The Dissertation Committee for Michelle Renee Robinson Certifies that this is the approved version of the following dissertation:

Advancement of Photodissociation Mass Spectrometry Methods for the Analysis of Protein Post-translational Modifications

Committee:

Jennifer S. Brodbelt, Supervisor

Richard M. Crooks

Kevin N. Dalby

Lauren J. Webb

Yan Jessie Zhang

Advancement of Photodissociation Mass Spectrometry Methods for the Analysis of Protein Post-translational Modifications

by

Michelle Renee Robinson, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2016

Dedication

For my parents Jim and Joanne

Acknowledgements

To start I would like to acknowledge my advisor, Jennifer Brodbelt, for giving me the opportunity to earn my PhD in your lab. I feel fortunate to have had access to state of the art instrumentation and to have been granted freedom in pursuing my research interests. Your willingness to initiate correspondence with other experts in the field on my behalf helped me overcome many road blocks in my research and gave me access to several interesting collaborations.

In particular, I am indebted to Professor Joshua Coon and his graduate student Catie Minogue for hosting me at the University of Wisconsin and providing hands on training for IMAC phosphopeptide enrichment. Also to Daniel Boutz, who showed me the process of cell culture and helped to grow enough HeLa to last through my many initial attempts at phosphoproteomics. The data analysis to determine phosphate retention was greatly aided by code written by Joyce Ho. Thanks for letting me distract you from your own dissertation. Thank you to Dr. Kevin Dalby and his graduate student Juliana Taliaferro for donating HCC70 cell lysates and to Dr. Jessie Zhang and her graduate student Josh Mayfield for providing CTD samples. To both groups, thank you for lending your expertise regarding protein phosphorylation and for providing interesting applications for my work.

Thank you to my labmates in the Brodbelt group over the years, especially Victoria Cotham and Dustin Klein. I hope you know how much your friendship has helped me during these last six years. Thanks for always coming through with useful suggestions and never letting me get too down on myself at times when research was slow to progress.

I have also been blessed with an incredible support system of friends outside of the lab. Special thanks to my roommates over the years including Jenny Knipe, Amanda Paine, and Rachel James for standing in as my Austin family. Thanks for the fun times and memories.

Finally I am so grateful to my family for all their love and endless support during my time in graduate school. You gave me the push that I needed to embark on my Austin adventure and have been instrumental in my success ever since. Mom, Dad, Chris, and Shannon, I could not have done this without you.

Advancement of Photodissociation Mass Spectrometry Methods for the Analysis of Protein Post-translation Modifications

Michelle Renee Robinson, Ph.D.

The University of Texas at Austin, 2016

Supervisor: Jennifer S. Brodbelt

Post-translational modifications (PTMs) are important for regulating protein structure and function. Despite significant progress for PTM analysis using liquid chromatography tandem mass spectrometry (LC-MS/MS), opportunities for new method development remain. The research presented in this dissertation promotes 193 nm ultraviolet photodissociation (UVPD) as an alternative activation technique for PTM analysis with specific utility for phosphorylated and sulfated peptides.

A novel *de novo* sequencing method with applications for unbiased PTM discovery was developed utilizing Lys-N proteolysis, N-terminal imidazolinylation, and UVPD to direct fragmentation for the formation of N-terminal ions. The N-terminal *a*, *b*, and *c* ions generated by UVPD were differentiated from one another by characteristic mass shifts. Sets of triplet peaks were used to distinguish N-terminal ions from confounding C-terminal ions and improve the accuracy of *de novo* sequencing.

UVPD was evaluated for the analysis of phosphopeptide cations and anions. Negative mode analysis was advantageous for the detection of casein peptides in high phosphorylation states, while positive mode proved more robust for global phosphoproteomic analysis of HeLa and HCC70 cell lysates. Compared to collisional

activation, the depth of coverage was lower using UVPD yet more extensive fragmentation and improved phosphate retention on products ions was achieved.

Phosphorylation mapping by LC-UVPD-MS was carried out in the C-terminal domain (CTD) of RNA polymerase II as a function of kinase treatment, ERK2 or TFIIH, and organism, yeast or fruit fly. Single phosphorylations on Ser2 or Ser5 in the consensus heptad, YSPTSPS, were observed across all experimental conditions. Analysis of the non-consensus fruit fly CTD revealed the significance of Tyr1 and Pro residues in the +1 position relative to Ser for phosphorylation to occur.

For sulfated peptides, negative mode UVPD yielded *a* and *x* ions that largely retained the labile sulfate modification which facilitated peptide sequencing and PTM localization. With appropriate MS/MS tools established, the next step towards global sulfoproteomics was the development of enrichment methods. Weak anion exchange (WAX) was applied for this purpose. Following carbamylation to neutralize primary amines which otherwise repel the anion exchanger; improved WAX retention was observed for sulfopeptides relative to a complex mixture of unmodified bovine serum albumin peptides.

Table of Contents

Chapter 1

Introduction

1.1 INTRODUCTION

Proteomics is the comprehensive study of an organism's proteins along with their related structure and function.[1–3] Both qualitative and quantitative information is required for complete proteome characterization which presents a significant challenge due to the high dynamic range and overall number of proteins within the proteome which far exceeds the number of protein coding genes based on splice variants, polymorphisms, and post-translational modifications (PTMs).[4] Beyond simply increasing the complexity of the proteome, PTMs are critical for regulating nearly all aspects of normal biological function from cellular differentiation, protein turnover and localization, protein-protein interactions, signaling cascades and DNA repair to protein degradation. Aberrant PTM expression has been linked to various diseases, and efforts to identify new biomarkers or therapeutic targets are ongoing.[5]

Liquid chromatography tandem mass spectrometry (LC-MS/MS) based techniques have been key for the advancement of modification-specific proteomics, but continued method development is required to address the unique challenges of PTM analysis.[6–10] While modified proteins are indeed ubiquitous, with more than 5% of the genes in the human genome encoding enzymes that are responsible for adding or removing the myriad of known PTMs, they are also transient and in low abundance owing to their function for dynamic protein regulation.[11,12] An added layer of complexity arises when specific protein functionality is dependent on the interplay or cross talk between multiple PTMs, as is the case for histone modifications.[13–17] Finally, the diverse array of chemical and physical properties that are associated with the ever growing

collection of PTMs continues to drive the development of new LC-MS/MS methodologies.

Two PTMs, phosphorylation[18] and sulfation[19] which are chief regulators of intracellular and extracellular protein-protein interactions, respectively, exemplify the need for improved characterization techniques. Both PTMs are labile, and phosphorylations are typically removed during tandem mass spectrometry (MS2), while the more labile sulfations are stripped upon both MS1 and MS2 analysis.[20,21] Sulfation analysis also suffers from a lack of effective methods for enrichment from biological matrices. The research presented in this dissertation aims to advance 193 nm ultraviolet photodissociation (UVPD) as an alternative activation method for improved identification and localization of PTMs with particular emphasis towards phosphorylation and sulfation.

1.2 BOTTOM-UP/SHOTGUN PROTEOMICS

The most widely used strategy for mass spectrometry-based protein analysis is the bottom-up LC-MS/MS method in which proteins are enzymatically digested into smaller peptides and chromatographically separated prior to online MS^n analysis. During the first stage of mass analysis (MS1), a full MS survey spectrum is acquired which shows all peptides at a given elution time point. Peptide ions from the MS1 spectrum are subsequently selected for MS2 in a data dependent manner, typically from most to least abundant, and activated to promote their fragmentation into diagnostic product ions which can be used to reconstruct the peptide sequence either by reference to an *in silico* database or by *de novo* sequencing. Database search is the more common approach, and many algorithms have been developed such as SEQUEST,[22] Mascot,[23] MassMatrix,[24]

OMSSA,[25] X!Tandem,[26] Byonic,[27] MaxQuant,[28] and MS Amanda[29] to automate peptide and ultimately protein identifications. Advanced informatics platforms are particularly critical for shotgun proteomics applications which profile very complex mixtures of proteins collected from crude tissues or whole cell lysates. The general bottom-up/shotgun proteomics workflow is illustrated in Figure 1.1.

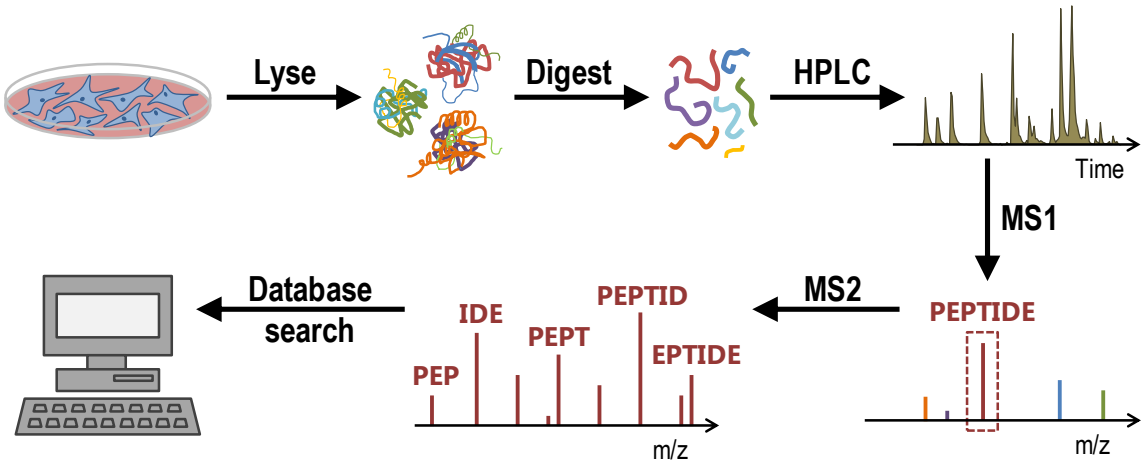


Figure 1.1 Workflow for bottom-up/shotgun proteomics

Advances in separation and mass spectrometry instrumentation have also greatly facilitated shotgun proteomics, and the identification of thousands of peptides and proteins from a single LC-MS/MS run is not only possible but exceedingly routine.[30–32] The advent of high resolution accurate mass (HRAM) mass spectrometers, most notably the Orbitrap,[33,34] has been particularly transformative, enabling better distinction between co-eluting peptides of similar *m/z*, better charge state determination, and more accurate quantitation.[35] These benefits of HRAM technologies have lead to significant performance gains for database search algorithms in terms of both the search speed and the accuracy of peptide identifications by narrowing the number of potential matches to a

given MS/MS spectrum through the use of tighter mass tolerance settings for precursor and/or product ions. The development of faster mass spectrometers with higher MS/MS acquisition rates and varying degrees of parallelization such as the Q-Exactive HF,[36] Orbitrap Fusion,[37] and Bruker Impact II Q-TOF,[38] has also improved the depth of coverage that is possible in shotgun proteomics. At an acquisition speed of 20 Hz, sampling 19 peptides per second, comprehensive characterization of the relatively simple yeast proteome was achieved in just over one hour.[39] Also, Mann and coworkers recently reported the largest breast tumor proteomic data set to date with a total of 10,135 identified proteins from 40 breast cancer samples in which more than 7,000 proteins were identified per sample.[40]

1.2.1 Modification Specific Proteomics

The general bottom-up/shotgun proteomics workflow described above may also be applied for the analysis of PTMs.[6,10] Because modified peptides are in low stoichiometric abundance in biological samples relative to unmodified peptides, incorporating an enrichment step prior to LC-MS/MS analysis at either the peptide or protein level can improve the depth of coverage for a particular modification of interest. Phosphorylation,[41,42,21] glycosylation,[43,44] and acetylation[45] have all benefited from enrichment by means of acid-base interactions, metal ion affinity, ion exchange, hydrophilic interaction, lectins, or antibodies, leading to the identification of thousands of modified peptides. Serial enrichment has also been reported in order to profile multiple classes of PTM from a single biological sample.[46]

Additionally, LC-MS/MS methods can be tailored to address the specific physicochemical properties of different modifications. This is especially advantageous

for the detection of peptides modified with labile PTMs for which conventional MS/MS strategies are less effective. For example, using electron transfer dissociation instead of the more commonly employed collisional dissociation results in improved modification retention for both phosphorylated[47–51] and glycosylated peptides.[52,53] Advanced multistage activation or decision tree programs, in which an MS3 event is triggered by the detection of MS2 generated characteristic neutral loss or reporter ions, have further optimized data dependant acquisition methods for PTM analysis.[54–57]

Bioinformatics tools have also been developed to specifically aid PTM characterization. For example, phosphorylation sites may be scored and localized using a variety of programs including Ascore,[58] SLoMo,[59] phosphoRS,[60] and LuciPHOr.[61] Software has also been developed that attempts to automate structural characterization of complicated N-linked glycopeptides.[62] There are also algorithms for predicting potential modification sites within proteins based on sequence motifs, such as the Sulfinator[63] and PredSulSite[64] for sulfation, which can provide a framework for selecting potential systems of study as well as a means for validating newly identified modification sites.

1.2.2 *De novo* Sequencing

For studies that are not directed towards a certain modification, *de novo* sequencing can provide the ultimate flexibility for PTM discovery because peptide sequences are derived without reference to a database.[65] Instead, the MS/MS fragmentation pattern is used to determine the peptide sequence based on mass differences between product ions that account for sequential cleavages along the peptide backbone.[66,67] In this way, all potential PTMs can be considered and appropriately assigned within peptides based on their characteristic mass additions without increasing

the search space or search time. Additionally novel or unexpected PTMs may be identified by *de novo* sequencing. A number of programs have been developed for automated *de novo* sequencing including PEAKS,[68] PepNovo,[69] NovoHMM,[70] MSnovo,[71] and Vonode,[72] yet *de novo* sequencing is less often used compared to database search methods. Differentiating confounding product ion series and poorly resolved product ions as well as inferring the amino acid composition in regions where sequence coverage is lacking present significant challenges for *de novo* sequencing even with advanced bioinformatics programs in place.[73,74] Isotopic labeling at one terminus of the peptide provides an effective means for differentiating product ion series.[75–78] Other methods have been developed to specifically bias fragmentation for a certain ion series in order to simplify spectral interpretation.[79–81] Although many of these strategies have improved *de novo* sequencing efforts, their success for PTM identification hinges on modification stability during ion activation and dissociation which is problematic for many labile PTMs. Also, without enrichment only the most abundant PTMs present in a biological sample will be detected which precludes *de novo* sequencing from global PTM analysis.

1.3 TANDEM MASS SPECTROMETRY NOMENCLATURE

The success of shotgun proteomics experiments using both *de novo* and database search methods is highly dependent on the ion activation technique that is used for peptide fragmentation. This is especially true for peptides that carry PTMs because extensive fragmentation is required to accurately pinpoint the modification to a single amino acid residue. Additionally for labile PTMs, fragmentation methods must be carefully selected to mitigate modification loss from product ions which would otherwise

prevent confident site localization. In order to accurately determine a peptide sequence from corresponding MS/MS data, the fragmentation pattern must be predicable and reproducible. To achieve this goal, dissociation methods have been developed that largely restrict fragmentation to specific bonds along the peptide backbone, and nomenclature has been established to classify the different types of fragment ions.[82,83] Figure 1.2 shows the fragmentation nomenclature and a representative set of product ions from a tetrapeptide in which the amino acid side chains are denoted as R_1-R_4.

