Method Development for Qualitative and Quantitative Study in Mass Spectrometry

Chang Xue

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

František Tureček, Chair

Bo Zhang

Champak Chatterjee

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Method Development for Qualitative and Quantitative Study in Mass Spectrometry

Chang Xue

Chair of the Supervisory Committee:

Klaus and Mary Ann Saegebarth Endowed Professor František Tureček

Department of Chemistry

Mass spectrometry has become the method of choice for lots of areas in recent years including protein study and clinical research. In this thesis, work from three projects covering three very different domains in the big area of mass spectrometry has been described. Chapter 2 to Chapter 4 presented the project of methodology development in chemical derivatization of peptides. In this project, custom-tunable tags prepared through solid phase synthesis increased sequence coverage of peptide ions in electron transfer dissociation for *de novo* sequencing, thus simplifies the characterization of proteins. Additionally, the application of developed stable isotopic tags pair can be used for quantitation of the proteins as well as for the simplified characterization of proteins. Chapter 5 presented the project of newborn screening in clinical study. In this project, a novel tandem mass spectrometry duplex assay for the clinical diagnosis

of neuronal ceroid lipofuscinoses (NCL) has been developed. The duplex assay recognized the deficient patients from healthy newborns by quantification of the enzyme activity in dried blood spot through the use of tandem mass spectrometer. Chapter 6 presented the result from protein-protein interaction project. Protein-protein interaction is essential to understand biological pathways. In this project, a second generation of protein interaction reporter BDD-PIR has been synthesized and applied to proteins for performance evaluation. This molecule has a low molecular weight and high solubility and is therefore expected to have improved performance than the first generation protein interaction reporter for *in vivo* application.

Dedication

This dissertation is dedicated to my parents Min Xue and Yuli Zhang, my husband Xin Chang and my daughter Tian-tian. Their love and encouragement is the greatest asset in my life.

Acknowledgement

First of all, I would like to express my sincere appreciation to my supervisor, Professor František Tureček. You are an amazing mentor as well as a wonderful scientist. Your knowledge and passion for science has greatly inspired me. Thank you for your support and guidance during the past years, they will benefit me for my whole life. I also would like to acknowledgement Professor James E. Bruce and Professor Michael H. Gelb for their support and opportunity provided in the protein-protein interaction project. Next, I would like to thank the great lab members in Tureček's group. Finally, I would like to sincerely thank my family and friends who have helped me a lot and I will always remember them deeply in heart.

Table of Contents

List of Abbreviations

A	Alanine
ACN	Acetonitrile
BOC	*Tert*-butoxycarbonyl
BSA	Bovine Serum Albumin
C	Cysteine
CI	Chemical Ionization
CID	Collisional Induced Dissociation
c-LINCL	Classical Late Infantile Neuronal Ceroid Lipofuscinoses
CV	Coefficients of Variances
D	Aspartic acid
DBS	Dried Blood Spots
DEA	Diethyl Amine
DMAP	4-Dimethylaminopyridine
DMF	*N,N*-dimethylformamide
DMSO	Dimethyl Sulfoxide
E	Glutamic acid
EIC	Extracted ion chromatogram
ELISA	Enzyme-linked Immunosorbent Assay
ESI	Electrospray Ionization
ETD	Electron Transfer Dissociation
F	Phenylalanine
FA	Formic acid
FAB	Fast Atom Bombardment
G	Glycine
H	Histidine
HuCNS-SC	Human Central Nervous Stem Cells
IAA	Iodoacetamide
INCL	Infantile Neuronal Ceroid Lipofuscinoses
IS	Internal Standard
K	Lysine
L	Leucine
LC	Liquid chromatography
LSD	Lysosomal Storage Diseases

MALDI	Matrix-assisted Laser Desorption Ionization
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
N	Asparagine
NCL	Neuronal Ceroid Lipofuscinosis
P	Proline
PIR	Protein Interaction Reporter
PPT I	Palmitoyl Protein Thioesterase I enzyme
PTM	Post Translational Modifications
R	Arginine
RT	Room temperature
S	Serine
SRM	Selected Reaction Monitoring
T	Threonine
TCEP	Tris(2-carboxyethyl) phosphine
TFA	Trifluoroacetic acid
TLC	Thin Layer Chromatography
TPP I	Tripeptidyl-peptidase 1
V	Valine
Y	Tyrosine

List of Schemes

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Mass Spectrometry

Mass spectrometry has been an important tool in physics and chemistry since the early part of the 1900s. It has expanded since then to its current application as the most powerful technique for analyzing proteomes due to the tremendous development in both science and instrumentation in the last two decades.

At a fundamental level, a mass spectrometer measures the mass-to-charge ratios (*m/z*) of ions, and therefore determines the molecular weight of the molecule based on the *m/z* value. This process consists of three steps.[1] First, molecules must be converted to gas phase ions, a step which poses a noticeable challenge for samples in a solid or liquid phase. Second, ions are separated by their unique m/z values via electric or magnetic fields in a component called a mass analyzer. Lastly, the separated ions and the abundance of each species with a specific m/z are detected. Therefore, though the individual parts can vary quite widely, all mass spectrometers are generally constituted of three functional components: an ion source, a mass separator, and an ion detector.

Mass spectrometry was initially used to discover fundamental aspects of atomic and molecular structure in early 1900s, resulting in the determination of atomic weights of elements and the discovery of stable isotopes. The landmark for the beginning of mass spectrometry is generally associated with the discovery of the electron by physicist J.J. Thomson, who used an electric field inside a cathode ray tube. This success led Thomson to invent a crude “mass spectrograph” to measure atomic weights of elements.[2] [3] The technique of mass spectrometry was used in more practical applications in the following few decades, such as being used for process control in petroleum processing to increase fuel quality. These applications encouraged fundamental research studies of gas-phase ionization and fragmentation in academia laboratories. Meanwhile, instrumentation has been developed as well. High resolution mass spectrometry, which is marked by improved mass accuracy and peak capacity, was

achieved by the production of the double-focusing magnetic sector instrument, which used an electric sector to correct for kinetic energy spread in ions before separation in the magnetic field.

Started in the late 1950s, efforts were made to use mass spectrometry to measure the molecular weight of small molecules to verify the structures.[4] In those days, the two commonly used methods for structural assignments of synthetic products are melting-point analysis, and through synthetic deconstruction of the molecule to something with known structure. However, the accuracy of melting-point analysis is highly dependent on the purity of the chemical, and the latter method is known to be time consuming. Since mass spectrometry does not require the tested compound to be pure, and it is a fast analytical method, the use of mass spectrometry for the characterization of small molecules and natural products has started to be widely spread.

The primary ionization method used in the 1950s was electron impact ionization, and this technology used a beam of energetic electrons to bombard volatilized molecules to create fragment ions. With this technology, the detection of molecules with limited volatility is achievable since they can still exhibit enough vapor pressure with electron impact ionization. However, this ionization method has a difficulty in ionizing biomolecules, which are polar and usually charged, and thus are not volatized even under low pressure.[5] Derivatization methods have been investigated to reduce the polarity and the charge on the biomolecules, but only peptides with lengths shorter than 4 amino acids are able to be detected by mass spectrometer even after derivatization when using electron impact ionization methods.

Two crucial technological advances in mass spectrometry turn the approach of using mass spectrometry to sequence peptides into practical reality. The first advance is the innovation of chemical ionization (CI), a soft ionization method which only induces very little fragmentation in peptide ions,[6] and therefore is advantageous over election impact ionization since CI preserves molecular ion information which is rarely available in election impact ionization. The other advance is the development of methods to interface gas chromatographs to mass spectrometers (GC-MS).[7] This development is significant since it allows analysis of a complicated mixture through the combination of two analytical methods enabling online separation followed by characterization.

The invention of new ionization methods has played an important role in the development of peptide sequencing analysis. Peptide derivatization is required to raise analyte volatility when using CI.

The innovation of another soft ionization method, fast atom bombardment (FAB) [8] [9] developed in the early 1980s, did not require derivatization and therefore was less labor intensive. However, the size of biomolecules that could be ionized with this technology was limited to thousands of daltons, which was advanced at that age but still far from robust. The softness of FAB led to limited fragmentations, necessitating the development of novel tandem instruments which added a separate fragmentation ability and tandem hybrids that combined different types of mass separators.

A revolution in the study of biomolecules was initiated by the invention of two novel ionization techniques in the late 1980s: matrix-assisted laser desorption ionization [10] (MALDI) and electrospray ionization (ESI).[11] MALDI is a technique similar to FAB, which requires a sample be prepared in a matrix and placed on a stage. It has the capability to accurately measure the mass of much larger molecules than FAB as singly charged ions, including proteins, carbohydrates and DNA.

ESI is very different from MALDI. This ionization technique enables the formation of multiply charged ions, which assures the *m/z* of multiply charged molecular ions falling in the range detectable by mass spectrometers. Additionally, interfacing liquid–phase separation to a mass spectrometer was achieved with ESI, allowing the convenient and robust analysis of complicated biological mixtures which are polar and charged.

Electron capture dissociation [12] (ECD) and electron transfer dissociation [13] (ETD) developed in early-2000s are two new dissociation methods collectively referred to as ExD. ExD induces fragmentation of cations by combining electrons with them leading to characteristic backbone cleavages. The mechanism of ExD has been investigated in detail by high–level theoretical calculations [14] [15] but is not yet fully understood. ExD is advantageous for the study of post-translation in proteins, since it preserves the post-translational modifications (PTM), a trait that makes it superior to CID for PTM analysis.

With the application of advance ionization methods, mass spectrometry made tremendous contributions to cell biology through the study of protein-protein interactions, organelles, cells and post-translational modifications, and also to the medicinal studies, agriculture studies as well as environmental studies. Much data has been collected for the complicated cell systems and finding the ways to analyze these huge datasets became the bottleneck for biological study. Multiple software search engines have been developed to identify proteins from peptide sequence databases using mass spectrometry data.

Various methods for sample pretreatment and enrichment to decease the complexity of biological sample have also been studied. The three projects I discuss in this thesis are all mass spectrometry related, while the focuses are different, including charge tag and quantitation study, clinical study and protein-protein interaction study. In **Chapters 2-4**, I am reporting the development of novel custom-tunable charge tags using solid phase synthesis to improve sequence coverage of modified peptides by ETD while decreasing the sample complexity, and also the development of an isotope labeled charge tag for quantitative study. In **Chapter 5**, I am reporting the development of a novel duplex tandem mass spectrometry assay for the clinical diagnosis of two diseases. In **Chapter 6**, I am reporting the development of a novel cleavable cross-linker (protein interaction reporter) for the discovery of protein-protein interaction in biological system for proteomics study.

1.2 Charge Tags in Mass Spectrometry

In recent years, mass spectrometry (MS) has become the method of choice for protein characterization and quantitation. Trypsin is the protease most commonly used for digestion of proteins into peptides. The digested peptides can then be separated by liquid chromatography (LC) and further sequenced by collisional induced dissociation (CID) based tandem mass spectrometry (MS/MS). Typically under CID conditions, tryptic peptide ions generate a number of different fragment ion series, where *N*-terminal (*b* ions) and *C*-terminal (*y* ions) (fragment ions) dominate the tandem mass spectra.[16]-[18] However, these mixtures of multiple fragment ion series can create rather complicated spectra which then involve several layers of potentially redundant information. Reducing spectra to the minimum information, ideally originating from an individual but entire ion series, will allow easier interpretation either by manual inspection or by automated search algorithms.[19]

In the past few decades, a number of strategies have been developed to influence peptide ion fragmentation creating simplified spectra in CID. The representative strategy for creating simplified spectra is to modify functional groups on peptides. Chemically assisted fragmentation (CAF) and 4-sulfophenyl isothiocyanate (SPITC) are two commonly used methods that simplify spectra by removing the presence of *N*-terminal fragments.[20][21] As an alternative to CID, there is the relatively new fragmentation technique electron transfer dissociation (ETD). ETD is a technique for peptide analysis that

provides good sequence coverage while also retaining post-translational modifications. Charge tags introduction into peptides has the potential effect of improving ionization, increasing charge states of the peptide ions, and increasing sequence coverage for *de novo* sequencing.

Our group has developed a 4-Dimethylaminopyridine (DMAP) charge tag with fixed charge, which labels the *N* terminus of the peptide and increases sequence coverage in ETD for *de novo* sequencing.[22] However, there are two drawbacks for using the DMAP tag. First, the reaction is performed in solution, and requires a large excess of reagents which are difficult to remove before ESI. Second, the reaction is not exclusively happening at the *N* terminus of the peptide, the amino group at the side chain of Lys also reacts. The first drawback is universal for any tagging reaction happening in the solution, which increases the sample complexity due to incomplete and side reactions. Therefore, we are developing a novel methodology to solve this problem.

The “capture and release” method using solid phase resins have been applied to derivatize and enrich *N*-linked glycopeptides,[23] carbohydrate,[24] and nucleotides.[25] Here, we develop an approach using solid phase synthesis to capture lysine terminated tryptic peptides on the resin and derivatize with the guanidine tag, followed by final release from the beads for analysis via mass spectrometry. In this approach, the sample complexity of modified tryptic peptides is highly reduced due to the use of the solid phase method. Meanwhile, lysine terminated tryptic peptides are modified with a basic guanidine tag, and are therefore expected to have higher sequence coverage than those that are underivatized. Finally, the custom-tunable tags can also be made with differential stable isotopes, and thus they can be applied in quantitative proteomics applications.

1.3 Application of Mass Spectrometry in Newborn Screening

Tremendous technological developments in tandem mass spectrometry during the past two decades boosted the development of newborn screening for inborn errors. In such diseases, the dominant tools of preventive medicine are screening procedures. The goal of these screening procedures is to detect the disease before the symptoms are manifest or at least in early stages, thus to prevent the onset or mitigate complications from the eventual symptoms of the disease.[26] Starts in 1968, the World Health Organization has published Wilson and Jungner criteria, which are the guidelines for the selection

of the diseases to be included in screening programs.[27] According to these guideline, a disease falls into the screening programs category only when it is treatable; and then its early treatment must be advantageous over later treatment. This rule requires the disease to have a long enough latency or period free of symptoms to make early intervention possible. However, the process of assay development for screening single or multiple diseases normally takes time. When the corresponding treatments are in development, eg.at phase I, II or III, there is a significant benefit to starting assay method development, which can promise the availability of the diagnosis assay by the time when the treatment receives approval from FDA and is open to the public.

Samples from dried blood spot (DBS) are widely used in current newborn screening labs all over the world. This universal newborn screening specimen was invented in the mid-20th century,[28]-[31] and is relatively simple: capillary blood is drawn by heel prick and dried on filter paper. This sample collection method is fast and convenient, and is especially economic when used on a large population, and therefore has become a widely used method by current standards.

There have been various methods available for newborn screening, includes radioimmunoassay, enzyme-linked immunosorbent assay (ELISA), fluoroimmunoassay, spectrophotometric assay and mass spectrometric assay. An ideal methodology for newborn screening should have the following major features: high sensitivity, high specificity, inexpensive, and it must support a high throughput. Mass spectrometric assays have advantages over others: first, other assay methods generally are not multiplexable with the exception of immunological assays; second, immunological assays only provide data for protein abundance, which is not useful in cases where nonfunctional proteins are produced.[32] Therefore, mass spectrometry based assays are more efficient and generally cost less with dried blood spot samples than other methods due to the multiplexability.

The lysosomal storage diseases (LSD) are a huge group of approximately 50 rare, inherited metabolic disorders that result from the deficiency or absence of specific enzymes which initiate the breakdown of undesirable materials in the cell.[33] Although each of them is a rare disease, the incidence of all LSDs has been predicted to be as high as 1:5200.[34] Therefore, there is significant benefit to develop assays to diagnosis LSDs. The detection of lysosomal enzyme activities in DBS is complicated due to the low concentration of enzymatic products as the protein content is natively low. A mass

spectrometry based approach has been developed to include the following four steps: first, a 3 mm punch is taken from a DBS and rehydrated in a buffer to extract the enzymes of interest; second, the substrate for the enzyme is added, and the mixture is incubated under a specific condition to allow enzymatic conversion of substrate to product; third, the sample is processed to remove the relatively large amounts of buffer salts, which will interfere with the electrospray ionization (ESI) process in the mass spectrometer; and finally, the sample is introduced into the ESI mass spectrometer operating in tandem mode by direct infusion, and the amount of enzymatically-generated product is selectively monitored along with an internal standard to determine the absolute amount of the reaction product.[32] The internal standard, which is chemically identical to or similar to the reaction product, is characterized in the mass spectrometer from the reaction product by having a different mass.

Tandem MS methodology is advantageous since it is a fast (requiring less than 2 min per patient sample), extremely sensitive (save for the cost of both reagents and patient sample), a highly multiplexable technique, and can provide for highly accurate quantitative information. Therefore, this methodology is used to develop the duplex assay as described in **Chapter 5**.

1.4 Protein-protein interaction

The elucidation of protein interactions is essential to understanding biological pathways. Most cellular functions are accomplished through large multi-protein complexes rather than single proteins. However, *in vivo* protein interaction identification and topological measurements are most challenging due to the fact that protein interactions are often quite transient in nature and the fact that no common physical properties exist for all protein complexes. Various methods have been developed to help understand protein interactions. The combination of cross-linking and mass spectrometry holds promise for *in vivo* interactions.[35-40] Cross-linking technology in principle provides protein interaction information and may yield unique and useful data on topological features of protein complexes. Unlike x-ray crystallography, cross-linking does not require large quantities of purified sample and does not have a protein mass range limitation as NMR does.[41] However, cross-linking sample analysis results are challenging to interpret because of the presence of multiple types of cross-linking products, small quantity of cross-linking products relative to other non-cross-linking peptide fragments after digestion, and the

complicated mass spectra associated with cross-linking peptides. As a result, currently only limited examples of the identification of peptides that were cross-linked in cells exists in the literature.[42]

Many different cross-linker molecules have been developed, including isotope labeled cross-linkers,[43][44] fluorophore containing cross-linkers,[45][46] affinity tag-containing cross-linkers,[47-49] and chemically cleavable cross-linkers.[46][47] Five years ago, the first generation of Protein Interaction Reporter (PIR) molecules were reported, which incorporate low collisional energy cleavable groups.[50] The key feature of PIR technology is that the cleavable bonds enable production of mass spectra that could readily be interpreted, based on the engineered cleavage of these cross-linkers and cross-linked products. The identification of *in vitro* protein interaction has been achieved with the application of phototype PIR, using the expected mass relationships between cross-linked products and their released peptides for a purified protein complex.[50] Based on the success of this first generation of PIR, the second generation of PIR was created with the addition of a biotin affinity tag, which enabled enrichment of cross-linked peptides, simplified the observed mass spectra and allowed the first unbiased *in vivo* protein interactions to be identified from live cells.[42] Given the success *in vitro* and *in vivo* applications, investment of additional efforts to refine the molecular design for future improved applications of PIR technology are warranted.

A novel type of collision-induced dissociation (CID) and in-source collision-induced dissociation (ISCID) cleavable Protein Interaction Reporter (PIR) cross-linker was designed and synthesized using solution phase chemistry. The molecule was designed to incorporate low energy cleavable bonds, biotin affinity enrichment capabilities, symmetrical arm structure, high aqueous solubility, low molecular weight, and improved chromatographic properties. In **Chapter 6**, the synthetic route and cross-linking performance analysis on peptide and protein standards are described. The new PIR reagent offers significant advantages for protein structure and interaction measurements in complex biological systems *in vivo*.

1.5 References

[1] Yates, J. A Century of Mass spectrometry: From Atoms to Proteomes. *Nat. Methods* **2011,** *8*, 633–637.

[2] Thomson, J. J. Cathode Rays. *Philos. Mag.* **1897,** *44*, 293-316.

[3] Grayson, M. (ed.) Measuring Mass: From Positive Rays to Proteins (Chemical Heritage Press, Philadelphia, **2002**).

[4] Biemann, K. Four Decades of Structure Determination by Mass Spectrometry: From alkaloids to Heparin. *J. Am. Soc. Mass Spectrom.* **2002,** *13*, 1254-1272.

[5] Biemann, K.; Gapp, G. ; Seibl, J. Application of Mass Spectrometry to Structure Problems. I. Amino Acid Sequence in Peptides. *J. Am. Chem. Soc.* **1959,** *81*, 2274-2275.

[6] Munson, M. S. B.; Field, F. H. Chemical Ionization Mass Spectrometry I. General introduction. *J. Am. Chem. Soc.* **1966,** *88*, 2621-2630.

[7] Ryhage, R. Use of a Mass Spectrometer as a Detector and Analyzer for Effluents Emerging from High Temperature Gas-Liquid Chromatography Columns. *Anal. Chem.* **1964,** *36,* 759-764.

[8] Barber, M.; Bordoli, R. S.; Sedgwick, R.D.; Tyler, A. N. FAB of Solids as an Ion Source in Mass Spectrometry. *Nature* **1981,** *293*, 270-275.

[9] Morris, H.R.; Panico, M.; Barber, M.; Bordoli, R.S.; Sedgwick, R.D.; Tyler, A. Fast Atom Bombardment: a New Mass Spectrometric Method for Peptide Sequence Analysis. *Biochem. Biophys. Res. Commun,* **1981,** *101,* 623–31.

[10] Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988,** *60*, 2299-2301.

[11] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, **1989,** *246*, 64-71.

[12] McLafferty, F.; Horn, D.M.; Breuker, K.; Ge, Y.; Lewis, M.A.; Cerda, B.; Zubarev, R.A.; Carpenter, B.K. Electron Capture Dissociation of Gaseous Multiply Charged Ions by Fourier-Transform Ion Cyclotron Resonance. *J. Am. Soc. Mass Spectrom.* **2001,** *12,* 245.

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

[14] Uggerud, E. Electron Capture Dissociation of the Disulfide Bond-a Quantum Chemical Model Study. *Int. J. Mass Spectrom.* **2004,** *234*, 45–50.

[15] Syrstad, E.A.; Tureček, F. Toward a General Mechanism of Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2005,** *16*, 208–224.

[16] Johnson, R. S.; Martin, S. A.; Biemann, K. Collision-Induced Fragmentation of $(M+H)^+$ Ions of Peptides. Side Chain Specific Sequence Ions. *Int. J. Mass Spectrom. Ion Processes.* **1988,** *86*, 137–154.

[17] Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

[18] Steen, H.; Mann, M. The ABC's (and XYZ's) of Peptide Sequencing. *Nat. Rev. Mol. Cell Biol.* **2004,** *5*, 699–711.

[19] Hennrich, M. L.; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J. R.; Mohammed, S. Effect of Chemical Modifications on Peptide Fragmentation Behavior upon Electron Transfer Induced Dissociation. *Anal. Chem.* **2009,** *81*, 7814– 22.

[20] Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96*, 7131–7136.

[21] Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate *De Novo* Sequencing of Lysine-terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000,** *14*, 2348–2356.

[22] Tian, Y.; Zhou, Y.; Elliott, S.; Aebersold, R.; Zhang, H. Solid-phase extraction of N-linked glycopeptides. *Nature Protocols*, **2007,** *2*, 334-339.

[23] Zimnicka, M.; Moss, C. L.; Chung, T. W.; Hui, R.; Tureček, F. Tunable Charge Tags for Electron-Based Methods of Peptide Sequencing: Design and Applications. *J.Am. Soc. Mass Spectrom.* **2012,** *23*, 608-620.

[24] Lohse, A.; Joergensen, M. R.; Martins, R.; Hindsgaul, O. Solid-phase Oligosaccharide Tagging: a Technique for Manipulation of Immobilized Carbohydrates *PCT Int. Appl.* **2006,** WO 2006084461 A1 20060817

[25] Kim, S.; Edwards, J. R.; Deng, L.; Chung, W.; Ju, J. Solid Phase Capturable Dideoxynucleotides for Multiplex Genotyping Using Mass Spectrometry. *Nucleic Acids Research* **2002,** *30*, e85/1-e85/6.

[26] Fingerhut, R.; Olgemöller, B. Newborn Screening for Inborn Errors of Metabolism and Endocrinopathies: an Update. *Anal. Bioanal. Chem.* **2009,** *393*, 1481–1497.

[27] Wilson, J. M. G., Jungner, G. Principles and Practice of Screening for Diseases. World Health Organization, Geneva **1968** Public health papers no 34.

[28] Fölling, A. Über Ausscheidung von Phenylbrenztraubensäure in den Harn als Stoffwechselanomalie in Verbindung mit Imbezillität. Hoppe-Seyler's. *Z Physiol. Chem.* **1934**, *227,* 169–176.

[29] Guthrie, R.; Susi, A. A Simple Phenylalanine Method for Detecting Phenylketonuria in Large Populations of Newborn Infants *Pediatrics*, **1963**, *32*, 338–343.

[30] Jervis, G. A. Phenylpyruvic Oligophrenia: Introductory Study of Fifty Cases of Mental Deficiency Associated with Excretion of Phenylpyruvic Acid. *Arch. Neurol. Psychiatry*, **1937**, *38*, 944–963.

[31] Jervis, G. A. *J. Ment. Sci.*, **1939**, *85*, 719–762.

[32] Gelb, M. H.; Tureček, F.; Scott, C. R.; Chamoles, N. A. Direct Multiplex Assay of Enzymes in Dried Blood Spots by Tandem Mass Spectrometry for the Newborn Screening of Lysosomal Storage Disorders *J. Inherit. Metab. Dis*. **2006***, 29,* 397–404.

[33] Lehotay, D. C.; Hall, P.; Lepage, J.; Eichhorst, J. C.; Etter, M. L.; Greenberg, C. R. LC–MS/MS Progress in Newborn Screening *Clinical Biochemistry* **2011,** *44*, 21–31.

[34] Sanderson, S.; Green, A.; Preece, M. A.; Burton, H. The Incidence of Inherited Metabolic Disorders in the West Midlands, *UK. Arch. Dis. Child.* **2006,** *91*, 896–899.

[35] Back, J. W.; De Jong, L.; Muijsers, A. O.; De Koster, C. G. Chemical Cross-linking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* **2003**, *331*, 303–313.

[36] Sinz, A. Chemical Cross-linking and Mass Spectrometry to Map Three-dimensional Protein Structures and Protein–protein Interactions. *Mass Spectrom. Rev.* **2006**, *25*, 663–682.[37] 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*, 5802–5806.

[38] Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L. N.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. Identification of Cross-linked Peptides from Large Sequence Databases. *Nat. Methods* **2008**, *5*, 315–318.

[39] Tang, X.; Yi, W.; Munske, G. R.; Adhikari, D. P.; Zakharova, N. L.; Bruce, J. E. Profiling the Membrane Proteome of Shewanella Oneidensis MR-1 with New Affinity Labeling Probes. *J. Proteome Res.* **2007**, *6*, 724–734.

[40] Zhang, H.; Tang, X.; Munske, G. R.; Zakharova, N.; Yang, L.; Zheng, C.; Wolff, M. A.; Tolic, N.; Anderson, G. A.; Shi, L.; Marshall, M. J.; Fredrickson, J. K.; Bruce, J. E. Identification of Protein-Protein Interactions and Topologies in Living Cells with Chemical Cross-linking and Mass Spectrometry *J. Proteome Res.* **2008**, *7*, 1712–1720.

[41] 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*, 1237–1247.

[42] Zhang, H.; Tang, X.; Munske, G. R.; Tolic, N.; Anderson, G. A.; Bruce, J. E. Identification of Protein-Protein Interactions and Topologies in Living Cells with Chemical Cross-linking and Mass Spectrometry *Mol. Cell. Proteomics* **2009**, *8*, 409–420.

[43] Mueller, D. R. ; Schindler, P. ; Towbin, H. ; Wirth, U. ; Voshol, H. ; Hoving, S. ; Steinmetz, M. O. Isotope-tagged cross-linking reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis. *Anal. Chem.* **2001**, *73*, 1927–1934.

[44] Chu, F.; Mahrus, S.; Craik, C. S.; Burlingame, A. L. Isotope-coded and Affinity-tagged Cross-linking (ICATXL): An Efficient Strategy to Probe Protein Interaction Surfaces. *J. Am. Chem. Soc.* **2006**, *128*, 10362–10363.

[45] Wine, R. N.; Dial, J. M. ; Tomer, K. B.; Borchers, C. H. Identification of Components of Protein Complexes Using a Fluorescent Photo-cross-linker and Mass Spectrometry. *Anal. Chem.* **2002**, *74*, 1939–1945.

[46] Yang, L.; Tang, X.; Weisbrod C.R.; Munske, G. R.; Eng, J.K.; Haller, P.D.; Kaiser, N.K.; Bruce, J.E. A Photocleavable and Mass Spectrometry Identifiable Cross-linker for Protein Interaction Studies. *Anal. Chem.* **2010**, *82*, 3556–3566.

[47] Trester-Zedlitz, M.; Kamada, K.; Burley, S. K.; Fenyoe, D.; Chait, B. T.; Muir, T. W. A Modular Cross-Linking Approach for Exploring Protein Interactions. *J. Am. Chem. Soc.* **2003**, *125*, 2416–2425.

[48] Hurst, G. B.; Lankford, T. K.; Kennel, S. J. Mass Spectrometric Detection of Affinity Purified Crosslinked Peptides *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 832–839.

[49] Sinz, A.; Kalkhof, S.; Ihling, C. Mapping Protein Interfaces by a Trifunctional Cross-Linker Combined with MALDI-TOF and ESI-FTICR Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1921–1931.

[50] 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*, 311–318.

Chapter 2
Development of Custom-Tunable Charge Tags for Multiplex Electron Transfer Dissociation of Peptides

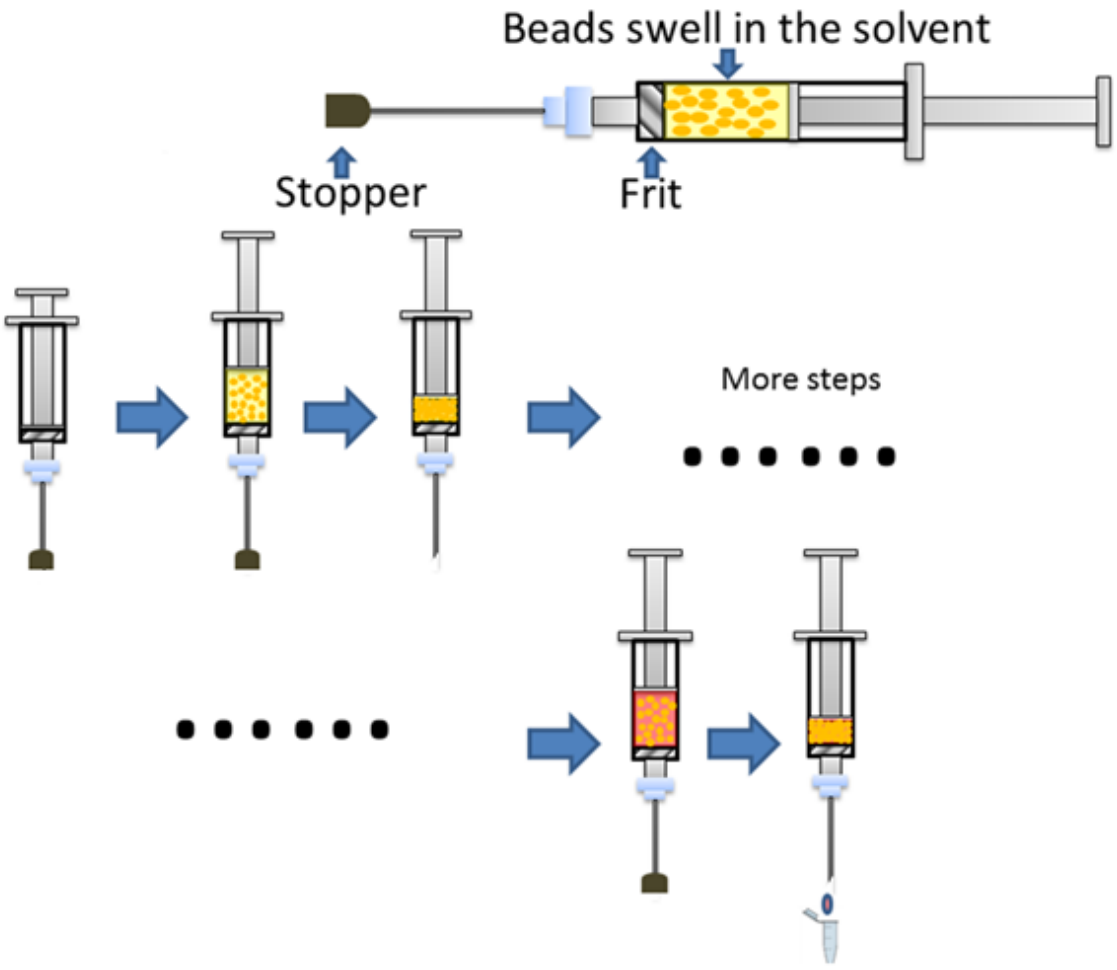
2.1 Introduction

In mass spectrometry, Electron-Transfer Dissociation (ETD) is a fragmentation technique [1] for peptide analysis that provides good sequence coverage while retaining post-translational modifications. These features make ETD a useful technique for studies of both regular peptides and modified peptides. ETD benefits from multiple charging to achieve efficient backbone fragmentation. Chemical derivatization, which introduces fixed-charge or readily chargeable groups,[2] can enhance multiple charging in electrospray. A few charge tags have been investigated on the basis of their synthetic convenience or chemical availability, e.g., trimethylammoniumalkyl,[3,4] 2,4,6-trimethylpyridinium,[5] tris(2,4,6-trimethoxyphenyl)phosphonium (TMPP),[6-8] and 2,2'-bipyridyl.[9] Guanidination methods,[10-16] which introduced a readily chargeable basic guanidine tag, can also potentially enhance charging.

In regards to ETD, an ideal charge tag would be able to [15][17] 1) enable the generation of a single but complete ion series to simply spectra; 2) enhance backbone N-C_α bond dissociations compared to underivatized peptides, and 3) be readily introduced into the peptide in a high yield. We designed new guanidine tags that selectively convert lysine residues to alkyl or aryl-modified homoarginines and allow facile and efficient solid-phase separation of lysine and arginine-terminated tryptic peptides, which simplified the peptide mixture. These new homo-arginine tags are compatible with the previously developed tunable charge tags (eg. DMAP [17]) for improved peptide charging and sequence coverage in ETD.

Solid phase synthesis is a method in which organic reactions are carried out on substrates that are covalently attached to the beads; and excess reactant or byproduct can be easily removed by washing. This technique has the advantage of high efficiency and easy purification, thus it was used to prepare guanidine-tagged peptides in our study. **Scheme 2.1** shows the way in which solid phase synthesis is performed. The beads are first stored inside the syringe and then swollen full in the solvent.

The coupling reagent is retracted into the syringe and the syringe is stirred or rotated for a period of time. When the reaction ends, the solvent inside the syringe contains the byproducts and excess reagents, and then the solution inside is ejected out of the syringe manually. After several washings with solvent, the modified beads inside the syringe are ready to be coupled with the next reagent. After all of the coupling steps, the modified peptide can be cleaved from the beads by strong acid and then dried with a stream of nitrogen.



Scheme 2.1 Reaction scheme of solid phase synthesis

Guanidination “strategy” is a method of introducing a readily chargeable basic guanidine tag into lysine containing peptide and can potentially enhance charging. During guanidination, the ε-amino group of lysine can be exclusively converted to homoarginine and the basicity of the peptide is increased. The

commonly used method, [10][12]-[15][18] shown in **Figure 2.1**, does have a limitation as histidine containing peptides undergo side reactions at the high pH needed for guanidination.

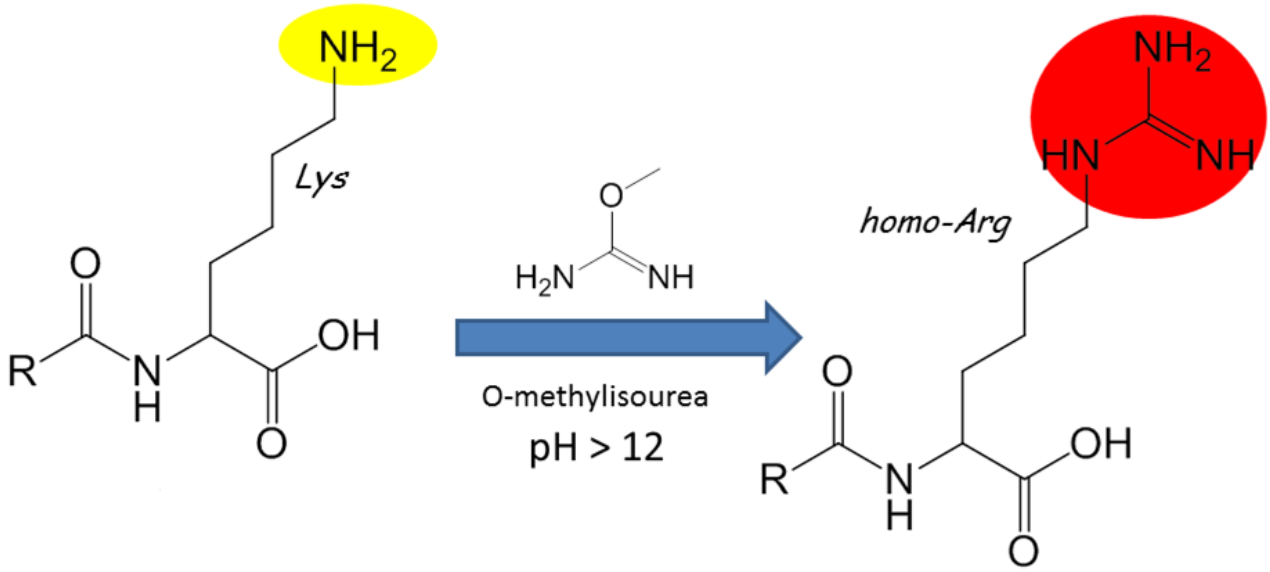


Figure 2.1 Commonly used method for guanidination

The solid phase strategy is used in our study to prepare custom tunable tags without the problems mentioned earlier. There are generally 4 steps for the preparation of the modified peptides which are derivatized by custom-tunable charge tags. First step is to prepare the lysine-specific reagent on a solid-phase resin. In this step, we can employ standard solid-phase peptide synthetic methods and make custom-tunable reagents. The second step is to enable the reaction of a mixture of lysine (K) and arginine (R) C-terminated tryptic peptides with the custom-tunable group already on the resin. After that, lysine-terminated peptides are derivatized and chemically bonded to the beads, while the arginine-terminated ones are washed off and analyzed separately. The last step is to elute lysine-derivatized peptides from the resin and to analyze the modified peptides by mass spectrometer. This methodology achieves the simplification of mixture analysis by separation of Lys (lysine) and Arg (argine) terminated tryptic peptides and the simplification of fragment ion assignment by anchoring one charge at the C-terminus. Additionally, this strategy shifts the *m/z* value of precursor and fragment ion in a custom-tunable fashion.

2.2 Experimental Section

2.2.1 Materials

All synthetic pentapeptides were custom-synthesized by CHI Scientific (Maynard, MA, USA) or NeoBioSci (Cambridge, MA, USA) and used as received. Disposable reaction vessels (empty syringe equipped with a frit, 3 mL) were purchased from CSPS Pharmaceuticals (San Diego, CA, USA). PAL-ChemMatrix resin was purchased from Biotage (Charlotte, NC, USA). Common chemicals (reagent grade) were purchased from Sigma-Aldrich (Milwaukee, WI, USA).

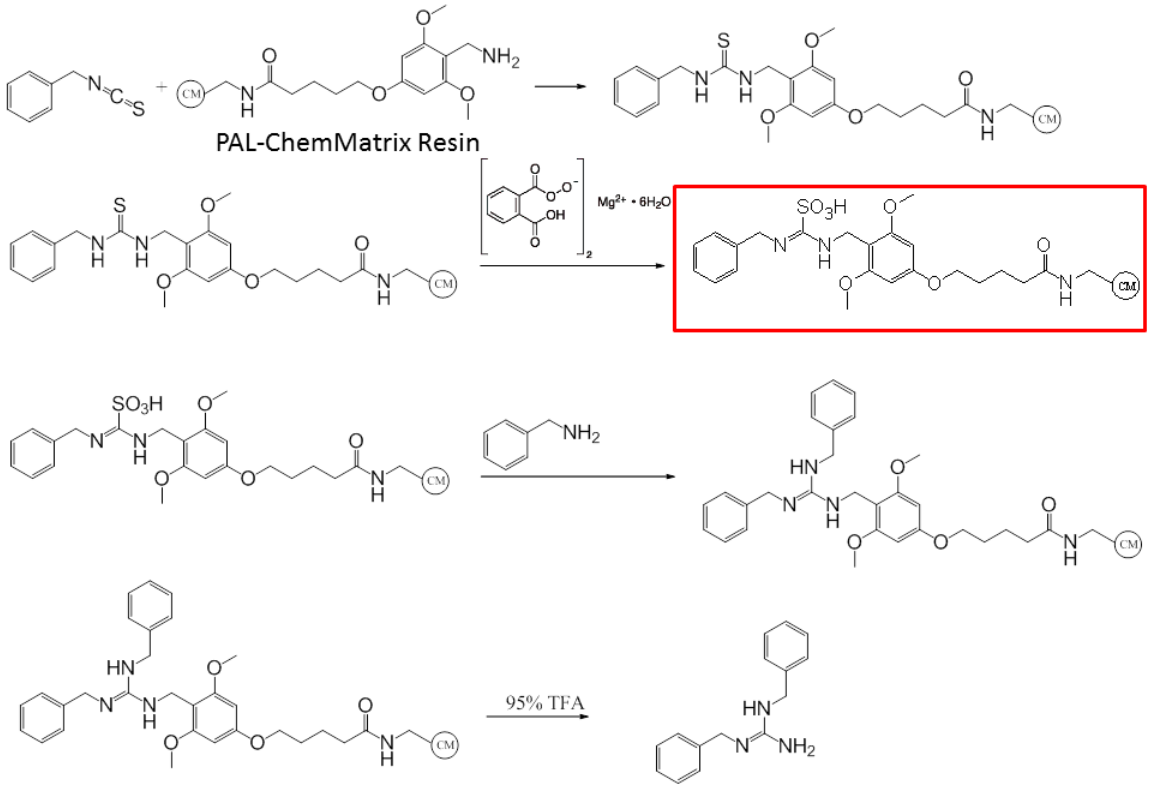
2.2.2 Preparation of the benzylguanidine tag precursor on beads

PAL-ChemMatrix resin (50 mg, 24 μmol) was weighed and put inside the frit equipped syringe and allowed to swell in dichloromethane (2 mL) for 1h. After sufficient swelling, the excess amount of dichloromethane was forced out of the syringe. Benzyl isothiocyanate (32 μl, 240 μmol) was diluted to 0.5 mL with dichloromethane and then was sucked into the syringe. The syringe was taped onto the rotary evaporator and rotated for 2 h. Following the reaction, excess reagent and byproducts were removed from the syringe by washing with dichloromethane (2.5 mL) five times (five minutes of rotation each time). Modified beads were further washed with dimethylformamide two times and then the excess amount of dimethylformamide was removed. Magnesium monoperoxyphthalate hexahydrate (0.0427 g, 86.4 μmol) was dissolved in 0.43 mL dimethylformamide, and then the solution was sucked into the syringe. The syringe was rotated for another 2 h and then washed with dimethylformamide five times (five minutes rotation each time). Then the syringe was stored in 4 °C, in which the syringe can be stored up to a few days while the beads are swollen well in dimethylformamide.

2.2.3 Use of model compound benzylamine for the test of the peptide coupling reaction

Benzylamine has been chosen as a model compound to react with the tag precursor on the beads so as to test the peptide coupling step. The guanidine tag precursor attached to the beads inside the syringe was washed with methanol two times (five minutes of rotation each time) and then the excess amount of methanol was removed. Benzylamine (48 μmol, 5.2 μL) was diluted with methanol (0.68 mL),

mixed with triethylamine (96 μmol, 13.4 μL), and then was sucked into the syringe. The syringe was stirred at room temperature (RT) for 64 h on a tube vortexer. After that, the syringe was washed with methanol five times (five minutes of rotation each time) and then washed with dichloromethane another five times (five minutes of rotation each time). The syringe, which now includes modified beads, was then dried over a mechanical vacuum pump for 10 minutes. A mixture (1 mL) of trifluoroacetic acid (TFA)/water/triethylsilane (95/2.5/2.5) was sucked into the syringe to cleave the product from the beads. After rotating for 1 h, the solution in the syringe was collected into a glass vial. The beads in the syringe were washed with TFA (1 mL) two additional times (five minutes of rotation each time). The washes were combined into the previous glass vial and dried with a stream of nitrogen. The product was identified with ESI-MS. The full synthetic route of benzylguanidine-tagged benzylamine is shown in **Scheme 2.2**.



