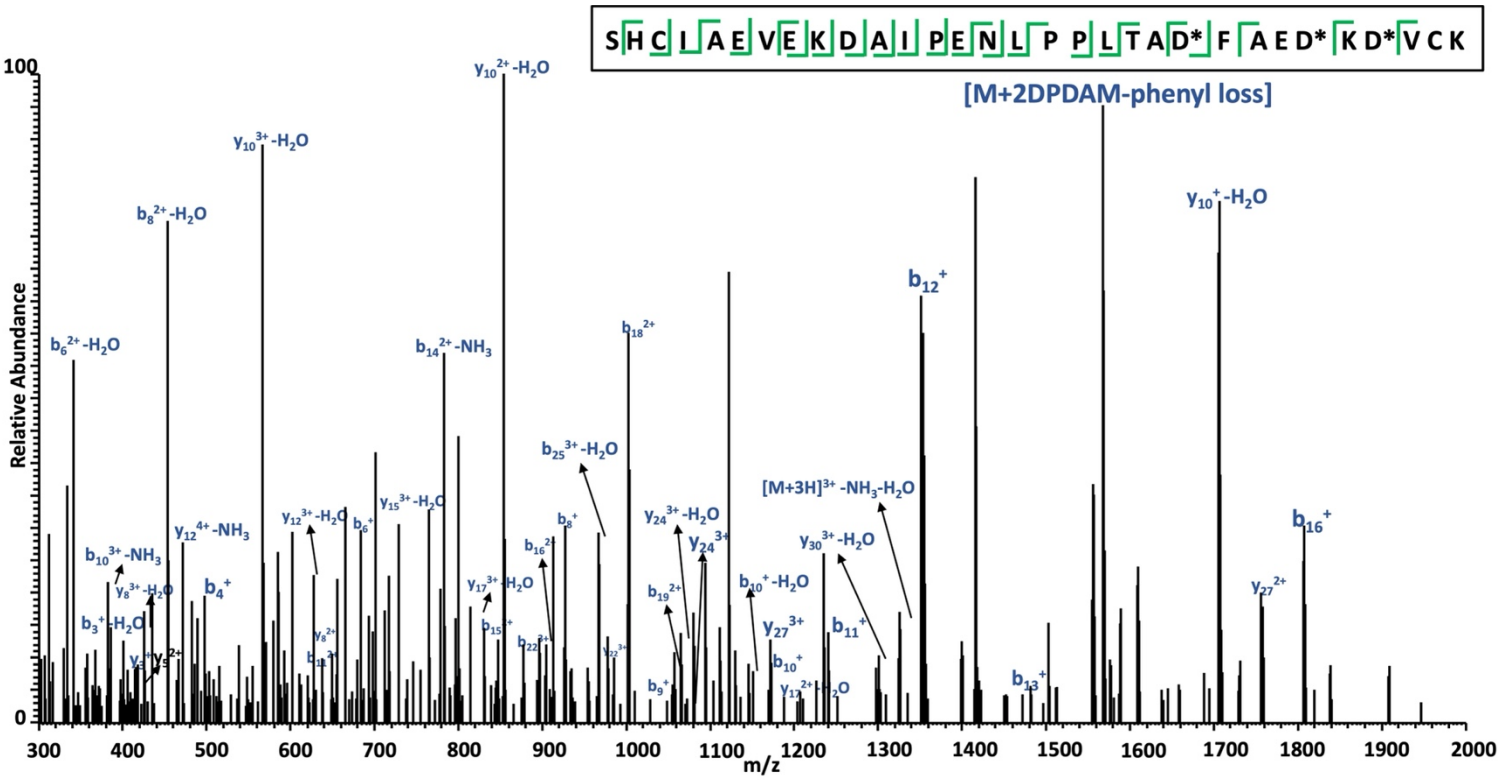


Figure 3-6 ESI-MS and MS/MS spectra of modified BSA with DPDAM. (A) BSA labeled with DPDAM at *m/z* 1002.4719 $[M+4H+DPDAM]^{4+}$ (B) MS/MS spectrum of modified peptide at *m/z* 1002.4719 $[M+4H+DPDAM]^{4+}$. *denotes DPDAM-modified sites

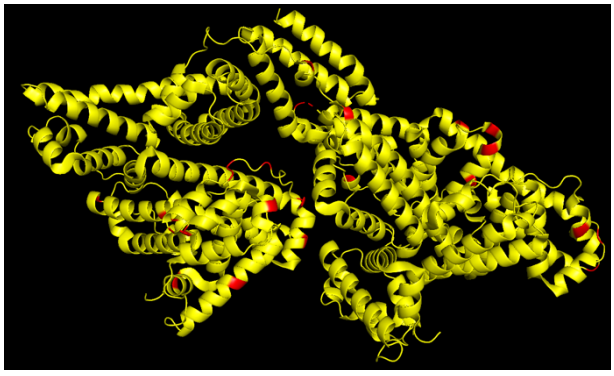


Figure 3-7 Crystal structure of BSA protein (PDB 3v03) The figure depicts modified sites of BSA protein with DPDAM.

