LARGE-SCALE IDENTIFICATIONS OF PROTEIN POST-TRANSLATIONAL MODIFICATIONS AND PROTEIN INTERACTIONS BY MASS SPECTROMETRY

By

ZIXIANG FANG

DISSERTATION

Submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy at

The University of Texas at Arlington

December 2020

Arlington, Texas

Supervising Committee:
Saiful M. Chowdhury, Supervising Professor
Subhrangsu S. Mandal
Kevin A. Schug
Krishnan Rajeshwar

Abstract

LARGE-SCALE IDENTIFICATIONS OF PROTEIN POST-TRANSLATIONAL MODIFICATIONS AND PROTEIN INTERACTIONS BY MASS SPECTROMETRY

Zixiang Fang, PhD

The University of Texas at Arlington, 2020

Supervising Professor: Saiful M. Chowdhury

Biological organisms depend on proteins to carry out essentially all cellular activities and functions. Therefore, proteins are structurally complex to perform all the intricate cellular processes and functionally dynamic to adapt to the changing environments. Protein diversity and complexity are further contributed by the attachments of different post-translational modifications (PTMs) and complicated inter-molecular interacting networks, which are extremely challenging to study by traditional techniques. New-generation mass spectrometry (MS) has become a major driving force that drastically expand detection limits and improve capabilities to provide unprecedented information for protein molecular and structural determinations. My research focuses on developing MS-based proteomics strategies and utilizing the separative power of liquid chromatography (LC) for studying low-abundance protein PTMs and protein interactions in large-scale biological samples.

There are two major protein PTMs that I investigate during my PhD research: prenylation and proteolysis. Both prenylation and proteolysis are important for initiating and regulating protein functions and activities while having significant implications in pathogenic development

of many diseases. To better identify the molecular constituents of these low-abundance but important protein modifications, I have been developing chemical and affinity approaches to target specific functional groups on these modifications for enrichment, as well as incorporating a gas-phase cleavable bond to prioritize the fragmentation and subsequent identifications of modification-containing peptides from complex cell lysate. The identification and enrichment study of prenylation are described in Chapter 2 and Chapter 3. The enrichment and identification strategy of protein proteolysis are introduced in Chapter 4. For the study of protein structures and interactions in Chapter 5, chemical cross-linking/MS strategy was applied to capture the potential interactive sites in proximity and improve the separation of cross-linked peptides by different column chemistries. We believe that these studies will facilitate the future proteomics research of protein PTMs and interactions.

Acknowledgements

I would like to express my most sincere gratitude towards my PhD mentor and research advisor Dr. Saiful Chowdhury, who accepts me into his lab, introduces me into the world of life science research, teaches me relevant research techniques and troubleshooting skills, and encourages me to work independently and think critically in tackling research challenges. It is truly my honor to have him guiding and supporting me throughout my PhD academic research.

I want to thank my committee members Dr. Subhrangsu S. Mandal, Dr. Kevin A. Schug and Dr. Krishnan Rajeshwar. They provide valuable feedbacks and suggestions on my research work, which helps me in better understanding and improving my research projects. From the discussions and interactions with them, I learn about what are the qualities needed to be a scientist and how to better present the science in my research. Last chapter of this dissertation in evaluating the separation of cross-linked peptides was a project in collaboration with Dr. Schug's lab, where their expertise in separation science provide important inputs in the completion and publication of my very first research article.

I am also thankful to my lab members in Dr. Chowdhury's lab for helping me started when I first joined the lab, and all the discussions and collaborations along the way. I feel very grateful and fortunate to be a graduate student in the chemistry and biochemistry department where I received tremendous academic and financial assistance, generous helps and supports from the faculties and staffs in the department. They are always being professional and willing to help whenever I encounter any difficulties or confusions. Special thanks to the Shimadzu Center for Advanced Analytical Chemistry (SCAAC) in UTA for the state-of-the-art instrumentations in facilitating my research, as well as the assistances and trainings from the friendly staffs.

October 25th, 2020

Table of Contents

List of Illustrations

List of Tables

List of Abbreviations

ABC	ammonium bicarbonate
ACN	acetonitrile
BH	biotin hydrazide
BM	biotin maleimide
BS^3	bis(sulfosuccinimidyl)suberate
BSA	bovine serum albumin
CID	collision-induced dissociation
DCM	dichloromethane
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
ESI	electrospray ionization
ETD	electron-transfer dissociation
FA	formic acid
FDR	false discovery rate
FTase	farnesyltransferase
GDP	guanosine diphosphate
GGTase	geranylgeranyltransferase
GTP	guanosine triphosphate
HPLC	high performance liquid chromatography
HVR	hypervariable region
IAM	iodoacetamide
IT	ion trap
LC	liquid chromatography
LIT	linear ion trap
MALDI	matrix-assisted laser desorption/ionization
mCPBA	meta-chloroperoxybenzoic acid
MS	mass spectrometry
MWCO	molecular weight cut-off
NLMS3	neutral loss MS^3
PBS	phosphate buffer saline
PFP	pentafluorophenyl
PSM	peptide spectrum match
PTM	post-translational modification
RI	reporter ion
RP	reversed phase
SCX	strong cation exchange
TFA	trifluoroacetic acid

TIC	total ion chromatogram
TOF	time-of-flight
XIC	extracted ion chromatogram

Chapter 1

Introduction

1.1 Protein post-translational modifications and protein interactions

1.1.1 Protein post-translational modifications

Proteins are a class of versatile, widespread, and abundant biomacromolecules central to the vitalities of virtually all living organisms. Commonly constructed by 20 amino acids through the linkage of peptide bonds, human proteins are estimated to reach millions in number, significantly more diverse and complex than 20000-30000 genes that encode them[1]. This significant increase of number from genes to proteins is mainly because a majority of proteins can undergo post-translational modifications (PTMs), modifications that covalently attach to the amino acid side chains or protein N-terminus/C-terminus for enabling proteins to mature structurally and functionally[2]. Recent release on UniProt protein database shows a total number of 633 different protein PTMs[3], among them the widely studied ones are phosphorylation, acetylation, acylation, acylation, glycosylation. PTMs are critical constituents of proteins in determining protein structures, stabilities, localizations, and other cellular activities. For example, disulfide is an important PTM formed by the coupling of two cysteinyl thiol groups to maintain the three-dimensional structures and cellular functions of proteins[4]. In addition, different PTMs are constantly coordinating and interacting to create a highly dynamic and flexible cellular environment, where interplay and orchestration between multiple PTMs are required for many cellular signaling pathways. For example, the activation and downstream signaling of epidermal growth factor receptor are regulated through a series of reversible phosphorylation, ubiquitination, and acetylation at different stages of signal transductions[5]. Furthermore, abnormality and disfunction of PTMs are frequently discovered to be the underlying molecular causes for a slew of human diseases including cancers and neurodegenerative diseases[6]. Therefore, the importance in understanding the implications and biological functions of these complex PTMs is hard to be overstated. My PhD research focuses on the study of two PTMs, prenylation and N-terminal proteolysis.

1.1.2 Protein prenylation

As one of the lipidations that increases the hydrophobicity of the modified proteins, protein prenylation is a type of irreversible, covalent modifications with the addition of a farnesyl group, a 15-carbon isoprenoid moiety or a geranylgeranyl group, a 20-carbon isoprenoid moiety. Protein prenylation is an enzymatic process that attach the prenyl group to the side chain of cysteine located at or near the C-terminus of proteins, most of which contain a conserved motif CAAX (C is cysteine, A is a aliphatic amino acid, and X can be one of several amino acids)[7]. The enzymes involved in these prenylation processes are farnesyltransferase (FTase) and geranylgeranyltransferase I (GGTase I), for the catalysis of farnesylation and geranylgeranylation, respectively[7,8]. After the prenylation, the AAX motif of the modified proteins can be further removed by Ras converting CAAX endopeptidase 1 and then the C-terminal carboxylic end is methylated, resulting in proteins with prenylated cysteine as the last amino acid at the methylated C-terminus[9]. Another enzyme geranylgeranyltransferase II (GGTase II), also termed Rab GGTase, targets a C-terminal CXC or CC sequence to modify one or both of the cysteines[9,10].

The earliest study of isoprenoid modification was in late 1970s when a fungal mating factor rhodotorucine A was found to contain farnesyl lipophilic chain and the related sequence was determined by Edman degradation[11]. The discovery of prenylation incorporation of mammalian proteins was later reported in the search of downstream products of mevalonate within the cholesterol biosynthesis pathway, by the radioactive labeling of mevalonate with ^{3}H in mouse embryo tissue cells[12]. Nuclear envelope protein lamin B was later reported to be modified by this mevalonate derivative, which is the farnesylation[13]. Many more protein substrates for prenylation had been since discovered including small GTP-binding Ras proteins that are important in regulating cell differentiations[8]. It is anticipated that over 0.5% of all animal proteins can be prenylated[14] and over 5000 proteins are predicted to be substrates of three prenyltransferases (FTase, GGTase I and GGTase II) according to C-terminal sequences in National Center for Biotechnology Institute (NCBI) protein database[15]. Non-canonical CXXXX sequence at protein C-terminus was also

reported to be prenylated which almost doubles the potential substrates of FTase[16]. Whereas only around 170 proteins are experimentally identified as prenylated to date according to UniProtKB reviewed human database[17], implicating that many more are yet to be discovered.

Intensive research efforts in improving the understanding of molecular and biological implications of prenylation are because of the apparent associations of prenylation with many human diseases, especially cancers[18]. Mutated Ras genes were among the first oncogenes reported from human cells but had been considered 'undruggable' due to the insufficient knowledge and understanding of the relevant functions and chemistry[19]. The connection of Ras mutations and cancers has since been gradually identified in more than 30% of human cancers[20], predominantly in pancreatic cancer[21], colorectal cancer[22]. Oncogenic mutation of Ras proteins disrupts the normal regulation of GTP (guanosine triphosphate) and GDP (guanosine diphosphate) binding and develop tumorigenic activities as the GTP-bound state persists to constantly activate the downstream signaling pathways of cell growth[23]. Prenylation of Ras proteins plays a key role in the membrane association activities of the pathway, which has been investigated as potential cancer therapeutic targets by developing and testing new inhibitors of FTase and GGTases as anti-Ras drugs[24]. Prenylation process is described to be involved in the development of many diseases including premature aging disease Hutchinson–Gilford progeria syndrome (HGPS), Alzheimer's disease, hepatitis and other infective diseases[25]. Contrast with the pivotal roles that prenylation plays in the pathogenic processes and disease developments, comprehensive and proteome-wide understanding of prenylated substrates still far from completion and requires continued research endeavors.

1.1.3 Protein proteolysis

One of the most ubiquitous protein PTMs is the proteolytic processing, mediated by over 560 proteases representing 2% of the human genome[26]. Biosynthesis of protein polypeptide chain generally starts with the manufacturing of methionine residue at the N-terminal end. For a majority of proteins, the proteolytic cleavage of the initiator methionine by methionine aminopeptidases is the earliest and most common

proteolytic event following protein synthesis[27]. One important function of proteolysis is the protein degradation, which can act as a "quality control" process for nascent proteins and a disposal mechanism that regulates the cellular stability and lifetime of proteins by trimming the protein N-terminus to specific amino acid residue[28]. Some protein N-terminus may contain signal peptides which are removed during the protein maturation processes. These sites of N-terminus removal and the sequences of the resultant mature proteins cannot be precisely predicted solely from the genome sequences. The different destabilizing effects of 20 amino acids at the protein N-terminus have been reported and they are termed 'N-degrons'[29]. Besides, proteolytic processing of proteins to subunits can dynamically change their cellular structures in order to maintain a myriad of normal functions, signaling and activities in response to external environment and stimulus. One such example is a family of proteases called caspases that cleave proteins at asparagine residue to activate the cytokine signaling pathways and regulate the cell apoptosis processes[30]. Proteolysis is also the driving force in different stages of cell cycle by regulating activities of kinases[31]. The participation of proteases in nearly all biological processes renders them important players in the development of many hereditary diseases and tumorigenesis, which is being specifically targeted by 5-10% drugs as potential therapeutics for these pathologies[26]. To better understand the disease mechanisms and reveal the protein proteolytic products as potential biomarkers, research has been focusing on the identifying specificity of proteases and their respective protein substrates.

1.1.4 Protein structures and protein interactions

The polypeptide sequence of a protein formed by a long chain of individual amino acid through peptide bonds is described as the primary structure. However, deciphering the primary sequence is not sufficient to depict the three-dimensional structures of proteins which are stabilized by many other weak and non-covalent interactions, shaping the proteins into higher secondary, tertiary and quaternary structural levels. Non-covalent interactions are also the major driving forces behind the protein-protein interactions, as almost all functions of proteins depending on the associations and communications with other

macromolecular complexes. The first and foremost challenge in understanding protein function is to establish the understanding of protein native structure and interaction network by the study of structural biology.

1.2 Mass spectrometry-based proteomics

1.2.1 Mass spectrometry instrumentations for proteomics

As one may assume by the diversity and complexity of protein structures, modifications and interactions, comprehensive understanding of proteins within the context of cellular and molecular biology is by no mean a simple task. Since the term 'proteome' was introduced back in 1995[32], the study of proteome, proteomics, has been in the spotlight of biological and biochemical research. The toolbox of techniques and methods for the visualization, identification and quantitation of proteins such as gel-electrophoresis, protein microarray, immunoassays, Edman sequencing and imaging techniques has been expanding, but the development of sensitive and specific mass spectrometry (MS) technique is by far one of the most powerful additions to the global, comprehensive proteomics research[33]. A typical mass spectrometer has three major components: ion source, analyzer and detector. The ion source ionizes the molecules of interest in the gas phase, followed by the separation of molecular ions in the analyzer according to their mass to charge ratio (m/z) and subsequent entrance to the detector for detection and signal amplification.

The most widespread ion sources for ionizing proteins and peptides are the two 'soft' ionization techniques: matrix-assisted laser desorption/ionization (MALDI)[34] and electrospray ionization (ESI)[35]. Soft ionization depicts the ionization process where the biomolecules remain intact so that the m/z of the native peptides or proteins can be monitored. MALDI source utilizes laser to irradiate the dry, crystalline sample prepared by mixing the analyte with an acidic organic compound called 'matrix' on a metal plate, where the matrix absorbs the laser energy, facilitates the isolation and the formation of molecular ions of proteins or peptides[36]. ESI ion source instead implements high voltage set-up to produce electric gradient where the analytes are carried through by solvent with desirable pH to form highly charged droplets in the gas phase

through desolvation process under atmospheric environment[37]. Compared to the predominantly singly charged molecular ions generated in MALDI, ESI process is capable of generating molecular ions in multiply charged states.

After ionization process, molecular ions are transmitted into the analyzer for ion separation. Commonly used mass analyzers for proteomics study including ion trap (IT), quadrupole (Q), time-of-flight (TOF), ion cyclotron resonance (ICR) and Orbitrap. Each different mass analyzer has its own characteristics and properties, some of the important metrics are mass accuracy, scan rate, sensitivity. To maximize the performance and capabilities of these analyzers, hybrid analyzers are more powerful and suitable for proteomic research. Hybrid mass spectrometer systems such as Q-TOF, QQQ, IT-Orbitrap enable the fragmentation functionalities to provide product ions specific to the peptide sequences, which are critical for confidently identifying the amino acid compositions of the peptides. This additional stage of fragmentation is called tandem mass spectrometry, which contains more than one level of MS acquisition. The molecular ion that corresponds to the m/z of intact peptide or protein produced in the MS^1 stage is called precursor ion to the product ions (or fragments ions) in the MS^2 stage. Further fragmentation of the fragments in the MS^2 level is also feasible in IT analyzers to generate MS^3 or even MS^n spectrum, which can be useful in providing more qualitative and quantitative information based on specific proteomic applications.

The fragmentation pattern in the tandem MS spectrum for identifying peptide sequence is dependent on the dissociation methods being used. The most universally utilized dissociation method is the collision-induced dissociation (CID)[38], where the dissociation of peptide is induced by collision with inert gas in the fragmentation chamber. The kinetic energy of ion is converted to internal energy which breaks the weakest bonds of the peptide chain, generally at the amide linkages in the peptide backbone to give rise to b-ions and y-ions (Figure 1-1)[39]. Another complementary technique to CID is electron-transfer dissociation (ETD) which involves reagent anions transferring the electrons to molecular cation which undergoes radical

mediated reaction that cleaves at the N-Cα linkage to produce c-ions and z-ions[40]. ETD is specially suitable for investigating peptide ion of a higher charge state (> +2) or containing CID-labile PTMs[41].

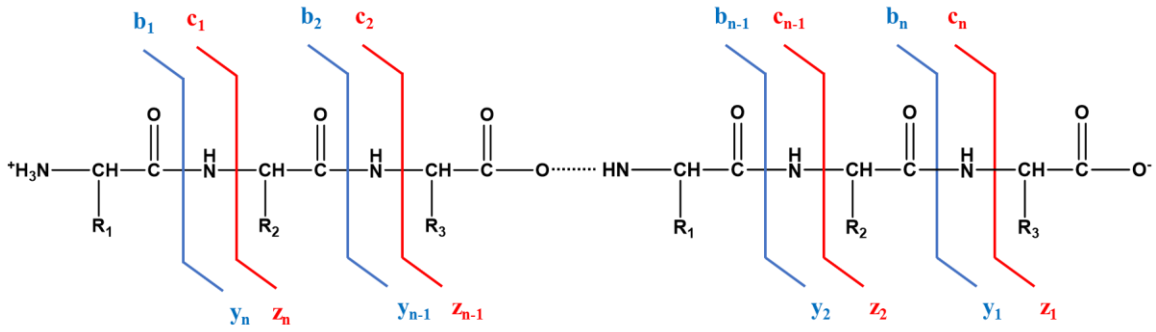


Figure 1-1. Fragmentation pattern of peptide under CID (b and y ions) and ETD (c and z ions)

When it comes to utilizing MS instruments for the identification of proteins, two general strategies are commonly and sometimes complementarily used: top-down approach and bottom-up approach[42]. Top-down proteomic approach analyzes the intact molecular composition of proteins by direct ionization and subsequent fragmentation, which normally requires a high-resolution mass analyzer for accurate acquisitions of protein masses. While top-down approach has the advantages of full protein coverage, protein isoform characterization and PTM identification, intact proteins are comparatively more difficult to be manipulated in solubilization, fractionation, separation, ionization, fragmentation and bioinformatic interpretation[43,44]. In comparison, bottom-up proteomic approach characterizes proteins by digesting them into shorter peptides using a protease (commonly trypsin). The peptides are identified after going through separation and fragmentation in the LC-MS/MS instrument and the resultant mass spectra are matched against a theoretical mass list generated from *in silico* digestion of a protein database[45]. The identified peptides are further grouped together to recover the sequence of a protein, similar to putting together pieces

to a puzzle. Overall, bottom-up approach is more widespread, and it has been the major approach for my PhD research.

1.2.2 Cross-linking mass spectrometry technology

The rapid advancement in MS instrumentation and analyzing techniques has open up vast possibilities of tackling the previous challenging difficulties in structural biology. Mapping protein structures and protein interactions have been traditionally achieved by X-ray crystallography[46] and nuclear magnetic resonance (NMR)[47], newer approaches such as cryo-electron microscopy (cryo-EM)[48] and cross-linking/MS[49] not only complement in protein characterizations but exhibit more promises in enabling structural interpretation from samples with limited quantity, large molecular size and high complexity. Cross-linking/MS has the advantages of simple sample preparation, structural capturing *in vivo* and capturing of weak and transient interactions, by the utilization of a cross-linker that contains two reactive groups targeting the side chains of amino acids on the surface of proteins under physiological environment[50]. The cross-linker can act as a molecular ruler that measures the distance between the reacted amino acid sites because the length of a cross-linker can be readily modified to constrain the distance for the reaction. Several types of cross-linked products are generated from a cross-linking experiment[51] which can provide different structural information (Figure 1-2): (1) Inter-cross-linked products provide information on the proximity of different amino acids, either within the same protein or from interacting proteins to provide low-resolution structural information of an individual protein or potential interactive sites on different proteins; (2) Dead-end cross-linked products have only one reactive group of the cross-linker attaching to an amino acid while the other reactive group is normally hydrolyzed. These products can provide information about the accessibility of amino acids on the surface of proteins; (3) Intra-cross-linked products can inform us about the accessibility of amino acids on the native proteins and the distance between the reactive sites; (4) Unmodified peptides are not reacted by the cross-linker and remain the same as digestion of the non-cross-linked proteins.

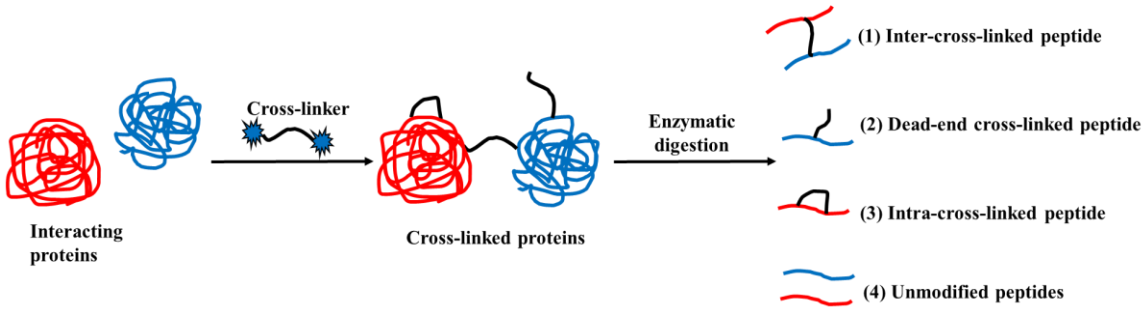


Figure 1-2. Different types of cross-linked peptides from a typical cross-linking reaction

The caveats on the application with cross-linking/MS technology are the further experimental and bioinformatic complication of the already complex samples in a general bottom-up proteomic study. Theoretically all peptides in a protein can be inter-, intra-, or dead-end cross-linked, with the inter-cross-linked products being the most structurally informative yet molecularly scarce (considering the higher abundance of unreacted or dead-end peptides). It has been a major challenge to separate the differently cross-linked products to improve the identification of inter-cross-linked peptides. An enrichable tag like biotin can be incorporated in the cross-linker to enrich the cross-linker modified peptides by biotin-avidin affinity chromatography. However, it still fails to differentiate the dead-end cross-linked products from inter-cross-linked products. The last chapter of this dissertation addresses this issue by evaluating unconventional stationary phases to separate different cross-linked products chromatographically.

1.2.3 Cleavable mass-spectrometry-based proteomics

Another advancement in improving the identification of cross-linked peptide is the rapid development of MS-cleavable cross-linkers to simplify the fragmentation pattern and provide particular signature fragment ion for the differentiation from other unmodified counterparts. MS-cleavable or gas-phase cross-linkers have certain cleavable sites which can be preferentially cleaved upon the ion activation in the fragmentation compared with the cleavage of peptide backbone. Among the earliest MS-cleavable cross-linkers being

developed were the a series of protein interaction reporters (PIR) utilizing Rink group, an acid cleavable group commonly adapted in the peptide synthesis and found to be gas-phase cleavable, as the low energy cleavable site within the cross-linkers[52,53]. Another widely utilized MS-cleavable bond is the sulfoxide group, which was discovered to undergo fragmentation through pathway of β-hydrogen elimination[54]. This unique property enables sulfoxide group to be incorporated into a CID-cleavable cross-linker called disuccinimidyl sulfoxide (DSSO) for structural elucidation of protein complex such as yeast proteasome[55]. Many other cleavable functional groups incorporated into the cross-linking applications are reviewed here[56].

The abilities of these MS-cleavable functional groups are also applied to the identification of low abundant PTMs. Phosphopeptides are derivatized with 2-dimethylaminoethanethiol and oxidized for the conversion to cleavable sulfoxide group in order to be detected in positive ion mode in triple quadrupole instrument[57]. Previously, our lab has develop a strategy for identifying low-abundance prenylation and differentiating the prenylation types by converting the thioether linkage to sulfoxide which can readily generate high-abundance signature fragments after neutral losses in the CID fragmentation[58,59]. Sulfoxide-based isobaric tags also exhibit more accurate quantitation in multiplexed experiments thanks to the low-energy fragmentation and high reporter ion yield[60,61].

1.3 Thesis organization

This dissertation will focus on the development of new MS-based approaches to study protein PTMs and protein interactions. Chapter 2 discusses on developing a MS-cleavable neutral loss MS^3 (NLMS3) strategy to identify low-abundance prenyl proteins from large-scale cell lysate. Chapter 3 continues to expand on the research of protein prenylation to explore the enrichment of prenylated peptides through selective biotinylation after epoxidation, as this can greatly reduce sample complexity during the MS acquisition. Another PTM of interest, protein proteolysis, is detailed in chapter 4. In this chapter, a novel amine-reactive MS-cleavable reagent PFP (pentafluorophenyl)-Rink-biotin is utilized for the large-scale

enrichment and identification of protein N-terminus and proteolytic products, facilitated by the bioinformatic strategy for automatic validation. New technique for separating cross-linked peptides is presented in Chapter 5, where we investigate the analytical aspect of how different stationary phases change the separative behaviors of different cross-linked products to optimize and improve the detection of low stoichiometric inter-cross-linked peptides.

Chapter 2

Neutral Loss MS^3 Strategy for Large-scale Profiling of Protein Prenylation

2.1 Abstract

Protein prenylation is an important post-translational modification (PTM) that regulates protein interactions, localizations, and signaling pathways in normal functioning of eukaryotic cells. It is also an essential process in the oncogenic developments of various cancers. Constant and rapid innovation of mass spectrometry-based instrumentations and techniques revolutionize the proteomics studies including modification-specific proteomics such as phosphorylation, glycosylation. However, direct identification of protein prenylation by mass spectrometry has been more challenging due to lower abundance, higher hydrophobicity, and lack of efficient enrichment techniques. To address these issues, a simple yet effective neutral loss MS^3 (NLMS3) based strategy is developed for the confident direct identification of protein prenylation, utilizing the preferential gas-phase CID fragmentation properties of prenyl thioether bond and sulfoxide moiety derivatized by the oxidation. For prenyl peptide, a signature fragment after neutral loss of prenyl group (loss of R group) is generated in high intensity in the MS^2 spectrum; For the same prenyl peptide after oxidation, a different signature fragment after neutral loss of oxidized prenyl group (loss of RSOH group) is observed in high abundance as well. These two distinct neutral losses can both be utilized for the identification of prenylation type, whereas the high abundance of both signature fragments enables further sequencing of the same prenyl peptide upon targeted MS^3 fragmentation. This dual-faceted NLMS3 strategy significantly improves the confidence in the identification of individual protein prenylation from large-scale samples, which enables the unambiguous identification of prenylated sites of the spiked low-abundance farnesyl peptide and native prenyl proteins from mouse cells, even without prior targeted enrichment during sample preparation. The ease of incorporating this strategy into prenylation study workflow and the minimum disruption to the biological lipidome are advantageous for unraveling unknown

native protein prenylation and promising in profiling and quantifying prenylome to better understand the molecular basis of this important PTM.

2.2 Introduction

As an important lipidation that regulates protein activities, interactions and localization, protein prenylation increases protein hydrophobicity to promote stable interactions of proteins with cellular membranes by attaching a 15-carbon farnesyl or a 20-carbon geranylgeranyl isoprenoid, generally to the cysteine located near the protein C-terminus[8,14]. The attachment of these two covalent modifications are irreversibly catalyzed by protein farnesyltransferase (FTase), protein geranylgeranyltransferase type I (GGTase-I) or protein geranylgeranyltransferase type II (GGTase-II)[62]. FTase and GGTase-I transfer the respective prenyl moiety to cysteine of C-terminal consensus motif CaaX ("C" is Cysteine, "a" represents aliphatic amino acid, X can be any amino acid), while GGTase-II, also known as Rab GGTase, instead targets the cysteine in the C-terminal CXC or CC motif[9]. Prenylation plays critical roles in the functions and activities of many proteins including nuclear lamin, Ras superfamily GTPases (Ras, Rho, Rab), subunits of heterotrimeric G proteins[63]. Most notably, prenylation of these protein initiates downstream signaling pathways that are responsible for the malignant activities of various human diseases such as pancreatic cancer[64], lung cancer[65,66] and Hutchinson–Gilford progeria syndrome (HGPS)[67,68], thus attracting intensive research interests in developing inhibitors of prenyltransferases in targeting the prenylation pathways for therapeutics[18,63]. Another research interest in protein prenylation is the identification of novel prenylation substrates, which is critical in deciphering the prenyl-dependent processes as well as evaluating the efficacy and underlying influences of prenyltransferase inhibitors[69]. It was estimated that 0.5% of total proteins in animal cells can be prenylated[14], with new substrates are constantly being discovered. Recently, non-canonical C-terminal sequence CXXXX was found to be prenylated as well, which could potentially doubles the protein substrates for prenylation in yeast and human proteomes.[16] Understanding the underlying molecular mechanisms and relevant biological activities of prenylome in cell functions of

normal and diseased states requires the discovery and identification of new prenylation substrates, potentially hundreds of which remains to be identified[18,70].

Detection of prenylation is first realized through autoradiography with ^{3}H-mevalonate labeling followed by immunoprecipitation or immunofluorescence, but drawbacks of radiochemical hazards, high cost, low throughput and inability to distinguish two different prenylation types render it unsuitable for global profiling of prenylation in complex samples[25,71]. Metabolic tagging-via-substrate strategy by incorporating synthetic isoprenoid analogues with unnatural enrichment functionalities was developed for targeted selection, enrichment and detection of prenylation within complex samples[72]. Numerous prenyl probes were subsequently designed and utilized for uncovering more substrates of FTase and GGTases.[73] However, challenges of this strategy include its limited applications in clinical samples and animal models, along with possible ambiguity in detection due to metabolic conversion of isoprenoid analogues to other lipids[25]. Besides, the prenylated substrates labeled and identified with synthetic isoprenoid analogues does not directly reflect the actual quantity of *in vivo* prenylated proteins in the biological system. Therefore, metabolic labeling strategy is not amenable to large-scale quantitation of native prenylome.

Mass spectrometry has grown into an indispensable technique in unraveling the molecular information of sophisticated biological systems. Advancements and improvements of mass spectrometric instrumentation has propelled qualitative and quantitative proteomics, which include the large-scale detection of low-abundant post-translational modifications (PTMs) such as phosphorylation and glycosylation[74]. However, only few studies reported on developing MS-based methods for prenylation characterization[75–78]. Similar to many other labile PTMs, the thioether linkage between cysteine and prenyl group is preferentially cleaved compared to peptide backbone cleavage. As a result, a neutral loss of 204 Da (farnesyl group) from precursor ions of farnesylated peptides was commonly observed under low-energy CID fragmentation[76–78]. This characteristic neutral loss was explored as a distinct identifier of farnesyl peptide for potentially pinpointing the site of prenylation. However, the selective cleavage of farnesyl group gives rise to a high

intensity of neutral loss fragment which complicates the interpretation of MS^2 spectra and adversely affects the efficient extraction of sequence information on the modified peptide. In the MALDI-TOF-TOF fragmentation study of farnesyl peptides, fragmentation pattern and intensity of the neutral loss were found varied with peptide sequence and charge state of precursors[76]. Another distinct property of the prenylated peptides is the high hydrophobicity compared to unmodified counterparts. This property was utilized in the multidimensional separation where low abundantly spiked farnesyl peptides in HeLa cell lysate matrix are strongly retained in hydrophobic stationary phase to be separated and identified, though no native prenylated protein was detected in the study[77]. More recently, an improved MS-cleavable strategy established for identifying and differentiating prenylated peptides by oxidation and epoxidation was developed[58]. This proof-of-concept method chemically converts the prenyl thioether linkage to sulfoxide group, which turns out to be a more labile gas-phase cleavable group and consistent in generating characteristic fragment after the neutral loss of RSOH (R = modified/unmodified prenyl side chain)[58,59]. This simple oxidation reaction simultaneously alleviates the high hydrophobicity and opens new opportunity to incorporate enrichment tags on the prenylated peptides for targeted enrichment.

Further exploring based on these prior studies, we develop a data-dependent neutral loss MS^3 (NLMS3) strategy for identifying prenylation in large-scale cell lysate samples without prior enrichment, through the identification of two individual signature neutral loss fragments and target them for further MS^3 sequencing (Figure 2-1a). This method was applied to the whole cell lysate for identifying the neutral loss of prenyl group from the prenylated peptides, as well as the neutral loss of RSOH from the oxidized prenyl group (Figure 2-2). NLMS3 based strategy had been successfully implemented for the characterization and quantitation of protein phosphorylation from large-scale samples[79,80]. Analogous to labile phosphorylation, the scissile nature of thioether group (prenyl) and sulfoxide group (oxidized prenyl) ensures that the neutral-loss-induced signature fragments are in high intensity among the fragments in the MS^2 spectra. Subsequently, charge-dependent signature fragments resulted from the respective prenyl neutral losses initiate MS^3 fragmentation specific to the signature fragment for interrogating the sequence of prenyl

peptides. Cross-examination by the combination of two signature fragments improves the confidence in prenylation discovery while minimizing the potential false-positive matches, which exploits the great potential of cleavable MS strategy in identifying and quantifying low stoichiometric prenylation from large-scale samples.