Scheme 2.2 Reaction scheme of benzylguanidine-tagged compound with the use of benzylamine as the model compound

2.2.4 Pentapeptide coupling reaction

A syringe with the guanidine tag precursor bonded onto beads was washed with a solution mixture of MeOH/H_2O (80/20) four times (five minutes of rotation each time) followed by removal of excess solvent. Synthetic pentapeptides AAXAK (X = F, H, N, D, K) (0.5 mg each, 1 μmol each) were dissolved in a 100 μL (each) solution of MeOH/H_2O (80/20). AASAK (0.5 mg, 1 μmol) was dissolved in a 200 μL solution of MeOH/H_2O (80/20). The two peptide solutions above were combined and mixed with triethylamine (96 μmol, 13.4 μL), and then sucked into the syringe. The syringe was stirred on the top of a tube vortexer at RT for 64 hours. Following the reaction, the pentapeptides that were still in solution and thus unreacted were collected for future analysis. Then the syringe containing the modified beads was washed with MeOH/H_2O (80/20) five times (five minutes of rotation each time) and then washed with diethyl ether another three times (five minutes of rotation each time). The syringe which contains the modified beads was then dried over a mechanical vacuum pump for 3 hours. Following this, the mixture (1.5 mL) of TFA/water/triethylsilane (95/2.5/2.5) was pulled into the syringe to cleave the product from the dried beads. After rotating for 1 hour, the solution in the syringe now contains the eluted modified peptides cleaved from the beads and was collected in the glass vial. The beads in the syringe were washed with TFA (1 mL) another two times (five minutes rotating each time). The eluted solutions were combined into the previous glass vial, dried with a stream of nitrogen and redissolved in 0.5 mL methanol. A portion of the modified peptides that dissolved in methanol were dried and then redissolved either in acetonitrile/water/acetic acid (49.5/49.5/1) for direct infusion to Bruker mass spectrometer and Thermo Fisher LTQ XL linear ion trap instrument with ETD; or in water/acetic acid (99/1) for further LC-MS on Waters Quattro Micro triple quadruple instrument.

2.2.5 Mass Spectrometry

Electron-transfer dissociation mass spectra were collected on a Thermo Fisher (San Jose, CA, USA) LTQ XL linear ion trap instrument, outfitted with a chemical ionization source for the generation of fluoranthene radical anion as ETD reagent. Precursor ions were mass isolated with a window of 0.6-2 *m/z* units to accommodate nearest ^{13}C isotopologues and allowed to react for 200 milliseconds with reagent ions.

ESI-LC-MS was performed on a Waters Quattro Micro triple quadruple instrument coupled with the system of 1525u LC pump and 2777 autosampler (Waters, Milford, MA). The mobile phase was mixed with solvent A (water/formic acid (99.9/0.1) by volume) and solvent B (acetonitrile/water/formic acid (94.9/5/0.1) by volume) and eluted at a flow rate of 0.2 mL/min according to the following linear gradient program: initial 1.1% B; 9 min 52.6% B; 12 min100% B; 14min 100% B; 14.1min 1.1% B; 22min, 1.1% B.

Bruker Esquire ion trap mass spectrometer (Bruker Daltonics, Billerica, MA) was used for the direct infusion mass spectrometer analysis. Mass spectrometer settings were as follows: capillary exit, 45.0 V; Skim 1, 30.0 V; Trap Drive, 55.0 V.

2.2.6 Study of the impact from temperature and reaction time to peptide coupling step

We prepared three identical syringes containing benzylguanidine tag precursor and started three parallel reactions in different reaction conditions: 1). RT, 64 hours; 2). 37 °C, 16 hours; 3). 37 °C, 64 hours. The three samples were processed in the same way as mentioned above and same aliquot of each sample was loaded on the LC column and then injected into a Waters Quattro Micro triple quadruple mass spectrometer for LC-MS analysis. A peptide mixture which contains AAXAK (X = F, H, N, D, K, S) was also analyzed by LC-MS as the standard for this study.

2.3 Results and Discussion

2.3.1 Benzylamine coupling reaction

Benzylamine contains a primary amino group at the end like lysine terminated peptide, and therefore was used as a model compound for lysine terminated peptides to react with the benzylguanidine tag precursor bonded on beads. An excess amount of benzylamine was used to react with the beads inside the syringe to ensure all guanidine tag precursor bonded on beads are fully reacted with benzylamine. Full MS spectrum of this product with electrospray ionization (ESI) shows intensive signal molecular ion peak at *m/z* 240.2 (**Figure 2.2**), which agrees well with the *m/z* of singly charged benzylguanidine-tagged benzylamine.

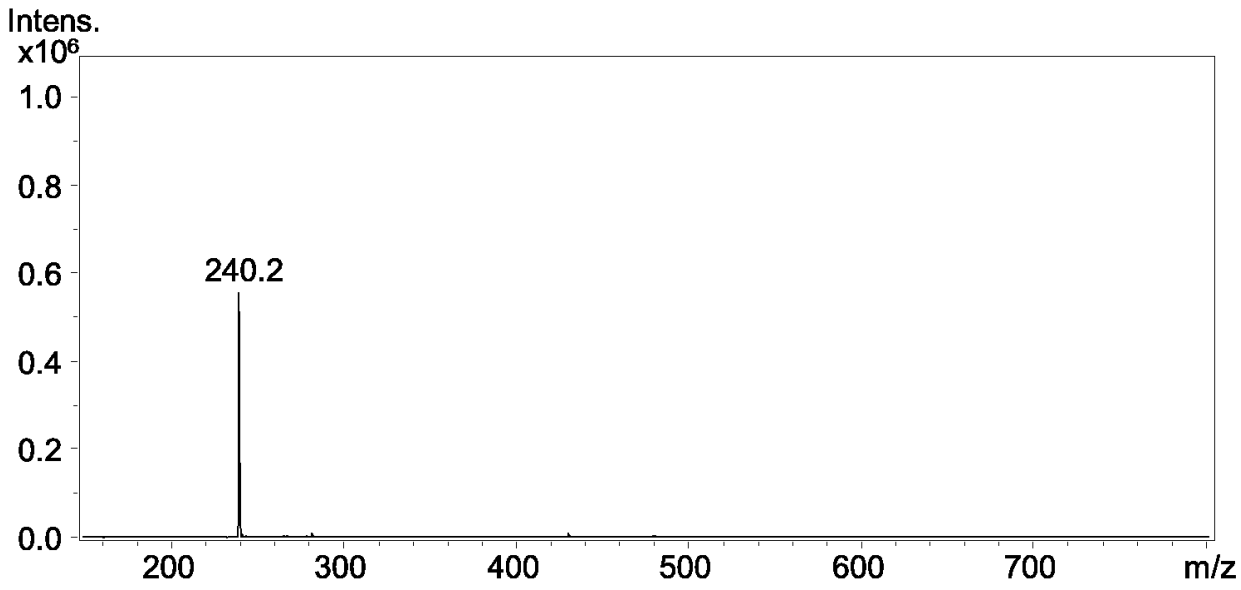


Figure 2.2 ESI-MS spectrum of the product from benzylamine coupling reaction

For MS/MS analysis (**Figure 2.3**) of the major peak (*m/z* = 240.0) in the full MS scan, a dominant peak (*m/z* = 108.1) corresponds to the loss of benzylguanidine-tag and a predominant peak (*m/z* = 181.0) corresponds to the benzyl group, indicating that the major peak of *m/z* = 240.0 in the full MS spectrum is benzylguanidine-tagged benzylamine.

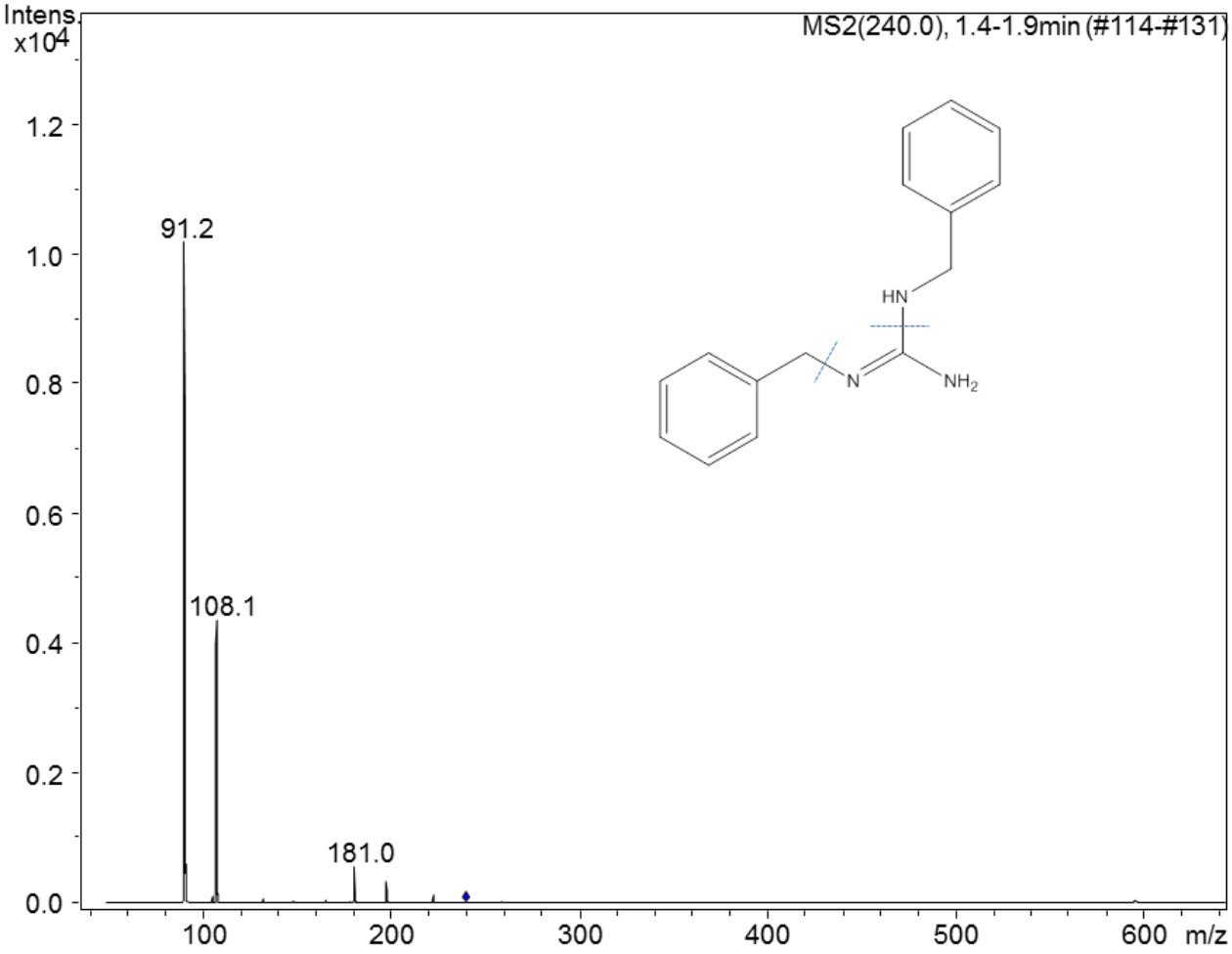


Figure 2.3 MS/MS spectrum of singly charged benzyl-guanidine tagged benzylamine (*m/z* =240.0)

2.3.2 Pentapeptides coupling reaction with benzyl-guanidine tag precursor

2.3.2.1 Efficiency of the method

All of the benzylguanidine-tag attached synthetic pentapeptides were analyzed and identified with LC-MS analysis. A clear separation of unmodified peptides from modified peptides bonded on beads is crucial for our future quantitation study, thus we are interested if unreacted pentapeptides can be efficiently removed by multiple washings. An aliquot of the mixture comtaining six synthetic pentapeptides was saved as the standard and set aside. Another aliquot of the benzylguanidine-tag derivatized pentapeptides cleaved from the beads was saved as well. Theoretically, if all of the unreacted peptides do not remain on the beads, they can be fully removed and will not be detected in the tag derivatized peptides sample. Both of the aliquots were loaded on column to do comparative LC-MS runs. The

amounts of each underivatized pentapeptide analyzed with the two LC-MS runs are compared. The results show that less than 1% of each type of unreacted peptide stays in the sample of benzylguanidine-tag derivatized pentapeptides, indicating that a majority of unreacted peptides at the end of the coupling step can be sufficiently removed by multiple washes.

The conversion ratio of pentapeptides being transformed into benzylguanidine-tag modified peptides was studied as well and calculated based on the comparison between the amounts of underivatized pentapeptide in the standard and those in the wash. The result is as below: AAXAK (X = F, 18.0%; H, 68.5%; N, 51.2%; D, 46.2%; K, 84.5%; S, 56.7%). This result shows that the conversion ratio has a median over 50% and could reach up until 84.5% in this pool, and therefore shows that the tag could be easily introduced into the peptide.

2.3.2.2 Test of reaction conditions for peptide coupling step

After all the benzylguanidine-tag attached pentapeptides were recognized from reaction conditions of 64 hours at room temperature, there arises an interest to determine if 64 hours is necessary for this reaction and if the reaction can be hastened by increasing the temperature. Therefore, parallel peptide coupling reactions were conducted under three conditions: 1) RT, 64 h; 2) 37 °C, 16 h; 3) 37 °C, 64 h. The amount of each benzylguanidine-tag modified peptide were quantified by LC-MS and compared among three reaction conditions (**Table 2.1**). The peak area of each modified peptide in condition 1 was standardized as 1.00, and the ratio of peak area in the other two conditions relative to that of condition 1 was calculated for each modified pentapeptide. Since both of the singly charged and doubly charged benzylguanidine-tagged peptides are identified in the mass spectra, the ratios of peak area in other conditions relative to condition 1 are calculated separately for each charge state. **Table 2.1** shows that there is not a big increase in the amount of modified pentapeptides when the temperature is increased to 37 °C while keeping the same reaction time of 64 h. Meanwhile, the time is an important consideration, and under the conditions of 37 °C, 16 h can be used instead of 64 h while still ensuring that a conversion of about 50% of the amount of modified peptides when compared to 64 h.

Modified peptide (2+)	1. RT 64h	2. 37 °C 16h	3. 37 °C 64h
AAFAK-tag	1.00	0.53	1.00
AAHAK-tag	1.00	0.54	0.85
AANAK-tag	1.00	0.66	1.00
AADAK-tag	1.00	0.73	1.09
AAKAK-tag	1.00	0.49	1.20
AASAK-tag	1.00	0.60	0.98

Modified peptide (1+)	1. RT 64h	2. 37 °C 16h	3. 37 °C 64h
AAFAK-tag	1.00	0.59	0.93
AAHAK-tag	1.00	0.51	0.92
AANAK-tag	1.00	0.60	1.06
AADAK-tag	1.00	1.32	1.34
AAKAK-tag	1.00	0.53	1.08
AASAK-tag	1.00	0.49	0.94

Table 2.1 Comparison of modified peptides quantities among three reaction conditions for separate charge states (z = 1, z = 2).

Table 2.2 shows the conversion ratio from pentapeptides to the benzylguanidine-tagged format. A known amount of pentapeptide before peptide coupling is saved as a standard and a corresponding amount from peptide solution after peptide coupling is saved as well to do quantitative LC-MS for an estimate of reaction yield. Since pentapeptides are observed both in singly charged and doubly charged formats, the conversion ratio is calculated separately based on each charge state. The first reaction condition, which has reaction time of 64 h at room temperature, yields a maximum amount of benzylguanidine-tag modified peptide among three reaction conditions. Therefore, this reaction condition is selected to be used in further studies.

conversion ratio for peptide coupling (2+)	1. RT 64h	2. 37 °C 16h	3. 37 °C 64h
AAFAK	18.0%	9.8%	4.3%
AAHAK	68.5%	62.4%	63.2%
AANAK	51.2%	36.4%	39.6%
AADAK	46.2%	33.1%	31.1%
AAKAK	84.5%	78.8%	79.6%
AASAK	56.7%	42.8%	45.6%

conversion ratio for peptide coupling (1+)	1. RT 64h	2. 37 °C 16h	3. 37 °C 64h
AAFAK	39.9%	33.0%	20.8%
AAHAK	76.4%	71.1%	77.8%
AANAK	61.7%	46.4%	48.6%
AADAK	61.0%	48.6%	48.2%
AAKAK	92.0%	87.5%	89.2%
AASAK	58.8%	46.8%	49.2%

Table 2.2 Conversion ratio of pentapeptides to the benzylguanidine-tag derivatized format.

2.3.2.3 Impact from chemical structure of guanidine tags on the ETD fragmentation pattern

Table 2.3 displays the chemical structure of three guanidine tag precursors and the corresponding structures of the derivatized peptides. Guanidination happens on lysine terminated tryptic peptides selectively at the ε-amino group of the lysine.

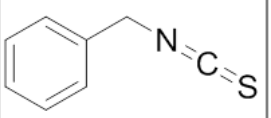
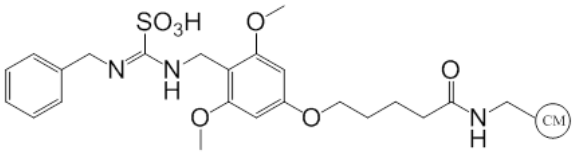
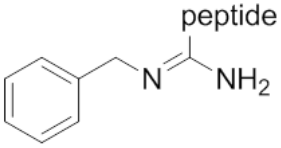
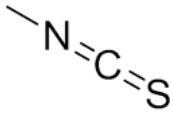
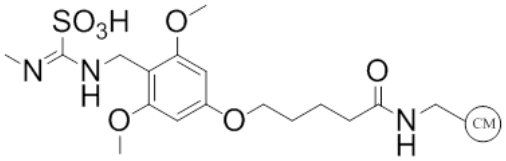
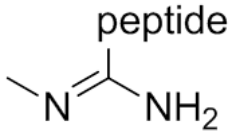
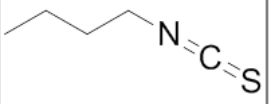
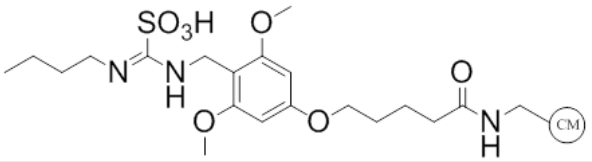
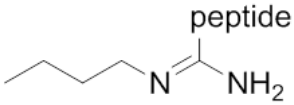
different starting material	modified beads awaiting coupling to peptides	tagged peptide
		
		
		

Table 2.3 Chemical structures of starting material, guanidine tags precursor and tagged peptide for benzylguanidine, methylguanidine and butylguanidine tags.

Figure 2.4 shows the ETD spectrum of doubly charged pentapeptide AAFAK and the guanidine tagged AAFAK. ETD of benzylguanidine-, methylguanidine- and butylguanidine-tagged peptide ions all result in a complete and dominant *z*-ion series, thus largely improving z-ion based peptide sequence coverage. Meanwhile, the intensity of *c* ions from guanidine tagged peptide ions decreases greatly when compared to the underivatized peptide. Therefore, complete but simplified spectra are acquired in this case. Loss of tags in two different types, i.e. tag1 and tag2, are observed but are not dominant peaks in the spectra.

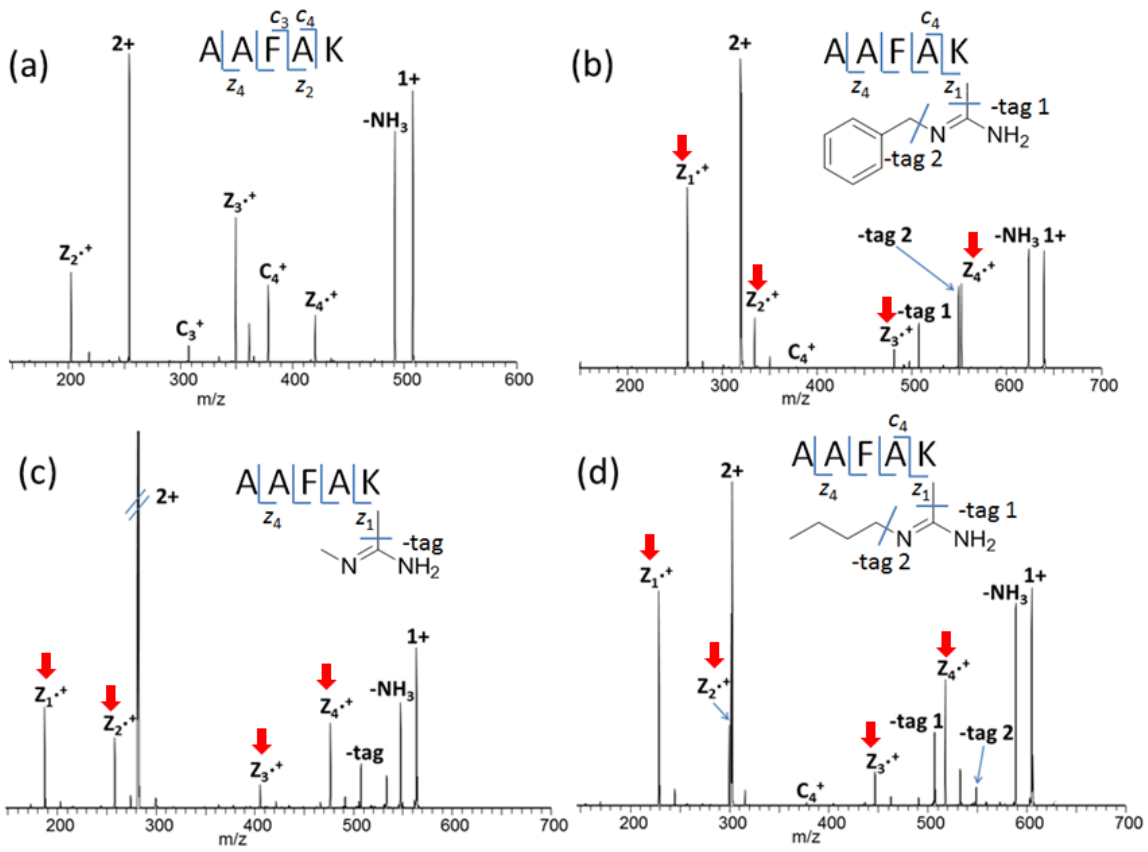


Figure 2.4 ETD spectra of doubly charged (a) underivatized peptide ion AAFAK; (b) benzylguanidine-tagged AAFAK; (c) methylguanidine-tagged AAFAK; (d) butylguanidine-tagged AAFAK

Figure 2.5 shows the ETD spectra of doubly charged underivatized peptide ion AAKAK and guanidine-tags modified AAKAK. Underivatized AAKAK has difficulty to be fully sequenced with a high confidence level since the z_2 ion is an extremely weak fragment peak. However, all three guanidine-tagged AAKAK peptides can be easily fully sequenced as all of them generate a complete series of *z* ions. As there are two lysines in the peptide sequence AAKAK, either of them can be labeled with the guanidine tag. Guanidine-tag modified peptide ions which are labeled at the third lysine and labeled at the fifth lysine have identical *m/z*, but the *m/z* of their fragment ions generated by ETD are not exactly the same. Thus from the ETD spectra, both sequence information and the information about the labeling position can be acquired at the same time.

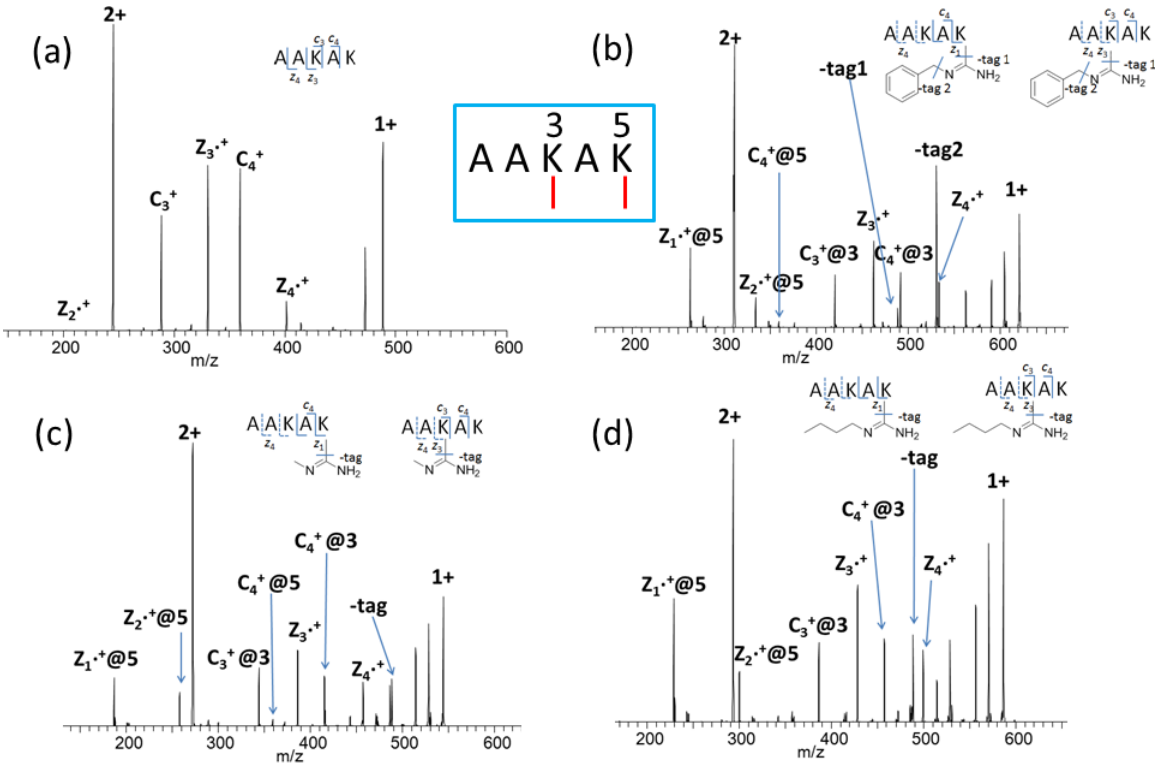


Figure 2.5 ETD spectra of doubly charged (a) underivatized peptide ion AAKAK; (b) benzylguanidine-tagged AAKAK; (c) methylguanidine-tagged AAKAK; (d) butylguanidine-tagged AAKAK

All six doubly charged benzylguanidine-tagged pentapeptide ions (AAXAK-benzylguanidine tag, X = F, H, N, D, K, S) generated a complete series of *z* ions with ETD, and therefore allows the benzylguanidine tag to be considered as the best guanidine tag in this study. Five doubly charged methylguanidine/butylguanidine-tagged peptide ions (AAXAK-methylguanidine tag/butylguanidine tag, X = F, H, N, K, S) also generated a complete series of *z* ions with ETD. However, in the case of AADAK, the intensity of z_3 fragment ion is very low in ETD spectrum of methylguanidine-tagged peptide ion (**Figure 2.6**); z_3 and z_4 ions are missing in the ETD spectrum of butylguanidine-tagged peptide ion. Therefore, benzylguanidine tag has the best performance among the three tags used to study our sample of six synthetic pentapeptides with a *C*-terminal lysine.

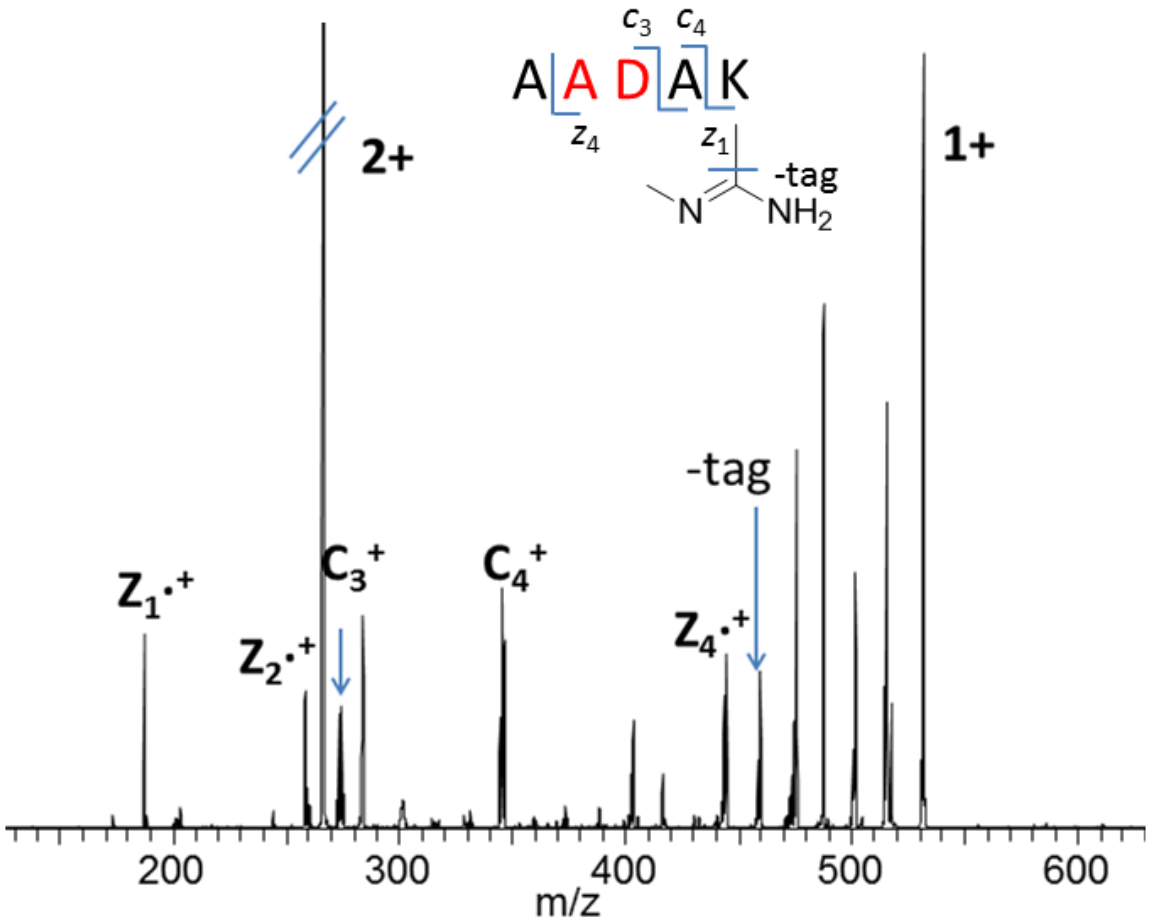


Figure 2.6 ETD spectrum of doubly charged methylguanidine-tagged AADAK

Since the guanidine tags developed here only targets at ε-amino group of lysine, the mass spectrometry analysis of tryptic peptide mixtures digested from proteins with either lysine (K) or arginine (R) *C*-terminations can be simplified by analyzing two portions with different terminals. The arginine terminated peptides can be washed off the beads and reacted with DMAP tag [17] later to have the tag located on the *N*-terminus. In contrast, the lysine terminated peptides are guanidine modified on the *C*-terminal lysine and will be retained on the beads so that they can be cleaved later for analysis on mass spectrometer separately from arginine terminated peptides.

2.4 Conclusion

Reducing fragmentation mass spectra to give the most succinct information, ideally originating from a single but complete ion series, will allow easier interpretation either by manual inspection or by automated search algorithms. Under ESI conditions, peptide fragment ion formation and observation is ultimately linked to protons. The “mobile” proton weakens the peptide bonds, helps to induce bond cleavage, and allows observation of fragment ions. Which fragment ions are observed is controlled by the average residence site of the protons. Change of the basicity or acidity will impact proton localization and thus the peptide fragmentation. The guanidine tags that we developed increase the basicity of the *C* terminus at lysine-terminated tryptic peptides, allow the generation of a full series of ***z*** ions, thus simplifying the ETD spectra of tryptic peptides. This will result in easier sequence identification of proteins. In our study, sequence coverage of tagged peptide ions with ETD generally increased when compared to those of underivatized ones. Meanwhile, all the benzyl-guanidine tag modified peptides generate a complete series of *z* ions with ETD, and benzylguanidine tag is considered to be the best tag in the case of six synthetic pentapeptides.

2.5 References

[1] 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. USA*. **2004**, *101*, 9528-9533.

[2] Roth, K. D. W.; Huang, Z.H.; Sadagopan, N.; Watson, J. T. Charge Derivatization of Peptides for Analysis by Mass Spectrometry. *Mass Spectrom. Rev.* **1999,** *17,* 255-274.

[3] Ren, D.; Julka, S.; Inerowicz, H. D.; Regnier, F. E. Enrichment of Cysteine-containing Peptides from Tryptic Digests Using a Quaternary Amine Tag. *Anal. Chem*. **2004,** *76,* 4522-4530.

[4] Gunawardena, H. P.; Gorenstein, L.; Erickson, D. E.; Xia, Y.; McLuckey, S. A. Electron Transfer Dissociation of Multiply Protonated and Fixed Charge Disulfide Linked Polypeptides. *Int. J. Mass Spectrom.* **2007,** *265,* 130-138.

[5] Vath, J. E.; Biemann, K. Microderivatization of Peptides by Placing a Fixed Positive Charge at the N-terminus to Modify High Energy Collision Fragmentation. *Int. J. Mass Spectrom. Ion Processes.* **1990,** *100*, 287-299.

[6] Shen, T. L.; Allison, J. Interpretation of Matrix-Assisted Laser Desorption/Ionization Postsource Decay Spectra of Charge-derivatized Peptides: Some Examples of Tris[(2,4,6- trimethoxyphenyl) phosphonium]-tagged Proteolytic Digestion Products of Phosphoenol-pyruvate Carboxykinase. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 145-152.

[7] Chamot-Rooke, J.; van der Rest, G.; Dalleu, A.; Bay, S.; Lemoine, J. The Combination of Electron Capture Dissociation and Fixed Charge Derivatization Increases Sequence Coverage for O-Glycosylated and O-Phosphorylated Peptides. *J. Am. Soc. Mass Spectrom.* **2007,** *18,* 1405-1413.

[8] Chamot-Rooke, J.; Malosse, C.; Frison, G.; Tureček, F. Electron Capture in Charge-tagged Peptides. Evidence for the Role of Excited Electronic States. *J. Am. Soc. Mass Spectrom.* **2007,** *18,* 2146-2161.

[9] Jones, J. W.; Sasaki, T.; Goodlett, D. R.; Tureček, F. Electron Capture in Spin-Trap Capped Peptides. An Experimental Example of Ergodic Dissociation in Peptide Cation-Radicals. *J. Am. Soc. Mass. Spectrom.* **2007,** *18,* 432-444.

[10] Brancia, F.L., Oliver, S.G., Gaskell, S.J. Improved Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis of Tryptic Hydrolysates of Proteins Following Guanidination of Lysine-containing Peptides. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2070–2073.

[11] Peters, E.C., Horn, D.M., Tully, D.C., Brock, A. A Novel Multifunctional Labeling Reagent for Enhanced Protein Characterization with Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2387–2392.

[12] Beardsley, R.L., Reilly, J.P. Optimization of Guanidination Procedures for MALDI Mass Mapping. *Anal. Chem*. **2002**, *74*, 1884–1890.

[13] Ji, C., Guo, N., Li, L. Differential Dimethyl Labeling of N-termini of Peptides after Guanidination for Proteome Analysis. *J. Proteome Res.* **2005**, *4*, 2099–2108.

[14] Warwood, S., Mohammed, S., Cristea, I.M., Evans, C., Whetton, A.D., Gaskell, S.J. Guanidination Chemistry for Qualitative and Quantitative Proteomics. *Rapid Commun. Mass Spectrom*. **2006**, *20*, 3245–3256.

[15] Hennrich, M.L., Boersema, P.J., van den Toorn, H., Mischerikow, N., Heck, A.J.R., Mohammed, S. Effect of Chemical Modifications on Peptide Fragmentation Behavior upon Electron Transfer Induced Dissociation. *Anal. Chem*, **2009**, *81*, 7814–7822.

[16] Miyashita, M., Hanai, Y., Awane, H., Yoshikawa, T., Miyagawa, H. Improving Peptide Fragmentation by N-terminal Derivatization with High Proton Affinity. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1130–1140.

[17] Zimnicka, M.; Moss, C. L.; Chung, T. W. ; Hui, R.; Tureček, F. Tunable Charge Tags for Electron-Based Methods of Peptide Sequencing: Design and Applications. *J.Am. Soc. Mass Spectrom.* **2012**, *23*, 608-620.

[18] Margolis, S.; Langdon, R. G. Chemistry and Metabolism of Macromolecules: Studies on Human Serum β_1-Lipoprotein: II. Chemical Modifications *J. Biol. Chem.* **1966**, *241*, 477-484.

Chapter 3
Comparison of the ETD Performance of Benzyl/Methyl/Butylguanidine-tagged Peptides

3.1 Introduction

After our initial performance evaluation of the benzyl/methyl/butyl guanidine tags for labeling of synthetic pentapeptides, we investigated the performance of these custom-tunable tags applied to peptides digested from native proteins. The protein Bovine Serum Albumin (BSA) contains multiple lysine (K) sites and can generate multiple lysine terminated peptides using a trypsin digestion, and is therefore selected for use in this study.

3.2 Experimental section

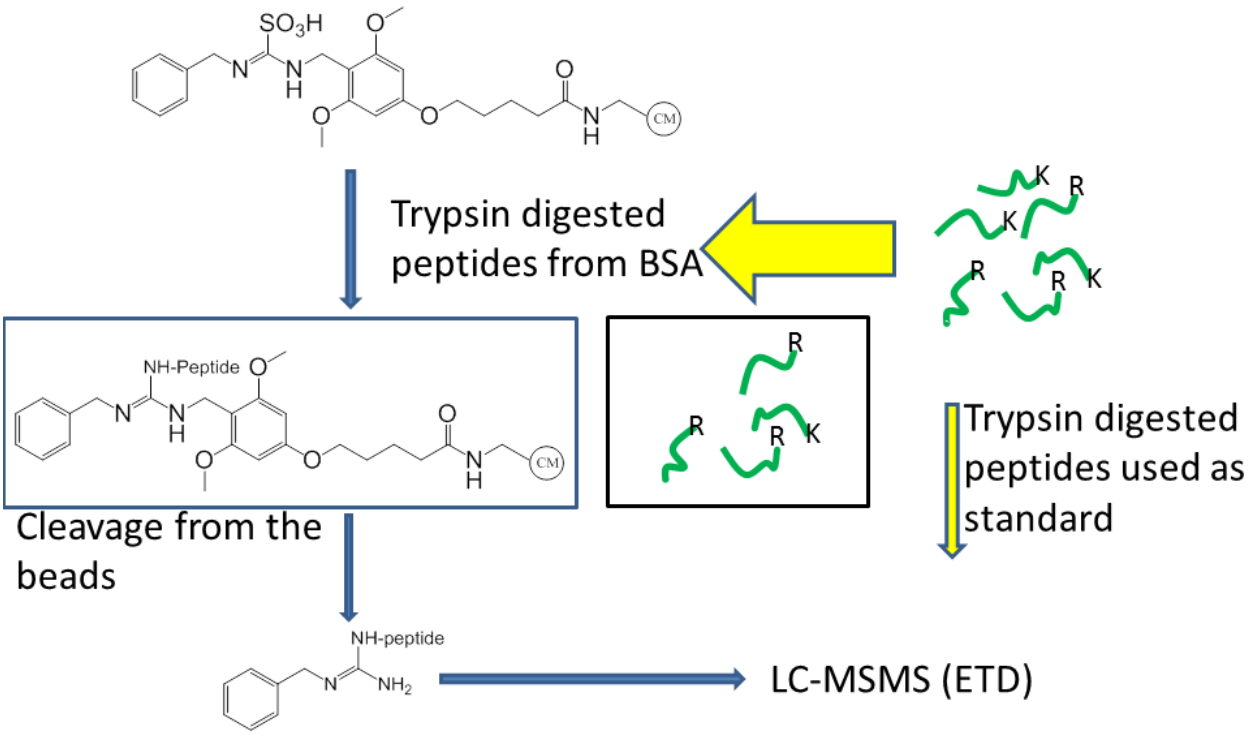
3.2.1 Materials

Bovine Serum Albumin (BSA) was purchased from Sigma-Aldrich (Milwaukee, WI, USA). Disposable reaction vessels (empty syringe equipped with a frit) were purchased from CSPS Pharmaceuticals (San Diego, CA, USA). PAL-ChemMatrix resin was purchased from Biotage (Charlotte, NC, USA). Trypsin Gold (Mass Spectrometry Grade) was purchased from Promega Corporation (Fitchburg, WI, USA). Column packing material Magic 5 µ 200 Å C18 AQ and Magic 5 µ 100 Å C18 AQ were purchased from Bruker-Michrom (Auburn, CA, USA). Self-Pack IntegraFrit Columns (Part Number IF360-100-50-N-5) were purchased from NewObjective Inc. (Woburn, MA, USA). Capillary tubing (Product Descriptor TSP075375-10M) was purchased from Polymicro Technologies (Phoenix, AZ, USA). Sep-Pak C18 (1 cc Vac Cartridge, 50 mg Sorbent per Cartridge, 55-105 µm Particle Size, Product Number: WAT054955) were purchased from Waters (Milford, MA, USA). Other common chemicals (reagent grade) were purchased from Sigma-Aldrich (Milwaukee, WI, USA).

3.2.2 Preparation method of benzylguanidine-tagged BSA tryptic peptides

The benzylguanidine-tag precursor was prepared in a disposable reaction vessel (syringe) according to the same method described previously, starting with 25 mg PAL-ChemMatrix Resin. Ammonium bicarbonate (0.1 M) in water was used as the solvent to prepare three separate stock solutions of 10 mg/mL BSA, 0.5 M tris(2-carboxyethyl) phosphine (TCEP) and 1 M iodoacetamide (IAA). Acetic acid (50 mM) in water was used to dissolve trypsin according to the Trypsin Gold Protocol provided from Promega. Four Eppendorf tubes of BSA tryptic peptides were prepared by the following method: In each tube, 91.3 µL of 10 mg/mL BSA was mixed with 821.3 µL 100 mM ammonium bicarbonate and then 9.1 µL of 0.5 M TCEP (final concentration: 5 mM) was added into the tube and the tube was incubated at 37 °C for 30 minutes to break disulfide bonds within and between proteins. Then, 9.1 µL of 1 M IAA (final concentration: 10 mM) was added to each tube and all the tubes were vortexed at RT for 10 minutes to enable covalent binding of IAA to cysteine in the protein. After that, 22.8 µL of 0.2 µg/µL trypsin (final trypsin: protein ratio of 1:200) was added to each tube and all the four tubes were incubated at 37 °C overnight. Following this digestion, each tube was acidified to pH = 3 tested by pH paper by adding formic acid (~ 7 µL). Trypsin digested peptides from two tubes were combined to be desalted with a single Sep-Pak C18 cartridge. Two Sep-Pak C18 cartridges were used according to the protocol provided from Waters as described in the following. First, the cartridge was conditioned with 1 mL of Acetonitrile (ACN) at a 5-10 mL/min flow rate. Second, the cartridge was equilibrated with 2 mL of ACN/H_2O/TFA (2/97.9/0.1) mixture at a 5-10 mL/min flow rate. Third, the sample was loaded from two eppendorf tubes into the cartridge with reloading each sample twice at 1 mL/min flow rate. Fourth, the cartridge was washed with 3 mL of ACN/H_2O/TFA (2/97.9/0.1) mixture at a 5-10 mL/min flow rate. Finally, BSA digested peptides in the cartridge were first eluted with 1 mL of ACN/H_2O/acidic acid (30/69.5/0.5) mixture then eluted with 1 mL of ACN/H_2O/acidic acid (69.5/30/0.5) mixture at a 1 mL/min flow rate. The 4 mL of eluates from the two cartridges were combined and were dried with a stream of nitrogen, and were redissolved in 0.8 mL of a MeOH/H_2O (50/50) solvent mixture. 8 µL of the peptide solution was taken out and saved as a standard for future LC-MS/MS analysis of BSA tryptic peptides standard. Triethylamine (6.7 µL, 48 µmoles) was mixed with the remaining BSA tryptic peptide solution, which was then sucked into the syringe with modified beads. The syringe was stirred at RT for 64 h. Following binding of the peptides to

the beads, the unreacted peptides left in the solution of the fritted syringe were collected for later analysis. Next, the beads in the syringe were washed with MeOH/H_2O (50/50) five times (five minutes stirring each time) and then washed with diethyl ether another three times (five minutes stirring each time). The syringe was then dried under a mechanical vacuum pump for 2 h. A mixture (0.75 mL) of TFA/ H_2O/triethylsilane (95/2.5/2.5) was introduced into the syringe to cleave the modified peptides from the beads. After rotating for 1 h, the solution in the syringe was collected in a glass vial. The beads in the syringe were washed with TFA (0.5 mL × 2) another two times (five minutes stirring each time) to enable the cleavage of the modified peptides from the beads into the solution. The washes were combined into the previous glass vial, dried under nitrogen and redissolved in 0.25 mL methanol. The 47 μL of the modified peptides dissolved in methanol were dried and then redissolved in 100 μL (estimated to give roughly 100 μM for modified digested peptides) of H_2O/acetic acid (99/1) for further data dependent LC-MS/MS analysis with ETD or CID fragmentations. The BSA tryptic peptide standard collected before was dried and then redissolved in 12.5 μL (estimated to give 100 μM for digested peptides) of H_2O/acetic acid (99/1). The work flow is shown in **Scheme 3.1**.