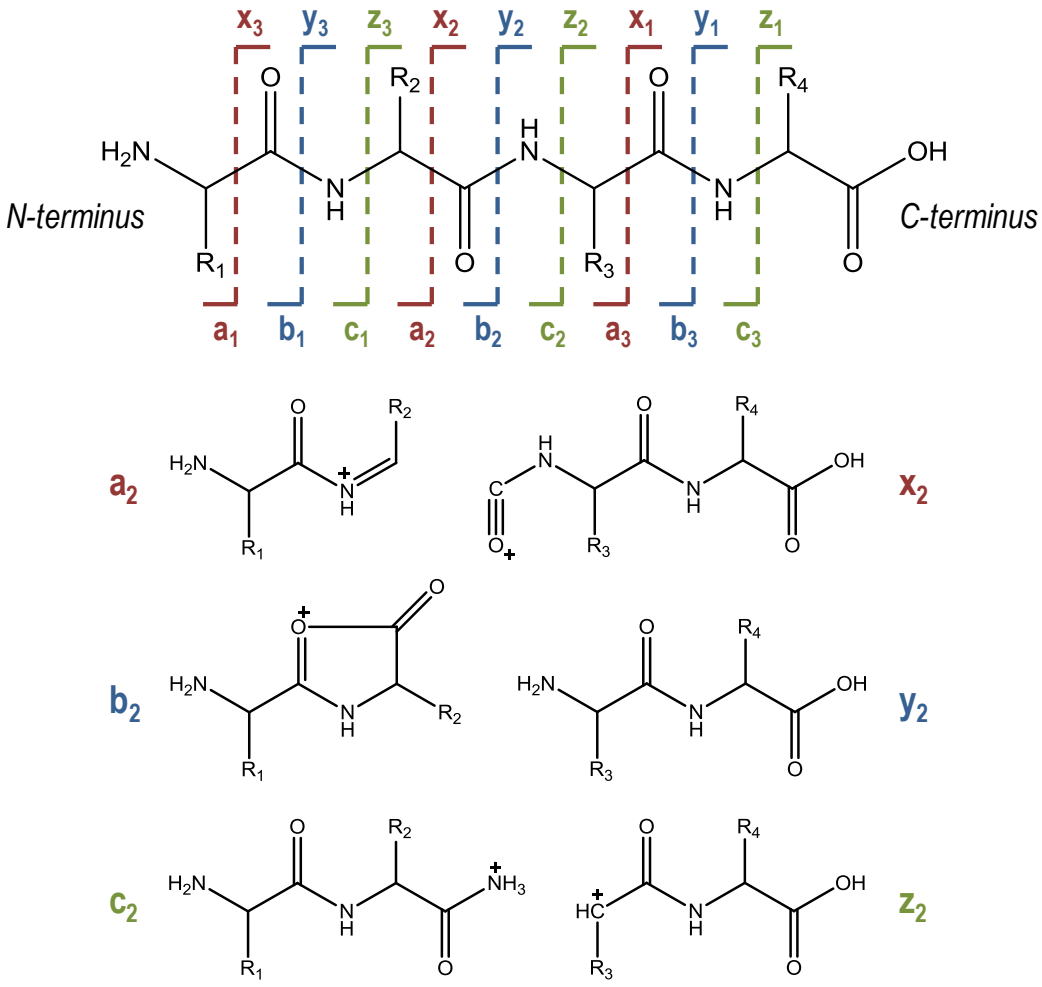


Figure 1.2 Peptide/protein fragmentation nomenclature and representative product ions

All peptides have the same core structure that is composed of three repeating bonds: C_α-C, C-N, and N-C_α (where C refers to the carbonyl carbon), and cleavage of these bonds produces complementary pairs of *a*/*x*, *b*/*y*, and *c*/*z* product ions, respectively. Ions that contain the N-terminus include *a*, *b*, and *c* ions, while *x*, *y*, and *z* ions contain the C-terminus. The number associated with each product ion indicates the number of amino acids that are present between the point of cleavage and the peptide terminus. Amino acid specific side chains branch off from the peptide backbone at the α-carbon, and thus successive cleavages along the backbone can be used to determine both the identity and connectivity of the amino acids that make up the peptide based on the characteristic mass of each side chain. Different activation techniques promote cleavage at different bonds in the peptide backbone.[84] Cleavage of the C-N amide bond to form *b* and *y* ions is the most common dissociation pathway. This is because the amide bond is the weakest bond in the peptide backbone and may be readily cleaved using slow heating activation methods such as collisional dissociation (CID, HCD) or infrared multiphoton dissociation (IRMPD). Alternatively, electron-based activation methods, which include electron transfer dissociation (ETD)[85] and electron capture dissociation (ECD),[86–88] generate *c* and *z* ions from cleavage at the N-C_α bond. The C_α-C bond may be cleaved using higher energy activation methods such as ultraviolet photodissociation, to form *a* and *x* type ions.[89]

Cleavage at more than one bond in the peptide backbone can lead to the formation of internal ions which lack both the N- and C-terminus, and these ions are annotated with the amino acid letter codes that correspond to the side chains that are included. Immonium ions are internal ions that contain only a single amino acid side chain. Other dissociation pathways observed for peptide ions include the neutral loss of small molecules such as water or ammonia as well as the loss of amino acid side chains which give rise to *d*, *w*, and *v* type satellite ions.[90,91] These amino acid side chain losses, pictured

in Figure 1.3, are formed from *a* and *x* type ions and can be useful for differentiating isobaric leucine and isoleucine.

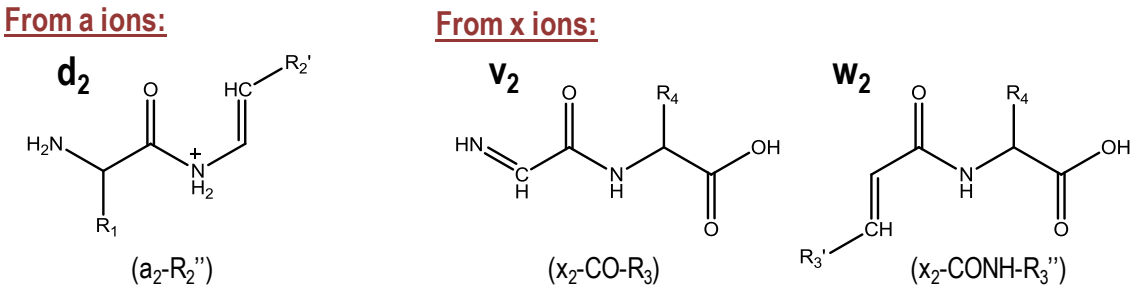


Figure 1.3 Side chain loss ions

1.3.1 Collision Induced Dissociation

Collision induced dissociation (CID) is the method of choice for peptide sequencing and comes standard on nearly all commercial mass spectrometers. Even as new activation methods are developed, CID remains popular for its ease of implementation and robust performance. During CID, peptides gain energy through multiple collisions with inert neutral gas molecules like helium, nitrogen, or argon until enough internal energy is acquired to break peptide bonds and produce *b* and *y* ions.[92]

Extensive mechanistic studies have helped to characterize the CID process, ultimately leading to the widely accepted mobile proton model in which backbone fragmentation proceeds through a charge site initiated mechanism.[93,94] Adequate proton mobility is key for efficient CID fragmentation and if the number of charge sequestering sites exceeds the number of ionizing protons, fragmentation becomes dictated by charge remote pathways in which preferential cleavage occurs C-terminal to acidic aspartic and glutamic acid residues.[95,96] Even under conditions of high proton mobility, other

preferential cleavages are observed using CID, most notably N-terminal to proline residues, which can suppress fragmentation at other sites along the peptide backbone and lower the overall peptide sequence coverage.[96–98]

Because cleavage occurs at the most labile bonds using CID, peptides modified with labile PTMs such as phosphorylation,[99,100] glycosylation,[52] and sulfation[101,102] undergo abundant undesirable modification loss from both precursor and product ions. CID in ion trap mass spectrometers is also limited by what is called the low mass cutoff (LMCO) in which the trajectories of low *m/z* ions become destabilized at the increased rf voltages applied for collisional activation leading to the removal of potentially informative product ions. A beam type variant of CID called higher energy collisional dissociation (HCD) overcomes the low mass cutoff problem because activation is independent of trapping parameters.[103,104] This feature is particularly advantageous for PTMs whose identification is aided by the detection of low mass reporter ions.[56,105] The higher energy deposition of HCD provides added utility for PTM analysis because cleavage is possible at bonds other than those that are most labile.[106]

1.3.2 Ultraviolet Photodissociation at 193 nm

Ultraviolet photodissociation (UVPD) is gaining popularity for peptide sequencing, and has been applied using a number of different excitation wavelengths for a growing number of proteomic applications.[89] The peptide backbone directly absorbs 157 nm and 193 nm photons, making these wavelengths the most widely applicable for peptide analysis. Absorption of a 157 nm or 193 nm photon promotes an electron into the excited state. Subsequent dissociation from excited electronic states leads to extensive fragmentation along the peptide backbone which produces all product ion types including

a, *b*, *c*, *x*, *y*, and *z* ions. Radial *a* and *x* ions (*a*• and *x*•) that are 1 Da heavier than conventional *a* and *x* ions are also observed.[91] These are formed by homolytic cleavage of the C_α-C bond and loss of a radical proton converts these ions to *a* and *x* ions as shown in Figure 1.4.

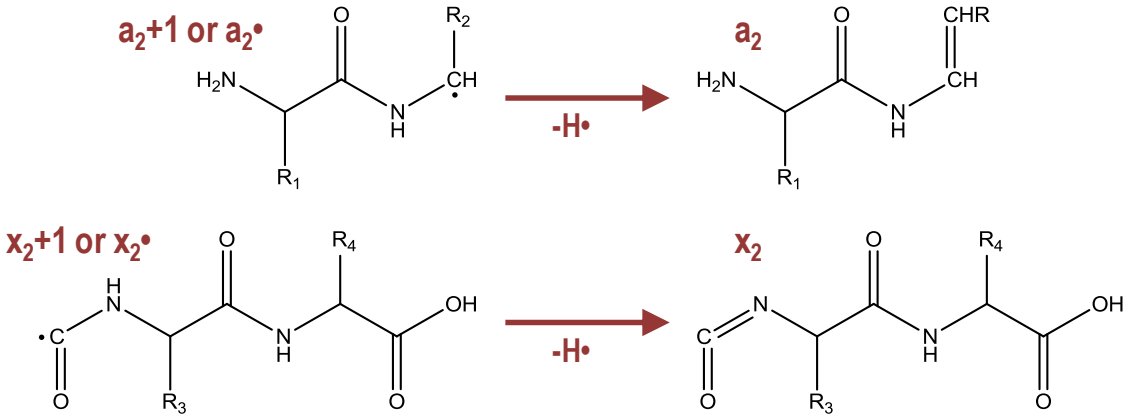


Figure 1.4 Formation of a+1/x+1 and a/x ions upon 193 and 157 nm UVPD

Because fragmentation is not directed by mobile protons, the efficiency of UVPD depends less on the peptide charge state compared to CID and preferential cleavages are less abundant. UVPD may also be applied for the analysis of peptide anions, which are less effectively characterized using CID.[107,108] The negative mode sequencing capabilities of UVPD have been used to identify and localize sulfation,[109,110] O-glycosylation,[111] and phosphorylation[112] PTMs which are all more stable in peptide anions.

These advantages of UVPD have prompted its implementation on a growing number of high performance mass spectrometers beyond the traditionally used ion trap and time of flight (TOF) instruments. Various Orbitrap instruments such as the Orbitrap Elite,[113] Orbitrap Q-Exactive,[114] and Orbitrap Fusion Tribid[115] have been equipped for

193 nm UVPD, and most recently photodissociation has been reported on a 15T FT-ICR (Fourier transform-ion cyclotron resonance) mass spectrometer.[116]

1.4 Overview of Chapters

Despite great strides in many aspects of modification specific proteomics, opportunities for continued method development remain particularly in the area of tandem mass spectrometry. There is currently no single MS/MS method that is universally applicable for sequencing all types of peptides, thus fragmentation must be tailored for specific applications. The research presented in this dissertation is aimed at improving the characterization of PTMs including phosphorylation and sulfation using 193 nm UVPD.

In chapter 3, Lys-N proteolysis, imidazolinylation of the resulting N-terminal lysine ε-amine, and UVPD were combined to generate simplified MS/MS spectra for improved *de novo* sequencing. For imidazolinylated Lys-N peptides, positive charges were effectively sequestered at the N-terminus which biased fragmentation for enhanced formation of N-terminal product ions while suppressing the formation of C-terminal ions. Using UVPD, all N-terminal ions were generated including *a*, *b*, and *c* ions, and their appearance as sets of regularly spaced triplet peaks further aided the distinction between N- and C-terminal ion series, leading to improved accuracy in de novo sequencing.

More directed PTM analysis was carried out in chapter 4 in which UVPD was evaluated for the analysis of phosphopeptides in both positive and negative ion modes. Negative ion mode offered the best phosphate retention; however, low sensitivity limited its application to peptides in high phosphorylation states that were otherwise not detected using positive mode. Positive UVPD was applied for phosphoproteomic analysis of HeLa and HCC70 cell lysates and benchmarked against the more conventional HCD activation.

More peptides and proteins were identified using HCD, but better phosphate retention on product ions was achieved using UVPD which can facilitate more confident phospho-site localization.

UVPD was next applied in chapter 5 to map phosphorylation sites within the C-terminal domain (CTD) of RNA polymerase II. Alternative proteolysis using proteinase K or chymotrypsin was required to digest the CTD which is composed of repeating units of the seven amino acid consensus sequence: YSPTSPS. Two different CTDs including the highly consensus yeast (*saccharomyces cerevisiae*) and highly divergent fruit fly (*drosophila melanogaster*) were analyzed following treatment with TFIIH and/or ERK2 kinases. Preferential phosphorylation was detected on Ser5 and Ser2. Deviations from the consensus sequence within the fruit fly CTD revealed the importance of certain amino acid residues and their relative positions within the heptad sequence, namely Tyr and Pro, for regulating phosphorylation events.

Sulfated peptides were analyzed using UVPD in chapter 6. Analysis was undertaken in the negative ion mode to maximize the stability of the labile sulfate modification on precursor and product ions, thus permitting accurate peptide sequencing and sulfate localization. The sulfate modification was stable on product ions over a range of UVPD conditions (laser energy and number of pulses), but considerable sulfate loss was observed from the undissociated precursor ion. This characteristic fragmentation behavior was ultimately useful for distinguishing sulfated peptides from similarly modified phosphorylated peptides. LC-UVPD-MS was applied for the analysis of bovine fibrinogen, a 340 kDa heterohexamer, and the sulfated peptide of interest was positively identified.

In chapter 7, weak anion exchange chromatography (WAX) was developed as an enrichment method for sulfated peptides to extend the negative LC-UVPD-MS strategy

established in chapter 6 for sulfoproteomic analysis on a global scale. Sulfopeptides were carbamylated to convert primary amines to less basic carbamates and thus improve their retention on WAX by the removal of interfering positive charges. Following carbamylation the desired shift to longer retention was observed for model sulfopeptides, which facilitated their separation from a matrix of unmodified peptides from bovine serum albumin.

1.5 REFERENCES

(1) Domon, B.; Aebersold, R. *Science* **2006**, *312* (5771), 212–217.
(2) Cox, J.; Mann, M. *Annu. Rev. Biochem.* **2011**, *80* (1), 273–299.
(3) Bensimon, A.; Heck, A. J. R.; Aebersold, R. *Annu. Rev. Biochem.* **2012**, *81* (1), 379–405.
(4) Zubarev, R. A. *PROTEOMICS* **2013**, *13* (5), 723–726.
(5) Karve, T. M.; Cheema, A. K. *J. Amino Acids* **2011**, *2011*, 1–13.
(6) Mann, M.; Jensen, O. N. *Nat Biotech* **2003**, *21* (3), 255–261.
(7) Nørregaard Jensen, O. *Curr. Opin. Chem. Biol.* **2004**, *8* (1), 33–41.
(8) Reinders, J.; Lewandrowski, U.; Moebius, J.; Wagner, Y.; Sickmann, A. *PROTEOMICS* **2004**, *4* (12), 3686–3703.
(9) Silva, A. M. N.; Vitorino, R.; Domingues, M. R. M.; Spickett, C. M.; Domingues, P. *Free Radic. Biol. Med.* **2013**, *65*, 925–941.
(10) Olsen, J. V.; Mann, M. *Mol. Cell. Proteomics* **2013**, *12* (12), 3444–3452.
(11) Walsh, C. *Posttranslational modification of proteins: expanding nature's inventory*; Roberts and Co. Publishers: Englewood, Colo, 2006.
(12) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. *Sci. Rep.* **2011**, *1*.
(13) Jensen, O. N. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (6), 391–403.
(14) Webb, K.; Bennett, E. J. *Nat. Methods* **2013**, *10* (7), 620–621.
(15) Huang, Y.; Xu, B.; Zhou, X.; Li, Y.; Lu, M.; Jiang, R.; Li, T. *Mol. Cell. Proteomics* **2015**, *14* (3), 761–770.
(16) Gajadhar, A. S.; White, F. M. *Curr. Opin. Biotechnol.* **2014**, *28*, 83–87.
(17) Wang, G. G.; Allis, C. D.; Chi, P. *Trends Mol. Med.* **2007**, *13* (9), 363–372.
(18) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. *Cell* **2006**, *127* (3), 635–648.
(19) Moore, K. L. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (35), 14741–14742.
(20) Monigatti, F.; Hekking, B.; Steen, H. *Posttranslational Modif. Proteomics* **2006**, *1764* (12), 1904–1913.
(21) Riley, N. M.; Coon, J. J. *Anal. Chem.* **2015**.
(22) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
(23) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20* (18), 3551–3567.
(24) Xu, H.; Freitas, M. A. *PROTEOMICS* **2009**, *9* (6), 1548–1555.
(25) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. *J. Proteome Res.* **2004**, *3* (5), 958–964.
(26) Craig, R.; Cortens, J. P.; Beavis, R. C. *J. Proteome Res.* **2004**, *3* (6), 1234–1242.
(27) Bern, M.; Kil, Y. J.; Becker, C. In *Current Protocols in Bioinformatics*; Baxevanis, A. D., Petsko, G. A., Stein, L. D., Stormo, G. D., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2012.
(28) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26* (12), 1367–1372.

(29) Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. *J. Proteome Res.* **2014**, *13* (8), 3679–3684.
(30) Yates, J. R. *J. Am. Chem. Soc.* **2013**, *135* (5), 1629–1640.
(31) Richards, A. L.; Merrill, A. E.; Coon, J. J. *Curr. Opin. Chem. Biol.* **2015**, *24*, 11–17.
(32) Mayne, J.; Ning, Z.; Zhang, X.; Starr, A. E.; Chen, R.; Deeke, S.; Chiang, C.-K.; Xu, B.; Wen, M.; Cheng, K.; Seebun, D.; Star, A.; Moore, J. I.; Figeys, D. *Anal. Chem.* **2016**, *88* (1), 95–121.
(33) Hardman, M.; Makarov, A. A. *Anal. Chem.* **2003**, *75* (7), 1699–1705.
(34) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *J. Mass Spectrom.* **2005**, *40* (4), 430–443.
(35) Mann, M.; Kelleher, N. L. *Proc. Natl. Acad. Sci.* **2008**, *105* (47), 18132–18138.
(36) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. *Mol. Cell. Proteomics* **2014**, *13* (12), 3698–3708.
(37) Senko, M. W.; Remes, P. M.; Canterbury, J. D.; Mathur, R.; Song, Q.; Eliuk, S. M.; Mullen, C.; Earley, L.; Hardman, M.; Blethrow, J. D.; Bui, H.; Specht, A.; Lange, O.; Denisov, E.; Makarov, A.; Horning, S.; Zabrouskov, V. *Anal. Chem.* **2013**, *85* (24), 11710–11714.
(38) Beck, S.; Michalski, A.; Raether, O.; Lubeck, M.; Kaspar, S.; Goedecke, N.; Baessmann, C.; Hornburg, D.; Meier, F.; Paron, I.; Kulak, N. A.; Cox, J.; Mann, M. *Mol. Cell. Proteomics* **2015**, *14* (7), 2014–2029.
(39) Richards, A. L.; Hebert, A. S.; Ulbrich, A.; Bailey, D. J.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. *Nat Protoc.* **2015**, *10* (5), 701–714.
(40) Tyanova, S.; Albrechtsen, R.; Kronqvist, P.; Cox, J.; Mann, M.; Geiger, T. *Nat. Commun.* **2016**, *7*, 10259.
(41) Thingholm, T. E.; Jensen, O. N.; Larsen, M. R. *PROTEOMICS* **2009**, *9* (6), 1451–1468.
(42) Macek, B.; Mann, M.; Olsen, J. V. *Annu. Rev. Pharmacol. Toxicol.* **2009**, *49* (1), 199–221.
(43) Ongay, S.; Boichenko, A.; Govorukhina, N.; Bischoff, R. *J. Sep. Sci.* **2012**, *35* (18), 2341–2372.
(44) Thaysen-Andersen, M.; Packer, N. H. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2014**, *1844* (9), 1437–1452.
(45) Zhang, K.; Tian, S.; Fan, E. *The Analyst* **2013**, *138* (6), 1628.
(46) Mertins, P.; Qiao, J. W.; Patel, J.; Udeshi, N. D.; Clauser, K. R.; Mani, D. R.; Burgess, M. W.; Gillette, M. A.; Jaffe, J. D.; Carr, S. A. *Nat. Methods* **2013**, *10* (7), 634–637.
(47) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. *Rapid Commun. Mass Spectrom.* **2000**, *14* (19), 1793–1800.
(48) Sweet, S. M. M.; Bailey, C. M.; Cunningham, D. L.; Heath, J. K.; Cooper, H. J. *Mol. Cell. Proteomics* **2009**, *8* (5), 904–912.
(49) Shi, S. D.-H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W. *Anal. Chem.* **2001**, *73* (1), 19–22.