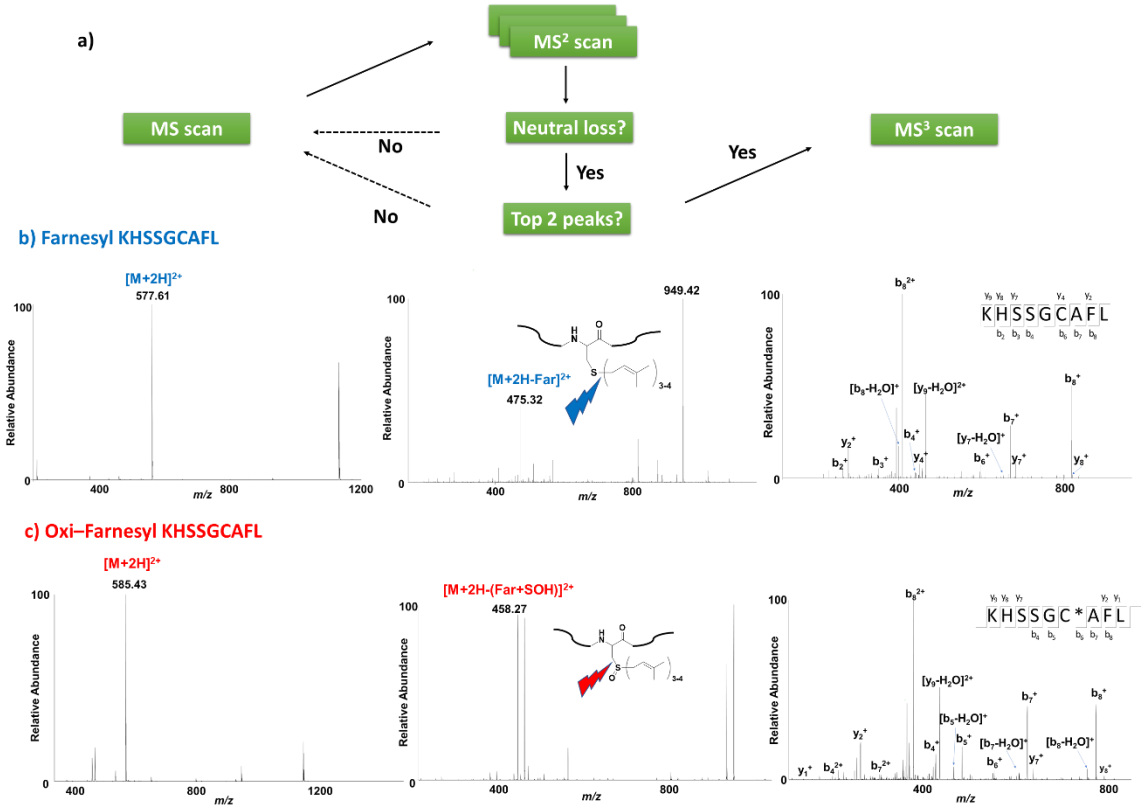


Figure 2-1. NLMS3 strategy for the identification of prenylated proteins

a) MS analysis decision tree for NLMS3 strategy. b) MS^1-MS^3 spectra for farnesyl KHSSGCAFL peptide; c) MS^1-MS^3 spectra for oxidized farnesyl KHSSGCAFL peptide

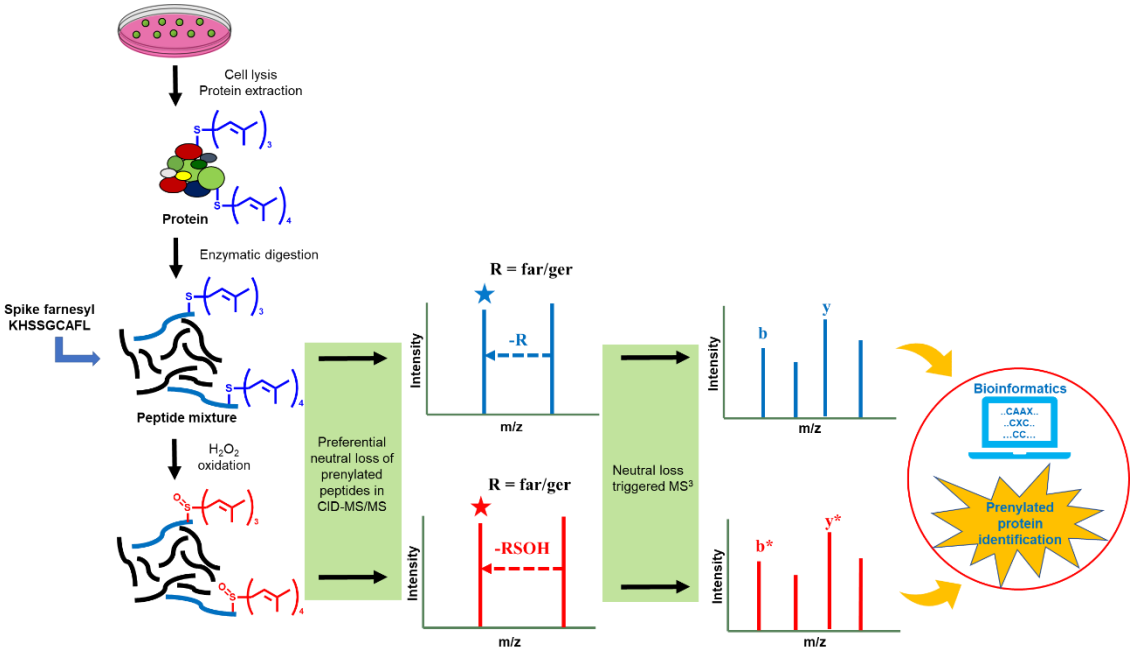


Figure 2-2. Workflow of NLMS3 strategy for the identification of native prenylated proteins from mammalian cells by two different signature neutral losses generated before and after oxidation

2.3 Materials and Methods

2.3.1 Materials

Murine raw 264.7 macrophage cells were a gift from Dr. Michael B. Fessler of the Immunity, Inflammation, and Disease Laboratory. Peptide KHSSGCAFL was purchased from AnaSpec. *Trans, trans*-farnesyl bromide, iodoacetamide, ammonium bicarbonate and dimethyl sulfoxide (DMSO) were purchased from Sigma Aldrich. Dithiothreitol (DTT) was acquired from BioRad. Dulbecco's Modified Essential Medium (DMEM), Iodoacetamide (IAM), methanol and acetonitrile (ACN) were obtained from VWR. Formic acid (FA), protease inhibitor, 30% hydrogen peroxide and C18 Tips were obtained from Thermo Fisher Scientific. Modified trypsin was purchased from Promega.

2.3.2 Prenylation and oxidation of standard peptide

Farnesyl KHSSGCAFL peptide was synthesized according to previously reported steps[58]. In short, KHSSGCAFL peptide was reacted with *trans, trans*-farnesyl bromide in 7 M ammonia in methanol at 4 °C in dark. Solvent was removed under vacuum and reconstituted with 0.1% FA. For oxidation of the farnesyl peptide, the farnesyl KHSSGCAFL peptide was subjected to oxidation with 0.5% H_2O_2 for 1 h at room temperature. All sample were desalted with C18 tips before mass spectrometric analysis.

2.3.3 Large-scale prenylation study of mouse raw macrophage

Raw 264.7 macrophage cells were cultured in DMEM supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin in a humidified atmosphere of 5% CO_2 maintained at 37 °C. The medium was replaced every 2-3 days until cells reached around 80% confluency. Cells were harvested after washing with phosphate buffer saline (PBS). Cell lysis was performed with 1X protease inhibitor as described previously[81] and the protein concentration was measured with bicinchoninic acid assay using bovine serum albumin standards. Protein extractions and digestions were modified from our previous studies[50,82]. In short, 100 μg of the extracted proteins from each raw macrophage biological replicate were precipitated by methanol/chloroform method and re-dissolved with 6 M guanidine hydrochloride in 50 mM ammonium bicarbonate, followed by reaction with 5 mM DTT and alkylation with 25 mM IAM in dark. The samples were then subjected to digestion with trypsin (1:50, w/w) overnight. The digestion was quenched by adding FA. Each raw macrophage replicate was split into 2 aliquots and each aliquot was spiked with farnesyl KHSSGCAFL. One aliquot was oxidized by adding H_2O_2 to a final concentration of 0.5% and mixed for 30 min at room temperature. All samples were desalted and reconstituted with 0.1% FA for mass spectrometric analysis.

2.3.4 Mass spectrometric analysis

Mass spectrometric analysis was performed in Thermo Velos Pro mass spectrometer. Farnesyl and oxidized farnesyl KHSSGCAFL peptides were analyzed by direct infusion. For the spiked study in raw macrophage cell lysate, Acclaim PepMap C18 column (150 mm × 75 μm, 3 μm) was used for the LC separation with either 90 min (MS^2 acquisition) or 135 min gradient (NLMS3 acquisition) using mobile phase A (0.1% FA in water) and mobile phase B (0.1% FA in 95% acetonitrile and 5% water) at a flow rate of 0.300 μL/min. In the 135 min gradient, mobile phase composition was held at 4.0% B for 3 min and increased to 50% B at 120 min, followed by 5 min at 95% B and equilibration back to 4% B for 10 min before next sample injection. Source voltage was 2.00 kV and capillary temperature was set to 275 °C. MS spectra were acquired from *m*/*z* 300 to 2000. Data-dependent CID-MS^2 spectra were collected for the 10 most abundant precursor ions in each MS scan at charge state ≥2 and isolated width of 2 Da. Other CID-MS^2 settings include 35.0% normalized collision energy, activation Q of 0.25, and activation time of 10 ms. Data-dependent NLMS3 was used for MS^3 data-dependent acquisition, the neutral losses were set to m/z 68 (Far 3+), m/z 102 (Far 2+), m/z 90.7 (Ger 3+), m/z 136 (Ger 2+) according to the m/z of farnesyl or geranylgeranyl group at doubly and triply charged states for the identification of prenyl peptides in the unoxidized samples; For the oxidized samples, neutral losses were monitored at m/z 84.3 (FarSHO 3+), m/z 126.5 (FarSHO 2+), m/z 107 (GerSHO 3+), m/z 160.5 (GerSHO 2+) based on the doubly and triply charged states of RSOH (R = farnesylation or geranylgeranylation). Neutral loss mass width was 1.50 by mass. Only the top 2 fragments in each MS^2 spectrum which also matched the neutral losses were selected for MS^3 fragmentation. Dynamic exclusion was enabled for the repeat count of 1, repeat duration of 30 s and exclusion duration of 30 s.

2.3.5 Mass spectrometric data analysis

MS^3 spectra of spiked macrophage samples raw data files were converted to mzML files by Proteome Discoverer 2.1 and searched in the MS-GF+[83,84] against Swiss-Prot reviewed mouse protein database

(accessed on 201912). Protein sequences of previously reported prenylated proteins (both intact and AAX removed protein C-terminus) along with the spiked peptide sequence KHSSGCAFL were manually added to the search. Precursor mass tolerance was set to 1 Da. Enzyme was set to trypsin. For unoxidized samples, cysteine carbamidomethylation, methionine oxidation and protein C-terminus methylation were set as dynamic modifications; For oxidized samples, dynamic modifications of cysteine carbamidomethylation, cysteine SH_2 loss, protein C-terminus methylation and methionine oxidation were allowed. All other parameters during the search were set as default. Peptides identified were applied a spectrum E-value cutoff of < $1e10^{-7}$. In the final reported results, only the MS^3 PSMs including cysteine in the identified sequences were considered as potential prenylated sequence and manually validated by the charge state of the corresponding neutral loss.

2.4 Results and Discussions

2.4.1 Identification of farnesyl peptide by two signature fragments and MS^3 fragmentation

The fragmentation pattern of synthetic farnesyl peptides has been investigated by MALDI-TOF/TOF and ESI-quadrupole/ion trap systems in the previous studies[76,77]. Although specific fragmentation pattern depends on the MS analyzer, peptide sequence and precursor charge state, a neutral loss equivalent to mass of farnesyl group was commonly observed from the farnesyl peptide precursors, resulting in farnesyl-specific signature fragment in CID-MS^2 spectra. This signature fragment can provide useful sequence characterization of the modified peptide upon further MS^3 fragmentation. An example showing here is the tandem MS analysis of a farnesyl peptide KHSSGCAFL in ESI-LTQ mass spectrometer analyzed by direct infusion (Figure 2-1b). Selective CID-MS/MS of doubly charged peptide precursor (m/z of 577.61) generates two major fragments in the MS^2 spectrum after the loss of farnesyl group in two different charge states while other fragments are observed in lower intensities. The annotated doubly charged fragment (m/z 475.32) is the fragment after neutral loss. The m/z difference of 102.29 between the MS^1 precursor and neutral loss fragment matches the calculated m/z of a farnesyl group in doubly charged state. The relative

high intensity enables further MS^3 fragmentation of this neutral loss fragment to produce rich b and y ion information for identifying the peptide sequence with high confidence.

In a parallel experiment, the same farnesyl peptide is oxidized with H_2O_2 which converts the prenyl thioether to sulfoxide, followed by tandem MS analysis. As demonstrated in the MS spectra of oxidized KHSSG(Far-)CAFL in Figure 2-1c, mono-oxidized doubly charged peptide precursor (m/z of 585.43) generates intense signature fragment (m/z 458.27) after the neutral loss of RSOH with the m/z difference of 127.16. Compared to MS^2 fragmentation pattern of the unoxidized prenyl peptides, the higher relative intensity of signature fragment after RSOH loss demonstrates higher lability of mono-oxidized prenyl peptides in CID-MS^2, which is consistent with the previous observation of RSOH neutral loss in MALDI-MS and ESI-MS systems[58,59]. MS^3 fragmentation of the RSOH neutral-loss-induced signature fragment again confirmed the same peptide sequence in high confidence.

Considering the two individual tandem MS analysis of the same farnesyl peptide in unoxidized and oxidized states, both signature fragments can be acquired with high intensity and provide unique gas-phase neutral loss characteristics, making them good diagnostic ions for predicting prenyl peptides. Further data dependent MS^3 of the two independent signature fragments each can independently identify the sequence of prenyl peptides. Therefore, we design a simple prenylation profiling strategy utilizing both cleavability of unoxidized and oxidized thioether prenyl linkage for unambiguously identifying protein prenylation type, characterizing the location of modification and determining the sequences of prenylated targets.

2.4.2 Signature fragments of quantitatively spiked farnesyl peptide in complex sample

Identification of farnesyl peptide and mono-oxidized farnesyl peptide through signature fragments were investigated in complex cell lysate. Raw macrophage cells are lysed, and protein were extracted as described in the method. After the enzymatic digestion, different amounts (10 µg, 1 µg, 0.1 µg, 0.01 µg) of farnesyl KHSSGCAFL peptide were spiked into 50 µg tryptic digest of raw macrophage cell lysate for monitoring the detection of signature fragments from the spiked peptide. Extracted ion chromatograms

(XICs) of the spiked peptide (1 Da window) are shown in Figure 2-3. The retention time of the spiked peptide (highlighted) was eluted consistently around 54 min of the gradient. In the CID-MS^2 analysis where the top 10 precursors of each mass spectrum were fragmented, the signature fragment of the farnesyl KHSSGCAFL peptide after farnesyl loss (m/z around 475) was observed from 10 µg spiked sample down to 0.1 µg spiked sample. For the spiked sample after oxidation, the signature fragment after the RSOH loss (m/z around 458) was also detected down to 0.1 µg spiked sample in the MS^2 spectrum. Both signature fragments are showing in high relative abundance in their corresponding MS^2 spectra, which could be readily selected for further MS^3 fragmentation and revelation of the peptide sequence. This study confirms that low amount of farnesyl peptide can be readily detected from a complex matrix background and identified by these characteristic signature fragments.

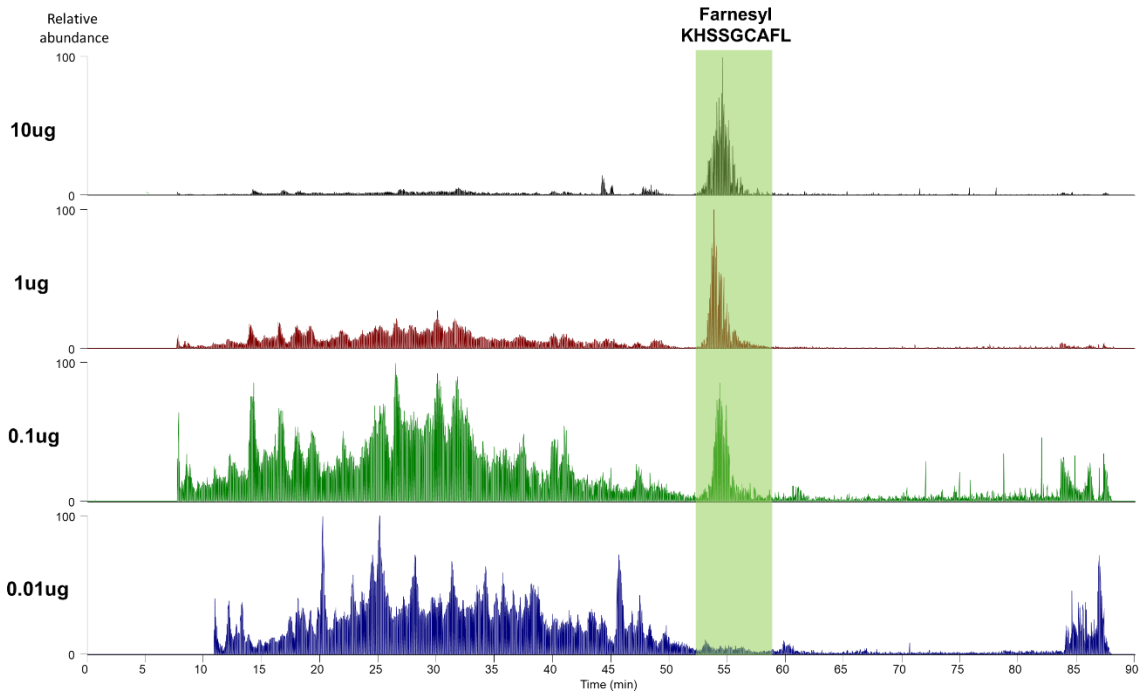


Figure 2-3. XICs (m/z 577.30 ± 0.5) of the spiked farnesyl KHSSGCAFL peptide in the raw macrophage digest

2.4.3 Advantages of applying NLMS3 strategy for large-scale prenylation study

Data-dependent NLMS3 strategy has been applied for profiling protein phosphorylation by monitoring the dominant product ion of 98 Da neutral loss ($-H_3PO_4$) in CID-MS^2 spectrum and subjecting it to MS^3 fragmentation for determining the modified locations and sequence information of phosphopeptides[79,80]. Both phosphorylation and prenylation are low-abundant PTMs, the difficulty of studying phosphopeptides are generally alleviated by various efficient enrichment techniques such as immobilized metal affinity chromatography[85] and TiO2 chromatography[86,87], whereas the methods for the enrichment and purification of native prenylated proteins are still under development. The high specificity of NLMS3 method to some degree can act as an enrichment strategy for targeting very minute quantity of prenylated peptides in the gas-phase. Furthermore, NLMS3 method can be innately more suitable for prenylation profiling than phosphorylation for the following reasons: (1) Phosphorylation occurs at serine, threonine and tyrosine residues which covers almost 16% of amino acids in current protein database[88], while as the only discovered prenylated site, cysteine has only 2.26% or lower occurrence in proteins[89]. The restriction of prenylated residue means that prenylation is less prone to mis-localization than phosphorylation when assigning prenylation sites during data interpretation. (2) Product ions of phosphorylation in serine, threonine and tyrosine are not created equally: predominant neutral losses vary depending on the modified amino acids: loss of H_3PO_4 is most prominent neutral loss for serine phosphopeptides, to a lesser extent in threonine phosphopeptides and least in tyrosine phosphopeptides whose loss can be exclusively HPO_3[88]. Significant proportion of phosphopeptides can be missed out if only the H_3PO_4 neutral loss is targeted in the MS^3 fragmentation. Although some alternative fragments are also observed (such as reduced charge loss or neutral loss of water/ammonia) in the NLMS3 analysis of prenylated peptide, the targeted signature fragments are consistently among the dominant fragments in the CID-MS^2 spectra, especially after the oxidation reaction[58,59]. (3) Another drawback of assigning phosphorylation sites by NLMS3 is the potential intramolecular rearrangement of phosphate group during the CID fragmentation, particularly when longer activation time is required for the fragmentation[90,91]. Competing fragmentation reaction of losing

HPO_3+H_2O also gives rise to a 98 Da neutral loss which further complicates the unambiguous identification and localization of phosphorylation, while identifying prenylation does not suffer from the same ambiguity related to the transfer of modification when assigning specific modified residue. (4) NLMS3 strategy offers confirmation on not only whether certain peptides have prenylation, but also the type of prenylation simply from the neutral loss a peptide precursor produces. This is generally difficult to achieve by other biochemical techniques. (5) The advantage of our NLMS3 strategy for prenylation over phosphorylation is that the identification of prenylation could potentially be achieved by two independent signature fragments generated from the same prenylated peptides through two individual yet parallel experiments. One signature fragment targets the prenyl neutral loss by analyzing prenylated peptides with NLMS3 method, whereas the other signature fragment targets RSOH neutral loss generated after the selective oxidation of prenyl group. If both signature fragments from the same prenylated protein can be confirmed, we can significantly reduce the false discovery matches and validate the authenticity of prenylated peptide, even when the prenylation identification solely depends on mass spectrometric analysis. Because neutral-loss-initiated MS^3 fragmentation only targets the highly abundant fragments in the MS^2, this method will be amenable to the discover of the prenylation with or without prior enrichment, taking advantage of high selectivity from the neutral loss scan mode without significantly compromising comprehensive protein identifications in regular shotgun proteomic experiment.

2.4.4 Profiling prenylome in complex sample by NLMS3 strategy

A NLMS3 strategy based on these two different signature fragments was designed for profiling native prenylation after the proof-of-concept study of spiked farnesyl peptide in the complex sample. NLMS3 strategy is applied for detecting the prenylation in the unoxidized and oxidized mouse raw macrophage tryptic digest, each with 0.1 μg farnesyl KHSSGCAFL peptide spiked. In the mass spectrometric method, the neutral losses were set to be R or RSOH losses (R = farnesylation or geranylgeranylation) as described in the method section including doubly and triply charged precursors. In addition, the tentative signature

fragments after the neutral losses need to be top 2 fragments in intensity within the MS^2 spectrum in order to be further fragmented. This ensures that the additional MS^3 acquisition is performed only on potential prenylated peptide ions which produces dominant signature fragments. For the data analysis, only the peptide spectrum matches (PSMs) in the MS^3 level are considered potential prenylated targets. From these MS^3 PSMs, the identification of prenylated peptides in the unoxidized samples will be matching against the peptide sequences without cysteinyl modification because the resultant signature fragment after neutral loss is equivalent to the peptide without prenylation; RSOH loss from the oxidized prenyl peptides is expected in the oxidized sample set, where the signature fragment has the mass that equals to the identified peptides losing SH_2 from the prenylation site. Therefore, loss of SH_2 was added as a variable modification of cysteine for identifying prenylated peptide in the oxidized samples by MS^3 sequencing. This modification of $-SH_2$ is also important for the localization of specific prenylation sites in cases where multiple cysteines are within the identified sequences. Furthermore, another validation of each MS^3 PSM was to compare the charge state of a potential prenylated peptide against the expected charge state of the neutral loss, where the mismatched results were considered false positives and discarded.

The spiked farnesyl KHSSGCAFL peptide was identified in both unoxidized and oxidized samples of all 3 raw macrophage biological replicates. Among all the replicates, the number of MS^3 PSMs of the spiked peptide is 17.6 ± 1.2 for the unoxidized samples and 17.3 ± 2.5 for the oxidized samples, which is rather consistent across the replicates. The range of PSM spectrum E-value is from 3.76e-09 to 1.26e-07 with the average of 2.91e-08 for the unoxidized farnesyl KHSSGCAFL peptides and from 5.04e-09 to 3.21e-07 with the average of 6.74e-08 for the oxidized ones. According to the spectrum E-value range in identifying this true prenylated peptide, the spectrum E-value cutoff for identifying the native prenylated peptides from the mouse raw macrophage cells was set to 1e-07 for both sample sets. After validating the neutral losses with the charge states, the identified native prenylated peptides from unoxidized and oxidized replicates were shown in Table 2-1 and Table 2-2, respectively. The C-terminus of nucleosome assembly protein 1-like-1 (NAP1L1) KDQNPAECK was identified to be farnesylated in both unoxidized and oxidized raw

macrophage sample sets. Tandem MS spectra of this peptide were shown in Figure 2-4. In the unoxidized sample, The MS^3 spectrum identified this C-terminus sequence confidently with the spectrum E-value of 8.51e-13. Precursor to this MS^3 PSM was the signature fragment (m/z 516.85) after the neutral loss of 102.16 corresponding to the farnesyl loss at +2 charge. Similar to the farnesyl KHSSGCAFL peptide, the singly charged fragment after the farnesyl loss was also observed; In the oxidized sample, the same peptide sequence with the $-SH_2$ modification on cysteine (C^*) was comprehensively confirmed with the spectrum E-value of 9.11e-12. The signature fragment was the most abundant fragment in the MS^2 spectrum with the RSOH neutral loss of 127.01 from the MS^1 precursor. Overall, 2 out of 3 unoxidized biological replicates and all 3 oxidized biological replicates identified the farnesylation of KDQNPAECK with the NLMS3 strategy. Another farnesyl peptide identified in high confidence (spectrum E-value of 1.21e-11) is the C-terminal sequence KEPSQPAECK from another nucleosome assembly protein 1-like-4 (NAP1L4), where the highly intense signature fragment (m/z 541.85) after RSOH neutral loss in the MS^2 spectrum and the confident coverage of the peptide sequence in the MS^3 spectrum were observed in the oxidized peptide (Figure 2-5). However, only 1 of 3 oxidized biological replicates reported this identification and the neutral loss counterpart of the same peptide in the unoxidized samples was not identified. Both NAP1L1 and NAP1L4 protein had been previously reported to be farnesylated in COS-1 cells[72] and EA.hy926 cells[92] with the C-terminal AAX unremoved. In total, 40 cysteine containing sequences after prenyl neutral losses were identified from the unoxidized samples and 22 cysteine containing sequences after RSOH neutral losses were identified from the oxidized samples. Besides the two farnesyl proteins mentioned above, the other identified peptides are internal sequences within the proteins. Although these internal peptides were not yet reported to be prenylated in the literature, they still provide us a potential list of prenylation proteins and could be beneficial in helping the future discovery of new prenylome.

#SpecFile	Precursor	Charge	Peptide	Protein	SpecEValue	EValue	Neutral loss
trypsin_RM_R3_prenyl.mzML	516.8497	2	K.KDQNPAECK.Q	sp|P28656|NP1L1_MOUSE	8.51E-13	1.61E-05	Far
trypsin_RM_R1_prenyl.mzML	516.8527	2	K.KDQNPAECK.Q	sp|P28656|NP1L1_MOUSE	4.02E-12	7.60E-05	Far
trypsin_RM_R3_prenyl.mzML	839.7902	3	R.TSDRRCSRSRDHKRSRSRDR.R	sp|Q5SUF2|LC7L3_MOUSE	1.24E-09	0.02436	Far
trypsin_RM_R3_prenyl.mzML	733.7039	2	R.LFLECDTHM+15.995WR.L	sp|Q811B1|XYLT1_MOUSE	2.69E-09	0.051492	Far
trypsin_RM_R2_prenyl.mzML	315.7417	2	K.PSPACR.N	sp|Q9EQS9|IGDC4_MOUSE	3.23E-09	0.057335	Far
trypsin_RM_R3_prenyl.mzML	485.8954	2	R.ACFKSIFR.I	sp|Q8BQS5|PAQR2_MOUSE	4.00E-09	0.075268	Far
trypsin_RM_R2_prenyl.mzML	791.1128	2	R.RTVRLRRRLPCR.L	sp|B1AXH1|NHSL2_MOUSE	5.19E-09	0.09968	Far
trypsin_RM_R2_prenyl.mzML	733.7535	2	R.LFLECDTHM+15.995WR.L	sp|Q811B1|XYLT1_MOUSE	5.45E-09	0.10411	Far
trypsin_RM_R1_prenyl.mzML	947.0651	3	K.FDSVVVNGFVCTKNIAHKKMNSC+57.021IK.N	sp|Q9Z1T6|FYV1_MOUSE	9.79E-09	0.19337	Far
trypsin_RM_R1_prenyl.mzML	713.1575	2	K.YLEM+15.995IYSMCKK.V	sp|Q9EPL8|IPO7_MOUSE	1.05E-08	0.20128	Far
trypsin_RM_R2_prenyl.mzML	485.8774	2	R.ACFKSIFR.I	sp|Q8BQS5|PAQR2_MOUSE	1.06E-08	0.20017	Far
trypsin_RM_R1_prenyl.mzML	485.8792	2	R.ACFKSIFR.I	sp|Q8BQS5|PAQR2_MOUSE	1.19E-08	0.22443	Far
trypsin_RM_R2_prenyl.mzML	634.8103	2	R.TCTLCKNSLMR.D	sp|P10820|PERF_MOUSE	1.29E-08	0.246	Far
trypsin_RM_R1_prenyl.mzML	592.4707	2	R.PASCVDLEPPR.I	sp|Q9R0M3|SRPX_MOUSE	2.01E-08	0.38449	Far
trypsin_RM_R1_prenyl.mzML	1041.882	3	K.LEFFDFTYDLNLCGLTEDPDLQVSAMK.H	sp|P97290|IC1_MOUSE	2.20E-08	0.43658	Far
trypsin_RM_R1_prenyl.mzML	1041.882	3	R.SGVGLGTKGGCASVIVSLTTQCPQDLTPYSGK.L	sp|Q6PDI5|ECM29_MOUSE	2.20E-08	0.43968	Far
trypsin_RM_R1_prenyl.mzML	745.6891	3	R.LVQSPNSYFMDVKC+57.021PGCYK.I	sp|Q6ZWU9|RS27_MOUSE	2.90E-08	0.56762	Far
trypsin_RM_R2_prenyl.mzML	585.8702	3	R.VIQQNCWDPEVRIR.E	sp|Q8R519|ACMSD_MOUSE	3.02E-08	0.58371	Far
trypsin_RM_R1_prenyl.mzML	597.4013	2	K.CKKCDGPC+57.021RK.V	sp|Q01279|EGFR_MOUSE	3.14E-08	0.59708	Far
trypsin_RM_R2_prenyl.mzML	782.5332	3	R.VLAAEAERSM+15.995LSPSGSCGPIKVK.T	sp|Q9ESZ8|GTF2I_MOUSE	3.65E-08	0.71829	Far
trypsin_RM_R1_prenyl.mzML	745.6008	3	K.RQAHLC+57.021VLASNCDEPMYVK.L	sp|P63323|RS12_MOUSE	3.72E-08	0.72705	Far
trypsin_RM_R2_prenyl.mzML	800.6904	2	K.HFKKPTYC+57.021NFCR.A	sp|Q91WG7|DGKG_MOUSE	3.76E-08	0.72224	Far
trypsin_RM_R2_prenyl.mzML	765.9774	2	R.RRRRCC+57.021RRRR.R	sp|P02319|HSP1_MOUSE	4.03E-08	0.76747	Far
trypsin_RM_R1_prenyl.mzML	890.3685	3	R.KDDRKKVDGKLLCWLCTLSYKR.V	sp|Q922G2|FA76A_MOUSE	4.23E-08	0.83195	Ger
trypsin_RM_R3_prenyl.mzML	853.5365	2	K.RELCKFYITGFCAR.A	sp|Q6ZPZ3|ZC3H4_MOUSE	4.24E-08	0.81921	Ger
trypsin_RM_R1_prenyl.mzML	868.7506	3	R.EGCPFC+57.021ADDRPC+57.021FVQEDKYLR.L	sp|Q8C419|GP158_MOUSE	4.50E-08	0.88269	Far
trypsin_RM_R1_prenyl.mzML	964.0487	3	-.M+15.995AAVETRVCETDGCSSEAKLQCPTCIK.L	sp|Q8BP48|MAP11_MOUSE	4.71E-08	0.93298	Far
trypsin_RM_R1_prenyl.mzML	636.4103	3	R.GFRRKCQALKPM+15.995YHR.V	sp|O54743|FOXF2_MOUSE	4.83E-08	0.93601	Ger
trypsin_RM_R1_prenyl.mzML	703.8911	3	K.PSCWNLVNDSVVMSLGEMK.G	sp|Q9QUR8|SEM7A_MOUSE	5.54E-08	1.083	Ger
trypsin_RM_R3_prenyl.mzML	724.0054	3	K.YLWTVCC+57.021HC+57.021GGKTKEAQK.I	sp|Q08460|KCMA1_MOUSE	5.58E-08	1.0879	Far
trypsin_RM_R2_prenyl.mzML	812.5187	2	K.CMSFTLNEQFM+15.995EK.Y	sp|Q9CQX5|CLDN1_MOUSE	5.61E-08	1.08	Far
trypsin_RM_R3_prenyl.mzML	745.6071	3	K.LSIQC+57.021YLRALDRCYAAYR.K	sp|P54869|HMCS2_MOUSE	6.01E-08	1.1718	Far
trypsin_RM_R3_prenyl.mzML	901.2025	2	K.EGQANLDPASCKSSHEK.R	sp|Q5DU28|PCX2_MOUSE	6.20E-08	1.2058	Ger
trypsin_RM_R1_prenyl.mzML	591.6504	2	R.ALCM+15.995VCGAEIR.S	sp|Q05AH6|SPNDC_MOUSE	6.37E-08	1.217	Far
trypsin_RM_R1_prenyl.mzML	963.7603	3	R.EILDLVCKRC+57.021FDELSPPCSEQHNR.E	sp|B2RX14|TUT4_MOUSE	6.63E-08	1.3063	Ger
trypsin_RM_R3_prenyl.mzML	895.6525	2	K.RCSRC+57.021QNVWYC+57.021CR.E	sp|Q99ML0|ZMY10_MOUSE	7.53E-08	1.4491	Ger
trypsin_RM_R1_prenyl.mzML	1066.066	2	K.KVMEECRRLQGEVQRLR.E	sp|Q9QY76|VAPB_MOUSE	7.90E-08	1.5374	Ger
trypsin_RM_R1_prenyl.mzML	788.022	3	R.M+15.995ECGAVIMNGCIYVTGGYSYSK.G	sp|Q6GQU2|KLH23_MOUSE	8.20E-08	1.6123	Far
trypsin_RM_R1_prenyl.mzML	1046.969	3	R.LPDPQILKSDTPC+57.021DDFTRC+57.021PTNNTCCK.L	sp|P28798|GRN_MOUSE	8.21E-08	1.626	Ger
trypsin_RM_R2_prenyl.mzML	475.5837	2	R.NNCWKER.Q	sp|Q149L6|DJB14_MOUSE	8.90E-08	1.6536	Far
trypsin_RM_R1_prenyl.mzML	750.0523	3	R.CDLRENLWLNLTDGSVLCGK.W	sp|Q5BKP2|UBP13_MOUSE	9.28E-08	1.817	Far
trypsin_RM_R3_prenyl.mzML	1050.787	3	R.CIHVVNKAILSCYRWFDKKHKNC+57.021MR.R	sp|Q059Y8|DCST1_MOUSE	9.51E-08	1.8773	Far
trypsin_RM_R1_prenyl.mzML	901.1918	2	R.VRESSCLALNDLLRGR.P	sp|Q6PDI5|ECM29_MOUSE	9.77E-08	1.8958	Ger
trypsin_RM_R2_prenyl.mzML	736.0808	2	K.AYGEACSGIRC+57.021QR.H	sp|O35118|GFRA3_MOUSE	9.89E-08	1.9045	Far

Table 2-1. Potential prenyl peptides identified from 3 unoxidized mouse raw macrophage samples