Scheme 3.1 Coupling of BSA tryptic peptides with the benzylguanidine-tag precursor

3.2.3 Performance evaluation of benzyl-guanidine tagged BSA tryptic peptides

A Thermo LTQ XL™ mass spectrometer with ETD (Thermo Scientific, Waltham, MA, USA) was coupled with a Waters nanoACQUITY UPLC system to perform LC-MS/MS sample analysis. An analytical column and trap column were prepared in house with materials mentioned above. The BSA tryptic peptides standard prepared above (12.5 µL) was diluted at the ratio of 1:50 (estimated to give 2 µM for digested peptide), then 1.3 µL from the diluted sample vial was injected to perform LC-MS/MS coupled with ETD. The benzyl-guanidine modified BSA tryptic peptides (100 µL) prepared before were diluted at a ratio of 1:50 (estimated to give roughly 2 µM for modified digested peptides), then 9.9 µL from this diluted sample vial was injected to perform LC-MS/MS with ETD. Within the parameters of the mass spectrometer, every cycle contains 6 scans, including 1 full MS scan and 5 subsequent data dependent MSMS ETD scans. The five most abundant peaks that were not in the exclusion list from the 1^{st} scan were picked and further dissociated with ETD from scans 2 to 6. Salient parameters of ETD fragmentation are shown as follows: ETD reaction time 200 milliseconds (ms), MS^n AGC target 1e4, reagent vial 1 AGC target 1e6, isolation width 3 units in m/z. The liquid chromatography (LC) was run with a binary solvent: solvent A is H_2O/formic acid (FA) (99.9/0.1), solvent B is ACN/FA (99.9/0.1). Sample was trapped at 2 µL/min with 95% A for 10 minutes and then separated with the following gradients: 1) for the sample of standard BSA tryptic peptides or the sample of unreacted (collected from the wash) BSA tryptic peptides, 0-30 min 5% B — 35% B; 30-31min 35% B — 75% B; 31-40 min 75% B; 40-41 min 75% B — 5% B; 41-60 min 5% B. 2) for benzyl-guanidine tagged BSA tryptic peptides, 0-60 min 5% B — 65% B; 60-61 min 65% B — 75% B; 61-70 min 75% B; 70-71 min 75% B — 5% B; 71-90 min 5% B.

The collected LC-MS/MS raw data files were converted to the mgf file format using MassSpecUtils (available at the UWPR website), searched with Mascot (Version 2.3.0) (www.matrixscience.com), and confirmed by manual inspection. The search was done with a database which includes the sequences from both BSA and yeast. Other search parameters were included as below: type of search MS/MS ion search, enzyme trypsin/P, fixed modifications carbamidomethyl (C), mass values monoisotopic, peptide mass tolerance ±1.5 Da, fragment mass tolerance ±1Da, maximum

missed cleavages 1, instrument type-ETD-Trap. In the search for hits of benzyl-guanidine modified peptides, the option of Benzyl-cx (K) was chosen as a variable modification.

3.2.4 Preparation method and performance evaluation of Methyl-guanidine/Butyl-guanidine tagged BSA tryptic peptides

Methylguanidine- and *n*-butylguanidine-tagged BSA tryptic peptides were prepared and analyzed by LC-MS/MS with ETD in the same way as described for benzylguanidine-tagged peptides above, with the only difference being the starting material was changed from benzyl isothiocyanate to methyl isothiocyanate and *n*-butyl isothiocyanate. The variable modification was changed from Benzyl-cx (K) to Methyl-cx (K) and Butyl-cx (K) correspondingly.

3.3 Results and Discussion

3.3.1 Improved sequence coverage in ETD for benzyl-guanidine tag modified peptides

Improved sequence coverage was found in the ETD of benzylguanidine-derivatized peptide ions when compared with the ETD of underivatized peptide ions. In the case of the peptide AEFVEVTK (**Figure 3.1**), the top spectrum is the MS/MS with ETD of doubly charged underivatized AEFVEVTK and the bottom spectrum is the MS/MS with ETD of doubly charged benzylguanidine-tagged AEFVEVTK. A complete series of ***z*** ions were observed in tandem MS of this tagged peptide ion, which simplified the spectra interpretation. In this case, in addition to the expected binding to the terminal lysine (K) the benzyl-guanidine tag was also found to attach to the *N*-terminal amino alanine (A) residue; however, only a very small portion of the tag reacted with the alanine NH_2 group. The extremely weak fragment peaks generated from the product with labeling position on A are marked by asterisks in the spectrum.

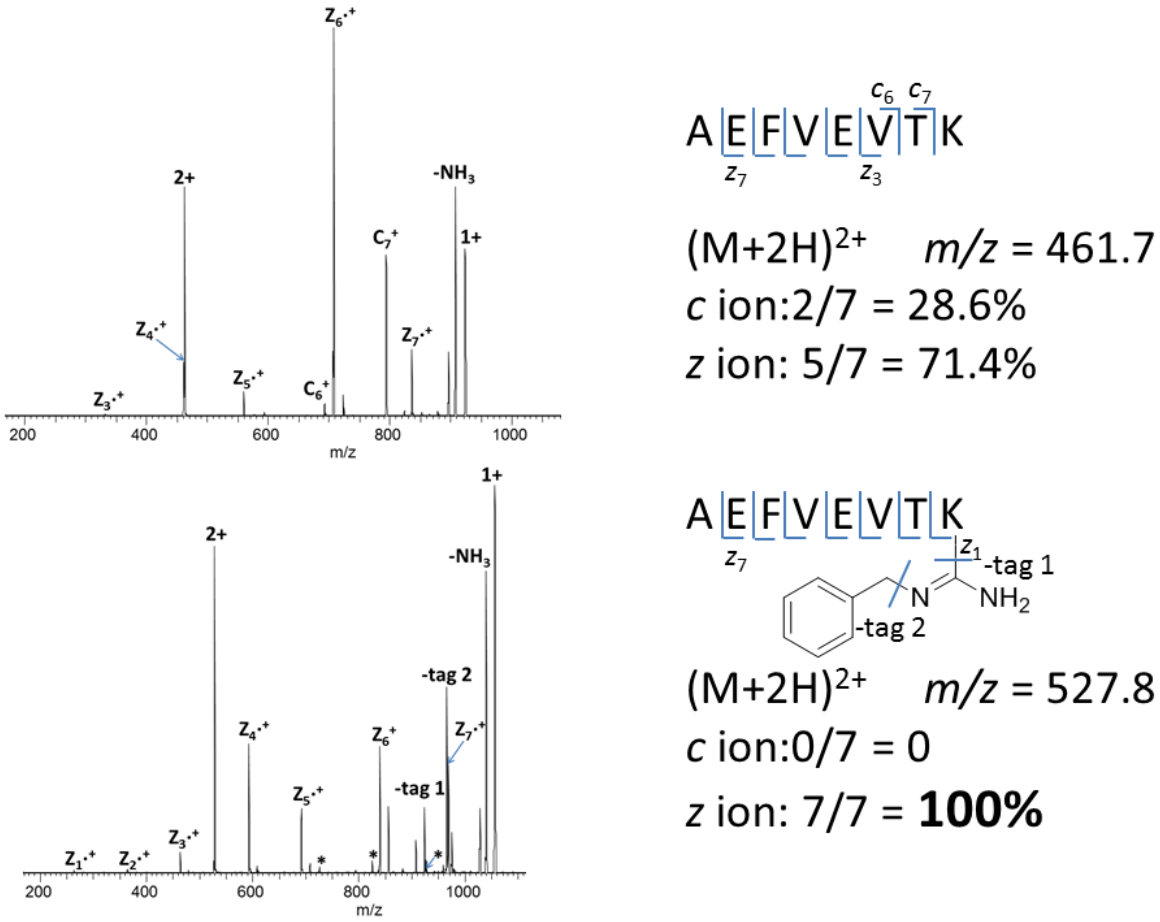


Figure 3.1 ETD spectra of doubly charged underivatized peptide ion AEFVEVTK (upper) and doubly charged benzylguanidine-tagged AEFVEVTK (lower). Peaks marked with an * are fragments from alanine tagged peptide ions.

The results from DLGEEHFK (**Figure 3.2**) provide another example that shows that the introduction of benzyl-guanidine tag improved the sequence coverage of modified peptide ions in ETD when compared with ETD of the underivatized peptide ions. The top figure is the MS/MS with ETD of triply charged underivatized DLGEEHFK and the bottom figure is the MS/MS with ETD of triply charged benzyl-guanidine tagged DLGEEHFK. Again, we observed a complete *z* ion series in the ETD spectra of benzylguanidine-tagged peptide ions, which enabled us to read the complete sequence of the peptide with full confidence. In contrast, when based on the ETD spectra of underivatized peptide ions, it was impossible to determine with full confidence whether the peptide was DLGEEHFK or LDGEEHFK, because the z_7 ion was missing. From this example, we can also see benzyl-guanidine tag doesn't

impede peptide backbone fragmentations by undesired side reactions, and does help the generation of small fragments like the z_1 ion.

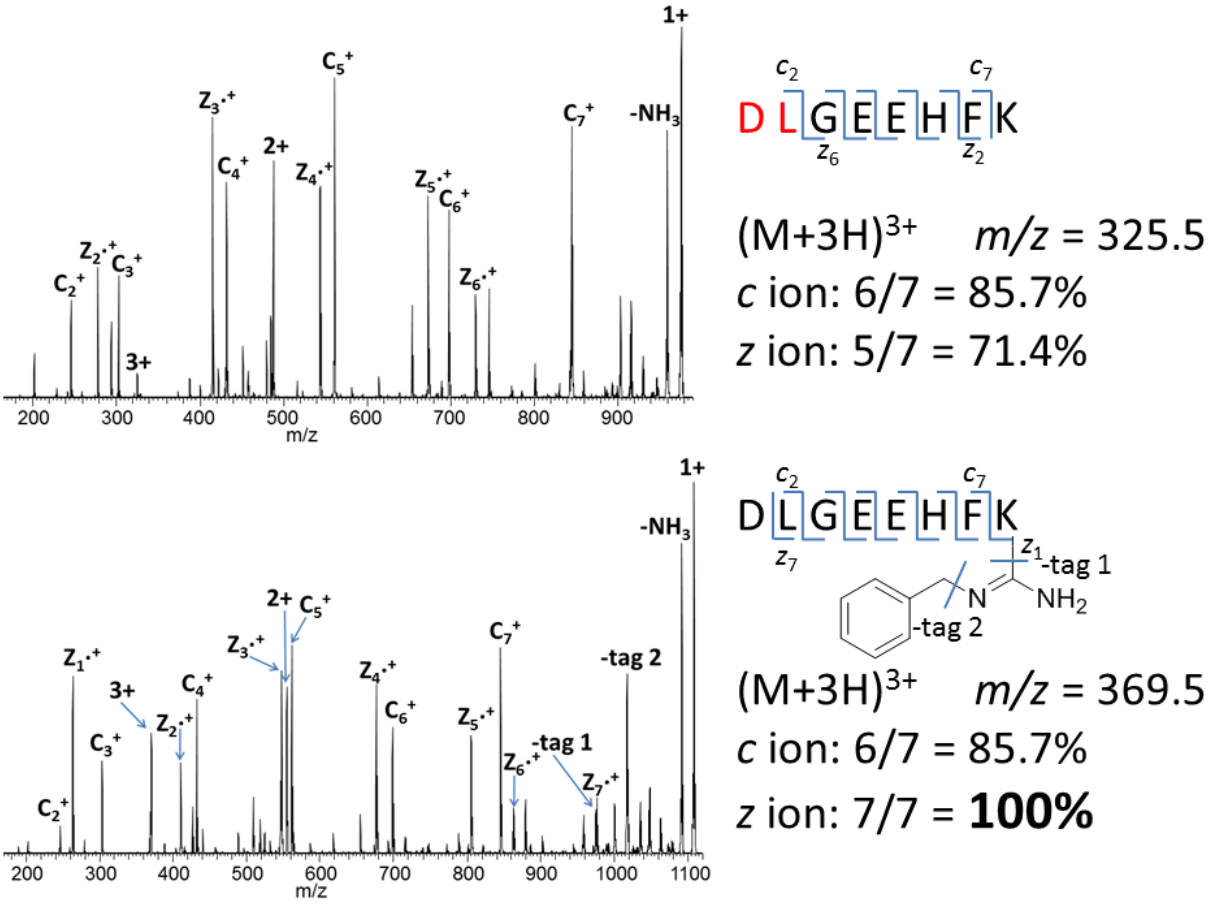


Figure 3.2 ETD spectra of triply charged underivatized peptide ion DLGEEHFK (upper) and triply charged benzylguanidine-tagged DLGEEHFK (lower).

The comparison between ETD of triply charged TCVADESHAGCEK and triply charged benzyl-guanidine tagged TCVADESHAGCEK are shown in **Figure 3.3**. The sequence coverage of the *z* ion fragments from the tagged peptide ions is 100% which is higher than the percentage of the untagged peptide ions, which is 83.3%. Notably for the z_3 and z_{12} fragments, a prominent *z*-90 fragment is observed instead of the full z fragment containing a carbamidomethylated cysteine (C) at its N terminus, due to the loss of •SCH_2CONH_2 (Sun, R.; Dong, M.; Song, C.; Chi, H.; Yang, B.; Xiu, L.; Tao, L..; Jing, Z.; Liu, C.;

Wang, L.; Fu, Y.; He, S. *J. Proteome. Res.* **2010**, *9*, 6354-6367). These peaks are marked with a red circle in the spectrum.

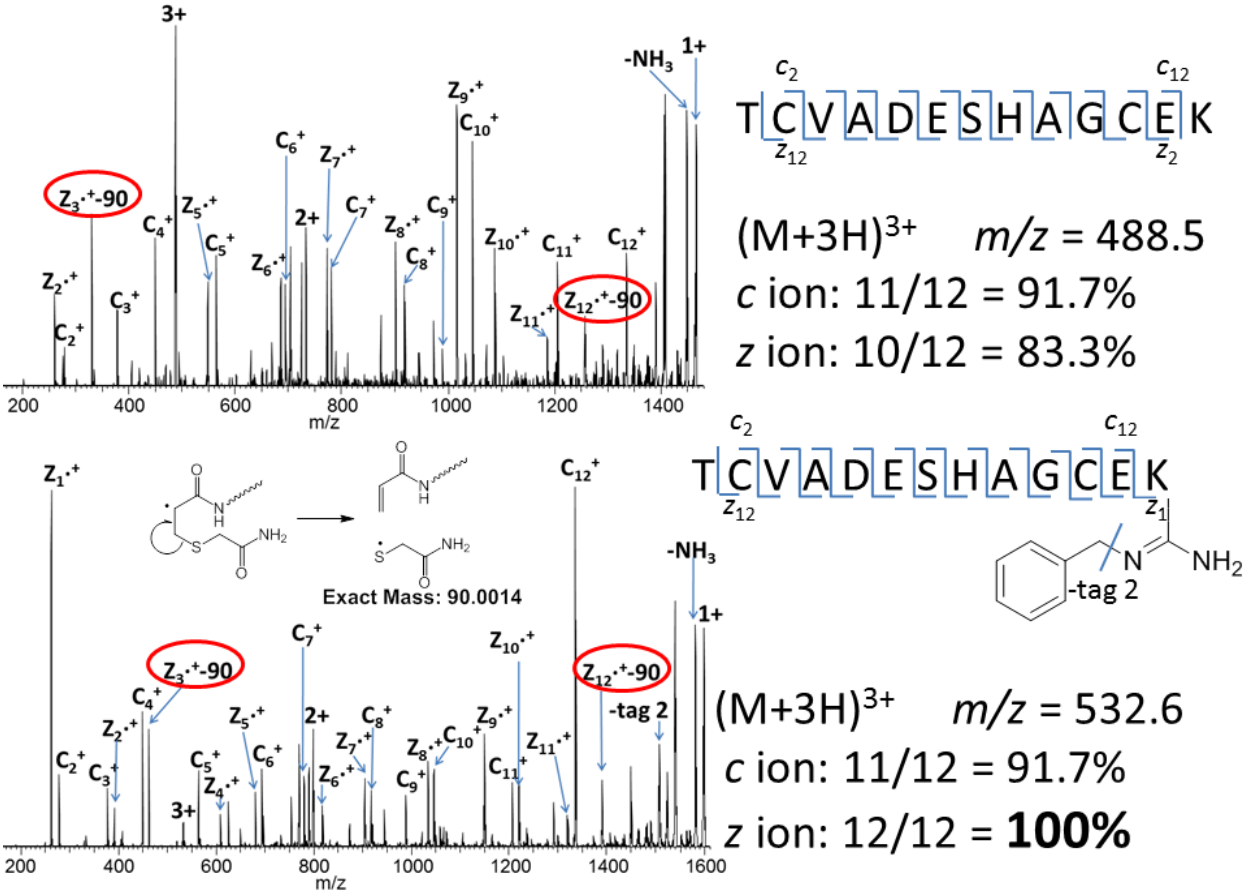


Figure 3.3 ETD spectra of triply charged underivatized peptide ion TCVADESHAGCEK (upper) and triply charged benzylguanidine-tagged TCVADESHAGCEK (lower). Red circles highlight the peaks corresponding to a prominent *z*-90 fragment is observed instead of the full *z* fragment.

The top spectrum in **Figure 3.4** is the MS/MS with ETD of triply charged underivatized 16 mer peptide ions RPCFSALTPDETYVPK and the bottom spectrum is the MS/MS with ETD of triply charged benzylguanidine-tagged 16 mer peptide ions RPCFSALTPDETYVPK. ETD does not cleave the backbone at the proline *N* terminus. Therefore, for the tagged peptide ions, there are 12 identifiable amino acids in addition to three prolines, and the sequence coverage of the *z* ions based on these remaining amino acids is 12/12 =100%.

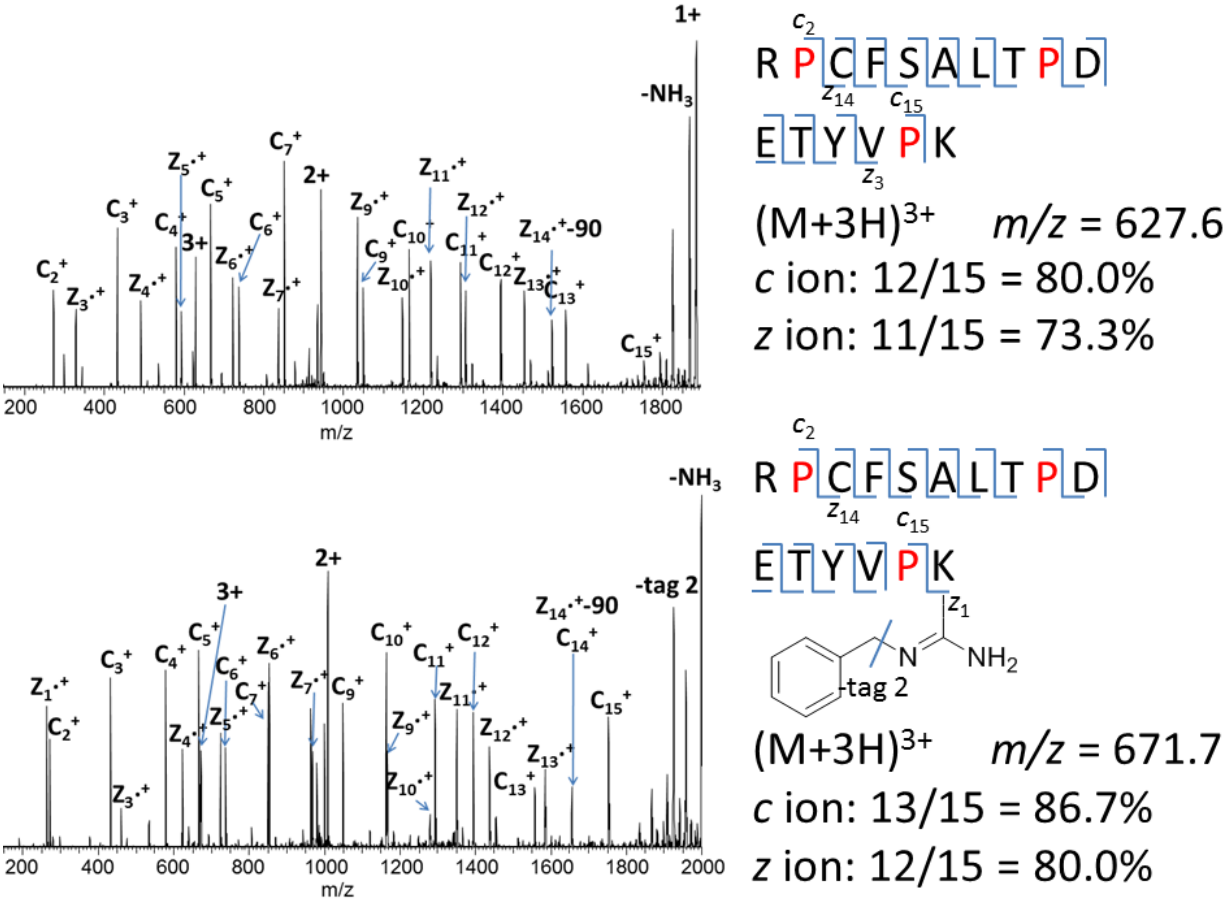


Figure 3.4 ETD spectra of triply charged underivatized peptide ion RPCFSALTPDETYVPK (upper) and triply charged benzylguanidine-tagged RPCFSALTPDETYVPK (lower).

There are 15 identified benzylguanidine-tagged BSA tryptic peptides from this study all of which are shown in **Table 3.1**. The K is in bold when there are multiple lysines in the sequence; the specific lysine which is labeled in the multiple K containing sequence is marked with an underline. The K terminated BSA tryptic peptides that resulted from a Mascot search included a total of 276 amino acids. Of the total 27 K terminated BSA tryptic peptides that could be experimentally expected as predicted by this research, 15 are benzyl-guanidine tagged. These 15 peptides contain 153 amino acids, therefore, the sequence coverage calculated in this way is 153/276 = 55%.

Peptide sequence	# of amino acids
LVTDLTK	7
AEFVEVTK	8
DLGEEHFK	8
SHCIAEVEK	9
SLHTLFGDELCK	12
TCVADESHAGCEK	13
DDPHACYSTVFDK	13
HPYFYAPELLYYANK	15
LFTFHADICTLPDTEK	16
RPCFSALTPDETYVPK	16
KFWG**K**	5
F**K**DLGEEHF**K**	10
L**K**ECCD**K**PLLE**K**	12
L**K**PDPNTLCDEF**K**	13
L**K**PDPNTLCDEF**K**ADE**K**	17

Table 3.1 List of identified benzylguanidine-tagged BSA tryptic peptides from LC-MS/MS in ETD. The K is in bold when there are multiple lysines in the sequence, and the specific lysine which is labeled in the multiple K containing sequence is marked with an underline

Ten identified benzylguanidine-tagged peptides only have one lysine in the sequence and are listed in **Table 3.2**. *Z* ions will be missing if the *N* terminus is proline. Considering this rule, it is evident that the sequence coverage of *z* ions is 100% for each of tagged peptides, and therefore the impact from the introduction of a benzyl-guanidine tag into the peptides is distinct.

Peptide sequence	# of amino acids	sequence coverage ($z\cdot$ ion)
LVTDLTK	7	6/6 = **100%**
AEFVEVTK	8	7/7 = **100%**
DLGEEHFK	8	7/7 = **100%**
SHCIAEVEK	9	8/8 = **100%**
SLHTLFGDELCK	12	11/11 = **100%**
TCVADESHAGCEK	13	12/12 = **100%**
DD**P**HACYSTVFDK	13	11/11 = **100%** (# of P: 1)
H**P**YFYA**P**ELLYYANK	15	12/12 = **100%** (# of P: 2)
LFTFHADICTL**P**DTEK	16	14/14 = **100%** (# of P: 1)
R**P**CFSALT**P**DETYV**P**K	16	12/12 = **100%** (# of P: 3)

Table 3.2 Sequence coverage of identified benzylguanidine-tagged BSA tryptic peptides from LC-MS/MS in ETD.

We have determined that the benzylguanidine tagging increases the sequence coverage of peptides with ETD. The next step was to investigate how the tagging impacts the sequence coverage of peptides with CID. Therefore a few peptides were selected and compared. As shown in **Table 3.3**, black means the sequence coverage keeps the same for the tagged and untagged peptide, red means the sequence coverage of the tagged peptide increases compared to the untagged peptides, and green means the sequence coverage of tagged peptide decreases compared to the untagged one. All 5 tagged peptides have higher or equal sequence coverage than unmodified peptides with ETD, and the introduction of benzylguanidine-tag does not have a negative impact for the fragmentation of peptide ions with CID.

peptide sequence	z	# of aa	sequence coverage (ETD)		sequence coverage (CID)	
			c ion	z ion	b ion	y ion
SLHTLFGDELCK	3	12	10/11=90.9%	10/11=90.9%	7/11=63.6%	9/11=81.8%
Benzyl-SLHTLFGDELCK	3	12	10/11=90.9%	11/11=100%	7/11=63.6%	7/11=63.6%
TCVADESHAGCEK	3	13	11/12=91.7%	10/12=83.3%	3/12 =25%	8/12=66.7%
Benzyl-TCVADESHAGCEK	3	13	11/12=91.7%	12/12=100%	6/12=50%	10/12=83.3 %
RPCFSALTPDETYVPK	3	16	12/15=80%	11/15=73.3%	9/15=60%	8/15=53.3%
Benzyl-RPCFSALTPDETYVPK	3	16	13/15=86.7%	12/15=80%	8/15=53.3%	6/15=40%
FKDLGEEHFK	3	10	8/9=88.9%	8/9=88.9%	7/9=77.8%	9/9=100%
Benzyl-FKDLGEEHFK@2&10	3	10	8/9=88.9%	9/9=100%	6/9=66.7%	8/9=88.9%
KQTALVELLK	3	10	9/9=100%	8/9=88.9%	9/9=100%	6/9=66.7%
KQTALVELLK@1	3	10	9/9=100%	8/9=88.9%	9/9=100%	8/9=88.9%

Table 3.3 Sequence coverage comparison between unmodified peptides and benzylguanidine-tag modified peptides with ETD or CID

For those singly modified peptides which have multiple lysines in the sequence, only the lysine which is closest to the *N*-terminus of the peptide is labeled. **Figure 3.5** is an example showing the ETD spectrum of triply charged benzyl-guanidine tagged KFWGK. There are two lysines in the sequence, one is at the *N*-terminus and the other is at the *C*-terminus, and only the lysine at the *N*-terminus is labeled. The number after the symbol "@" indicates the labeling position from left (*N*-terminus) to right (*C*-terminus) in the peptide sequence. Due to the isobaric nature of isomers, the mass of the labeled peptide ions are the same whether the labeling position is at either @1 or @5, however, the labeling position can subsequently be identified because of the difference of the mass of their fragment with ETD. In the case of KFWGK, the labeling position is the K at the *N*-terminus. A possible explanation for this observation is that the combination of the NH_2 group from the lysine and the NH_2 group from the *N*-terminus is more basic when K is closer to *N*-terminus, and therefore the K closer to *N*-terminus of the peptide is more easily to react with the functional group of $–SO_3H$ on the tag.

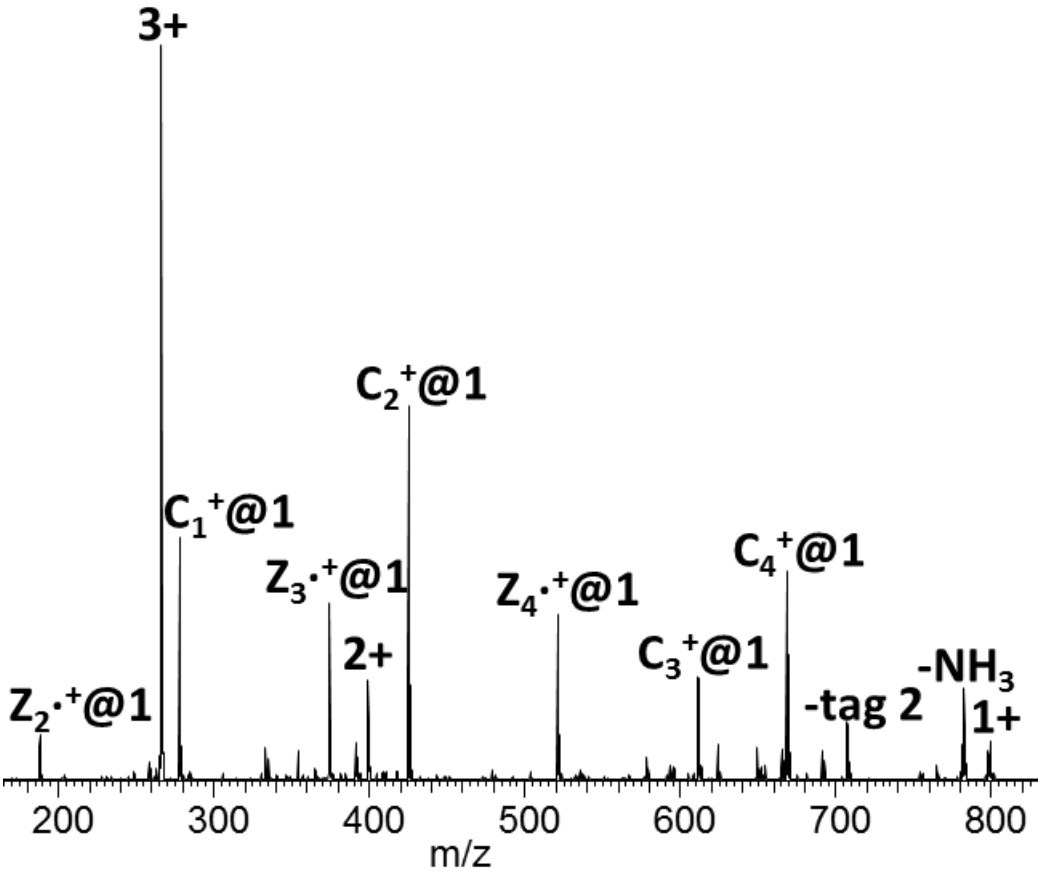


Figure 3.5 ETD spectra of triply charged benzylguanidine-tagged peptide ion KFWGK. The number after the symbol "@" describes the lysine position which was labeled, starting from the *N*-terminus.

Therefore, the sequence coverage for *z* ions with multiple lysines in the sequence of singly benzylguanidine-tagged peptide ions is different from those with a single lysine in the sequence. As shown in **Table 3.4**, all singly tagged peptides with multiple K in the sequence are missing the z_1 ion in the ETD spectra while still possessing all other *z* ions expected. Consistent with the explanation above, this is due to the increased basicity of functional groups close to the *N*-terminus due to the proximity of the introduced tag, therefore favoring the charge to be located on position close to the *N*-terminus rather than that of *C*-terminus.

Peptide sequence	# of amino acids	sequence coverage ($z\cdot$ ion)
KFWG**K**@1	5	$z_1\cdot$ is missing only
F**K**DLGEEHF**K**@2	10	$z_1\cdot$ is missing only
L**K**ECCD**K**PLLE**K**@2	12	$z_1\cdot$ is missing ($z_{10}\cdot$ and $z_{11}\cdot$ are outside of mass range)
L**K**PDPNTLCDEF**K**@2	13	$z_1\cdot$ is missing only
L**K**PDPNTLCDEF**K**ADE**K**@2	17	$z_1\cdot$ is missing ($z_{16}\cdot$ is outside of mass range)

Table 3.4 Sequence coverage of previously identified benzylguanidine-tagged BSA tryptic peptides which have multiple K in their sequence.

From the ETD spectra of tagged peptide ions, we can observe that the peaks representing the fragment generated from the dissociation of the full tag (-tag 1) or the partial tag (-tag 2) from the peptide ions (**Table 3.5**). For the single lysine containing peptide ions that are tagged, the peak corresponding to a loss of tag 1 disappear gradually with an increase of number of amino acids. All the peptide ions with less than 13 amino acids in the sequence generate a peak which loses the full tag, and all the peptide ions with more than 13 amino acids in the sequence do not have peaks representing the loss of the full tag with ETD. Meanwhile, tagged peptide ions with multiple lysines lose only part of the tag during dissociation.

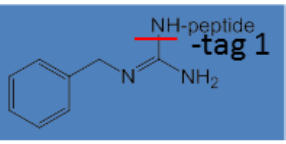
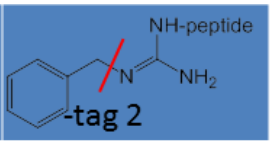
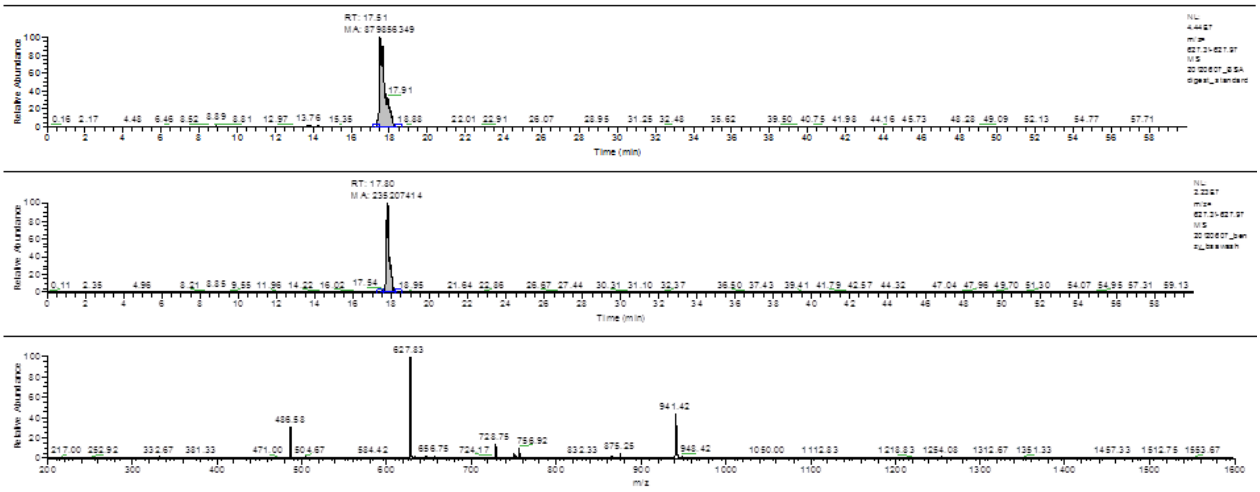
Peptide sequence	# of amino acids		
LVTDLTK	7	*	*
AEFVEVTK	8	*	*
DLGEEHFK	8	*	*
SHCIAEVEK	9	*	*
SLHTLFGDELCK	12	*	*
DDPHACYSTVFDK	13	*	*
TCVADESHAGCEK	13		*
HPYFYAPELLYYANK	15		*
LFTFHADICTLPDTEK	16		*
RPCFSALTPDETYVPK	16		*
KFWGK@1	5		*
FKDLGEEHFK@2	10		*
LKPDPNTLCDEFK@2	13		*

Table 3.5 Loss of tags from the benzylguanidine-tagged peptide ions in ETD. Observed fragmentation channels are marked with an asterisk (*).

In order to have an estimate for the reaction yield of the whole coupling process, parallel LC-MS/MS runs of standard BSA tryptic peptides and unreacted BSA tryptic peptides were performed on LTQ-Orbitrap to quantify specific peak area. Since 1% of the tryptic peptides were set aside as the peptide standards, 1/99 of the unreacted BSA tryptic peptides by volumn were collected as well. Following the proper dilution, the same volume of liquid from the two above vials were dried, and redissolved in H_2O/acidic acid (99/1) to be loaded on the column for LC-MS/MS. Following this parallel runs, the coupling yield for each identified lysine terminated ended tagged peptide was calculated. **Figure 3.6** shows an example to demonstrate how the yield calculation was performed. RPCFSALTPDETYVPK is a peptide from the list of identified tagged peptides. This underivatized peptide ion has multiple charge states in the spectra, and since the triply charged form is the most intense one, is therefore used for the purpose of quantitation. In this example, the peak area for triply charged RPCFSALTPDETYVPK peptide ion from the MS scan of the BSA tryptic standard is 879856349, and the peak area for triply charged

RPCFSALTPDETYVPK peptide ion from the MS scan of uncoupled BSA tryptic peptides is 235207414, thus the estimated reaction yield is calculated to be 1-235207414/879856349 = 73.3%.



RPCFSALTPDETYVPK	m/z	mass range	rt/min	peak area from the standard	peak area from the unreacted	yield
3+	627.64	627.31-627.97	17.8	879856349	235207414	73.3%

Figure 3.6 Reaction yield calculation of benzylguanidine-tagged BSA tryptic peptide ions with a sequence of RPCFSALTPDETYVPK.

Reaction yield was calculated using the method above for each identified benzyl-guanidine tagged peptide, and the result is shown in **Table 3.6**. Only the peptides which have a good quality of peaks are taken into consideration for the calculation of the reaction yield. Several peptides which have poor signal due to the ion suppression from other ions at the same elution time are not considered. The average of the reaction yields is calculated to be 75.0% from the 8 tagged peptides on the table. The ratio of the peak area (unreacted/standard) of R ended BSA tryptic peptides were also calculated. Theoretically, arginine terminated peptides are removed and will not retain on the beads during the wash. Peptides AWSVAR and DAFLGSFLYEYSR have the peak area ratio of 0.93 and 0.92,which are close to 1, meaning that they did not retain on the beads during the wash as we expected. However, peptides YLYEIAR and HPEYAVSVLLR are found to have the ratio of 0.23 and 0.47, meaning that they retained on the beads and were not fully recovered during the wash.

Peptides	charge state	yield
AEFVEVTK	2	67.6%
RPCFSALTPDETYVPK	3	73.3%
FKDLGEEHFK@2	3	71.7%
LKPDPNTLCDEFK@2	3	76.9%
DDPHACYSTVFDK	3	77.6%
HPYFYAPELLYYANK	3	78.0%
LKPDPNTLCDEFKADEK@2	4	78.4%
SLHTLFGDELCK	3	76.5%

Table 3.6 Reaction yield of benzylguanidine-tagged BSA tryptic peptides.

Derivatization of the benzyl-guanidine tag to the K ended peptides decreases the polarity of the peptide, giving them higher affinity for the column media, therefore the retention time of the benzyl-guanidine tagged peptides is longer than that of underivatized peptides. The delay of the retention time caused by the introduction of the tag ranges from 3.4 minutes to 10.9 minutes, with an average of 6.5 minutes. **Table 3.7** and **Table 3.8** show the retention time of the unmodified peptide and modified peptide identified in two LC-MS/MS runs done on different days. The peptides in each table are listed according to the retention time of their unmodified version in an ascending order. It is clear that, generally, the

retention time of the modified version are also in the ascending order with an exception of DLGEEHFK and DDPHACYSTVFDK, which are marked in bold.

peptide sequence	retention time of unmodified peptide (min)	retention time of benzyl-guanidine modified peptide (min)
SHCIAEVEK	8.9	16.2
DLGEEHFK	9.3	**19.0**
FKDLGEEHFK@2	9.6	18.4
AEFVEVTK	13.5	22.1
DDPHACYSTVFDK	16.1	**23.7**
LKPDPNTLCDEFK@2	18.0	22.1
RPCFSALTPDETYVPK	19.6	23.7
SLHTLFGDELCK	21.4	26.8

Table 3.7 Comparison of retention times between unmodified peptides and modified peptides in LC-MS/MS at an earlier day

peptide sequence	retention time of unmodified peptide (min)	retention time of benzyl-guanidine modified peptide (min)
LKECCDKPLLEK@2	8.8	13.6
KFWGK@1	9.0	16.8
LVTDLTK	9.0	19.9
LKPDPNTLCDEFKADEK@2	15.8	19.9
LFTFHADICTLPDTEK	21.8	25.2
HPYFYAPELLYYANK	25.0	29.2

Table 3.8 Comparison of retention times between unmodified peptides and modified peptides in LC-MS/MS at a later day.

3.3.2 Comparison of benzyl/methyl/butyl-guanidine tagged BSA tryptic peptides

The lists of identified tagged peptides modified by the methyl-guanidine tag and butyl-guanidine tag are clarified in **Table 3.9** and **Table 3.10**. The lysines are bold when there are multiple lysines in the sequence, meanwhile, the K which is modified in the multiple K containing peptide sequence is marked with underline. Additionally, there are two italicized peptides here: LCVLHEK and TPVSEK. Methyl-guanidine tagged LCVLHEK, butyl-guanidine tagged TPVSEK, and butyl-guanidine tagged LCVLHEK are identified. Both of the two peptides are not identified in the mascot search of BSA tryptic digest standard which either means the molecular ion is not selected for tandem MS or means the ETD fragmentation of these two underivatized peptides is not efficient enough to be identified. If the later is the case, it is clear the introduction of guanidine tag to the peptides can help identify more peptides. The identified butyl-guanidine tagged peptide KVPQVSTPTLVEVSR which terminates with R has a K in the sequence. K is tagged then and the tagged peptide is identified in Mascot search.

Peptide sequence modified by methyl-guanidine tag	# of amino acids
DLGEEHF<u>K</u>	8
NECFLSH<u>K</u>	8
HLVDEPQNLI<u>K</u>	11
YICDNQDTISS<u>K</u>	12
RPCFSALTPDETYVP<u>K</u>	16
LCVLHE<u>K</u>	7
CCAADD**<u>K</u>**EACFAVEGP**K**	17
L**<u>K</u>**PDPNTLCDEF**K**ADE**K**	17
F**<u>K</u>**DLGEEHF**K**	10
F**<u>K</u>**DLGEEHF**<u>K</u>**	10

Table 3.9 List of identified methylguanidine-tagged BSA tryptic peptide.

Peptide sequence modified by butyl-guanidine tag	# of amino acids
LVTDLT**K**	7
ATEEQL**K**	7
AEFVEVT**K**	8
DLGEEHF**K**	8
NECFLSH**K**	8
SHCIAEVE**K**	9
YICDNQDTISS**K**	12
DDPHACYSTVFD**K**	13
RPCFSALTPDETYVP**K**	16
*TPVSE**K***	6
*LCVLHE**K***	7
KVPQVSTPTLVEVSR	15
L**K**ECCD**K**PLLE**K**	12
L**K**PDPNTLCDEF**K**	13
F**K**DLGEEHF**K**	10
F**K**DLGEEHF**K**	10

Table 3.10 List of identified butylguanidine-tagged BSA tryptic peptide.