(50) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proc. Natl. Acad. Sci.* **2007**, *104* (7), 2199–2204.
(51) Creese, A.; Cooper, H. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (9), 1263–1274.
(52) Hogan, J. M.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *J. Proteome Res.* **2005**, *4* (2), 628–632.
(53) Zhu, Z.; Su, X.; Clark, D. F.; Go, E. P.; Desaire, H. *Anal. Chem.* **2013**, *85* (17), 8403–8411.
(54) Villén, J.; Beausoleil, S. A.; Gygi, S. P. *PROTEOMICS* **2008**, *8* (21), 4444–4452.
(55) Ulintz, P. J.; Yocum, A. K.; Bodenmiller, B.; Aebersold, R.; Andrews, P. C.; Nesvizhskii, A. I. *J. Proteome Res.* **2009**, *8* (2), 887–899.
(56) Saba, J.; Dutta, S.; Hemenway, E.; Viner, R. *Int. J. Proteomics* **2012**, *2012*, 1–7.
(57) Wu, S.-W.; Pu, T.-H.; Viner, R.; Khoo, K.-H. *Anal. Chem.* **2014**, *86* (11), 5478–5486.
(58) Beausoleil, S. A.; Villén, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. *Nat. Biotechnol.* **2006**, *24* (10), 1285–1292.
(59) Bailey, C. M.; Sweet, S. M. M.; Cunningham, D. L.; Zeller, M.; Heath, J. K.; Cooper, H. J. *J. Proteome Res.* **2009**, *8* (4), 1965–1971.
(60) Taus, T.; Köcher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. *J. Proteome Res.* **2011**, *10* (12), 5354–5362.
(61) Fermin, D.; Walmsley, S. J.; Gingras, A.-C.; Choi, H.; Nesvizhskii, A. I. *Mol. Cell. Proteomics* **2013**, *12* (11), 3409–3419.
(62) Dallas, D. C.; Martin, W. F.; Hua, S.; German, J. B. *Brief. Bioinform.* **2013**, *14* (3), 361–374.
(63) Monigatti, F.; Gasteiger, E.; Bairoch, A.; Jung, E. *Bioinformatics* **2002**, *18* (5), 769–770.
(64) Huang, S.-Y.; Shi, S.-P.; Qiu, J.-D.; Sun, X.-Y.; Suo, S.-B.; Liang, R.-P. *Anal. Biochem.* **2012**, *428* (1), 16–23.
(65) Ma, B.; Johnson, R. *Mol. Cell. Proteomics* **2012**, *11* (2), O111.014902–O111.014902.
(66) Hughes, C.; Ma, B.; Lajoie, G. A. In *Proteome Bioinformatics*; Hubbard, S. J., Jones, A. R., Eds.; Humana Press: Totowa, NJ, 2010; Vol. 604, pp 105–121.
(67) Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. *PROTEOMICS* **2010**, *10* (4), 634–649.
(68) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.
(69) Frank, A.; Pevzner, P. *Anal. Chem.* **2005**, *77* (4), 964–973.
(70) Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. *Anal. Chem.* **2005**, *77* (22), 7265–7273.
(71) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. *Anal. Chem.* **2007**, *79* (13), 4870–4878.
(72) Pan, C.; Park, B. H.; McDonald, W. H.; Carey, P. A.; Banfield, J. F.; VerBerkmoes, N. C.; Hettich, R. L.; Samatova, N. F. *BMC Bioinformatics* **2010**, *11* (1), 118.
(73) Song, Y.; Yu, M. *Inf. Process. Lett.* **2015**, *115* (2), 377–381.

(74) Chi, H.; Chen, H.; He, K.; Wu, L.; Yang, B.; Sun, R.-X.; Liu, J.; Zeng, W.-F.; Song, C.-Q.; He, S.-M.; Dong, M.-Q. *J. Proteome Res.* **2013**, *12* (2), 615–625.
(75) Münchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem.* **2000**, *72* (17), 4047–4057.
(76) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. *Anal. Chem.* **2005**, *77* (19), 6300–6309.
(77) Hennrich, M.; Mohammed, S.; Altelaar, A.; Heck, A. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965 – 1965.
(78) Cagney, G.; Emili, A. *Nat Biotech* **2002**, *20* (2), 163–170.
(79) Robotham, S. A.; Kluwe, C.; Cannon, J. R.; Ellington, A.; Brodbelt, J. S. *Anal. Chem.* **2013**, *85* (20), 9832–9838.
(80) Miyashita, M.; Hanai, Y.; Awane, H.; Yoshikawa, T.; Miyagawa, H. *Rapid Commun. Mass Spectrom.* **2011**, *25* (9), 1130–1140.
(81) Madsen, J. A.; Brodbelt, J. S. *Anal. Chem.* **2009**, *81* (9), 3645–3653.
(82) Roepstorff, P. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601.
(83) Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. *Anal. Chem.* **1987**, *59* (21), 2621–2625.
(84) Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (1), 30–51.
(85) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci.* **2004**, *101* (26), 9528–9533.
(86) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.
(87) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72* (3), 563–573.
(88) Zubarev, R. A. *Curr. Opin. Biotechnol.* **2004**, *15* (1), 12–16.
(89) Brodbelt, J. S. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.
(90) Johnson, R. S.; Martin, S. A.; Biemann, K. *Int. J. Mass Spectrom. Ion Process.* **1988**, *86*, 137–154.
(91) Weidong Cui; Thompson, M. S.; Reilly, J. P. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1384–1398.
(92) McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **1992**, *3* (6), 599–614.
(93) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118* (35), 8365–8374.
(94) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35* (12), 1399–1406.
(95) Gu, C.; Tsaprailis, G.; Breci, L.; Wysocki, V. H. *Anal Chem* **2000**, *72* (23), 5804–5813.
(96) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. *Anal. Chem.* **2005**, *77* (18), 5800–5813.
(97) Breci, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H. *Anal. Chem.* **2003**, *75* (9), 1963–1971.
(98) Vaisar, T.; Urban, J. *J. Mass Spectrom.* **1996**, *31* (10), 1185–1187.

(99) Palumbo, A. M.; Smith, S. A.; Kalcic, C. L.; Dantus, M.; Stemmer, P. M.; Reid, G. E. *Mass Spectrom. Rev.* **2011**, *30* (4), 600–625.
(100) Brown, R.; Stuart, S.; Houel, S.; Ahn, N.; Old, W. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (7), 1128–1142.
(101) Nemeth-Cawley, J. F.; Karnik, S.; Rouse, J. C. *J. Mass Spectrom.* **2001**, *36* (12), 1301–1311.
(102) Wolfender, J.-L.; Chu, F.; Ball, H.; Wolfender, F.; Fainzilber, M.; Baldwin, M. A.; Burlingame, A. L. *J. Mass Spectrom.* **1999**, *34* (4), 447–454.
(103) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nat. Methods* **2007**, *4* (9), 709–712.
(104) McAlister, G. C.; Phanstiel, D. H.; Brumbaugh, J.; Westphall, M. S.; Coon, J. J. *Mol. Cell. Proteomics* **2011**, *10* (5), O111.009456–O111.009456.
(105) Henriksen, P.; Wagner, S. A.; Weinert, B. T.; Sharma, S.; Bacinskaja, G.; Rehman, M.; Juffer, A. H.; Walther, T. C.; Lisby, M.; Choudhary, C. *Mol. Cell. Proteomics* **2012**, *11* (11), 1510–1522.
(106) Nagaraj, N.; D'Souza, R. C. J.; Cox, J.; Olsen, J. V.; Mann, M. *J. Proteome Res.* **2010**, *9* (12), 6786–6794.
(107) Madsen, J. A.; Xu, H.; Robinson, M. R.; Horton, A. P.; Shaw, J. B.; Giles, D. K.; Kaoud, T. S.; Dalby, K. N.; Trent, M. S.; Brodbelt, J. S. *Mol. Cell. Proteomics* **2013**, *12* (9), 2604–2614.
(108) Waugh, R. J.; Bowie, J. H. *Rapid Commun. Mass Spectrom.* **1994**, *8* (2), 169–173.
(109) Han, S.-W.; Lee, S.-W.; Bahar, O.; Schwessinger, B.; Robinson, M. R.; Shaw, J. B.; Madsen, J. A.; Brodbelt, J. S.; Ronald, P. C. *Nat Commun* **2012**, *3*, 1153.
(110) Pruitt, R. N.; Schwessinger, B.; Joe, A.; Thomas, N.; Liu, F.; Albert, M.; Robinson, M. R.; Chan, L. J. G.; Luu, D. D.; Chen, H.; Bahar, O.; Daudi, A.; De Vleesschauwer, D.; Caddell, D.; Zhang, W.; Zhao, X.; Li, X.; Heazlewood, J. L.; Ruan, D.; Majumder, D.; Chern, M.; Kalbacher, H.; Midha, S.; Patil, P. B.; Sonti, R. V.; Petzold, C. J.; Liu, C. C.; Brodbelt, J. S.; Felix, G.; Ronald, P. C. *Sci. Adv.* **2015**, *1* (6), e1500245–e1500245.
(111) Madsen, J. A.; Ko, B. J.; Robotham, S. A.; Xu, H.; Horton, A. P.; Iwashkiw, J. A.; Shaw, J. B.; Feldman, M. F.; Brodbelt, J. S. *Anal Chem* **2013**, *85* (19), 9253–9261.
(112) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. *PROTEOMICS* **2011**, *11* (7), 1329–1334.
(113) Vasicek, L.; Ledvina, A.; Shaw, J.; Griep-Raming, J.; Westphall, M.; Coon, J.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (6), 1105–1108 – 1108.
(114) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. *Anal. Chem.* **2016**, *88* (4), 2303–2310.
(115) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (1), 1044–1051.
(116) Shaw, J. B.; Robinson, E. W.; Paša-Tolić, L. *Anal. Chem.* **2016**, *88* (6), 3019–3023.

Chapter 2

Experimental Methods

2.1 MASS SPECTROMETRY

The development of electrospray ionization (ESI) has revolutionized the field of proteomics by providing a means for converting large nonvolatile biomolecules into gas phase ions.[1] In the process of ESI, a voltage is applied to an analyte solution which is then sprayed from a capillary or emitter tip to form an aerosol of charged droplets containing solution phase analyte molecules. As solvent evaporates from the droplets, charges become condensed until the columbic repulsion exceeds the Rayleigh limit, causing the ejection of gas phase ions. ESI is a soft ionization technique that is rarely implicated for peptide fragmentation, and instead intact molecular ions are typically observed across multiple charge states in the mass spectrum.

ESI is now widely coupled with mass spectrometers, and the research in this dissertation was conducted using several different ESI-MS instruments including linear ion trap (LIT)[2] and Orbitrap[3] mass analyzers. In LITs, ions are confined axially by rf fields and radially by DC voltages that are applied at separate ends of the rods. Mass selective ion ejection from the trap is achieved by applying appropriate rf voltages to the rods. The four rod geometry of LITs enables much greater ion trapping capacity compared to 3D quadrupole ion traps. In contrast, Orbitraps are advantageous for their high resolution and mass accuracy. During Orbitrap analysis, ions exhibit simultaneous orbital and translational motion around/along an inner spindle electrode housed within an outer barrel electrode, and the resulting image current is converted to a mass spectrum via Fourier transform. The following sections highlight more specifically the instrumentation that was used in this dissertation.

2.1.1 Thermo Fisher Scientific LTQ XL Linear Ion Trap

A Thermo Fisher Scientific LTQ XL mass spectrometer (San Jose, CA) equipped for electron transfer dissociation and photodissociation[4,5] was used for the analysis of Lys-N peptides. The spray voltage was maintained at 4 kV, the capillary temperature was 180 °C, and the MS1 automated gain control (AGC) was 3E4.

2.1.2 Thermo Fisher Scientific Velos Pro Dual Linear Ion Trap

A Velos Pro dual linear ion trap mass spectrometer (Thermo Fisher Scientific, San Jose, Ca) with CID, HCD, and UVPD[4,5] functionality was operated in both positive and negative mode for the analysis of CTD and casein peptides eluting from nano LC, and also for sulfopeptides from capillary LC. For nano ESI, 1.8-2 kV was applied in positive mode and 1.5-1.7 kV was applied in negative mode. The temperature of the heated capillary was 275 °C in positive mode and increased to 300-325 °C in negative mode to improve desolvation. For negative ESI, the source was operated at 4 kV and 10 units of sheath gas. An MS1 AGC between 1E4 and 2E4 was used for all experiments.

2.1.3 Oribtrap Elite Mass Spectrometer

A modified Thermo Fisher Scientific Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany)[6,7] was used for the analysis of model sulfopeptides and bovine fibrinogen peptides by both positive and negative ESI. In positive mode, the ESI source was operated at 3.5 kV with a sheath gas flow of 4 units. In negative mode, the heated ESI source (HESI) was used, and the optimal spray parameters were more variable with HESI temperatures ranging from 40-60 °C, source voltage 2.3-3 kV, and

sheath gas flow between 25-40 units. For both polarities the Orbitrap MS1 AGC target was 1E6.

2.1.4 Orbitrap Fusion Tribrid Mass Spectrometer

An Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific Instruments, Bremen, Germany) modified for UVPD[8] was used for the analysis of HeLa and HCC70 lysates and also the fruit fly CTD. Orbitrap detection was used for both MS1 and MS2 measurements at resolving powers of 60K and 15K (at *m/z* 200), respectively. The MS1 AGC target was 2E5 and spectra were collected over *m/z* 400-1500.

Intact mass analysis was carried out for the fruit fly CTD at a resolving power of 240K at *m/z* 200. The maximum number of informative spectra were averaged together prior to Xtract deconvolution at a S/N threshold of 3 to improve the accuracy of the deconvolved mass.

2.2 LIQUID CHROMATOGRAPHY

The success of shotgun proteomics is also heavily reliant on techniques for efficient peptide separation. The subsequent sections describe the reverse phase separation methods that were utilized in this dissertation.

2.2.1 Dionex Ultimate 3000

All separations were carried out on a Dionex Ultimate 3000 LC either at nano (0.3 µL/min) or capillary (4 µL/min) flow rates. Positive mode analysis was exclusively undertaken using nano LC (nLC) in which mobile phase A (MP A) was 0.1% formic acid in water, and mobile phase B (MP B) was 0.1% formic acid in acetonitrile (ACN). Nano

and capillary flow rates were used for negative mode MS analysis, and various different mobile phase systems were employed which will be described in section 2.2.1.2.

Initial nLC separations were carried out by direct injection onto a 15 cm x 75 µm Acclaim PepMap RSLC nano column packed with 2 µm C18. Using this set-up, bovine serum albumin (BSA) Lys-N peptides were separated by a gradient of 5-50% B over 55 min. Reproducibility, robustness, and throughput were increased by implementing a preconcentration nLC set-up in which peptides were loaded onto a short trap column at microliter flow rates and subsequently separated by switching the analytical column in line with the trap and applying a gradient. In-house column packing was adopted for both trap and analytical columns using 100 µm integrafrit and 75 µm integrated emitter picofrit capillaries (New Objective, Woburn, MA), respectively. Traps were packed to ~3.5 cm with 5 µm Michrom magic C18, and analytical columns were packed to 15-20 cm with either 3 µm Michrom magic C18 or 3.5 µm Waters Xbridge BEH C18 (Milford, MA). UPLC columns containing 1.8 µm, 120 Å UChrom C18 (nanoLCMS Solutions, Gold River, CA) were also fabricated specifically for cell lysate applications.

2.2.1.1 Positive Mode LC-MS/MS Analysis

Yeast CTD peptide separations were carried out using the trap and elute set-up described above. Peptides were loaded onto the trap column for 5 min at 5 µL/min in aqueous loading solvent containing 2% ACN, and 0.1% formic acid. A multi-step gradient was applied for separation in which the percent B was increased from 2-15% during the first 15 min and further increased to 35% during the last 5 min. The LC conditions for fruit fly CTD separations were analogous to those described for the yeast

CTD, except that peptides were loaded directly onto the C18 analytical column and separated over 60 min using a gradient from 2-40% B.

The separation of HeLa and HCC70 cell lysates was carried out by direct injection onto a 30 cm UPLC column heated to 60 °C inside a custom column oven.[9] HeLa was separated by an 80 min gradient in which the percentage of MP B was increased from 2-25% B during the first 65 min and from 25-40% B during the final 15 min. The same gradient steps were used for HCC70 but the increase from 2-25% B was carried out more gradually over 118 min.