#SpecFile	Precursor	Charge	Peptide	Protein	SpecEValue	EValue	Neutral loss
trypsin_RM_R1_oxi.mzML	499.9052	2	K.KDQNPAEC-33.988K.Q	sp|P28656|NP1L1_MOUSE	9.11E-12	1.73E-04	FarSHO
trypsin_RM_R1_oxi.mzML	541.846	2	K.KEPSQPAEC-33.988K.Q	tr|B7ZNL2|B7ZNL2_MOUSE	1.21E-11	2.30E-04	FarSHO
trypsin_RM_R2_oxi.mzML	499.8549	2	K.KDQNPAEC-33.988K.Q	sp|P28656|NP1L1_MOUSE	1.26E-10	0.002394	FarSHO
trypsin_RM_R1_oxi.mzML	1012.676	2	K.TWRWGPGGSGAILLVNC-33.988DR.D	sp|Q9Z185|PADI1_MOUSE	1.09E-08	0.2136	FarSHO
trypsin_RM_R1_oxi.mzML	648.3754	3	K.NCLERC-33.988LDHAPYCPLCK.E	sp|D3YY23|LONF1_MOUSE	1.11E-08	0.21655	FarSHO
trypsin_RM_R3_oxi.mzML	879.2972	2	K.C+57.021IMRPLSKDAC-33.988SRVR.S	sp|Q8BHP7|OLM2A_MOUSE	1.46E-08	0.28253	GerSHO
trypsin_RM_R1_oxi.mzML	492.7716	3	K.MGGSFLIC-33.988SKLKAR.F	sp|Q60936|COQ8A_MOUSE	1.67E-08	0.32182	FarSHO
trypsin_RM_R3_oxi.mzML	499.8488	2	K.KDQNPAEC-33.988K.Q	sp|P28656|NP1L1_MOUSE	2.76E-08	0.52329	FarSHO
trypsin_RM_R3_oxi.mzML	644.9882	3	R.GDIM+15.995C-33.988YYTLTEKFIPR.R	sp|A2A6M5|CACO2_MOUSE	2.78E-08	0.53882	GerSHO
trypsin_RM_R1_oxi.mzML	612.2222	3	K.NASQGPASECRFFC-33.988VPR.D	sp|Q8R507|FKTN_MOUSE	3.76E-08	0.73172	FarSHO
trypsin_RM_R1_oxi.mzML	917.4702	3	R.EC-33.988GWGFTQKSDLIQHQRTHTREK.-	sp|Q96EQ9|PRDM9_MOUSE	3.80E-08	0.74778	FarSHO
trypsin_RM_R3_oxi.mzML	583.8335	3	K.HKVC-33.988MDLRANLKQVK.K	sp|P13412|TNNI2_MOUSE	4.37E-08	0.8461	GerSHO
trypsin_RM_R1_oxi.mzML	314.3017	3	R.VAC+57.021C-33.988EHQK.A	sp|Q0KK56|F184B_MOUSE	4.62E-08	0.86821	FarSHO
trypsin_RM_R1_oxi.mzML	608.4167	2	K.FRGICFLC-33.988YK.T	sp|Q9JLF7|TLR5_MOUSE	4.94E-08	0.93953	FarSHO
trypsin_RM_R2_oxi.mzML	460.7818	3	R.C+57.021AATGIC-33.988CSPDGCR.T	sp|P35454|NEU1_MOUSE	5.03E-08	0.97093	GerSHO
trypsin_RM_R2_oxi.mzML	460.7818	3	R.SSWVC-33.988DGDNDC+57.021R.D	sp|O88307|SORL_MOUSE	5.03E-08	0.96477	GerSHO
trypsin_RM_R3_oxi.mzML	761.6338	3	K.HC-33.988DDSYTPRDKLWQQLRR.G	sp|P36552|HEM6_MOUSE	5.06E-08	0.98638	FarSHO
trypsin_RM_R1_oxi.mzML	932.7505	3	K.KLIC-33.988YC+57.021RIRGCKRRERVFGTC+57.021R.N	sp|P28312|DEFA5_MOUSE	5.80E-08	1.1396	FarSHO
trypsin_RM_R2_oxi.mzML	1084.569	3	R.RSSVKKIEKM+15.995KEKRQAKYSENKLKC-33.988TK.A	sp|Q812A2|SRGP3_MOUSE	5.96E-08	1.1798	FarSHO
trypsin_RM_R2_oxi.mzML	608.3958	2	K.FRGICFLC-33.988YK.T	sp|Q9JLF7|TLR5_MOUSE	7.31E-08	1.3921	FarSHO
trypsin_RM_R2_oxi.mzML	1044.008	3	M.VEIKKIC-33.988CIGAGYVGGPTCSVIAHM+15.995CPEIR.V	sp|O70475|UGDH_MOUSE	8.00E-08	1.5906	GerSHO
trypsin_RM_R2_oxi.mzML	933.5905	3	R.RHQRTHTGPPTPPC-33.988PTC+57.021GFRCCAPR.P	sp|Q6PD29|ZN513_MOUSE	8.82E-08	1.7423	GerSHO
trypsin_RM_R2_oxi.mzML	933.5905	3	R.TGYGGTSCIQIPVLC-33.988LLVNGDPNTLER.I	sp|Q9JJH7|TRPM5_MOUSE	8.82E-08	1.7476	GerSHO
trypsin_RM_R3_oxi.mzML	855.1259	3	K.GITC-33.988QRVTETLAAWNDHRPIC+57.021R.A	sp|Q923L3|CSMD1_MOUSE	9.43E-08	1.8538	FarSHO
trypsin_RM_R3_oxi.mzML	612.0981	3	R.DRTGRC-33.988CKGYAWISPR.K	sp|Q9JMB1|THEG_MOUSE	9.50E-08	1.8449	FarSHO

Table 2-2. Potential prenyl peptides identified from 3 oxidized mouse raw macrophage samples

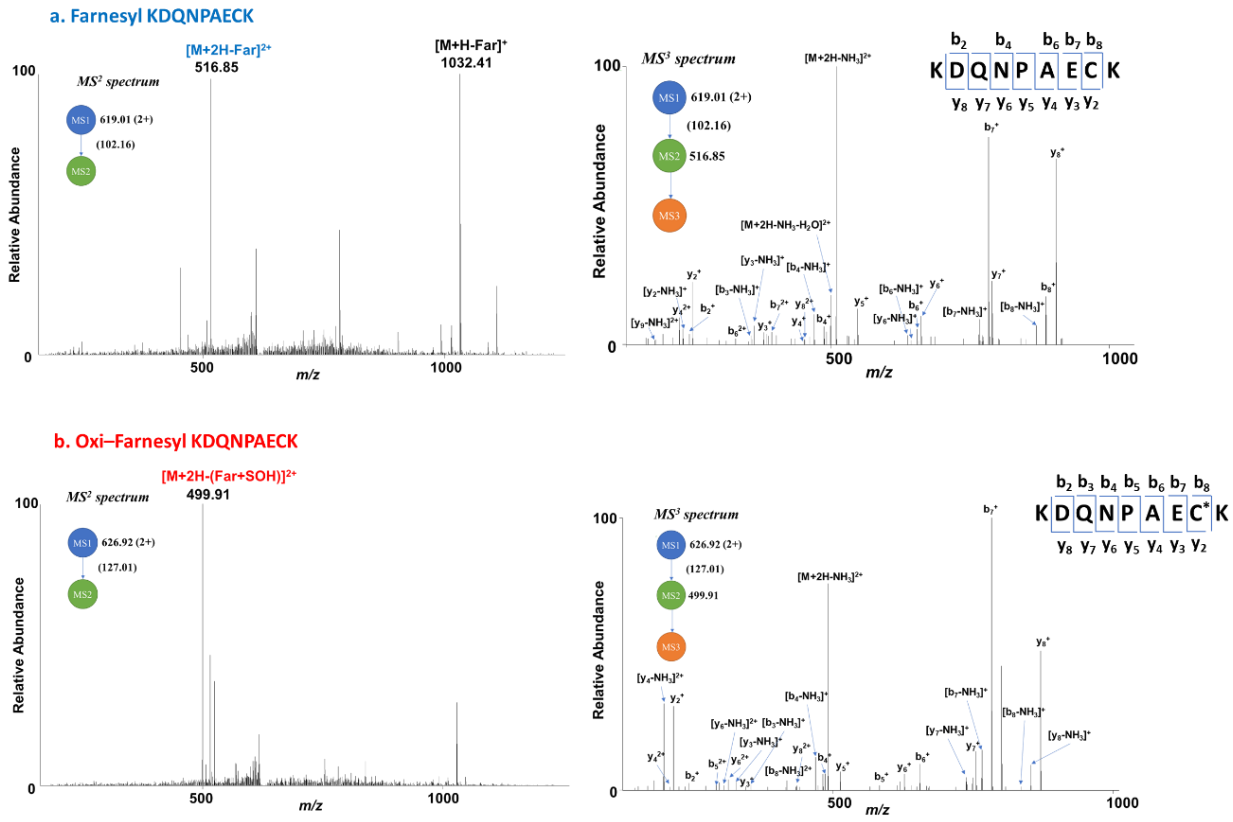


Figure 2-4. MS^2 and MS^3 spectra of native farnesyl KDQNPAECK identified from mouse raw macrophage cells

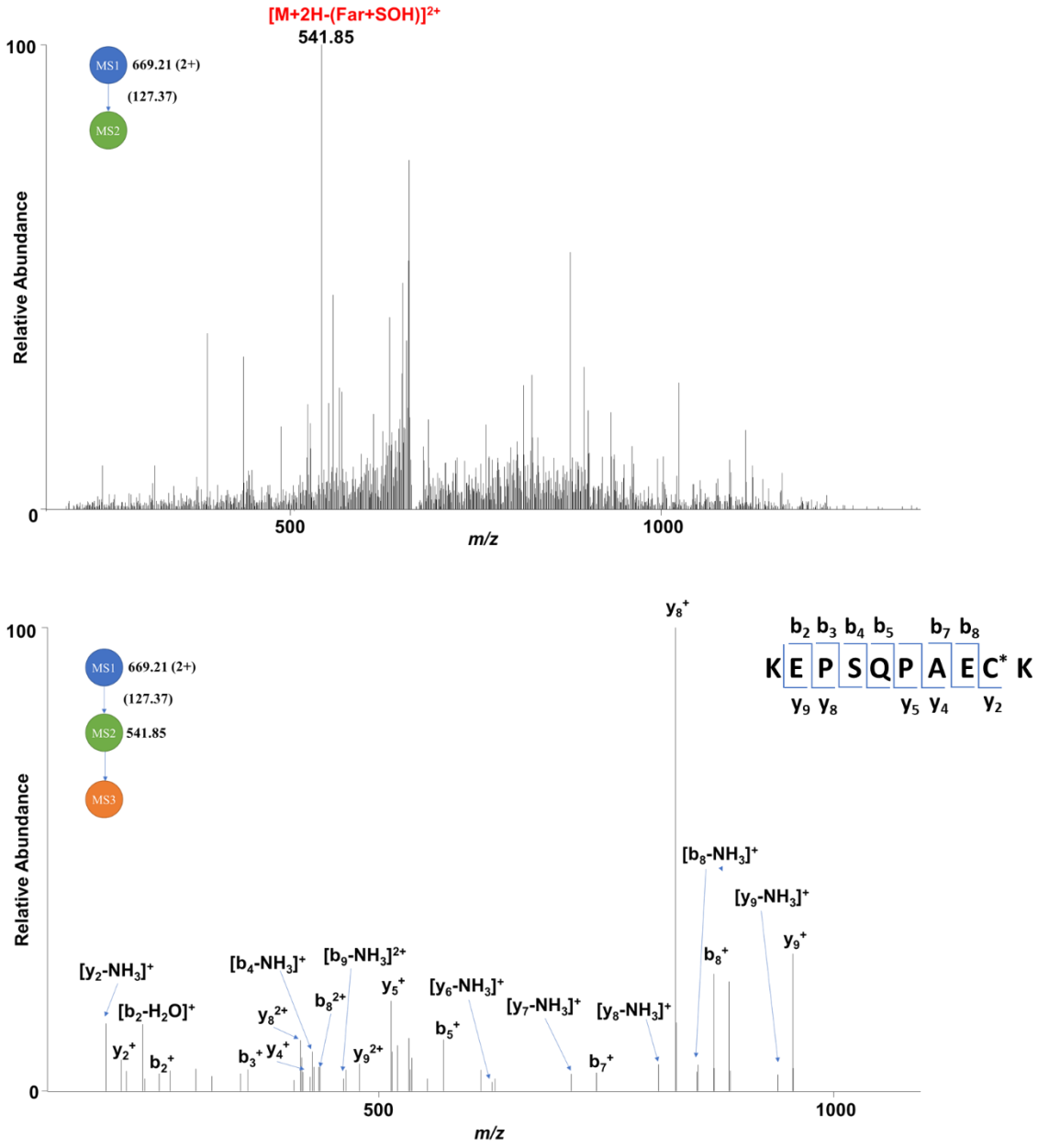


Figure 2-5. MS^2 and MS^3 spectra of native farnesyl KEPSQPAECK identified from oxidized mouse raw macrophage cells

2.5 Conclusion

In this study, we explore the potential of applying a neutral loss-based strategy for the identification of an important PTM: prenylation. Large-scale investigation of phosphorylation had been carried out successfully by NLMS3, though some shortcomings affect its widespread applications. NLMS3 strategy for studying prenylation does not encounter the similar limitations because of more stable gas-phase chemistry of the prenyl groups and less propensity to misidentification. What truly makes our NLMS3 strategy for prenylation discovery unique comparing to NLMS3 application in phosphorylation is the incorporation of two independent signature fragments to greatly improve the confidence in identifying the prenylation type, prenylation site and prenylation sequence. To the best of our knowledge, this is the first time C-terminus of natively prenylated proteins are identified from whole cell lysate by mass spectrometry without prior enrichment. With the promising preliminary results from large-scale prenylation study by NLMS3, we are working on addressing the following aspects to improve the number and reproducibility of prenylated proteins identified: (1) C-terminal sequences generated by trypsin digestion are mostly singly charged which are commonly excluded for tandem mass analysis, alternative enzymes such as GluC or chymotrypsin may produce more identifiable C-terminus; (2) Some prenylated sequences have rather extreme hydrophobicity (such as double geranylgeranylations) which can be challenging to elute from the C18 stationary phases even after the oxidation; (3) Although we show that this method can detect prenylated proteins without prior enrichment, we are still working on developing an efficient enrichment method specifically targeting prenylated proteins as this will ultimately be essential for identifying more novel native prenylome from large-scale biological samples; (4) With the capability to detect native prenylated proteins, quantitation of this important protein modification by isotopic labeling becomes possible. We believe that the ease of incorporating this NLMS3 strategy into MS-based prenylation workflows and its ability to discover new native protein prenylation are beneficial for the next breakthrough in global prenylome research.

Chapter 3

Targeted Labeling Strategy for the Enrichment and Fragmentation Study of Prenylated C-terminal Peptides

3.1 Abstract

Protein prenylation plays pivotal roles in the transforming activities of various oncogenic proteins. Therefore, it is being intensively investigated as targets for potential anticancer drugs. However, only a few mass spectrometry-based methods were reported for the direct detection of protein prenylation because of its high hydrophobicity, low abundance, and high propensity of neutral loss during collision-induced dissociation (CID). To develop an enrichment strategy for this low-abundant post-translational modification (PTM) and facilitate the MS fragmentation study of this labile PTM, we develop and evaluate the strategy for the enrichment of prenylated peptides by directly derivatizing the prenyl group with biorthogonal biotinylated reagent, followed by the fragmentation study of the labeled peptides by tandem MS fragmentations. Proof-of-concept study was performed on the labeling and identification of customized peptides that consist of CAAX motifs mimicking the consensus sequence of native prenylation. The strategy was also applied to the enrichment and identification of farnesylated polybasic C-terminal sequence of KRas4B, a Ras isoform that commonly found mutated in cancers but has been challenging in its confident identification by mass spectrometry primarily because of its polybasic hypervariable nature.

3.2 Introduction

Prenylation is an irreversible lipid post-translational modification (PTM) consisting of either three or four isoprene units to attach to the cysteine residue near the C-termini of proteins, termed farnesylation or geranylgeranylation, respectively. Most farnesylated proteins contain C-terminal CAAX consensus sequence (C = Cysteine, A = aliphatic amino acids, X = any amino acid). It was quantitatively determined that around 2% of all proteins, potentially all proteins with C-terminal cysteine, are prone to prenylation[93]. Though hundreds of proteins have been experimentally examined to be prenylated, among which are small GTPases such as Ras, Rho, Rab proteins, the discovery of new prenylated protein substrates is still an

ongoing task. Prenylation not only plays important biological roles in protein trafficking through a series of protein interaction pathways and signaling cascades in normal cells, but also mediates the cellular survival, proliferation and metastasis associated with various cancers, inflammatory diseases and premature disorders[7]. Farnesyltransferase (FTase) and geranylgeranyltransferase (GGTase) inhibitors are being developed as potential therapeutic drugs that directly target the prenylation process and transduction pathways[25,94]. GTPases Ras isoforms (HRas, NRas and KRas) are commonly observed to be mutated in multiple malignant cancers of pancreas, colon and lung, varying from 30%-90% depending on the specific tumor type[95]. Unlike other Ras isoforms that depend on dual lipidations (prenylation plus palmitoylation) for stable membrane localization, KRas4B isoform harbors a polybasic hypervariable region (HVR) to associate electrostatically with negatively-charged region of the cell membrane[96,97]. This distinct structural and interactive heterogeneity of KRas4B enables its unique binding with calmodulin and activation of alternative proliferative pathway, which was suspected to be a key factor in the more frequent occurrence of KRas4B mutations in various cancer types[98].

Nearly all protein PTMs are presented in low stoichiometric quantity compared to unmodified counterparts. Selective and specific study of these low-abundant PTMs require enrichment or purification techniques that efficiently target the modification itself before downstream analysis. The traditional method to successfully evaluate and identify prenylated proteins is the radioactive labeling of cells with ^{3}H-mevalonic acid that incorporates ^{3}H-substituted prenyl diphosphates into the prenylated protein substrates, which were later detected by western blot or autoradiography[99]. Although this method has been widely applied for prenylation identification, it can suffer from radioactive hazard, low throughput and inability for global prenylation study[25]. Global detection of farnesylation later was achieved by tagging-via-substrate approach which metabolically incorporates synthetic farnesyl analogues to facilitate the biotinylation of the potential farnesylated proteins[72]. This approach shows better sensitivity than radiolabeling method and stimulates a myriad of researches to develop various analogues and probes with different functionalities such as fluorophore, azido or alkynyl reactive groups and affinity tags to accommodate different enrichment and

detection measures, including the proteomic analysis of enriched prenylated proteins by MS[25,73,100,101]. New improvements and optimizations from the original design such as utilizing dual probe tagging strategy was reported recently[92]. Nonetheless, this innovative strategy has certain limitations such as unknown perturbations to the endogenous biological system when inhibiting natural prenyl biphosphate, *in vivo* metabolism of these analogs into the formation of non-native membrane and other lipids in cell, as well as challenges for qualitatively and quantitatively studying natural prenylome in biological systems, especially for tissue samples[102–104].

Rapid advancements and improvements in LC-MS instrumentations have greatly propelled MS-based proteomics for analyzing samples with increasing complexity, which can be better applied to the targeted study of low abundant PTMs like prenylation. Direct fragmentation studies of prenylated peptides revealed that neutral loss of farnesyl or geranylgeranyl moiety can be used for the identification of prenylated peptides[75,76], even in the complex HeLa cell lysate without prior enrichment[77]. Further improvement to produce more hydrophilic, identifiable prenyl derivatives was achieved by subjecting the prenyl groups to chemical oxidation or epoxidation, which generates more consistent neutral loss of RSOH (R is the modified farnesyl chain; S is from cysteine on the modified site; O is oxygen; H is hydrogen) upon CID fragmentation[58]. Identities of prenyl peptides were further confirmed complementarily by alternative ETD fragmentation[59]. The epoxy derivatives of unsaturated prenyl group from the epoxidation reaction open a promising opportunity for directly and specifically targeting the modification for enrichment. To continue the investigation of prenylation in tandem MS following these previous findings, our study focuses on the enrichment aspect by the targeted labeling of prenylation after epoxidation and detection by signature fragment of the derivatized epoxidized prenylated peptides, a new systematic MS-based approach for the confident identification of protein prenylation.

3.3 Materials and Methods

3.3.1 Materials

Cysteine containing peptides REKKFFCAIL, KHSSGCAFL and GKKKKKKSKTKC were customized and synthesized by Genscript Corp (Piscataway Township, NJ). Bovine serum albumin (BSA), myoglobin from equine heart, ubiquitin from bovine enrythrocytes, *trans, trans*-farnesyl bromide, meta-chloroperoxybenzoic acid (mCPBA), iodoacetamide, ammonium bicarbonate (ABC), dimethyl sulfoxide (DMSO), biotin maleimide (BM), biotin hydrazide (BH) and trifluoroacetic acid (TFA) were purchased from Sigma Aldrich (St. Louis, MO). Pierce monomeric avidin agarose was purchased from Thermo Scientific (Waltham, MA). Hydrogen peroxide (30%), 7 M ammonia in methanol and dichloromethane (DCM) were obtained from Alfa Aesar (Tewksbury, MA). Acetonitrile (ACN) was acquired from VWR (Radnor, PA). Dithiothreitol (DTT) was from Biorad (Hercules, CA), and protease used in the study was trypsin (Promega, Madison, WI).

3.3.2 Prenylation and derivatization of peptides

Synthesis of epoxidized farnesylated (epoxy-far) peptides and protein adopts the method as previously described[58,59]. Briefly, individual peptide was mixed with excess of *trans, trans*-farnesyl bromide in 7 M ammonia in methanol under 4 °C in dark for farnesylation. Solvent was removed by speed vacuum concentration and the farnesylated peptide was subjected to epoxidation with mCPBA in DCM.

Biotinylation of epoxidized samples was achieved with either BM or BH. BM and BH were prepared fresh right before use by dissolving with minimum amount of DMSO. For BM labeling, the epoxy-far peptide was first mixed with 2 mM DTT solution in 50 mM ABC solution and then reacted with 4 mM BM in PBS (pH 7.4); For BH labeling, the epoxidized peptide samples were directly reacted with 5 mM BH in PBS (pH 5.0). The samples were reconstituted in 0.1% formic acid for LC-MS analysis.

3.3.3 Enrichment study of the prenylated peptides from peptide digest

To demonstrate the selective enrichment of farnesylated GKKKKKKSKTKC peptide by this strategy, the peptide was spiked into protein digest of BSA, myoglobin and ubiquitin in 1:1:1:1 molar ratio. The proteins were digested as previous procedures[50]. The combined sample underwent epoxidation and biotinylation as described above. The sample was then reconstituted in PBS (pH 7.4) and mixing with monomeric avidin beads for 2 h under room temperature. After peptide binding, the beads were washed extensively and sequentially with PBS and ultrapure water to remove the non-binding peptides. After the washing steps, the beads were incubated with the elution buffer ACN/H_2O/TFA (50/50/0.4, v/v/v) for 1 h at room temperature. Supernatant was separated from the beads by centrifugation, which was then evaporated, desalted, and reconstituted in 0.1% formic acid for LC-MS analysis.

3.3.4 Mass spectrometric analysis

The samples were analyzed in a UHPLC nano system (Dionex UltiMate 3000) coupled with Velos Pro mass spectrometer (Thermo Fisher Scientific, Waltham, MA). Acclaim PepMap C18 column (150 mm × 75 μm, 3 μm) was utilized for separation. Source voltage was set to 2.00 kV and capillary temperature was set to 275 °C. MS scan was from 200 to 2000 *m*/*z* mass range. Data-dependent CID-MS^2 fragmentation was performed using CID activation of 35.0% normalized collision energy for the 10 most abundant precursor ions in each MS scan with charge state ≥2 at the isolated width of 1.5 Da, activation Q of 0.25, and activation time of 30 ms. For MS^3 acquisition, data was collected from the top 3 most abundant precursor ions in each MS^2 scan by CID activation of 45.0% normalized collision energy, activation Q of 0.25, and activation time of 30 ms. In alternative ETD fragmentation the reaction time was set to 120 ms.

3.4 Results and discussions

3.4.1 Biotinylation and identification of the farnesylated peptides

Three peptides were selected to investigate the enrichment and fragmentation of C-terminal farnesylated peptides. REKKFFCAIL and KHSSGCAFL peptides contain CAAX motif that simulate the consensus sequence of commonly prenylated proteins. GKKKKKKSKTKC peptide is the polybasic C-terminal sequence of KRas4B after the CAAX motif was processed by Ras-converting CAAX endopeptidase 1 (RCE1) which removes the AAX residues[24]. The peptides were first farnesylated and epoxidized according previous procedures, where the conversion of prenyl thioether linkage to sulfoxide linkage was observed[58,59]. Prenyl-specific reaction was achieved by the targeted biotinylation of epoxy-far peptides through chemical derivatization of epoxy groups formed on the unsaturated sites of the prenyl groups. This means that at least one epoxy site must be present for the prenylated peptides to be labeled through the attack of biotinylated nucleophiles on the epoxy groups. Multiple epoxidized products were generated in various abundances after the reaction of peptides with mCPBA oxidant, with the observation that the doubly-oxidized product being the most abundant product among all oxidized products. Two biotinylation strategies were established for the enrichment of farnesylated peptide after epoxidation as shown in Figure 3-1. Initially the two-step reaction of epoxy-far REKKFFCAIL peptide with DTT and then BM was evaluated. In the first step one thiol end of the DTT opens the epoxy ring for attachment, while the other thiol end remains reactive for the further attachment of BM through thiol-maleimide Michael coupling reaction (Figure 3-1a). The labeled peptide was identified by MS analysis where precursors were subjected to MS^2 and MS^3 multistage tandem fragmentation.

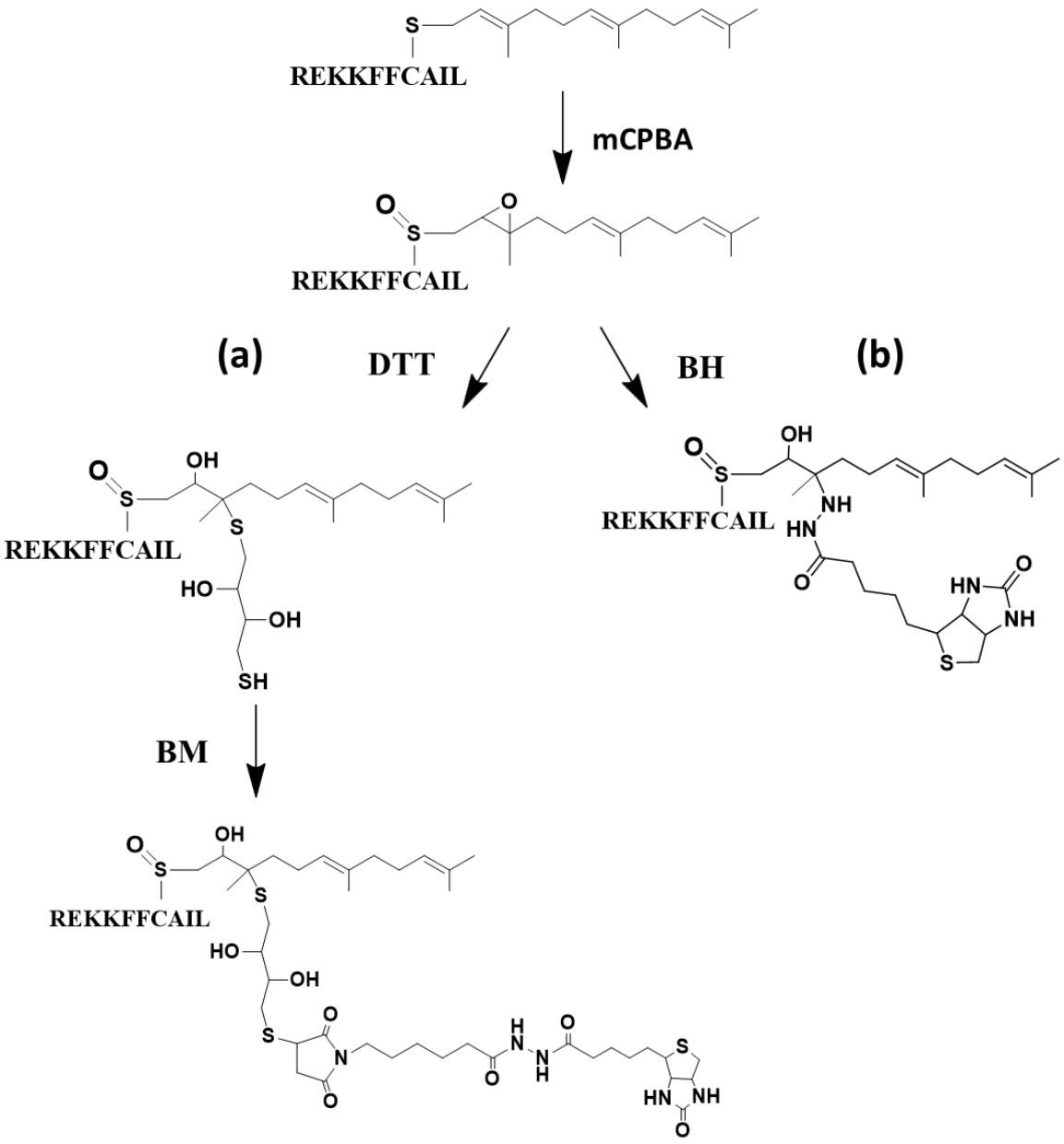


Figure 3-1. Biotinylation strategies for the enrichment of epoxy-Far REKKFFCAIL (a) Derivatize with DTT and BM; (b) Derivatize with BH.

The BM labeled REKKFFCAIL peptide precursor $[M+Far+DTT+2O+BM]^{3+}$ (m/z 699.62) was fragmented and gave rise to a highly intense fragment m/z 611.03^{2+} after the signature loss of RSOH (Figure 3-2a). This signature fragment is the same m/z as the fragmentation of epoxidized REKKFFCAIL after the removal of derivatized cysteinyl side chain, which converts the modified cysteine group to dehydroalanine residue by β-elimination[54]. This is in accordance with the previous observed signature loss in the epoxidized peptide[58,59], which confirms that the attachment of the extended biotin group does not change the preferential loss of RSOH group in the gas-phase fragmentation. However, it is important to mention that the biotin group is readily protonated in the ionization process and thus the signature fragment will be in reduced charge state after the loss of charged biotinylated side chain, instead of the neutral loss as in the case of oxidized or epoxidized prenyl peptides. Furthermore, this signature fragment in the MS^2 was subjected to MS^3 fragmentation to generate the b and y ions of the sequence for peptide identification, where good coverage of the peptide sequence was observed (Figure 3-2b). Most MS^3 fragments of REKKFFCAIL peptide are b ions which can be explained by the basic residues near the N-terminal side of the peptide.

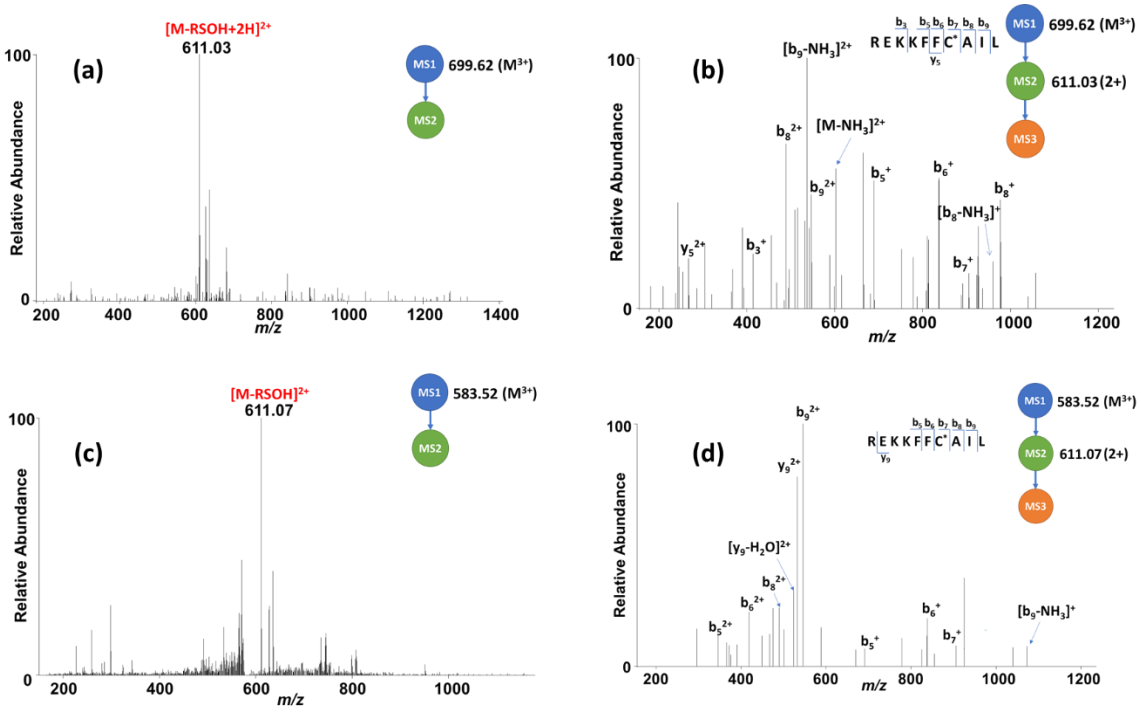


Figure 3-2. Tandem MS analysis of the derivatized epoxy-far REKKFFCAIL peptide

(a) MS^2 spectrum after DTT and BM reaction; (b) MS^3 spectrum of signature fragment after DTT and BM reaction; (c) MS^2 spectrum after BH reaction; (d) MS^3 spectrum of signature fragment after BH reaction.

The second derivatization strategy was performed by reacting the same epoxidized peptide directly with BH, another nucleophile for targeted biotinylation (Figure 3-1b). Similarly, the BH labeled precursor $[M+Far+2O+BH]^{3+}$ (m/z 583.52) was also producing high-abundant signature fragment 611.07^{2+} after the loss of RSOH (Figure 3-2c), as well as the CID fragmentation pattern of prevalent b ions in the MS^3 spectrum of the signature fragment for peptide sequencing (Figure 3-2d). For both labeling experiments, the epoxy-far REKKFFCAIL peptide was successfully biotinylated, as the RSOH signature loss still applies to the peptide after the biotinylation, indicating that the same RSOH strategy can be used for targeted enrichment and identification of the prenylated peptides, along with the acquisition of substrate

sequence information by multistage MS^3 fragmentation. In comparison of the two biotinylation strategies, one extra step of reacting with DTT is required for biotinylation labeling with BM which can result in lower yield in the final biotinylated peptide. The excess DTT reagent can also compete for the reaction with BM. In order to prevent the sample loss in this extra labeling step, BH was selected as the biotinylated reagent for studying another farnesylated peptide KHSSGCAFL, where the labeling product with BH was identified in CID by the high-abundant signature fragment 481.54^{2+} after the RSOH loss as well (Figure 3-S1).

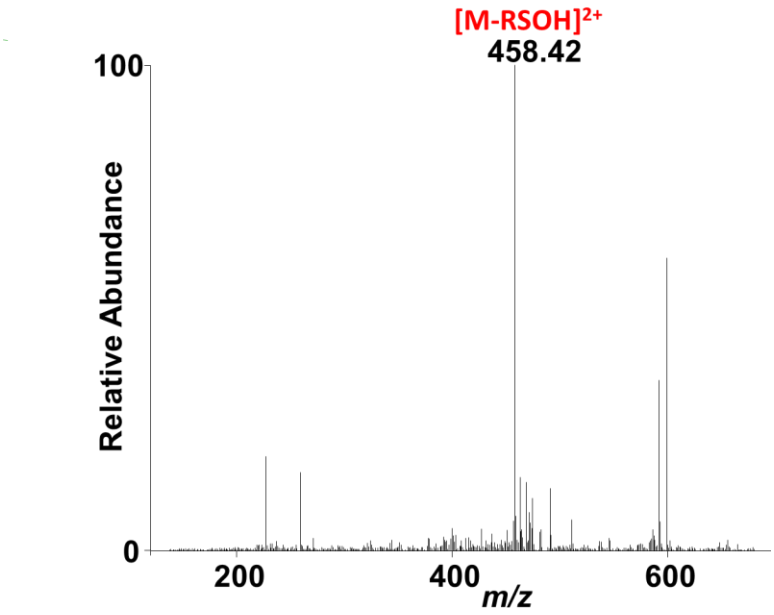


Figure 3-S1. MS^2 analysis of the derivatized epoxy-far KHSSGCAFL peptide

The polybasic C-terminal sequence of KRas4B GKKKKKKSKTKC was also experimented by the BH labeling strategy and analyzed with tandem MS fragmentation. After the reaction, the peptide precursor observed was $[M+Far+2O+BH]^{4+}$ (m/z 472.00). Upon the CID-MS/MS fragmentation, a dominant fragment 453.29^{3+} was identified to match the signature fragment in reduced charge state after the RSOH

loss (Figure 3-3a). The MS^3 spectrum of 453.29^{3+} exhibits a decent coverage of peptide backbone but most of the peptide fragment ions are in low intensity (Figure 3-3b), presumably because the highly charged sequence affect the efficient fragmentation in CID[105]. Therefore, for the complementary confirmation of the peptide sequence, ETD fragmentation was also performed on the $[M+Far+2O+BH]^{4+}$ precursor and resulted in higher fragmentation efficiency of the same peptide sequence (Figure 3-3c), showing the successful biotinylation of KRas4B C-terminal peptide confirmed by complementary CID and ETD fragmentation.