Figure 3.7 shows all the identified guanidine tagged BSA tryptic peptides and how they overlap with each other. These results show that the performance of each tag is different as each of them identify their own unique set of peptides. Five unique benzyl-guanidine tagged peptides, four unique butyl-guanidine tagged peptides and three unique methyl-guanidine tagged peptides are identified in the Mascot search. Meanwhile, all the three guanidine tags have common hits. Three peptides shared in the middle are observable with all three tags. The overlap is greatest between benzyl-guanidine tag and butyl-guanidine tag which share six identifiable peptides, then between methyl-guanidine tag and butyl-guanidine tag which share three, and then between benzyl-guanidine tag and methyl-guanidine tag where the coincidence is least and there is only a single shared peptide. This might due to steric effect, as the benzyl group has a greater similarity with butyl group than it does with a methyl group. Overall, a combination of 23 modified peptides can be identified by the Mascot search and manual inspection, with

an overall coverage of 236 amino acids. Mature BSA contains 583 amino acids. When all peptides identified by the Mascot search are mapped into the full sequence of BSA, the total sequence coverage is 40.5% then. There are 26 peptides in the BSA tryptic standard that are identified by the Mascot search which encompasses 276 amino acids. Nineteen peptides of these twenty-six peptides are identified following our tagging procedure, which contain 193 amino acid. Therefore, 70% of peptides in BSA tryptic peptide standard are modified by guanidine tags and are subsequently identified.

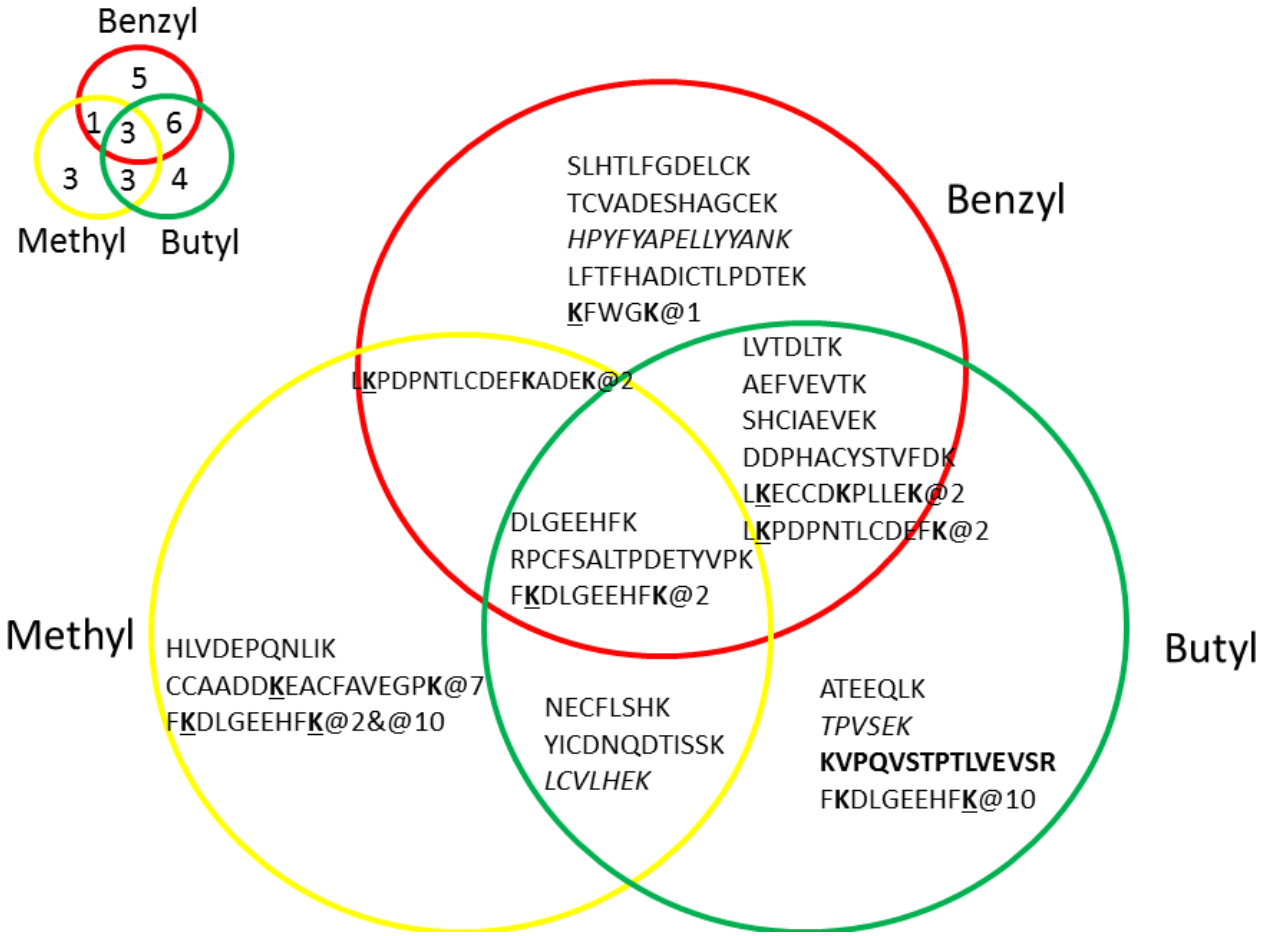


Figure 3.7 Comparison among identified benzyl/methyl/butylguanidine-tagged BSA tryptic peptides.

Comparison of sequence coverage for z ions of three types of guanidine tagged peptide ions with ETD are also studies and shown in **Table 3.11**. This table only lists the sequence of the peptides which are identified to be modified by at least two different guanidine tags. Several modified peptides are

identified with multiple charge states, and only the sequence coverage of the peptide ions with the same charge state are used to compare the performance of the tags. All observed benzyl-guanidine tagged peptides have 100% sequence coverage. Three out of the four methyl-guanidine tagged peptides are identified with 100% sequence coverage, while eight out of the nine identified butyl-guanidine tagged peptides have 100% sequence coverage. Again, the performance of the benzyl-guanidine tag is better than that of either the methyl-guanidine or the butyl-guanidine tag from the aspect of sequence coverage of the modified peptide ions in ETD.

Peptide sequence	# of amino acids	sequence coverage ($z\cdot$ ion)		
		Benzyl	Methyl	Butyl
RPCFSALTPDETYVPK	16	12/12 = **100%** (# of P: 3) (3+)	12/12 = **100%** (3+)	12/12 = **100%** (3+)
DLGEEHFK	8	7/7 = **100%** (3+)	/	7/7 = **100%** (3+)
LVTDLTK	7	6/6 = **100%** (2+)	/	6/6 = **100%** (2+)
AEFVEVTK	8	7/7 = **100%** (2+)	/	7/7 = **100%** (2+)
SHCIAEVEK	9	8/8 = **100%** (3+)	/	8/8 = **100%** (3+)
DDPHACYSTVFDK	13	11/11 = **100%** (# of P: 1) (3+)	/	11/11 = **100%** (3+)
LCVLHEK	7	/	6/6 = **100%** (3+)	6/6 = **100%** (3+)
NECFLSHK	8	/	7/7 = **100%** (3+)	7/7 = **100%** (3+)
YICDNQDTISSK	12	/	7/11 = 64% (2+)	9/11 = 82% (2+)
# of 100%		6/6	3/4	8/9

Table 3.11 Sequence coverage of identified guanidine tagged BSA tryptic peptides in ETD.

In the case of observed benzyl-guanidine tagged peptides, as previously mentioned, if there are multiple lysine sites in the sequence, only the lysine closest to the *N*-terminus is labeled. When it turns to the case of methylguanidine tag, identified CCAADD**K**EACFAVEGP**K** also agrees with this, as it has two lysine in the sequence, and only the K at site 7 (from left to right) is tagged. However, this trend is not without its exceptions, such as the case for the peptide FKDLGEEHFK. As **Figure 13** shows, F**K**DLGEEHF**K** is observed with all three tags, which with derivatization of only the first lysine in the

sequence *generally following* the same observation rule. However, F**K**DLGEEHF**K** is additionally found in the case of the butyl-guanidine tag, and double guanidination F**K**DLGEEHF**K** is found when the methyl-guanidine tag is used. This phenomenon might due to steric effect as well, as the methyl-guanidine tag is the smallest tag, thus the two methyl-guanidine tags within the same bead would have the ability to simultaneously tag two different sites in the same peptide. **Figure 3.8** shows the ETD spectra of quadruply charged double methyl-guanidine tagged F**K**DLGEEHF**K**, and its complete *z* ion series from the z_1 ion to the z_8 ion with the exception of z_9 ion, which is unknown as it is out of mass range. **Figure 3.9** shows the ETD spectra of a mixture of triply charged butyl-guanidine tagged F**K**DLGEEHF**K** and F**K**DLGEEHF**K**. The ions in the spectra marked with black color are the unique fragments generated from F**K**DLGEEHF**K**, and the ions marked with red color are the unique fragments generated from F**K**DLGEEHF**K**. The ions marked with blue color are generated from either F**K**DLGEEHF**K** or F**K**DLGEEHF**K**, since the mass of the fragments are the same when tagged either site. The spectra demonstrate the existence of a mixture of F**K**DLGEEHF**K** and F**K**DLGEEHF**K**, and the majority is still F**K**DLGEEHF**K**, where the labeling site which is closer to the *N*—terminus.

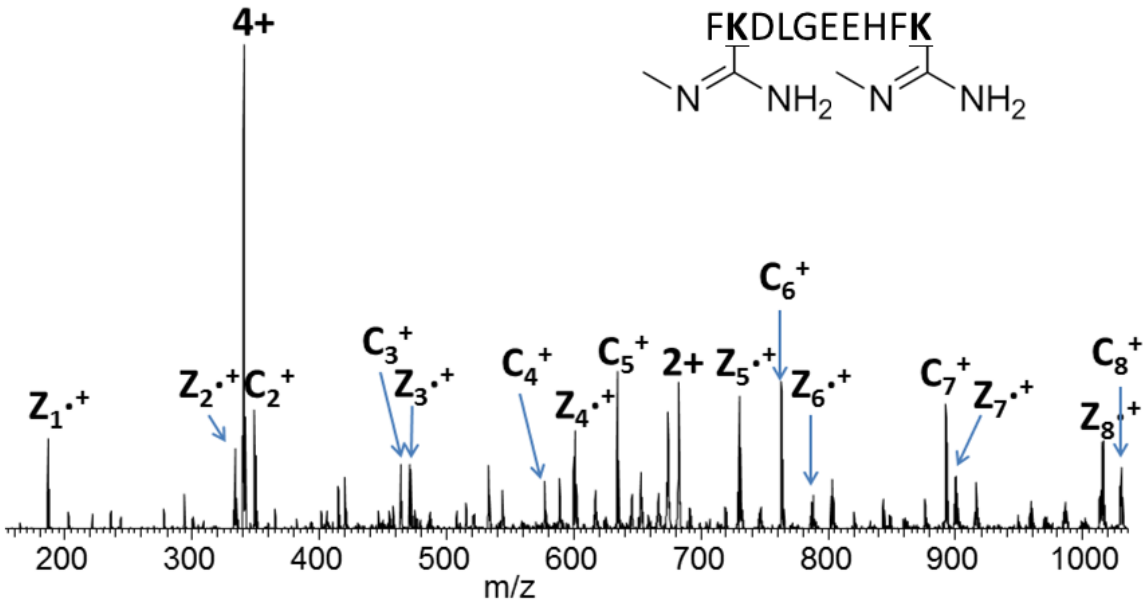


Figure 3.8 ETD spectra of quadruply charged methylguanidine doubly tagged F**K**DLGEEHF**K**

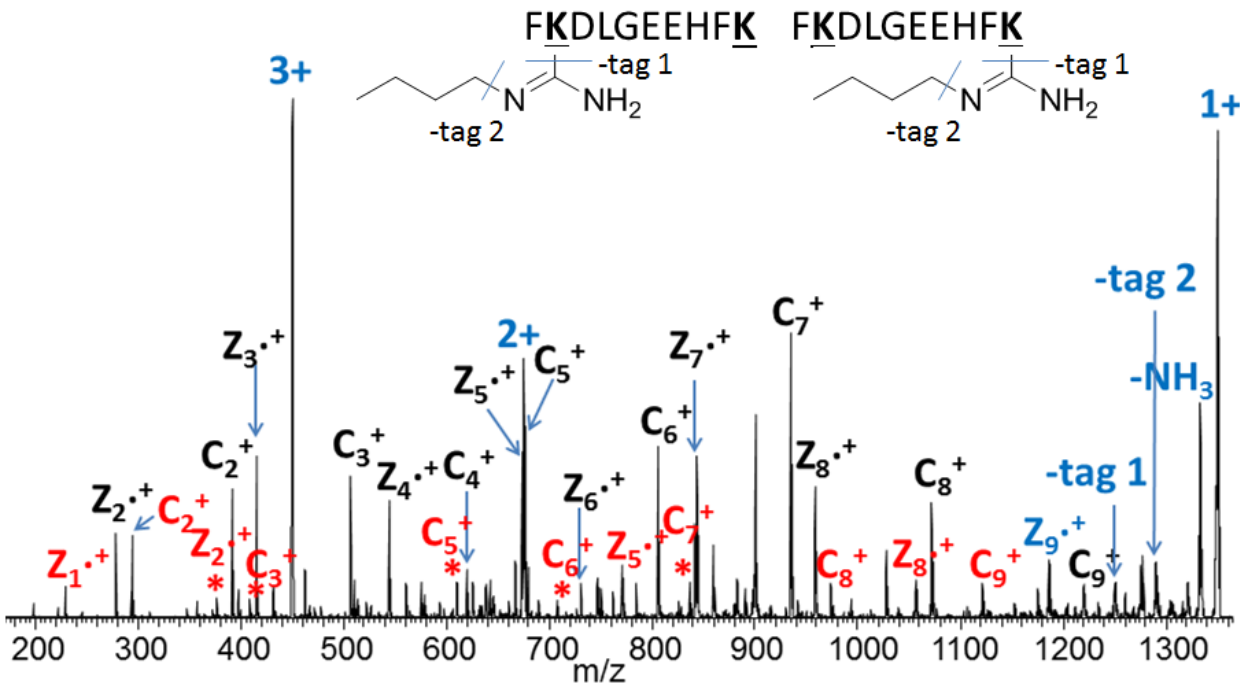


Figure 3.9 ETD spectra of a mixture of triply charged butylguanidine-tagged F**K**DLGEEHF**K** (fragments in black and blue) and triply charged butylguanidine-tagged F**K**DLGEEHF**K** (fragments in red and blue)

3.4 Conclusion

In this study, application of benzylguanidine tag enables the remarkable increase in sequence coverage of BSA tryptic peptides, and further results in much easier interpretation of mass spectra and peptide sequencing. Methylguanidine tag and butylguanidine tag have been applied to label BSA tryptic peptides as well, and all three tags mentioned above have their unique characteristic since they are able to tag tryptic peptides with different sequences. However, the benzylguanidine tag still has a better performance over both of the other two tags.

Chapter 4
Quantitative Study with the Pair of d0/d7-Benzylguanidine Tags

4.1 Introduction

The introduction of benzylguanidine tag has been demonstrated to simplify the mass spectra and remarkably improve sequence coverage of peptide ions by ETD. There is interest to know if the tag can be applied for quantitative study as well. In addition to the benzylguanidine tag, which is called d_0-benzylguanidine tag in this chapter for the comparison purpose, we designed and synthesized its isotopic analog d_7-benzylguanidine tag which is chemically identical but differs in the isotopic composition. The pair of tags were designed to have a reasonable mass difference, high enough to have resolved doublet peaks from tagged peptide pairs in a low-resolution mass spectrometer, and low enough to have the doublet peaks elute at the same time during a column separation. Additionally the pair of tags were designed to have the same labeling characteristic with peptides, and also show very similar fragmentation patterns for the tagged peptide pairs. Seven hydrogens (H) were replaced with seven deuterons (D) in the tag, on the benzene ring and the benzylic H, to maximize the mass difference of the two tags while minimizing the elution time difference of the tagged peptide pairs.

4.2 Experimental Section

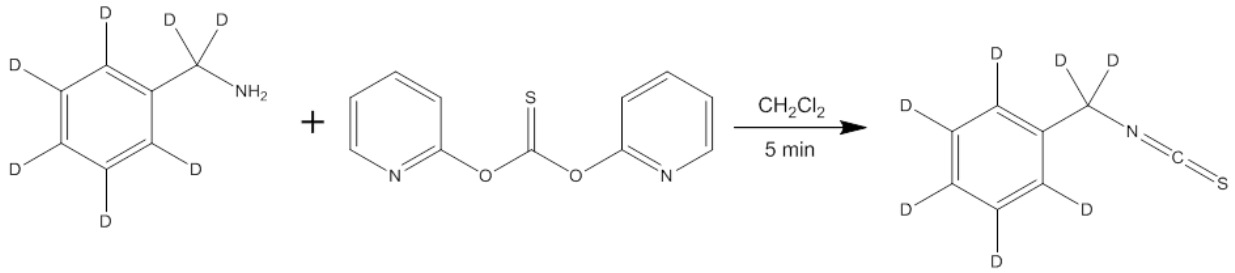
4.2.1 Materials

Bovine serum albumin (BSA) and β-Casein were purchased from Sigma-Aldrich (Milwaukee, WI, USA). Disposable reaction vessels (Empty syringe equipped with frit) were purchased from CSPS Pharmaceuticals (San Diego, CA, USA). PAL-ChemMatrix Resin was purchased from Biotage (Charlotte, NC, USA). Benzyl-d_7-amine was purchased from C/D/N Isotopes Inc. (Pointe-Claire, QC, Canada). Trypsin Gold (Mass Spectrometry Grade) and rLys-C (Mass Spectrometry Grade) were purchased from Promega Corporation (Fitchburg, WI, USA). Column packing material Magic 5 μ 200 Å C18 AQ and column packing material Magic 5 μ 100 Å C18 AQ were purchased from Bruker-Michrom (Auburn, CA, USA). Self-Pack IntegraFrit Columns (Part Number IF360-100-50-N-5) was purchased from NewObjective Inc. (Woburn, MA, USA). Capillary tubing (Product Descriptor TSP075375-10M) was

purchased from Polymicro Technologies (Phoenix, AZ, USA). Sep-Pak C18 solid phase extraction cartridges (1 cc Vac Cartridge, 50 mg Sorbent per Cartridge, 55-105 µm Particle Size, Product Number: WAT054955) were purchased from Waters (Milford, MA, USA). Common chemicals (reagent grade) were purchased from Sigma-Aldrich (Milwaukee, WI, USA).

4.2.2 Preparation method and analysis of d_0- and d_7-benzylguanidine-tagged BSA/β-Casein tryptic peptides

4.2.2.1 Preparation of d_0- and d_7-benzylguanidine tags precursors on the beads inside the syringe



Scheme 4.1 Synthesis route of d_7-benzyl-isothiocyanate

D_7-benzyl-isothiocyanate was prepared by my coworker Aleš Marek according to the method of Kim, S etd [1] as shown in **Scheme 4.1**. D_7-benzylamine was mixed with di-2-pyridyl thionocarbonate at mole ratio of 1:1.02 in dichloromethane and left to react for 10 min. Reaction progress was monitored by thin layer chromatography (TLC) plate with ethyl acetate: hexane =1:2 as developing solvent. The Rf value of d_7-benzyl-isothiocyanate on TLC plate is 0.83. The compound was purified by column chromatography (SiO_2, Ethyl acetate: hexane = 1:2) giving a yield of 99%. The resulting white powder d_7-benzyl-isothiocyanate was stored in the -20 °C freezer. D_7-benzylguanidine tag precursor and d_0-benzylguanidine tag precursor were prepared in parallel in two separate syringes starting with 25 mg PAL-ChemMatrix Resin according to the preparation method for the d_0-benzylguanidine tag described before in **Chapter 2**.

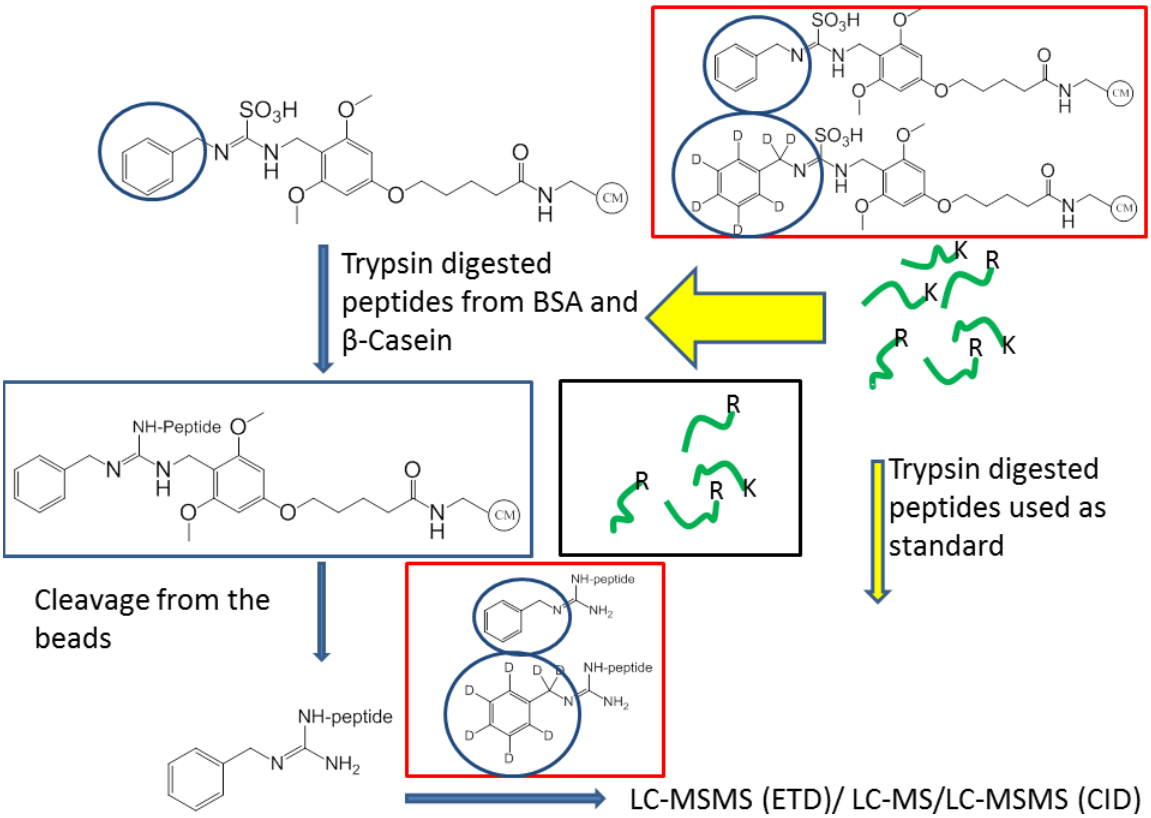
4.2.2.2 Preparation of BSA/β-Casein tryptic peptides

Ammonium bicarbonate (0.1 M) in water was used as the solvent to prepare four separate stock solutions of 10 mg/mL BSA, 3.57 mg/mL β-Casein, 0.5 M tris(2-carboxyethyl) phosphine (TCEP) and 1 M iodoacetamide (IAA). Acetic acid (50 mM) was used to dissolve trypsin according to the Trypsin Gold Protocol provided from Promega. Eight eppendorf tubes of BSA and β-Casein tryptic peptides solution were prepared according to the following method: In each tube, 91.3 µL of 10 mg/mL BSA, 91.3 µL of 3.57 mg/mL β-Casein (mole ratio of BSA to β-Casein is 1 to 1), 730 µL ammonium bicarbonate (0.1 M), and 9.1 µL of 0.5M TCEP (final concentration: 5 mM) were added into the tube and were incubated at 37 °C for 30 minutes to break disulfide bonds within and between proteins. Then, 9.1 µL of 1 M IAA (final concentration: 10 mM) was added to each tube and agitated at RT for 10 minutes to encourage IAA capping of cysteine residues in the protein. After cysteine capping, 12.4 µL of 1 µg/µL trypsin (final trypsin: protein ratio of 1:100) was added to each tube and all eight tubes were incubated at 37 °C overnight. Following this digestion, each tube was acidified to pH = 3 (verified with pH paper) by adding formic acid (~ 7 µL). Digested peptides from every two tubes were combined to be desalted with a single Sep-Pak C18 cartridge. Four Sep-Pak C18 cartridges were used according to the same protocol mentioned earlier in Chapter 3. Eluates (8 mL) collected from four cartridges were combined, dried with a stream of nitrogen and redissolved in 1.6 mL of a MeOH/H_2O (50/50) solvent mixture. 16 µL of the peptide solution were taken out and saved as a standard for future LC-MS/MS analysis of BSA/*β*-Casein tryptic peptides. The remaining solution was separated into two equal portions.

4.2.2.3 Peptide modification by d_0- and d_7-benzylguanidine tag

Triethylamine (6.7 µL, 48 µmoles) was introduced into each portion, and additional triethylamine may be added as needed to attain pH higher than 10. Each mixture was then sucked into a syringe, and both of the syringes were stirred at RT for 64 h. Following this, the unreacted peptides were collected for future analysis. Next the beads in the syringe were washed with a MeOH/H_2O (50/50) solvent mixture for five times (five minutes stirring each time) and then washed with diethyl ether for another three times (five minutes stirring each time). The syringe was then dried under a mechanical vacumn pump for 2.5 h. A cleavage solution consisting 0.75 mL of TFA/water/triethylsilane (95/2.5/2.5) was introduced into each

syringe to cleave the modified peptides from the beads. After rotated stirring for 1 h, the solution in the syringe was collected in a glass vial. The beads in each syringe were then washed with TFA (0.5 mL × 2) another two times (five minutes rotated stirring each time). The washes were combined into the previous glass vial, dried over nitrogen and redissolved in a 0.25 mL solution of methanol/water (20/80). 16 µL of the modified peptides dissolved in methanol were dried and then redissolved in 37.5 µL (estimated to give roughly 100 µM for modified digested peptides) of H_2O/acetic acid (99/1) for further data dependent LC-MS/MS analysis by ETD or CID. The BSA/β-Casein tryptic peptides standard collected before was dried then redissolved in 37.5 µL (estimated to give 100 µM for digested peptides) of H_2O/acetic acid (99/1). The work flow is shown in **Scheme 4.2**.



Scheme 4.2 Coupling of BSA/ *β*-Casein tryptic peptides with d_0/d_7-benzylguanidine tag precursor

4.2.2.4 Performance evaluation of d_0 and d_7-benzylguanidine-tagged BSA/β-Casein tryptic peptides

A Thermo LTQ XL™ mass spectrometer with ETD (Thermo Scientific, Waltham, MA, USA) and a Thermo LTQ-Orbitrap XL mass spectrometer (Thermo Scientific, Waltham, MA, USA) are each coupled with a Waters nanoACQUITY UPLC system to perform LC-MS/MS sample analysis. An analytical column and a trap column were prepared in house with the packing materials mentioned above in section *4.2.1*. A volume of 10 µL from the BSA/β-Casein tryptic peptides standard prepared above (37.5 µL) was diluted at the ratio of 1 to 5, then 1.0 µL (estimated 20 µM for digested peptide) from the diluted sample vial was injected to perform LC-MS/MS on both the instrument of LTQ-ETD and LTQ-Orbitrap. The 9.4 µL of the d_0-benzyl-guanidine modified BSA/β-Casein tryptic peptides (0.25 mL) and 9.4 µL of the d_7-benzyl-guanidine modified BSA/β-Casein tryptic peptides (0.25 mL) prepared before were dried and redissolved in 100 µL solvent of H_2O/acetic acid (99/1) (estimated to give roughly 20 µM for modified digested peptides) separately. Then 30 µL from each vial were combined and mixed into another vial (estimated to give roughly 10 µM for modified digested peptides). For the single tag modified sample, 4.5 µL was injected to run three runs, including LC-MS/MS with ETD, LC-MS on LTQ and LC-MS/MS with CID on LTQ-Orbitrap. For the mixture (60 µL) of two different tags-modified peptides mixed at the ratio of 1:1 as mentioned above, 9 µL was injected to run exactly the same set of three runs as for the single tag modified peptides. The parameters set for mass spec and the LC gradient are the same as what were mentioned in **Chapter 3**.

The collected raw data were converted to mgf file using MassSpecUtils (available at UWPR http://proteomicsresource.washington.edu/protocols06/MassSpecUtils.php), searched with Mascot (Version 2.3.0) (www.matrixscience.com) and confirmed with manual inspection and the use of Importer, an application software which is developed in house. The search database includes the sequences of BSA, yeast proteins, and β-Casein. Search parameters for LC-MS/MS runs done in LTQ-ETD were included as follows: type of search MS/MS ion search, enzyme trypsin/P, fixed modifications carbamidomethyl (C), mass values monoisotopic, peptide mass tolerance ±1.5 Da, fragment mass tolerance ±1 Da, maximum missed cleavages 2, instrument type-ETD-Trap, variable modification Phospho (ST). Additional variable modification of Benzyl-cx (K) and Benzyl-d7-cx (K) can be chosen depend on the data of the run loaded. The setting of the search parameters for LC-MS/MS runs done in

LTQ-Orbitrap is very similar as above with the exception of peptide mass tolerance ±0.4 Da, instrument type ESI-Trap.

4.2.3 Preparation of 1) d_0-benzylguanidine-, 2) d_7-benzylguanidine- and 3) the mixture of d_0/d_7-benzylguanidine- tagged BSA/β-Casein rLysC digested peptides

4.2.3.1 Preparation of 1) d_0-benzylguanidine, 2) d_7-benzylguanidine and 3) the mixture of d_0/d_7-benzylguanidine tags precursors on the beads inside the syringe

D_0-benzylguanidine tag precursor, d_7-benzylguanidine tag precursor and the mixture of d_0/d_7-benzylguanidne tag precursor with the ratio of 1 to 1 were prepared in parallel in three syringes according to the same method described previously in section *4.2.2* all starting with 25 mg PAL-ChemMatrix Resin.

4.2.3.2 Preparation of BSA/β-Casein rLysC digested peptides

A buffer (pH = 8.5) containing 25 mM Tris-HCl (by mixing Trizma and Trizma HCl), 1 mM EDTA and 6 M urea was used to prepare two separate stock solutions of 10 mg/mL BSA and 3.57 mg/mL *β*-Casein. DI water was used as the solvent to prepare stock solutions of 0.5 M tris(2-carboxyethyl) phosphine (TCEP) and 1 M iodoacetamide (IAA). Resuspension buffer came with rLysC from Promega was used to dissolve rLysC according to the Protocol provided from the company. Four Eppendorf tubes of BSA and *β*-Casein rLysC digested peptides were prepared according to the following method: In each tube, 55.3 µL of 10 mg/mL BSA and 55.3 µL of 3.57 mg/mL *β*-Casein (mole ratio of BSA to *β*-Casein is 1 to1) were mixed with 90 µL buffer (pH = 8.5) which contains 25 mM Tris-HCl, 1 mM EDTA and 6M urea. Then 2 µL of 0.5M TCEP (final concentration: 5 mM) was added into the tube and all tubes were incubated at 37 °C for 30 minutes. Then, 2 µL of 1 M IAA (final concentration: 10 mM) was added to each tube and all four tubes were agitated at RT for 10. After that, 1.1 mL of buffer which contains 25 mM Tris-HCl, 1 mM EDTA (pH = 8.5) were added into each tube to decrease the concentration of urea to be lower than 1 M. Then, 75 µL of 0.2 µg/µL r-LysC (final ratio of r-LysC: protein by weight is 1:50) was added to each tube and all the four tubes were incubated at 37 °C overnight (16 hours). After digestion, each tube was acidified to pH = 3 (verified with pH paper) by adding formic acid (~ 5 µL). Digested peptides in every two tubes were combined to be desalted with a single Sep-Pak C18 cartridge. Two Sep-Pak C18

cartridges were used according to the same protocol mentioned earlier in Chapter 3. A volume of 4 mL of elutes from four cartridges were combined and were dried over nitrogen, and were redissolved with 3 mL solution of MeOH/H_2O (50/50). An aliquot (30 μL) of the solution were taken out and saved as the standard for LC-MS/MS analysis of BSA/*β*-Casein tryptic peptides, and the remaining solution were separated into three equal portions.

4.2.3.3 Peptide modification by 1) d_0-benzylguanidine, 2) d_7-benzylguanidine and 3) the mixture of d_0/d_7-benzylguanidine tags

After triethylamine was added into each portion and the mixture was sucked into each syringe, samples inside all three syringes went through peptide coupling reaction, washes, drying and acid cleavage following the procedure mentioned earlier in *section 4.2.2.3*. After rotated stirring for 1 h, the solution in each syringe was collected in a glass vial. The beads in each syringe were washed with TFA (0.8 mL × 2) another two times (five minutes stirring each time). The washes were combined into the previous glass vial, dried under nitrogen and redissolved in 0.25 mL MeOH/H_2O (50/50). A volume of 47 μL of the dissolved modified peptides were dried and then redissolved in 100 μL (estimated roughly 20 μM for modified digested peptides) of H_2O/acetic acid (99/1) for further data dependent LC-MS/MS analysis in CID (each injection volume is 4.5 μL). 10 μL of the BSA/*β*-Casein tryptic peptides standard collected before was dried and then redissolved in 37.5 μL (estimated 20 μM for digested peptides) of H_2O/ACN/acetic acid (94/5/1) for data dependent LC-MS/MS analysis with CID (each injection volume is 1.0 μL).

4.2.3.4 Performance evaluation of d_0, d_7 and the mixture of d_0/d_7-benzylguanidine-tagged BSA/β-Casein r-LysC digested peptides

Mass Spectrometer Thermo LTQ-Orbitrap XL (Thermo Scientific, Waltham, MA, USA) is coupled with Waters nanoACQUITY UPLC system to perform LC-MS/MS sample analysis. An analytical column and trap column were prepared in house.

Five consecutive LC-MS/MS runs were performed from five sample vials which contains 1). Underivatized peptides standards; 2). d_0-benzylguanidine-tagged peptides; 3). d_7-benzylguandine-tagged

peptides; 4). mixture of d_0- and d_7-benzylguandine-tagged peptides with the ratio of 1:1; 5). Tagged peptides which were derivatized by the mixture of d_0- and d_7-benzylguanidine tag precursor at 1:1 ratio at the very beginning. The parameters for the setting of mass spec and the LC gradient are the same as what were mentioned in **Chapter 3**.

The collected raw data were processed by MassSpecUtils and Mascot in the same way as mentioned in *4.2.2.4*, and confirmed with manual inspection and the use of Importer. Search parameters in Mascot for LC-MS/MS runs done in LTQ-Orbitrap are as follows: type of search MS/MS ion search, enzyme Lys-C, fixed modifications carbamidomethyl (C), mass values monoisotopic, peptide mass tolerance ±0.4 Da, fragment mass tolerance ±1Da, maximum missed cleavages 2, instrument type ESI-Trap, variable modification Phospho (ST). Additional variable modification of Benzyl-cx (K) and Benzyl-d7-cx (K) can be chosen depend on the data of the run loaded.

4.2.4 Development and user guide for Importer 1.0

4.3.3.1 Working Flow & Capabilities

The software tool Importer 1.0 was developed as an application to identify pairs of doublet peaks from liquid chromatography mass spectra of two isotopic tags modified peptides mixture.

There are mainly two steps to generate the list of doublet peak pairs in the work flow.

Step 1: Export Thermo Raw files to FT1 text files using raxport which is available online http://code.google.com/p/raxport/ The FT1 text file is input file for Importer.

An example showing FT1 text file format is as below:

```
H       Extractor Raxport v3.3
H       m/z     Intensity Resolution      Baseline  Noise     Charge
H       Instrument Model  LTQ Orbitrap XL
S       1       1
I       RetentionTime     0.00325
I       ScanType FT-MS1
I       ScanFilterFTMS + p NSI Full ms [300.00-1600.00]
300.06152       1557.98 67004   7.24    390.91  0
300.07932       2514.50 52204   7.24    390.91  0
300.18051       720.50  74904   7.24    390.91  0
```

S line: This is scan #1.

I line: Retention time is at 0.00325 minute; the scan type is FT-MS1; the filter text is "FTMS + p NSI Full ms [300.00-1600.00]", which means FTMS acquisition with positive nanospray ionization for full MS1 scan from 300 Da to 1600 Da

Data block: Each row represents a peak. High-resolution MS data contain six columns representing "m/z", "Intensity", "Resolution", "Baseline", "Noise", and "Charge". Low-resolution MS data contain two columns representing "m/z" and "Intensity".

Step 2: Open a Windows console and run Importer 1.0 with the input of the following format:

```
Importer_x86 [FT1 file name] [output file name] [parameter file name] [y_lb] [ry_lb] [ry_ub] [# of column in FT1 file]
```

[FT1 file name] is the name of FT1 text file generated in step1; [output file name] is the name of the output text file set by the user, and the pairs list generated is saved in this file; [y_lb] is intensity threshold, and all data points with an intensity lower than this cutoff value are filtered out and will not participate in the comparison. [parameter file name] is the name of a text file set by the user which includes the specific requirement for the pairs to be selected, and its format is as following:

z_1	lower bound of $\Delta m/z$ at z_1	upper bound of $\Delta m/z$ at z_1
z_2	lower bound of $\Delta m/z$ at z_2	upper bound of $\Delta m/z$ at z_2
.	.	.
.	.	.
.	.	.

It contains three columns; the first column lists chosen charge states (*z*) of interest, the second and third column lists the lower and upper bound of $\Delta m/z$ and determines the interval of $\Delta m/z$ for each charge state. The user can set the interval according to the specific needs. The same charge state pairs which satisfy this condition can then be selected to compare their intensity ratio. The exception is the data points of which *z* cannot be identified and shows 0 in charge state column in the FT1 text file. If the user is interested in these data points as well, they can add a line ”0 -1 -1” in the parameter file to enable the comparison of these data points with other data points regardless of charge state. [ry_lb] and [ry_ub] are lower and upper bound of an interval that the intensity ratio of a pair needs to fall into to be selected. [# of column in FT1 file] is either 6 when the data is collected in a high-resolution instrument; or 2 when the data is collected in a low-resolution instrument.

Here is how Importer works: The information including *m/z*, Intensity and charge *z* of each data point from FT1 text file is extracted. Importer then filters out data points lower than cutoff value [y_lb]. In each scan, a pair with the same charge state can be selected and saved in [output file name] only when $\Delta m/z$ of the pair falls into the interval clarified in text file [parameter file name], while simultaneously the intensity ratio falls into the interval determined by [ry_lb] and [ry_ub]. Additionally, Importer gives a summarization of minimum, maximum and average of noise of all data points, and can be useful for future setting of the intensity threshold by user when additional runs are needed.

Importer is a standalone application which does not depend on search engines like Mascot. Therefore, it provides complementary information, and the results from Importer and Mascot can be combined to increase the confidence level to identify a modified peptide.

4.3.3.2 Example use of Importer

Here is an example showing the setting of the parameter.txt and a command using this text file:

Parameter.txt:

```
2 3.482 3.562
3 2.308 2.388
4 1.721 1.801
```

The command line is as below:

```
Importer_x86  Run1.FT1  pair.txt  parameter.txt  10000 0.8 1.25 6
```

This command line means FT1 text file Run1.FT1 which is generated from a high-resolution instrument data is loaded. The list of the pairs selected after Importer search is stored in pair.txt. The data points with intensity lower than 10000 will be filtered out and do not participate in the comparison. For data points with $z = 2$, a pair can only be selected when $\Delta m/z$ of the pair falls in the interval of [3.482, 3.562], and while the intensity ratio falls in the interval of [0.8, 1.25] as well. For data points with $z = 3$, a pair can be selected only when $\Delta m/z$ of the pair falls in the interval of [2.308, 2.388], and the intensity ratio falls in the interval of [0.8, 1.25]. For data points with $z = 4$, a pair can be selected only when $\Delta m/z$ of the pair falls in the interval of [1.721, 1.801], and the intensity ratio falls in the interval of [0.8, 1.25].

The output has the format as below:

```
Summarizing ...
TOTAL_SEC_NUM = 2593
TOTAL_ELE_NUM = 5822539
MIN_NOISE = 364.05
MAX_NOISE = 167327.53
MEAN_NOISE = 26808.96

Existing charge state array number = 25
charge state list
0       1       9       2       3       8       13      19      4       11
17      10      16      6       5       12      15      23      7       14
18      20      22      25      21

TOTAL_SEL_SEC_NUM = 543
TOTAL_SEL_PAIR_NUM = 1901
```

The output test file pair.txt has the format as below:

2086	18.35429	2	208	344.68640	386886.78	73604	38.43	2647.32	2	227	348.20819	310825.75	73204	38.39	2648.42	2
2134	18.66949	1	879	537.88989	304460.59	58301	332.31	3082.54	3	902	540.23798	333043.53	58001	334.56	3090.17	3
2134	18.66949	2	882	538.22467	203592.14	57604	332.63	3083.63	3	903	540.57227	227063.17	58204	334.88	3091.25	3

Each row represents a pair identified. The first and the second column are the scan number and retention time of the pair. The third column represents n^{th} pair identified in the scan. Values from 4^{th} to 10^{th} column and values from 11^{th} to 17^{th} column provide information of two peaks in the pair correspondingly. The 4^{th} column provides the position of the data point in the scan. The 5^{th} column to 10^{th} column represents "*m/z*", "Intensity", "resolution", "baseline", "noise" and "charge" correspondingly.

4.3 Results and Discussion

4.3.1 Performance comparison of d_0-benzylguanidine tagged peptide and d_7-benzylguanidine tagged peptide during fragmentation by ETD

Since the d_0-benzylguanidine tag has been demonstrated to be a very useful tag for improving sequence coverage with ETD, its isotopic analog d_7-Benzylguanidine tag is prepared as well and is evaluated for performance with ETD. D_7-benzylguanidine tag has the identical chemical structure as d_0-benzylguanidine tag, with an exception of all benzyl hydrogens are replaced with their isotopic form deuterium; therefore we expect to see its performance as comparable to d_0-benzylguanidine tag. The mixture of BSA and β-Casein tryptic peptides have been prepared and divided into two equal portions to react with d_0 or d_7-benzylguanidine tag in parallel. Following the acid cleavage, two parallel LC-MS/MS runs with ETD were performed for the vial containing d_0-benzylguanidine-tagged tryptic peptides and the vial containing d_7-benzylguanidine-tagged tryptic peptides. Fragmentation patterns of d_0 and d_7 labeled peptide ions by ETD are then compared and revealed to be highly comparable. **Figure 4.1** is an example

showing the comparison of the ETD spectra of triply charged SLHTLFGDELCK labeled by d_0-benzylguanidine tag and the ETD spectra of triply charged SLHTLFGDELCK labeled by d_7-benzylguanidine tag. SLHTLFGDELCK is a tryptic peptide digested from BSA. A complete set of *z* ions are observed in both of the spectra, and meanwhile, the same set of *c* ions including c_3 to c_{11} ions are observed in both spectra. In addition to that, the relative intensities of each fragment ion peak from two spectra are generally quite similar.