2.2.1.2 Negative Mode LC-MS/MS Analysis

Fruit fly CTD peptides were analyzed by nLC in negative mode. Methanol (MeOH) was used instead of ACN for mobile phase B and 0.1% trifluoroethanol (TFE) was added to all mobile phases instead of formic acid. Peptides were loaded for 3 min at 5 µL/min in water containing 2% MeOH and 0.1% TFE. A 50 min linear gradient from 2-90% B was used for separations. For negative nLC-MS analysis of alpha and beta casein phosphopeptides ammonium acetate (NH_4OAc) containing mobile phases were used. MP A was 5 mM NH_4OAc in water, and MP B was 5 mM NH_4OAc in 90% MeOH. The pH of MP A was adjusted to eight using ammonium hydroxide. Following 4 min loading at 5 µL/min, separation proceeded over a 45 min linear gradient from 2-45% B.

Performing separations at capillary flow rates and directing the LC eluent through the ESI source provided improved negative mode spray stability (but with some loss in sensitivity). Sulfopeptides were separated by capillary LC using a 3 x 150 mm Agilent Zorbax Extend-C18 column packed with 3.5 µm particles (Santa Clara, CA). MP A was 5 mM NH_4OAc in water, and MP B was 5 mM NH_4OAc in 90% MeOH. Steeper gradients

were often required using MeOH due to its weaker eluting strength relative to ACN, but a fast gradient from 25-60% B over 15 min proved effective for separating a mixture of sulfopeptides. Alternatively, ACN was used instead of MeOH along with the NH_4OAc mobile phase modifiers for sulfopeptide separations following carbamylation reaction and WAX fractionation. In this case, MP A was 5 mM NH_4OAc in water (pH 8) and MP B was 5 mM NH_4OAc in 85% acetonitrile. Separations were carried out on the Zorbax column using a linear gradient from 5-45% B over 45 min.

2.3 ION ACTIVATION

CID, HCD and UVPD were applied for peptide activation and dissociation. The MS2 AGC targets were set to 1E4 on the LTQ and Velos Pro ion trap instruments; 1E4 (positive mode) and 5E4 (negative mode) on the Orbitrap Elite; and 1E5 using the Orbitrap Fusion. Typical MS2 isolation widths ranged from 3 Da down to 1.6 Da using quadrupole isolation on the Orbitrap Fusion.

For protein digests analyzed by LC-MS, data dependant acquisition (DDA) was used to systematically select and dissociate peptides. Using this method, an MS1 spectrum was first collected to survey the available peptides at a given time in the separation, followed by a series of MS2 events for ions selected in order of abundance (beginning with the most abundant). After acquiring a user defined number of MS2 spectra, the DDA cycle is repeated. For most experiments the MS1 spectrum covered *m/z* 400-2000. Exceptions were for cell lysate analysis on the Orbitrap Fusion which used a narrower range from *m/z* 400-1500 and yeast CTD analysis on the Velos Pro which used a wider range from *m/z* 300-2000. The number of MS2 spectra acquired between each MS1 spectrum was both sample and instrument dependant. For complex peptide mixtures

derived from large proteins such as fibrinogen and BSA, ten MS2 were acquired. For simpler mixtures from casein or CTD protein digests, MS2 was acquired for the top eight or five ions, respectively. Using the Orbitrap Fusion, DDA was performed in a top speed mode in which the maximum number of MS2 spectra are acquired during a 3s cycle time.

2.3.1 Collisional Activation

For CID and HCD, normalized collision energies (NCE) between 30-35% were most often applied for peptide fragmentation, although greater NCE values up to 55% were used specifically for sulfopeptides in an effort to generate fragmentation beyond the neutral loss of SO_3. For CID on the LTQ, 30 ms of activation using a q value of 0.25 was necessary for optimal fragmentation. Faster CID (10 ms) was possible on the Orbitrap instruments, but HCD provided far superior speed based on sub millisecond activation times (0.1 ms on the Orbitrap Elite) and also alleviated the low mass cut off problem.

2.3.1 Photodissociation at 193 nm

All mass spectrometers mentioned in the preceding sections were coupled with a Coherent ExciStar XS excimer laser (Santa Clara, Ca) operated at 193 nm and 500 Hz to generate one pulse every 2 ms.[4–8] On the LIT instruments, UVPD was triggered by setting the CID NCE to zero. Since UVPD functions independently of activation q, the q value was decreased to 0.1 in order to improve the detection of low *m/z* ions. UVPD on the Orbitrap Elite was performed in the HCD cell, and thus a voltage must be applied to direct ions into the cell for UVPD fragmentation. Using 1% NCE, efficient ion transfer was achieved without sufficient energy deposition to cause unwanted activation/dissociation prior to UVPD. On the Orbitrap Fusion the UVPD controls were

built into the instrument software for more streamlined analysis. In initial UVPD studies of Lys-N peptides, a single 8 mJ laser pulse was used for activation. In all subsequent studies, 1-2 pulses at 2-3 mJ were effective for peptide fragmentation using UVPD.

2.4 CHEMICALS

Peptides KLVFFAEDVGS, KYGVSVQDI, KPLLIIAEDVEGEY, KLVANNTRL, KVPRNQDWL, KMVELVHFL, and KTMTESSFYSNMLA were purchased from AnaSpec (Fremont, CA); KGAIIGLM, GDFEEIPEEsYLQ, and NsYsYGWMDF-NH2 were from Sigma (St. Louis, MO); GlpQDsYTGWMDF-NH2, RDsYTGWNleDF-NH2, Ac-DpYVPML-NH2, RRLIEDAEpYAARG-NH2, Ac-IpYGEF-NH2, and TSTEPQpYQPGENL were from American Peptide Company (Sunnyvale, Ca); and sYGGFL was from Phoenix Pharmaceuticals (Burlingame, Ca). The proteins BSA, alpha casein, and beta casein were obtained from Sigma, while bovine fibrinogen was obtained from Calbiochem.

The proteases metalloendopeptidase Lys-N, from *Grifola frondasa*, was purchased from Associates of Cape Code (E. Falmouth, MA); Lys-C and chymotrypsin were from Promega (Madison, WI); and trypsin was obtained from Promega, Sigma, or Thermo Fisher Scientific (Grand Island, NY). All other solvents, chemicals, and reagents were obtained from Sigma, Thermo Fisher Scientific, or EMD Millipore (Temecula, CA).

2.5 SAMPLE PREPARATION

2.5.1 Cell Culture

HeLa and HCC70 stable shRNA scramble cells were cultured in accordance with ATCC guidelines and maintained at 37 °C and 5% CO_2. Dulbecco's Modified Eagle

Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) was used for HeLa cells. RPMI media supplemented with 10% FBS, 2 mM glutamine, 1% antibiotic/antimycotic (10 U/mL penicillin, 10 µg/mL streptomycin, 0.025 µg/mL amphotericin B), and 1 µg/mL puromycin was used for HCC70. HeLa cells were lysed by sonication for 30 s with 1 s alternating on and off cycles at 20% amplitude in buffer containing 50 mM ammonium bicarbonate (pH 8), 8 M urea, 1 mM sodium orthovanadate, 100 mM sodium fluoride, 10 mM sodium pyrophosphatase, 1 µM microcystin-LR, 100 nM calyculin A, complete EDTA-free protease inhibitor mix (Roche), and phosSTOP phosphatase inhibitor mix (Roche). HCC70 cells were lysed by freezing and thawing in the same lysis buffer described for HeLa. Protein concentration was measured by Bradford assay.

2.5.2 Protein Processing

For proteins with disulfide linkages, reduction and alkylation were carried out prior to enzymatic digestion. Disulfide bonds were reduced using 5 mM dithiothreitol (DTT) at 55 °C for 30-45 min. The resulting free thiols were then capped by alkylation using 15 mM iodoacetamide (IAM) at room temperature in the dark for 30-45 min after which an additional aliquot of DTT was added to quench the reaction.

For trypsin digestion, protein solutions were buffered at pH 7.5-8 using 50-100 mM ammonium bicarbonate or Tris HCl. The ratio of trypsin to protein (w/w) ranged from 1:20 to 1:50, and digestion proceeded overnight at 37 C.

For HeLa and HCC70 cell lysates, Lys-C digestion was first applied directly in the cell lysis buffer in order to preserve phosphorylations during protein processing. A 1:200 ratio of Lys-C to protein was used for digestion, and after 2-4 hrs of incubation at

37 °C the urea concentration was diluted to 1.5 M by addition of 50 mM Tris HCl, 5 mM $CaCl_2$. Trypsin was then added in a 1:50 ratio and digestion continued overnight at 37 °C.

Serial digestion by trypsin and proteinase K was applied to the yeast GST-CTD. The first stage of digestion using trypsin occurred overnight at 37 °C in 50 mM ammonium bicarbonate and 1.6 M urea using a 1:50 ratio of enzyme to protein. The digest solution was then passed through a 10 KDa molecular weight cutoff (MWCO) filter to isolate and buffer exchange the CTD 26mer into 50 mM Tris HCl containing 10 mM $CaCl_2$ (pH 8). Proteinase K was added to the CTD solution in a 1:100 ratio and digestion proceeded overnight at 37 °C.

Chymotrypsin was used in a 1:50 enzyme to protein ratio to digest the fruit fly GST-CTD. The digest buffer was 100 mM Tris HCl containing 10 mM $CaCl_2$ (pH 8). After overnight digestion at room temperature, proteolysis was quenched by the addition of 0.5% trifluoroacetic acid.

2.5.3 Peptide Derivatization

Imidazolinylation of the lysine ε-amine of Lys-N peptides was carried out by diluting 10 µg of peptide into 30 µL of 1M Na_2CO_3 and 45 µL of 1M 2-methylthio-2-imidazoline hydroiodide followed by incubation at 55°C for 12 hours. Reactions were desalted using Thermo Pepclean C_{18} spin columns and reconstituted to10 µM for model peptides and 1 µM for protein digests.

Carbamylation was applied to sulfated peptides to convert all available primary amines to carbamates. Urea was added directly to peptide solutions to a concentration of 8 M and the samples were incubated at 80 °C for 4 hours. No sample clean-up was

carried out after carbamylation and instead reaction solutions were diluted to 500 µL in 50 mM ammonium chloride in preparation for weak anion exchange.

2.6 ENRICHMENT

2.6.1 Immobilized Metal Affinity Chromatography

Phosphopeptide enrichment from whole cell lysates was achieved by immobilized metal affinity chromatography (IMAC) using Ni-nitrilotriacetic acid (Ni-NTA) magnetic agarose beads (Qiagen, Germantown, MD).[10] The beads were prepared for IMAC by 1 hr of shaking in 40 mM EDTA (pH 7.5) to remove metal ions by chelation, followed by thorough washing with water. Next Fe^{3+} was incorporated onto the surface of the beads by shaking for 1 hr in 100 mM $FeCl_3$. Washing with IMAC loading buffer composed of 80% ACN, 0.15% TFA removed excess $FeCl_3$ and conditioned the beads for sample loading. Protein digests were added to the beads in IMAC loading buffer and shaking proceeded for 1 hr to bind phosphopeptides. Three washes in loading buffer were performed to removed non-phosphorylated peptides, and phosphopeptides were subsequently eluted by vortex shaking for 1 min and 15 sec in 100 µL of 50% ACN, 0.7% NH_4OH (pH ~11). The solution of IMAC enriched phosphopeptides was immediately neutralized by addition of 50 µL of 4% formic acid.

2.6.2 Hydroxyapatite

Phosphopeptides from alpha and beta casein were enriched using ceramic hydroxyapatite (HAP) type I, 20 µm resin (Bio Rad, Hercules, CA). Digested peptides in 50 mM Tris HCl were mixed with HAP resin in a 1:5 (w/w) ratio in a fritted centrifuge column. Phosphopeptide binding to HAP occurred for 1 hr at room temperature with

shaking. To remove unmodified peptides, the loading/binding solution was removed by centrifugation at 1200 rcf, and the resin was washed using 200 µL aliquots of 50 mM Tris HCl containing 20% ACN. To elute bound phosphopeptides, the HAP resin was exposed to 1 M KH_2PO_4 (pH 7.8) for 15 min with shaking. Desalting on either C18 or graphite solid phase extraction columns followed.

2.6.3 Weak Anion Exchange

Weak anion exchange (WAX) using diethylaminoethyl (DEAE)-Sephadex A-25 (Sigma, St. Louis, MO) resin was applied for sulfopeptide enrichment. WAX columns were prepared by adding 20 µg of pre-swollen DEAE-Sephadex suspended in 50 mM ammonium chloride (NH_4Cl) to a fritted SPE column. A second frit was added to secure the resin bed. Several milliliters of 50 mM NH_4Cl loading buffer were used to condition the column followed by sample loading in the same buffer. For peptide fractionation, 500 µL aliquots of increasing NH_4Cl concentration were passed through the WAX column and separately collected. Samples were desalted on C18 stage tips constructed according to published protocols using Empore C18 extraction disks (3M, Minneapolis, MN).[11]

2.7 AUTOMATED PEPTIDE SEQUENCING

Several programs were used to interpret the results of bottom up LC-MS/MS experiments. Certain search parameters, such as those related to enzymes, were applied independent of search algorithm. For both trypsin and chymotrypsin, three missed cleavages were allowed and the P-rule, which prohibits cleavage N-terminal to Pro, was applied. Cleavage C-terminal to Lys and Arg was specified for trypsin and C-terminal to Phe, Tyr, Trp, Leu, and Met for chymotrypsin. The minimum peptide length was set to 5

amino acids. Advanced search settings for each program are described in the sections that follow.

2.7.1 PEAKS

PEAKS Studio 5.3 was used to *de novo* sequence tryptic BSA peptides that were analyzed by LC-MS-CID. Prior to *de novo* sequencing, separate MS/MS scans were merged when precursor ion *m/z* values and chromatographic retention times agreed within tolerance limits of 1 Da and 5 min, respectively. The range of allowed precursor charge states was +1 to +3. Peak centroiding, charge deconvolution, and deisotoping were used for all MS/MS spectra. For *de novo* sequencing, the mass error tolerances for parent and fragment ions were both set to 0.5 Da. Carbamidomethylation of cysteine was a fixed modification. Scoring thresholds of total local confidence (TLC) ≥ 5 and average local confidence (ALC) (%) ≥ 50 were used to filter *de novo* peptide sequences.

2.7.2 Proteome Discoverer

Proteome Discoverer 1.3 was used for positive mode UVPD and HCD analysis of IMAC enriched tryptic peptides from HeLa and HCC70 lysates. A non-fragment filter was applied in UVPD searches to remove precursor peaks from MS2 spectra within a 1 Da mass window prior to Sequest database search against the forward and reverse uniprot human database. For UVPD, *a, b, c, x, y,* and *z* ions were used for spectrum matching while only a, *b,* and *y* ions were used for HCD. For both UVPD and HCD neutral losses from *a, b,* and *y* ions were considered. Other database search parameters used for both UVPD and HCD were: 10 ppm precursor mass tolerance; 0.02 Da fragment mass tolerance; N-terminal acetylation, methionine oxidation, and serine/threonine/tyrosine

phosphorylation dynamic modifications; carbamidomethyl cysteine static modifications. PSM (peptide spectral match) validation was carried out using Percolator, and phosphoRS 3.0 was used for phosphosite localization. Filters were applied post search to select only rank 1, high confidence PSMs and an isoform probability of 75% was required for phosphosite localization.

Proteome Discoverer 2.0 was used to interpret the results from positive UVPD analysis of the fruit fly CTD. Sequest database search was performed as described for HeLa and HCC70 with the exception of the enzyme which was chymotrypsin and the FASTA database which was comprised of fruit fly CTD sequences. Instead of using Percolator, matches were filtered based on a maximum Delta Cn of 0.05 using the fixed value PSM validator. Strict and relaxed target FDR settings were 0.01 and 0.05 respectively for both PSMs and peptides. Phospho-site localization was performed using ptmRS, which is analogous to phosphoRS but with added utility for localizing any PTM. Isoform confidence probabilities of 99% from ptmRS were required for phospho-site localization.

2.7.3 MassMatrix

The MassMatrix database search algorithm (version MassMatrix Xtreme 3.0.10.16) was used to interpret the UVPD fragmentation of peptide anions.[12–15] Searches were performed against forward and reverse FASTA databases which contained only the relevant protein sequences for each study, and peptide matching was based on *a*, *x*, *c*, *z*, and *y* type product ions.

For HRAM Orbitrap analysis of bovine fibrinogen, the peptide mass tolerance was 20 ppm and the fragment mass tolerance was ±0.02 Da. Sulfated tyrosine and

pyroglutamate from glutamate were variable modifications and iodoacetamide derivatization of cysteine was a fixed modification. Score thresholds for pp and pp_{tag} were defined as 5.0 and 1.3, respectively.

For low resolution ion trap UVPD analysis of phosphopeptides from alpha casein, beta casein, and fruit fly CTD the peptide mass and fragment mass tolerances were set to ±1.00 Da and ±0.80 Da, respectively. Variable modifications were phosphorylation of serine, threonine, and tyrosine residues. The minimum pp and pp_{tag} scores were 4.0 and 1.0 respectively for casein peptides and 6.0 and 3.0 respectively for fruit fly CTD peptides.