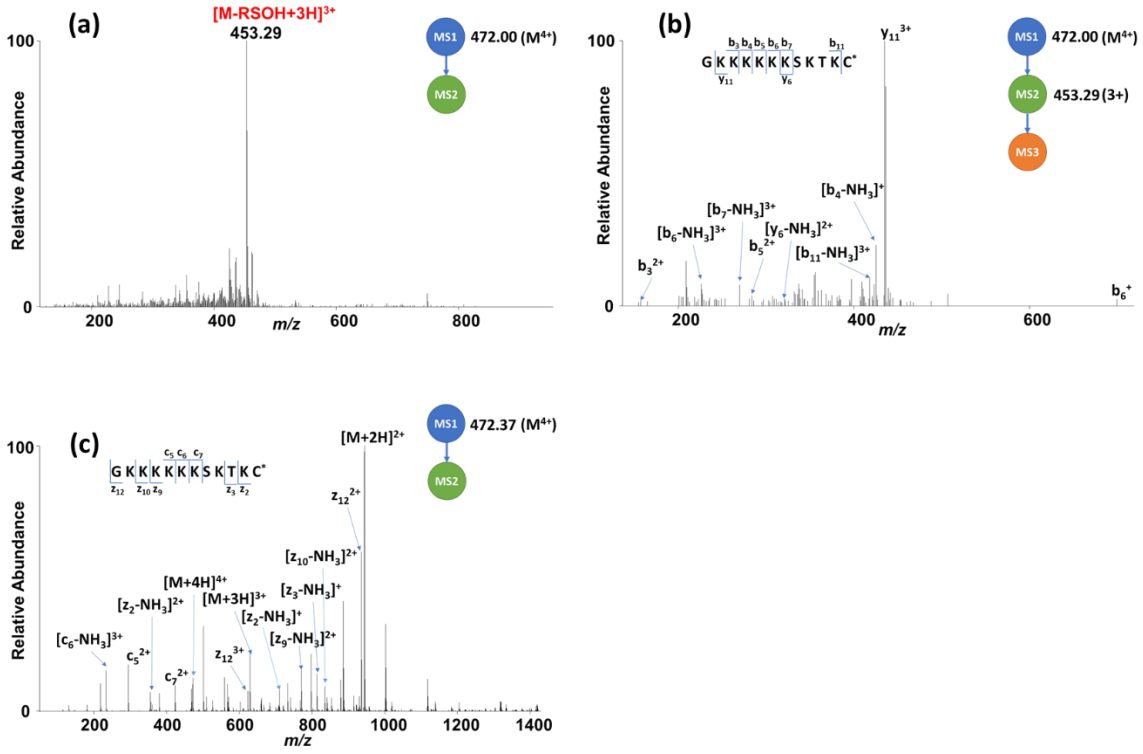


Figure 3-3. Tandem MS analysis of the epoxy-far GKKKKKKSKTKC peptide after BH derivatization

(a) CID-MS^2 spectrum of the peptide precursor; (b) CID-MS^3 spectrum of the signature fragment; (c) ETD-MS^2 spectrum of the peptide precursor

3.4.2 Enrichment of KRas4B C-terminus by targeting peptide farnesylation

As a proof-of-concept study, KRas4B C-terminal peptide was investigated to be enriched from the tryptic digest of protein standards by the BH labeling strategy. The farnesylated GKKKKKKSKTKC peptide was spiked into the digested peptide pool of BSA, myoglobin and ubiquitin. The spiked sample was then epoxidized and biotinylated, followed by the enrichment of biotin-tagged peptides with monomeric avidin beads. The same sample after the epoxidation and biotinylation but without enrichment was analyzed as the control sample. The extracted ion chromatograms (XICs) of control and the enriched samples (aliquoted right before enrichment steps) are shown in Figure 3-4. The extracted m/z range was 1 Da window of the GKKKKKKSKTKC peptide precursor $[M+Far+2O+BH]^{4+}$ (m/z 472.30) and we observed the emergence of peaks between 58 to 62 min in the enriched sample compared to the control sample. These peaks were identified to be the biotinylated GKKKKKKSKTKC peptide according to the analysis of tandem mass spectra which also corresponded to the retention time of the precursor that generated the signature loss fragment. In the MS^2 spectrum of this enriched peptide, precursor ion $[M+Far+2O+BH]^{4+}$ (m/z 472.04) generated signature fragment of 453.28^{3+} in highest intensity (Figure 3-S2a). Further selection and fragmentation of the signature fragment by MS^3 matched the previous CID pattern of the biotinylated GKKKKKKSKTKC peptide in the peptide labeling experiment (Figure 3-S2b). The peptide fragments identified in ETD-MS^2 of the precursor ion also matched consistently with the above annotated results of the peptide labeling (Figure 3-S2c). These results exhibit that this biotinylation enrichment strategy can specifically target the farnesyl peptide and facilitate the selective identification of this modification with great confidence.

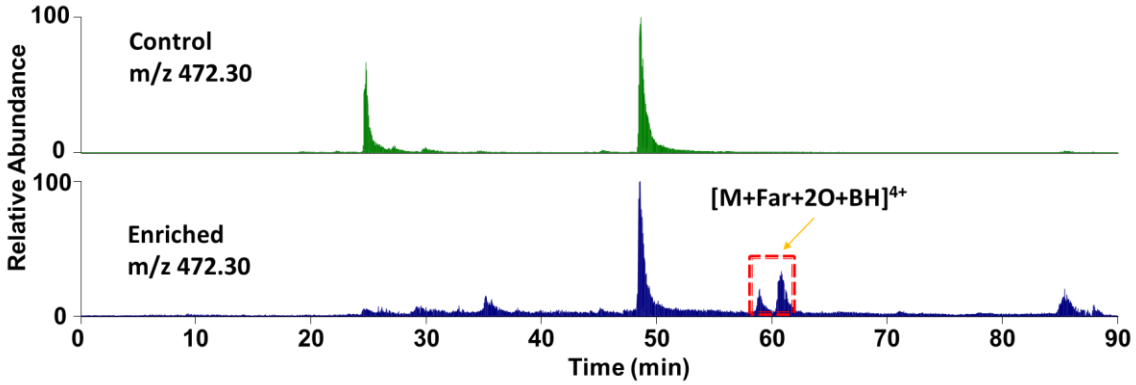


Figure 3-4. XICs of control (non-enriched) and enriched epoxy-far GKKKKKKSKTKC peptide after BH derivatization

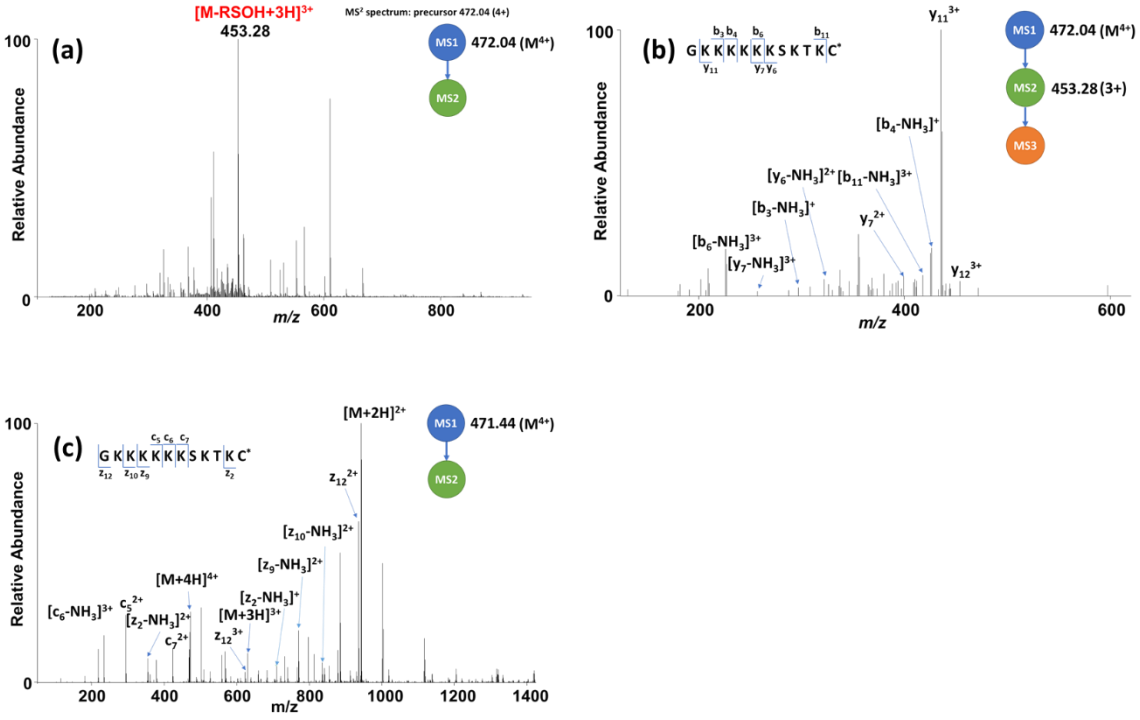


Figure 3-S2. Tandem MS analysis of the derivatized epoxy-far GKKKKKKSKTKC peptide after targeted enrichment (a) CID-MS^2 spectrum of the peptide precursor; (b) CID-MS^3 spectrum of the signature fragment; (c) ETD-MS^2 spectrum of the peptide precursor

3.5 Conclusion

In this study, we design and demonstrate an effective method for targeting the prenyl peptides with selective chemical derivatization and subsequent enrichment through biotin-avidin affinity binding. Moreover, the fragmentation characteristics of the epoxidized prenyl peptides after the derivatization and enrichment were shown to consistently undergo predominant RSOH loss in CID, while for highly charged prenyl peptides like the KRas4B C-terminus, ETD can be utilized complementarily for more confident confirmation. To our knowledge this is the first time a MS study was conducted to reveal the fragmentation behaviors of extreme polybasic HVR (8 lysine in a 12-amino-acid sequence) in KRas4B C-terminus by both CID and ETD. Now that we have optimized the chemical reactions and mastered the fragmentation pattern of derivatized prenylation in the peptide level, we are investigating this strategy in the selective labeling and enrichment of prenylated proteins, in particular for the large-scale application in biological systems to target native prenylome for confident characterization.

Chapter 4

MS-cleavable Strategy for Selective Enrichment and Identification of Protein N-terminus and Proteolytic Products

4.1 Abstract

Proteolysis is one of the most important protein post-translational modifications (PTMs) that influences the function, activity, and structure of nearly all proteins during their lifetime. This irreversible PTM is regulated and catalyzed by proteases through hydrolysis reaction in the process of protein maturation or degradation. The identification of proteolytic substrates is pivotal in understanding the specificity of proteases and the physiological role of proteolytic process. To facilitate the targeted enrichment and identification of low-abundant proteolytic products, we devise a strategy incorporating a novel biotinylated reagent PFP (pentafluorophenyl)-Rink-biotin to specifically target, enrich and identify protein N-terminus. Within the PFP-Rink-biotin reagent, a MS-cleavable feature is designed to assist in the unambiguous confirmation of the enriched protein N-terminus. This is the first reported study of identifying protein N-terminus by MS-cleavable feature widely adopted in studying low-abundant protein PTMs and cross-linking/MS. The proof-of-concept study was performed with multiple standard proteins whose N-terminus were successfully modified, enriched and identified by reporter ion (RI) in the MS/MS fragmentation, along with the determination of N-terminal sequences by multistage tandem MS of the complementary fragment generated after the cleavage of MS-cleavable bond. The enrichment and identification of N-terminus by this strategy was also applied to *Escherichia. coli* cells for large-scale study.

4.2 Introduction

In vivo protein biosynthesis follows a stepwise fashion starting from N-terminal end towards the C-terminal end, usually with the N-terminal amino acid being a methionine(Met)[27]. Among the earliest modifications during or following the manufacturing of most proteins are the removal of initiating Met residue and the acetylation of N-terminal α-amino group, both of which are crucial in protein structure, stability, function

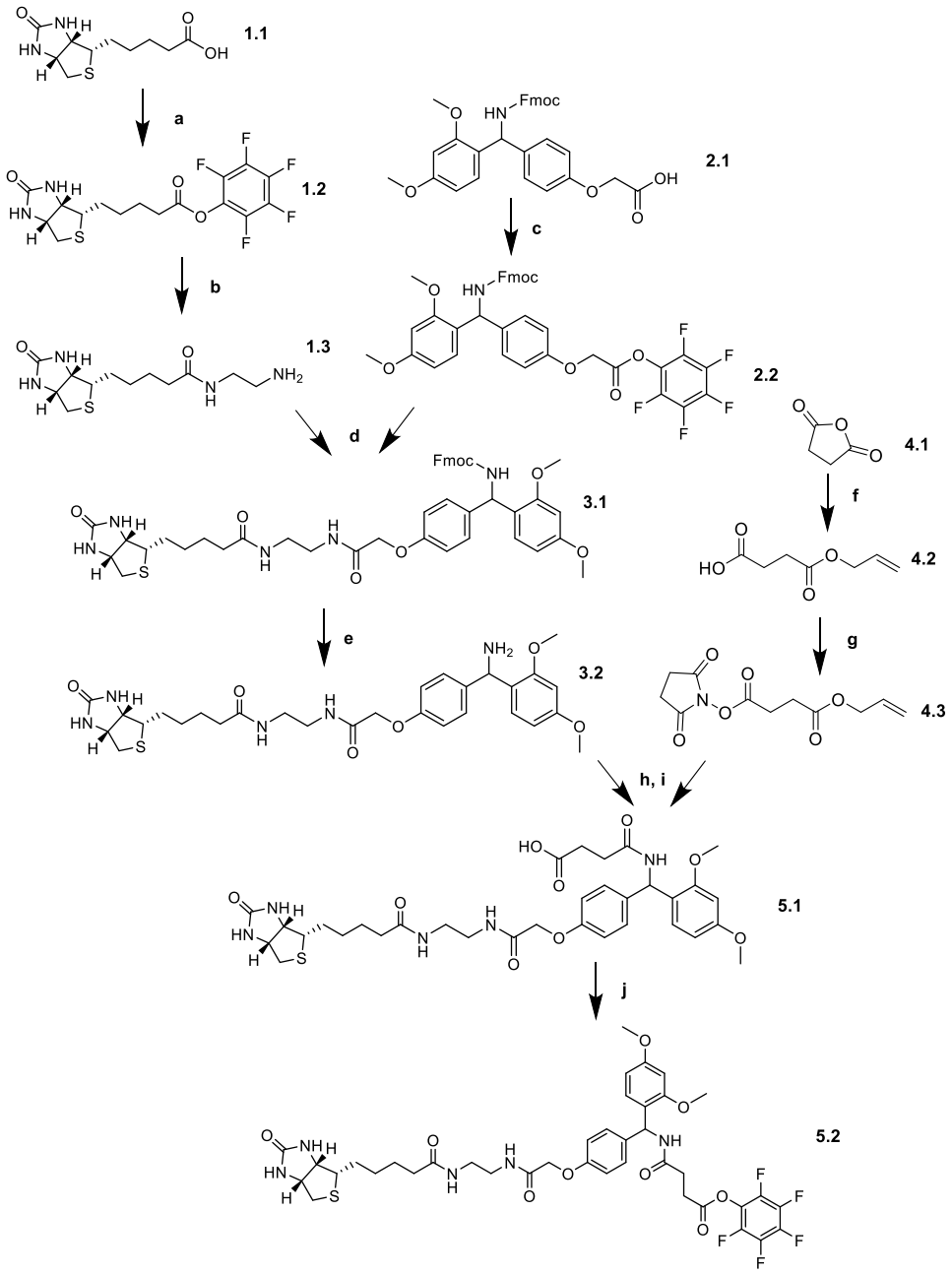
and localization[27,106,107]. As the two most ubiquitous protein N-terminal modifications, around 80% of soluble proteins in human have acetylated N-terminus[108] while all the proteins undergo proteolysis during their maturation or degradation processes[109]. Proteolysis is an irreversible and a highly selective hydrolytic reaction of peptide bond catalyzed by more than 560 proteases encoded by over 2% of human genome[26]. Proteolysis used to be considered predominantly as protein-degrading and disposing procedures, but gradually it has been discovered to play pivotal roles of precisely adjusting and finetuning the protein sequences to construct protein structures, regulate protein functions and control protein cellular localizations in a myriad of biological events[109]. The intracellular proteolytic network of caspases regulates and activates the transmission of inflammatory and apoptotic signals to control many important pathways, one of such examples is the blood coagulation cascade[30]. Proteolysis is also intimately involved in significant biological implications and pathological conditions of various cancers, chronic inflammations, neurodegenerative and cardiovascular diseases. It is estimated that 5-10% of all drug targets were designed for anti-proteolytic therapies[110–112]. Because proteolysis occurs co-translationally or post-translationally, the functional N-terminus of proteins cannot be directly inferred from genome sequence or transcriptome sequence. To reveal the pathways and mechanisms of proteolysis, proteolytic substrates need to be determined and the specifically cleaved positions by the proteases need to be identified through the study of protein degradomics[26,109,113].

Each protein has only one protein N-terminal sequence while numerous internal peptides from enzymatic digestion are generated through bottom-up proteomics approach. In silico endopeptidase Arg-C digestion of human proteome estimates an average of 17.5 internal peptides per N-terminal peptide[114]. This challenge will be more pronounced and overwhelming for identifying N-terminus and pertinent proteolytic products of low-abundant proteins in large-scale biological samples of increased complexity and dynamic range. Therefore, before the mass spectrometric analysis, certain enrichment strategies of the original protein N-terminus are essential for their confident identifications, which is conventionally achieved by either positive or negative selection strategies. Positive selection approaches specifically label the α-amine of the

protein N-terminus with an enrichment group while minimizing or preventing the mislabeling of lysine ε-amines, whose abundance is of higher orders of magnitude, either through chemical modifications (for example, by guanidination to block lysine ε-amine)[115–120] or enzymatically-assisted modification (subtiligase has high specificity for N-terminal α-amine)[121,122]. After sample digestion, the N-terminus of the proteins can be selectively purified by targeting the enrichment group of the N-terminus to be separated from the interfering internal peptides. On the contrary, negative selection approaches, notably the combined fractional diagonal chromatography (COFRADIC)[114,123] and terminal amine isotopic labeling of substrates(TAILS)[124,125], instead target the internal peptides for depletion after the blockage of all primary amines in the proteins, thus the N-terminal sequences of original proteins including the naturally blocked protein N-terminus can be retained for LC-MS/MS analysis. Negative selection strategies are advantageous in studying protein N-terminus with native modifications including N-terminal acetylation. However, enrichment by negative selection tends to result in higher sample complexity from the co-enrichment of free and modified protein N-terminus[126] and requires better enrichment efficiency for removing the more abundant internal peptides[127], which can be inferior in selectively studying proteolytically processed protein N-terminus compared to positive selection[26].

Ideally, a successful N-terminal positional proteomics experiment will only detect a single peptide from each protein. This can be problematic in the confident bioinformatic analysis of the respective protein because the lack of complementary confirmation of associated internal peptides from the same protein, which can in turn compromise the confidence of identifying protein N-terminus. The issue of false N-terminus discovery can be further exacerbated if the enrichment is incomplete or contains non-specific interactions, during the process of targeting the labeled N-terminus in positive selection approach or depleting the internal peptides in negative selection approach. Although these problems can be alleviated by vigorous validations and strict data filtering processes including the utilization of multiple search engines, narrower mass tolerance for identifications, repetitive confirmation of same peptide of different charge states in multiple replicates[124], these criteria may simultaneously filter out low-abundant protein N-

terminus. Here, we design a direct strategy to improve the confidence in the identification of proteolytic products by developing and utilizing a MS-cleavable reagent PFP(pentafluorophenyl)-Rink-biotin (Scheme 4-1) in positive enrichment for investigating proteolytic protein N-terminus of low abundance. The chemical reaction occurs as free protein N-terminus were targeted by PFP group and simultaneously biotinylated for downstream enrichment by removing non-labeled internal peptides. The confident N-terminus identification is enabled by the generation of a fixed signature reporter ion (RI) and high-intensity complementary fragment upon the CID-MS^2 fragmentation, followed by further MS^3 fragmentation on the complementary fragment to reveal the sequence information of the enriched protein N-terminus. This strategy incorporates the MS-cleavable feature widely applied in improving the identification capabilities and lowering false-discovery rate of low-abundant peptides in proteomics applications of protein post-translational modification (PTM) study[58] and cross-linking chemical biology[56]. To facilitate the automatic N-terminus confirmation, an in-house Java-based software NTermFinder was developed to validate the identified protein N-terminal peptides by the signature fragments and mass correlations among peptide precursor ion, complementary fragment and the scan numbers during MS acquisition. This strategy aims to achieve unambiguous identification of protein N-terminus and proteolytic products, especially in complex sample analyzed by a low MS-resolution instrument.



Scheme 4-1. Synthesis of PFP-Rink-biotin

4.3 Materials and methods

4.3.1 Materials and reagents

Ubiquitin of bovine erythrocytes, lysozyme of chicken and β-lactoglobulin of bovine milk, ammonium bicarbonate, sodium bicarbonate, dimethyl sulfoxide (DMSO), trifluoroacetic acid (TFA), chloroform and guanidine hydrochloride were purchased from Sigma-Aldrich (MO). O-methylisourea sulfate was from TCI America (OR). Iodoacetamide (IAM), methanol, acetonitrile (ACN) were obtained from VWR (PA). Formic acid (FA), 3K MWCO protein concentrators, monomeric avidin and streptavidin agarose resins were obtained from Thermo Scientific (IL). Dithiothreitol (DTT) was acquired BioRad (CA), trypsin was from Promega (WI) and Endoproteinase GluC was from New England Biolabs (MA). 1,1,1,3,3,3-Hexafluoroisopropanol (HFIP) was from Oakwood Chemical (SC). Bacterial cell lysis kit was acquired from Goldbio (MO). Synthesis of PFP-Rink-biotin was previously reported[128] and shown in Scheme 4-1. An Aries Filterworks (NJ) water system supplies all the high purity water used for preparing aqueous solutions.

4.3.2 Enrichment and identification of standard protein N-terminus

Preliminary experiment of labeling and enriching protein N-terminus was investigated for identifying the N-terminus of standard proteins: ubiquitin, lysozyme and β-lactoglobulin. First, guanidination procedure was performed as previously described to block ε-amines of these proteins[129]. Briefly, O-methylisourea sulfate was dissolved in water and mixed with 1 mM, 5 μL of standard proteins. pH of the reaction solution was adjusted to between 10-11 with the addition of NaOH and then incubated at 65˚C for 30 min. After guanidination, pH of the protein samples was adjusted to around 7. PFP-rink-biotin was dissolved in DMSO and added to the guanidinated proteins to a final concentration of 2.5 mM. The α-amine labeling reaction was proceeded for 2 hours at 37 ˚C. The excess PFP-rink-biotin was quenched by adding Tris-HCl buffer to a final concentration of 20 mM and removed by either 3K MWCO[50] or methanol-chloroform precipitation[82,130], followed by the disulfide cleavages with DTT and alkylation with IAM in dark. After the

labeling, sample was digested either with trypsin or GluC (1:50 w/w) overnight in 50 mM ammonium bicarbonate. The sample was re-concentrated by removing the solvent and reconstituted in pH 7.4 PBS buffer, followed by mixing with monomeric avidin resins for 1 h under room temperature. After peptide binding, the resins were washed extensively with PBS and ultrapure water sequentially to remove the non-binding peptides. Once done with the washing steps, the beads were incubated with the elution buffer ACN/H_2O/TFA (50/50/0.4, v/v/v) for 1 h at room temperature. Supernatant was separated from the beads by centrifugation, and was subsequently evaporated, desalted and reconstituted in 0.1% formic acid for LC-MS^n analysis.

4.3.3 Enrichment and identification of proteolytic products from *E. coli* cell lysate

The *Escherichia. coli (E. coli)* top 10 cell was a gift from Dr. Shawn Christensen's lab. Cell lysis was performed with bacterial cell lysis kit according to manufacturer's manual and the protein concentration was measured with bicinchoninic acid assay using BSA standards. Aliquots of *E. coli* cell lysate were dissolved with 6 M guanidine hydrochloride containing 50 mM sodium bicarbonate. Followed by reaction with 4 mM DTT and then alkylation with 12 mM IAM in dark. Alkylation reaction was quenched by 4 mM DTT solution followed by guanidination reaction at 4 °C for 24 h as previously reported(after adjusting to pH 10-11)[120]. Proteins were then purified by methanol-chloroform method, reconstituted in PBS (pH 7.4) and reacted with 4 mM PFP-rink-biotin for 2 h at 37 ˚C. After the reaction, excess PFP-Rink-biotin was quenched by adding Tris-HCl buffer and removed by methanol-chloroform purification. The samples were subjected to digestion with trypsin (1:50, w/w) overnight. The samples were then reconstituted in PBS buffer and then with mixing with PBS pre-washed streptavidin agarose beads for 1 h at room temperature. After peptide binding, the filtrate is removed, and the beads were washed extensively and sequentially with PBS and ultrapure water to remove the non-binding peptides. Once done with the washing steps, the beads were incubated with the elution buffer HFIP for 5 min at room temperature[131]. The filtrate was retained, evaporated, desalted and reconstituted in 0.1% formic acid for LC-MS^n analysis.

4.3.4 LC-MS^n analysis

All samples were analyzed in a Dionex UltiMate 3000 nano-UHPLC system hyphened with nano-ESI-linear ion trap (LIT) Thermo Velos Pro mass spectrometer (Thermo Fisher Scientific, Waltham, MA). For standard protein enriched samples, Acclaim PepMap C18 column (150 mm × 75 μm, 3 μm) was used for the LC separation with a 90 min gradient (mobile phase A: 0.1% FA in water; mobile phase B: 0.1% FA in 95% acetonitrile, 5% water; flow rate: 0.300 μL/min; 90 min gradient: 0–3 min 4.0%B, 3–80 min 4.0–50.0% B, 80–80.1 min 50–95% B, 80.1–85 min 95% B, 85–85.1 min 95–4% B, 85.1–90 min 4% B). Source voltage was 2.20 kV and capillary temperature was set to 275 °C. MS data was obtained from 300 to 2000 *m*/*z* mass range. Data-dependent MS/MS spectra were collected for the 10 most abundant precursor ions in each MS scan upon fragmentation (charge state ≥2; isolated width of 2 Da; min signal required: 200) using CID activation with 35.0% normalized collision energy, activation Q of 0.25, and activation time of 30 ms. For MS^3 data-dependent acquisition, precursor mass range was set according to the mass range of the analyte and MS^3 spectra were collected from the top 2 most abundant precursor ions upon fragmentation (charge state ≥2; isolated width of 2 Da; min signal required: 50) using CID. The activation was set with 45.0% normalized collision energy, activation Q of 0.25, and activation time of 30 ms.

For *E. Coli* enriched samples, Acclaim PepMap C18 column (500 mm × 75 μm, 3 μm) was used for the LC separation with same mobile phases as above with a flow rate of 0.200 μL/min in two separate gradient separations (Gradient 1: 0–5 min 4.0%B, 5–6 min 4.0–20.0% B, 6–145 min 20–70% B, 145–146 min 70-95% B, 146–150 min 95% B, 150–151 min 95-4% B, 151–170 min 4% B; Gradient 2: 0–4 min 4.0%B, 4–120 min 4.0–50.0% B, 120–145 min 50–80% B, 145–145.1 min 80-95% B, 145.1–150 min 95% B, 150–151 min 95-4% B, 151–170 min 4% B). Source voltage was 2.00 kV and capillary temperature was set to 275 °C. MS data was obtained from 400 to 2000 *m*/*z* mass range in rapid scanning mode. Data-dependent MS/MS spectra were collected for the 5 most abundant precursor ions in each MS scan upon fragmentation (charge state ≥2; isolated width of 2 Da; min signal required: 10000) using CID activation with 35.0%

normalized collision energy, activation Q of 0.25, and activation time of 10 ms. Top 2 ions (excluding m/z 569.2) in each MS^2 spectrum were selected for MS^3 data-dependent acquisition (charge state ≥2; isolated width of 2 Da; min signal required: 500) using CID activation of 45.0% normalized collision energy, activation Q of 0.25, and activation time of 10 ms.

4.3.5 MS data analysis

The MS raw files were converted to MGF format and mzML format (only MS^3 spectra) in Proteome Discoverer 2.1. MGF files were imported to in-house Java software NTermFinder to scan for RI (m/z 569.2) at ± 1 Da mass threshold for all the MS scans. MS^3 mzML file was analyzed in MS-GF+ against respective databases with the parameter settings as follow[83,84]: static modification of cysteine carbamidomethylation (+57.02146 Da) and lysine guanidination (+42.021798 Da); Variable modification of methionine oxidation (+15.99492 Da). Additional variable modifications from reagent labeling of protein N-terminus (+99.032027 Da) and methionine-removed protein N-terminus (-32.007458 Da) were added. The MS-GF+ identified PSMs were validated by NTermFinder based on the RI intensity, scan number, the m/z correlations among the RI, complementary fragment, and precursors as described in more details in discussion.

4.4 Results and discussions

4.4.1 Development of MS-cleavable reagent PFP-Rink-biotin for N-terminus labeling

MS-cleavable reagents are increasingly gaining popularities in the research of low-abundant proteins and protein PTMs by producing unique signatures such as reporter ion or neutral losses upon tandem MS fragmentation, which assist in the unambiguous identifications[58,59] and more precise quantifications[60] of these otherwise overshadowed targets. This unique feature of MS-cleavable reagent enables it to be an attractive strategy for identifying the protein N-terminus which are dramatically outnumbered by internal peptides. Hence, we are designing to improve the confidence in characterizing protein N-terminus by

incorporating a CID-cleavable Rink moiety, which has been formerly utilized in solid phase peptide synthesis[132] and novel cross-linker developments[133,134], in our design of new generation enrichment tag targeting primary amines. N-Hydroxysuccinimide (NHS) esters are conventionally used as reactive group towards primary amines[135]. However, pentafluorophenyl (PFP) ester was reported less subjective to hydrolysis and having higher reactivity with amines compared to NHS ester[136]. Combined with the enrichment tag biotin, PFP-Rink-biotin reagent is constructed and synthesized as shown in Scheme 4-1. Within the PFP-Rink-biotin reagent, the key feature for improving the confidence in N-terminus identification is the Rink group. Upon CID activation, the labile Rink group is preferentially cleaved under lower energy compared to the cleavage of peptide bonds, which readily gives rise to a high-abundant RI peak from the labeled peptides. This feature helps to alleviate the common issue of false-positive identification due to the binding and coelution of non-biotinylated peptides along with the biotinylated peptides. Within the same MS^2 spectrum, a complementary fragment that consists of the N-terminal peptide derivatized with the remnant of PFP-Rink-biotin after the loss of RI will be simultaneously generated in high intensity. This reagent-modified fragment can be further subjected to MS^3 data-dependent fragmentation to generate peptide bond fragmentation for the sequencing of original or proteolytic protein N-terminus facilitated by specific bioinformatic strategy (Figure 4-1).

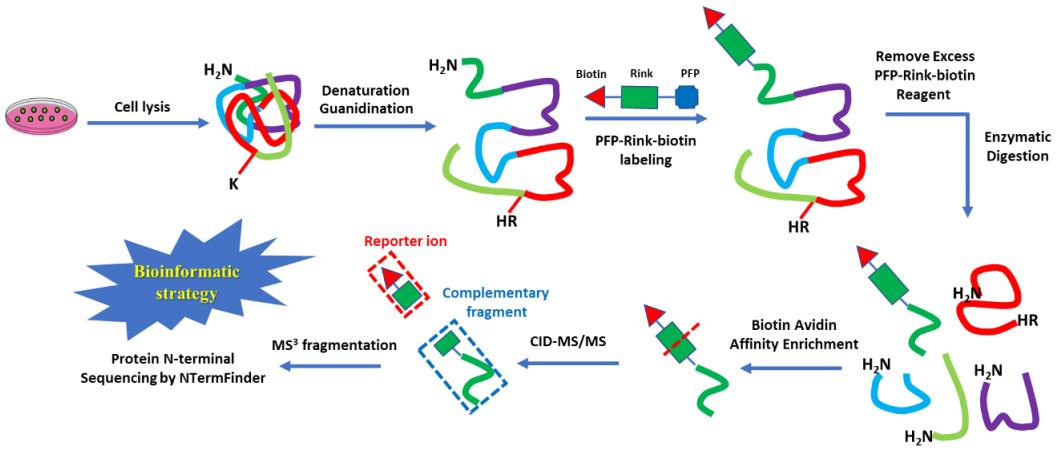


Figure 4-1. Workflow of PFP-Rink-biotin-based enrichment and identification of protein N-terminus

4.4.2 Evaluation of PFP-Rink-biotin in the enrichment and identification of protein N-terminus

The MS-cleavable feature of PFP-Rink-biotin in identifying the biotin-tagged protein N-terminus was investigated using three standard proteins: ubiquitin, β-lactoglobulin and lysozyme. For the model proteins we study, lysozyme[137] and β-lactoglobulin[138] were discovered containing signal peptides that are selectively removed during protein maturation. According to the N-terminal sequences of these model proteins, trypsin (for digesting ubiquitin and β-lactoglobulin) and GluC (for digesting lysozyme) were selected to obtain adequate lengths of the protein N-terminus for MS analysis against the known N-terminal sequence in the established protein database.

The enrichment of lysozyme N-terminus was first evaluated. For the N-terminus characterization of lysozyme, GluC digestion is more favorable over tryptic digestion because of its N-terminal lysine residue. After the labeling reaction, quenching and removal of PFP-Rink-biotin were performed before the digestion of protein for achieving desired enrichment of the lysozyme N-terminal peptide. The quenching of PFP reactive group deactivates its reactivity towards internal peptides to prevent the mislabeling and coelution of internal peptides during enrichment. The removal of excess PFP-Rink-biotin is as essential after the deactivation step due to competition for limited avidin binding sites leading to low binding efficiency of the labeled peptides and adversely affecting their subsequent enrichment. The reagent removal step was compared and evaluated by two different purification techniques: commercial protein concentrator molecular weight cut-off (MWCO) purification and methanol-chloroform purification. In the total ion chromatogram (TIC) of MWCO purification (Figure 4-S1a), two dominant peaks (arrowed) were observed and identified to be from the unreacted and hydrolyzed PFP-Rink-biotin reagent, whereas the two N-terminal peptide precursors (M^{2+} m/z 802.80 and M^{3+} m/z 535.50) were not present in the extracted ion chromatogram (XIC) and the lysozyme N-terminus cannot be identified in the MS data (Figure 4-S2a and Figure 4-S2b). The same protein sample was alternatively precipitated by methanol-chloroform purification

after reacting with the reagent, followed by the same digestion procedures. In this TIC, the two dominant reagent peaks existed in the MWCO treated sample were significantly reduced by methanol-chloroform purification (Figure 4-S1b). In the XIC the two precursors (M^{2+} m/z 802.80 and M^{3+} m/z 535.50) emerged in almost identical retention time (Figure 4-S2c and Figure 4-S2d) and identified to be the labeled lysozyme N-terminus by the tandem MS analysis. This experiment clearly shows methanol-chloroform purification is more advantageous over the commercial MWCO purification in removing the excess unreacted reagent to help enriching the protein, especially considering the widespread applications of methanol-chloroform in the purification of proteins from large-scale biological samples[82].