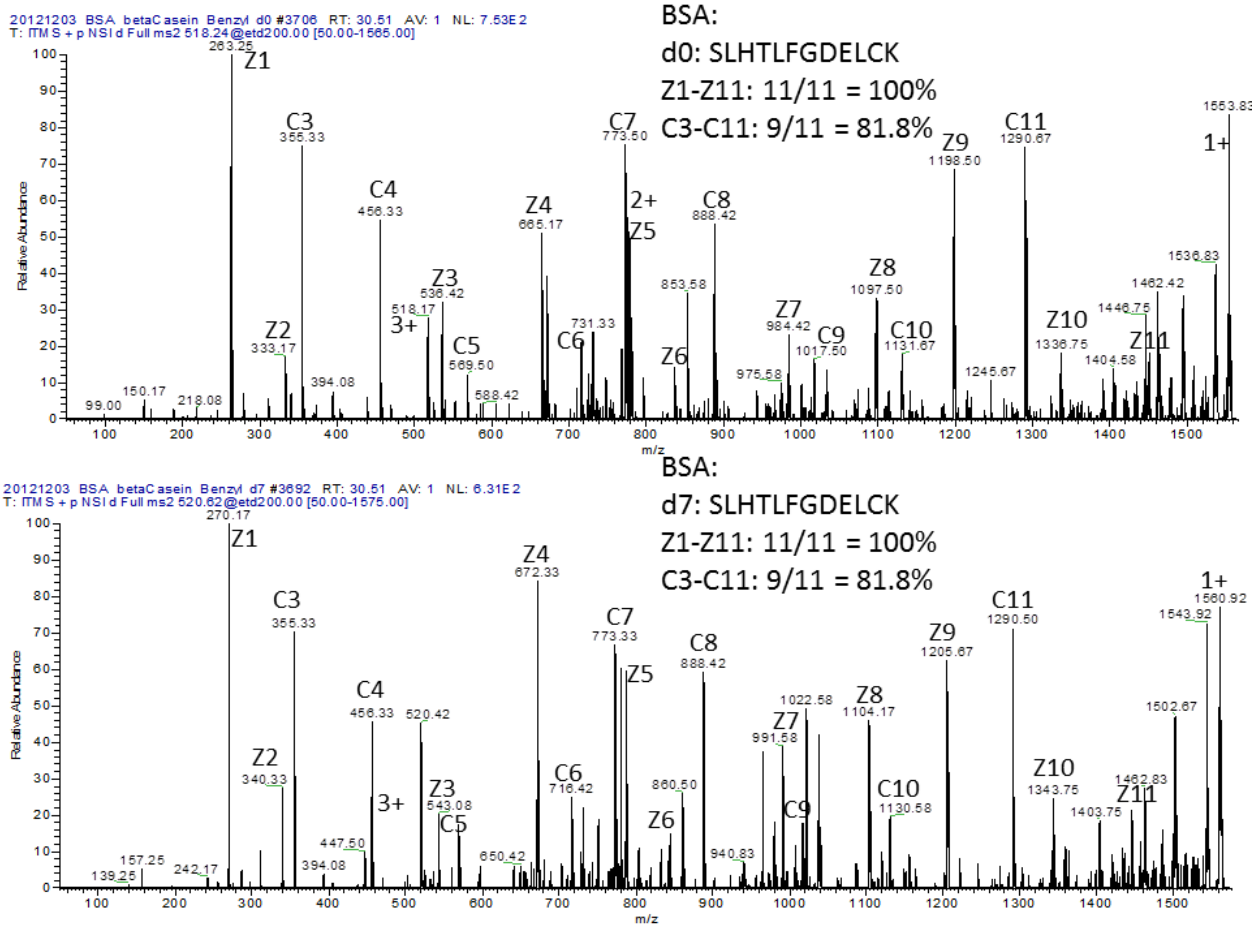


Figure 4.1 Comparison of the fragmentation performance with ETD between triply charged d_0-benzylguandine-tagged SLHTLFGDELCK (upper) and triply charged d_7-benzylguanidine-tagged SLHTLFGDELCK (lower)

Figure 4.2 shows another example comparing the performance between doubly charged d_0-benzylguandine-tagged VLPVPQK and doubly charged d_7-benzylguanidine-tagged VLPVPQK with ETD. VLPVPQK is a tryptic peptide digested from β-Casein. The sequence coverage of *z* ions is 100% with the

expected absence of z_3 and z_5 ions due to the prolines again for both of the labeled peptides, and similar fragmentation patterns are observed in this case as well. Clearly, ETD fragmentation pattern of the modified peptide derivatized by d_7-benzylguanidine tag is the same as that of the modified peptide derivatized by d_0-benzylguanidine tag as we expect to see.

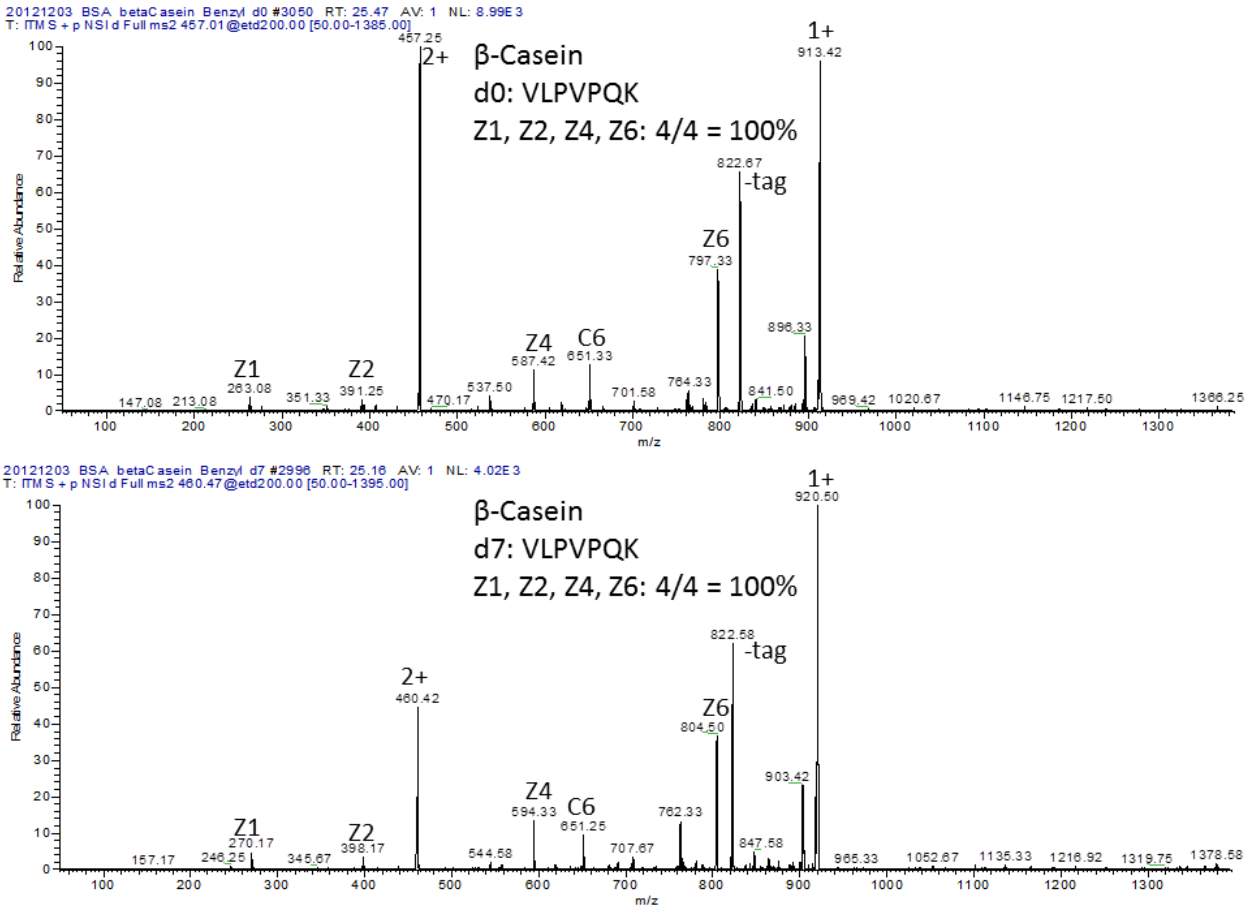


Figure 4.2 Comparison of the performance with ETD between doubly charged d_0-benzylguandine-tagged VLPVPQK (upper) and doubly charged d_7-benzylguanidine-tagged VLPVPQK (lower)

4.3.2 Performance evaluation of d_0/d_7-benzylguanidine tags with tryptic BSA/β-Casein peptides for quantitative study

In order to see if the d_0/d_7 tag pairs are useful for quantitative study, d_0-benzylguanidine tag precursor and d_7-benzylguanidine tag precursor have been prepared in parallel and allowed to react with

the same amount of tryptic peptides. The two vials of modified tryptic peptides labeled by d_0-benzylguanidine tag and d_7-benzylguanidine tag separately have been mixed at the ratio of 1:1 and loaded on the column to do LC-MS or LC-MS/MS. LC-MS has been performed on the low resolution instrument LTQ coupled with UPLC, while LC-MS/MS has been performed on high resolution instrument LTQ-Orbitrap coupled with UPLC. From these, comparisons of peak areas between d_0- and d_7-benzylguanidine-labeled tryptic peptide ions in full MS scans can be made to see if there is quantifiable behavior in this d_0/d_7-benzylguanidine-tagged peptides (1:1) mixture. **Figure 4.3** shows the extracted ion chromatogram (EIC) of doubly charged d_0-benzyl-guanidine tagged VLPVPQK (upper) and EIC of doubly charged d_7-benzyl-guanidine tagged VLPVPQK (lower) from the LC-MS run on LTQ. The ratio of peak area of doubly charged d_0-benzyl-guanidine tagged VLPVPQK to the peak area of doubly charged d_7-benzyl-guanidine tagged VLPVPQK is 1.15, which is essentially 1:1, as expected.

LC-MS of d0/d7 mixture (1:1)
EIC of d0 labeled VLPVPQK (2+) : 16321906
EIC of d7 labeled VLPVPQK (2+) : 14224820
ratio of area (d0/d7) = 16321906/14224820 = 1.15≈ 1

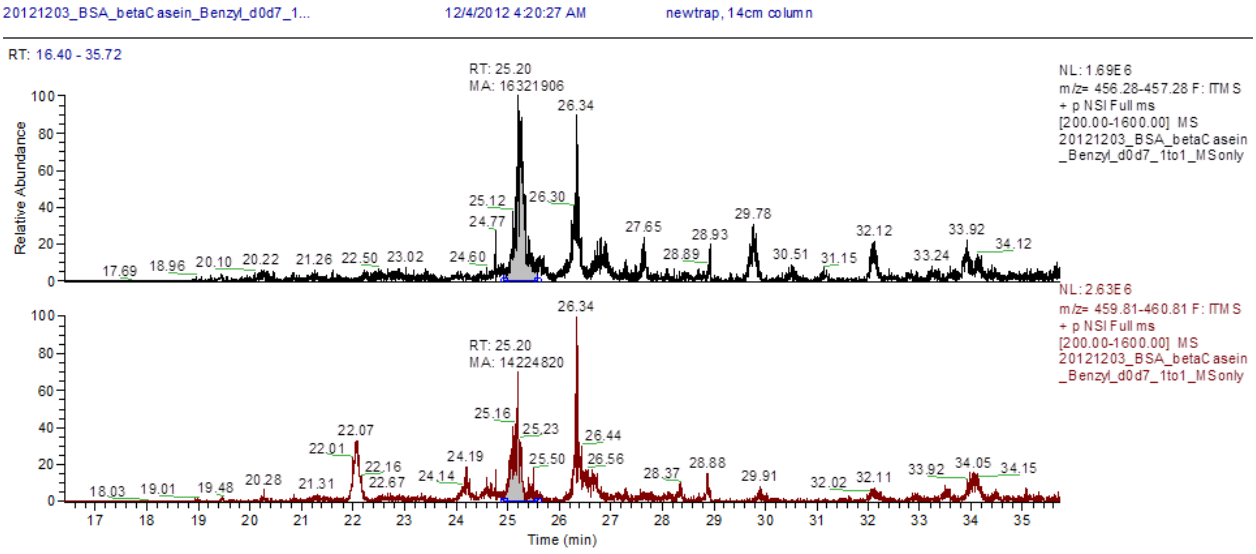


Figure 4.3 EIC of doubly charged d_0-benzylguanidine-tagged VLPVPQK (upper) and doubly charged d_7-benzylguanidine-tagged VLPVPQK (lower) from LC-MS on LTQ

Figure 4.4 shows the EIC of doubly charged d_0-benzylguanidine-labeled VLPVPQK and EIC of doubly charged d_7-benzylguanidine-labeled VLPVPQK in full MS scans from the LC-MS/MS run on high resolution instrument LTQ-Orbitrap. The ratio of peak area of doubly charged d_0-benzylguanidine-tagged VLPVPQK to the peak area of doubly charged d_7-benzylguanidine-tagged VLPVPQK is 1.16 which is close to 1 as well, and is very similar to the number from the LTQ instrument. Thus this case shows d_0/d_7-benzylguanidine tag pairs are very useful for the quantitative study.

LC-MSMS (Orbitrap) of d0/d7 mixture (1:1)
EIC of d0 labeled VLPVPQK (2+) : 501137664
EIC of d7 labeled VLPVPQK (2+) : 432527223
ratio of area (d0/d7) = 501137664/432527223 = 1.16≈ 1

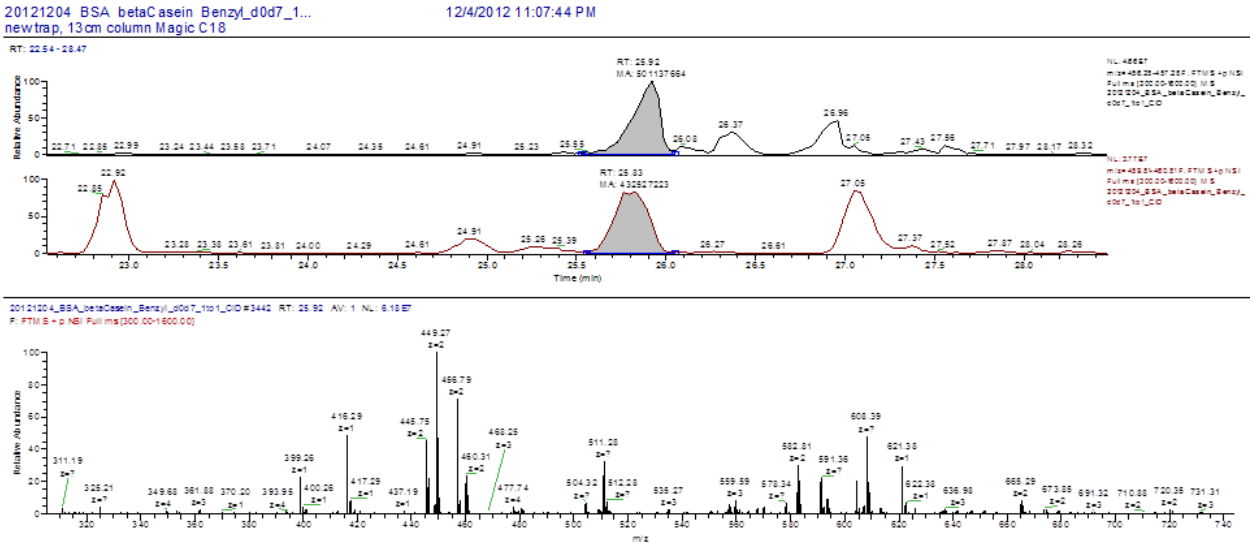


Figure 4.4 EIC of doubly charged d_0-benzylguanidine-tagged VLPVPQK (upper) and doubly charged d_7-benzylguanidine-tagged VLPVPQK (lower) on LTQ-Orbitrap

Mascot search is conducted for the raw isotopic data collected on the high resolution instrument LTQ-Orbitrap for the sample which is the 1:1 mixture of d_0- and d_7-benzylguanidine-tagged tryptic peptides, and a list of derivatized peptides identified by Mascot is generated. Some peptides in the list are not identified with both tags and it is hard to tell if that is due to the limitation of mass spectrometer or the

limitations of Mascot's search algorithm. Meanwhile, some peptides in the list have very bad scores, with a *p* value much higher than 0.05. Since a substantial length of time is required for manual inspection and validation for each hit in the list, I started the development of a software tool which is called Importer, to search for the doublet pairs with an intensity ratio close to 1:1. After checking with Importer for the hits from the Mascot output list, the results are shown from **Table 4.1** to **Table 4.4**. "Good pair" indicates the pair of d_0-benzylguanidine-tagged peptide and the corresponding d_7-benzylguanidine-tagged peptide are both in the list with a good score (*p* value is less than 0.05) from Mascot search. "Bad pair" indicates that at least one tagged peptide of the pair has a *p* value higher than 0.05, or one of the labeled formats is not identified at all in Mascot search. **Table 4.1** shows the sequence of the good peptide pairs which are identified from Mascot search and then also validated through the use of Importer.

	length	start	end	sequence
1	5	524	528	AFDEK
2	8	123	130	NECFLSHK
3	9	310	318	SHCIAEVEK
4	10	548	557	KQTALVELLK
5	10	66	75	LVNELTEFAK
6	10	35	44	FKDLGEEHFK
7	12	89	100	SLHTLFGDELCK
8	13	76	88	TCVADESHAGCEK
9	12	298	309	LKECCDKPLLEK
10	13	139	151	LKPDPNTLCDEFK
11	15	437	451	KVPQVSTPTLVEVSR
12	16	508	523	RPCFSALTPDETYVPK
13	16	168	183	RHPYFYAPELLYYANK

Table 4.1 Validated good pairs from Mascot with application of Importer

Table 4.2 shows the list of bad pairs in the result list from Mascot search but are verified in Importer. All the peptides in **Table 4.2** either are identified in Mascot with a bad score or only one tagging format is identified. However, since the $\Delta m/z$ between the same charge state peaks of modified peptides within the same pair in full MS scan is a fixed value, and the intensity ratio of the pair of doublet peaks is

close to 1:1, all of the peptide pairs which have these characteristics are validated in Importer. Clearly, Importer has an important role in verification of isotopic tagged peptides.

	length	start	end	sequence
1	5	156	160	KFWGK
2	7	198	204	GACLLPK
3	7	257	263	LVTDLTK
4	7	483	489	LCVLHEK
5	8	37	44	DLGEEHFK
6	9	549	557	QTALVELLK
7	8	413	420	QNCDQFEK
8	9	460	468	CCTKPESER
9	11	210	220	EKVLASSARQR
10	10	300	309	ECCDKPLLEK
11	12	286	297	YICDNQDTISSK
12	12	375	386	EYEATLEECCAK
13	13	236	248	AWSVARLSQKFPK
14	13	387	399	DDPHACYSTVFDK
15	14	184	197	YNGVFQECCQAEDK
16	15	169	183	HPYFYAPELLYYANK
17	6	123	128	EMPFPK (β-Casein)
18	7	185	191	VLPVPQK (β-Casein)
19	8	113	120	VKEAMAPK (β-Casein)

Table 4.2 Bad pairs from the result list of Mascot search but verified in Importer

Table 4.3 lists the bad pairs which are discovered from Mascot search but are not validated in Importer, therefore they are excluded from the peptide list. Importer validated 19 peptide pairs while excluding 9 peptide pairs from the bad pairs list generated from Mascot search. Therefore, Importer has an important role in differentiation of the right pairs from the wrong pairs in the bad pairs list from Mascot search.

	length	start	end	sequence
1	6	341	346	NYQEAK
2	12	569	580	TVMENFVAFVDK
3	12	106	117	ETYGDMADCCEK
4	14	198	211	GACLLPKIETMREK
5	13	118	130	QEPERNECFLSHK
6	21	508	528	RPCFSALTPDETYVPKAFDEK
7	21	400	420	LKHLVDEPQNLIKQNCDQFEK
8	6	361	386	HPEYAVSVLLRLAKEYEATLEECCAK
9	16	1	16	MKVLILACLVALALAR

Table 4.3 Bad pairs from the Mascot search and not identified in Importer

There are an additional three peptides which are validated with manual inspection but are not able to be identified in the Importer. Their sequences are TPVSEKVTK, NECFLSHKDDSPDLPK and LKPDPNTLCDEFKADEK. The reason is the charge states of the peaks are not resolved automatically in the MS scan due to the poor pattern of the isotopic peaks, and then the information of the corresponding peaks is not extracted and converted into the input file for Importer to analyze. There are six additional peptide pairs identified and validated with manual inspection as shown in **Table 4.4**. All the six peptides are not in the list from the Mascot search; however the information of the mass and its corresponding sequence are available after manual calculation based on Mascot search results for other samples (eg. BSA/β-Casein tryptic peptides). Since the high resolution instrument LTQ-Orbitrap is used in this study, the *m/z* of the doublet peaks that are of interest are used to check against the result list from Importer, and then validated by manual inspection. Importer is advantageous in the case when the period of elution time of the tagged peptide is extremely short. Even when the tagged peptides only appear for a few seconds, and the mass spectrometer does not select the peak to perform MS/MS, the modified peptide pair appears as doublet peaks in the full MS scans, this pair still can be found by search with Importer. As Importer is not dependent on Mascot, and is therefore complementary to the Mascot search. In the case of mutation of the protein, it is very difficult to be identified by Mascot, since the fragmentation pattern of the peaks changes dramatically to make it nearly impossible for Mascot to identify the hit. However, as

long as the high resolution mass is known, the hits can be identified by Importer since Importer only looks for the pair of doublet peaks which have fixed $\Delta m/z$ and reasonable peak intensity ratio. There is an identified tagged peptide which was digested from β-Casein in **Table 4.4** and it has a phosphorylation site in the sequence. **Table 4.5** lists all validated d_0/d_7 benzyl-guanidine tag modified peptide pairs which are tagged after trypsin digestion of BSA/β-Casein. There are 38 modified peptide pairs from BSA and there are 4 modified peptide pairs from β-Casein.

	length	start	end	sequence
1	7	562	568	ATEEQLK
2	8	131	138	DDSPDLPK
3	8	249	256	AEFVEVTK
4	11	402	412	HLVDEPQNLIK
5	12	106	117	ETYGDMADCCEK
6	16	48	63	FQ**S**EEQQQTEDELQDK (β-Casein)

Table 4.4 Additional peptide pairs identified with Importer and manual inspection.

	length	start	end	sequence
1	5	524	528	AFDEK
2	5	156	160	KFWGK
4	7	198	204	GACLLPK
5	7	257	263	LVTDLTK
6	7	483	489	LCVLHEK
7	7	562	568	ATEEQLK
8	8	123	130	NECFLSHK
9	8	37	44	DLGEEHFK
10	8	413	420	QNCDQFEK
11	8	131	138	DDSPDLPK
12	8	249	256	AEFVEVTK
13	9	310	318	SHCIAEVEK
14	9	549	557	QTALVELLK
15	9	460	468	CCTKPESER
16	9	490	498	TPVSEKVTK
17	10	548	557	KQTALVELLK
18	10	66	75	LVNELTEFAK
19	10	35	44	FKDLGEEHFK
20	10	300	309	ECCDKPLLEK
21	11	210	220	EKVLASSARQR
22	11	402	412	HLVDEPQNLIK
23	12	89	100	SLHTLFGDELCK
24	12	298	309	LKECCDKPLLEK
25	12	286	297	YICDNQDTISSK
26	12	375	386	EYEATLEECCAK
27	12	106	117	ETYGDMADCCEK
28	13	76	88	TCVADESHAGCEK
29	13	139	151	LKPDPNTLCDEFK
30	13	236	248	AWSVARLSQKFPK
31	13	387	399	DDPHACYSTVFDK
32	14	184	197	YNGVFQECCQAEDK
33	15	437	451	KVPQVSTPTLVEVSR
34	15	169	183	HPYFYAPELLYYANK
35	16	123	138	NECFLSHKDDSPDLPK
36	16	508	523	RPCFSALTPDETYVPK
37	16	168	183	RHPYFYAPELLYYANK
38	17	139	155	LKPDPNTLCDEFKADEK
1	6	123	128	EMPFPK (β-Casein)
2	7	185	191	VLPVPQK (β-Casein)
3	8	113	120	VKEAMAPK (β-Casein)
4	16	48	63	FQSEEQQQTEDELQDK (β-Casein)

Table 4.5 List of all validated d_0/d_7 benzylguanidine-tag modified peptide pairs derivatized after trypsin digestion of BSA/β-Casein.

Figure 4.5 shows an example of pair identification by Importer. The upper figure is the EIC of doubly charged d_0-benzylguanidine-tagged LKPDPNTLCDEFK (z = 2), the lower figure is the EIC of doubly charged d_7-benzylguanidine-tagged LKPDPNTLCDEFK (z = 2). The seven straight lines represent the pair being identified in 7 different scans as the ratio of intensities in the 7 scans all fall in the range of 0.8 to 1.25. They are scan 3724 (27.43min), scan 3730 (27.46min), scan 3736 (27.49min), scan 3790 (27.78min), scan 3796 (27.81min), scan 3802 (27.84min) and scan 3808 (27.87min).

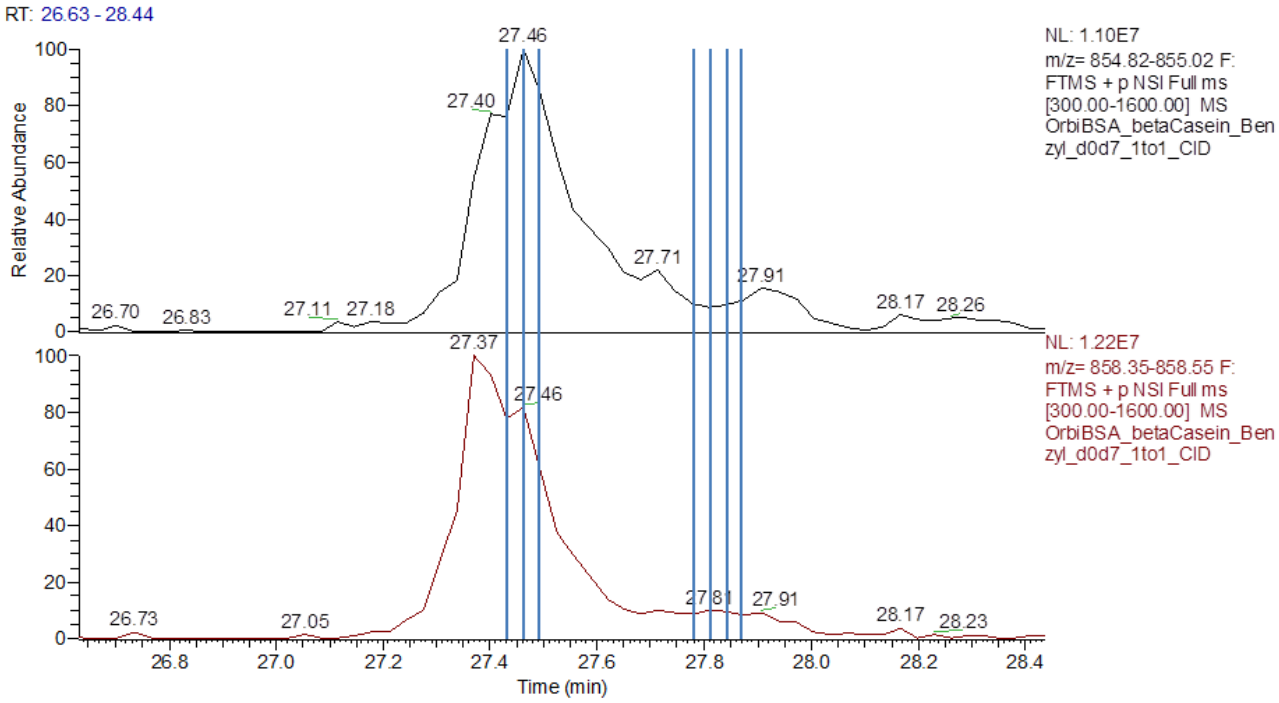


Figure 4.5 D_0/d_7 benzylguanidine tag modified doubly charged LKPDPNTLCDEFK validated by Importer. The seven straight lines represent the pair being identified in 7 different scans as the ratio of intensities in the 7 scans all fall in the range of 0.8 to 1.25. They are scan 3724 (27.43min), scan 3730 (27.46min), scan 3736 (27.49min), scan 3790 (27.78min), scan 3796 (27.81min), scan 3802 (27.84min) and scan 3808 (27.87min).

4.3.4 Identification of premixed d_0/d_7-benzylguanidine tag modified peptide pairs digested from BSA/β-Casein with rLys-C before modification

The protein mixture of BSA and β-Casein were digested with rLys-C and then separated into two equal portions to be labeled by d_0-benzylguanidine tag or d_7-benzylguandine tag in parallel. The tagged

peptides were then postmixed at a ratio of 1:1 before being loaded on the column to perform LC-MS/MS with CID analysis. Importer search and manual inspection were performed after Mascot search for the LC-MS/MS data of the mixture, and 21 d_0/d_7-benzylguanidine-tagged peptide pairs digested from BSA before modification were verified as shown in the left of **Table 4.6**. In addition, the pair of d_0/d_7-benzylguanidine tagged VLPVPQK from β-Casein was identified as well.

The mixture of d_0-benzylguanidine tag precursor and d_7-benzylguanidine tag precursor premixed at the ratio of 1:1 was prepared as well and then used to tag the rLys-C digested peptides. Labeled peptides were analyzed with LC-MS/MS and subsequent Mascot search. Importer search and manual inspection were used afterwards, and 9 d_0/d_7-benzylguanidine-tagged peptide pairs from BSA were verified as shown in the right of **Table 4.6**. All the peptide pairs that are identified in the right list are marked as bold in both of the lists. Clearly, the list on the right is a subset of the list on the left, and this is due to two possible reasons. The first reason is when the mixture of tag precursors were used to label the peptides, half amount of each tag starting material were mixed and used when labeled the same amount of the peptides as before, therefore, the peptides have less chance to meet the tag precursors. The second possibility is that some peptides have preferences to react with one type of benzylguanidine tag over the other, resulting in the identification of a single format of tagged peptide instead of d_0/d_7-benzylguanidine tagged peptide pair.

There are 21 d_0/d_7-benzylguanidine-tagged BSA rLys-C digested peptide pairs identified as shown in **Table 4.6** (left), and they contain 240 AA. Since mature BSA contain 583 AA in the sequence, the sequence coverage calculated in the way is 41.2%. There are 477 AA in the BSA sequence are identified with the LC-MS/MS analysis of the rLys-C digested standard, since 240 of 477 are identified, the sequence coverage calculated in this way is 50.3%. There is one d_0/d_7-benzylguanidine tag modified peptide digest from β-Casein identified, and there are only 4 more peptides identified in the peptide digest standard. AVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV is the peptide at the very end of β-Casein and does not contain K, thus is unable to be derivatized by guanidine tag. Phosphorylated peptide FQ**S**EEQQQTEDELQDK is identified in the underivatized form instead of in the derivatized form in the run for sample which should only contain modified peptides in theoretical. This might indicate that this peptide is not easy to be modified and also requires stronger solvent to be removed by wash. Phosphorylated

d_0/d_7-benzylguanidine tag modified peptide pair with a sequence of FQ**S**EEQQQTEDELQDK is identified when trypsin is used. However, that pair is not identified by Mascot, instead, it is identified with the combination of Importer search and manual inspection. Another two identified peptide IHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSK and peptide EMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSK have K in the sequence, however since both of them are really long peptides (49 AA, 62 AA), that might be reason that they can't be labeled or identified.

	length	start	end	Sequence
1	7	229	235	**FGERALK**
2	8	131	138	DDSPDLPK
3	8	37	44	DLGEEHFK
4	8	29	36	**SEIAHRFK**
5	9	310	318	**SHCIAEVEK**
6	10	66	75	LVNELTEFAK
7	11	402	412	**HLVDEPQNLIK**
8	12	286	297	**YICDNQDTISSK**
9	12	375	386	**EYEATLEECCAK**
10	13	387	399	**DDPHACYSTVFDK**
11	13	139	151	**LKPDPNTLCDEFK**
12	14	184	197	YNGVFQECCQAEDK
13	18	438	455	**VPQVSTPTLVEVSRSLGK**
14	17	421	437	LGEYGFQNALIVRYTRK
15	7	257	263	LVTDLTK
16	10	588	597	EACFAVEGPK
17	13	76	88	TCVADESHAGCEK
18	13	118	130	QEPERNECFLSHK
19	17	101	117	VASLRETYGDMADCCEK
20	8	249	256	AEFVEVTK
21	12	89	100	SLHTLFGDELCK
22	7	185	191	VLPVPQK (β-Casein)

	Length	start	end	sequence
1	7	229	235	**FGERALK**
2	8	29	36	**SEIAHRFK**
3	9	310	318	**SHCIAEVEK**
4	11	402	412	**HLVDEPQNLIK**
5	12	286	297	**YICDNQDTISSK**
6	12	375	386	**EYEATLEECCAK**
7	13	387	399	**DDPHACYSTVFDK**
8	13	139	151	**LKPDPNTLCDEFK**
9	18	438	455	**VPQVSTPTLVEVSRSLGK**

Table 4.6 Comparison between verified tagged peptide pairs from the mixture of d_0-benzylguanidine-tagged peptides and d_7-benzylguanidine-tagged peptides postmixed at the ratio of 1:1 right before loading on LC-MS/MS (left) and verified tagged peptide pairs from the sample labeled with premixed d_0/d_7-benzylguanidine tag precursor at the very beginning at the ratio of 1:1 (right).

The rLys-C was used instead of trypsin to digest the protein mixture of BSA/β-Casein. The rLys-C, which is similar to a native Lys-C and less expensive than Lys-C, cleaves at the carboxyl side of lysine residues with exceptional specificity, while trypsin specifically cleaves at the carboxylic side of lysine and arginine residues. Therefore, theoretically the application of the rLys-C alone will generate longer peptides when compared to the trypsin. The longest tagged tryptic peptide sequence identified as a pair is composed of 17 amino acid residues, while the longest tagged rLys-C digested peptide sequence identified as a pair is composed of 18 residues. The median number of residues in the sequence of modified peptide pairs is 10 when the trypsin is used, and is 12 when the rLys-C is used. The comparison of the length and the sequence of the tagged peptide pairs either digested with trypsin or rLys-C are listed in **Table 4.7**. The trypsin is still performing better than rLys-C although arginine terminated peptides are not labeled since the tag only targets at the lysine site. When trypsin is used, there are 38 modified peptide pairs identified from BSA and there are 4 modified peptide pairs from β-Casein. When rLys-C is used, there are 21 modified peptide pairs identified from BSA and 1 modified peptide pair from β-Casein. The sequence coverage of mature BSA is 54.2% when trypsin is used and is 41.2% when rLys-C is used; the combined sequence coverage is 65.0% which shows substantial overlap in residues identified.

	length	start	end	sequence
1	5	524	528	AFDEK
2	5	156	160	KFWGK
4	7	198	204	GACLLPK
5	7	257	263	LVTDLTK
6	7	483	489	LCVLHEK
7	7	562	568	ATEEQLK
8	8	123	130	NECFLSHK
9	8	37	44	DLGEEHFK
10	8	413	420	QNCDQFEK
11	8	131	138	DDSPDLPK
12	8	249	256	AEFVEVTK
13	9	310	318	SHCIAEVEK
14	9	549	557	QTALVELLK
15	9	460	468	CCTKPESER
16	9	490	498	TPVSEKVTK
17	10	548	557	KQTALVELLK
18	10	66	75	LVNELTEFAK
19	10	35	44	FKDLGEEHFK
20	10	300	309	ECCDKPLLEK
21	11	210	220	EKVLASSARQR
22	11	402	412	HLVDEPQNLIK
23	12	89	100	SLHTLFGDELCK
24	12	298	309	LKECCDKPLLEK
25	12	286	297	YICDNQDTISSK
26	12	375	386	EYEATLEECCAK
27	12	106	117	ETYGDMADCCEK
28	13	76	88	TCVADESHAGCEK
29	13	139	151	LKPDPNTLCDEFK
30	13	236	248	AWSVARLSQKFPK
31	13	387	399	DDPHACYSTVFDK
32	14	184	197	YNGVFQECCQAEDK
33	15	437	451	KVPQVSTPTLVEVSR
34	15	169	183	HPYFYAPELLYYANK
35	16	123	138	NECFLSHKDDSPDLPK
36	16	508	523	RPCFSALTPDETYVPK
37	16	168	183	RHPYFYAPELLYYANK
38	17	139	155	LKPDPNTLCDEFKADEK
1	6	123	128	EMPFPK (β-Casein)
2	7	185	191	VLPVPQK (β-Casein)
3	8	113	120	VKEAMAPK (β-Casein)
4	16	48	63	FQSEEQQQTEDELQDK (β-Casein)

	length	start	end	sequence
1	7	229	235	FGERALK
2	7	257	263	LVTDLTK
3	8	131	138	DDSPDLPK
4	8	37	44	DLGEEHFK
5	8	29	36	SEIAHRFK
6	8	249	256	AEFVEVTK
7	9	310	318	SHCIAEVEK
8	10	66	75	LVNELTEFAK
9	10	588	597	EACFAVEGPK
10	11	402	412	HLVDEPQNLIK
11	12	286	297	YICDNQDTISSK
12	12	375	386	EYEATLEECCAK
13	12	89	100	SLHTLFGDELCK
14	13	387	399	DDPHACYSTVFDK
15	13	139	151	LKPDPNTLCDEFK
16	13	76	88	TCVADESHAGCEK
17	13	118	130	QEPERNECFLSHK
18	14	184	197	YNGVFQECCQAEDK
19	17	421	437	LGEYGFQNALIVRYTRK
20	17	101	117	VASLRETYGDMADCCEK
21	18	438	455	VPQVSTPTLVEVSRSLGK
1	7	185	191	VLPVPQK (β-Casein)

Table 4.7 Comparison of d_0/d_7-benzylguanidine-tagged peptides pairs with the use of trypsin (left) or Lys-C (right) for digestion of protein mixture of BSA and *β*-Casein.

4.4 Conclusion

An isotopic pair of d_0- and d_7-benzylguanidine tags has been developed for quantitative study of proteins. ETD fragmentation patterns of derivatized peptides modified by d_7-benzylguanidine tag are demonstrated to have remarkable performances which are comparable to those modified by d_0-benzylguanidine tag. LC-MS/MS with CID on Orbitrap has been used for qualitative and quantitative study

of isotopic tags modified digested peptides. An application tool, Importer, has been developed to search for doublet peaks in full MS scan, which are corresponding to the modified peptide ions tagged by d_0- and d_7-benzylguanidine tags; which provides data that is complementary to the search engine Mascot.

4.5 Reference

[1] Kim, S.; Yi, K. Y. Di-2-Pyridyl Thionocarbonate. A New Reagent for the Preparation of Isothiocyanates and Carbodiimides. *Tetrahedron Lett.* **1985,** *26*, 1661-1664

4.6 Acknowledgement

Thanks to Aleš Marek for the synthesis of D_7-benzyl-isothiocyanate.

Chapter 5
A Novel Duplex Tandem Mass Spectrometry Assay for the Clinical Diagnosis of Neuronal Ceroid Lipofuscinoses (NCL)

5.1 Introduction

The neuronal ceroid lipofuscinosis (NCLs) are a group of hereditary neurodegenerative diseases primarily affecting children and adolescents who share similar clinical features, including seizures, mental regression, visual loss, behavior changes, movement disorders, and shortened life expectancy [1]. All NCL disorders are revealed by an accumulation of lipopigments in neural and peripheral tissues, eventually leading to death. The different types of NCLs are categorized by the age at which signals and symptoms first appear. Infantile Neuronal Ceroid Lipofuscinoses (INCL, also called NCL I), which is also known as the Santavuori-Heltia disease and begins between about 6 months and 2 years of age, is the most severe form of NCL. Classical Late Infantile Neuronal Ceroid Lipofuscinoses (c-LINCL, also called NCL II), which is known as Jansky-Bielschowsky disease and begins between ages 2 and 4, is the most common form of NCL.

NCL I is raised by a mutation in the CLN1 gene, which encodes the Palmitoyl Protein Thioesterase I enzyme (PPT I).

NCL II is a group of rare diseases and affects 0.36 to 0.46 per 100,000 live births [2]. NCL II typically exposed clinically at age 2 to 4 years with myoclonus, ataxia, impaired speech, and developmental regression [3]. Seizures are often the first manifestation, but variability in the time of onset and the appearance of the collective symptoms are observed. A gradual decline in visual ability follows, resulting in blindness by age 4 to 6 years. Affected children usually become wheelchair-bound between 4 and 6 years. Toward the late stages of the disease, feeding becomes difficult, resulting in loss of weight. Death happens by age 8 to 12 years [3-5].

NCL II results from mutations in the TPP I gene (previously named *CLN2* gene) [6], which encodes the 46-kDa tripeptidyl-peptidase 1(TPP I). TPP I is a lysosomal serine protease that functions in

the acidic milieu of the lysosome to remove groups of three amino acids from the amino terminus of proteins. The rat version of the enzyme is specific towards substrates possessing Ala-Ala-Phe or Gly-Pro-Met tripeptides at the *N*-terminus, whereas Ala-Ala-Phe amino acid sequence is the most favorable for the human TPP I [7] [8].

A clinical trial has been conducted to evaluate the safety and preliminary efficacy (phase 1) of human central nervous stem cells (HuCNS-SC) implanted into the cortex and lateral ventricles of patients with advanced neuronal ceroid lipofuscinoses, a cell therapy which can potentially synthesize and secrete TPP 1 and PPT 1 enzymes (Deficiency of PPT I enzyme also results in NCL). Therefore, a strong interest in screening both TPP I and PPT 1 deficiency simultaneously has arisen. Detection methods, e.g. fluorescence assays, have been reported to test the activity of TPP I enzyme [9-12]. However, current fluorescence-based screening protocols have limitations primarily because of variable background fluorescence, and for NCL I, it requires the addition of a coupling enzyme to yield a fluorescent compound. Meanwhile, a more advanced technique, electrospray ionization tandem mass spectrometry (ESI-MSMS) has been used in newborn screening labs to quantify the level of metabolites associated with treatable diseases using dried blood spots (DBS) collected from newborns [13]. Thus, in order to circumvent the limitations of the current method, we designed and synthesized Ala-Ala-Phe containing substrate for TPP I enzyme and developed an assay which is able to monitor TPP I enzyme activity quantitatively by triple quadruple mass spectrometer. Additionally, we developed the first duplex tandem mass spectrometry assay with a single injection for clinical diagnosis of both NCL I and NCL II in DBS from newborns.

5.2 Experimental Section

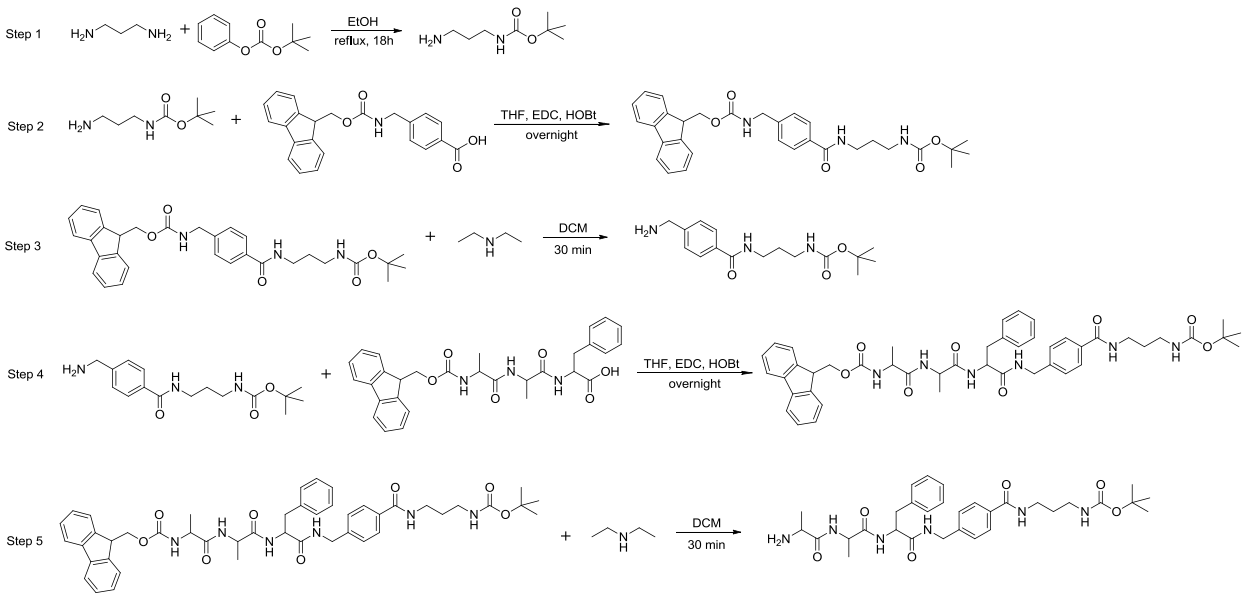
Using a previously developed synthesis route for the substrates and internal standard as well as a previously developed set of assay parameters for TPP I, my goal was to prepare bulk amount of internal standard and substrate for TPP I, and to modify the current assay method to detect TPP I activity for TPP I. Development of a duplex assay method of both TPP I and PPT I is the final objective of the project.

5.2.1 Materials

Fmoc-4-aminomethyl benzoic acid and *N*-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride were purchased from AnaSpec Inc. Fmoc-AAF was purchased from Lifetein LLC. Other starting materials were purchased from Sigma-Aldrich and used without further purification. Electrospray ionization mass spectra were acquired on a Bruker Esquire LC00066 mass spectrometer.

5.2.2 Synthesis of the TPP I substrate

The 5-step synthesis route of TPP I substrate is listed in the **Scheme 5.1**, and bulk amount of substrate are prepared starting from step 2 according to the method in the **Scheme 5.1**.