2.8 REFERENCES

(1) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. *Science* **1989**, *246* (4926), 64–71.
(2) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (6), 659–669.
(3) Makarov, A. *Anal. Chem.* **2000**, *72* (6), 1156–1162.
(4) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Anal. Chem.* **2008**, *80* (13), 4807–4819.
(5) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.
(6) Vasicek, L.; Ledvina, A.; Shaw, J.; Griep-Raming, J.; Westphall, M.; Coon, J.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (6), 1105–1108 – 1108.
(7) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.
(8) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (1), 1044–1051.
(9) Richards, A. L.; Hebert, A. S.; Ulbrich, A.; Bailey, D. J.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. *Nat Protoc.* **2015**, *10* (5), 701–714.
(10) Jackson, S. S.; Coughlin, E. E.; Coon, J. J.; Miyamoto, S. *Protein Expr. Purif.* **2013**, *92* (1), 48–53.
(11) Rappsilber, J.; Mann, M.; Ishihama, Y. *Nat. Protoc.* **2007**, *2* (8), 1896–1906.
(12) Xu, H.; Freitas, M. *BMC Bioinformatics* **2007**, *8* (1), 133.
(13) Xu, H.; Yang, L.; Freitas, M. *BMC Bioinformatics* **2008**, *9* (1), 347.
(14) Xu, H.; Freitas, M. A. *PROTEOMICS* **2009**, *9* (6), 1548–1555.
(15) Xu, H.; Freitas, M. A. *J. Proteome Res.* **2008**, *7* (7), 2605–2615.

Chapter 3

193 nm Ultraviolet Photodissociation of Imidazolinylated Lys-N Peptides for De Novo Sequencing[1]

3.1 OVERVIEW

The goal of many MS/MS *de novo* sequencing strategies is to generate a single product ion series that can be used to determine the precursor ion sequence. Most methods fall short of achieving such simplified spectra, and the presence of additional ion series impede peptide identification. The present study aims to solve the problem of confounding ion series by enhancing the formation of "golden" sets of *a*, *b*, and *c* ions for sequencing. Taking advantage of the characteristic mass differences between the golden ions allows N-terminal fragments to be readily identified while other ion series are excluded. By combining the use of Lys-N, an alternate protease, to produce peptides with lysine residues at each N-terminus with subsequent imidazolinylation of the ε-amino group of each lysine, peptides with highly basic sites localized at each N-terminus are generated. Subsequent MS/MS analysis by using 193 nm ultraviolet photodissociation (UVPD) results in enhanced formation of the diagnostic golden pairs and golden triplets that are ideal for *de novo* sequencing.

[1] Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. 193 nm Ultraviolet Photodissociation of Imidazolinylated Lys-N Peptides for De Novo Sequencing. *Analytical Chemistry*. **2012**, *84*, 2433–2439.
JAM and JSB provided mentorship and reviewed the manuscript prior to publication.

3.2 INTRODUCTION

The development of mass spectrometry techniques for analyzing biological samples, in particular tandem MS using collisional induced dissociation (CID), has played a vital role in establishing modern day proteomics.[1] Despite ongoing advancement in instrumentation and a flurry of new activation methods, CID of chromatographically-separated tryptic peptides followed by *in silico* database searching remains the gold standard for bottom-up analysis. While this method has proven useful for routine protein analysis, it cannot be applied to proteins from organisms having unsequenced genomes. Additionally a general decrease in performance occurs as samples become more complex. For example, confident identification of proteins having post-translational modifications remains difficult by *in silico* algorithms,[2] especially if specialized software and the highest performing (and most expensive) mass spectrometers are not available.[3]

These limitations of *in silico* database searching have prompted the development of *de novo* sequencing algorithms, which attempt to reconstruct peptide sequences based on MS/MS data without reference to a database. Several *de novo* programs have been developed to analyze CID data including Lutefisk,[4] PEAKS,[5] DACSIM,[6] NovoHMM,[7] PepNovo,[8] EigenMS,[9] and MSNovo[10]. Because these algorithms identify amino acids based on the difference between consecutive product ions of a given series, complete backbone fragmentation is ideal for unambiguous identification. However, full sequence coverage is often not obtained, especially when using CID which exhibits dominant loss of labile groups as well as preferential cleavage at proline and acidic residues.[11–13]

By using alternative activation methods instead of conventional CID, a greater degree of backbone fragmentation can be achieved. Photon-based activation methods are becoming increasingly well established[14–17] and promote greater diversity in the

fragmentation pathways compared to the predominant *b/y* ions exhibited up on CID. Ultraviolet photodissociation (UVPD) at 193 nm in a linear ion trap has particular merit because of the extensive backbone fragmentation that results in the formation of *a, b, c, x, y,* and *z* ions.[18] The formation of *d*, v, and *w* ions from side-chain losses is also useful for differentiating isobaric or near isobaric amino acids.

While more extensive fragmentation providing a larger array of product ions helps to ensure a complete series for *de novo* sequencing, the presence of multiple ion series can complicate spectral interpretation. Several strategies for differentiating complementary ion series have been reported, such as those entailing sample fractionation and selective labeling at the N- or C-terminus of only one fraction .[19–23] Fragments that include the modified terminus are easily identified based on the mass shift between peaks in the spectra of the unmodified versus modified fractions. An alternative approach involves collection of separate MS/MS spectra using different activation methods that yield inherently different results, i.e. production of *b* and *y* ions for CID compared to *c* and *z* ions for electron based dissociation and *a* and *x ions* for UVPD at 157 nm. Through comparison of complementary spectra, "golden pairs" of ions, defined as *a*/*b, b*/*c, a*/*c,* or *x*/*y, y*/*z, x*/*z* pairs, [24] can be identified based on characteristic mass differences.[24,25]

An alternative to differentiating ion series that originate from opposite termini is to eliminate ion series derived from one terminus of the peptide altogether. Addition of negatively charged modifications at the peptide N-terminus neutralizes the positive charge to effectively produce spectra consisting of predominantly C-terminal ions.[26–28] Conversely, the N-terminal series can be enhanced through modification of the N-terminus with moieties having high proton affinity.[29] Careful tuning of the proton affinity at the N-terminus is necessary so that the ionizing proton is free to move along the

peptide backbone but is ultimately captured at the N-terminus to form *b* ions upon CID. A similar strategy involves the use of an alternate protease, Lys-N, which produces peptides with lysines at the N-termini.[30] When a single basic residue resides at the N-terminus of a peptide, the proton is effectively sequestered, and simplified spectra comprised of mostly *b* ions for CID and *c* ions for electron transfer dissociation (ETD) are produced.[31] Further enhancement of N-terminal ions can be achieved following guanidination or imidazolinylation to increase the basicity of the lysine's ε-amino group,[32,33] thus ensuring immobilization of the proton.

In the present study, 193 nm UVPD is applied to imidazolinylated Lys-N peptides to generate clean spectra consisting of "golden triplet" sets of *a*, *b*, and *c* ions and/or "golden pair" sets of *a* and *b* ions that are well-suited for *de novo* sequencing. The richer diversity of diagnostic ions afforded by 193 nm UVPD allows the three N-terminal ion series to be readily identified based on characteristic mass differences to the exclusion of other fragment ions. Unlike other "golden pair" techniques, the complete N-terminal ion series can be generated in a single spectrum, thus eliminating the need for multiple activation events and cross spectral comparison.

3.3 EXPERIMENTAL

3.3.1 Materials

Model peptides were obtained from AnaSpec (Fremont, CA) with the exception of KGAIIGLM, which was purchased from Sigma-Aldrich (St. Louis, MO). Bovine serum albumin (BSA), proteomics grade trypsin, dithiothreitol, iodoacetamide, and 2-methylthio-2-imidazoline hydroiodide were also acquired from Sigma-Aldrich. Metalloendopeptidase Lys-N, isolated from *Grifola frondasa*, was purchased from

Associates of Cape Code (E. Falmouth, MA). Sodium carbonate was purchased from EMD Chemicals (Gibbstown, NJ) and all other reagents and solvents were obtained from Fisher Scientific (Fairlawn, NJ).

3.3.2 Sample Preparation

For disulfide bond reduction 5 µL of a 200 mM solution of dithiothreitol (DDT) in 100 mM NH_4HCO_3 was added to 20 µg of BSA, and the mixture was incubated at 55 °C for 45 min. Subsequent alkylation was performed by addition of 4 µL of 1M iodoacetamide. Following reaction for 30 min at room temperature in the dark, 20 µL of DTT was added to quench the reaction. For digestion with trypsin, 1 µg of trypsin was added to the reduced and alkylated BSA, and the mixture was incubated at 37 °C overnight. Protein digestion with Lys-N was performed using a 25:1 protein to enzyme ratio in 100 mM NH_4HCO_3. The reaction mixture was incubated overnight at 37 °C. For imidazolinylation of the lysine amine groups, 10 µg of protein digest or model peptide was diluted in 30 µL of 1M Na_2CO_3 and allowed to react with 45 µL of 1M 2-methylthio-2-imidazoline hydroiodide at 55°C for 12 hours. Prior to analysis, reaction mixtures were cleaned up using Thermo Pepclean C_{18} spin columns and diluted to the appropriate concentrations, 10 µM for model peptides and 1 µM for protein digests.

3.3.3 MS, UVPD, and LC

Model peptides were analyzed by direct infusion ESI-MS using a Thermo Fisher Scientific LTQ XL mass spectrometer (San Jose, CA). The spray voltage and capillary temperature settings were kept constant at 4 V and 180 °C respectively for all mass spectrometry experiments. The automated gain control (AGC) remained at 3.0×10^4 for

MS and 1×10^4 for MS^n scans. For UVPD, a Coherent ExiStar XS ArF excimer laser (Santa Clara, Ca) was coupled to the LTQ as previously described,[18,34] and 193 nm photons were delivered at a frequency of 500 Hz. The pulse duration was 5 ns with an energy of 8 mJ. A single pulse during an activation period of 30 µs with q=0.1 induced fragmentation. For CID experiments, the activation q was increased to 0.25 and normalized collision energy was chosen such that the precursor ion remained at about 10 percent relative abundance in the MS/MS spectrum (to facilitate identification of the precursor ion).

Protein digests were separated on a Dionex Ultimate 3000 RSLCnano system using an Acclaim PepMap RSLC nano column of length 15 cm and inner diameter 75 µm packed with 2 µm C18 stationary phase. The system was operated at a flow rate of 0.3 µL/min using mobile phases A and B consisting of water and acetonitrile respectively, each containing 0.1% formic acid. The following LC program, lasting 70 minutes for each sample, was used for all separations. After 5 minutes running at 5% B, the percent of mobile phase B increased linearly to 50% over 55 minutes. At the end of the gradient, the %B was increased to 80% for 5 minutes and subsequently decreased to 5% for the final five minutes in preparation for the next run. Data dependent UVPD and CID scans were performed by first collecting a full mass spectrum of the m/z range 400-2000. Next, UVPD or CID was performed on the top ten most abundant peaks from the full mass spectrum. The same UVPD parameters used for direct infusion ESI-MS of model peptides were also used for LC-MS-UVPD experiments. For CID, a normalized collision energy of 35 was applied during a 30 ms activation period.

3.3.4 *De Novo* Sequencing of Protein Digest

De novo sequencing was performed on the results from LC-MS-CID analysis of the tryptic digest of BSA using PEAKS Studio 5.3. Prior to *de novo* sequencing, MS/MS data refinement was performed as follows: (1) Separate MS/MS scans were merged when the precursor *m/z* values agreed within an error tolerance of 1 Da and fell within a 5 min retention time window (defined as the maximum difference of retention time between two spectra to be merged). (2) The minimum and maximum precursor charge states were fixed at 1+ and 3+ respectively. (3) Peak centroiding, charge deconvolution, and deisotoping functions were applied to all MS/MS scans.

For *de novo* sequencing, the parent and fragment mass error tolerances were both set to 0.5 Da. Carbamidomethylation of cysteine was selected as a fixed modification. *De novo* peptide candidates were filtered based on amino acid total local confidence (TLC) and average local confidence (ALC). The TLC is the sum of the local confidence scores which define the likelihood that a particular amino acid is present at a particular position in a *de novo* peptide. Local confidence scores are determined based on several factors including the peak abundance, mass errors, and coexistence of other supporting peaks. ALC is calculated by dividing the TLC by the total peptide length. Only peptide sequences having TLC $\geq$ 5 and ALC (%) $\geq$ 50 were kept and cross-checked against known tryptic BSA peptides to either confirm identifications or reject false positives.

3.4 RESULTS AND DISCUSSION

Via a combination of imidazolinylation of the N-terminus and using UVPD as an alternative to conventional CID, our goal was to enhance the formation of triplet *a*, *b*, *c* ions from peptides. In theory, the presence of low abundances of C-terminal ions should not affect the identification and utilization of the triplet N-terminal (*a*,*b*,*c*) series for *de*

novo strategies, as the latter ions are identified based on characteristic mass differences. Thus, as long as the N-terminal triplet sets are present, all other ions can be systematically excluded from use in *de novo* sequencing algorithms. All the same, there are several reasons that suppression of the formation of C-terminal series is more ideal. Simplified spectra that solely show triplet *a, b, c* sets lend themselves to more rapid manual data interpretation without the need for more tedious manual exclusion of C-terminal ions. Additionally, spectral overlap can be more problematic for N- and C-terminal ions that have similar *m/*z values, especially when high mass accuracy measurements are not feasible. Finally, the total product ion abundance that is split three ways to yield *a*, *b*, and *c* ions should be greater in the absence of C-terminal ions because the ion current will not be further sub-divided among additional fragmentation channels. These benefits of generating golden triplets also hold true for golden pairs (when the triplet series is not present). In the following sections, the impact of imidazolinylation on the enhancement of golden triplets and pairs by UVPD is discussed, as well as the distributions of *a*, *b*, and *c* ions and the influence of charge state of the peptide.

3.4.1 Imidazolinylation of N-terminal Lysine Peptides

Eight model peptides, all having lysine at their N-termini, were modified through an imidazolinylation reaction shown schematically in **Figure 3.1.**

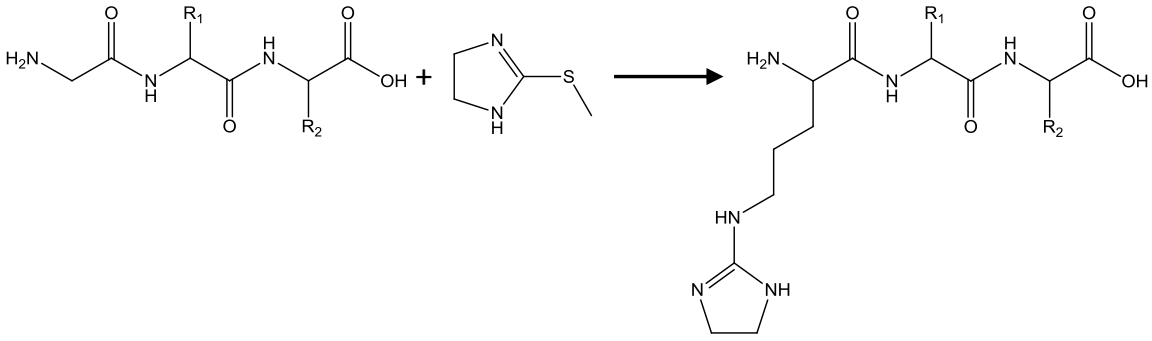


Figure 3.1 Imidazolinylation reaction

The mass spectra of the model peptides following derivatization show both one and two successive additions of 68 Da, the latter suggesting the double addition of imidazole with the second imidazole presumably attached at the N-terminus. Following incubation, all peptides were converted to either the singly or doubly imidazolinylated products with no unreacted peptides detected in the full mass spectra. Activation of the singly and doubly modified peptides by UVPD promoted the same types of fragmentation pathways, however with each N-terminal product ion shifted by 68 Da for the ions arising from the doubly modified peptides relative to the singly modified peptides. This consistent and uniform mass shift confirmed the location of the second imidazole at the N-terminus. While a mixture of modified peptides resulted from the imidazolinylation reaction (and even occasionally triply-modified peptides of very low abundance), the single addition at the lysine was the far more dominant of the products (typically 70% singly modified and 20% doubly modified based on the peak intensities of these two major products) and is the primary product of interest for the remainder of the study.

3.4.2 CID versus UVPD for Formation of N-terminal Triplets

The eight model peptides as both their protonated imidazolinylated and underivatized forms were subjected to CID and UVPD. Representative spectra obtained upon CID and UVPD of singly charged underivatized and imidazolinylated KLVFFAEDVGS are shown in **Figure 3.2**. Companion bar graphs to the right of each spectrum display the distributions of *a*, *b*, and *c* fragment ions relative to their amino acid position in the sequence.