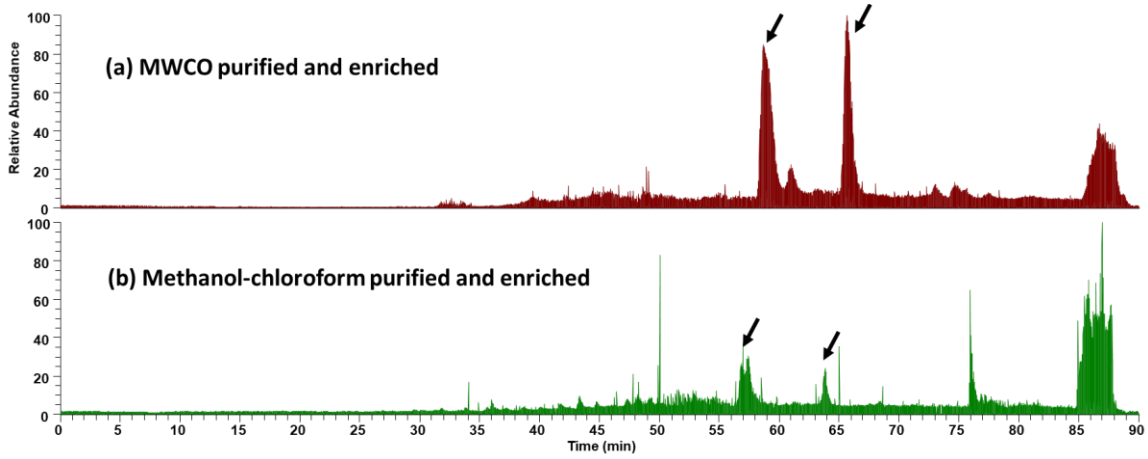


Figure 4-S1. TIC of lysozyme protein N-terminus after purification

(a) MWCO; (b) methanol-chloroform purification

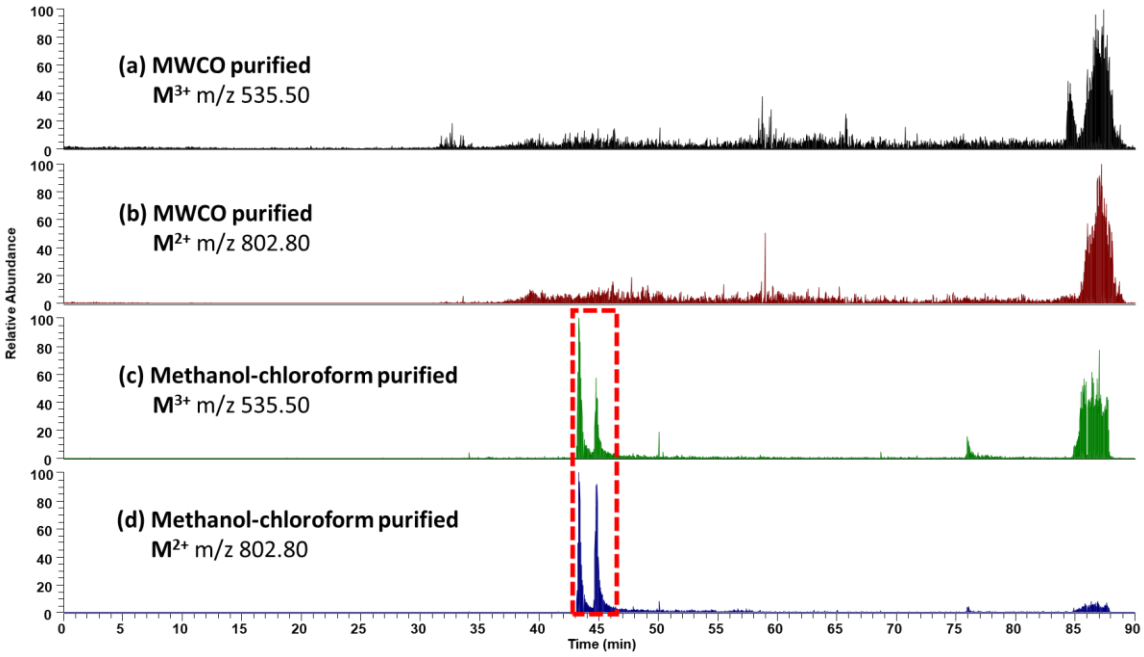


Figure 4-S2. XICs of lysozyme protein N-terminus after different purifications

(a) MWCO m/z 535.5; (b) MWCO m/z 802.8 (c) methanol-chloroform purification m/z 535.5; (d) methanol-chloroform purification m/z 802.8

Two corresponding molecular ions 802.97 (M^{2+}) and 536.18 (M^{3+}) were confirmed by matching against theoretical fragments of the lysozyme N-terminal sequence K_GVFGRC_CE (C_C represents carbamidomethylated cysteine) by tandem MS analysis. In the MS^2 spectrum of 802.97 (M^{2+}) precursor (Figure 4-2a), three dominant fragments were generated, one of which was the RI (m/z 569.22). The other fragment (labeled as $[M\text{-}RI+H]^+$ in the MS^2 spectrum) was identified to be the complementary fragment (m/z 1036.42) to the reporter ion, which was confirmed to match the m/z of [M-RI] in reduced charge state. The lowest intensity of the three being the neutral loss of ammonia from the complementary fragment. The MS^2 spectrum of 536.18 (M^{3+}) also generated highly intense complimentary fragment of 518.88 and the RI

peak (Figure 4-2b). While the relative intensity of RI was around 25% relative intensity of the base peak, which was lower compared to around 90% relative intensity from $M^{2+}$of the same peptide upon fragmentation, showing charge-dependent fragmentation pattern in producing the RI. Further MS^3 spectrum of the complimentary fragment (Figure 4-2c) was manually annotated against the N-terminal sequence of lysozyme, which exhibited confident coverage of the sequence confirmation. Note that mass addition of remnant fragment to the peptide was calculated for all the complementary fragments and the resultant fragments in the MS^3 spectrum. Although different relative intensities of RI and complimentary fragments were observed, this study shows that the same peptide with different charge states helps to further confirm the protein N-terminus. In this particular case, the higher abundance of RI from M^{2+} helps to identify the enriched peptide utilizing the signature feature of PFP-Rink-biotin, while the higher abundance of complimentary fragment from M^{3+} results in better sequence annotation, both of which can complement and contribute to the identification of lysozyme N-terminus.

After the optimization and evaluation of lysozyme N-terminus, the tandem MS fragmentation study was also performed with ubiquitin and β-lactoglobulin N-terminus. Although the guanidination step modified lysine to homoarginine within the proteins, tryptic digestion was previously proven to be effective in cleaving the C-terminal of homoarginine[129]. MS^2 and MS^3 spectra of the enriched ubiquitin protein N-terminal peptides precursor were extracted. Ubiquitin N-terminus MQIFVK$_G$ modified by PFP-Rink-biotin was generated upon the cleavage after the first guanidinated lysine (K_G), with the m/z of 737.79 (M^{2+}) in the full MS spectrum. The CID spectrum of M^{2+} peptide precursor produced RI and complimentary fragment ion (m/z 906.54), both in high relative intensities (Figure 4-3a). The complimentary fragment ion was subjected to MS^3 fragmentation to identify the ubiquitin N-terminus sequence shown in Figure 4-3b. This result matched with the previously Edman-degradation-determined complete amino acid sequence where the N-terminal methionine is not removed in the mature form of ubiquitin[139]. Similarly, identification of the β-lactoglobulin N-terminal sequence after tryptic digestion was achieved similarly by labeling with PFP-Rink-biotin reagent and MS analysis with multistage activation (Figure 4-S3).

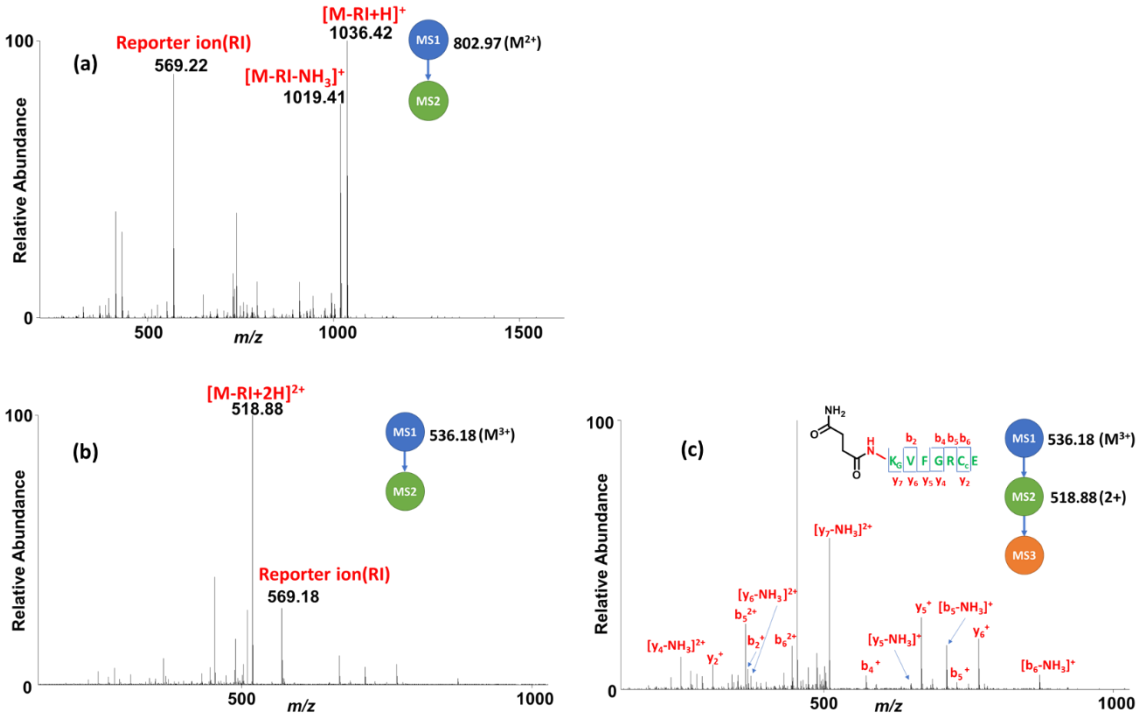


Figure 4-2. Identification of lysozyme labeled protein N-terminus (after GluC digestion) by MS^n analysis

(a) MS^2 spectrum of M^{2+} precursor; (b) MS^2 of M^{3+} precursor; (c) MS^3 spectrum of complementary fragment 518.88(2+)

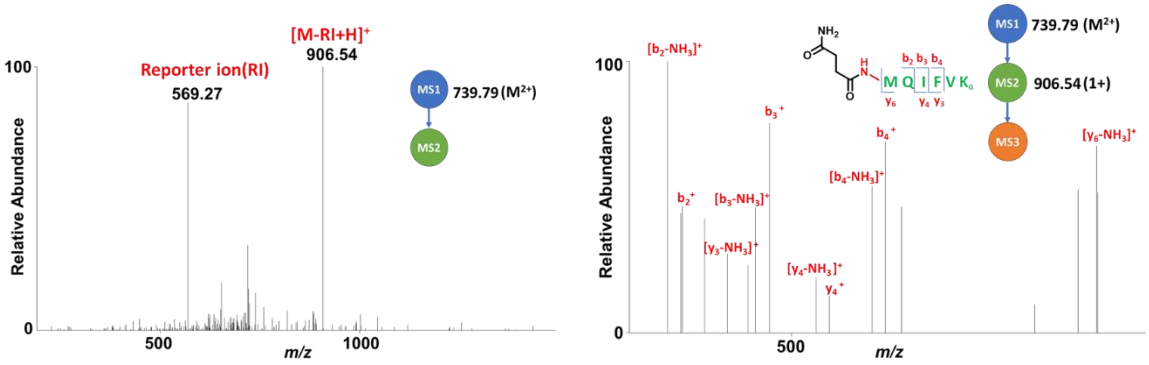


Figure 4-3. Identification of labeled ubiquitin protein N-terminus (after trypsin digestion) by MS^n analysis

(a) MS^2 spectrum of M^{2+} precursor; (b) MS^3 spectrum of complementary fragment 906.54(1+)

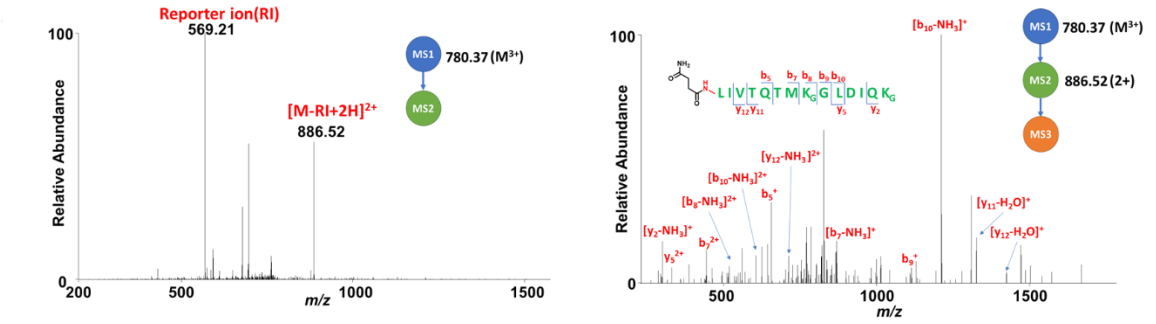


Figure 4-S3. Identification of β-lactoglobulin protein N-terminus by MS^n analysis after trypsin digestion

(a) MS^2 spectrum of M^{3+} precursor; (b) MS^3 spectrum of complementary fragment 886.52(2+)

To summarize, the MS analysis of the labeled N-terminus of standard proteins shows that the RI can be consistently and predominantly produced across different modified peptides, usually one of the highest fragments observed, exhibiting low energy pathway of cleavage than most of peptide bond fragmentations. This signature RI is confirmed to be suitable as a common feature for identifying the N-terminal peptides. Complementary fragment coexists as one of the most intense fragments among the CID-MS^2 spectra,

paired with the RI as identifiers of the derivatized N-terminal peptides. Furthermore, this complimentary fragment contains the complete N-terminal sequence, which can be further subjected to multistage data-dependent MS^3 fragmentation for peptide sequencing. The enrichment strategy was confirmed to be efficient after optimization and proceeded for large-scale N-terminus enrichment and analysis.

4.4.3 Bioinformatic strategy for the application of PFP-Rink-biotin in large-scale N-terminus identification

Strategy for protein N-terminus identification in large-scale by PFP-Rink-biotin was implemented as shown in Figure 4-1. The proteins from cell lysis were first subjected to denaturation and guanidination for blocking the ε-amines of lysine. Near quantitative labeling of lysine guanidination (> 99%) was previously achieved[120] and was adapted in our experiment. The N-terminus that contain free α-amines were then labeled by PFP-Rink-biotin reagent. After the labeling of protein N-terminus, excess reagent was quenched and removed by methanol-chloroform precipitation before enzymatic digestion to prevent mislabeling of internal peptides. the biotinylated N-terminal peptides can be selectively targeted and isolated by biotin-avidin affinity enrichment, and the sample complexity was significantly reduced before subsequent tandem MS and bioinformatic analysis.

To facilitate the bioinformatic analysis and automatic validation in large-scale N-terminus identification, an in-house Java script NTermFinder was developed and incorporated to streamline the filtering and selection of N-terminal peptides from MS analysis (Figure 4-4). Raw MS files of enriched samples were converted to MGF (Mascot Generic File) format which were subsequently imported into NTermFinder to select spectra containing m/z of RI (569.2 ± 1 Da) with at least 20% relative abundance within the spectra. The scan numbers of these spectra were denoted as $MS^2Scan\#$. On the other hand, the MS^3 spectra from the same Raw file was converted to mzML format and then searched in MS-GF+, where the protein N-terminal peptides were identified, and the corresponding scan numbers were reported as $MS^3Scan\#$. Furthermore, a 3-step validation of the identified N-terminal PSMs from MS-GF+ was performed in the NTermFinder: (1) Scan number validation: scan numbers of MS^2 and MS^3 spectra have to match $MS^3Scan\# \leq MS^2Scan\# + N$,

because only top *N* precursors from each MS^2 scan were subjected to MS^3 fragmentation in the LCMS method; (2) Precursor of each MS^3 N-terminal PSM should be present as one of the fragments in the corresponding MS^2 spectrum; (3) The MS^1 precursor of the identified MS^3 PSM should be within 1 Da of the theoretical result calculated by (*Precursor***Charge* + 569.2+1.0078) /(*Charge*+1), where *Precursor* is the MS^3 precursor. Both *Precursor* and *Charge* were extracted from the MS-GF+ output results. The rationale behind this validation step is that in MS^2 spectra these complementary fragments were always found to be one charge less than the respective charges of their precursors in the full MS spectra (as confirmed in the analysis of standard proteins N-terminus), due to the protonation of biotin group within the RI in the same spectrum. This reduce-charge effect is incorporated as a criterion in distinguishing the positive results of protein N-terminus from potential interferences. The final results after the validation were reported as confident protein N-terminus identifications.

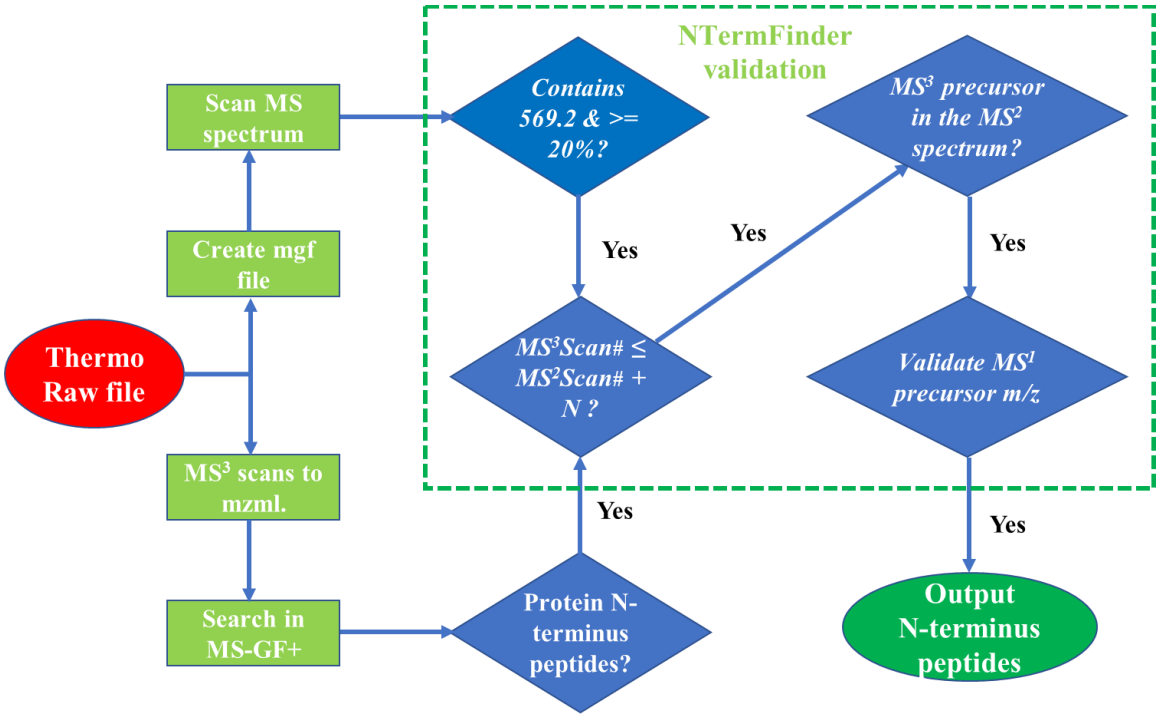


Figure 4-4. Bioinformatic approach for the validation of PFP-Rink-biotin labeled protein N-terminus

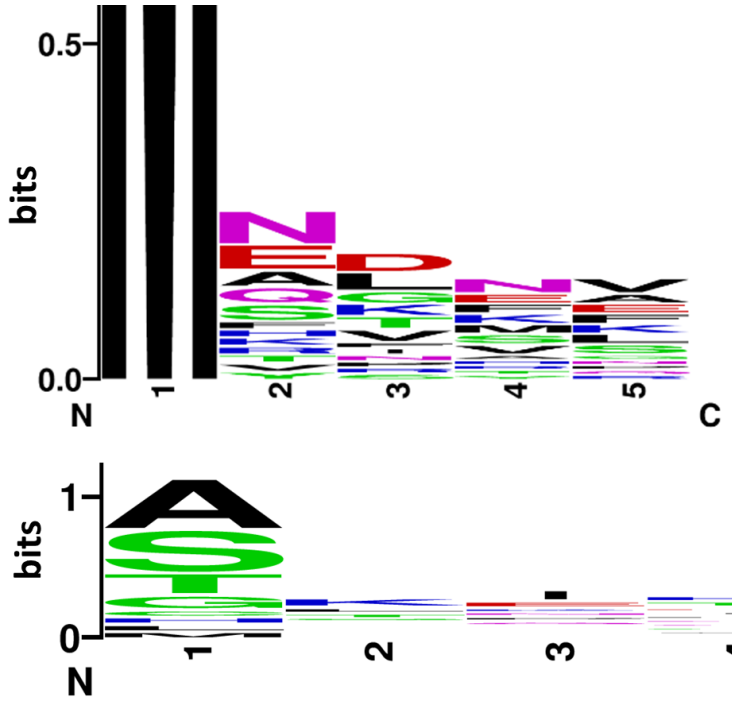
4.4.4 Application of PFP-Rink-biotin to *E. coli* cell lysate

Proof-of-concept large-scale N-terminomic study was carried out for the N-terminus enrichment of *E. coli* cell lysate. During this study, the binding of labeled N-terminus peptides with streptavidin and elution by HFIP was found to be more effective compared to binding with monomeric avidin and elution by ACN. In the LC-MS/MS analysis, two LC gradients with high ACN elution were performed to account for the increased hydrophobicity due to the attachment of bulky Rink-biotin group to the N-terminal peptides. In addition, to reduce the selection of background noises for the fragmentation in MS^2, the minimum signal threshold for MS/MS fragmentation was increased to 10000 (see the method section for details). The resultant MS data files were searched according to the bioinformatic strategy above and the total number of N-terminus identified from the sample runs was 42 after the in-house bioinformatic approach along with manual removal of repetitive sequences and validation of the modifications identified (Table 4-1).

N-terminal Peptide	Protein	N-terminal Peptide	Protein
-.+99.032M+15.995ARYFRRRK+42.022FC+57.021R.F	sp|B1XDV2|RS18_ECODH	-.+99.032MYVVSTK+42.022.Q	sp|B1X7I7|GATY_ECODH
-.+99.032M+15.995ELK+42.022K+42.022LM+15.995GHISIIPDYR.Q	sp|P28917|YDCC_ECOLI	-.+99.032MYVVSTK+42.022.Q	sp|B1X7I7|GATY_ECODH
-.+99.032M+15.995FK+42.022RRYVTLLPLFVLLAAC+57.021SSK+42.022.P	sp|P41052|MLTB_ECOLI	-.-32.007MGFTTR.V	sp|P32720|ALSC_ECOLI
-.+99.032M+15.995K+42.022DK+42.022VYK+42.022.R	sp|P0AFC0|NUDB_ECOLI	-.-32.007MGFTTR.V	sp|P32720|ALSC_ECOLI
-.+99.032M+15.995NTEATHDQNEALTTGARLR.N	sp|B1XAZ1|RODZ_ECODH	-.-32.007MGIFSR.F	sp|P0AFM6|PSPA_ECOLI
-.+99.032M+15.995RVNLLITM+15.995IIFALIWPVTALR.A	sp|P77694|ECPD_ECOLI	-.-32.007MHSLQR.K	sp|P37051|PURU_ECOLI
-.+99.032M+15.995SGFFQRLFGK+42.022.D	sp|P39293|YJFK_ECOLI	-.-32.007MSK+42.022IVK+42.022.I	sp|B1XDI9|ENO_ECODH
-.+99.032MAK+42.022APIRARK+42.022RVRK+42.022.Q	sp|B1X6E9|RS11_ECODH	-.-32.007MSK+42.022IVK+42.022.I	sp|B1XDI9|ENO_ECODH
-.+99.032MENFK+42.022HLPEPFR.I	sp|B1X9T7|TNAA_ECODH	M.+99.032AAK+42.022DVK+42.022.F	sp|B1XDP7|CH60_ECODH
-.+99.032METTQTSTIASK+42.022DSR.S	sp|P0AAD6|SDAC_ECOLI	M.+99.032AK+42.022GQSLQDPFLNALR.R	sp|B1XDS3|HFQ_ECODH
-.+99.032MHPMLNIAVR.A	sp|P0ADG4|SUHB_ECOLI	M.+99.032AK+42.022GQSLQDPFLNALR.R	sp|B1XDS3|HFQ_ECODH
-.+99.032MHPMLNIAVR.A	sp|P0ADG4|SUHB_ECOLI	M.+99.032ALNLQDK+42.022QAIVAEVSEVAK+42.022.G	sp|B1XBY7|RL10_ECODH
-.+99.032MNDSEFHR.L	sp|B1XAH3|CYAY_ECODH	M.+99.032AQQTPLYEQHTLC+57.021GAR.M	sp|P27248|GCST_ECOLI
-.+99.032MNDSEFHR.L	sp|B1XAH3|CYAY_ECODH	M.+99.032ASENMTPQDYIGHHLNNLQLDLR.T	sp|B1X9W6|ATP6_ECODH
-.+99.032MNFEGK+42.022IALVTGASR.G	sp|P0AEK2|FABG_ECOLI	M.+99.032ATPHINAEMGDFADVVLMPGDPLR.A	sp|B1XFJ4|DEOD_ECODH
-.+99.032MNLHEYQAK+42.022.Q	sp|B1X6Q8|SUCC_ECODH	M.+99.032ATVSMR.D	sp|B1XD38|RS2_ECODH
-.+99.032MQGSVTEFLK+42.022PR.L	sp|P0A7Z4|RPOA_ECOLI	M.+99.032C+57.021GIVGAIAQR.D	sp|P17169|GLMS_ECOLI
-.+99.032MQLNSTEISELIK+42.022QR.I	sp|B1X9W2|ATPA_ECODH	M.+99.032C+57.021GIVGAIAQR.D	sp|P17169|GLMS_ECOLI
-.+99.032MSIVVK+42.022.N	sp|B1XCN7|NORV_ECODH	M.+99.032LYIDK+42.022ATILK+42.022FDLEM+15.995LK+42.022K+42.022.H	sp|P0AET5|HDED_ECOLI
-.+99.032MTDMNILDLFLK+42.022.A	sp|P0ABU9|TOLQ_ECOLI	M.+99.032M+15.995VQHLK+42.022RRPLSRYLK+42.022DFK+42.022.H	sp|P52627|FLIZ_ECOLI
-.+99.032MVSNASALGR.N	sp|P0AC44|DHSD_ECOLI	M.+99.032SK+42.022EHTTEHLR.A	sp|P64581|YQJD_ECOLI

Table 4-1. 42 unique protein N-terminus were identified from *E. coli* cells

The predominant proteolytic event in *E. coli* is the excision of Met by methionine aminopeptidase (MetAP) after the deformylation of the N-terminus[140]. In this catalytic event the Met being removed is at P1 position whereas the new N-terminal amino acid (penultimate position in the original sequence) is the P1' position according to Schechter and Berger nomenclature[141]. The efficiency of Met removal is mostly depending on the size of side-chain in P1' amino acid based on the *in vitro* study with synthetic peptides: amino acid of larger size-chain generally has lower MetAP catalytic efficiency, therefore protein N-terminus with small amino acid residue as P1' such as Ala, Gly, Pro, Ser and Cys have near complete Met excision, whereas cleavage of penultimate Val and Thr could depend on the P2'-P4' amino acids while even lower cleavage efficiency was observed for Phe, Arg, Tyr, Trp[142,143]. The identified N-terminal sequence motifs[144] were analyzed and summarized in Figure 4-5. Overall, the result of identified *E. coli* N-terminus from the enrichment experiment matches well with the expectation of this size-dependent proteolytic event, except for two Met intact peptides with P1' Ala potentially due to partial Met removal. In addition, nearly half of all N-terminal Met identified were cleaved in our experiment, which is consistent with previously reported control study[145]. Though the total number of identifications is comparably lower than the other enrichment methods reported in the literature due to the longer acquisition time required for MS^3 scanning, it still provides a confident validation of the enriched N-terminal peptides even with a low MS-resolution ion trap instrument.



	A	E	F	K	N	R	S	H	Q	T	V	Y	G	C	L	M	Total
Met Intact	2	3	1	1	4	1	2	1	2	1	1	1	0	0	0	0	20
Met Removed	7	0	0	0	0	0	6	1	0	3	0	0	2	1	1	1	22
																	42

Figure 4-5. N-terminal sequence motif for the identified *E. coli* proteins

(a) Met retained; (b) Met removed; (c) Summary of penultimate amino acids

4.5 Conclusion

In this study we explored the potentials of implementing MS-cleavable strategy for the identifications of proteolytic N-terminus after enrichment. The design of this new cleavable labeling reagent PFP-Rink-biotin inherently provides a fixed signature RI generated in high intensity for the confirmation of the labeled peptides, with the sequence information being extracted and analyzed by multistage MS fragmentation. Further validation of the N-terminus authenticity was facilitated by the correlations among precursors/product ions from MS^1 to MS^3 levels and automated by our custom high-throughput

bioinformatic approach catering to the result files of MS-GF+ searches. Large-scale application with this strategy is still under development with the ongoing focus being on the optimization of enrichment and LCMS acquisition methods for identifying new proteolytic substrates from different organisms to improve our understanding of proteolytic mechanisms in normal and diseased biological systems.

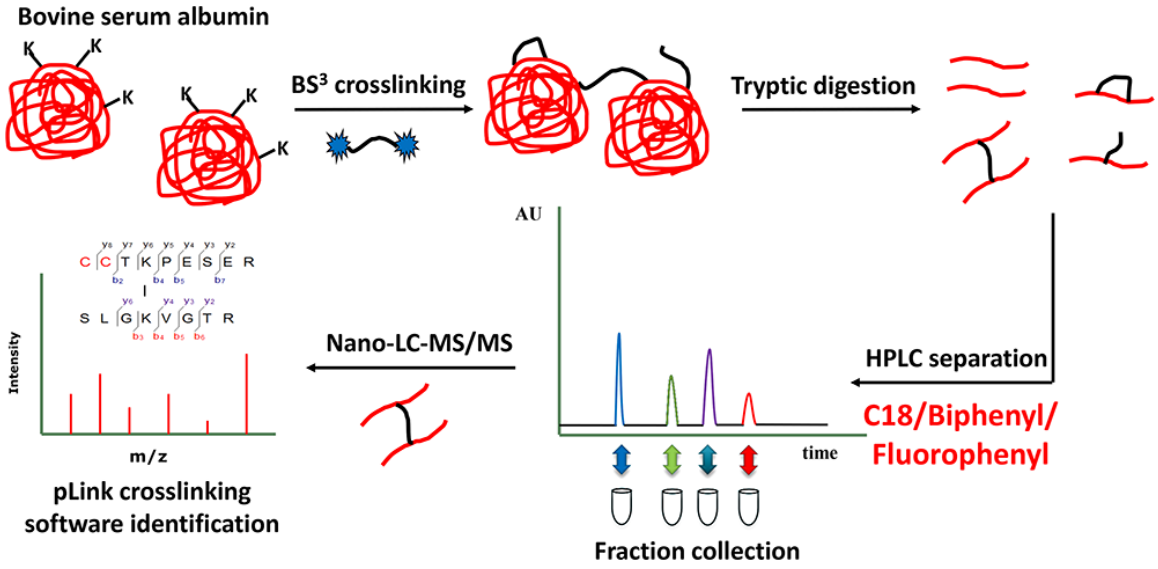
Chapter 5

Evaluation of Different Stationary Phases in the Separation of Inter-Cross-Linked Peptides[1]

Authors: Zixiang Fang, Yehia Z. Baghdady, Kevin A. Schug, and Saiful M. Chowdhury

[1] Reprinted with permission from Fang, Z.; Baghdady, Y. Z.; Schug, K. A.; Chowdhury, S. M. *J. Proteome Res.* 2019, 18, 4, 1916–1925. Copyright 2019 American Chemical Society.

5.1 Abstract



Chemical cross-linking coupled with mass spectrometry (MS) is becoming a routinely and widely used technique for depicting and constructing protein structures and protein interaction networks. One major challenge for cross-linking/MS is the determination of informative low-abundant inter-cross-linked products, generated within a sample of high complexity. A C18 stationary phase is the conventional means for reversed-phase (RP) separation of inter-cross-linked peptides. Various RP stationary phases, which provide different selectivities and retentions, have been developed as alternatives to C18 stationary phases. In this study, two phenyl-based columns, biphenyl and fluorophenyl, were investigated and compared with a C18 phase for separating BS^3 (bis(sulfosuccinimidyl)suberate) cross-linked bovine serum albumin (BSA) and myoglobin by bottom-up proteomics. Fractions from the three columns were collected and analyzed in a linear ion trap (LIT) mass spectrometer for improving detection of low abundant inter-cross-linked peptides. Among these three columns, the fluorophenyl column provides additional ion-exchange interaction and exhibits unique retention in separating the cross-linked peptides. The fractioned data was analyzed in pLink, showing the fluorophenyl column consistently obtained more inter-cross-linked peptide identifications than both C18 and biphenyl columns. For the BSA cross-linked sample, the identified inter-

cross-linked peptide numbers of the fluorophenyl to C18 column are 136 to 102 in “low confident” results and 11 to 6 in “high confident” results. The fluorophenyl column could potentially be a better alternative for targeting the low stoichiometric inter-cross-linked peptides.