Scheme 5.1 Synthesis route of the TPP I substrate

Step 1: *tert*-Butyl phenyl carbonate (9.7 mL, 0.05 mol) is added to a solution of 1,3-diaminopropane (3.7 mL, 0.05 mol) in 35 mL ethanol in a 200-mL, single-necked, round-bottomed flask equipped with a stirring bar and reflux condenser [14]. The reaction mixture is heated gently to reflux overnight ensuring that the temperature falls in the range of 80-85 °C, resulting in a yellow solution. The reaction mixture is cooled to RT and the solution concentrated to approximately 7-8 mL using rotary evaporator, which leaves a yellow solution. Water (50 mL) is added and pH is adjusted to approximately

3 by careful addition of aqueous HCl followed by extraction with CH_2Cl_2 (80 mL X 3). The aqueous layer is adjusted to pH 13 by the addition of 2M NaOH and extracted with CH_2Cl_2 (100 mL X 5). The combined organic extracts are concentrated using rotary evaporator to afford 2.8 g of the yellow oil. Yellow oil crystallized within ~ 1 h. Yield: 33% MW: 174.2 MS: *m/z* = 175.0 $[M+H]^+$

Step 2: A solution of Fmoc-4-aminomethyl benzoic acid (74.7 mg, 0.2 mmol, 1 eq) in a 10 mL of anhydrous tetrahydrofuran is cooled to 0 °C on ice [15][16]. *N*-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (42.2 mg, 0.22 mmol, 1.1 eq) and 1-hydroxybenzotiazole (27.1 mg, 0.22 mmol, 1.1 eq) are added, and the suspension is stirred at 0 °C for 30 min. A solution of *N*-Boc-1,3-diaminopropane (34.8 mg, 0.2 mmol, 1 eq) in 2 mL of *N,N*-dimethylformamide (DMF) is added dropwise to the suspension. The reaction mixture is allowed to warm to RT and stirred overnight. The reaction mixture is concentrated by rotary evaporation. The residue is taken up in ethyl acetate (100 mL), and then washed with 1M HCl (80 mL) and water (80 mL X 3). The organic solvent is evaporated using rotary evaporator. The solid product is purified by crystallization from isopropyl alcohol to give the pure compound. MW = 529.4 MS: *m/z* = 552.4 $[M+Na]^+$

Step 3: A mixture of Fmoc-Protected product synthesized in Step 2 (80 mg, 0.15 mmol) is dissolved in 6 mL of diethyl amine (DEA)/dichloromethane mixture (V/V = 50/50) [4]. The mixture is magnetically stirred till white colored solution turns yellow (approximately 30min). The organic solvent is evaporated by rotary evaporation and the residual mass (yellow oil) is purified by column chromatography (SiO_2, particle size 40-63 μm) using stepped gradient of ethyl acetate/methanol(100/0 → 80/20), with 1% triethylamine. The 80/20 fraction is evaporated in the rotary evaporator. Sample is left in the desiccator for 1-2 days to afford crystals that have yellow tint. Yield: 82%. MW: 307.3 MS: *m/z* = 308.3 $[M+H]^+$; *m/z* = 615.3 $[2M+H]^+$

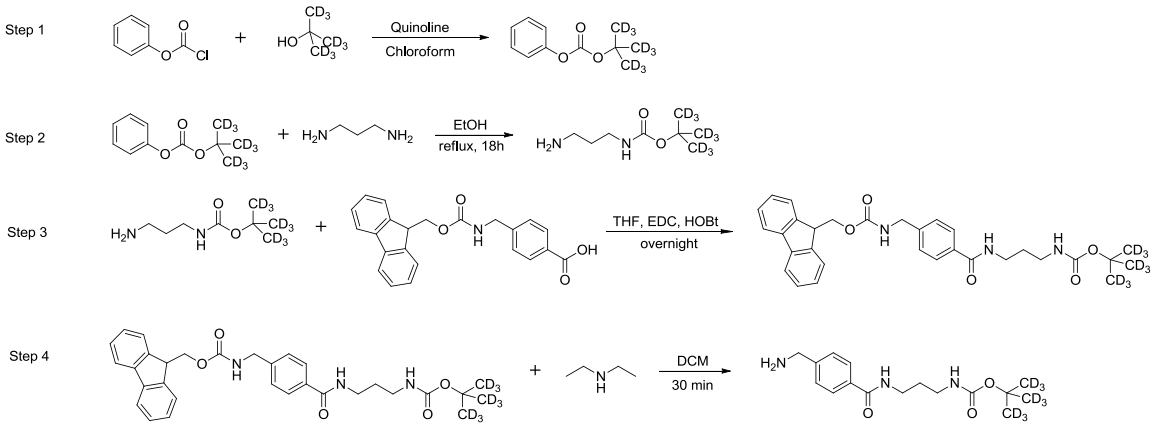
Step 4: A solution of Fmoc-AAF (4 mg, 0.008 mmol, 1 eq) in a 2 mL of anhydrous tetrahydrofuran is cooled to 0 °C on ice. *N*-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (1.6 mg, 0.0085 mmol, 1.1 eq) and 1-hydroxybenzotiazole (1.15 mg, 0.0085 mmol, 1.1 eq) are added, and the suspension is stirred for 30 min at 0 °C. A solution of product from step 3 (2.5 mg, 0.008 mmol, 1 eq) in 200 μL of DMF is added dropwise to the suspension. The reaction mixture is allowed to warm to room temperature and stirred overnight. The reaction mixture is concentrated by rotary evaporation. The residue is taken up

in ethyl acetate (5 mL) and washed with 1M HCl (5 mL), then with water (5 mL X 3). The organic solvent is evaporated using rotary evaporator. The solid product is washed with hot isopropanol, filtered and dried. Product is white solid. Yield: 65% MW: 818.9

Step 5: A mixture of Fmoc-protected product synthesized in Step 4 (4.1 mg, 0.005 mmol) is dissolved in 1 mL of DEA/dichloromethane (V/V = 50/50) mixture. The mixture is magnetically stirred until white colored solution turns yellow (approximately 30min). The organic solvent is evaporated by rotary evaporation and the residual mass (yellow oil) is purified by column chromatography (SiO_2, particle size 40-63 μm) using dichloromethane/ethyl acetate (V/V = 50/50), with 1% triethylamine followed by dichloromethane/ethyl acetate/methanol (V/V/V = 40/40/20) with 1% triethylamine to elute the substrate. Yield: 79%. MW: 596.6. MS: *m/z* 597.7 $(M+H)^+$.

5.2.3 Synthesis of the TPP I internal standard

The 4-step synthesis route of TPP I internal standard (IS) is listed in the **Scheme 5.2**, and bulk amount of IS are prepared starting from step 3 according to the method in **Scheme 5.2**.



Scheme 5.2 Synthesis route of the TPP I internal standard

Step 1: The d_9-*tert*-Butyl alcohol (0.128 mL, 1.20 mmol, 1 eq) and quinoline (0.141 mL, 1.20 mmol, 1eq) are combined in a dry screw cap vial with 1 mL of anhydrous dichloromethane. Phenyl

chloroformate (0.155 mL, 1.2 mmol, 1eq) is then added dropwise. Reaction is allowed to proceed overnight. d_9-*tert*-Butyl phenyl carbonate is purified on silica using a hexane/ethyl acetate gradient up to hexane/ethyl acetate (V/V = 50/50). ^{1}H NMR 400 MHz ($CDCl_3$) δ 1.54 (9H, s), 6.97-7.50 (5H, m) MS *m/z* 226.1 $(M+Na)^+$.

Step 2: The compound synthesized in the first step (9.3 mL, 0.05 mol) is added to a solution of 1,3-diaminopropane (3.7 mL, 0.05 mol) in 35 mL EtOH in a 200-mL, single-necked, round-bottomed flask equipped with a stirring bar and reflux condenser. The reaction mixture is heated gently to reflux overnight ensuring that the temperature falls in the range of 80-85 °C, resulting in a yellow solution. The reaction mixture is cooled to room temperature and the solution concentrated to approximately 7-8 mL by rotary evaporation, which leaves a yellow solution. Water (50 mL) is added and pH is adjusted to approximately 3 by careful addition of aqueous HCl followed by extraction with CH_2Cl_2 (80 mL X 3). The aqueous layer is adjusted to pH 13 by the addition of 2M NaOH and extracted with CH_2Cl_2 (100 mL X 5). The combined organic extracts are concentrated using rotary evaporator to yield 2.8 g of the yellow oil. Yellow oil crystallized within 1 h or so. Yield: 33% MW: 183.2 MS: *m/z* 184.1 $(M+H)^+$

Step 3: A solution of Fmoc-4-aminomethyl benzoic acid (0.1494 g, 0.4 mmol, 1 eq) in 20 mL of anhydrous tetrahydrofuran is cooled to 0 °C on ice. *N*-(3-Dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (0.0844 g, 0.44 mmol, 1.1 eq) and 1-hydroxybenzotiazole (0.0542 g, 0.44 mmol, 1.1 eq) are added, and the suspension is stirred for 30 min at 0 °C. A solution of *N*-d_9-Boc-1,3-diaminoproane (0.0732 g, 0.4 mmol, 1 eq) in 4 mL of DMF is added drop by drop to the suspension. The reaction mixture is allowed to warm to room temperature and stirred overnight. The reaction mixture is concentrated using rotary evaporator. The residue is taken up in 200 mL of ethyl acetate and washed with 1M HCl (160 mL), then water (160 mL X 3). The organic solvent is evaporated using rotary evaporator. The solid product is purified by crystallization from isopropyl alcohol to give 0.1513 g compound. Yield: 70% MW = 538.3

Step 4: A mixture of Fmoc-Protected compound synthesized in Step 3 (0.0807 g, 0.15 mmol) is dissolved in 6 mL of DEA/dichloromethane mixture (V/V = 50/50). The mixture is magnetically stirred till white colored solution turns yellow (approximately 30min). The organic solvent is evaporated using rotary evaporator and the residual mass (yellow oil) is purified by column chromatography (SiO_2, particle size

40-63 μm) using stepped gradient of ethyl acetate: methanol (100/0 → 80/20), 1% triethylamine. The right fractions checked with mass spectrometer are evaporated in the rotary evaporator to give 0.0250 g internal standard. Yield: 53% MW: 316.3. MS: *m/z* 317.3 $(M+H)^+$.

5.2.4 Assay procedure for screening of TPP I deficiency

TPP I substrate solution was prepared by dissolving 5 mg TPP I substrate in 1.5 mL of dimethyl sulfoxide (DMSO). 1 mg TPP I IS was dissolved in 1.5 mL assay buffer, which contained 0.15 M NaCl, 0.1 M sodium acetate (pH = 4 adjusted with acetic acid). This concentrated TPP I IS solution was further diluted with stock assay buffer to make a 300 μM IS solution. Assay cocktail was a 100 μL mixture composing of 7.2 μL of substrate solution (40 nmol), 5 μL of 300 μM IS solution (1.5 nmol) and additional 87.8 μL of stock assay buffer. A 3 mm dried blood spot (DBS) was placed in a 1.5 mL polypropylene Eppendorf tube containing 100 μL assay cocktail. The solution was then vortexed briefly and incubated for 10 h at 37 ºC in a thermostated air shaker at 250 rpm. After incubation, 10 μL of 15% trifluorocetic acid was added to quench the assay followed by 200 μL of 1 M NaOH, and then 800 μL ethyl acetate. The mixture was vortexed, centrifuged, and 750 μL of top layer were collected to dry down. Then the dried residue was reconstituted in 200 μL of acetonitrile/water (V/V =50/50) with 1% acetic acid for analysis by MS.

5.2.5 *Dwell time study*

Dwell time is a setting in the mass spectrometer that indicating the length of time that is spent collecting data at a particular mass. The previous setting for dwell time was 0.01 s. The values of coefficients of variances (CV) were compared between under the condition of 0.01 s and under condition of 0.1 s using the same sample. The result, which will be discussed in section *5.3.6,* shows 0.1 s is better than 0.01 s, and therefore 0.1 s is used in the further study.

5.2.6 Duplex assay procedure for the screening of both TPP I and PPT I deficiency

Dried blood spots (DBS) collected from deficient patient and healthy people were obtained from newborn screening laboratories and were used in our study as the biological sample. Two 3 mm punches of a dried blood spot from the same patient were assayed separately for screening of TPP I and PPT I deficiency. One DBS was placed in an Eppendorf tube containing 100 µL of TPP I assay cocktail including 1.5 nmol of TPP I IS and 40 nmol of TPP I substrate. The other DBS was added to the other Eppendorf tube containing PPT I IS and PPT I substrate. 5 TPP I deficiency, 5 PPT I deficiency and 40 healthy patient samples were processed. Samples for TPP I deficiency screening and samples for PPT I deficiency screening were incubated simultaneously at 37 °C for 10 h in parallel. After incubation, samples for PPT I deficiency screening were processed by Mariana in the way described in her thesis. Samples for TPP I deficiency screening were quenched with 15% trifluoroacetic acid. 200 µL of 1 M NaOH was added followed by the addition of 800 µL ethyl acetate. After vortex and centrifuge, 700 µL of top layer was then taken to dry down and combined with the samples for PPT I deficiency screening, and further analyzed with MS/MS. In each sample, varying amounts of product are produced during the incubation based on the level of activity of the TPP I/PPT I inside. Therefore, activity of both TPP I and PPT I in each sample was calculated based on the ratio of intensity of the fragment ion of IS to the intensity of the fragment ion of product during MS/MS.

5.2.7 Mass Spectrometry

ESI-MS/MS was carried out on a Waters Quattro Micro tandem quadruple (quadrupole-hexapole-quadrupole) instrument using a positive ionization mode and selected reaction monitoring (SRM). Sample injection was 10 µL for each analysis and flow injection was 0.2 mL/min of acetonitrile/water (V/V =80/20) with 0.1% Formic Acid solution. The mass transitions monitored were selected based on the peptide fragmentation pattern, e.g m/z 317.2→ m/z 209.4 and m/z 308.2→ m/z 208.4 for TPP I-Internal Standard (TPP I-IS) and TPP I-Product (TPP I-P) respectively as shown in **Figure 5.1**.

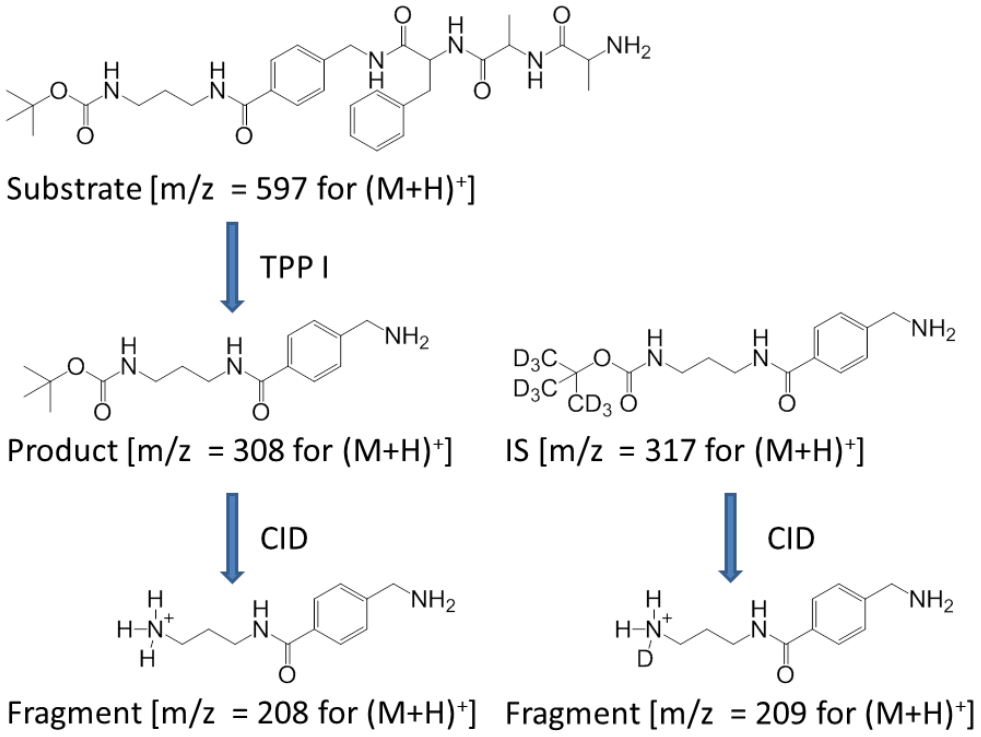


Figure 5.1 MS/MS Fragmentation pattern of the TPP I assay product and IS.

Mass spectrometer settings were: Capillary voltage 3.5 kV; Cone 15 V; Extractor 4 V; RF lens, 0.3 V; Source temperature, 130ºC; Desolvation temperature, 250ºC; Cone gas flow, 25 L/h; Desolvation gas flow, 500 L/h; LM 1 resolution, 10; HM resolution, 14.5; Ion energy 2, 2eV; Entrance, 2 V; Collision, 10 eV; Exit, 2 V; LM 2 resolution, 15; HM 2 resolution, 15; Multiplier, 650 V; Gas cell pirani pressure, 2.13 e-3 mbar; dwell time 100 milisecond, Inter delay, 0.1.

5.3 Results and Discussion

5.3.1 Assay principle for screening of TPP I deficiency

Since it is known that the human TPP I is a lysosomal serine protease that functions in the acidic milieu of the lysosome to specifically remove Ala-Ala-Phe at the *N*-terminal; theoretically, TPP I enzyme in the blood of newborns also removes Ala-Ala-Phe from the substrate which contains this tripeptide in the end to form TPP I product. When a sufficient amount of substrate (*m/z* = 597) is added into the DBS of newborns, healthy newborns with efficient TPP I in their blood will have quantitative amount of TPP I

product (*m/z* = 308), while TPP I deficient newborns (NCL II patients) have barely any amount of TPP I product. TPP I IS (*m/z* = 317), which is chemically identical but mass different to TPP I product, has been prepared and put into the assay solution as well for the calculation of TPP I activity in DBS. **Figure 5.1** shows how the TPP I activity is specifically quantified by single reaction monitoring (SRM) in the triple quadrupole mass spectrometer. TPP I Product ion (*m/z* = 308) is selected in the first quadrupole and further dissociated in second quardupole by collisional induced dissociation (CID). After dissociation, the most abundant fragment ion (*m/z* = 208) is selected and monitored in the third quadrupole. The peak intensity of the specific fragment ion (m/z = 208) was measured quantitatively this way. Meanwhile, the peak intensity of fragment ion (m/z = 209) of IS was also measured and used as the standard. TPP I internal standard ion (*m/z* = 317) is selected in the first quadrupole and dissociated in second quardupole by CID. After dissociation, the most abundant fragment ion (*m/z* = 209) is mass filtered to be monitored in the third quadrupole.

5.3.2 Response ratio of the TPP I-Product to the TPP I-Internal Standard

The amount of product formed from the substrate was calculated using the SRM intensity ratio of the product to the internal standard, the known concentration of the internal standard and the response ratio. The response ratio used for the calculation was obtained from the result shown in **Figure 5.2**. TPP I Product and TPP I internal standard are both synthesized, and known amount of them are used for the experiment of calculating the response ratio. As shown is **Figure 5.2**, the x-axis value of each data point shows the actual ratio of Product/Internal standard, and the y-axis value of each data point is the ratio of Product/Internal standard that is measured on “triple quadrupole” (quadrupole-hexapole-quadrupole) mass spectrometer. It is found in **Figure 5.2** that the response ratio follows the linear trendline and shows good linearity (R^2 = 0.9987).

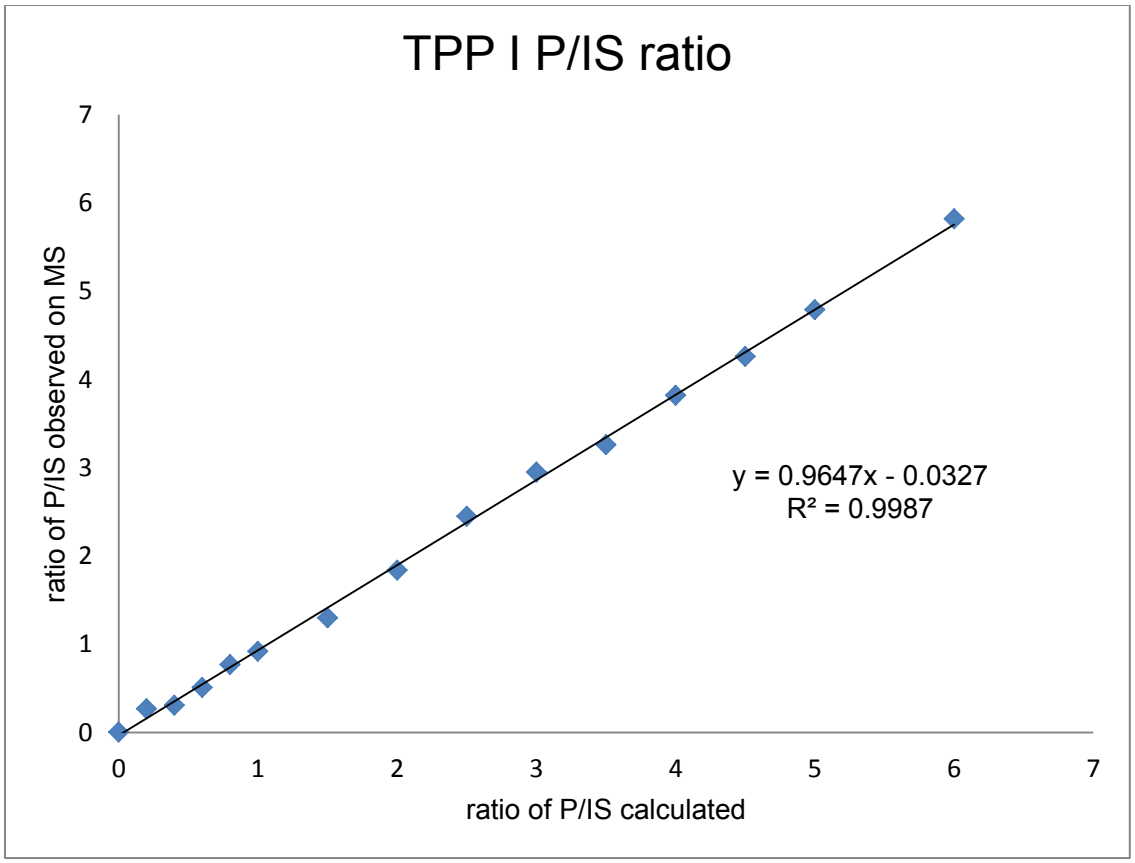


Figure 5.2 Response ratio of the TPP I-Product to the TPP I-Internal Standard

5.3.3 pH study of the TPP I assay

The value of pH to use is crucial in the assay development since the enzyme is sensitive and can be fully deactivated in an inappropriate environment. Since it is known that TPP I functions in an acidic environment, pH dependence of the TPP I assay was examined in detail from pH 3 to 5. As shown in **Figure 5.3**, the optimum pH of the TPP I assay buffer was found to be 4.0. However, the optimum pH of PPT I is 7.0, and has barely any activity in the acidic situation, therefore, the TPP I assay and PPT I assay were incubated separately in parallel instead of being mixed for the duplex assay.

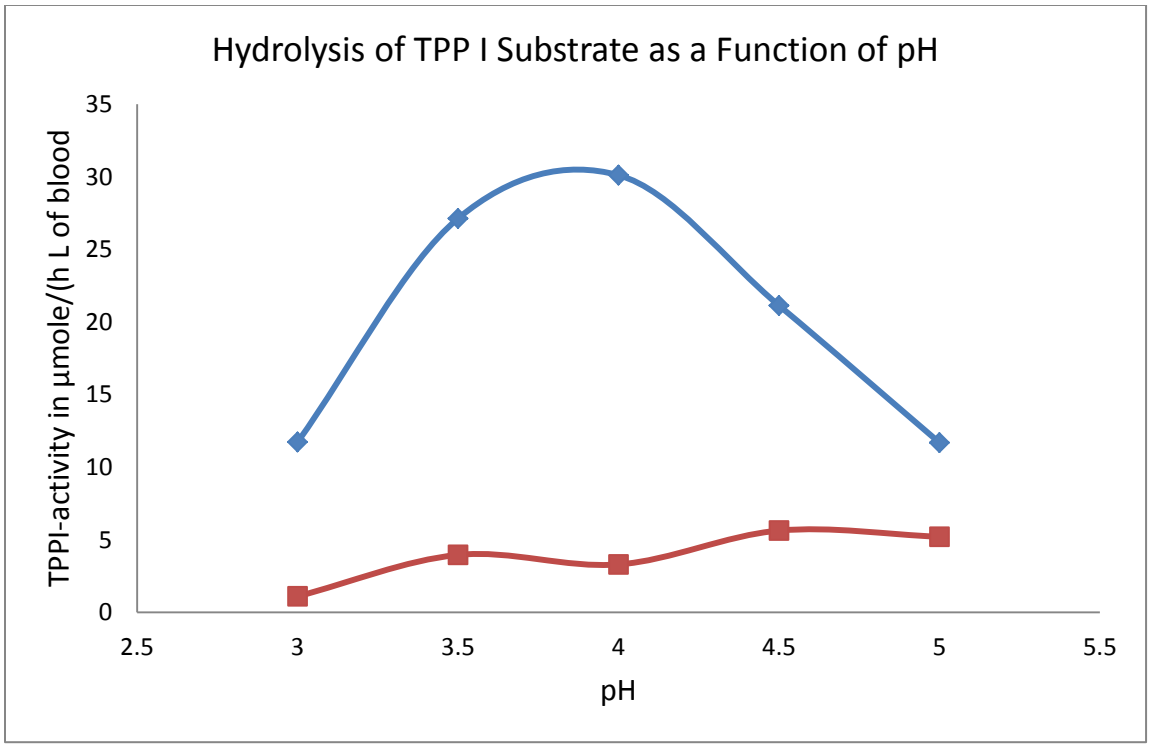


Figure 5.3 pH dependence of the TPPI enzyme activity (blue line: blood assay; red line: blank).

5.3.4 Impact of incubation time to the activity of TPP I

Enzymatic Activity was calculated as µmol/(h L of Blood), based on the amount of product formed (µmol), the incubation time (h) and the volume of blood (3.2 µL of blood, which is calculated from the estimated volume of a blood spot (10 µL) and the punch/DBS area ratio). Enzyme kinetics was studied for the TPP I assay as illustrated in **Figure 5.4**. An incubation time of 16 hours was selected for the study of this single assay, and an incubation time of 10 hours was used for the duplex assay to be compatible with PPT I.

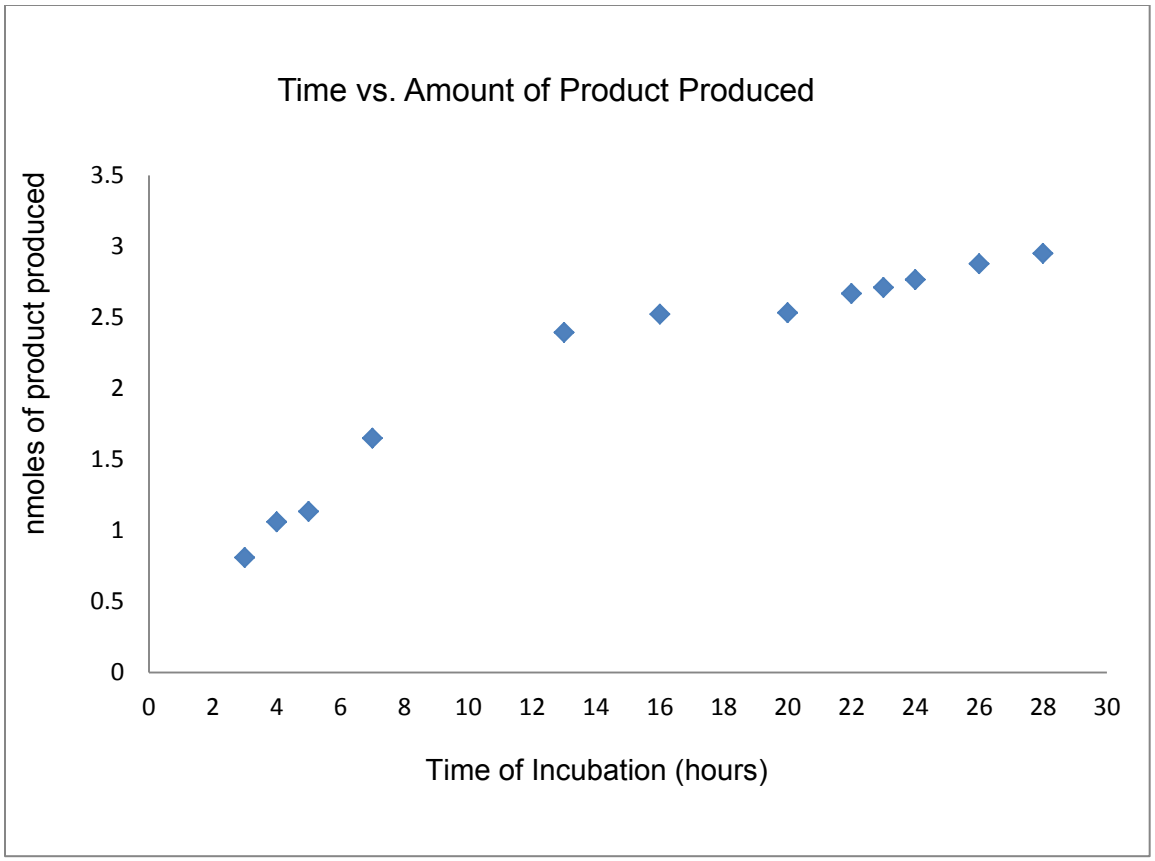


Figure 5.4 TPP I enzyme activity as a function of the incubation time.

5.3.5 Michaelis-Menten kinetics study for TPP I assay

Michaelis-Menten parameters were also evaluated for the TPP I enzyme and it was found that K_M = 53.3 mM and V_{MAX} = 45.99 µmol/h*L. A concentration of 400 µM of the substrate was used in an attempt to minimize the effect of inhibitors present in blood.

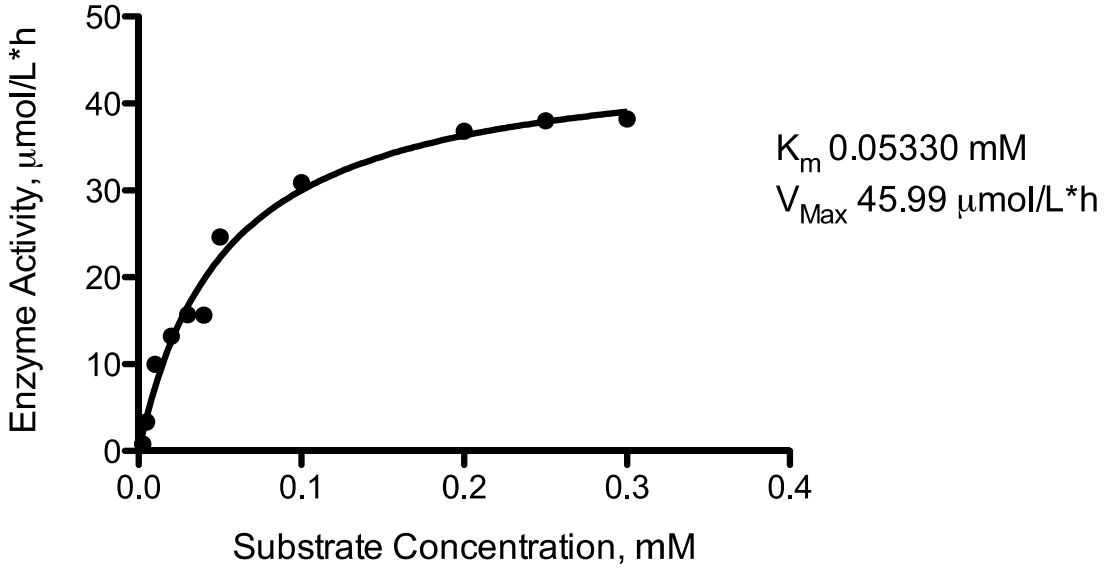


Figure 5.5 Dependence of enzymatic activity on substrate concentration.

5.3.6 Dwell time study for TPP I assay

The result of dwell time study shows that when dwell time is changed from 0.01 s to 0.1 s, the coefficient of variance changes from 3.8% to 1.4%. It is clear that the modification of the dwell time from 0.01 s to 0.1 s effectively decreases coefficient of variance, increasing the accuracy of data collected.

5.3.7 TPP I Assay for random newborns

After optimizing the parameters, the TPP I assay was performed on DBS samples of 54 random newborns, and 10 diagnosed TPP I enzyme deficiency (NCL II) patients (**Figure 5.6**) collected from the Washington state lab. Unaffected newborns showed a range of activity from 33.8 to 86.5 μmol/(h L of blood) and an average of 58.4 μmol/(h L of blood) after a background correction with blank samples; NCL II patients displayed a range between 0.02 to 0.83 μmol/(h L of blood) and an average of 0.37 μmol/(h L of blood) after a blank correction. Blanks were carried out combining all the components of the assay but replacing the DBS with a filter paper punch. The results show a clear distinction between NCL II affected patients and healthy newborns, which demonstrates the efficiency of the assay.

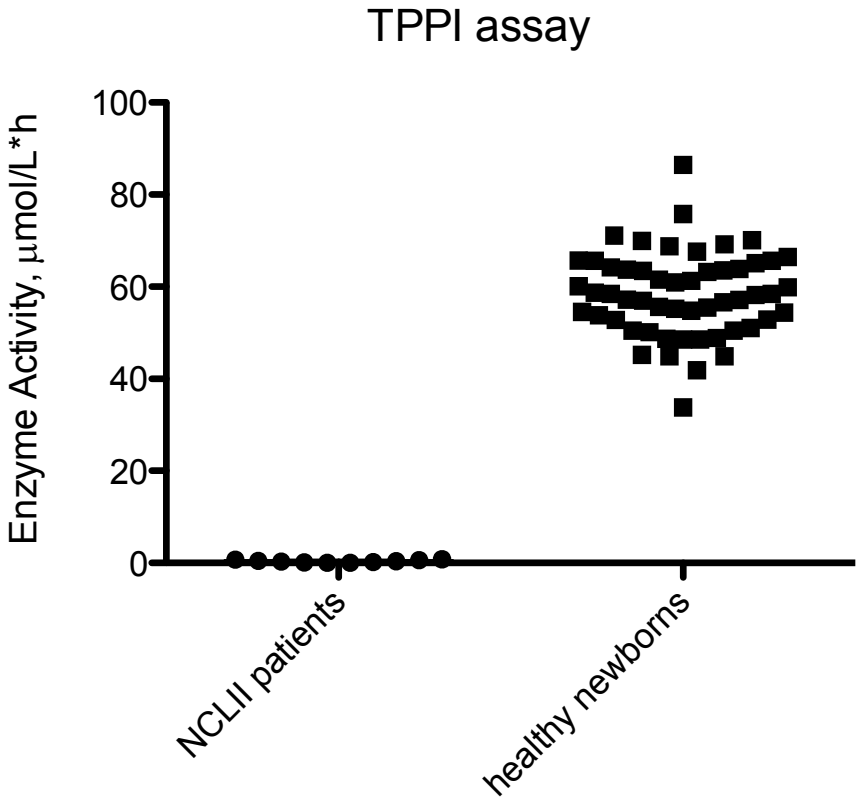


Figure 5.6 TPP I enzyme activity of random newborns and NCL-II patients.

TPP I assay precision was calculated using a DBS from a control sample. Intraassay variation coefficient of variation (CV) was 1.4% (n = 5) involving five injections of the sample from the incubation of a single blood spot. Interassay variation CV was 8.74% and involved 15 injections from different blood spots.

5.3.8 Duplex assay for detection of TPP I and PPT I deficiency in random newborns

Figure 5.7, below, shows the result of the duplex assay, which contains healthy newborns, NCL II patients (TPP I deficiency) and NCL I patients (PPT I deficiency). The blank has an average of 2.51 µmol/(h L of blood), the NCL II patients displayed a mean activity of 2.72 µmol/(h L of blood). Healthy newborns has the mean of 65.7 µmol/(h L of blood), which is 24 times as high as those of NCL II patients.

NCL I patients has the mean of 26.8 μmol/(h L of blood), which is at the low end of but still within the normal range.

The duplex assay was developed by combining TPP I assay and PPT I assay in a single mass spectrometry injection. Due to different optimal pH for the two enzymes, the assays cannot be incubated together; therefore they are combined after the sample workup into a single mass spectrometry injection. The results of TPP I activity in the duplex assay are shown in **Figure 5.7**. 40 Random healthy newborns samples had an activity range of 42.2 to 109.0 μmol/(h L of blood) with a mean activity of 65.7 μmol/(h L of blood), 5 NCL I (PPT I deficiency) affected newborns had a range of activity of 23.6 to 29.8 μmol/(h L of blood)and a mean activity of 26.8 μmol/(h L of blood), which is at the low end of the normal range. TPP I activity of 5 NCL II (TPP I deficiency) affected newborns ranged between 2.59 and 2.86 μmol/(h L of blood) with an average activity of 2.72 μmol/(h L of blood). Blanks of the assay were once again evaluated and the range of activity was 2.41 to 2.71 μmol/(h L of blood) with an average activity of 2.51 μmol/(h L of blood). Clearly, the duplex assay works efficiently for the detection and diagnosis of TPP I deficiency. Healthy newborns has the mean of 65.7 μmol/(h L of blood), which is 24 times as high as those of NCL II patients.

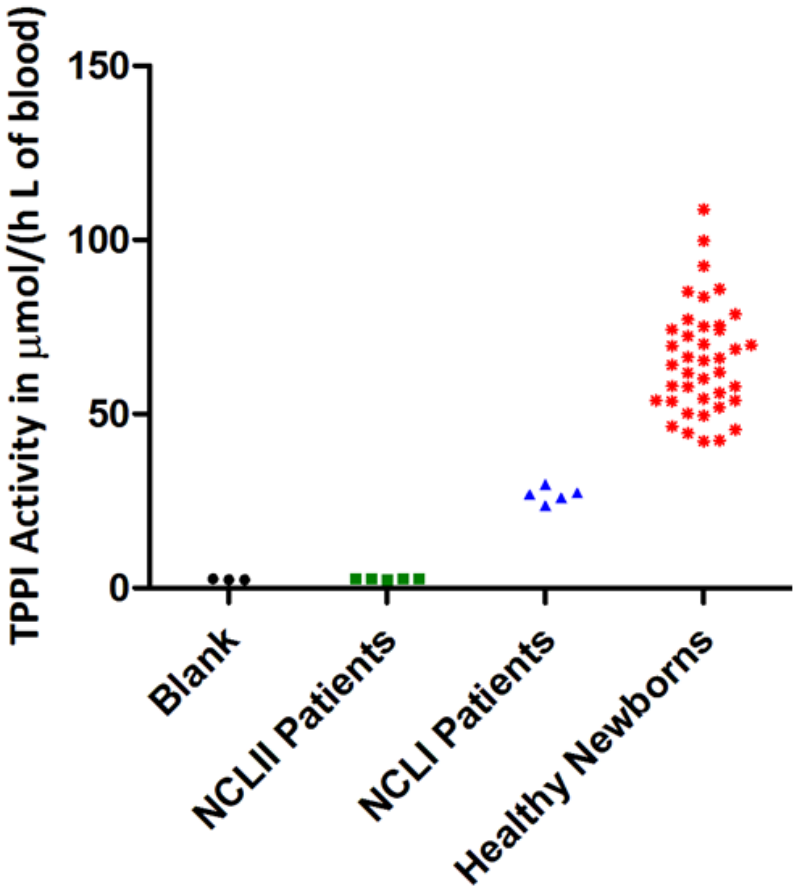


Figure 5.7 TPP I Enzyme activity of random newborns, NCL I (PPT I deficiency) and NCL II (TPP I deficiency) affected patients.

5.4 Conclusion

In order to conquer the limitation of current fluorescence-based screening method for NCL I (PPT I deficiency) and NCL II (TPP I deficiency), a duplex tandem mass spectrometry assay method with a single injection is developed for clinical diagnosis of both diseases in DBS from newborns. Due to different pH conditions for the two enzymes, two assay cocktails were incubated for 10 hours in parallel and were combined into a single injection onto a tandem quadrupole mass spectrometer after the sample workup. A set of mass-orthogonal enzymatic products and internal standards are used for SRM and quantitation of two enzyme activities. DBS samples from 40 healthy patients, 5 TPP I deficiency patients and 5 PPT I deficiency patients were tested and the efficiency were demonstrated.

5.5 References

[1] Goebel, H. H.; Mole, S. E.; Lake, B. D. (eds) *The Neuronal Ceroid Lipofuscinoses (Batten Disease): Biomedical and Health Research*, IOS Press, Amsterdam, **1999**

[2] Wisniewski, K. E.; Kida, E.; Golabek, A. A.; Kacemarski, W.; Connell, F.; Zhong, N. Neuronal Ceroid Lipofuscinoses: Classificaion and Diagnosis. *Adv. Genet.* **2001,** *45,* 1-34.

[3] Haltia, M. The Neurocal Ceroid-lipofuscinoses. *J. Neuropathol. Exp. Neurol*. **2003**, *62*, 1-13.

[4] Williams, R. E.; Gottlob, I.; Lake, B. D.; Goebel, H. H.; Winchester, B. G.; Wheeler, R. B. CLN2: Classic Late Infantile NCL. In *the Neurocal Ceroid lipofuscinoses (Battern Disease),*H. H. Goebel, ed. (Amsterdam: IOS Press), pp. 37-53.

[5] Jalanko, A.; Braulke, T. Neuronal Ceroid Lipofuscinosis. *Biochimica et Biophysica Acta, Molecular Cell Research.* **2009**, *1793*, 697-709.

[6] Pal, A.; Kraetzner, R.; Gruene, T.; Grapp, M.; Schreiber, K.; Grønborg, M.; Urlaub, H.; Becker, S.; Asif, A. R.; Gärtner, J.; Sheldrick, G. M.; Steinfeld, R. Structure of Tripeptidyl-peptidase I Provides Insight into the Molecular Basis of Late Infantile Neuronal Ceroid Lipofuscinosis*. J. Biol. Chem.* **2009**, *284*, 3976-3984.

[7] Vines, D.; Warburton, M. J. Purification and Characterisation of a Tripeptidyl Aminopeptidase I from Rat Spleen. *Biochim. Biophys. Acta* **1998,** *1384*, 233-242.

[8] Page, A. E.; Fuller, K.; Chambers, T. J.; Warburton, M. J. Purification and Characterization of a Tripeptidyl Peptidase I from Human Osteoclastomas: Evidence for its Role in Bone Resorption. *Arch. Biochem. Biophys*. **1993***, 306*, 354-359.

[9] Ivanov, I.; Tasheva, D.; Todorova, R.; Dimitrova, M. Synthesis and use of 4-peptidylhydrazido-N-hexyl-1,8-naphthalimides as fluorogenic histochemical substrates for dipeptidyl peptidase IV and tripeptidyl peptidase I. *Eur. J. Med. Chem*. **2009**, *44*, 384-392.

[10] Sohar, I.; Lin, L.; Lobel, P. Enzyme-based diagnosis of classical late infantile neuronal ceroid lipofuscinosis: comparison of tripeptidyl peptidase I and pepstatin-insensitive protease assays. *Clin. Chem*. **2000**, *46*, 1005-1008.

[11] Tian, Y.; Sohar, I.; Taylor, J. W.; Lobel, P. Determination of the Substrate Specificity of Tripeptidyl-peptidase I Using Combinatorial Peptide Libraries and Development of Improved Fluorogenic Substrates. *J. Biol. Chem.* **2006**, *281*, 6559-6572

[12] Lukacs, Z.; Santavuori, P.; Keil, A.; Steinfeld, R.; Kohlschütter, A. Rapid and Simple Assay for the Determination of Tripeptidyl Peptidase and Palmitoyl Protein Thioesterase Activities in Dried Blood Spots. *Clin. Chem*. **2003**, *49*, 509-511.

[13] Wolfe, B. J.; Blanchard, S.; Sadilek, M.; Scott, C. R.; Tureček, F.; Gelb, M. H. Tandem Mass Spectrometry for the Direct Assay of Lysosomal Enzymes in Dried Blood Spots: Application to Screening Newborns for Mucopolysaccharidosis II (Hunter Syndrome) *Anal. Chem*. **2011**, *83*, 1152-1156.