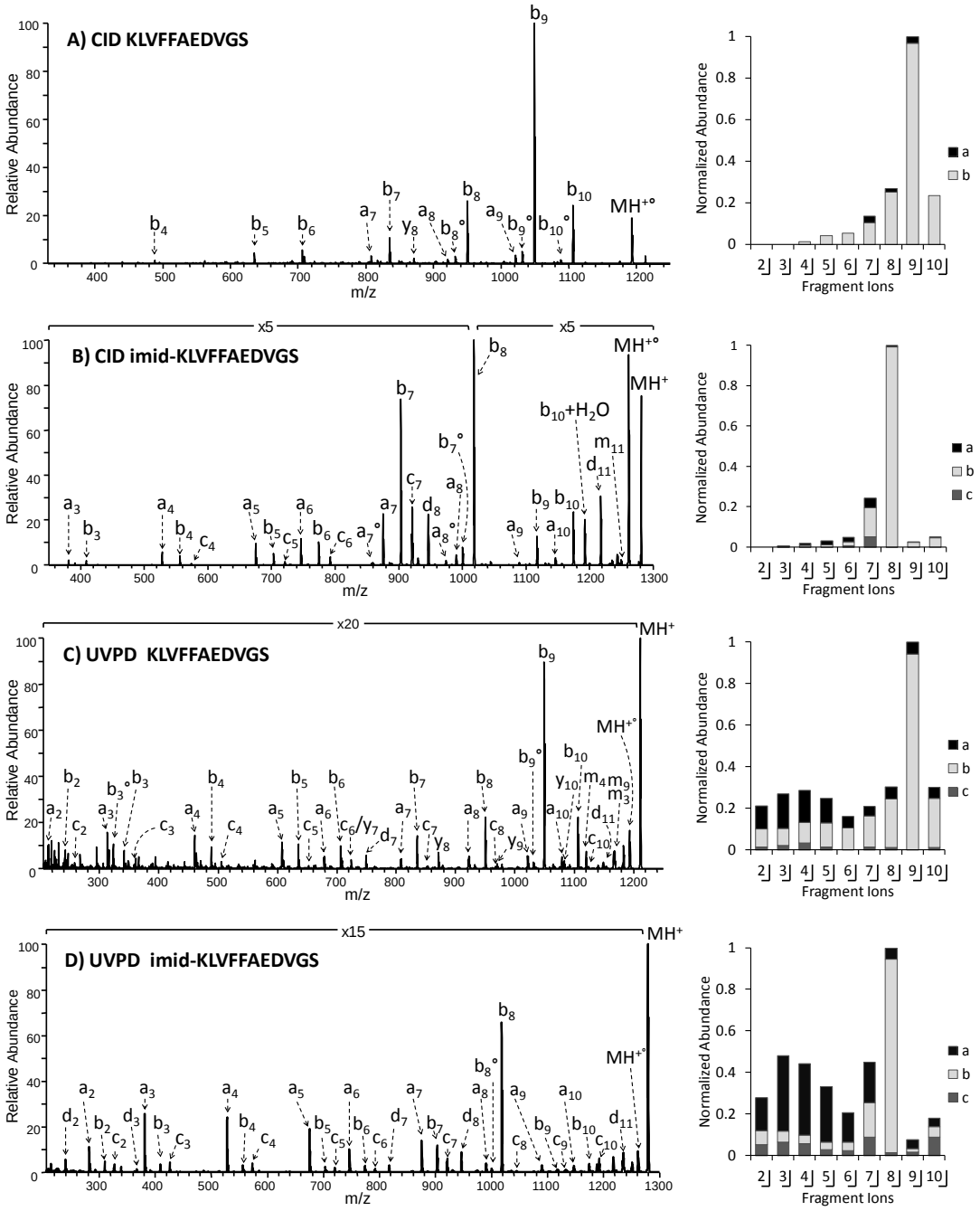


Figure 3.2 MS/MS spectra of 1+ KLVFFAEDVGS: (A) CID of underivatized, (B) CID of imidazolinylated, (C) UVPD of underivatized, and (D) UVPD of imidazolinylated. Peaks labeled with a "°" sign denote loss of H_2O or NH_3. Companion bar graphs displaying distribution of N-terminal ion abundances are shown beside each spectrum with y-axis normalized to 1.

Upon conventional CID, *b* ions dominated the spectrum (**Figure 3.2A**), and the other two N-terminal series of fragment ions (*a*, *c*) were poorly represented with only a few *a* ions of low abundance and no *c* ions observed. After imidazolinylation, CID again produced a nearly complete set of *b* ions, and their relative abundances differed greatly from those observed for the underivatized peptide (**Figure 3.2B**). The product ion distribution bar graphs in **Figures 3.2A-B** show that for underivatized KLVFFAEDVGS the b_9 ion was the most abundant, but following imidazolinylation the b_8 product corresponding to cleavage C-terminal to aspartic acid was clearly favored. This type of preferential cleavage is observed when CID is applied to peptides in which the number of highly basic sites exceeds the number of ionizing protons, thus restricting proton mobility and promoting alternative charge remote fragmentation.[12,35,33,29] Therefore, the presence of the abundant b_8 ion in **Figure 3.2B** indicates that the imidazole moiety effectively sequesters the ionizing proton to such an extent that other CID pathways are substantially reduced. Evidence for this is seen in the supporting ion series, which include a nearly complete set of *a* ions as well as several *c* ions having isotopic distributions characteristic of radical $c + 1$ ions, all of which have abundances of less than 5% relative to the dominant b_8 ion.

UVPD at 193 nm provided a far greater array of *a*, *b*, and *c* ions (**Figures 3.2C-D**). Imidazolinylated KLVFFAEDVGS yielded a particularly unique UVPD spectrum (**Figure 3.2D**) showing complete sets of *a*, *b*, and *c* ions along with some additional *d* ions that arise from side-chain losses from *a* ions. Some preference for b_8 formation is observed as was seen in the CID spectrum of imidazolinylated KLVFFAEDVGS, but the far greater array and abundances of *a* ions reflects the different mechanism of high energy UVPD. The UVPD spectrum of underivatized KLVFFAEDVGS also shows a prominent series of N-terminal ions; however, the appearance of *y* ions, y_7 to y_{10}, indicates

that unmodified lysine is not sufficiently basic to suppress the formation of C-terminal ions, the latter which are not observed for the imidazolinylated peptide (**Figure 3.2D**). The triplet series of N-terminal fragment ions was readily identified based on the characteristic 28 Da separation between consecutive *a* and *b* ions and the 17 Da separation between consecutive *b* and *c* ions. The bar graph distribution shown in **Figure 3.2D** illustrates the increase in *a* and *c* ion abundances relative to the analogous bar graph for the underivatized peptide in **Figure 3.2C**. Moreover, differences in the relative abundances of *a* and *b* ions were also evident upon comparison of the N-terminal ion distribution graphs in **Figure 3.2C-D** which show a consistent shift from dominant *b* ions for underivatized KLVFFAEDVGS to dominant *a* ions for imidazolinylated KLVFFAEDVGS upon UVPD.

3.4.3 Assessment of Imidazolinylation/UVPD Strategy

UVPD alone is capable of generating the entire array of peptide fragment ions which includes the key *a, b,* and *c* ions, but further enhancement of the N-terminal series can be achieved by placing a highly basic site at the peptide N-terminus through Lys-N digestion and imidazonlinylation. In general, UVPD of the underivatized peptides showed similar but in many cases less complete sets of N-terminal ions relative to UVPD of the corresponding imidazolinylated peptides. **Table 3.1** compares the percent sequence coverage afforded by each *a*, *b*, and *c* ion series for the eight model peptides, for which percent sequence coverage (adapted from [36]) is defined by:

$$\%\ Sequence\ Coverage = \frac{total\ \#\ of\ observed\ fragment\ ions}{total\ \#\ of\ possible\ fragment\ ions\ -\ 1} \times 100$$

A)	Underivatized 1+			Imidazolinylated 1+		
	a	b	c	a	b	c
KGAIIGLM	100	100	83	100	100	100
KLVFFAEDVGS	100	100	**78**	100	100	100
KYGVSVQDI	86	86	86	100	100	100
KPLLIIAEDVEGEY	92	100	92	100	100	92
KLVANNTRL	86	86	57	100	100	100
KVPRNQDWL	71	**86**	86	100	**86**	**86**
KMVELVHFL	86	100	86	100	100	100
KTMTESSFYSNMLA	**75**	100	58	100	100	92

B)	Underivatized 2+			Imidazolinylated 2+		
	a	b	c	a	b	c
KGAIIGLM	100	100	**33**	100	100	83
KLVFFAEDVGS	89	100	**67**	100	100	**67**
KYGVSVQDI	71	100	43	86	100	100
KPLLIIAEDVEGEY	**83**	100	**75**	92	100	**50**
KLVANNTRL	**57**	57	14	86	100	86
KVPRNQDWL	71	**57**	43	57	**43**	**14**
KMVELVHFL	71	86	57	**71**	86	71
KTMTESSFYSNMLA	**50**	83	42	92	92	82

Table 3.1 Percent sequence coverage provided by *a*, *b*, and *c* fragment ion sets for UVPD of imidazolinylated and underivatized peptides in (A) 1+ and (B) 2+ charge states. Numbers reported in **bold font** indicate omission of fragment ions due to an inability to distinguish between different ions with similar m/z. 100% sequence coverage is defined as a complete set of fragment ions ranging from a_2 (b_2, c_2) to a_n (b_n, c_n), where n is the total number of possible fragment ions. Gray cells highlight sequence coverages of less than 100%.

The first N-terminal product ions corresponding to a_1, b_1, or c_1 were not included in the calculation of percent sequence coverage (i.e. subtraction of 1 in the denominator) because lysine is exclusively the first amino acid for each peptide created using the Lys-

N protease, and thus the N-terminal residue is already known. Overall, the sequence coverage obtained upon UVPD improved for all three series of fragment ions following imidazolinylation (an average improvement of 13%, 4% and 18% for *a*, *b*, and *c* ions respectively), with the increase in percent sequence coverage being most dramatic for the *c* ion series. The precursor ion charge state also plays a significant role in the formation and distributions of the N-terminal product ions. For both imidazolinylated and underivatized peptides, the best sequence coverage for golden *a*, *b*, and *c* ions obtained upon UVPD is observed for singly charged peptides, with golden pairs of *a* and *b* ions more typically observed for higher charge states (see example in **Figures 3.3 and 3.4**).

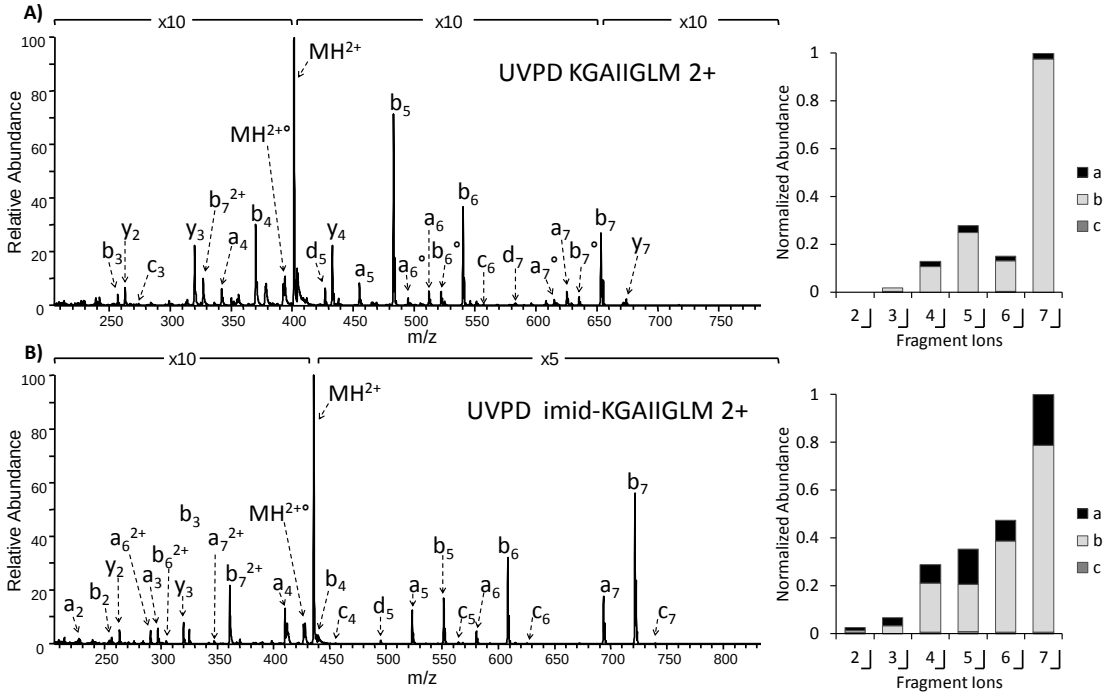


Figure 3.3 MS/MS showing UVPD of (A) underivatized and (B) imidazolinylated KGAIIGLM 2+. Peaks labeled with a "°" sign denote loss of H_2O or NH_3. Companion bar graphs displaying distribution of N-terminal ion abundances are shown beside each spectrum with y-axis normalized to 1.

Table 3.1B shows that the *b* ions maintain relatively high sequence coverage even for the higher charge states due to an increase in proton mobility which facilitates *b* ion formation,[37] but the higher proton mobility does not enhance *a* ions nor *c* ions, the latter formed by an alternative route.[38] Interestingly, for the 2+ peptides, increases in sequence coverage are achieved following imidazolinylation due to the increase in basicity of the N-terminus, thus causing at least one proton to be more strongly retained at the N-terminus. The high *a/b* sequence coverage of over 85% for six out of eight imidazolinylated 2+ peptides allows successful *de novo* sequencing of doubly charged precursor ions. The incidence of spectral overlap (i.e. inability to distinguish between different fragment ions having similar *m/z* values), shown by bold values in **Table 3.1**, also increases for the doubly charged peptides because a greater number of product ions are observed for precursors of higher charge state.

UVPD of triply charged ions, as exemplified in **Figure 3.4** for KLVANNTRL, results in the same trends reported for doubly charged ions but with more notable decreases in the *a*/*b* sequence coverage as a result of the additional mobile proton (which allow greater opportunities for formation of fragment ions containing the C-terminus).

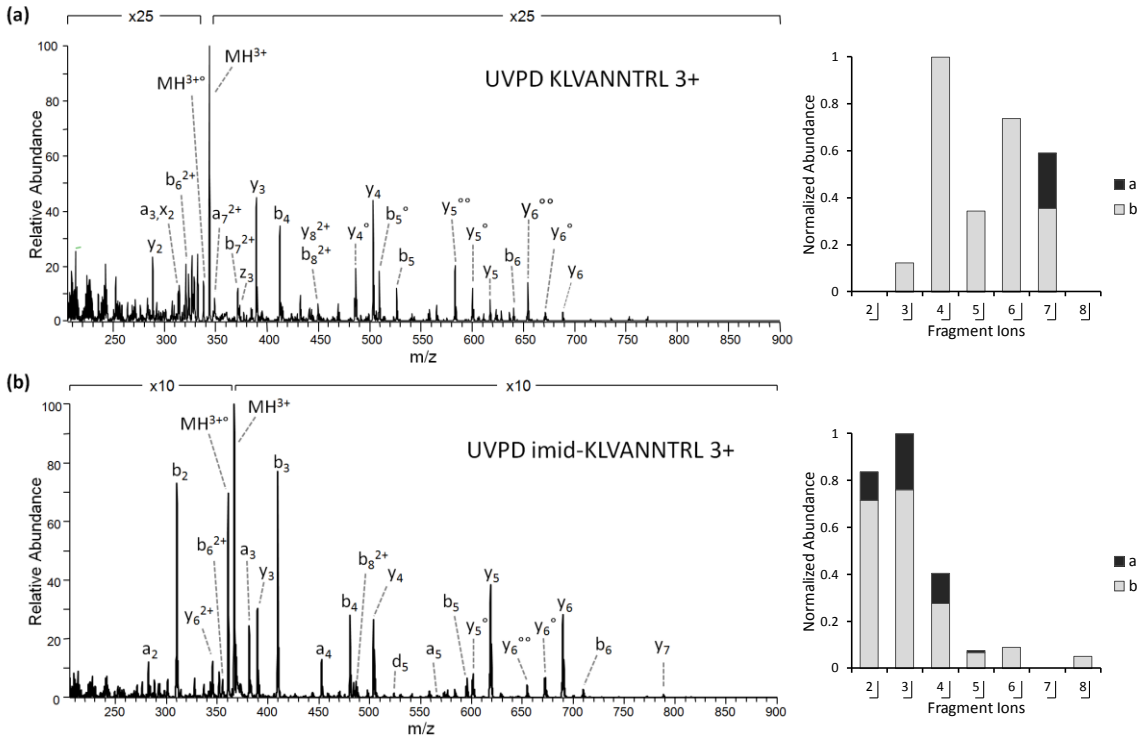


Figure 3.4 MS/MS spectra showing (a) UVPD of underivatized KLVANNTRL 3+, and (b) UVPD of imidazolinylated KLVANNTRL 3+. Peaks labeled with a "°" sign denote loss of H_2O or NH_3. Companion bar graphs displaying distribution of N-terminal ion abundances are shown beside each spectrum with y-axis normalized to 1.

To further illustrate the shifts in *a* and *b* ion abundances following imidazolinylation, the ratios of the normalized average abundances of *a* and *b* ions for each of the eight peptides upon UVPD were calculated and plotted in **Figure 3.5**. Ratios greater than 1.0 or less than 1.0 indicate dominant *a* ion or dominant *b* ion series, respectively. For seven out of the eight singly protonated peptides, the *a* ions increased in abundance relative to the *b* ions after derivatization (**Figure 3.5A**). The one exception, KLVANNTRL, shows the opposite trend, and this is attributed to the formation of a single unusually abundant *b* ion (b_5). If this anomalously abundant product ion is

excluded, the *a*/*b* ratio increases to 1.12. The shift in the dominance of *a* versus *b* ions is attributed to a decrease in proton mobility that is experienced upon imidazolinylation.[37] When the proton mobility is lost, other fragmentation channels are increasingly competitive. With UVPD, the bond between the carbonyl-carbon and the α-carbon is photolytically cleaved to form *a* and *x* ions,[39] with *a* ions being the major product normally observed for Lys-N peptides. Moreover, addition of a second proton substantially enhances the formation of *b i*ons along with the complementary *y* ions. For this reason, *b* ions remain dominant compared to other N-terminal ions. This outcome is illustrated in **Figure 3.5B** for doubly charged peptides by the *a*/*b* ratio that never exceeds 1.0 regardless of the lysine modification. In short, the abundance of the *a* ions increased relative to the *b* ions after imidazolinylation, but the *a* ions never gain dominance as observed for the singly charged peptides.