5.2 Introduction

Chemical cross-linking is on its way to becoming a routine and straightforward technique for exploring protein interactions and protein structures. In a cross-linking study, a cross-linker is used to chemically react, derivatize, and covalently link the side chains of specific amino acids, such as lysine, cysteine, or other charged residues, between interacting proteins under physiological conditions.[146–148] An efficient cross-linking reaction can capture and stabilize macromolecules when two reactive side chains of interacting proteins are within the reacting range of the cross-linker.[149] By customizing the reactive groups and the span of a cross-linker, the accessibility of certain amino acids in proteins and the distance of protein interactions can be deduced and utilized for determining the low-resolution structures of proteins and their interacting partners.[149–151] The simplicity of experimental setup and the capability to capture weak/transient interactions render cross-linking a great complementary approach to traditional high-resolution techniques, such as X-ray crystallography or nuclear magnetic resonance (NMR).[152–154] Coupled with sensitive mass-spectrometric detection, chemical cross-linking is developed to be a very powerful tool for identifying protein interactions and unravelling the structures of proteins or protein complexes.[155–161]

Despite many advantages with respect to deducing protein structures, chemical cross-linking still faces certain challenges and limitations. The major drawback is the low abundance of cross-linked proteins. Chemical cross-linking studies are generally performed in the fashion of “bottom-up” proteomics, which begins with the cross-linking reaction of proteins in biological context, followed by enzymatic digestion of proteins into peptides and further analysis of these peptides by LC-MS/MS.[146,162] These cross-linked peptides generated by reacting with cross-linkers are not equally informative.[151,162] Three main types of cross-linked peptides can be formed after cross-linking and enzymatic digestion of proteins:[163] (1) intra-

cross-linked peptides; (2) dead-end cross-linked peptides; and (3) inter-cross-linked peptides. An intra-cross-link or "loop-link" is the cross-linking of two reactive groups within the same peptide sequence; a dead-end cross-link or "mono-link" is the cross-linking of one reactive group, while the other reactive end is deactivated; an inter-cross-link is the cross-linking of two functional groups of two different peptide sequences. The information on protein conformation and interaction is predominately acquired from the inter-cross-linked products. However, inter-cross-linked peptides usually account for the smallest portion, while dead-end counterparts prevail following execution of the cross-linking reaction.[162] Enzymatic dissection generates enormous amount of unmodified peptides, which further lowers the relative abundance of inter-cross-linked peptides. On the other hand, the interpretation and analysis of cross-linking MS data are considered much more time-consuming and computationally intensive than traditional proteomic analysis.[150] The difficulty in interpreting MS data associated with cross-linking experiments is due to the combinatorial increase from a theoretical non-cross-linked peptide number of n to a theoretical cross-linked peptide number of n^2.[162] This requires more computational capacity in hardware and complicated searching algorithms in software for cross-linking analysis, especially for samples with high biological complexity. Many software tools addressing the data analysis of cleavable or noncleavable cross-linking/MS have been developed, such as MetaMorpheus,[164] X-linkX,[158] Xlink Identifier,[165] SIM-XL,[166] Kojak,[167] Stavrox,[168] and pLink.[169] However, without efficient enrichment of inter-cross-linked peptides prior to mass spectrometric analysis, the processing of MS data to confidently identify inter-cross-linked peptides remains a challenging goal.

Strategies had been designed for reducing complexity in cross-linked samples and enriching inter-cross-linked peptides. Trifunctional cross-linkers, which incorporate an affinity tag like a biotin moiety, can be specifically targeted and selectively enriched by immobilized avidin.[147,170,171] Dual cleavable cross-linker DUCCT was developed to pinpoint the inter-cross-linked peptides by simultaneous CID and ETD fragmentations.[150] After the cross-linking reaction, non-cross-linked peptides can be removed before the recovery of cross-linked peptides from avidin. However, this approach cannot differentiate dead-end cross-

linked peptides from inter-cross-linked peptides, and the nonspecific binding of unmodified peptides may occur. Meanwhile, the biotin group can also be fragmented in tandem MS. This further complicates MS analysis and increases false discovery rate (FDR). Structurally, inter-cross-linked peptides can be considered as the sum of two individual linear peptides fixated by a cross-linker. These products are structurally having higher molecular mass and longer average sequence than dead-end cross-linked and unmodified peptides. Thus, size-exclusion chromatography (SEC) is applicable to enrich and separate inter-cross-linked peptides from their non-cross-linked counterparts.[156,172] Moreover, inter-cross-linked peptides bear higher charges in an acidic environment because they possess more chargeable groups within the combined peptide sequence, relative to non-cross-linked species. Thus, strong cation exchange (SCX) chromatography is also feasible to separate highly charged inter-cross-linked peptides from other types of peptides.[153,173,174] Fritzsche et al. compared the number of inter-cross-linked peptides of BS^3 cross-linked BSA with or without SCX enrichment and observed almost an 8-fold increase in inter-cross-linked-peptide identifications after enrichment.[173] More recently, two-dimensional (2D) SCX approaches were developed to further assist in the separation of inter-cross-linked peptides. Buncherd et al. explored a chemical-cleavable cross-linker and utilized 2D diagonal SCX chromatography for enriching inter-cross-linked peptides.[174] The determination of inter-cross-linked peptides was achieved by mapping the relation of intact cross-linked peptide before and after chemical cleavage. Another 2D SCX approach was reported by Tinnefeld et al.[153] In that study, LysC digestion of cross-linked protein was performed to generate long, highly charged peptides, which were preferentially retained in the first dimension SCX separation. Upon the first SCX fractionation, the lengthy cross-linked peptides were subjected to tryptic digestion and followed by a second SCX separation and fraction collection to achieve efficient enrichment.

Alkyl-based stationary phases are the most widely utilized bonded phases in RP high-performance liquid chromatography (HPLC) for the separation of cross-linked peptides. All the aforementioned enrichment methods from the literature inevitably included C18 separation during the separation of cross-linked peptides, regardless of the sample complexity. In a C18 column the cross-linked peptides are separated

based on their relative hydrophobicity. In general, the hydrophobicity of a peptide increases as the number of amino acids in the peptide sequence increases. However, the increase of basic ionizable functional groups such as histidine, lysine, and arginine in longer peptides will increase the number of accessible charge states and yield more interactions with charged stationary phases under acidic condition. Another major difference between the inter-cross-linked peptides and others is that the geometric arrangements of inter-cross-linked peptides are changed when covalently linked. This conformational distinction of inter-cross-linked peptides has the potential to be exploited for targeted separation using stationary phases with shape selectivity.

Many alternative stationary phases to C18 have been developed. These stationary phases have been compared and shown variations in selectivity and retention for multiple classes of analytes,[175] but they have not been fully explored for separating complex mixtures of macromolecules, particularly cross-linked macromolecules. Here, we assessed C18, biphenyl, and fluorophenyl columns for their ability in selectively separating and enriching inter-cross-linked peptides of BSA cross-linked by BS^3 cross-linker. We evaluated and compared the performance of three commercially available column chemistries according to chromatographic retention and number of identified inter-cross-linked peptides by the cross-linking software "pLink1".[169] This work aims to provide insights in improving separation and enrichment of the low abundant inter-cross-linked peptides by alternative stationary phases.

5.3 Materials and methods

5.3.1 Materials and reagents

Bovine serum albumin (BSA), myoglobin of equine heart, Tris·HCl, ammonium bicarbonate (ABC), and formic acid (FA) were purchased from Sigma-Aldrich (St. Louis, MO). Methanol and acetonitrile were obtained from VWR (Radnor, PA). BS^3 cross-linker was obtained from Thermo Scientific (Rockford, IL). 3K MWCO protein concentrators (Pierce, IL) were used to concentrate the protein after the cross-linking reaction. Dithiothreitol (Biorad, CA), iodoacetamide (Sigma- Aldrich, MO), and trypsin (Promega,

Madison, WI) were obtained. A Milli-Q-filtrated water (resistivity >18 MΩ) by Aries Filterworks (West Berlin, NJ) filtration system was used to prepare all aqueous solutions in this study.

5.3.2 Cross-linking of BSA

BS^3 cross-linker was prepared by dissolving to 50 mM solution in water. BSA was reacted with BS^3 cross-linker in 1:50 molar ratio in pH 7.3 phosphate buffer saline (PBS). Different concentrations of cross-linked samples were achieved by increasing the amount of BSA used for the cross-linking reaction. The reaction was performed under room temperature for 30 min in a constantly mixing thermomixer. The cross-linking reaction was quenched by adding Tris·HCl buffer (pH 8.8) to a final concentration of 40 mM. The sample was transferred to a 3K MWCO protein concentrator to remove excess cross-linker by washing with 50 mM ABC solution under 15000 G centrifugation. Then the cross-linked BSA sample was subjected to in-solution digestion.

5.3.3 In-solution digestion

The cross-linked BSA protein was first reduced with 10 mM dithiothreitol solution and then alkylated with 55 mM iodoacetamide in dark. The digestion was carried out by incubating the sample with modified trypsin in 1:100 (w/w) trypsin/protein ratio at 37 °C overnight. After that, the enzymatic digestion was stopped by adding FA. The digested sample was dried by a speed vacuum concentrator and resuspended in 0.1% FA solution for LC separation and mass spectrometric analysis.

5.3.4 LC-MS/MS analysis

Analysis of the cross-linked BSA samples using C18, biphenyl, and fluorophenyl columns was performed on a Prominence HPLC system coupled to an electrospray ionization (ESI) -ion trap (IT) -time-of-flight (TOF) mass spectrometer (Shimadzu, Japan). The chromatographic separation in the HPLC-ESI-IT-TOF was achieved using three Restek (Bellefonte, PA) columns: Raptor C18 column (100 × 2.1 mm, 2.7 μm), Raptor Biphenyl column (100 × 2.1 mm, 2.7 μm), and Force Fluorophenyl column (100 × 2.1 mm, 3

μm) in a 60 min gradient (flow rate: 0.2 mL/min; gradient: 0-2 min 5% B, 2-40 min 5-50% B, 40-40.01 min 50-100% B, 40.01-50 min 100% B, 50-50.01 min 100-5% B, 50.01-60 min 5% B; mobile phase A: 0.1% FA in water; mobile phase B: 0.1% FA in 95% acetonitrile, 5% water). MS scan mass range for MS and MS/MS was set from m/z 200 to 2000. Data-dependent acquisition was performed for the fragmentation of peptides. Five MS events (one MS scan and four MS/MS scans) were performed. Event 1 was full MS scan. Event 2 to 5 (4 MS/MS scan events) selected the four most abundant ion of each event 1 scan for further fragmentation (charge state ≥ 2; isolated width of 3 Da; precursor ion threshold: 20000). Collision induced dissociation (CID) energy was set to 60% and collision gas was 50%. Dynamic exclusion was set to 10 s.

Improved detection of cross-linked samples was carried out by collecting fractions from three columns. Each fraction was collected using the previous HPLC system. The fractions were analyzed in a Dionex UltiMate 3000 nano-UHPLC system coupled with nano-ESI-linear ion trap (LIT) Thermo Velos Pro mass spectrometer (Thermo Fisher Scientific, Waltham, MA). The separation in the nano-LC-ESI-LIT was achieved by Acclaim PepMap C18 column (150 mm × 75 μm, 3 μm) in a 70 min gradient separation (flow rate: 0.300 μL/min; 70 min gradient: 0-3 min 4.0% B, 3-50 min 4.0-50.0% B, 50-50.1 min 50-100% B, 50.1-55 min 100% B, 55-55.1 min 100-4% B, 55.1-70 min 4% B; mobile phase A: 0.1% FA in water; mobile phase B: 0.1% FA in 95% acetonitrile, 5% water). MS data was obtained from 200 to 2000 m/z mass range. Data-dependent MS/MS spectra were collected from the five most abundant precursor ions upon fragmentation (charge state ≥2; isolated width of 2 Da; min signal required: 500) using CID activation with 35.0% normalized collision energy, activation Q of 0.25, and activation time of 30 ms.

5.3.5 Identification of inter-cross-linked peptides by pLink

The original MS data files were converted to mgf files by Mascot (for Shimadzu lcd files) or Proteome Discoverer (for Thermo raw files). Identification of inter-cross-linked peptides was performed using pLink1 cross-linking software.[169] The FASTA database of BSA sequence was acquired from Uniprot. The

parameters for pLink search were as follow: sample1.spectra. instrument = CID; sample1.spectra. format = mgf; mod. fixed.1 = Carbamidomethyl[C]; mod. variable.1 = Oxidation[M]; linker.name1 = BS3; noninterexport = true; evalue_max = 0.1. Mass tolerances of precursor and product ion searching are either 100 or 1000 ppm depending on the type of mass spectrometer used for detection. The other parameters were set as default. Results of the identification were annotated with b and y ions of peptide sequence and visualized by pLabel software, a component of the pLink software.

5.4 Results and discussion

5.4.1 Evaluation of the separation of cross-linked BSA in three columns

Currently, alkyl C18-bonded stationary phases are still the standard stationary phase in RPLC for separating cross-linked samples. These generally show consistent performance in separating hydrophobic compounds, but they are found not to be as efficient for certain compounds, specifically those that are more hydrophilic in nature.[176] Many alternative stationary phases have been developed to improve retention and selectivity of targeted analytes.[177] These generally rely on the introduction of addition functional units into the stationary phase that can then interact with analytes through a myriad of intermolecular forces. Phenyl-based stationary phases were developed to provide π-π interaction mechanisms, as well as to generate separative driving forces from polarity and shape selectivity.[178,179] Biphenyl stationary phases have been shown to exhibit increased retention and enhanced selectivity toward polarizable or hydrophilic aromatic compounds. Fluorophenyl stationary phases have been shown to provide increased retention for basic hydrophilic compounds. Additional interactive properties of the fluorophenyl column were previously investigated in the applications of off-line fractionations on Hela cells,[180] orthogonality in 2D-LC separations,[181] and separation of metabolites in urine samples.[182] Fluorine atoms incorporated in the fluorophenyl stationary phase act as electron-withdrawing groups and induce enhanced interactions between the aromatic group and positively charged compounds. More importantly, the shape selectivity from the phenyl columns could be beneficial in separating the “H”-shaped inter-cross-linked peptides from

unmodified linear peptides or dead-end cross-linked peptides. Thus, these two columns were selected to evaluate their potential for improved separation of cross-linked peptides, relative to a traditional C18 phase. To the best of our knowledge, no prior literature reported on the separation of cross-linked peptides by biphenyl and fluorophenyl phases, though the structurally unique inter-cross-linked peptides appears appealing and promising in separation by alternative stationary phases other than C18.

Different concentrations of BSA cross-linked samples were initially investigated in HPLC-ESI-IT-TOF mass spectrometer by these three columns. Three concentrations (0.6 μg/μL, 1.7 μg/μL and 5.0 μg/μL, measured by Thermo Scientific Nanodrop 2000c instrument) of the cross-linked BSA samples were prepared in triplicate. Each replicate was injected three times for each column and performed MS acquisition as described in the Materials and methods. The numbers of inter-cross-linked peptides identified are summarized in Table 5-1, and the list is shown in Table 5-S1. The 1.7 μg/μL samples yielded the most identifications of inter-cross-linked peptides among all three columns and were further investigated to compare the performance of the three columns for separating inter-cross-linked peptides (total of nine runs for each column).

Column	**Number of inter-crosslinked identification**		
	0.6 μg/μL	**1.7 μg/μL**	**5.0 μg/μL**
C18	1.6±0.7	5.4±1.1	3.8±0.8
Biphenyl	1.8±0.4	5.1±0.9	3.1±0.8
Fluorophenyl	1.6±0.5	4.4±1.4	2.7±0.5
Total IDs	2	10	6

Table 5-1. Number of inter-cross-linked peptides identified under different concentrations

Column	Concentration (μg/μL)	number	Inter-cross-linked peptides
Biphenyl	0.6	2	LAKEYEATLEECCAK(3)-ALKAWSVAR(3)
			CCTKPESER(4)-SLGKVGTR(4)
C18	0.6	2	LAKEYEATLEECCAK(3)-ALKAWSVAR(3)
			CCTKPESER(4)-SLGKVGTR(4)
Fluorophenyl	0.6	2	LAKEYEATLEECCAK(3)-ALKAWSVAR(3)
			CCTKPESER(4)-SLGKVGTR(4)
Biphenyl	1.7	8	CCTKPESER(4)-SLGKVGTR(4):0
			CCTKPESER(4)-LSQKFPK(4):0
			LAKEYEATLEECCAK(3)-ALKAWSVAR(3):0
			LCVLHEKTPVSEK(7)-CASIQKFGER(6):0
			LFTFHADICTLPDTEKQIK(16)-KQTALVELLK(1):0
			RDTHKSEIAHR(5)-FKDLGEEHFK(2):0
			FKDLGEEHFK(2)-DTHKSEIAHR(4):0
			DTHKSEIAHR(4)-LSQKFPK(4):0
C18	1.7	8	CCTKPESER(4)-SLGKVGTR(4):0
			CCTKPESER(4)-LSQKFPK(4):0
			LAKEYEATLEECCAK(3)-ALKAWSVAR(3):0
			LCVLHEKTPVSEK(7)-CASIQKFGER(6):0
			LFTFHADICTLPDTEKQIK(16)-KQTALVELLK(1):0
			RDTHKSEIAHR(5)-FKDLGEEHFK(2):0
			VHKECCHGDLLECADDR(3)-ALKAWSVAR(3):0
			FKDLGEEHFK(2)-DTHKSEIAHR(4):0
Fluorophenyl	1.7	7	CCTKPESER(4)-SLGKVGTR(4):0
			CCTKPESER(4)-LSQKFPK(4):0
			LAKEYEATLEECCAK(3)-ALKAWSVAR(3):0
			LCVLHEKTPVSEK(7)-CASIQKFGER(6):0
			RDTHKSEIAHR(5)-FKDLGEEHFK(2):0
			FKDLGEEHFK(2)-DTHKSEIAHR(4):0
			KVPQVSTPTLVEVSR(1)-HKPKATEEQLK(4):0
Biphenyl	5	5	AEFVEVTKLVTDLTK(8)-ALKAWSVAR(3):0
			CCTKPESER(4)-LSQKFPK(4):0
			CCTKPESER(4)-SLGKVGTR(4):0
			FKDLGEEHFK(2)-DTHKSEIAHR(4):0
			RDTHKSEIAHR(5)-FKDLGEEHFK(2):0
C18	5	5	CCTKPESER(4)-LSQKFPK(4):0
			CCTKPESER(4)-SLGKVGTR(4):0
			FKDLGEEHFK(2)-DTHKSEIAHR(4):0
			RDTHKSEIAHR(5)-FKDLGEEHFK(2):0
			PLLEKSHCIAEVEKDAIPENLPPLTADFAEDK(5)-DVCKNYQEAKDAFLGSFLYEYSR(10):0
Fluorophenyl	5	3	CCTKPESER(4)-SLGKVGTR(4):0
			FKDLGEEHFK(2)-DTHKSEIAHR(4):0
			RDTHKSEIAHR(5)-FKDLGEEHFK(2):0

Table 5-S1. List of inter-cross-linked peptides identified in cross-linked samples of different concentrations

Total ion current (TIC) chromatograms obtained from 1.7 μg/μL samples by three columns are shown in Figure 5-1. Chromatograms of three biological replicates from the same column are consistent in peak distributions, confirming the reproducibility of the cross-linking method and column performances. When comparing the separative performances of each column based on chromatograms, most peaks eluted within the first 30 min for the C18 and biphenyl columns, while the peaks were more broadly distributed and retained until 35 min on the fluorophenyl phase, using the same mobile phase conditions. The fluorophenyl phase showed unique retention and separative pattern, enabling elongated retention for certain peptides in the cross-linked BSA sample over a longer time range.

Table 5-2 summarizes six inter-cross-linked peptides that were mutually identified from the pLink analysis by all three columns. Comparing the retention times of these inter-cross-linked peptides, they were eluted around the same retention times in C18 and biphenyl columns. These inter-cross-linked peptides were slightly less retained on the biphenyl column compared to the conventional C18 column, with a maximum difference of 0.8 min. In the separation on the fluorophenyl column, the peptides were significantly more retained compared with the C18 and biphenyl phases. The retentions of the cross-linked peptides were increased by between 2.5 min to more than 4 min in retention time in the fluorinated phase. Similar retention shifts of non-cross-linked peptides (unmodified and dead-end cross-linked peptides) were also observed (Table 5-S2). From the differences in retention times we find that the inter-cross-linked peptides are retained in C18 and biphenyl columns similarly, whereas they exhibited more interactions in fluorophenyl column. This could be related to the charge-based interactions between the negative dipole in fluorine and positively charged inter-cross-linked peptides. The fact that inter-cross-linked peptides are more interactive with the fluorophenyl phase could be beneficial for improving their separation and identification, relative to the use of a traditional C18 column.

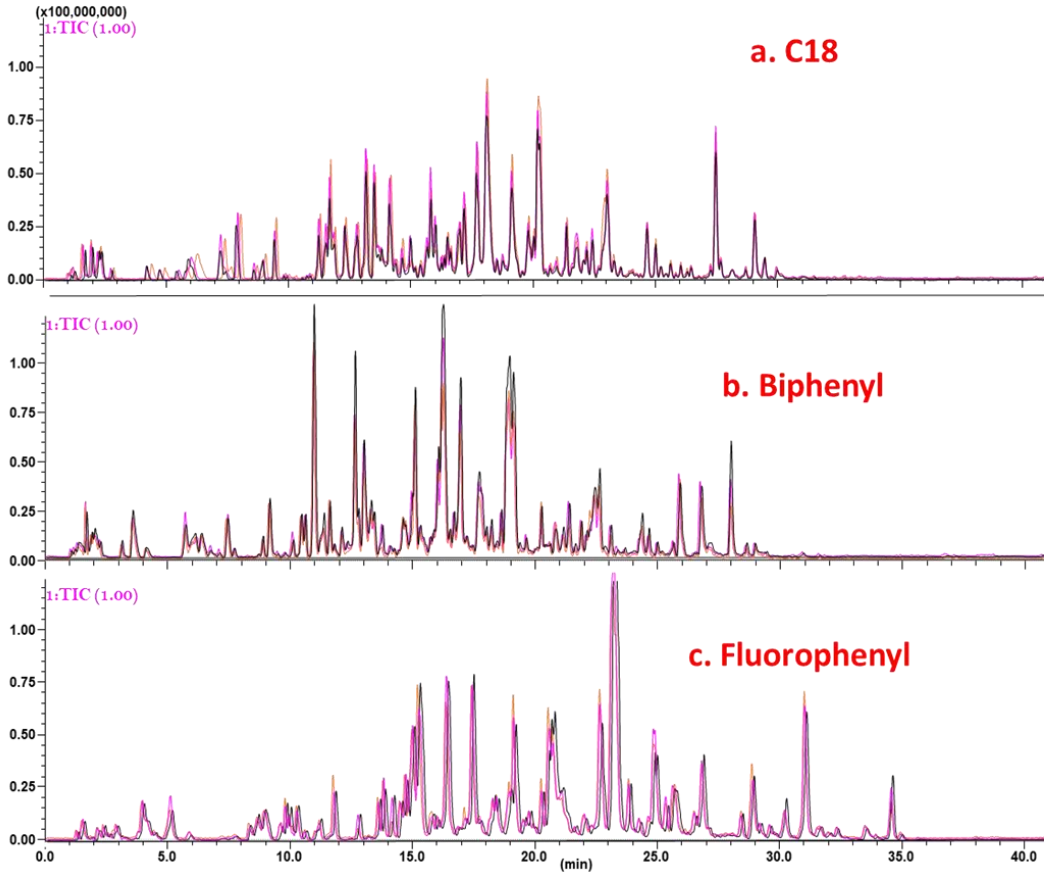


Figure 5-1. Chromatograms of cross-linked BSA sample

(a) C18 column, (b) biphenyl column and (c) fluorophenyl column for 3 replicates (Pink: replicate1, Orange: replicate2, Black: replicate3).

Inter-crosslinked peptides		Biphenyl	C18	Fluorophenyl
CC-SL	CCTKPESER(4)-SLGKVGTR(4):0	13.3 min	13.5 min	16.0 min
CC-LS	CCTKPESER(4)-LSQKFPK(4):0	14.6 min	14.6 min	17.3 min
LA-AL	LAKEYEATLEECCAK(3)-ALKAWSVAR(3):0	20.6 min	20.8 min	24.9 min
LC-CA	LCVLHEKTPVSEK(7)-CASIQKFGER(6):0	18.0 min	18.6 min	21.9 min
LF-KQ	LFTFHADICTLPDTEKQIK(16)-KQTALVELLK(1):0	24.7 min	25.8 min	N/A
RD-FK	RDTHKSEIAHR(5)-FKDLGEEHFK(2):0	18.0 min	18.6 min	22.0 min
VH-AL	VHKECCHGDLLECADDR(3)-ALKAWSVAR(3):0	17.8 min	N/A	N/A
FK-DT	FKDLGEEHFK(2)-DTHKSEIAHR(4):0	15.0 min	15.8 min	18.9 min
KV-HK	KVPQVSTPTLVEVSR(1)-HKPKATEEQLK(4):0	N/A	N/A	20.0 min
DT-LS	DTHKSEIAHR(4)-LSQKFPK(4):0	N/A	14.9 min	N/A

Table 5-2. Retention time comparison of the identified inter-cross-linked peptides (N/A implies that the peptide was not detected in certain column)

Non-cross-linked peptides	Avg. Retention time (min)			Retention time increase in percent	
	BP	C18	FP	C18 from BP	FP from C18
AEFVEVTK:0	13.03	14.08	16.40	8.05%	16.43%
CASIQK:0	2.19	2.74	4.45	24.97%	62.22%
CCAADDKEACFAVEGPK:0	14.96	14.83	17.61	-0.89%	18.74%
CCTESLVNR:0	10.56	11.80	13.59	11.78%	15.17%
DAFLGSFLYEYSR:0	28.23	28.84	31.42	2.14%	8.97%
DAIPENLPPLTADFAEDKDVCK:0	22.23	22.79	25.71	2.55%	12.82%
DDPHACYSTVFDK:0	15.95	16.21	18.80	1.66%	15.97%
DLGEEHFK:0	11.33	12.35	15.25	9.06%	23.47%
ECCDKPLLEK:0	10.83	11.46	14.52	5.82%	26.65%
ECCHGDLLECADDR:0	13.38	14.60	17.05	9.13%	16.78%
ECCHGDLLECADDRADLAK:0	14.97	16.38	19.06	9.43%	16.35%
ETYGDMADCCEK:0	12.33	10.66	12.41	-13.53%	16.43%
FKDLGEEHFK:0	12.91	13.86	16.95	7.34%	22.30%
GLVLIAFSQYLQQCPFDEHVK:0	27.99	28.96	34.49	3.45%	19.10%
HLVDEPQNLIK:0	15.18	16.35	19.25	7.74%	17.70%
HPEYAVSVLLR:0	18.17	19.93	24.51	9.67%	23.00%
HPYFYAPELLYYANK:0	24.31	24.59	30.15	1.19%	22.57%
IETMR:0	6.37	7.77	10.27	22.07%	32.14%
KVPQVSTPTLVEVSR:0	16.95	17.58	20.62	3.69%	17.29%
LCVLHEK:0	9.13	11.21	14.61	22.77%	30.39%
LFTFHADICTLPDTEK:0	21.83	22.13	26.44	1.38%	19.47%
LGEYGFQNALIVR:0	22.58	22.98	26.89	1.75%	17.01%
LKECCDKPLLEK:0	10.90	11.67	15.20	7.03%	30.28%
LVNELTEFAK:0	18.85	20.12	23.25	6.75%	15.53%
LVTDLTK:0	10.99	13.03	14.93	18.55%	14.56%
MPCTEDYLSLILNR:0	25.77	27.28	30.96	5.86%	13.49%
NECFLSHK:0	10.37	11.58	14.80	11.63%	27.86%
NECFLSHKDDSPDLPK:0	14.84	15.37	18.52	3.54%	20.51%
QNCDQFEK:0	8.82	8.95	11.20	1.48%	25.14%
RHPEYAVSVLLR:0	16.26	18.20	23.26	11.94%	27.83%

RHPYFYAPELLYYANK:0	22.32	22.79	28.95	2.10%	27.00%
RPCFSALTPDETYVPK:0	19.01	19.13	22.60	0.64%	18.14%
SEIAHR:0	1.64	1.84	3.93	12.68%	113.07%
SHCIAEVEK:0	7.44	9.44	11.78	26.87%	24.90%
SHCIAEVEKDAIPENLPPLTADFAEDKDVCK:0	21.34	22.32	25.53	4.60%	14.41%
SLHTLFGDELCK:0	19.06	20.44	24.87	7.21%	21.69%
TCVADESHAGCEK:0	6.11	7.93	9.73	29.70%	22.72%
TPVSEK:0	1.93	2.30	4.14	18.92%	79.78%
TVMENFVAFVDK:0	24.94	25.96	29.55	4.06%	13.83%
YICDNQDTISSK:0	11.52	12.28	13.47	6.61%	9.66%
YNGVFQECCQAEDK:0	15.39	15.28	17.66	-0.72%	15.54%
ADLAKYICDNQDTISSK(5):0	21.03	21.88	24.30	4.04%	11.04%
ALKAWSVAR(3):0	21.31	21.72	25.70	1.94%	18.33%
DTHKSEIAHR(4):0	11.76	12.77	15.02	8.53%	17.66%
FKDLGEEHFK(2):0	19.10	19.78	23.75	3.53%	20.06%
KQTALVELLK(1):0	23.01	24.85	28.38	8.03%	14.18%
KVPQVSTPTLVEVSR(1):0	21.26	21.64	23.88	1.80%	10.34%
LCVLHEKTPVSEK(7):0	17.58	18.47	21.86	5.07%	18.37%
LKHLVDEPQNLIK(2):0	20.21	21.30	24.58	5.39%	15.40%
LSQKFPK(4):0	17.70	17.94	21.13	1.34%	17.79%
SLGKVGTR(4):0	15.36	16.38	18.51	6.63%	13.01%
SLHTLFGDELCKVASLR(12):0	27.01	28.51	33.49	5.53%	17.49%
TCVADESHAGCEKSLHTLFGDELCK(13):0	22.07	23.10	26.40	4.66%	14.32%
VHKECCHGDLLECADDR(3):0	16.03	16.97	20.10	5.84%	18.44%
VHKECCHGDLLECADDRADLAK(3):0	16.69	17.76	21.06	6.40%	18.61%
			Average	**7.26%**	**22.40%**

Table 5-S2. Retention comparison of non-cross-linked peptides

From pLink software, search results of both inter-cross-linked and non-cross-linked peptides (unmodified and dead- end cross-linked peptides) were reported and compared in evaluating the column performance. In the data-dependent mode of MS analysis, the more abundant precursor ions have higher chances of being selected for the fragmentation. Thus, the better a certain type of peptides is separated within the column, the more identifications can be achieved and reported. We expect the column with the most inter-cross-linked peptide identifications having the best separation of inter-cross-linked peptides. In contrast, the naturally more abundant non-cross-linked peptides are undesired because they act as “matrix” in the sample and coelute with more informative inter-cross-linked peptides, competing for fragmentation. The less identification of non-cross-linked peptides shows that the “matrix” is more depleted, and the cross-linked peptides could be relatively preferred. When we investigated the number of non-cross-linked peptides identified in total 9 runs from each of three columns, the biphenyl column had 48.0 ± 2.8, the C18 column achieved 52.0 ± 5.0 while fluorophenyl column only had 40.1 ± 3.3 in non-cross-linked peptide identifications (data not shown). The number of non-cross-linked peptides identified in the fluorophenyl

column is statistically less than the other two columns which helps to reduce the detection and fragmentation of undesired non-cross-linked peptides.

In terms of inter-cross-linked results, the number of identifications for the biphenyl, C18, and fluorophenyl column in three replicates from nine runs were calculated to be 5.1 ± 0.9, 5.4 ± 1.1, and 4.4 ± 1.2, respectively. Considering the deviation in relation to number of identified peptides, there is no significant advantage for any column in obtaining more inter-cross-linked identifications, presumably due to the low sensitivity of MS instrument. Therefore, this result is not conclusive for determining the column performance. A nano-LC-ESI system would be preferred for separating and detecting the low abundant cross-linked samples with high dynamic range. However, there are no commercially available capillary biphenyl or fluorophenyl columns compatible with the nano-LC system available to date. We are in the process of constructing the column with fluorophenyl phase.

5.4.2 Fraction Collections and Analysis by Nano-UHPLC-ESI-LIT

According to the previous analysis of inter-cross-linked results, three columns performed almost identical in identifying inter-cross-linked peptides under regular HPLC-ESI-IT-TOF mass spectrometer, regardless of the retention differences on the different phases. The evaluation of columns was mainly hampered by low sensitivity considering the numbers of inter-cross-linked peptides identified were less than previously reported.[153,183,184] Therefore, system with higher sensitivity is necessary for more appropriate evaluation of performances of three columns. We improved the sensitivity of detection by collecting the fractions from each column and analyzing the fractions in a more sensitive nano-UHPLC-ESI-LIT instrument. During the sample fractionation, the MS setting in LC-ESI-IT-TOF using the three columns were maintained the same. Based on the chromatograms (Figure 5-1), fractions were collected on the retention times containing the time span when the inter-cross-linked peptides were eluted from the columns. For C18 and biphenyl columns, most peaks were eluted within the first 30 min. Thus, the fractions were collected from 0 to 31 min of the LC runs for C18 and biphenyl columns. To reduce analysis time of fractions, the collection of

fractions was performed by combining first 10 min eluent as one fraction and every 3 min interval for retention time 10-31 min. A total of 8 fractions (0-10 min; 10-13 min; 13-16 min; 16-19 min; 19-22 min; 22-25 min; 25-28 min; 28-31 min) were collected from separations performed on the C18 and biphenyl columns. Fraction collection for the fluorophenyl column was performed in the same fashion, except that the collection time was from 0 to 37 min in accordance with the extended retention times. A total of 10 fractions (0-10 min; 10-13 min; 13-16 min; 16-19 min; 19-22 min; 22-25 min; 25-28 min; 28-31 min; 31-34 min; 34-37 min) were collected for the fluorophenyl column. These fractions were dried, reconstituted, and analyzed in the nano-UHPLC-ESI-LIT system as described in the method.

The increase of sensitivity was reflected in the increasing matching of fragments for confident identification of inter-cross-linked peptides (Figure 5-2). After nano-LCMS analysis and pLink search, the number of inter-cross-linked peptides identified from each fraction of each column was determined. Under the same searching parameters, the number of inter-cross-linked peptides identified had increased significantly after collecting fractions. By comparing the number of unique peptides, the fractions from the C18 column resulted in the identification of 59 unique inter-cross-linked peptides, the fractions from the biphenyl column returned 82 unique inter-cross-linked peptides, and fractions from the fluorophenyl column resulted in 102 unique inter-cross-linked peptides (Figure 5-3a).[185] Both the biphenyl and fluorophenyl phases returned more inter-cross-linked peptides than the C18 column.