[14] Pitterkow, M.; Lewinsky R.; Christensen, Selective Synthesis of Carbamate Protected Polyamines Using Alkyl Phenyl Carbonates, J. B. *Synthesis*, **2002**, *15*, 2195-2202.

[15] Szabó, P.T.; Kele, Z. Electrospray Mass Spectrometry of Hydrophobic Compounds Using Dimethyl Sulfoxide and Dimethylformamide as Solvents. *Rapid Commun Mass Spectrom.* **2001**, *15*, 2415-2419.

[16] Blanchard, S.; Sadilek, M.; Scott, C.R.; Tureček, F.; Gelb, M.H. Tandem Mass Spectrometry for the Direct Assay of Lysosomal Enzymes in Dried Blood Spots: Application to Screening Newborns for Mucopolysaccharidosis I. *Clin Chem.* **2008**, *54*, 2067-2070.

5.6 Acknowledgement

Thanks to Tatyana for the development of the synthesis route of the substrate and internal standard, and the assay method for TPP I. Thanks to Mariana for the assay development of PPT I and working together with me to develop the duplex assay method for both TPP I and PPT I.

Chapter 6

Synthesis and Application of a Low MW, High Solubility Protein Interaction Reporter —BDD-PIR

6.1 Introduction

A new cross-linker Bis(2,5-dioxopyrrolidin-1-yl)4,4,16,16-tetramethyl-6,14-dioxo-10-(2-(5-(2-oxohexahydro-*1H*-thieno[3,4-d]imidazol-4-yl)pentanamido)ethyl)-5,15-dioxa-7,10,13-triazanonadecane-1,19-dioate PIR (BDD-PIR, **Figure 6.1**) has been developed that will further improve on the advantages of PIR technology and provide additional benefits with its unique structure.

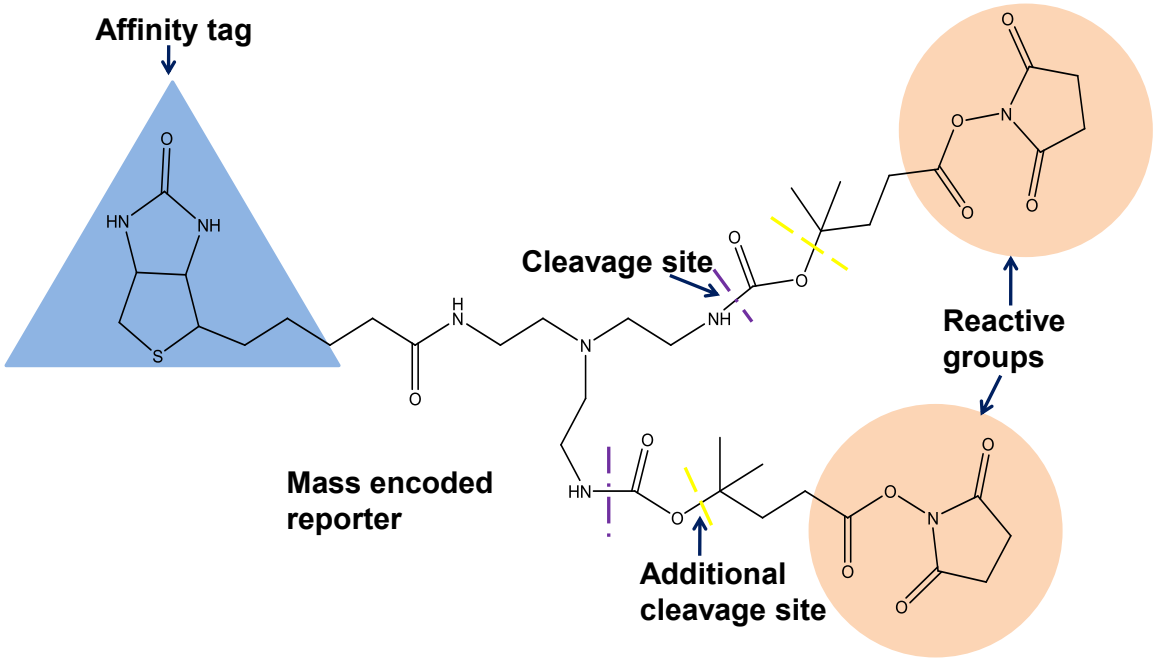


Figure 6.1 Chemical structure of BDD-PIR.

We chose a tertiary amine building block rather than a tertiary carbon as the cross-linker core to eliminate the chiral center and the resultant undesirable effects on chromatographic separation of cross-linked peptides due to diastereomeric mixtures. Also, the tertiary amine within BDD-PIR molecule can be protonated at neutral pH, thus increasing water solubility and improving labeling efficiency that benefits *in vivo* applications which are performed in a water-based environment. Furthermore, the molecular weight of BDD-PIR is also significantly lower than previous PIR molecules, further improving solubility and likely enhancing cell penetration, which will also enable improved *in vivo* applications. Finally, the BDD-PIR reporter ion is produced predominantly in the 1+ charge state with additional collisional energy and detected as m/z = 373.24, which is large enough to avoid the low mass cut-off of most ion trap mass spectrometers, thus making it applicable to commonly used instruments.

A critical feature of the BDD-PIR is its lack of chiral centers and its symmetry between cross-linker arms. The latter is important since it means that all released peptides will be produced with the same, known mass modification, which enables a more efficient database search. The lack of chiral centers means that all identical cross-linked peptides will elute from the HPLC column as a single peak, making it easier for the protein interaction to be detected. Compared to previous PIR molecules, this could account for as much as a 4-fold improvement in observed peak heights for identical cross-linked peptide amounts. In summary, the design of BDD-PIR offers high aqueous solubility, low MW, improved chromatographic properties, and much greater sensitivity. Here we describe the solution phase synthesis of this novel PIR molecule and present our initial applications of this molecule with model peptides and protein standards.

6.2 Experimental Section

6.2.1 Materials

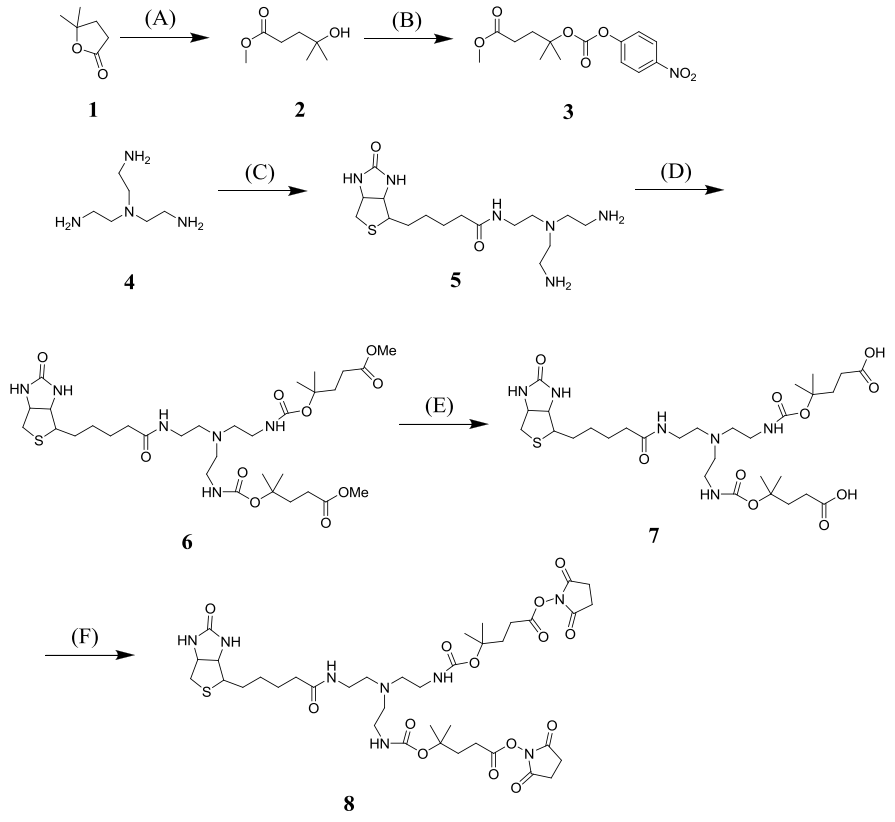
Starting materials were purchased from Aldrich, ACROS, EMD, Fisher or VWR and used without further purification. 5,5'-Dimethyl-dihydro-furan-2-one was purchased from Astatech (Bristol, PA). Unless otherwise indicated, all anhydrous solvents were purchased and used as obtained. Anhydrous reactions were performed under a dried nitrogen atmosphere in oven-dried glassware. Reactions were monitored by thin layer chromatography (TLC) using silica gel 60 F-254 (0.25 mm) plates with detection by UV light.

^{1}H NMR spectra were recorded in diluted $CDCl_3$ solutions at 300 MHz. Chemical shifts are reported in parts per million (δ) downfield from tetramethylsilane (TMS). Coupling constants (J) are reported in Hz. Electrospray ionization mass spectra were acquired either on a Bruker Esquire LC00066 mass spectrometer or on a Thermo LTQ mass spectrometer. Flash chromatography was carried out on silica gel (40-63 μm).

Reverse phase HPLC was performed on an automated Waters Alliance HPLC system using a Partisil 10 ODS-3 column (9.4 mm × 250 mm, Whatman Inc.).

6.2.2 Synthesis of BDD-PIR

BSS-PIR has been prepared in a 7-step convergent synthesis with 5 linear steps as shown in **Scheme 6.1**.



Scheme 6.1 Reaction scheme of BDD-PIR. Reagents and conditions: (A) (i) NaOH, EtOH/H_2O, 2 h, reflux; (ii) citric acid, H_2O; (iii) TMSD, ether, 2 h, 0 ^oC, 45% (B) *p*-nitrophenolchloroformate, anhydrous pyridine, anhydrous CH_2Cl_2, 0 ^oC-RT, overnight, 62% (C) NHS-Biotin, anhydrous DMF, 2 h, RT, 84% (D) anhydrous DMF, *i*-Pr_2NEt, compound **3**, RT, 2 h, 52% (E) (i)1 M NaOH, H_2O, RT, 4 h; (ii) 0.1 M HCl, 100% (F) TFA-NHS, anhydrous pyridine, RT, 1 h, 25%

Methyl 4-hydroxy-4-methylpentanoate (Compound 2). A solution of compound 5,5'-dimethyl-dihydro-furan-2-one [1] (**1**, 0.500 g, 4.386 mmol) and NaOH (0.209 g, 5.225 mmol) in ethanol/water (v/v = 2/1, 6.63 mL) was heated under reflux for 2 h. Solvents were removed on a rotary evaporator. The dried product was redissolved in water (1.27 mL) at 0 °C. The solution was acidified to pH 4-5 by addition of citric acid (ca. 0.158 g) at 0 °C. Excess NaCl was added to give a saturated solution, and the mixture was extracted with diethyl ether (four times). The combined organic layers were dried over Na_2SO_4 and concentrated to give the unstable 4-hydroxy-4-methylpentanoic acid, which was immediately reacted with (trimethylsilyl)diazomethane solution (2.0 M in diethyl ether) (4.386 mL) in ether at 0 °C for 2 h. The resulting mixture was concentrated and purified by flash chromatography eluting with 15% EtOAc in hexane to afford compound **2** (0.290 g, 45% yield): ^{1}H NMR 300 MHz ($CDCl_3$) δ 3.67 (3H, s), 2.43 (2H, t, J = 7.8 Hz), 1.81 (2H, t, J = 7.8 Hz), 1.22 (6H, s). MS *m/z* 169.1 $(M+Na)^+$.

Methyl 4-methyl-4-((4-nitrophenoxy)carbonyloxy)pentanoate (Compound 3). Compound **2** (0.357 g, 2.45 mmol) was dissolved in 15 mL of anhydrous CH_2Cl_2. Dry pyridine (0.590 mL, 7.32 mmol) was added, and the reaction mixture was cooled to 0 °C [2]. In a separate flask *p*-nitrophenylchloroformate (0.616 g, 3.05 mmol) was dissolved in dry CH_2Cl_2 (2.2 mL), and this solution was slowly added under nitrogen atmosphere. The reaction mixture was allowed to stir overnight under N_2 while warming up to room temperature. The crude reaction mixture was washed three times with NaH_2PO_4 (0.4 M, pH 4.6), the organic layer was dried over anhydrous Na_2SO_4 and the solvent was removed on a rotary evaporator. The residue was purified by flash chromatography eluting with 6% EtOAc to afford compound **3** (0.469g, 62% yield): ^{1}H NMR 300 MHz ($CDCl_3$) δ 8.289 (2H, d, J = 9.1 Hz), 7.388 (2H, d, J = 9.1 Hz), 3.72 (3H, s), 2.50 (2H, t, J = 7.9 Hz), 2.21 (2H, t, J = 7.9 Hz), 1.60 (6H, s). MS *m/z* 334.0 $(M+Na)^+$.

Mono-biotinylated tris(2-aminoethyl)amine (Compound 5). Compound **4** Tris(2-aminoethyl)amine (0.4400 mL, 29.40 mmol) was diluted with anhydrous DMF (2 mL). A solution of NHS-Biotin (0.1000 g, 2.93 mmol) in anhydrous DMF (2 mL) was slowly added and the reaction mixture was stirred at RT for 2 h. The DMF was removed in a centrifugal concentrator (Speed-Vac), and the crude mixture was acidified to pH = 4 with addition of 1% TFA and then purified using reverse phase HPLC with

detection at 245 nm with a gradient of 5% MeOH to 30% MeOH (with 0.08% trifluoroacetic acid) at 1 mL/min over 30 min, and compound **5** (0.0920 g, 84%) eluted at 16% MeOH. MS *m/z* 373.3 $(M+H)^+$.

Compound 6. Compound **5** (0.0400 g, 0.1072 mmol, 1 equiv.) was dissolved in anhydrous DMF (1.03 mL) and *N,N*-diisopropylethylamine (0.1123 mL, 0.6431 mmol, 6 equiv.) was added. After stirring for a few minutes, a solution of compound **3** (0.1333 g, 0.4288 mmol, 4 equiv.) in anhydrous DMF (1.03 mL) was added.[3] The reaction mixture was stirred for 2 h at RT. The crude mixture was brought to dryness in a Speed-Vac then purified with reverse phase HPLC with detection at 230 nm with a gradient of 20% MeOH to 45% MeOH (with 0.08% trifluoroacetic acid) at 1 mL/min over 40 min then 45% MeOH to 70% MeOH (with 0.08% trifluoroacetic acid) at 1 mL/min over 20 min, and compound **6** (0.0403 g, 52% yield) eluted at 63% MeOH. MS *m/z* 717.5 $(M+H)^+$.

Dicarboxylic acid (Compound 7). The methyl ester compound **6** (0.0248g, 0.0346 mmol) was hydrolyzed to the free acid by treatment with 1 M NaOH (3 mL) in water at RT for 4 h.[4] The crude mixture was acidified to pH = 5-6 by the addition of 0.1 M HCl, and then concentrated to dryness in a Speed-Vac. The crude mixture was further desalted with a C18Sep-Pak cartridge (Waters, Milford, MA) and concentrated to dryness in a Speed-Vac to obtain the purified compound **7** (0.0239 g, 100% yield). MS *m/z* 689.6 $(M+H)^+$.

Compound 8 (BDD-PIR) (Figure 6.1). Trifluoroacetic anhydride (2.08 mL, 0.015 mol, 1.5 equiv.) was added dropwise to a solution of *N*-hydroxysuccinimide (NHS) (1.1509 g, 0.010 mol, 1.0 equiv.) in anhydrous tetrahydrofuran (THF) (10 mL).[5] The reaction mixture was stirred overnight at RT, and was concentrated to dryness in a Speed-Vac to obtain quantitative TFA-NHS, which was stored at -80 ºC prior to use. The free acid compound **7** (0.0400 g, 0.0581 mmol, 1 equiv) was suspended in anhydrous pyridine (2 mL) to which TFA-NHS (0.2000 g, 0.9479 mmol, 16 equiv) was added at RT. The reaction mixture in a 5 mL round bottom flask was stirred for 1 h and then the solvent was removed on a rotary evaporator. The crude mixture was purified with reverse phase HPLC with detection at 235 nm with a gradient of 20% MeOH to 45% MeOH (with 0.08% trifluoroacetic acid) at 1 mL/min over 40 min then 45% MeOH to 70% MeOH (with 0.08% trifluoroacetic acid) at 1 mL/min over 20 min, and compound **8** (0.0130 g, 25.4%) eluted at 50% MeOH. MS *m/z* 883.4 $(M+H)^+$.

6.2.3 Cross-Linking of Standard Peptides and Analysis

An aliquot (1 μL) of 10 mM Bradykinin (10 nmol) was mixed with 1 μL of 10 mM Angiotensin II (10 nmol) in water and dried in a Sped-Vac overnight.[6] An aliquot (1 μL) of 10 mM BDD-PIR in dimethyl sulfoxide (DMSO) (10 nmol) was added into the dried peptide mixture, and then was stirred in the shaker with 1000 rpm at RT overnight.

An electrospray-ion trap LTQ-XL (Thermo Fisher Scientific Inc., Waltham, MA) mass spectrometer was used for tandem mass spectrometry (MS/MS) analysis of cross-linked peptides. Multiple collision-induced dissociation (CID) energies were used to cleave the observed cross-linked peptides to test the characteristic of the cleavage efficiency of BDD-PIR. A LTQ-FT Ultra hybrid Fourier-transform mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA) coupled with an Ultra Performance Liquid Chromatography (UPLC, Waters nanoAcquity, Milford, MA) was used to test the cleavage performance of the cross-linked peptides under in-source collision-induced dissociation (ISCID) mode.

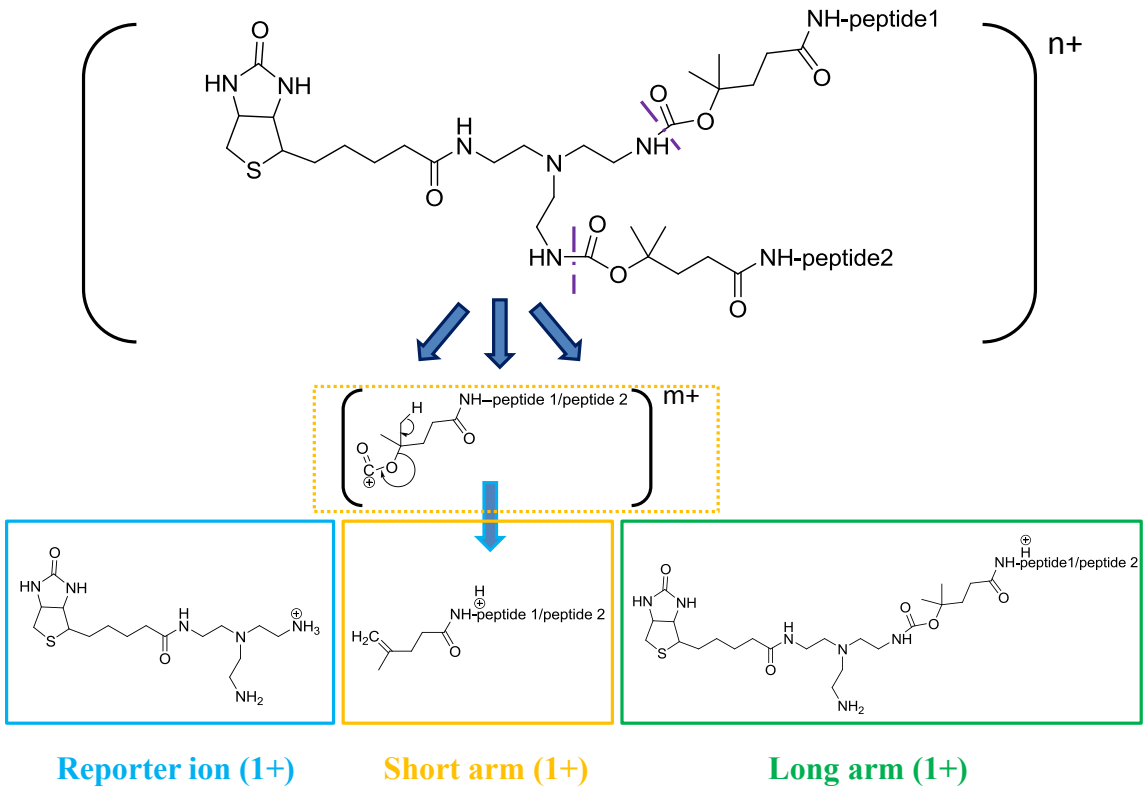
6.2.4 Cross-linking of proteins using bovine serum albumin (BSA) as a model system

The BDD-PIR cross-linked BSA sample was prepared using the modified protocol based on that reported by Rinner and Aebersold.[7] BSA (180 μL, 1 mg/mL) in phosphate buffered saline (PBS) and 20 μL BDD-PIR (10 mM) in DMSO were mixed together, and the sample was shaken at 35 °C for 30 min. The sample was then reduced with tris(2-carboxyethyl) phosphine (TCEP, 1 μmol), alkylated with iodoacetamide (IAA, 2 μmol) and digested with trypsin (2 μg) overnight at 37 °C. The sample was then acidified to pH = 3 (pH paper) with the addition of formic acid and brought to dryness in a Speed-Vac. A C18Sep-Pak cartridge (Waters, Milford, MA) was used to purify the protein digest. The cross-linked products were then fractionated with UPLC and detected with a high mass accuracy instrument—LTQ-FT mass spectrometer. A 32-cm long C18 column was prepared in-house by packing a fused silica capillary (360 μm × 75 μm) with MAGIC C18AQ 100Å 5U beads (Michrom Bioresources, Inc., Auburn, CA). A 2-cm long trap column was similarly made by packing a fused silica capillary (360 μm × 100 μm) with MAGIC C18AQ 200Å 5U beads. The following 3-step LC gradient program was employed in LC runs: 1) Peptide elution from 5% to 50% solvent B for 90 minutes, 2) Column wash at 80% solvent B for 20

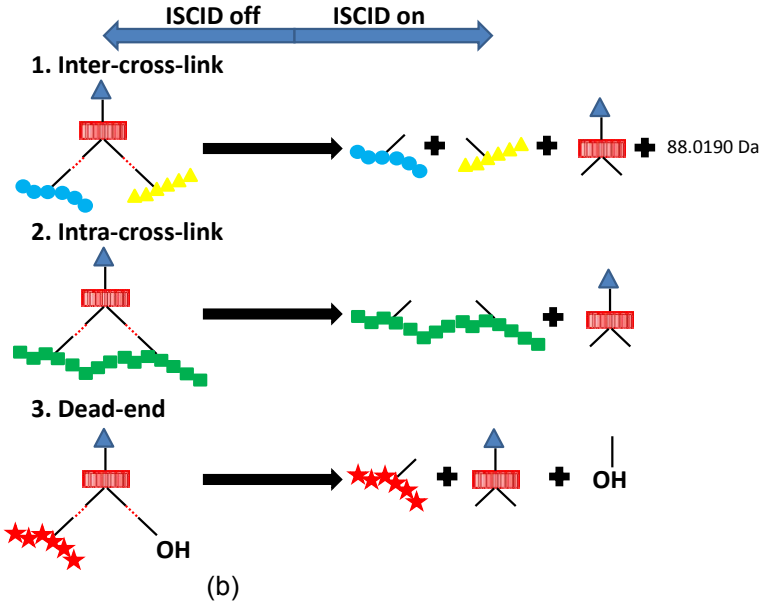
minutes, and 3) Column re-equilibration at 5% solvent B for 25 minutes (solvent A: 0.1% formic acid in DI water; solvent B: 0.1% formic acid in acetonitrile). The MS instrument resolution was set at 25000 to obtain high mass accuracy data. Three LC-MS or LC-MS/MS runs were conducted to get accurate and detailed cross-linking pair information as described below:

a. LC-MS identification of BDD-PIR modified peptides

The LC-MS run was performed for identification of inter-cross-links. Each MS scan cycle contains two full MS scans in the settings. The parameters in the settings are identical for the two full MS scans in one cycle with the exception that ISCID voltage is on in one scan and is off in the other scan. When the ISCID voltage is off, the inter-cross-link is intact and the parent ion can be detected, and when the ISCID voltage is on, the inter-cross-link is released as either two short arm peptides plus a reporter ion, or one long arm peptide and one short arm peptide (**Figure 6.2 (a)**) which are then detectable in this scan. Therefore, ISCID voltage is switched between on and off in adjacent scans to enable the discovery of the inter-cross-link relationship between the parent ions and their released peptides. The ISCID voltage was modified to 75 V at the scan event with ISCID switched on. The software tool Blinks [8] was used to identify the PIR relationships. For pairs identified by Blinks as inter-cross-link pairs which need further confirmation, the *m/z* value and the LC elution time of the released peptides and those of parent ions were used to generate two mass-and-time targeted inclusion lists for released peptides and parent ions respectively. The lists were then used in the following two LC-MS/MS runs correspondingly to perform one run for short/long arm identification and the other run for validation of the parent ions with two separate inclusion lists mentioned above on the same column. The run for short/long arm identification is used to acquire the sequence information of two arms, and the run for parent ion validation is used to acquire the sequence information of parent ions (**Figure 6.1, 6.2**).



(a)



(b)

Figure 6.2 (a) Cleavage of inter-peptide cross-linked species. (b) Mathematical relationships between cross-linked parent ion and released peptides: inter-cross-links = released peptide A + released peptide B + reporter; intra-cross-links = doubly modified released peptide + reporter; dead-end = released peptide + reporter + OH

b. LC-MS/MS identification of released peptides from inter-peptide cross-links

A mass-and-time targeted list of the released peptides, which was generated from Blinks identified (not verified yet) cross-linked pairs, was used in the following LC-MS/MS run for the identification of released peptides based on sequencing information. The MS scan cycle was set to include one full MS scan followed by five MS/MS scans with ISCID switched on in all six scans, which enabled the constant release of peptides from cross-links to be targeted so as to acquire sequencing information. Dynamic exclusion repeat and exclusion duration were both set to 15 seconds. The CID energy was set as 35 V. Mascot server (Version: 2.3.01) was used to identify released peptides for the data processing.

c. LC-MS/MS parent ion validation

The mass-and-time targeted list of the Blinks identified (not verified yet) cross-linked parent ions was used in the LC-MS/MS validation run with each MS scan cycle set to include one full MS scan followed by five MS/MS scans. The ISCID is switched off in all scan events. The dynamic exclusion repeat and exclusion duration were both set as 15 seconds and the CID energy was set as 20 V.

BDD-PIR contains a biotin group which allows for the affinity enrichment of cross-linked peptide pairs. After cross-linking of intact BSA, Amicon Ultra -0.5 mL 10K Da molecular weight cut off centrifugal filters (Millipore, Billerica, MA) were used (diluted to 0.5 mL with ammonium bicarbonate solution, spun at 14,000×g for 10 min, then reversed upside down in a new filter, spun at 1,000×g for another 2 min, repeat spin cycles 3 times) to remove excess unreacted cross-linker before trypsin digestion, then UltraLink monomeric avidin resin (Pierce, Rockford, IL) was used to perform the biotin affinity capture according to the manufacturer's protocol. The cross-linking product was then fractionated with UPLC and detected with a high mass accuracy instrument—LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA). The MS instrument resolution was set at 30000 and the LC gradient program was modified as below to acquire a better separation within a shorter time: 1) Peptide elution from 10% to 40% solvent B for 70 minutes, 2) Column wash with 80% solvent B for 15 minutes, and 3) Column re-

equilibration at 10% solvent B for 25 minutes (solvent A: 0.1% formic acid in DI water; solvent B: 0.1% formic acid in acetonitrile).

The first protocol mentioned above in the first paragraph of section *6.2.4* for cross-linking BSA with BDD-PIR was also applied to cross-link BSA with the commercialized DSS cross-linker (Pierce, Rockford, IL) as the control. Another BSA cross-linking experiment using a 10 fold concentrated DSS (100 mM) in DMSO was also performed to test the impact of an excess amount of DMSO to the cross-linking reaction. The data of both experiments were collected using the same instrument—UPLC coupled with LTQ-FT.

6.2.5 Data Analysis

The software tool Blinks [8] was used to identify the potential cross-link relationships. Relationship mass tolerance was set to 10 ppm, and the number of minimum data points was set to 5. Then the data from first LC-MS/MS run to identify released peptides were searched against the *Yeast* database spiked with BSA (6,765 sequences; 3,041,726 residues) using Mascot. Three missed protease cleavages were allowed as possibilities, the precursor error tolerance was set as 25 ppm, and the fragmentation error tolerance was 0.6 Da. The remaining tag from the BDD-PIR cross-linker (mass = 96.0581) was used as a variable modification on lysine residues and the protein *N*-terminus. Meanwhile, manual inspection of the mass spectra from the parent ion validation LC-MS/MS run was used to exclude any incorrect cross-link pairs.

The high mass accuracy LC-MS/MS data from the BSA cross-linking with DSS experiments were analyzed using the search engine *xQuest*.[7]

6.3 Results and Discussion

6.3.1 PIR concept and strategy

BDD-PIR is our 2nd generation of PIR cross-linker. It retains the core function of PIRs while having additional functionality based on its unique symmetrical structure and cleavage sites. A detailed review of the PIR concept was recently published by Tang et al.[9] Briefly, there are three types of cross-linked products: inter-, intra- and dead end labeled cross-links, and each can be differentiated via the

engineered mass relationships differences between the cross-linked products (parent ion) and released intact peptides. The incorporation of labile bonds in the PIR cross-linker is the key factor in this strategy. Two alternative scan types with ISCID on and off were used in the LC-MS run to discover the potential PIR relationships. With ISCID turned off, the accurate masses and charge states of all parent ions were recorded. With ISCID energy implemented, parent ions were cleaved in the mass spectrometer to release two individual peptide chains either in the form of short arm tagged peptide (the peptide plus the small tag, see the structure in the yellow rectangle of **Figure 6.2 (a)** for the case of BDD-PIR, C_6H_9O, exact mass = 97.0648 Da) or the long arm tagged peptide (the peptide plus the larger tag, see the structure in the green rectangle of **Figure 6.2 (a)** for the case of BDD-PIR, $C_{23}H_{40}N_6O_5S$, exact mass = 512.2781 Da). As the cleavage site in the PIR cross-linker is more labile than the peptide backbone, the parent ion is cleaved specifically without peptide backbone fragmentation. The mass relationships between the cross-linked products and the released peptide ions can be used to identify inter-peptide cross-linked products using the mass of the parent ion A-PIR-B obtained from ISCID-off scan and the sum of neutral masses of released peptide A and B plus the reporter ion mass obtained from ISCID-on scan. The MS data were analyzed using Blinks to identify whether there are masses that fit this relationship. Once found, the MS peaks were reported as a cross-linked candidate. In the following LC-MS/MS run used for the identification of the released peptides, ISCID was implemented continuously to assure that short arms were released throughout the run. The MS cycle which includes one full MS scan followed by five MS/MS scans was used to identify the peptides released. The CID energy was set as 35 V to assure that the peptide bonds were fragmented in this case, and the collected data were searched with Mascot to identify sequences of all of the released peptides. A subsequent LC-MS/MS run was used for parent-product relationship validation. All the parent ions on the list generated from BLinks were targeted under a specific CID voltage to enable the cleavage of the labile bonds in the cross-linker moiety of the parent ion without the peptide backbone fragmentation. All putative cross-linked pairs were identified with the correct mass relationships to validate assignment of all cross-linked pairs. The ISCID was turned off for the entire run to have the intact parent ions detected in the first full MS, and have CID done on all of the intact parent ions in the next five MS/MS scans.

6.3.2 Design of BDD-PIR

a. PIR function

BDD-PIR has a new labile functional group when compared to all the previous PIR molecules [10]. The labile functional group is based on the carbamate *tert*-butoxycarbonyl (BOC) group (widely used in peptide chemistry) since the BOC group is known to undergo facile fragmentation in the mass spectrometer [10] [11]. BDD-PIR contains an "internal BOC" group in which one of the methyl group of *t*-butyl in BOC-amide is substituted as the site of protein attachment to the BDD-PIR backbone. Under energetic activation conditions in the mass spectrometer, either or both of the carbamate groups cleave, generating one long arm and one short arm, or two short arms. In the BDD-PIR inter-cross-links, the long arm was generated from the cleavage at one carbamate bond of cross-links, and the short arm was generated from the complementary side of the carbamate bond with additional loss of C(O)O. **Figure 6.2** provides an example showing the chemical structures of released short arm (z = 1), long arm (z = 1) and reporter ion. When the inter-cross-linking parent ion was cleaved to generate reporter ion and two short arms, an additional mass loss of 88.0190 Da (C_2O_4) should be considered in accounting for the mass relationship.

b. Symmetrical structure

BDD-PIR is symmetrical with two identical arms that are connected to the central tertiary amine nitrogen. If BDD-PIR contained a chiral center, the species in which one arm contains peptide A and the other arm contains peptide B, the structure would not be identical to the isobaric adduct in which the same two peptides were attached to the reagent in the reverse orientation; the two adducts would be diastereomers and they could resolve into two peaks during UPLC. This would create additional complexity in the analysis as well as lead to a reduction in sensitivity of cross-linked peptide detection. Although the tertiary amine in the core of BDD-PIR adopts a tetrahedral geometry, tertiary amines (and ammoniums) are well known to rapidly invert their geometry. Thus, for any given pair of cross-linked peptides (i.e. peptides A and B in the above example), there is only one cross-linked chemical species, thus leading to only a single UPLC peak.

The key to BDD-PIR and other PIR reagents is the cleavage of cross-linked species into single

peptides. The interpretation of the MS/MS spectra becomes nearly impossible without the process of absolute cleavage of cross-linked species into two single peptides components since there are multiple combination possibilities and not all peptide bond cleavages are found in the MSMS spectra in general.

Additionally, unlike isotopically coded cross-linking technology which allows identification of cross-linked peptides based on doublet peaks, PIR technology allows identification based on the relationship formed between the high accuracy mass of the parent ion measured in the ISCID absent scan and the mass of two released peptides measured with ISCID on. Therefore, increased signal for the cross-linked parent ion and the released peptides of BDD-PIR can be useful for *in vivo* applications.[6]

c. Affinity enrichment

Avidin affinity chromatography was used to enrich those tryptic peptides containing covalently attached BDD-PIR while leaving those tryptic peptides which are not attached to this reagent. Thus, ion suppression from an overwhelming amount of interfering free peptides is highly reduced and BDD-PIR-containing peptides are concentrated. This facilitates an easier detection of BDD-PIR containing cross-links in the mass spectrum.

6.3.3 Performance of BDD-PIR

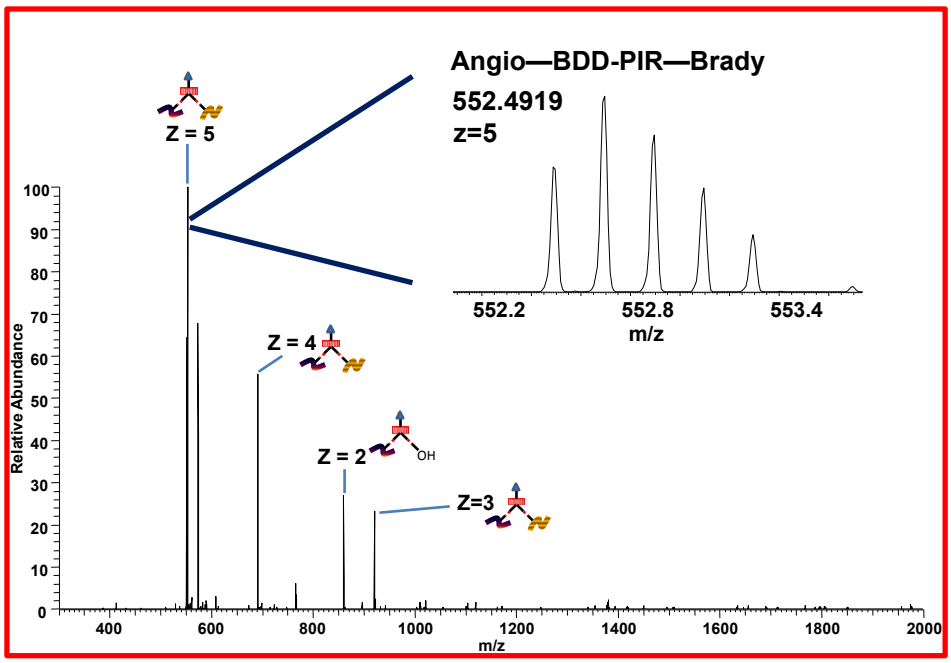
a. Cleavage of BDD-PIR containing peptides

The specific fragmentation properties of BDD-PIR allow cross-linked peptides to be released and measured. Peptide backbone bonds typically require a higher accelerating voltage for fragmentation than the voltage needed for cleaving BDD-PIR cross-linked parent ions into two single peptides and a reporter ion. The cleavage efficiency of the cross-linker can be determined by comparing the gap between the lowest energy needed to fragment peptide bonds and that to cleave the labile bond in the PIR cross-links. A larger gap can yield higher specificity of cross-linker cleavage. Thus, cleavage efficiency was first studied by applying different accelerating CID voltages (0-80 V) on an inter-peptide cross-links—Bradykinin-BDD-PIR—Bradykinin. In this experiment, the MS/MS fragmentation pattern of parent ion was monitored and compared when increasing CID voltages. As a result, the parent ion started to fragment at

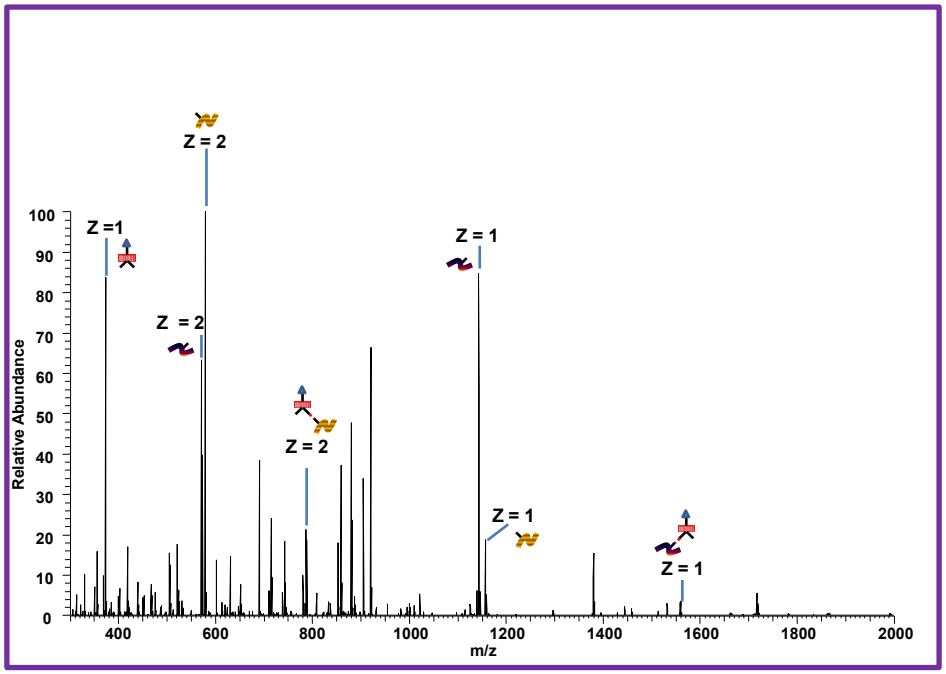
CID = 10 V and completely fragmented at CID = 14 V. The ISCID voltage was also investigated and was found to be optimal at 75 V.

b. Analysis of BDD-PIR Cross-Linked Peptides

The peptide mixture Angiotensin II (A) and Bradykinin (B) was incubated with BDD-PIR, yielding cross-linked products A—BDD-PIR—A, A—BDD-PIR—B, and B— BDD-PIR—B. As these peptides do not have lysine in the sequence, the *N*-terminus of each peptide is the only reactive site for this cross-linking reaction. **Figure 6.3** shows an example of the mass relationships used to identify the A—BDD-PIR—B cross-linked species based on two consecutive mass spectra acquired with ISCID on and off during the LC-MS run. With ISCID off, the parent ion of A—BDD-PIR—B (marked as "Angio—BDD-PIR—Brady") was detected with multiple charge states (z = 3, 4, 5). With ISCID on, the peaks for the "Angio—BDD-PIR—Brady" disappeared, and the released Angiotensin and Bradykinin peptides ions generated from the cleavage of Angio—BDD-PIR—Brady cross-linked peptide were detected in the forms of short arm ions and long arm ions. Meanwhile, the reporter ion (z = 1) was also detected as a big peak in the low-mass range, indicating the cleavage of cross-linked species. As shown in **Figure 6.3**, the mass relationships allow assignment of Angio—BDD-PIR—Brady cross-linked product. With high mass accuracy measurement, the error between calculated and measured product mass is 0.47 ppm. High mass accuracy measurement allows the assignment of peptide cross-linked relationship with high confidence as described elsewhere [12]. Similarly, both the intra-peptide-cross-link and dead-end products were identified based on their specific mass relationships (**Figure 6.2 (b)**).



(a)



(b)

Figure 6.3 MS of Angiotensin II—BDD-PIR—Bradykinin inter-peptide cross-linked species. (a) The low energy mass spectra with ISCID off. (b) The high energy mass spectra with the ISCID on.

Cross-linked products A—BDD-PIR—A and B—BDD-PIR—B were also identified in the same LC-MS run (data not shown). For more complex systems, software BLinks was used to identify the cross-links.

c. Analysis of BDD-PIR Cross-linked BSA

BSA was chosen as the model to test the cross-linking performance of BDD-PIR when applied in proteins. An excess of BDD-PIR was mixed with BSA, digested with trypsin, enriched with avidin beads, and then analyzed with a UPLC coupled high resolution mass spectrometer to finalize the identification of potential mass relationship among cross-links. **Figure 6.4** shows the full workflow for the identification of the inter-cross-linked product AEFVEVTKLVTDLTK—BDD-PIR— ALKAWSVAR (A—BDD-PIR—B). The modified lysine residues labeled by BDD-PIR are underlined in the sequences. Overlapping of the two MS spectra (**Figure 6.4 (a)**) shows the notable difference of the peak assignment under two different energy levels while ISCID is turned off in the top spectrum and ISCID is turned on at the bottom spectrum. Inter-cross-linked product peaks dominated in the low energy ISCID absent mode while peaks from released short arms and the reporter ion were present only in the high energy ISCID on mode. The relationships between the mass of the parent ion peak in **Figure 6.4 (b-1)** and the mass of the released peptides peaks in **Figure 6.4 (b-2)** were identified in BLinks. Once identified, ions of the released peptides were targeted with LC-MS/MS and further identified in Mascot as AEFVEVTKLVTDLTK and ALKAWSVAR (**Figure 6.4 (b-2)**). The cross-linked parent ion was also targeted in a separate LC-MS/MS run to validate relationships. **Figure 6.4 (b-1)** shows that the parent ion A—BDD-PIR—B was selected and targeted to do tandem MS, then generated short arm and long arm of both A and B, finally validating this cross-link.