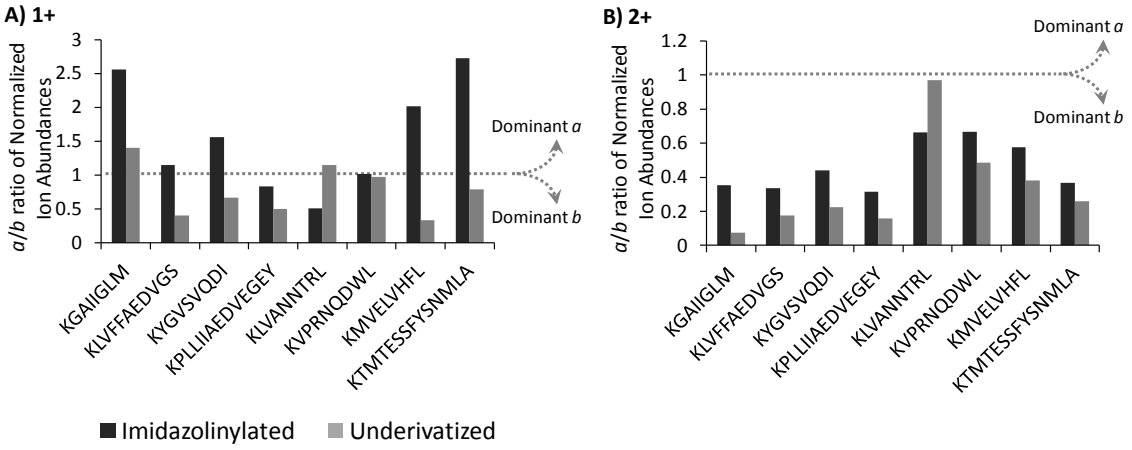


Figure 3.5 Comparison of *a* and *b* ion abundances for UVPD of (A) 1+ and (B) 2+ imidazolinylated and underivatized peptides. Ion abundances are normalized to the most abundant *a*, *b*, or *c* ion. Normalized abundances are averaged across all *a* ions (*b* ions) for each peptide and then ratios are calculated.

Figure 3.6 compares the total product ion distribution of N- and C-terminal ions in terms of the summed percentage of N-terminal versus C-terminal ions for the singly and doubly protonated peptides upon UVPD. For the underivatized peptides, a greater percentage of the product ions arise from the C-terminus (on average 27%). The percentage of C-terminal ions decreased to less than 3% for six of the peptides following imidazolinylation. The other two peptides, KLVANNTRL and KKVPRNQDWL, showed greater portions of C-terminal ions, 8% and 13% respectively, likely because each of these peptides contains a basic arginine residue that is able to compete with the imidazolinylated lysine for retention of a proton. The C-terminal series is less effectively suppressed upon UVPD of doubly charged peptides, as summarized in **Figure 3.6B**. With the addition of more charges to the peptide backbone, the likelihood of a proton residing on the C-terminal side of the cleavage site in the peptide backbone is greater.

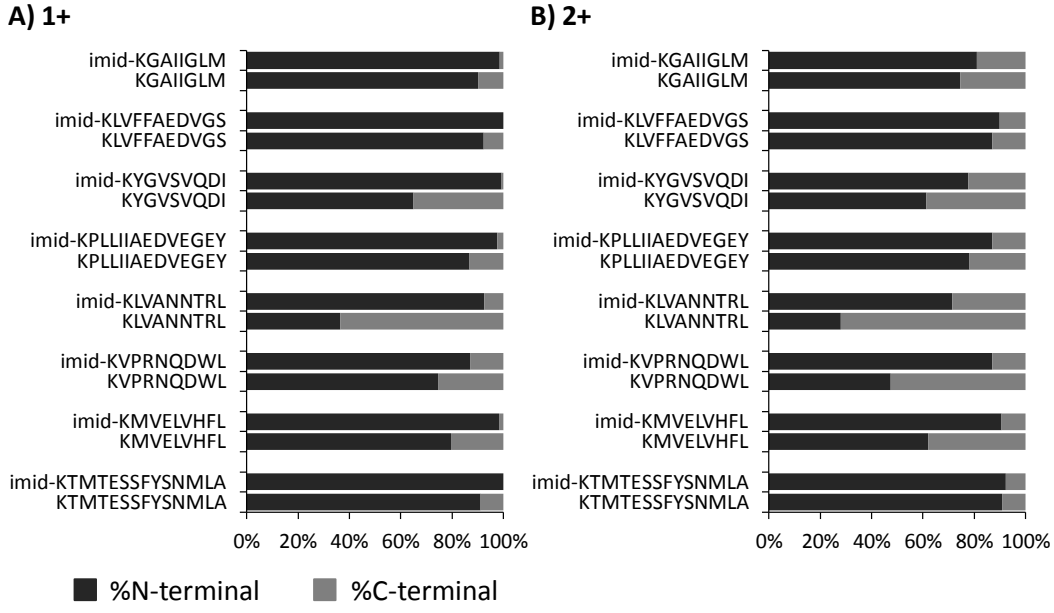


Figure 3.6 Total fragment ion distribution in terms of percentage of N- and C-terminal ions for UVPD of (A) 1+ and (B) 2+ underivatized and imidazolinylated peptides

3.4.4 LC-MS/MS Analysis of Lys-N Digested BSA

Based on the consistent formation of *a*, *b*, and *c* triplets or *a* and *b* pairs upon UVPD of the series of imidazolinylated model peptides, the strategy was extended for an LC-MS/MS proteomic workflow - in this case demonstrated for bovine serum albumin in conjunction with Lys-N digestion.

Peptides	**Charge State**	**a**	**b**	**c**
KPLLE	1+	100	100	100
KAFDE	1+	100	100	100
KLVTDLT	1+	100	100	100
KAEFVEVT	1+/2+	100/100	100/100	67/33
KDLGEEHF	1+/2+	100/67	100/83	83/17
KQTALVELL	1+	100	71	86
KGACLLP	1+	80	60	40
KFGERAL	1+	60	60	20
KIETMRE	1+	20	40	40
KTVMENFVAFVD	2+	100	100	10
KEYEATLEECCA	2+	80	90	10
KYICDNQDTISS	2+	60	100	0
KATEEQL	2+	80	80	20
KLVVSTQTALA	2+	78	89	11
KLVNELTEFA	2+	88	88	38
KSEIAHRF	2+	83	83	33
KSLHTLFGDELC	2+	70	80	40
KTCVADESHAGCE	2+	45	64	27
KLGEYGFQNALIVRYTR	2+	47	40	7
KYNGVFQECCQAED	2+	42	58	8
KDDSPDLPKL	2+	11	33	22
KVPQVSTPTLVEVSRSLG	3+	50	81	13
KVASLRETYGDMADCCE	3+	40	73	20
KGLVLIAFSQYLQQCPFDEHV	3+	32	47	16
KECCHGDLLECADDRADLA	3+	47	47	6
KQEPERNECFLSH	3+	35	64	0

Table 3.2 Peptides identified by UVPD from a BSA Lys-N digest with corresponding charge state and *a*, *b*, *c* sequence coverage (%)

Table 3.2 lists the peptides identified in the digest along with their charge states and *a*, *b*, and *c* sequence coverages. In total 26 peptides accounting for 289 amino acids (out of 607) were identified to give a sequence coverage of approximately 50%. As expected based on the model peptide experiments described above, the most complete triplet sets of *a*, *b*, *c* ions are achieved for singly charged peptides with several of the Lys-N peptides yielding a complete or nearly complete golden set. Among these peptides is singly charged KAEFVEVT, shown in **Figure 3.7**, that only lacks two N-terminal *c* ions (c_5, c_7) but for which the entire series of *a* and *b* ions is detected. While the singly charged peptides show the cleanest golden triplet sets, they only account for about one third of the total peptide population, and the doubly charged peptides are also heavily relied upon for *de novo* sequencing. Many of these peptides maintain high sequence coverage of *a* and *b* ions, but the abundances of *c* ions are diminished. The *a* and *b* series constitute a golden pair, and complementary *c* ions are useful to validate the pairs but are not necessarily needed for *de novo* sequencing.

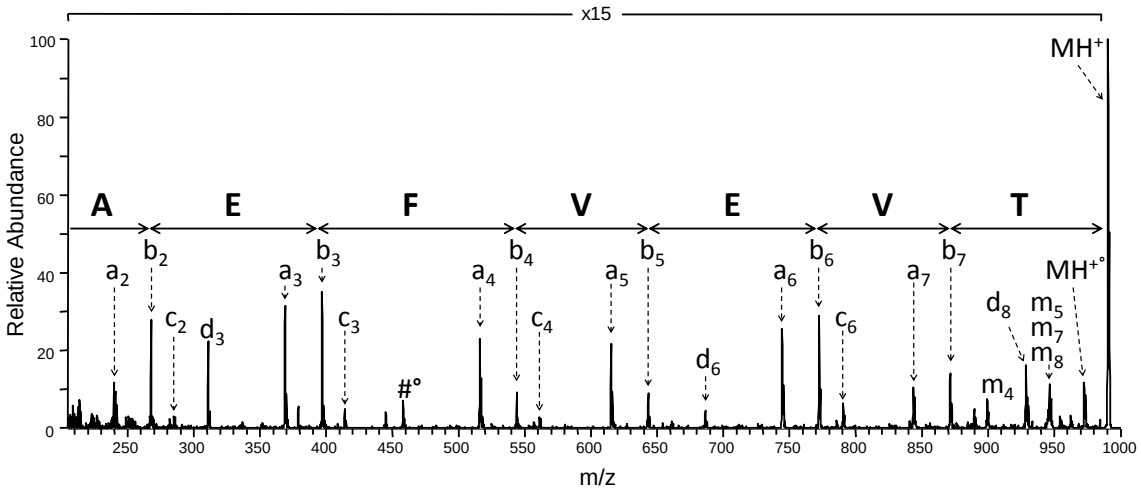


Figure 3.7 UVPD of singly charged imidazolinylated Lys-N peptide KAEVFVEVT from BSA

The percentages of golden triplets and pairs as well as single (or absent) product ions are shown for the Lys-N BSA peptides in **Figure 3.8**.

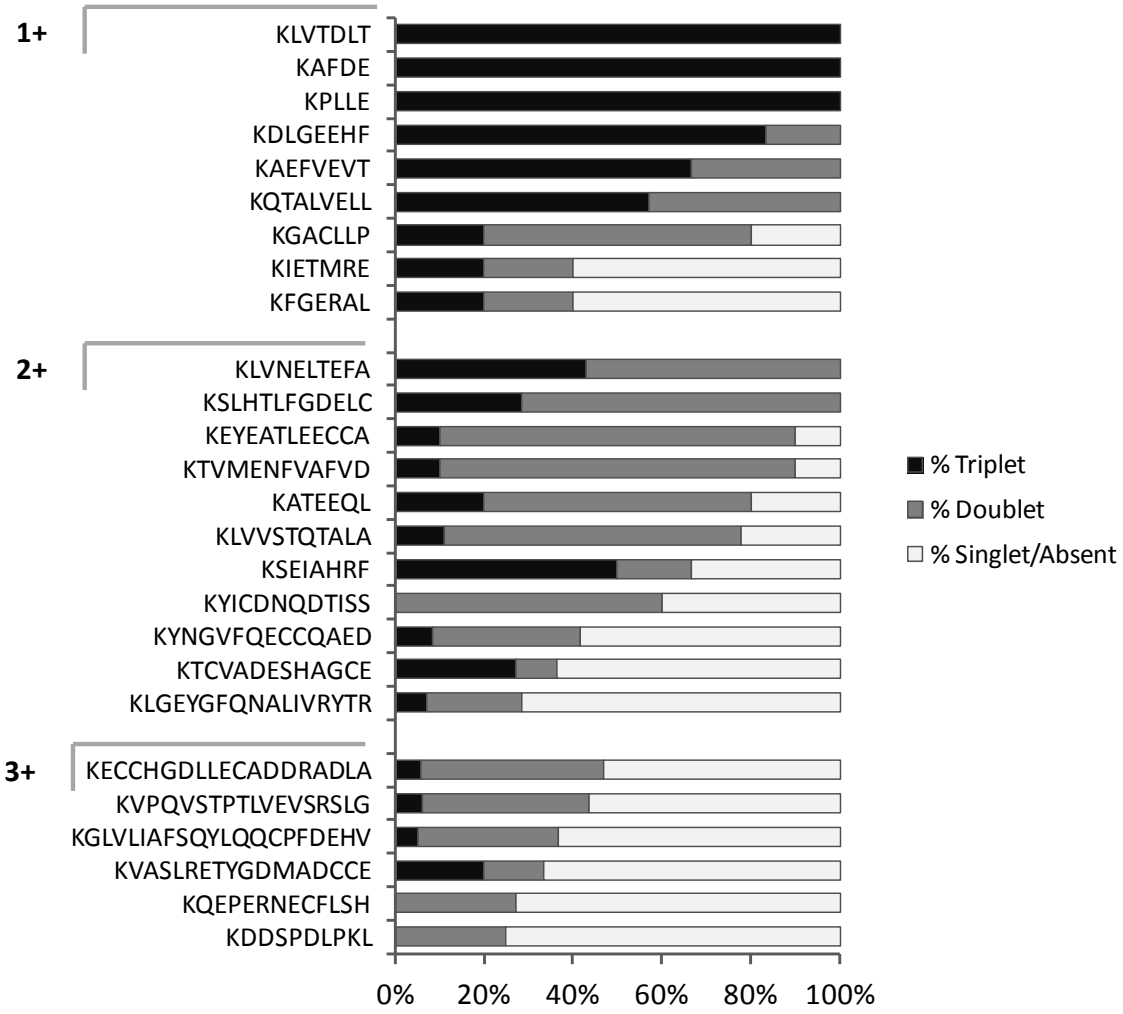


Figure 3.8 Percentages of golden ion triplets, golden ion pairs, and singlet/absent ions for BSA lys-N peptides obtained by UVPD after imidazolinylation of the digest. Singly, doubly, and triply charged peptides are grouped in the top, middle, and bottom portions of the graph.

Singly and doubly charged peptides show a large percentage of product ions that are part of golden sets. Exceptions occur for peptides that have internal basic residues presumably because these basic sites sequester the ionizing proton away from the N-terminal lysine. All of the observed triply charged peptides contain a second basic site, and consequently individual product ions (i.e. *a* or *b* ions but not both) are more common than golden sets, the latter which account for less than 50% of the total product ion population. Furthermore, the triply charged peptides generally exhibit poor sequence coverage for all N-terminal ions including *b* ions. The increased length of the triply charged peptides, many of which contain missed cleavages, may account for the diminished sequence coverages. Missed cleavages that result in an internal imidazolinylated lysine produce results in an erosion in the formation of the golden series similar to that observed for peptides containing arginine, thus reinforcing the importance of the proteolytic efficiency in generating Lys-N terminated peptides.

As a comparison to our UVPD/imidazolinylation strategy for *de novo* sequencing, a tryptic BSA digest was analyzed by LC-MS/MS (using conventional CID as the activation method) followed by *de novo* sequencing using PEAKS Studio 5.3. *De novo* sequencing based on CID data derived from non-derivatized tryptic peptides is representative of the universal standard, and therefore these results provide a benchmark for assessing improvements to *de novo* sequencing based on the novel Lys-N/imidazolinylation/UVPD protocol. PEAKS was chosen instead of other *de novo* sequencing software based on its superior performance in terms of the number of accurate peptide identifications and ease of use.[40] After applying filters based on the total local confidence (TLC) and average local confidence (ALC) of amino acid assignments, a total of 105 *de novo* peptides were identified, listed in **Table 3.3**.

Sequence	TLC	ALC (%)	z	Local Confidence (%)
KQTALVELLK	9.1	91	2	74 84 92 93 96 95 95 97 98 82
LVNELTEFAK	9.1	91	2	94 94 95 96 96 97 96 92 93 53
WPVSVLLR	7.1	88	2	84 91 92 87 88 95 98 69
MQTALVELLK	8.7	87	2	92 87 91 93 93 94 91 88 89 52
*LVNELTEFAK	8.6	86	2	86 85 90 94 95 95 88 85 87 45
GFAVSVLLR	7.7	85	2	74 73 84 91 86 88 95 98 72
LVTDLTK	5.8	83	1	75 72 74 92 95 96 72
EVCTEDYLSLLLNR	11.5	82	2	61 56 73 96 97 98 86 98 97 97 99 72 72 43
MTCAEDYLSLLLNR	11.5	82	2	69 64 79 94 95 97 80 97 96 97 95 69 69 39
TLYLVQDLMETDLYK	12.2	81	2	57 55 84 89 91 84 86 87 96 95 95 80 78 87 49
TATVSLPR	6.5	81	2	91 90 74 71 84 80 90 65
APLMTYAVSVLLR	10.5	81	3	36 40 40 80 96 92 92 95 94 96 97 99 84
CCTESLVNR	7.2	80	2	9 96 76 96 96 96 89 89 69
EPCTEDYLSLLLNR	11.1	80	2	62 53 69 92 95 95 75 96 97 97 96 70 70 38
RQTALVELLK	7.9	79	2	49 49 96 93 93 96 93 83 86 47
TVMENFVAFVDK	9.5	79	2	43 42 63 95 92 96 97 97 97 81 83 54
WFLGSFLYEYSR	9.3	78	2	50 86 28 98 97 97 98 85 85 60 85 56
**QTALVELLK	7	77	2	25 27 82 91 94 93 96 98 86
HAQENFVAFVGK	9.2	77	2	57 55 80 86 85 92 92 94 97 73 74 31
LCVLHEQTPVSTR	9.8	76	3	88 88 98 97 92 92 85 95 68 54 46 49 25
YLCDNQDTLSSK	9.1	76	2	41 73 45 96 73 78 96 95 96 98 99 9
HLVDESCNLLK	8.2	75	2	90 89 98 98 86 61 57 32 63 84 58
CLYELAR	5.2	74	2	79 79 46 68 66 95 83
LWYGFQNELLVR	8.9	74	2	0 0 92 98 98 70 63 91 97 98 98 78
FQDLGEEHFK	7.3	73	2	52 48 98 98 78 78 96 77 53 42
DWYGFQNALLVR	8.6	72	2	57 41 23 26 91 85 86 94 96 96 97 66
CHGGGTFK	5.7	72	2	84 83 71 52 50 67 94 68
FLALGSFLYEYSR	9.2	71	2	32 31 70 95 97 95 95 97 76 75 44 72 40
CVVGGPLR	5.7	71	2	51 51 77 80 81 81 82 60