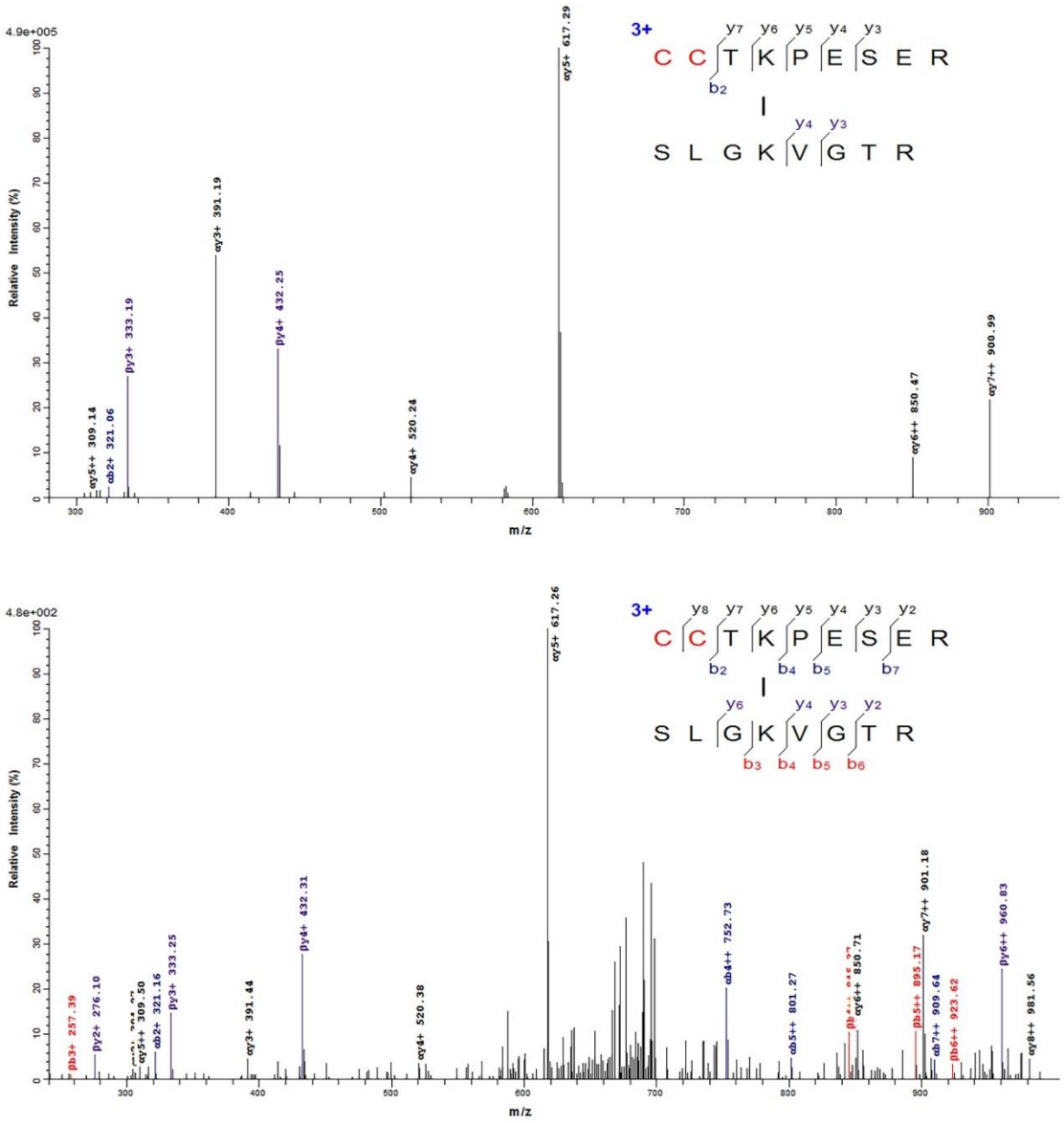


Figure 5-2. The matching of b and y ions of inter-cross-linked peptide CCTKPESER(4)-SLGKVGTR(4) from regular LC-ESI-MS/MS (top spectrum) and nano LC-ESI-MS/MS (bottom spectrum), extracted from pFind software

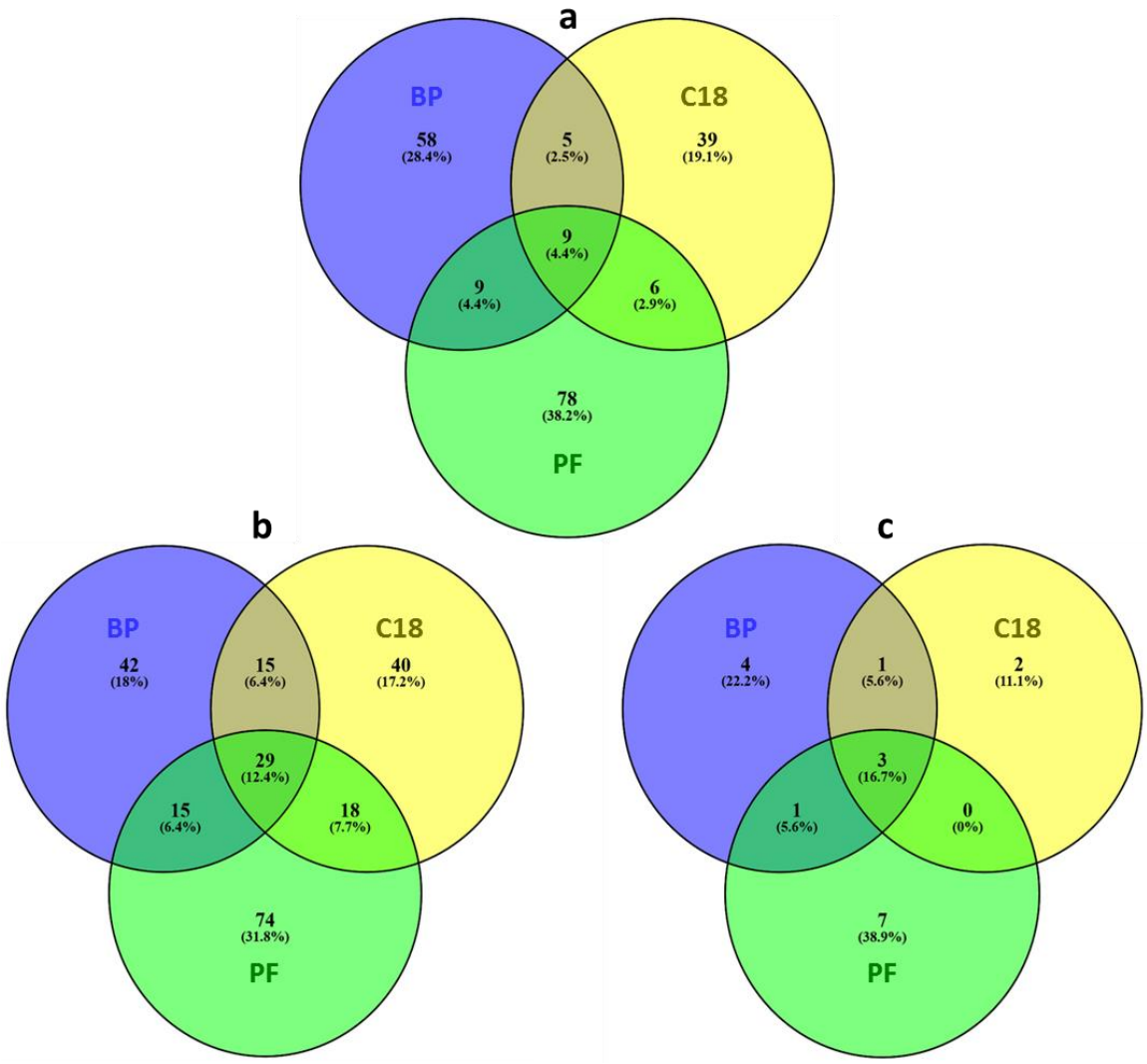


Figure 5-3. Total number of inter-cross-linked peptides identified by three columns after collecting fractions with different parameters and filtering (a. search with 100 ppm precursor and fragment ion tolerance; b. search with 1000 ppm precursor and fragment ion tolerance; c. search with 1000 ppm precursor and fragment ion tolerance with re-filtering with confident parameters in pLink

However, a very low percentage of the total inter-cross-linked peptides is mutually identified by all three columns as seen in the Venn diagram. We suspect the reason could be the relatively low resolution in the ion-trap-type mass analyzer, and if using the same searching parameters as TOF instrument, some inter-cross-linked peptide candidates could be filtered by the limited tolerance of precursor and fragment ions. Thus, we reanalyze the same data with a higher tolerance of 1000 ppm for precursor and fragment ion matching. The overall identified cross-linked peptides for all columns had increased from 204 to 233, and more importantly, the mutually identified inter-cross-linked peptides increased to 12.4% in percentage from 4.4% in the first search (Figure 5-3a, b). Fluorophenyl column still yields the greatest number of inter-cross-linked identifications, identified 136 compared to 102 from C18 and 101 from biphenyl column, outperforming the other two columns by the increase of more than 30% in inter-cross-linked peptides identified.

During each analysis by the pLink software, multiple result files are generated representing different degrees of filtering by the default 5% FDR. One result of the search is generated after filtering each spectra file independently, from which the reported inter-cross-linked numbers are shown in Figure 5-3a, b. There is also a refiltered result file combining all of the independent filtered result files and generating a combined result file which provides more confident identifications. When comparing the combined results from each column, the identification numbers are largely reduced. Only 18 inter-cross-linked peptides in total are reported in the "high confident" results from all columns studied. These identified peptides are listed in Table 5-3. Regardless of the stricter filter, the greatest number of inter-cross-linked peptides came from the fluorophenyl column, reporting 11 out of 18 totals identified against 6 from C18 column and 9 from biphenyl column. Furthermore, for validating the plausibility in forming the reported inter-cross-linked peptides and evaluating the columns in identifying "true" cross-linked peptides, distance between two reacted lysines of each cross-linked peptide was calculated in reference to the crystal structure of BSA from the protein data bank (entry: 3V03, Figure 5-S1). The maximum distance between two α-carbons of BS^3 cross-linked lysine is reported to be 27.4 Å, taking into account the span of BS^3 cross-linker, two lysine

side chains, and side-chain flexibility.[173] A total of seven of the fluorophenyl identified peptides are within the calculated limit of BS^3 cross-linker with 4 for biphenyl column and 3 for C18 column (Table 5-3). This confirms that fluorophenyl column is able to achieve better separation and identification of the "true" cross-linked peptides.

Column	Inter-cross-linked peptides	Cross-linked distance(Å)
BP	CCTKPESER(4)-SLGKVGTR(4):0	**13.4**
BP	LAKEYEATLEECCAK(3)-ALKAWSVAR(3):0	**13.6**
BP	GACLLPKIETMR(7)-SLGKVGTR(4):0	**19.7**
BP	CCTKPESER(4)-LSQKFPK(4):0	**20.2**
BP	CCTKPESER(4)-PKATEEQLK(2):0	35.3
BP	PKATEEQLK(2)-LSQKFPK(4):0	47.1
BP	KFWGK(1)-HKPK(2):0	51.7
BP	PKATEEQLK(2)-KFWGK(1):0	54.4
BP	PKATEEQLK(2)-LKECCDK(2):0	58.5
C18	CCTKPESER(4)-SLGKVGTR(4):0	**13.4**
C18	LAKEYEATLEECCAK(3)-ALKAWSVAR(3):0	**13.6**
C18	CCTKPESER(4)-LSQKFPK(4):0	**20.2**
C18	TPVSEKVTK(6)-LSQKFPK(4):0	35.5
C18	AEFVEVTKLVTDLTK(8)-HKPK(2):0	55.1
C18	PKATEEQLK(2)-LKECCDK(2):0	58.5
FP	CCTKPESER(4)-SLGKVGTR(4):0	**13.4**
FP	LAKEYEATLEECCAK(3)-ALKAWSVAR(3):0	**13.6**
FP	NYQEAKDAFLGSFLYEYSR(6)-ALKAWSVAR(3):0	**14.4**
FP	ALKAWSVARLSQK(3)-LSQKFPK(4):0	**15.1**
FP	ALKAWSVAR(3)-LSQKFPK(4):0	**15.1**
FP	CCTKPESER(4)-LSQKFPK(4):0	**20.2**
FP	ADEKKFWGK(5)-KFWGK(1):0	**27.2**
FP	DTHKSEIAHR(4)-ADEKK(4):0	31.3
FP	PKATEEQLK(2)-EKVLASSAR(2):0	38.4
FP	PKATEEQLK(2)-LSQKFPK(4):0	47.1
FP	RDTHKSEIAHR(5)-QIKK(3):0	56.4

Table 5-3. Combined results of inter-cross-linked peptide and their cross-linked distances. Plausible cross-linked distances are bolded

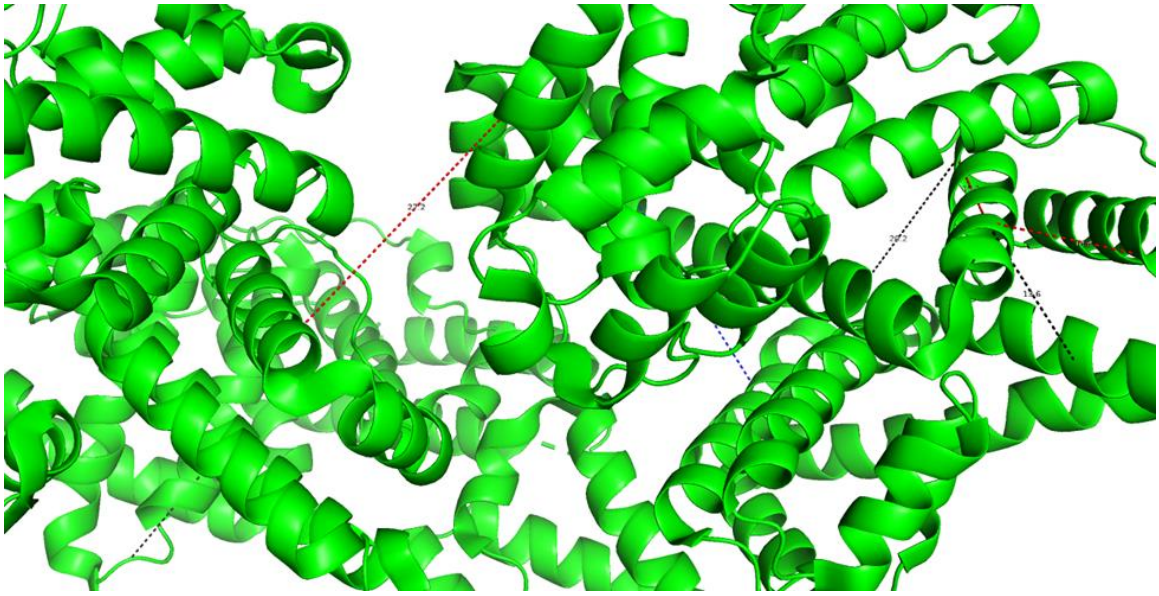


Figure 5-S1. 3D structure of BSA with inter-cross-linked peptides identified (black: found in all 3 columns; blue: only found in biphenyl column; red: only found in fluorophenyl column)

In addition, myoglobin from equine heart (around 17 kDa molecular weight) was also studied by the fractionation method as in BSA fractionation. There is no remarkable improvement in biphenyl column compared to C18 column; thus, C18 column and fluorophenyl columns are compared by the "high confident" results from pLink software. Fluorophenyl continues to outperform the C18 column by identifying 5 inter-cross-linked peptides, from which 4 out of 5 within the maximum distance of BS3 cross-linking (Figure 5-S2). For the C18 column, only one inter-cross-linked peptide which cross-linked in the same Lys147 position (possibly from dimer) was reported (Table 5-S3). The superiority of fluorophenyl column to C18 column in cross-linking separation is further confirmed.

Nano-LC-MS improved the overall sensitivity to address low abundant cross-linked peptides, which assisted in appraising the three columns in separating and identifying inter-cross-linked peptides. From the experimental results, both phenyl-based columns are proven to be comparable with the C18 column. Particularly, the fluorophenyl column exhibits significant advantages in identifying important inter-cross-

linked peptides and “stacking” the interfering non-cross-linked peptides. These stationary phases are proven suitable to substitute for the commonly used linear alkyl stationary phase in separating and enriching cross-linked peptides, especially for their applications in online highly sensitive nano-LC-ESI-MS systems.

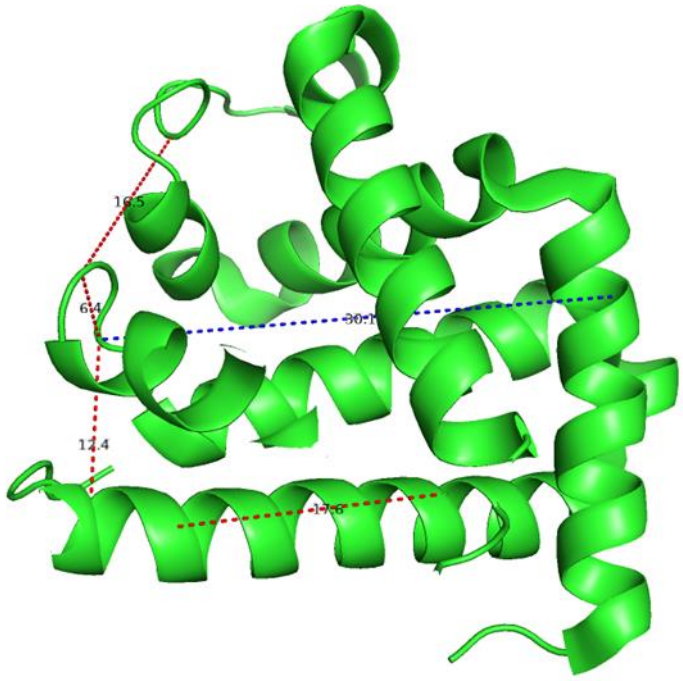


Figure 5-S2. 3D structure of myoglobin (horse) with inter-cross-linked identified (blue: only found in C18 column; red: only found in fluorophenyl column)

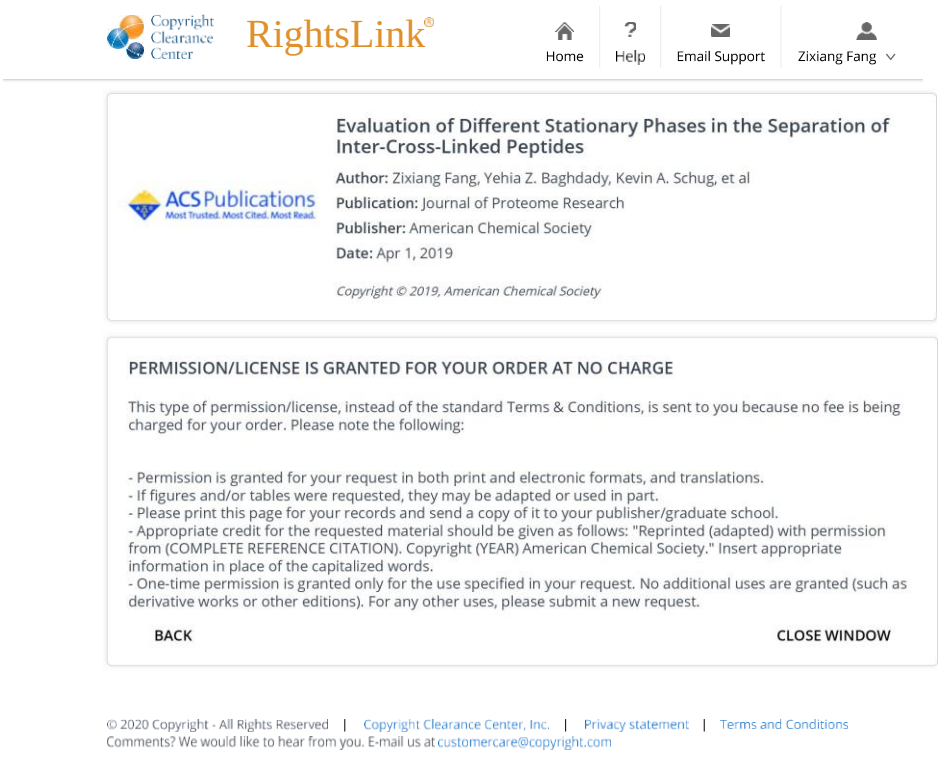
Column	Inter-cross-linked peptides	Cross-linked distance(Å)
C18	NDIAAKYKELGFQG(6)-YKELGFQG(2):0	dimer
FP	KKGHHEAELK(1)-NDIAAKYK(6):0	17.6
FP	PLAQSHATKHK(9)-FKHLK(2):0	16.5
FP	PLAQSHATKHK(9)-HKIPIK(2):0	6.4
FP	YKELGFQG(2)-HKIPIK(2):0	12.4
FP	MGLSDGEWQQVLNVWGKVEADIAGHGQEVLIR(17)-HKIPIK(2):0	30.1

Table 5-S3. Inter-cross-linked peptides from myoglobin identified by C18 and fluorophenyl columns

5.5 Conclusion

Although various approaches had been applied for enriching inter-cross-linked peptides for their improved identification, it is still necessary to improve and optimize the existing enrichment methods. Two commercialized columns (biphenyl and fluorophenyl), which provided different selectivities in separating polar and hydrophilic compounds, were compared with a versatile C18 column in separating the protein digests of BS^3 cross-linked BSA. It was shown that the fluorophenyl column has extended retention toward the inter-cross-linked peptides. Through the analysis of fractions from three columns, it is shown that the fluorophenyl column increased inter-cross-linked peptide identifications while it decreased the non-cross-linked peptide identification, possibly because of the additional electrostatic interactions between cross-linked peptides and the fluorophenyl stationary phase. The potential of the fluorophenyl column as a competitive substitute in chemical cross-linking research was demonstrated. In future, fluorophenyl based capillary columns will be investigated and evaluated in the cross-linking study with higher biological complexity; this will also circumvent the laborious fractionations and possible uncertainty and errors in this process.

5.6 Copyright and permission



Copyright Clearance Center RightsLink®

Home | Help | Email Support | Zixiang Fang

ACS Publications
Most Trusted. Most Cited. Most Read.

Evaluation of Different Stationary Phases in the Separation of Inter-Cross-Linked Peptides

Author: Zixiang Fang, Yehia Z. Baghdady, Kevin A. Schug, et al
Publication: Journal of Proteome Research
Publisher: American Chemical Society
Date: Apr 1, 2019

Copyright © 2019, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK CLOSE WINDOW

Chapter 6

General Conclusions

The identifications of two important PTMs, prenylation and proteolysis, are improved by targeted chemical modifications for enrichment and systematically developed MS-cleavable strategies. For the identification of prenylation, prenyl peptide and oxidized prenyl peptide generate two unique signature fragments in the CID-MS^2 spectra for unambiguously identifying the type and modification site of prenylation within the protein, with the sequence of the modified peptide revealed by NLMS3 strategy. This targeted strategy can successfully identify native prenylated proteins from mammalian cell lysate even without prenylation enrichment during sample preparation. But for the purpose of reducing the sample complexity and improving the number of prenylated proteins identified, an efficient enrichment strategy remains to be beneficial. Therefore, an enrichment approach is achieved through biotinylation of prenylated peptides after epoxidation of the prenyl moieties. Preliminary results of enrichment study show that the C-terminal prenylated peptide of protein KRas4B is selectively labeled by this approach, enriched from the peptide pool of digested standard proteins, and confidently identify by tandem MS fragmentations.

Targeted enrichment and identification of proteolytic protein N-terminus is also investigated in this dissertation. Biotin-containing reagent PFP-Rink-biotin is utilized in reacting with primary amines of original or proteolytic protein N-terminus. The derivatized protein N-terminus are enriched by positive selection and identified through the generation of a fixed-mass RI, along with further fragmentation of the complimentary fragment for the sequencing of N-terminal peptides. This enables the targeted study of enrichment and identification of N-terminal sequence of standard proteins. Moreover, an in-house bioinformatic software is developed to automatically locate and validate the identified N-terminal PSMs from numerous results generated by regular shotgun proteomics searches with large protein database. Proteins from *E. coli* cell lysate were extracted, and 42 native protein N-terminus confidently identified by this strategy.

In terms of improving the current cross-linking/MS strategy for deciphering protein structures and protein interactions, proof-of-concept study of BSA cross-linked by commercial amine reactive cross-linker BS^3 was performed and the separation of cross-linked products were evaluated by comparing the number of inter-cross-linked peptides identified from 3 different stationary phases: C18, biphenyl and fluorophenyl. Both biphenyl and fluorophenyl columns show more inter-cross-linked peptides identified compared to the commonly used C18 column, with fluorophenyl column outperforms the other two by further increasing the retentions of inter-cross-linked peptides, potentially thanks to additional affinity of electrostatic interactions with highly charged inter-cross-linked peptides.

In summary, this dissertation explored the targeted identifications of low-abundance PTMs and protein interactions from complex matrices and large-scale biological samples. The research area discussed in this dissertation is constantly and continuously being developed and innovated in the hope of answering important biological questions and uncovering the underlying mechanisms in relation with diseases. We hope our research can shed lights on the MS-based strategies that are available for the study of PTMs and protein interactions. Although my graduate research only covers a very small subset of the grand research endeavors in the field, our lab will keep exploring large-scale applications of these approaches and simultaneously developing new methods to address new challenges and contribute to the future development and applications in large-scale proteomics study.

References

(1) Muñoz, J.; Heck, A. J. R. From the Human Genome to the Human Proteome. *Angew. Chemie Int. Ed.* **2014**, *53* (41), 10864–10866.

(2) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chemie Int. Ed.* **2005**, *44* (45), 7342–7372.

(3) Controlled vocabulary of posttranslational modifications (PTM) https://www.uniprot.org/docs/ptmlist.txt (accessed Apr 9, 2020).

(4) Hogg, P. J. Disulfide Bonds as Switches for Protein Function. *Trends Biochem. Sci.* **2003**, *28* (4), 210–214.

(5) Deribe, Y. L.; Pawson, T.; Dikic, I. Post-Translational Modifications in Signal Integration. *Nat. Struct. Mol. Biol.* **2010**, *17* (6), 666–672.

(6) Xu, H.; Wang, Y.; Lin, S.; Deng, W.; Peng, D.; Cui, Q.; Xue, Y. PTMD: A Database of Human Disease-Associated Post-Translational Modifications. *Genomics. Proteomics Bioinf.* **2018**, *16* (4), 244–251.

(7) Wang, M.; Casey, P. J. Protein Prenylation: Unique Fats Make Their Mark on Biology. *Nat. Rev. Mol. Cell Biol.* **2016**, *17* (2), 110–122.

(8) Zhang, F. L.; Casey, P. J. Protein Prenylation: Molecular Mechanisms and Functional Consequences. *Annu. Rev. Biochem.* **1996**, *65* (1), 241–269.

(9) Clarke, S. Protein Isoprenylation and Methylation at Carboxyl-Terminal Cysteine Residues. *Annu. Rev. Biochem.* **1992**, *61* (1), 355–386.

(10) Schafer, W. R.; Rine, J. Protein Prenylation: Genes, Enzymes, Targets, and Functions. *Annu. Rev. Genet.* **1992**, *26* (1), 209–237.

(11) Kamiya, Y.; Sakurai, A.; Tamura, S.; Takahashi, N.; Abe, K.; Tsuchiya, E.; Fukui, S.; Kitada, C.; Fujino, M. Structure of Rhodotorucine A, a Novel Lipopeptide, Inducing Mating Tube Formation in Rhodosporidium Toruloides. *Biochem. Biophys. Res. Commun.* **1978**, *83* (3), 1077–1083.

(12) Schmidt, R. A.; Schneider, C. J.; Glomset, J. A. Evidence for Post-Translational Incorporation of a Product of Mevalonic Acid into Swiss 3T3 Cell Proteins. *J. Biol. Chem.* **1984**, *259* (16), 10175–10180.

(13) Wolda, S. L.; Glomset, J. A. Evidence for Modification of Lamin B by a Product of Mevalonic Acid. *J. Biol. Chem.* **1988**, *263* (13), 5997–6000.

(14) Marshall, C. J. Protein Prenylation: A Mediator of Protein-Protein Interactions. *Science* **1993**, *259* (5103), 1865–1866.

(15) Maurer-Stroh, S.; Koranda, M.; Benetka, W.; Schneider, G.; Sirota, F. L.; Eisenhaber, F. Towards Complete Sets of Farnesylated and Geranylgeranylated Proteins. *PLoS Comput. Biol.* **2007**, *3* (4), e66.

(16) Blanden, M. J.; Suazo, K. F.; Hildebrandt, E. R.; Hardgrove, D. S.; Patel, M.; Saunders, W. P.; Distefano, M. D.; Schmidt, W. K.; Hougland, J. L. Efficient Farnesylation of an Extended C-Terminal C(x)3X Sequence Motif Expands the Scope of the Prenylated Proteome. *J. Biol. Chem.* **2018**, *293* (8), 2770–2785.

(17) UniProtKB 2020_05 prenylation results. https://www.uniprot.org/uniprot/?query=reviewed:yes keyword:%22Prenylation [KW-0636]%22 (accessed Apr 13, 2020).

(18) Berndt, N.; Hamilton, A. D.; Sebti, S. M. Targeting Protein Prenylation for Cancer Therapy. *Nat. Rev. Cancer* **2011**, *11* (11), 775–791.

(19) Stephen, A. G.; Esposito, D.; Bagni, R. K.; McCormick, F. Dragging Ras Back in the Ring. *Cancer*

Cell **2014**, *25* (3), 272–281.

(20) Shah, S.; Brock, E. J.; Ji, K.; Mattingly, R. R. Ras and Rap1: A Tale of Two GTPases. *Semin. Cancer Biol.* **2019**, *54*, 29–39.

(21) Nissim, S.; Leshchiner, I.; Mancias, J. D.; Greenblatt, M. B.; Maertens, O.; Cassa, C. A.; Rosenfeld, J. A.; Cox, A. G.; Hedgepeth, J.; Wucherpfennig, J. I.; Kim, A. J.; Henderson, J. E.; Gonyo, P.; Brandt, A.; Lorimer, E.; Unger, B.; Prokop, J. W.; Heidel, J. R.; Wang, X.-X.; Ukaegbu, C. I.; Jennings, B. C.; Paulo, J. A.; Gableske, S.; Fierke, C. A.; Getz, G.; Sunyaev, S. R.; Wade Harper, J.; Cichowski, K.; Kimmelman, A. C.; Houvras, Y.; Syngal, S.; Williams, C.; Goessling, W. Mutations in RABL3 Alter KRAS Prenylation and Are Associated with Hereditary Pancreatic Cancer. *Nat. Genet.* **2019**, *51* (9), 1308–1314.

(22) Ntai, I.; Fornelli, L.; DeHart, C. J.; Hutton, J. E.; Doubleday, P. F.; LeDuc, R. D.; van Nispen, A. J.; Fellers, R. T.; Whiteley, G.; Boja, E. S.; Rodriguez, H.; Kelleher, N. L. Precise Characterization of KRAS4b Proteoforms in Human Colorectal Cells and Tumors Reveals Mutation/Modification Cross-Talk. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (16), 4140–4145.

(23) Pylayeva-Gupta, Y.; Grabocka, E.; Bar-Sagi, D. RAS Oncogenes: Weaving a Tumorigenic Web. *Nat. Rev. Cancer* **2011**, *11* (11), 761–774.

(24) Cox, A. D.; Der, C. J.; Philips, M. R. Targeting RAS Membrane Association: Back to the Future for Anti-RAS Drug Discovery? *Clin. Cancer Res.* **2015**, *21* (8), 1819–1827.

(25) Brioschi, M.; Martinez Fernandez, A.; Banfi, C. Exploring the Biochemistry of the Prenylome and Its Role in Disease through Proteomics: Progress and Potential. *Expert Rev. Proteomics* **2017**, *14* (6), 515–528.

(26) Doucet, A.; Butler, G. S.; Rodríguez, D.; Prudova, A.; Overall, C. M. Metadegradomics: Toward in Vivo Quantitative Degradomics of Proteolytic Post-Translational Modifications of the Cancer

Proteome. *Mol. Cell. Proteomics* **2008**, *7* (10), 1925–1951.

(27) Bradshaw, R. A.; Brickey, W. W.; Walker, K. W. N-Terminal Processing: The Methionine Aminopeptidase and Nα-Acetyl Transferase Families. *Trends Biochem. Sci.* **1998**, *23* (7), 263–267.

(28) Varshavsky, A. The N-End Rule Pathway of Protein Degradation. *Genes to Cells* **1997**, *2* (1), 13–28.

(29) Varshavsky, A. N-Degron and C-Degron Pathways of Protein Degradation. *Proc. Natl. Acad. Sci.* **2019**, *116* (2), 358–366.

(30) Salvesen, G. S.; Dixit, V. M. Caspases: Intracellular Signaling by Proteolysis. *Cell* **1997**, *91* (4), 443–446.

(31) King, R. W.; Deshaies, R. J.; Peters, J.-M.; Kirschner, M. W. How Proteolysis Drives the Cell Cycle. *Science* **1996**, *274* (5293), 1652–1659.

(32) Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L.; Humphery-Smith, I. Progress with Gene-Product Mapping of the Mollicutes:Mycoplasma Genitalium. *Electrophoresis* **1995**, *16* (1), 1090–1094.

(33) Chandramouli, K.; Qian, P.-Y. Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. *Hum. Genomics Proteomics* **2009**, *1* (1).

(34) Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988**, *60* (20), 2299–2301.

(35) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246* (4926), 64–71.

(36) Karas, M.; Krüger, R. Ion Formation in MALDI: The Cluster Ionization Mechanism. *Chem. Rev.* **2003**, *103* (2), 427–440.

(37) Gaskell, S. J. Electrospray: Principles and Practice. *J. Mass Spectrom.* **1997**, *32* (7), 677–688.

(38) Wells, J. M.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Methods Enzymol.* **2005**, *402*, 148–185.

(39) Biemann, K.; Papayannopoulos, I. A. Amino Acid Sequencing of Proteins. *Acc. Chem. Res.* **1994**, *27* (11), 370–378.

(40) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528–9533.

(41) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal. Chem.* **2018**, *90* (1), 40–64.

(42) Chait, B. T. CHEMISTRY: Mass Spectrometry: Bottom-Up or Top-Down? *Science* **2006**, *314* (5796), 65–66.

(43) Kelleher, N. L. Peer Reviewed: Top-Down Proteomics. *Anal. Chem.* **2004**, *76* (11), 196 A-203 A.

(44) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down Proteomics: Facts and Perspectives. *Biochem. Biophys. Res. Commun.* **2014**, *445* (4), 683–693.

(45) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394.

(46) Shi, Y. A Glimpse of Structural Biology through X-Ray Crystallography. *Cell* **2014**, *159* (5), 995–1014.

(47) Markwick, P. R. L.; Malliavin, T.; Nilges, M. Structural Biology by NMR: Structure, Dynamics, and Interactions. *PLoS Comput. Biol.* **2008**, *4* (9), e1000168.

(48) Frank, J. Single-Particle Imaging of Macromolecules by Cryo-Electron Microscopy. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31* (1), 303–319.

(49) Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.* **2016**, *41* (1), 20–32.

(50) Fang, Z.; Baghdady, Y. Z.; Schug, K. A.; Chowdhury, S. M. Evaluation of Different Stationary Phases in the Separation of Inter-Cross-Linked Peptides. *J. Proteome Res.* **2019**, *18* (4), 1916–1925.

(51) Sinz, A. Chemical Cross-Linking and Mass Spectrometry for Mapping Three-Dimensional Structures of Proteins and Protein Complexes. *J. Mass Spectrom.* **2003**, *38* (12), 1225–1237.

(52) Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. Mass Spectrometry Identifiable Cross-Linking Strategy for Studying Protein-Protein Interactions. *Anal. Chem.* **2005**, *77* (1), 311–318.

(53) Chowdhury, S. M.; Munske, G. R.; Tang, X.; Bruce, J. E. Collisionally Activated Dissociation and Electron Capture Dissociation of Several Mass Spectrometry-Identifiable Chemical Cross-Linkers. *Anal. Chem.* **2006**, *78* (24), 8183–8193.