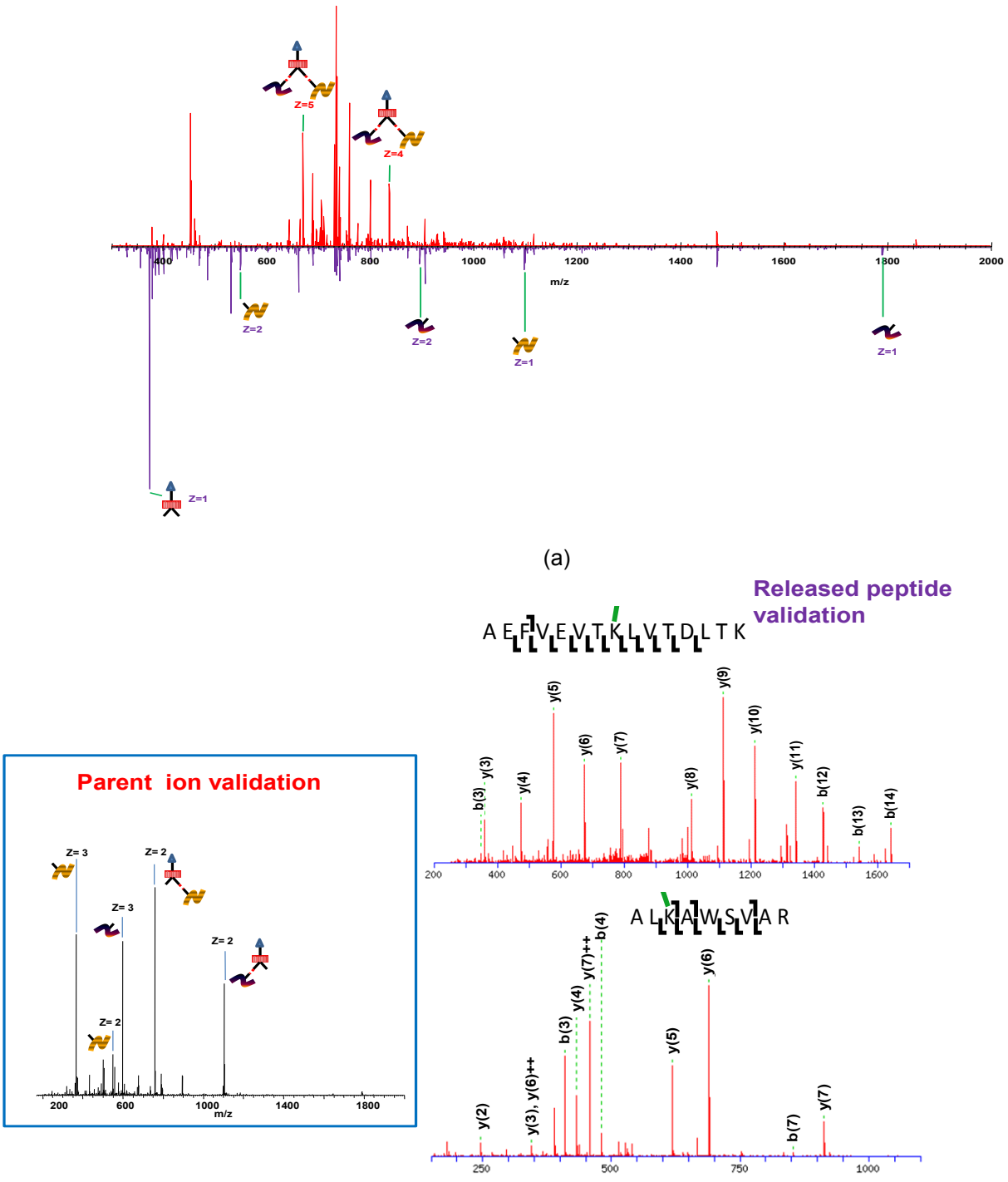


Figure 6.4. Full workflow of the identification of inter-cross-links AEFVEVTKLVTDLTK—BDD-PIR—ALKAWSVAR (A—BDD-PIR—B) in BSA cross-linking experiment with BDD-PIR cross-linker. **Figure 6.4 (a)** shows the overlap of the low energy mass spectra with ISCID off and on, indicating the notable difference of the peaks pattern between low energy (top) and high energy spectra (bottom). The relationship was identified in BLinks as the mass of the short arm ions in high energy spectra and the mass of parent ions in low energy spectra can be fit into the inter-cross-link relationship. After that, parent validation spectra shown in **Figure 6.4 (b-1)** validates the cross-links A—BDD-PIR—B, while **Figure 6.4 (b-2)** shows the sequence identified peptides A and B by Mascot.

BDD-PIR contains a biotin functional group which allows the enrichment of tryptic peptides containing BDD-PIR attached at one end. Molecular weight cut off centrifugal filters were used to remove the excess BDD-PIR reagent, now as the hydrolyzed form with both NHS esters hydrolyzed.

In total, 19 inter-cross-link pairs were identified and validated (**Table 6.1**). Manual inspection of the data resulted in the identification of one additional cross-linked species, marked with an asterisk in **Table 6.1**. This pair was not identified using BLinks, because the signal intensity of the precursor ion was too low. Previous cross-linking studies on BSA have identified the same cross-linked peptide pair [7]. This pair was validated in our research when the specific parent ion was targeted to fragment. As expected, reporter ion, short arm (z = 2) and long arm (z = 2) ions of both peptides were generated. For the short arms identified in Mascot with p value greater than 0.05, manual inspection was used to confirm the identity (**Table 6.2**). The identified cross-linked pairs agreed well with previous results [7] and our results when we used DSS as the control to cross-link BSA. Partial inter-cross-links identified were mapped to the model structure of BSA (**Figure 6.5**). The distances between identified cross-linked sites were measured with the software Molsoft ICM browser (version: 3.6-1i), and the histogram shown in **Figure 6.6** was prepared based on these measurements. In general, the histogram showed that the BDD-PIR cross-linker is expected to span a range from 3 Å to 22.3 Å with a median length of 14.8 Å, which will be highly useful for *in vivo* studies. Our previous PIR cross-linker BRINK was limited by its large size where its maximum span is 44 Å. However, the maximum span of the BDD-PIR was measured in the software ChemDraw Ultra 12.0 as 27.5 Å, and it is much lower than the value of BRINK as we anticipated.

cross-linked parent neutral mass	released peptide A mass	released peptide B mass	error (ppm)	cross-linking	distance (Å)
2458.3578	1085.6183	912.541	3.00	EKVLTSSAR—BDD-PIR—SLGKVGTR	14.6
2488.3724	1085.6111	942.5598	1.87	EKVLTSSAR—BDD-PIR—LSQKFPK	14.7
2634.2928	1261.5361	912.544	-2.10	CCTKPESER—BDD-PIR—SLGKVGTR	20.4
2805.5940	1248.7444	1096.6421	4.22	LVTDLTKVHK—BDD-PIR—ALKAWSVAR	13.1
3075.8067	1703.0685	912.5203	-3.60	QIKKQTALVELLK—BDD-PIR—SLGKVGTR	19.8,17.1
3093.5343	1344.6715	1288.6526	0.65	FKDLGEEHFK—B DD-PIR—DTHKSEIAHR	10.7
3466.7300	1909.8809	1096.6421	2.13	LAKEYEATLEECCAK—BDD-PIR—ALKAWSVAR	14.9
3765.7848	2208.9374	1096.6421	1.51	VHKECCHGDLLECADDR—BDD-PIR—ALKAWSVAR	11.5
3952.9823	2396.1513	1096.6487	6.27	NYQEAKDAFLGSFLYEYSR—BDD-PIR—ALKAWSVAR	12.3
4008.0294	2635.2889	912.543	3.64	DDSPDLPKLKPDPNTLCDEFK—BDD-PIR—SLGKVGTR	12.6, 18.0
4322.2582	2624.2946	1237.775	18.0	QNCDQFEKLGEYGFQNALIVR—BDD-PIR—KQTALVELLK	20.4
4636.3000	2635.2949	1540.8196	5.17	DDSPDLPKLKPDPNTLCDEFK—BDD-PIR—FWGKYLYEIAR	17.5, 16.6
4688.3295	3542.783	685.3707	3.83	DLGEEHFKGLVLIAFSQYLQQCPFDEHVK—BDD-PIR—ADEKK	4.5
4763.3921	3542.79	760.4325	5.16	DLGEEHFKGLVLIAFSQYLQQCPFDEHVK—BDD-PIR—KFWGK	13.9
4766.2258	2396.1521	1909.8996	6.92	NYQEAKDAFLGSFLYEYSR—BDD-PIR—LAKEYEATLEECCAK	8.2
4819.4602	2624.2904	1735.0017	4.91	QNCDQFEKLGEYGFQNALIVR—BDD-PIR—KVPQVSTPTLVEVSR	22.3
5098.4201	3541.581	1096.6421	2.89	LAKEYEATLEECCAKDDPHACYSTVFDK—BDD-PIR—ALKAWSVAR	14.9
5414.5839	3868.781	1085.6155	0.38	RHPYFYAPELLYYANKYNGVFQECCQAEDK—B DD-PIR—EKVLTSSAR	3.0
2807.3616*				CCTKPESER—BDD-PIR—EKVLTSSAR	16.9

Table 6.1 Identified Inter-cross-linked pairs from BDD-PIR cross-linking BSA experiment
* indicates the cross-links that was not identified by BLinks due to the low signal in the mass spectra.

Observed *m/z*	Observed mass	calculated mass	delta mass (ppm)	expect	peptide sequence
343.6900	685.3654	685.3646	1.20	0.47	ADEKK
1019.4925	2036.9704	2036.9724	-0.98	0.024	ADLAKYICDNQDTISSK
549.3251	1096.6356	1096.6393	-3.36	0.00004	ALKAWSVAR
646.3300	1290.6454	1290.6390	4.97	0.000095	CASIQKFGER
1262.5400	1261.5327	1261.5431	-8.24	0.022	CCTKPESER
879.4272	2635.2597	2635.2727	-4.93	0.011	DDSPDLPKLKPDPNTLCDEFK
886.6980	3542.7629	3542.7544	2.40	0.047	DLGEEHFKGLVLIAFSQYLQQCPFDEHVK
645.3329	1288.6512	1288.6524	-0.93	0.008	DTHKSEIAHR
543.8102	1085.6058	1085.6080	-2.06	0.00025	EKVLTSSAR
673.3416	1344.6686	1344.6714	-2.08	0.0005	FKDLGEEHFK
771.4093	1540.8041	1540.8078	-2.40	0.039	FWGKYLYEIAR
742.9030	1483.7915	1483.7891	1.62	0.039	GACLLPKIETMR
381.2206	760.4267	760.4272	-0.67	0.041	KFWGK
883.4871	1764.9596	1764.9603	-0.40	0.000015	KFWGKYLYEIAR
835.4579	1668.9012	1668.9028	-0.96	0.00021	KFWGKYLYEIAR
1238.7676	1237.7603	1237.7645	-3.39	0.033	KQTALVELLK
868.4996	1734.9846	1734.9880	-1.96	0.0055	KVPQVSTPTLVEVSR
955.9425	1909.8705	1909.8801	-5.03	0.000068	LAKEYEATLEECCAK
1181.5323	3541.5752	3541.5727	0.71	0.047	LAKEYEATLEECCAKDDPHACYSTVFDK
818.4408	1634.8670	1634.8702	-1.96	0.0008	LCVLHEKTPVSEK
821.9777	1641.9409	1641.9454	-2.74	3.5e-6	LKHLVDEPQNLIK
1394.7227	2787.4308	2787.4265	1.52	0.0021	LKHLVDEPQNLIKQNCDQFEK
472.2859	942.5572	942.5538	3.53	0.025	LSQKFPK
902.0937	2703.2592	2703.2520	2.66	0.011	LVNELTEFAKTCVADESHAGCEK
1249.7533	1248.7460	1248.7442	1.48	0.48	LVTDLTKVHK
1538.7307	3075.4469	3075.4464	0.16	0.0043	LVTDLTKVHKECCHGDLLECADDR
1199.0773	2396.1400	2396.1324	3.15	3.5e-6	NYQEAKDAFLGSFLYEYSR
852.5347	1703.0549	1703.0596	-2.76	0.0029	QIKKQTALVELLK
875.7608	2624.2606	2624.2693	-3.32	0.00065	QNCDQFEKLGEYGFQNALIVR
457.2757	912.5368	912.5393	-2.74	0.044	SLGKVGTR
1021.5377	2041.0608	2041.0666	-2.84	0.0024	SLHTLFGDELCKVASLR
987.4427	2959.3062	2959.3150	-2.97	0.000032	TCVADESHAGCEKSLHTLFGDELCK
1194.9037	3581.6892	3581.6953	-1.70	0.0083	TCVADESHAGCEKSLHTLFGDELCKVASLR
1105.4735	2208.9325	2208.9351	-1.18	0.00026	VHKECCHGDLLECADDR
781.8870	1561.7595	1561.7592	0.19	0.0022	VTKCCTESLVNR

Table 6.2 Identified released peptides from BDD-PIR cross-linking BSA experiment

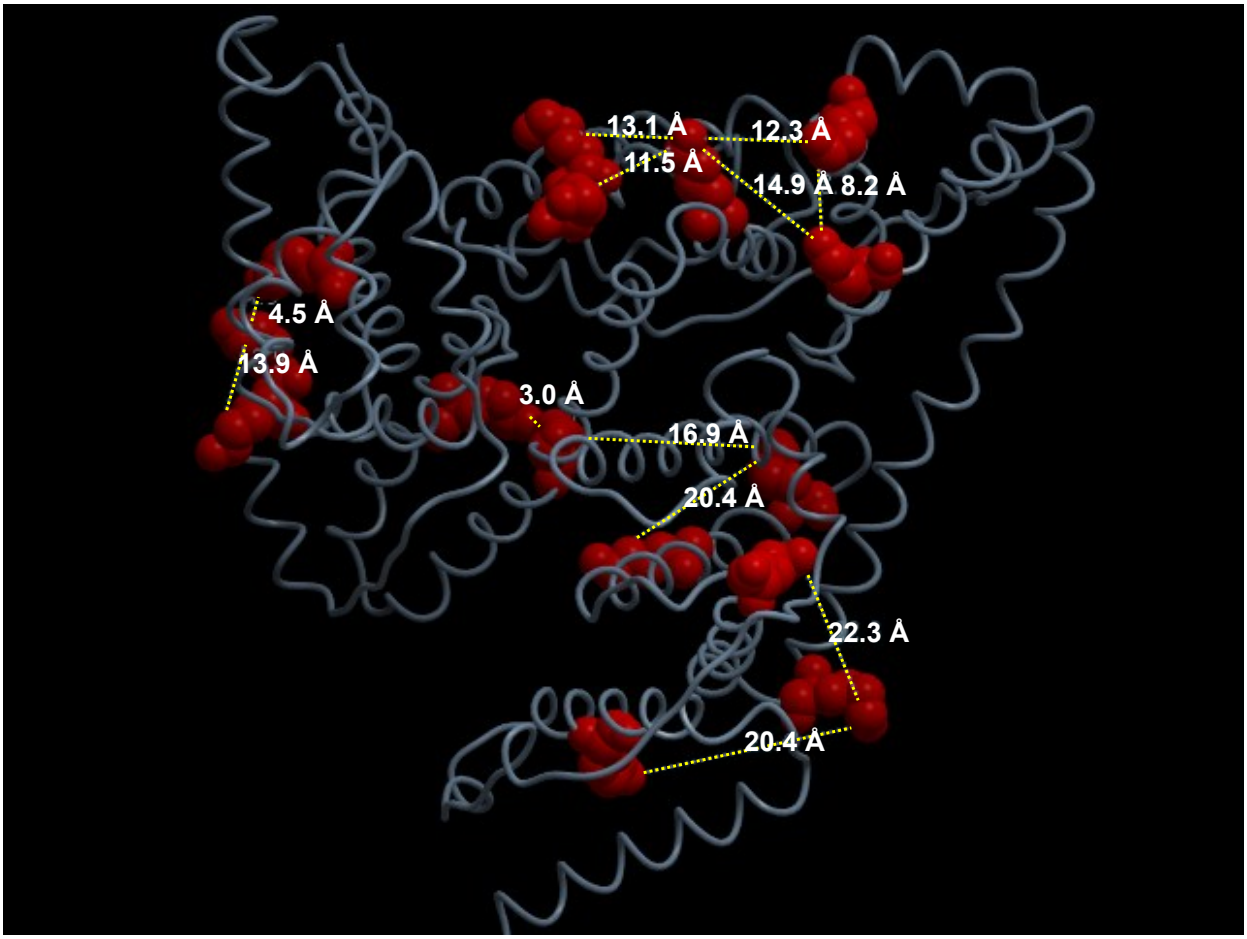


Figure 6.5 A subset of inter-cross-links identified from bovine serum albumin with BDD-PIR cross-linker. Red amino acids indicate lysines that have been found to be cross-linked to one another as indicated by connecting yellow dash lines.

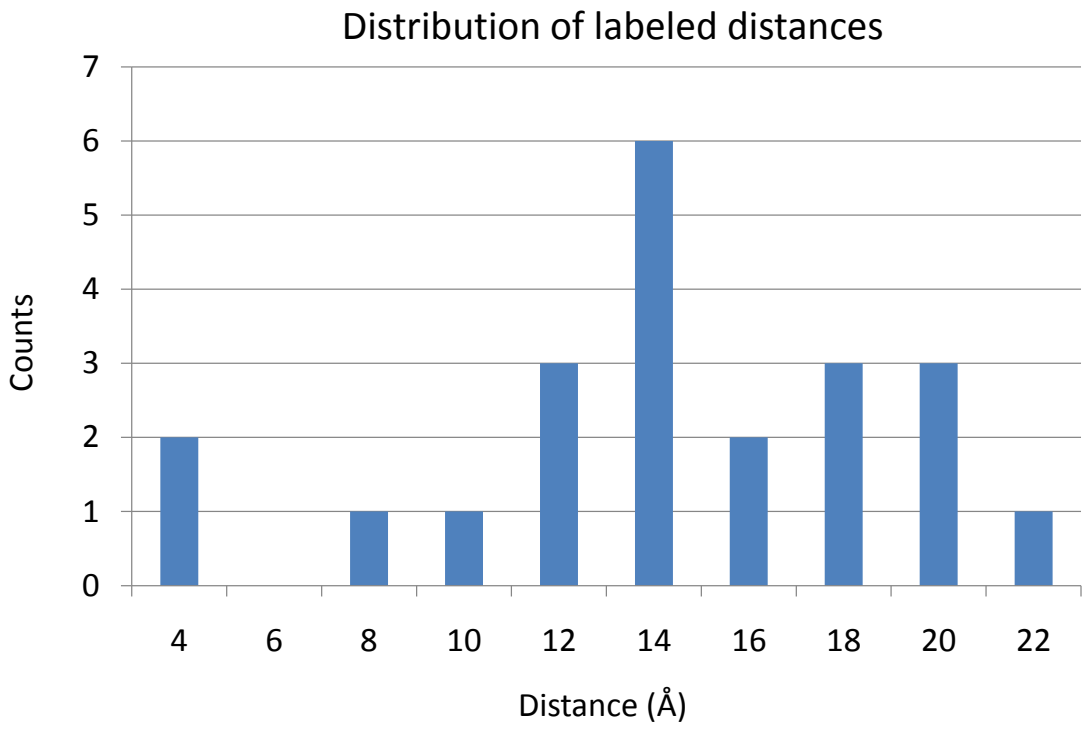


Figure 6.6 Distribution of cross-linked distances for all the identified inter-cross-links in standard protein BSA.

Our previous PIR-cross-linkers typically generated both z = 1 and z = 2 reporter ions, thus frequently released z = 1 peptides. The distinct advantage of the BDD-PIR cross-linker is its cross-links generate z = 1 reporter ions only, and therefore higher charge states rather than z = 1 of the released peptide ions are more likely to be released. Therefore, CID of the released peptide ion is likely to generate more useful fragment peaks and this helps the peptide ion to be sequenced in the Mascot search engine.

6.4 Conclusion

A novel mass spectrometer identifiable PIR cross-linker BDD-PIR was designed and synthesized using solution phase chemistry. Cleavage of the carbamate labile bonds in BDD-PIR generates a specific

fragmentation pattern for inter-peptide cross-links as demonstrated with peptide standards and a model protein. BDD-PIR is compatible with existing software tools and allows increased cleavage specificity as compared with previous PIR molecules. The incorporation of the biotin moiety into BDD-PIR enables enrichment of reagent-linked peptides in complex systems. The symmetrical structural design of BDD-PIR increases the sensitivity for detection of protein interactions in a mass spectrometer. These properties, combined with the smaller size and improved aqueous solubility of BDD-PIR will be highly beneficial for *in vivo* studies on proteins and complexes in cells.

6.5 References

[1] Ward, D. E.; Gai, Y.; Lazny, R.; Pedras, M.S. Probing Host-selective Phytotoxicity: Synthesis of Destruxin B and Several Natural Analogues. *J. Org. Chem.* **2001,** *66,* 7832-7840.

[2] Paramonov, S. E.; Bachelder, E. M.; Beaudette, T. T.; Standley, S. M.; Lee, C. C.; Dashe, J.; Fréchet, JMJ. Fully Acid-Degradable Biocompatible Polyacetal Microparticles for Drug Delivery. *Bioconjugate Chem.* **2008,** *19,* 911–919.

[3] André, S.; Specker, D.; Bovin, N.V.; Lensch, M.; Kaltner, H.; Gabius, HJ.; Wittmann, V. Carbamate-linked lactose: design of clusters and evidence for selectivity to block binding of human lectins to (neo)glycoproteins with increasing degree of branching and to tumor cells. *Bioconjugate Chem.* **2009,** *20,* 1716–1728.

[4] Lewis, I.; Bauer, W.; Albert, R.; Chandramouli, N.; Pless, J.; Weckbecker, G.; Bruns, C. A Novel Somatostatin Mimic with Broad Somatotropin Release Inhibitory Factor Receptor Binding and Superior Therapeutic Potential. *J. Med. Chem.* **2003,** *46,* 2334-2344.

[5] Rao, T.S.; Nampalli, S.; Sekher, P.; Kumar, S. TFA-NHS as Bifunctional Protecting Agent: Simultaneous Protection and Activation of Amino Carboxylic Acids. *Tetrahedron Lett.* **2002**, *43,* 7793–7795.

[6] Yang, L.; Tang, X.; Weisbrod C.R.; Munske, G. R.; Eng, J.K.; Haller, P.D.; Kaiser, N.K.; Bruce, J.E. A Photocleavable and Mass Spectrometry Identifiable Cross-linker for Protein Interaction Studies. *Anal. Chem.* **2010**, *82*, 3556–3566.

[7] Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L. N.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. Identification of Cross-linked Peptides from Large Sequence Databases. *Nat. Methods* **2008**, *5*, 315–318.

[8] Hoopmann, M. R.; Weisbrod, C. R.; Bruce, J. E. Improved Strategies for Rapid Identification of Chemically Cross-linked Peptides. *J. Proteome Res.* **2010**, *9*, 6323-6333.

[9] Tang, X.; Bruce, J. E. A New Cross-linking Strategy: Protein Interaction Reporter (PIR) Technology for Protein-protein Interaction Studies. *Mol. BioSyst*., **2010**, *6*, 939-947.

[10] Blanchard, S.; Sadilek, M.; Scott, C.R.; Tureček, F.; Gelb, M. H. Tandem Mass Spectrometry for the Direct Assay of Lysosomal Enzymes in Dried Blood Spots: Application to Screening Newborns for Mucopolysaccharidosis I *Clin Chem*. **2008**, *54*, 2067-2070.

[11] Wang, D.; Wood, T; Sadilek, M; Scott, C.R.; Tureček, F; Gelb, M.H. Tandem Mass Spectrometry for the Direct Assay of Enzymes in Dried Blood Spots. *Clin Chem*. **2007**, *53*, 137-140.

[12] Anderson, G. A.; Tolic, N.; Tang, X.; Zheng, C.; Bruce, J. E. Informatics Strategies for Large-scale Novel Cross-linking Analysis. *J. Proteome Res.* **2007**, *6*, 3412-3421.

Complete Bibliography

1. Yates, J. A Century of Mass spectrometry: From Atoms to Proteomes. *Nat. Methods* **2011,** *8*, 633–637.
2. Thomson, J. J. Cathode Rays. *Philos. Mag.* **1897,** *44*, 293-316.
3. Grayson, M. (ed.) Measuring Mass: From Positive Rays to Proteins (Chemical Heritage Press, Philadelphia, **2002**).
4. Biemann, K. Four Decades of Structure Determination by Mass Spectrometry: From alkaloids to Heparin. *J. Am. Soc. Mass Spectrom.* **2002,** *13*, 1254-1272.
5. Biemann, K.; Gapp, G.; Seibl, J. Application of Mass Spectrometry to Structure Problems. I. Amino Acid Sequence in Peptides. *J. Am. Chem. Soc.* **1959,** *81*, 2274-2275.
6. Munson, M. S. B.; Field, F. H. Chemical Ionization Mass Spectrometry I. General introduction. *J. Am. Chem. Soc.* **1966,** *88*, 2621-2630.
7. Ryhage, R. Use of a Mass Spectrometer as a Detector and Analyzer for Effluents Emerging from High Temperature Gas-Liquid Chromatography Columns. *Anal. Chem.* **1964,** *36,* 759-764.
8. Barber, M.; Bordoli, R. S.; Sedgwick, R.D.; Tyler, A. N. FAB of Solids as an Ion Source in Mass Spectrometry. *Nature* **1981,** *293*, 270-275.
9. Morris, H.R.; Panico, M.; Barber, M.; Bordoli, R.S.; Sedgwick, R.D.; Tyler, A. Fast Atom Bombardment: a New Mass Spectrometric Method for Peptide Sequence Analysis. *Biochem. Biophys. Res. Commun,* **1981,** *101,* 623–31.
10. Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988,** *60*, 2299-2301.
11. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, **1989,** *246*, 64-71.
12. McLafferty, F.; Horn, D.M.; Breuker, K.; Ge, Y.; Lewis, M.A.; Cerda, B.; Zubarev, R.A.; Carpenter, B.K. Electron Capture Dissociation of Gaseous Multiply Charged Ions by Fourier-Transform Ion Cyclotron Resonance. *J. Am. Soc. Mass Spectrom.* **2001**, *12,* 245.
13. Syka, J.E.; Coon, J.J.; Schroeder, M.J.; Shabanowitz, J.; Hunt, D.F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry .*Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101*, 9528–9533.
14. Uggerud, E. Electron Capture Dissociation of the Disulfide Bond-a Quantum Chemical Model Study. *Int. J. Mass Spectrom.* **2004,** *234*, 45–50.

15. Syrstad, E.A.; Tureček, F. Toward a General Mechanism of Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2005,** *16*, 208–224.

16. Johnson, R. S.; Martin, S. A.; Biemann, K. Collision-Induced Fragmentation of $(M+H)^+$ Ions of Peptides. Side Chain Specific Sequence Ions. *Int. J. Mass Spectrom. Ion Processes.* **1988,** *86*, 137–154.

17. Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

18. Steen, H.; Mann, M. The ABC's (and XYZ's) of Peptide Sequencing. *Nat. Rev. Mol. Cell Biol.* **2004,** *5*, 699–711.

19. Hennrich, M. L.; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J. R.; Mohammed, S. Effect of Chemical Modifications on Peptide Fragmentation Behavior upon Electron Transfer Induced Dissociation. *Anal. Chem.* **2009,** *81*, 7814– 22.

20. Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96*, 7131–7136.

21. Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate *De Novo* Sequencing of Lysine-terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000,** *14*, 2348–2356.

22. Tian, Y.; Zhou, Y.; Elliott, S.; Aebersold, R.; Zhang, H. Solid-phase extraction of N-linked glycopeptides. *Nature Protocols*, **2007,** *2*, 334-339.

23. Zimnicka, M.; Moss, C. L.; Chung, T. W.; Hui, R.; Tureček, F. Tunable Charge Tags for Electron-Based Methods of Peptide Sequencing: Design and Applications. *J.Am. Soc. Mass Spectrom.* **2012,** *23*, 608-620.

24. Lohse, A.; Joergensen, M. R.; Martins, R.; Hindsgaul, O. Solid-phase Oligosaccharide Tagging: a Technique for Manipulation of Immobilized Carbohydrates *PCT Int. Appl.* **2006,** WO 2006084461 A1 20060817

25. Kim, S.; Edwards, J. R.; Deng, L.; Chung, W.; Ju, J. Solid Phase Capturable Dideoxynucleotides for Multiplex Genotyping Using Mass Spectrometry. *Nucleic Acids Research* **2002,** *30*, e85/1-e85/6.

26. Fingerhut, R.; Olgemöller, B. Newborn Screening for Inborn Errors of Metabolism and Endocrinopathies: an Update. *Anal. Bioanal. Chem.* **2009,** *393*, 1481–1497.

27. Wilson, J. M. G., Jungner, G. Principles and Practice of Screening for Diseases. World Health Organization, Geneva **1968** Public health papers no 34.

28. Fölling, A. Über Ausscheidung von Phenylbrenztraubensäure in den Harn als Stoffwechselanomalie in Verbindung mit Imbezillität. Hoppe-Seyler's. *Z Physiol. Chem.* **1934**, *227,* 169–176.

29. Guthrie, R.; Susi, A. A Simple Phenylalanine Method for Detecting Phenylketonuria in Large Populations of Newborn Infants *Pediatrics*, **1963**, *32*, 338–343.

30. Jervis, G. A. Phenylpyruvic Oligophrenia: Introductory Study of Fifty Cases of Mental Deficiency Associated with Excretion of Phenylpyruvic Acid. *Arch. Neurol. Psychiatry*, **1937**, *38*, 944–963.

31. Jervis, G. A. *J. Ment. Sci.*, **1939**, *85*, 719–762.

32. Gelb, M. H.; Tureček, F.; Scott, C. R.; Chamoles, N. A. Direct Multiplex Assay of Enzymes in Dried Blood Spots by Tandem Mass Spectrometry for the Newborn Screening of Lysosomal Storage Disorders *J. Inherit. Metab. Dis*. **2006***, 29,* 397–404.

33. Lehotay, D. C.; Hall, P.; Lepage, J.; Eichhorst, J. C.; Etter, M. L.; Greenberg, C. R. LC–MS/MS Progress in Newborn Screening *Clinical Biochemistry* **2011,** *44*, 21–31.

34. Sanderson, S.; Green, A.; Preece, M. A.; Burton, H. The Incidence of Inherited Metabolic Disorders in the West Midlands, *UK. Arch. Dis. Child.* **2006,** *91*, 896–899.

35. Back, J. W.; De Jong, L.; Muijsers, A. O.; De Koster, C. G. Chemical Cross-linking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* **2003**, *331*, 303–313.

36. Sinz, A. Chemical Cross-linking and Mass Spectrometry to Map Three-dimensional Protein Structures and Protein–protein Interactions. *Mass Spectrom. Rev*. **2006**, *25*, 663–682.[37] 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*, 5802–5806.

37. Rinner, O.; Seebacher, J.; Walzthoeni, T.; Mueller, L. N.; Beck, M.; Schmidt, A.; Mueller, M.; Aebersold, R. Identification of Cross-linked Peptides from Large Sequence Databases. *Nat. Methods* **2008**, *5*, 315–318.

38. Tang, X.; Yi, W.; Munske, G. R.; Adhikari, D. P.; Zakharova, N. L.; Bruce, J. E. Profiling the Membrane Proteome of Shewanella Oneidensis MR-1 with New Affinity Labeling Probes. *J. Proteome Res.* **2007**, *6*, 724–734.

39. Zhang, H.; Tang, X.; Munske, G. R.; Zakharova, N.; Yang, L.; Zheng, C.; Wolff, M. A.; Tolic, N.; Anderson, G. A.; Shi, L.; Marshall, M. J.; Fredrickson, J. K.; Bruce, J. E. Identification of Protein-

Protein Interactions and Topologies in Living Cells with Chemical Cross-linking and Mass Spectrometry *J. Proteome Res.* **2008**, *7*, 1712–1720.

40. 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*, 1237–1247.

41. Zhang, H.; Tang, X.; Munske, G. R.; Tolic, N.; Anderson, G. A.; Bruce, J. E. Identification of Protein-Protein Interactions and Topologies in Living Cells with Chemical Cross-linking and Mass Spectrometry *Mol. Cell. Proteomics* **2009**, *8*, 409–420.

42. Mueller, D. R. ; Schindler, P. ; Towbin, H. ; Wirth, U. ; Voshol, H. ; Hoving, S. ; Steinmetz, M. O. Isotope-tagged cross-linking reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis. *Anal. Chem.* **2001**, *73*, 1927–1934.

43. Chu, F.; Mahrus, S.; Craik, C. S.; Burlingame, A. L. Isotope-coded and Affinity-tagged Cross-linking (ICATXL): An Efficient Strategy to Probe Protein Interaction Surfaces. *J. Am. Chem. Soc.* **2006**, *128*, 10362–10363.

44. Wine, R. N.; Dial, J. M. ; Tomer, K. B.; Borchers, C. H. Identification of Components of Protein Complexes Using a Fluorescent Photo-cross-linker and Mass Spectrometry. *Anal. Chem.* **2002**, *74*, 1939–1945.

45. Yang, L.; Tang, X.; Weisbrod C.R.; Munske, G. R.; Eng, J.K.; Haller, P.D.; Kaiser, N.K.; Bruce, J.E. A Photocleavable and Mass Spectrometry Identifiable Cross-linker for Protein Interaction Studies. *Anal. Chem.* **2010**, *82*, 3556–3566.

46. Trester-Zedlitz, M.; Kamada, K.; Burley, S. K.; Fenyoe, D.; Chait, B. T.; Muir, T. W. A Modular Cross-Linking Approach for Exploring Protein Interactions. *J. Am. Chem. Soc.* **2003**, *125*, 2416–2425.

47. Hurst, G. B.; Lankford, T. K.; Kennel, S. J. Mass Spectrometric Detection of Affinity Purified Crosslinked Peptides *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 832–839.

48. Sinz, A.; Kalkhof, S.; Ihling, C. Mapping Protein Interfaces by a Trifunctional Cross-Linker Combined with MALDI-TOF and ESI-FTICR Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1921–1931.

49. 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*, 311–318.

50. Roth, K. D. W.; Huang, Z.H.; Sadagopan, N.; Watson, J. T. Charge Derivatization of Peptides for Analysis by Mass Spectrometry. *Mass Spectrom. Rev.* **1999,** *17,* 255-274.

51. Ren, D.; Julka, S.; Inerowicz, H. D.; Regnier, F. E. Enrichment of Cysteine-containing Peptides from Tryptic Digests Using a Quaternary Amine Tag. *Anal. Chem*. **2004,** *76,* 4522-4530.

52. Gunawardena, H. P.; Gorenstein, L.; Erickson, D. E.; Xia, Y.; McLuckey, S. A. Electron Transfer Dissociation of Multiply Protonated and Fixed Charge Disulfide Linked Polypeptides. *Int. J. Mass Spectrom.* **2007,** *265,* 130-138.

53. Vath, J. E.; Biemann, K. Microderivatization of Peptides by Placing a Fixed Positive Charge at the N-terminus to Modify High Energy Collision Fragmentation. *Int. J. Mass Spectrom. Ion Processes.* **1990,** *100*, 287-299.

54. Shen, T. L.; Allison, J. Interpretation of Matrix-Assisted Laser Desorption/Ionization Postsource Decay Spectra of Charge-derivatized Peptides: Some Examples of Tris[(2,4,6- trimethoxyphenyl) phosphonium]-tagged Proteolytic Digestion Products of Phosphoenol-pyruvate Carboxykinase. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 145-152.

55. Chamot-Rooke, J.; van der Rest, G.; Dalleu, A.; Bay, S.; Lemoine, J. The Combination of Electron Capture Dissociation and Fixed Charge Derivatization Increases Sequence Coverage for O-Glycosylated and O-Phosphorylated Peptides. *J. Am. Soc. Mass Spectrom.* **2007,** *18,* 1405-1413.

56. Chamot-Rooke, J.; Malosse, C.; Frison, G.; Tureček, F. Electron Capture in Charge-tagged Peptides. Evidence for the Role of Excited Electronic States. *J. Am. Soc. Mass Spectrom.* **2007,** *18,* 2146-2161.

57. Jones, J. W.; Sasaki, T.; Goodlett, D. R.; Tureček, F. Electron Capture in Spin-Trap Capped Peptides. An Experimental Example of Ergodic Dissociation in Peptide Cation-Radicals. *J. Am. Soc. Mass. Spectrom.* **2007,** *18,* 432-444.

58. Brancia, F.L., Oliver, S.G., Gaskell, S.J. Improved Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis of Tryptic Hydrolysates of Proteins Following Guanidination of Lysine-containing Peptides. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2070–2073.

59. Peters, E.C., Horn, D.M., Tully, D.C., Brock, A. A Novel Multifunctional Labeling Reagent for Enhanced Protein Characterization with Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2387–2392.

60. Beardsley, R.L., Reilly, J.P. Optimization of Guanidination Procedures for MALDI Mass Mapping. *Anal. Chem*. **2002**, *74*, 1884–1890.

61. Ji, C., Guo, N., Li, L. Differential Dimethyl Labeling of N-termini of Peptides after Guanidination for Proteome Analysis. *J. Proteome Res.* **2005**, *4*, 2099–2108.

62. Warwood, S., Mohammed, S., Cristea, I.M., Evans, C., Whetton, A.D., Gaskell, S.J. Guanidination Chemistry for Qualitative and Quantitative Proteomics. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 3245–3256.

63. Miyashita, M., Hanai, Y., Awane, H., Yoshikawa, T., Miyagawa, H. Improving Peptide Fragmentation by N-terminal Derivatization with High Proton Affinity. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1130–1140.

64. Margolis, S.; Langdon, R. G. Chemistry and Metabolism of Macromolecules: Studies on Human Serum β_1-Lipoprotein: II. Chemical Modifications *J. Biol. Chem.* **1966**, *241*, 477-484.

65. Kim, S.; Yi, K. Y. Di-2-Pyridyl Thionocarbonate. A New Reagent for the Preparation of Isothiocyanates and Carbodiimides. *Tetrahedron Lett.* **1985,** *26*, 1661-1664

66. Goebel, H. H.; Mole, S. E.; Lake, B. D. (eds) *The Neuronal Ceroid Lipofuscinoses (Batten Disease): Biomedical and Health Research*, IOS Press, Amsterdam, **1999**

67. Wisniewski, K. E.; Kida, E.; Golabek, A. A.; Kacemarski, W.; Connell, F.; Zhong, N. Neuronal Ceroid Lipofuscinoses: Classificaion and Diagnosis. *Adv. Genet.* **2001,** *45,* 1-34.

68. Haltia, M. The Neurocal Ceroid-lipofuscinoses. *J. Neuropathol. Exp. Neurol*. **2003**, *62*, 1-13.

69. Williams, R. E.; Gottlob, I.; Lake, B. D.; Goebel, H. H.; Winchester, B. G.; Wheeler, R. B. CLN2: Classic Late Infantile NCL. In *the Neurocal Ceroid lipofuscinoses (Battern Disease),*H. H. Goebel, ed. (Amsterdam: IOS Press), pp. 37-53.

70. Jalanko, A.; Braulke, T. Neuronal Ceroid Lipofuscinosis. *Biochimica et Biophysica Acta, Molecular Cell Research.* **2009**, *1793*, 697-709.

71. Pal, A.; Kraetzner, R.; Gruene, T.; Grapp, M.; Schreiber, K.; Grønborg, M.; Urlaub, H.; Becker, S.; Asif, A. R.; Gärtner, J.; Sheldrick, G. M.; Steinfeld, R. Structure of Tripeptidyl-peptidase I Provides Insight into the Molecular Basis of Late Infantile Neuronal Ceroid Lipofuscinosis*. J. Biol. Chem.* **2009**, *284*, 3976-3984.

72. Vines, D.; Warburton, M. J. Purification and Characterisation of a Tripeptidyl Aminopeptidase I from Rat Spleen. *Biochim. Biophys. Acta* **1998,** *1384*, 233-242.

73. Page, A. E.; Fuller, K.; Chambers, T. J.; Warburton, M. J. Purification and Characterization of a Tripeptidyl Peptidase I from Human Osteoclastomas: Evidence for its Role in Bone Resorption. *Arch. Biochem. Biophys*. **1993***, 306*, 354-359.

74. Ivanov, I.; Tasheva, D.; Todorova, R.; Dimitrova, M. Synthesis and use of 4-peptidylhydrazido-N-hexyl-1,8-naphthalimides as fluorogenic histochemical substrates for dipeptidyl peptidase IV and tripeptidyl peptidase I. *Eur. J. Med. Chem*. **2009**, *44*, 384-392.

75. Sohar, I.; Lin, L.; Lobel, P. Enzyme-based diagnosis of classical late infantile neuronal ceroid lipofuscinosis: comparison of tripeptidyl peptidase I and pepstatin-insensitive protease assays. *Clin. Chem*. **2000**, *46*, 1005-1008.

76. Tian, Y.; Sohar, I.; Taylor, J. W.; Lobel, P. Determination of the Substrate Specificity of Tripeptidyl-peptidase I Using Combinatorial Peptide Libraries and Development of Improved Fluorogenic Substrates. *J. Biol. Chem.* **2006**, *281*, 6559-6572

77. Lukacs, Z.; Santavuori, P.; Keil, A.; Steinfeld, R.; Kohlschütter, A. Rapid and Simple Assay for the Determination of Tripeptidyl Peptidase and Palmitoyl Protein Thioesterase Activities in Dried Blood Spots. *Clin. Chem*. **2003**, *49*, 509-511.

78. Wolfe, B. J.; Blanchard, S.; Sadilek, M.; Scott, C. R.; Tureček, F.; Gelb, M. H. Tandem Mass Spectrometry for the Direct Assay of Lysosomal Enzymes in Dried Blood Spots: Application to Screening Newborns for Mucopolysaccharidosis II (Hunter Syndrome) *Anal. Chem*. **2011**, *83*, 1152-1156.

79. Pitterkow, M.; Lewinsky R.; Christensen, Selective Synthesis of Carbamate Protected Polyamines Using Alkyl Phenyl Carbonates, J. B. *Synthesis*, **2002**, *15*, 2195-2202.

80. Szabó, P.T.; Kele, Z. Electrospray Mass Spectrometry of Hydrophobic Compounds Using Dimethyl Sulfoxide and Dimethylformamide as Solvents. *Rapid Commun Mass Spectrom.* **2001**, *15*, 2415-2419.

81. Blanchard, S.; Sadilek, M.; Scott, C.R.; Tureček, F.; Gelb, M.H. Tandem Mass Spectrometry for the Direct Assay of Lysosomal Enzymes in Dried Blood Spots: Application to Screening Newborns for Mucopolysaccharidosis I. *Clin Chem.* **2008**, *54*, 2067-2070.

82. Ward, D. E.; Gai, Y.; Lazny, R.; Pedras, M.S. Probing Host-selective Phytotoxicity: Synthesis of Destruxin B and Several Natural Analogues. *J. Org. Chem.* **2001,** *66,* 7832-7840.

83. Paramonov, S. E.; Bachelder, E. M.; Beaudette, T. T.; Standley, S. M.; Lee, C. C.; Dashe, J.; Fréchet, JMJ. Fully Acid-Degradable Biocompatible Polyacetal Microparticles for Drug Delivery. *Bioconjugate Chem.* **2008,** *19,* 911–919.

84. André, S.; Specker, D.; Bovin, N.V.; Lensch, M.; Kaltner, H.; Gabius, HJ.; Wittmann, V. Carbamate-linked lactose: design of clusters and evidence for selectivity to block binding of human lectins to (neo)glycoproteins with increasing degree of branching and to tumor cells. *Bioconjugate Chem.* **2009,** *20,* 1716–1728.

85. Lewis, I.; Bauer, W.; Albert, R.; Chandramouli, N.; Pless, J.; Weckbecker, G.; Bruns, C. A Novel Somatostatin Mimic with Broad Somatotropin Release Inhibitory Factor Receptor Binding and Superior Therapeutic Potential. *J. Med. Chem.* **2003,** *46,* 2334-2344.

86. Rao, T.S.; Nampalli, S.; Sekher, P.; Kumar, S. TFA-NHS as Bifunctional Protecting Agent: Simultaneous Protection and Activation of Amino Carboxylic Acids. *Tetrahedron Lett.* **2002**, *43,* 7793–7795.

87. Hoopmann, M. R.; Weisbrod, C. R.; Bruce, J. E. Improved Strategies for Rapid Identification of Chemically Cross-linked Peptides. *J. Proteome Res.* **2010**, *9*, 6323-6333.

88. Tang, X.; Bruce, J. E. A New Cross-linking Strategy: Protein Interaction Reporter (PIR) Technology for Protein-protein Interaction Studies. *Mol. BioSyst*., **2010**, *6*, 939-947.

89. Wang, D.; Wood, T; Sadilek, M; Scott, C.R.; Tureček, F; Gelb, M.H. Tandem Mass Spectrometry for the Direct Assay of Enzymes in Dried Blood Spots. *Clin Chem*. **2007**, *53*, 137-140.

90. Anderson, G. A.; Tolic, N.; Tang, X.; Zheng, C.; Bruce, J. E. Informatics Strategies for Large-scale Novel Cross-linking Analysis. *J. Proteome Res.* **2007**, *6*, 3412-3421.

Vita

Chang Xue was born in Nanjing, Jiangsu Province, China. She completed her Bachelor of Science in Pharmaceutical Science from Tianjin University, Tianjin, China in July 2008. She then enrolled in University of Washington in Seattle in 2009 and then received her Master of Science in Chemistry in 2011. In March 2014, she received the degree of Doctor of Philosophy in Chemistry from University of Washington.