3.5 Conclusion

Amino acid chemo-selective modification offers an effective approach for the investigation of protein structure and function. It can facilitate the identification of potential drug targets. New conjugation reagents and processes are therefore increasingly needed to enable selective binding to amino acids other than lysine, cysteine, and arginine. Our aim in this study is to target aspartic acid and glutamic acid residues, which are prevalent in proteins. We have demonstrated proof of concept in this study in model peptides and proteins. We have found that the majority of tagged acidic residues were located at the surface. The peptide sequences were confirmed using CID and HCD fragmentation. This work can further be extended in exploring the selectivity of DPDAM reagent on different forms of acidic residues and under different pH conditions. We hope that by using this method, the profiling of reactive carboxyl residues across the proteome can be done.

References

(1) Kalia, J.; Raines, R. Advances in Bioconjugation. *Current Organic Chemistry* **2010**, *14* (2), 138-147. DOI: 10.2174/138527210790069839 (acccessed 2023-08-03T06:50:54).

(2) Mendoza, V. L.; Vachet, R. W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrometry Reviews* **2009**, *28* (5), 785-815. DOI: 10.1002/mas.20203 (acccessed 2023-07-29T07:37:12).

(3) Zacharias, A.; Chowdhury, S. Developing an efficient Tyrosine Reactive Crosslinker (TRCL) for protein crosslinking. American Chemical Society (ACS): 2023.

(4) Leitner, A.; Joachimiak, L. A.; Unverdorben, P.; Walzthoeni, T.; Frydman, J.; Förster, F.; Aebersold, R. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proceedings of the National Academy of Sciences* **2014**, *111* (26), 9455-9460. DOI: 10.1073/pnas.1320298111 (acccessed 2022-10-10T18:31:11).

(5) Ma, N.; Hu, J.; Zhang, Z.-M.; Liu, W.; Huang, M.; Fan, Y.; Yin, X.; Wang, J.; Ding, K.; Ye, W.; et al. 2H-Azirine-Based Reagents for Chemoselective Bioconjugation at Carboxyl Residues Inside Live Cells. *Journal of the American Chemical Society* **2020**, *142* (13), 6051-6059. DOI: 10.1021/jacs.9b12116 (acccessed 2023-07-29T07:52:20).

(6) Furrow, M. E.; Myers, A. G. A General Procedure for the Esterification of Carboxylic Acids with Diazoalkanes Generated in Situ by the Oxidation of N-tert-Butyldimethylsilylhydrazones with (Difluoroiodo)benzene. *Journal of the American Chemical Society* **2004**, *126* (39), 12222-12223. DOI: 10.1021/ja0459779 (acccessed 2023-07-29T07:48:01).

(7) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; et al. SARS-CoV-2 Cell Entry Depends on

ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181* (2), 271-280.e278. DOI: 10.1016/j.cell.2020.02.052 (acccessed 2023-08-03T23:41:26).

(8) Ozono, S.; Zhang, Y.; Ode, H.; Sano, K.; Tan, T. S.; Imai, K.; Miyoshi, K.; Kishigami, S.; Ueno, T.; Iwatani, Y.; et al. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nature Communications* **2021**, *12* (1). DOI: 10.1038/s41467-021-21118-2 (acccessed 2023-08-03T23:46:16).

(9) Majumdar, P.; Niyogi, S. SARS-CoV-2 mutations: the biological trackway towards viral fitness. *Epidemiology and Infection* **2021**, *149*, 1-19. DOI: 10.1017/s0950268821001060 (acccessed 2023-08-03T23:49:46).

(10) Mcgrath, N. A.; Andersen, K. A.; Davis, A. K. F.; Lomax, J. E.; Raines, R. T. Diazo compounds for the bioreversible esterification of proteins. *Chemical Science* **2015**, *6* (1), 752-755. DOI: 10.1039/c4sc01768d (acccessed 2023-07-29T06:13:13).

Biographical information

Ashima Mehta grew up in a small town in Haryana, India. She completed her Bachelor of Science degree from the University of Delhi, India. She obtained her Master of Science degree from Jawaharlal Nehru University, Delhi, India. She moved to the United States in 2021 and pursued her Master of Science degree from the University of Texas at Arlington (UTA) and worked on selective covalent labeling in proteins and their analysis using mass spectrometry technique. She received her Masters degree in Chemistry from UTA in August 2023.