(54) Chowdhury, S. M.; Munske, G. R.; Ronald, R. C.; Bruce, J. E. Evaluation of Low Energy CID and ECD Fragmentation Behavior of Mono-Oxidized Thio-Ether Bonds in Peptides. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (3), 493–501.

(55) Kao, A.; Chiu, C.; Vellucci, D.; Yang, Y.; Patel, V. R.; Guan, S.; Randall, A.; Baldi, P.; Rychnovsky, S. D.; Huang, L. Development of a Novel Cross-Linking Strategy for Fast and Accurate Identification of Cross-Linked Peptides of Protein Complexes. *Mol. Cell. Proteomics* **2011**, *10* (1), M110.002212.

(56) Sinz, A. Divide and Conquer: Cleavable Cross-Linkers to Study Protein Conformation and Protein–Protein Interactions. *Anal. Bioanal. Chem.* **2017**, *409* (1), 33–44.

(57) Steen, H.; Mann, M. A New Derivatization Strategy for the Analysis of Phosphopeptides by Precursor Ion Scanning in Positive Ion Mode. *J Am Soc Mass Spectrom* **2002**, *13* (8), 996–1003.

(58) Bhawal, R. P.; Sadananda, S. C.; Bugarin, A.; Laposa, B.; Chowdhury, S. M. Mass Spectrometry Cleavable Strategy for Identification and Differentiation of Prenylated Peptides. *Anal. Chem.* **2015**, *87* (4), 2178–2186.

(59) Bhawal, R. P.; Shahinuzzaman, A. D. A.; Chowdhury, S. M. Gas-Phase Fragmentation Behavior of Oxidized Prenyl Peptides by CID and ETD Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (4), 704–707.

(60) Stadlmeier, M.; Bogena, J.; Wallner, M.; Wühr, M.; Carell, T. A Sulfoxide-Based Isobaric Labelling Reagent for Accurate Quantitative Mass Spectrometry. *Angew. Chemie Int. Ed.* **2018**, *57* (11), 2958–2962.

(61) Virreira Winter, S.; Meier, F.; Wichmann, C.; Cox, J.; Mann, M.; Meissner, F. EASI-Tag Enables Accurate Multiplexed and Interference-Free MS2-Based Proteome Quantification. *Nat. Methods* **2018**, *15* (7), 527–530.

(62) Casey, P. J.; Seabra, M. C. Protein Prenyltransferases. *J. Biol. Chem.* **1996**, *271* (10), 5289–5292.

(63) Gelb, M. H.; Brunsveld, L.; Hrycyna, C. A.; Michaelis, S.; Tamanoi, F.; Van Voorhis, W. C.; Waldmann, H. Therapeutic Intervention Based on Protein Prenylation and Associated Modifications. *Nat. Chem. Biol.* **2006**, *2* (10), 518–528.

(64) Court, H.; Amoyel, M.; Hackman, M.; Lee, K. E.; Xu, R.; Miller, G.; Bar-Sagi, D.; Bach, E. A.; Bergö, M. O.; Philips, M. R. Isoprenylcysteine Carboxylmethyltransferase Deficiency Exacerbates

KRAS-Driven Pancreatic Neoplasia via Notch Suppression. *J. Clin. Invest.* **2013**, *123* (11), 4681–4694.

(65) Sjogren, A.-K. M.; Andersson, K. M. E.; Liu, M.; Cutts, B. A.; Karlsson, C.; Wahlstrom, A. M.; Dalin, M.; Weinbaum, C.; Casey, P. J.; Tarkowski, A.; Swolin, B.; Young, S. G.; Bergo, M. O. GGTase-I Deficiency Reduces Tumor Formation and Improves Survival in Mice with K-RAS-Induced Lung Cancer. *J. Clin. Invest.* **2007**, *117* (5), 1294–1304.

(66) Liu, M.; Sjogren, A.-K. M.; Karlsson, C.; Ibrahim, M. X.; Andersson, K. M. E.; Olofsson, F. J.; Wahlstrom, A. M.; Dalin, M.; Yu, H.; Chen, Z.; Yang, S. H.; Young, S. G.; Bergo, M. O. Targeting the Protein Prenyltransferases Efficiently Reduces Tumor Development in Mice with K-RAS-Induced Lung Cancer. *Proc. Natl. Acad. Sci.* **2010**, *107* (14), 6471–6476.

(67) Young, S. G.; Meta, M.; Yang, S. H.; Fong, L. G. Prelamin A Farnesylation and Progeroid Syndromes. *J. Biol. Chem.* **2006**, *281* (52), 39741–39745.

(68) Gordon, L. B.; Massaro, J.; D'Agostino, R. B.; Campbell, S. E.; Brazier, J.; Brown, W. T.; Kleinman, M. E.; Kieran, M. W.; Progeria Clinical Trials Collaborative. Impact of Farnesylation Inhibitors on Survival in Hutchinson-Gilford Progeria Syndrome. *Circulation* **2014**, *130* (1), 27–34.

(69) Hougland, J. L.; Hicks, K. A.; Hartman, H. L.; Kelly, R. A.; Watt, T. J.; Fierke, C. A. Identification of Novel Peptide Substrates for Protein Farnesyltransferase Reveals Two Substrate Classes with Distinct Sequence Selectivities. *J. Mol. Biol.* **2010**, *395* (1), 176–190.

(70) Sebti, S. M. Protein Farnesylation: Implications for Normal Physiology, Malignant Transformation, and Cancer Therapy. *Cancer Cell* **2005**, *7* (4), 297–300.

(71) Berndt, N.; Sebti, S. M. Measurement of Protein Farnesylation and Geranylgeranylation in Vitro, in Cultured Cells and in Biopsies, and the Effects of Prenyl Transferase Inhibitors. *Nat. Protoc.* **2011**, *6* (11), 1775–1791.

(72) Kho, Y.; Kim, S. C.; Jiang, C.; Barma, D.; Kwon, S. W.; Cheng, J.; Jaunbergs, J.; Weinbaum, C.; Tamanoi, F.; Falck, J.; Zhao, Y. A Tagging-via-Substrate Technology for Detection and Proteomics of Farnesylated Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (34), 12479–12484.

(73) Hannoush, R. N.; Sun, J. The Chemical Toolbox for Monitoring Protein Fatty Acylation and Prenylation. *Nat. Chem. Biol.* **2010**, *6* (7), 498–506.

(74) Mann, M.; Jensen, O. N. Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.* **2003**, *21* (3), 255–261.

(75) Suzuki, T.; Ito, M.; Ezure, T.; Shikata, M.; Ando, E.; Utsumi, T.; Tsunasawa, S.; Nishimura, O. Protein Prenylation in an Insect Cell-Free Protein Synthesis System and Identification of Products by Mass Spectrometry. *Proteomics* **2007**, *7* (12), 1942–1950.

(76) Hoffman, M. D.; Kast, J. Mass Spectrometric Characterization of Lipid-Modified Peptides for the Analysis of Acylated Proteins. *J. Mass Spectrom.* **2006**, *41* (2), 229–241.

(77) Wotske, M.; Wu, Y.; Wolters, D. A. Liquid Chromatographic Analysis and Mass Spectrometric Identification of Farnesylated Peptides. *Anal. Chem.* **2012**, *84* (15), 6848–6855.

(78) Kassai, H.; Satomi, Y.; Fukada, Y.; Takao, T. Top-down Analysis of Protein Isoprenylation by Electrospray Ionization Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometry; the Mouse T? Protein. *Rapid Commun. Mass Spectrom.* **2005**, *19* (2), 269–274.

(79) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villén, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. Large-Scale Characterization of HeLa Cell Nuclear Phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (33), 12130–12135.

(80) Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Færgeman, N. J.; Mann, M.; Jensen, O. N. Quantitative Phosphoproteomics Applied to the Yeast Pheromone Signaling Pathway. *Mol. Cell.*

Proteomics **2005**, *4* (3), 310–327.

(81) Kamal, A. H. M.; Chakrabarty, J. K.; Udden, S. M. N.; Zaki, M. H.; Chowdhury, S. M. Inflammatory Proteomic Network Analysis of Statin-Treated and Lipopolysaccharide-Activated Macrophages. *Sci. Rep.* **2018**, *8* (1), 164.

(82) Shahinuzzaman, A. D. A.; Chakrabarty, J. K.; Fang, Z.; Smith, D.; Kamal, A. H. M.; Chowdhury, S. M. Improved In-solution Trypsin Digestion Method for Methanol–Chloroform Precipitated Cellular Proteomics Sample. *J. Sep. Sci.* **2020**, jssc.201901273.

(83) Kim, S.; Gupta, N.; Pevzner, P. A. Spectral Probabilities and Generating Functions of Tandem Mass Spectra: A Strike against Decoy Databases. *J. Proteome Res.* **2008**, *7* (8), 3354–3363.

(84) Kim, S.; Pevzner, P. A. MS-GF+ Makes Progress towards a Universal Database Search Tool for Proteomics. *Nat. Commun.* **2014**, *5*.

(85) Andersson, L.; Porath, J. Isolation of Phosphoproteins by Immobilized Metal (Fe3+) Affinity Chromatography. *Anal. Biochem.* **1986**, *154* (1), 250–254.

(86) Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R. Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns. *Anal. Chem.* **2004**, *76* (14), 3935–3943.

(87) Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jørgensen, T. J. D. Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. *Mol. Cell. Proteomics* **2005**, *4* (7), 873–886.

(88) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. Phosphopeptide Fragmentation and Analysis by Mass Spectrometry. *J. Mass Spectrom.* **2009**, *44* (6), 861–878.

(89) Miseta, A.; Csutora, P. Relationship Between the Occurrence of Cysteine in Proteins and the

Complexity of Organisms. *Mol. Biol. Evol.* **2000**, *17* (8), 1232–1239.

(90) Palumbo, A. M.; Reid, G. E. Evaluation of Gas-Phase Rearrangement and Competing Fragmentation Reactions on Protein Phosphorylation Site Assignment Using Collision Induced Dissociation-MS/MS and MS 3. *Anal. Chem.* **2008**, *80* (24), 9735–9747.

(91) Palumbo, A. M.; Smith, S. A.; Kalcic, C. L.; Dantus, M.; Stemmer, P. M.; Reid, G. E. Tandem Mass Spectrometry Strategies for Phosphoproteome Analysis. *Mass Spectrom. Rev.* **2011**, *30* (4), 600–625.

(92) Storck, E. M.; Morales-Sanfrutos, J.; Serwa, R. A.; Panyain, N.; Lanyon-Hogg, T.; Tolmachova, T.; Ventimiglia, L. N.; Martin-Serrano, J.; Seabra, M. C.; Wojciak-Stothard, B.; Tate, E. W. Dual Chemical Probes Enable Quantitative System-Wide Analysis of Protein Prenylation and Prenylation Dynamics. *Nat. Chem.* **2019**, 1.

(93) Epstein, W. W.; Lever, D.; Leining, L. M.; Bruenger, E.; Rilling, H. C. Quantitation of Prenylcysteines by a Selective Cleavage Reaction. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (21), 9668–9670.

(94) Berndt, N.; Hamilton, A. D.; Sebti, S. M. Targeting Protein Prenylation for Cancer Therapy. *Nat. Rev. Cancer* **2011**, *11* (11), 775–791.

(95) Bos, J. L. Ras Oncogenes in Human Cancer: A Review. *Cancer Res.* **1989**, *49* (17), 4682–4689.

(96) Dharmaiah, S.; Bindu, L.; Tran, T. H.; Gillette, W. K.; Frank, P. H.; Ghirlando, R.; Nissley, D. V; Esposito, D.; McCormick, F.; Stephen, A. G.; Simanshu, D. K. Structural Basis of Recognition of Farnesylated and Methylated KRAS4b by PDEδ. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (44), E6766–E6775.

(97) Hancock, J. F.; Paterson, H.; Marshall, C. J. A Polybasic Domain or Palmitoylation Is Required in

Addition to the CAAX Motif to Localize P21ras to the Plasma Membrane. *Cell* **1990**, *63* (1), 133–139.

(98) Nussinov, R.; Tsai, C.-J.; Jang, H. Oncogenic Ras Isoforms Signaling Specificity at the Membrane. *Cancer Res.* **2018**, *78* (3), 593–602.

(99) Berndt, N.; Sebti, S. M. Measurement of Protein Farnesylation and Geranylgeranylation in Vitro, in Cultured Cells and in Biopsies, and the Effects of Prenyl Transferase Inhibitors. *Nat. Protoc.* **2011**, *6* (11), 1775–1791.

(100) Charron, G.; Li, M. M. H.; MacDonald, M. R.; Hang, H. C. Prenylome Profiling Reveals S-Farnesylation Is Crucial for Membrane Targeting and Antiviral Activity of ZAP Long-Isoform. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (27), 11085–11090.

(101) Nguyen, U. T. T.; Guo, Z.; Delon, C.; Wu, Y.; Deraeve, C.; Fränzel, B.; Bon, R. S.; Blankenfeldt, W.; Goody, R. S.; Waldmann, H.; Wolters, D.; Alexandrov, K. Analysis of the Eukaryotic Prenylome by Isoprenoid Affinity Tagging. *Nat. Chem. Biol.* **2009**, *5* (4), 227–235.

(102) Hentschel, A.; Zahedi, R. P.; Ahrends, R. Protein Lipid Modifications-More than Just a Greasy Ballast. *Proteomics* **2016**, *16* (5), 759–782.

(103) Couvertier, S. M.; Zhou, Y.; Weerapana, E. Chemical-Proteomic Strategies to Investigate Cysteine Posttranslational Modifications. *Biochim. Biophys. Acta - Proteins Proteomics* **2014**, *1844* (12), 2315–2330.

(104) Tate, E. W.; Kalesh, K. A.; Lanyon-Hogg, T.; Storck, E. M.; Thinon, E. Global Profiling of Protein Lipidation Using Chemical Proteomic Technologies. *Curr. Opin. Chem. Biol.* **2015**, *24*, 48–57.

(105) Toorn, H. W. P. van den; Mohammed, S.; Gouw, J. W.; Breukelen, B. van; Heck, A. J. R. Targeted SCX Based Peptide Fractionation for Optimal Sequencing by Collision Induced, and Electron

Transfer Dissociation. *J. Proteomics Bioinform.* **2008**, *01* (08), 379–388.

(106) Polevoda, B.; Sherman, F. The Diversity of Acetylated Proteins. *Genome Biol.* **2002**, *3* (5), reviews0006.1.

(107) Brown, J. L.; Roberts, W. K. Evidence That Approximately Eighty per Cent of the Soluble Proteins from Ehrlich Ascites Cells Are Nalpha-Acetylated. *J. Biol. Chem.* **1976**, *251* (4), 1009–1014.

(108) Van Damme, P.; Arnesen, T.; Gevaert, K. Protein Alpha-N-Acetylation Studied by N-Terminomics. *FEBS J.* **2011**, *278* (20), 3822–3834.

(109) López-Otín, C.; Overall, C. M. Protease Degradomics: A New Challenge for Proteomics. *Nat. Rev. Mol. Cell Biol.* **2002**, *3* (7), 509–519.

(110) Prudova, A.; Keller, U. auf dem; Butler, G. S.; Overall, C. M. Multiplex N-Terminome Analysis of MMP-2 and MMP-9 Substrate Degradomes by ITRAQ-TAILS Quantitative Proteomics. *Mol. Cell. Proteomics* **2010**, *9* (5), 894–911.

(111) Mason, S. D.; Joyce, J. A. Proteolytic Networks in Cancer. *Trends Cell Biol.* **2011**, *21* (4), 228–237.

(112) Quesada, V.; Ordonez, G. R.; Sanchez, L. M.; Puente, X. S.; Lopez-Otin, C. The Degradome Database: Mammalian Proteases and Diseases of Proteolysis. *Nucleic Acids Res.* **2009**, *37* (Database), D239–D243.

(113) McQuibban, G. A. Inflammation Dampened by Gelatinase A Cleavage of Monocyte Chemoattractant Protein-3. *Science* **2000**, *289* (5482), 1202–1206.

(114) 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. *Nat. Biotechnol.* **2003**, *21* (5), 566–569.

(115) Timmer, J. C.; Enoksson, M.; Wildfang, E.; Zhu, W.; Igarashi, Y.; Denault, J.-B.; Ma, Y.; Dummitt, B.; Chang, Y.-H.; Mast, A. E.; Eroshkin, A.; Smith, J. W.; Tao, W. A.; Salvesen, G. S. Profiling Constitutive Proteolytic Events in Vivo. *Biochem. J.* **2007**, *407* (1), 41–48.

(116) Xu, G.; Shin, S. B. Y.; Jaffrey, S. R. Global Profiling of Protease Cleavage Sites by Chemoselective Labeling of Protein N-Termini. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (46), 19310–19315.

(117) Yamaguchi, M.; Nakazawa, T.; Kuyama, H.; Obama, T.; Ando, E.; Okamura, T.; Ueyama, N.; Norioka, S. High-Throughput Method for N-Terminal Sequencing of Proteins by MALDI Mass Spectrometry. *Anal. Chem.* **2005**, *77* (2), 645–651.

(118) Ji, C.; Guo, N.; Li, L. Differential Dimethyl Labeling of N-Termini of Peptides after Guanidination for Proteome Analysis. *J. Proteome Res.* **2005**, *4* (6), 2099–2108.

(119) Enoksson, M.; Li, J.; Ivancic, M. M.; Timmer, J. C.; Wildfang, E.; Eroshkin, A.; Salvesen, G. S.; Tao, W. A. Identification of Proteolytic Cleavage Sites by Quantitative Proteomics. *J. Proteome Res.* **2007**, *6* (7), 2850–2858.

(120) Kim, J.-S.; Dai, Z.; Aryal, U. K.; Moore, R. J.; Camp, D. G.; Baker, S. E.; Smith, R. D.; Qian, W.-J. Resin-Assisted Enrichment of N-Terminal Peptides for Characterizing Proteolytic Processing. *Anal. Chem.* **2013**, *85* (14), 6826–6832.

(121) 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* (5), 866–876.

(122) Wildes, D.; Wells, J. A. Sampling the N-Terminal Proteome of Human Blood. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (10), 4561–4566.

(123) Staes, A.; Impens, F.; Van Damme, P.; Ruttens, B.; Goethals, M.; Demol, H.; Timmerman, E.; Vandekerckhove, J.; Gevaert, K. Selecting Protein N-Terminal Peptides by Combined Fractional Diagonal Chromatography. *Nat. Protoc.* **2011**, *6* (8), 1130–1141.

(124) 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. *Nat. Biotechnol.* **2010**, *28* (3), 281–288.

(125) Kleifeld, O.; Doucet, A.; Prudova, A.; auf dem Keller, U.; Gioia, M.; Kizhakkedathu, J. N.; Overall, C. M. Identifying and Quantifying Proteolytic Events and the Natural N Terminome by Terminal Amine Isotopic Labeling of Substrates. *Nat. Protoc.* **2011**, *6* (10), 1578–1611.

(126) Yeom, J.; Ju, S.; Choi, Y.; Paek, E.; Lee, C. Comprehensive Analysis of Human Protein N-Termini Enables Assessment of Various Protein Forms. *Sci. Rep.* **2017**, *7* (1), 6599.

(127) Chen, L.; Shan, Y.; Weng, Y.; Sui, Z.; Zhang, X.; Liang, Z.; Zhang, L.; Zhang, Y. Hydrophobic Tagging-Assisted N-Termini Enrichment for In-Depth N-Terminome Analysis. *Anal. Chem.* **2016**, *88* (17), 8390–8395.

(128) Yepremyan, A. *SYNTHESES OF NOVEL PEPTIDE LINKERS*; 2015.

(129) Chowdhury, S. M.; Munske, G. R.; Yang, J.; Zhukova, D.; Nguyen, H.; Bruce, J. E. Solid-Phase N-Terminal Peptide Enrichment Study by Optimizing Trypsin Proteolysis on Homoarginine-Modified Proteins by Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2014**, *28* (6), 635–644.

(130) Wessel, D.; Flügge, U. I. A Method for the Quantitative Recovery of Protein in Dilute Solution in the Presence of Detergents and Lipids. *Anal. Biochem.* **1984**, *138* (1), 141–143.

(131) WO2016120247A1 - Methods for a quantitative release of biotinylated peptides and proteins from

streptavidin complexes - Google Patents https://patents.google.com/patent/WO2016120247A1/ja (accessed Mar 26, 2020).

(132) Rink, H. Solid-Phase Synthesis of Protected Peptide Fragments Using a Trialkoxy-Diphenyl-Methylester Resin. *Tetrahedron Lett.* **1987**, *28* (33), 3787–3790.

(133) Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. Mass Spectrometry Identifiable Cross-Linking Strategy for Studying Protein−Protein Interactions. *Anal. Chem.* **2005**, *77* (1), 311–318.

(134) Tang, X.; Bruce, J. E. A New Cross-Linking Strategy: Protein Interaction Reporter (PIR) Technology for Protein-Protein Interaction Studies. *Mol. Biosyst.* **2010**, *6* (6), 939–947.

(135) Kalkhof, S.; Sinz, A. Chances and Pitfalls of Chemical Cross-Linking with Amine-Reactive N-Hydroxysuccinimide Esters. *Anal. Bioanal. Chem.* **2008**, *392* (1–2), 305–312.

(136) Wang, J.; Zhang, R.-Y.; Wang, Y.-C.; Chen, X.-Z.; Yin, X.-G.; Du, J.-J.; Lei, Z.; Xin, L.-M.; Gao, X.-F.; Liu, Z.; Guo, J. Polyfluorophenyl Ester-Terminated Homobifunctional Cross-Linkers for Protein Conjugation. *Synlett* **2017**, *28* (15), 1934–1938.

(137) CANFIELD, R. E. THE AMINO ACID SEQUENCE OF EGG WHITE LYSOZYME. *J. Biol. Chem.* **1963**, *238*, 2698–2707.

(138) Braunitzer, G.; Chen, R.; Schrank, B.; Stangl, A. [The Sequence of Beta-Lactoglobulin (Author's Transl)]. *Hoppe. Seylers. Z. Physiol. Chem.* **1973**, *354* (8), 867–878.

(139) Schlesinger, D. H.; Goldstein, G.; Niall, H. D. Complete Amino Acid Sequence of Ubiquitin, an Adenylate Cyclase Stimulating Polypeptide Probably Universal in Living Cells. *Biochemistry* **1975**, *14* (10), 2214–2218.

(140) Solbiati, J.; Chapman-Smith, A.; Miller, J. L.; Miller, C. G.; Cronan, J. E. Processing of the N Termini of Nascent Polypeptide Chains Requires Deformylation Prior to Methionine Removal. *J.*

Mol. Biol. **1999**, *290* (3), 607–614.

(141) Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27* (2), 157–162.

(142) Hirel, P. H.; Schmitter, J. M.; Dessen, P.; Fayat, G.; Blanquet, S. Extent of N-Terminal Methionine Excision from Escherichia Coli Proteins Is Governed by the Side-Chain Length of the Penultimate Amino Acid. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (21), 8247–8251.

(143) Frottin, F.; Martinez, A.; Peynot, P.; Mitra, S.; Holz, R. C.; Giglione, C.; Meinnel, T. The Proteomics of N-Terminal Methionine Cleavage. *Mol. Cell. Proteomics* **2006**, *5* (12), 2336–2349.

(144) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14* (6), 1188–1190.

(145) Bienvenut, W. V.; Giglione, C.; Meinnel, T. Proteome-Wide Analysis of the Amino Terminal Status of Escherichia Coli Proteins at the Steady-State and upon Deformylation Inhibition. *Proteomics* **2015**, *15* (14), 2503–2518.

(146) Sinz, A. Chemical Cross-Linking and Mass Spectrometry for Mapping Three-Dimensional Structures of Proteins and Protein Complexes. *J. Mass Spectrom.* **2003**, *38* (12), 1225–1237.

(147) 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. *Anal. Chem.* **2009**, *81* (13), 5524–5532.

(148) Sinz, A. Chemical Cross-Linking and Mass Spectrometry to Map Three-Dimensional Protein Structures and Protein-Protein Interactions. *Mass Spectrom. Rev.* **2006**, *25* (4), 663–682.

(149) Gingras, A. C.; Gstaiger, M.; Raught, B.; Aebersold, R. Analysis of Protein Complexes Using Mass

Spectrometry. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (8), 645–654.

(150) 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. *Anal. Chem.* **2016**, *88* (20), 10215–10222.

(151) Leitner, A. Cross-Linking and Other Structural Proteomics Techniques: How Chemistry Is Enabling Mass Spectrometry Applications in Structural Biology. *Chem. Sci.* **2016**, *7* (8), 4792–4803.

(152) Singh, P.; Panchaud, A.; Goodlett, D. R. Chemical Cross-Linking and Mass Spectrometry As a Low-Resolution Protein Structure Determination Technique. *Anal. Chem.* **2010**, *82* (7), 2636–2642.

(153) Tinnefeld, V.; Venne, A. S.; Sickmann, A.; Zahedi, R. P. Enrichment of Cross-Linked Peptides Using Charge-Based Fractional Diagonal Chromatography (ChaFRADIC). *J. Proteome Res.* **2017**, *16* (2), 459–469.

(154) 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. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (11), 5802–5806.

(155) Lasker, K.; Forster, F.; Bohn, S.; Walzthoeni, T.; Villa, E.; Unverdorben, P.; Beck, F.; Aebersold, R.; Sali, A.; Baumeister, W. Molecular Architecture of the 26S Proteasome Holocomplex Determined by an Integrative Approach. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (5), 1380–1387.

(156) Herzog, F.; Kahraman, A.; Boehringer, D.; Mak, R.; Bracher, A.; Walzthoeni, T.; Leitner, A.; Beck, M.; Hartl, F.-U.; Ban, N.; Malmstrom, L.; Aebersold, R. Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-Linking and Mass Spectrometry. *Science* **2012**, *337* (6100), 1348–1352.

(157) Kao, A.; Randall, A.; Yang, Y.; Patel, V. R.; Kandur, W.; Guan, S.; Rychnovsky, S. D.; Baldi, P.; Huang, L. Mapping the Structural Topology of the Yeast 19S Proteasomal Regulatory Particle Using Chemical Cross-Linking and Probabilistic Modeling. *Mol. Cell. Proteomics* **2012**, *11* (12), 1566–1577.

(158) Liu, F.; Rijkers, D. T. S.; Post, H.; Heck, A. J. R. Proteome-Wide Profiling of Protein Assemblies by Cross-Linking Mass Spectrometry (Vol 12, Pg 1179, 2015). *Nat. Methods* **2017**, *14* (5), 539.

(159) Navare, A. T.; Chavez, J. D.; Zheng, C. X.; Weisbrod, C. R.; Eng, J. K.; Siehnel, R.; Singh, P. K.; Manoil, C.; Bruce, J. E. Probing the Protein Interaction Network of Pseudomonas Aeruginosa Cells by Chemical Cross-Linking Mass Spectrometry. *Structure* **2015**, *23* (4), 762–773.

(160) Shi, Y.; Pellarin, R.; Fridy, P. C.; Fernandez-Martinez, J.; Thompson, M. K.; Li, Y. Y.; Wang, Q. J.; Sali, A.; Rout, M. P.; Chait, B. T. A Strategy for Dissecting the Architectures of Native Macromolecular Assemblies. *Nat. Methods* **2015**, *12* (12), 1135-+.

(161) Aufderheide, A.; Beck, F.; Stengel, F.; Hartwig, M.; Schweitzer, A.; Pfeifer, G.; Goldberg, A. L.; Sakata, E.; Baumeister, W.; Forster, F. Structural Characterization of the Interaction of Ubp6 with the 26S Proteasome. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (28), 8626–8631.

(162) Petrotchenko, E. V; Borchers, C. H. Crosslinking Combined with Mass Spectrometry for Structural Proteomics. *Mass Spectrom. Rev.* **2010**, *29* (6), 862–876.

(163) Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. MS2Assign, Automated Assignment and Nomenclature of Tandem Mass Spectra of Chemically Crosslinked Peptides. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (8), 834–850.

(164) Lu, L.; Millikin, R. J.; Solntsev, S. K.; Rolfs, Z.; Scalf, M.; Shortreed, M. R.; Smith, L. M. Identification of MS-Cleavable and Noncleavable Chemically Cross-Linked Peptides with MetaMorpheus. *J. Proteome Res.* **2018**, *17* (7), 2370–2376.

(165) Du, X. X.; Chowdhury, S. M.; Manes, N. P.; Wu, S.; Mayer, M. U.; Adkins, J. N.; Anderson, G. A.; Smith, R. D. Xlink-Identifier: An Automated Data Analysis Platform for Confident Identifications of Chemically Cross-Linked Peptides Using Tandem Mass Spectrometry. *J. Proteome Res.* **2011**, *10* (3), 923–931.

(166) Lima, D. B.; de Lima, T. B.; Balbuena, T. S.; Neves-Ferreira, A. G. C.; Barbosa, V. C.; Gozzo, F. C.; Carvalho, P. C. SIM-XL: A Powerful and User-Friendly Tool for Peptide Cross-Linking Analysis. *J Proteomics* **2015**, *129*, 51–55.

(167) Hoopmann, M. R.; Zelter, A.; Johnson, R. S.; Riffle, M.; MacCoss, M. J.; Davis, T. N.; Moritz, R. L. Kojak: Efficient Analysis of Chemically Cross-Linked Protein Complexes. *J. Proteome Res.* **2015**, *14* (5), 2190–2198.

(168) Gotze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C. H.; Krauth, F.; Fritzsche, R.; Kuhn, U.; Sinz, A. StavroX-A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (1), 76–87.

(169) Yang, B.; Wu, Y. J.; Zhu, M.; Fan, S. B.; Lin, J. Z.; Zhang, K.; Li, S.; Chi, H.; Li, Y. X.; Chen, H. F.; Luo, S. K.; Ding, Y. H.; Wang, L. H.; Hao, Z. Q.; Xiu, L. Y.; Chen, S.; Ye, K. Q.; He, S. M.; Dong, M. Q. Identification of Cross-Linked Peptides from Complex Samples. *Nat. Methods* **2012**, *9* (9), 904-+.

(170) Zheng, C. X.; Yang, L.; Hoopmann, M. R.; Eng, J. K.; Tang, X. T.; Weisbrod, C. R.; Bruce, J. E. Cross-Linking Measurements of In Vivo Protein Complex Topologies. *Mol. Cell. Proteomics* **2011**, *10* (10).

(171) Tan, D.; Li, Q.; Zhang, M. J.; Liu, C.; Ma, C. Y.; Zhang, P.; Ding, Y. H.; Fan, S. B.; Tao, L.; Yang, B.; Li, X. K.; Ma, S. C.; Liu, J. J.; Feng, B. Y.; Liu, X. H.; Wang, H. W.; He, S. M.; Gao, N.; Ye, K. Q.; Dong, M. Q.; Lei, X. G. Trifunctional Cross-Linker for Mapping Protein-Protein Interaction

Networks and Comparing Protein Conformational States. *Elife* **2016**, *5*.

(172) Leitner, A.; Reischl, R.; Walzthoeni, T.; Herzog, F.; Bohn, S.; Forster, F.; Aebersold, R. Expanding the Chemical Cross-Linking Toolbox by the Use of Multiple Proteases and Enrichment by Size Exclusion Chromatography. *Mol. Cell. Proteomics* **2012**, *11* (3), M111 014126.

(173) Fritzsche, R.; Ihling, C. H.; Gotze, M.; Sinz, A. Optimizing the Enrichment of Cross-Linked Products for Mass Spectrometric Protein Analysis. *Rapid Commun. Mass Spectrom.* **2012**, *26* (6), 653–658.

(174) Buncherd, H.; Roseboom, W.; Ghavim, B.; Du, W.; de Koning, L. J.; de Koster, C. G.; de Jong, L. Isolation of Cross-Linked Peptides by Diagonal Strong Cation Exchange Chromatography for Protein Complex Topology Studies by Peptide Fragment Fingerprinting from Large Sequence Databases. *J Chromatogr A* **2014**, *1348*, 34–46.

(175) Appulage, D. K.; Wang, E. H.; Carroll, F.; Schug, K. A. Automated Screening of Reversed-Phase Stationary Phases for Small-Molecule Separations Using Liquid Chromatography with Mass Spectrometry. *J. Sep. Sci.* **2016**, *39* (9), 1638–1647.

(176) Buszewski, B.; Noga, S. Hydrophilic Interaction Liquid Chromatography (HILIC)-a Powerful Separation Technique. *Anal. Bioanal. Chem.* **2012**, *402* (1), 231–247.

(177) Jardim, I. C. S. F.; Maldaner, L.; Lourenco, J.; Fioravanti, L. M. A.; Collins, C. H. Some New Selective Stationary Phases for RP-HPLC. *J. Sep. Sci.* **2010**, *33* (19), 2917–2929.

(178) Brindle, R.; Klaus, A. Stationary Phases with Chemically Bonded Fluorene Ligands: A New Approach for Environmental Analysis of Pi-Electron Containing Solutes. *J. Chromatogr. A* **1997**, *757* (1–2), 3–20.

(179) Shollenberger, D.; Cramer, H.; Bell, D. S. Evaluation of Retention and Selectivity Using Biphenyl

Stationary Phases. *Lc Gc Eur.* **2017**, *30* (6), 314–320.

(180) Grassetti, A. V; Hards, R.; Gerber, S. A. Offline Pentafluorophenyl (PFP)-RP Prefractionation as an Alternative to High-PH RP for Comprehensive LC-MS/MS Proteomics and Phosphoproteomics. *Anal. Bioanal. Chem.* **2017**, *409* (19), 4615–4625.

(181) Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. Orthogonality of Separation in Two-Dimensional Liquid Chromatography. *Anal. Chem.* **2005**, *77* (19), 6426–6434.

(182) Yong, Y. S.; Chong, E. T. J.; Chen, H. C.; Lee, P. C.; Ling, Y. S. A Comparative Study of Pentafluorophenyl and Octadecylsilane Columns in High-Throughput Profiling of Biological Fluids. *J. Chinese Chem. Soc.* **2017**, *64* (6), 699–710.

(183) Huang, B. X.; Kim, H. Y.; Dass, C. Probing Three-Dimensional Structure of Bovine Serum Albumin by Chemical Cross-Linking and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (8), 1237–1247.

(184) Schmidt, R.; Sinz, A. Improved Single-Step Enrichment Methods of Cross-Linked Products for Protein Structure Analysis and Protein Interaction Mapping. *Anal. Bioanal. Chem.* **2017**, *409* (9), 2393–2400.

(185) Oliveros, J. C. An interactive tool for comparing lists with Venn Diagrams https://bioinfogp.cnb.csic.es/tools/venny/ (accessed Oct 23, 2020).

Biography

Zixiang Fang was born in Guangdong, China. He received his bachelor's degree from Guangdong University of Technology in Applied Chemistry in 2012. He joined the chemistry PhD program of Department of Chemistry and Biochemistry in the University of Texas at Arlington (UTA) in fall 2015. Under the supervision of Dr. Saiful Chowdhury, his PhD research focuses on developing MS-based strategies for proteomics study, specifically on the study of protein PTMs and protein interactions. He represented UTA to attend and present his research in the annual American Society of Mass Spectrometry (ASMS) conferences from 2017-2020. He worked in Bayer during the summer of 2019 as Regulatory Mass Spec Analyst Intern. He also received Mav Science Graduate Research Fellowship from College of Science in summer of 2020. He is expected to graduate with the PhD degree in bioanalytical chemistry from UTA in fall 2020.