Table 3.3 PEAKS *de novo* peptides assigned from LC-MS-CID of tryptic BSA

Table 3.3 Continued

QDMADDKEACFAVEGPK	12	71	2	33 27 29 82 83 81 59 96 97 93 92 93 94 91 64 67 9
AEFVEVTK	5.6	70	2	63 61 92 80 49 48 88 72
MFVSVLLR	5.5	69	2	45 43 86 89 85 82 85 34
FQDLGEEHFK	6.8	68	3	70 40 44 51 88 91 96 83 63 49
RHPEHPVSVLLR	8.1	68	2	91 97 97 96 40 38 35 45 47 80 89 49
SMEVEVTQLVTDLTK	10	67	2	22 16 17 89 38 37 93 89 97 96 96 93 80 82 49
RPPELLYYANK	7.3	66	2	69 72 74 74 75 92 74 39 60 62 34
KLHLVDEPKNLLK	8.6	66	2	32 36 92 98 99 99 94 50 34 32 52 87 53
CTVADESHAGCEK	8.5	65	2	26 25 95 95 94 97 95 96 94 58 26 26 17
FETSVLLR	5.2	65	2	53 40 43 72 72 91 94 53
RALLAGNR	5.2	65	2	79 35 35 63 73 71 74 87
TSEAWSVAR	5.8	64	2	35 34 89 91 26 85 76 83 52
NYVLHEKTWVVR	7.7	64	2	30 32 69 97 97 96 81 84 14 54 67 40
SSLSSVVGR	5.7	64	2	74 58 51 67 68 72 69 73 36
QAAVSVLLR	5.7	63	2	35 35 37 87 92 91 81 81 23
LVHVTTTRR	5.7	63	2	93 93 95 83 44 24 25 47 59
LCVLHEQTPVSEK	8.2	63	2	15 15 18 97 98 98 91 97 88 86 39 45 24
KEYFADYEER	6.3	63	2	74 87 28 56 66 64 83 40 46 81
FSALTPEDDTFPK	8.1	63	2	27 26 86 96 98 89 22 21 39 54 79 96 75
FTGTSEEYLSLLLNR	9.4	63	2	75 53 24 24 64 77 65 52 59 97 97 97 57 57 33
LYGDALLVR	5.6	62	2	77 76 53 29 30 58 91 91 51
KPQGGGMYTLK	6.8	62	2	70 60 41 72 82 92 95 69 33 38 24
MRVTEEYLSLLLNR	8.7	62	2	20 20 24 86 94 88 63 61 94 95 94 47 47 27
KPEEGFAHASSK	7.4	62	2	59 21 21 37 73 54 51 69 85 85 85 94
GCNELTEFAK	6.2	62	2	51 51 70 54 91 95 78 50 51 19
DCVLHEQTAFSTR	8	62	2	80 80 91 93 94 91 60 58 69 25 17 19 18
SLHTLFGDELCK	7.4	61	1	32 27 72 91 97 97 95 92 66 18 29 14
MRDSEEYLSLLLNR	8.6	61	2	27 20 14 28 91 95 63 66 97 97 97 62 62 33
FCLMSTFNLCK	6.7	61	2	76 67 33 35 89 88 88 76 43 43 24

Table 3.3 Continued

MCLSMDYLSLLLNR	8.4	60	2	20 20 15 40 58 94 79 95 94 98 97 48 48 29
RGKVDNQDTLSSK	7.6	59	2	13 11 11 46 53 70 85 93 94 89 89 93 9
KKTYSLALLK	5.8	58	2	60 82 94 33 24 32 45 81 86 40
GFHTLFGDELCK	7	58	2	13 13 15 84 98 97 92 89 65 55 31 38
KLECCDPKLLEK	6.9	58	2	67 70 88 59 60 63 63 23 29 61 65 38
KVPKVSTPTLVEVSR	8.6	58	2	70 85 92 64 92 93 86 27 16 17 75 59 30 31 19
QELRETYTAMADCCAK	9.2	58	2	40 31 30 38 93 96 80 35 35 95 96 97 68 22 36 19
WLYGFQKGLLVR	6.9	57	2	16 17 14 96 97 58 11 15 90 98 98 75
CSVDEAARGPLK	6.8	57	2	76 66 68 80 70 67 33 22 24 33 51 89
VPAANGSFLYEHNR	8	57	2	20 20 27 77 71 92 89 89 95 68 53 24 42 23
SPCTGAATR	5.1	57	2	77 77 76 41 27 30 38 44 95
AEFVEVNDLVTHLTK	8.5	57	2	79 78 86 90 96 84 0 34 67 36 35 61 38 40 20
MEAYYLLPEAYFYPFK	9	56	2	58 54 72 96 96 77 65 33 32 80 78 70 20 18 18 27
FADEQLAPAR	5.6	56	2	60 58 96 73 45 76 71 34 23 26
ALQADSSVAR	5.6	56	2	84 83 87 83 51 50 23 22 40 37
GNYNLVGGRCK	6.2	56	2	83 65 66 47 85 94 65 28 28 28 23
YLCHNKDTLSSK	6.7	56	2	75 84 53 38 14 19 22 80 84 92 95 9
AFEVDDLDLVTDLTK	8.3	56	3	54 46 34 34 67 65 73 93 89 83 84 48 21 21 14
FWLGSFLYEYTR	6.6	55	2	25 18 15 97 98 97 97 79 53 26 40 13
*LVNELTEFAK	5.5	55	1	64 63 80 87 89 80 22 21 21 18
WVSTPTLPMWR	6	55	2	63 81 89 80 68 67 78 36 22 0 16
LVNELTGYVR	5.5	55	2	72 63 41 46 76 78 67 39 42 19
GVEATPGNLEVTK	7	54	3	41 37 64 63 33 32 33 50 52 93 78 80 44
GSTLTNNAVSVLLR	7.6	54	3	22 25 38 90 50 14 15 21 78 89 89 89 91 42
ALYGTVEVSR	5.4	54	2	18 15 17 54 74 85 91 67 70 42
MMPAELKTVMENFVAFWR	9.7	54	2	26 21 29 30 52 63 67 68 70 88 70 44 82 87 93 29 23 17
CDKFSLVNSPPDR	6.9	53	2	89 76 14 18 85 92 90 67 47 39 25 28 11
TAVVEQAAVDALLR	7.4	53	3	42 25 24 25 36 30 46 87 91 54 48 80 92 52

Table 3.3 Continued

DRHADLCTLPGQCK	7.3	52	2	16 15 13 87 59 56 80 66 70 83 65 40 53 25
RPVDKSGFAEPK	6.3	52	2	79 87 86 83 75 47 43 26 26 21 30 16
NNHLEALLNR	5.2	52	2	26 16 16 36 64 71 93 76 76 45
FWLGSPGLGLEYTR	7.3	52	2	24 0 88 96 72 70 71 73 29 30 87 27 43 13
GYTEAVNEHMENFVAFKR	9.3	52	3	60 56 63 65 32 32 73 48 50 71 81 80 61 23 23 39 45 23
WKTGGRKVQLLVR	6.7	52	3	48 40 14 31 35 20 20 27 71 94 96 97 73
KLFLLSLLLNR	5.7	52	3	63 79 24 24 24 29 43 89 73 74 43
VTFKGFQATALLVR	7.2	52	2	45 24 22 27 72 51 50 29 28 32 86 95 96 62
HVSSRSCHGDLLECWDR	8.8	52	3	67 60 30 30 20 20 14 23 87 93 98 97 95 53 25 25 35
KLECCHQPLLEK	6.2	51	2	82 90 89 61 14 14 14 67 63 42 43 29
MNFPLQTQPLLTR	6.7	51	2	77 70 40 21 22 66 93 78 75 66 18 18 16
LPDTLVHAAVPK	6.1	51	2	30 29 24 46 60 68 41 40 49 66 82 73
GSEDLGGGALDTR	6.6	51	2	35 35 62 85 81 71 51 38 36 53 30 47 29
CPQDQFEQLGEYGFQNALVLR	10.7	51	2	11 11 11 22 23 87 30 62 94 86 88 97 96 93 79 55 33 27 18 20 12
RGYLDAHFTFLQHMR	7.6	51	2	41 38 39 77 57 54 56 62 92 88 78 0 25 25 18
YNTFARSCMENFVAFVDK	9.1	50	3	12 12 14 49 15 15 27 20 85 44 71 92 97 97 90 61 62 34
KSSMTMTYFEEK	6	50	2	68 76 58 56 68 60 60 60 26 18 17 29
AQFASVEVSR	5	50	2	49 21 21 40 52 75 77 60 64 36
TTKEVAVPSPRR	6	50	2	48 38 37 46 85 96 78 54 27 27 34 25

Table 3.3 PEAKS *de novo* peptides assigned from LC-MS-CID of tryptic BSA. TLC and ALC are listed along with the peptide charge and local percent confidence scores of amino acid assignments. Peptide confidence scales with ACL as follows: 90% is very high confidence, >80% is high, 60%-80% is medium, and <60% is low. Rows highlighted in gray show sequences that match to BSA tryptic peptides. Sequences marked with an asterisk denote a repetitive identification while sequences marked with a double asterisk denote redundant sequence information (i.e. a peptide that is encompassed by another larger peptide having missed cleavages)

From this list ten peptides corresponded to BSA peptides, while the rest were false positives. Included in this set of ten peptides were three peptides which provided redundant protein sequence information based on repeat identifications of the same peptide and overlapping amino acid sequences in peptides having different numbers of missed cleavages, for example QTALVELLK and KQTALVELLK. Further manual inspection of the CID spectra and assessment of the PEAKS-assigned peptides in a manner similar to that undertaken for interpretation of the UVPD spectra reveals 11 additional peptides, listed in **Table 3.4**, having sequences that closely agree with BSA peptides except for mis-assigned isobaric or nearly isobaric amino acids that cannot reliably be distinguished using low mass accuracy data. After manually-assisted *de novo* sequencing, 208 different amino acids were identified accounting for 35% of the total BSA sequence.

PEAKS de novo peptide	Actual BSA peptide	%correct	Miss assignments
EV(228.111)CTEDYLSL**L**LNR	**MP(228.093)**CTEDYLSL**I**LNR	86	I/L + missed pair
W(186.079FLGSFLYEYSR	**DA(186.064)**FLGSFLYEYSR	85	missed pair
LCVLHE**Q**TPVS**TR(257.149)**	LCVLHE**K**TPVS**EK(257.138)**	85	Q/K + missed pair
YLCDNQDT**L**SSK	YICDNQDT**I**SSK	92	I/L
F**Q**DLGEEHFK	F**K**DLGEEHFK	90	Q/K
DW(301.106)YGFQNAL**L**VR	**LGE(299.1482)**YGFQNAL**I**VR	77	I/L + missed consecutive three
RHPE**HP(234.112)**VSVLLR	RHPE**YA(234.100)**VSVLLR	83	missed pair
KLHLVDEP**K**NL**L**K	**LK**HLVDEP**Q**NL**I**K	85	Q/K + I/L + flip-flop
CTVADESHAGCEK	**TC**VADESHAGCEK	85	flip-flop
KLECCDPKLLEK	**LK**ECCDKPLLEK	83	flip-flop
KVP**K**VSTPTLVEVSR	KVP**Q**VSTPTLVEVSR	93	Q/K

Table 3.4 Manually assisted PEAKS *de novo* sequencing. Amino acids in bold and underlined denote incorrect assignments with the molecular weight listed in parentheses. I/L and Q/K denote mis-assignment of Ile and Leu or Gln and Lys residues. "Flip-flop" under the mis-assignment column denotes correct amino acids that are assigned the wrong location in the sequence.

3.5 CONCLUSION

The present study demonstrates the ability to generate golden sets of *a*, *b*, and *c* ions through 193 nm UVPD activation of imidazolinylated Lys-N peptides. UVPD of unmodified Lys-N peptides can provide a similar set of N-terminal ions; however, a greater number of C-terminal ions are formed and the N-terminal ion distribution is skewed in favor of the *b* ion series. Following imidazolinylation, the *a* and *c* ion series increase in abundance with the *a* ions gaining dominance for singly charged peptides. The quality of the golden triplet series decreases with increasing precursor ion charge state because a second proton makes additional fragmentation channels more accessible due to greater charge mobility. In conjunction with the Lys-N protease, this methodology was applied to a BSA digest for which 50% sequence coverage was achieved. A high percentage of the total product ion population was comprised of golden sets for the singly and doubly charged peptides upon UVPD, making them amenable to the proposed *de novo* sequencing strategy. We anticipate that using a customized *de novo* algorithm that caters more specifically to the type of fragment ions generated upon UPVD of imidazolinylated lys-N peptides could result in even more accurate *de novo* sequencing results. We are currently investigating new automated spectral interpretation strategies that can utilize the unique peptide fragmentation chemistry defined in this study.

3.6 REFERENCES

(1) Domon, B.; Aebersold, R. *Science* **2006**, *312* (5771), 212–217.
(2) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21* (3), 255.
(3) Aguiar, M.; Haas, W.; Beausoleil, S. A.; Rush, J.; Gygi, S. P. *J Proteome Res* **2010**, *9* (6), 3103–3107.
(4) Taylor, J. A.; Johnson, R. S. *Anal. Chem.* **2001**, *73* (11), 2594–2604.
(5) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.
(6) Zhang, Z. *Anal. Chem.* **2004**, *76* (21), 6374–6383.
(7) Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. *Anal. Chem.* **2005**, *77* (22), 7265–7273.
(8) Frank, A.; Pevzner, P. *Anal. Chem.* **2005**, *77* (4), 964–973.
(9) Bern, M.; Goldberg, D. *J. Comput. Biol.* **2006**, *13* (2), 364–378.
(10) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. *Anal. Chem.* **2007**, *79* (13), 4870–4878.
(11) Qin, J.; Chait, B. T. *J Am Chem Soc* **1995**, *117* (19), 5411–5412.
(12) Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. *J Am Chem Soc* **1999**, *121* (22), 5142–5154.
(13) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. *Anal. Chem.* **2005**, *77* (18), 5800–5813.
(14) Reilly, J. P. *Mass Spectrom. Rev.* **2009**, *28* (3), 425–447.
(15) Brodbelt, J. S.; Wilson, J. J. *Mass Spectrom. Rev.* **2009**, *28* (3), 390–424.
(16) Ly, T.; Julian, R. R. *Angew. Chem. Int. Ed.* **2009**, *48* (39), 7130–7137.
(17) Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (2), 197–206.
(18) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.
(19) Cagney, G.; Emili, A. *Nat Biotech* **2002**, *20* (2), 163–170.
(20) Münchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem.* **2000**, *72* (17), 4047–4057.
(21) Brancia, F. L.; Montgomery, H.; Tanaka, K.; Kumashiro, S. *Anal. Chem.* **2004**, *76* (10), 2748–2755.
(22) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. *Anal. Chem.* **2005**, *77* (19), 6300–6309.
(23) Hennrich, M.; Mohammed, S.; Altelaar, A.; Heck, A. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965 – 1965.
(24) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proc. Natl. Acad. Sci.* **2000**, *97* (19), 10313–10317.
(25) Zhang, L.; Reilly, J. P. *Anal. Chem.* **2010**, *82* (3), 898–908.
(26) Madsen, J. A.; Brodbelt, J. S. *Anal. Chem.* **2009**, *81* (9), 3645–3653.
(27) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136.

(28) Vasicek, L.; Wilson, J.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384 – 384.
(29) Miyashita, M.; Hanai, Y.; Awane, H.; Yoshikawa, T.; Miyagawa, H. *Rapid Commun. Mass Spectrom.* **2011**, *25* (9), 1130–1140.
(30) Hohmann, L.; Sherwood, C.; Eastham, A.; Peterson, A.; Eng, J. K.; Eddes, J. S.; Shteynberg, D.; Martin, D. B. *J. Proteome Res.* **2009**, *8* (3), 1415–1422.
(31) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. *Mol. Cell. Proteomics* **2009**, *8* (4), 650–660.
(32) Hennrich, M. L.; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J. R.; Mohammed, S. *Anal. Chem.* **2009**, *81* (18), 7814–7822.
(33) Carabetta, V.; Li, T.; Shakya, A.; Greco, T.; Cristea, I. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (6), 1050–1060 – 1060.
(34) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Anal. Chem.* **2008**, *80* (13), 4807–4819.
(35) Gu, C.; Tsaprailis, G.; Breci, L.; Wysocki, V. H. *Anal Chem* **2000**, *72* (23), 5804–5813.
(36) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Anal Chem* **2006**, *79* (2), 477–485.
(37) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35* (12), 1399–1406.
(38) Zhang, L.; Reilly, J. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (7), 1378–1390 – 1390.
(39) Weidong Cui; Thompson, M. S.; Reilly, J. P. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1384–1398.
(40) Bringans, S.; Kendrick, T. S.; Lui, J.; Lipscombe, R. *Rapid Commun. Mass Spectrom.* **2008**, *22* (21), 3450–3454.

Chapter 4

193 nm Ultraviolet Photodissociation Mass Spectrometry for Phosphopeptide Characterization in the Positive and Negative Ion Modes

4.1 Overview

Advances in liquid chromatography tandem mass spectrometry (LC-MS/MS) have permitted phosphoproteomic analysis on a grand scale, but ongoing challenges specifically associated with confident phosphate localization continue to motivate the development of new fragmentation techniques. In the present study, ultraviolet photodissociation (UVPD) at 193 nm is evaluated for the characterization of phosphopeptides in both positive and negative ion modes. Compared to the more standard higher energy collisional dissociation (HCD), UVPD provided more extensive fragmentation with improved phosphate retention on product ions. Negative mode UVPD showed particular merit for detecting and sequencing highly acidic phosphopeptides from alpha and beta casein, but was not as robust for larger scale analysis due to lower ionization efficiencies in the negative mode. HeLa and HCC70 cell lysates were analyzed by both UVPD and HCD. While HCD identified more phosphopeptides and proteins compared to UVPD, the unique matches from UVPD analysis could be combined with the HCD data set to improve the overall depth of coverage compared to either method alone.

4.2 INTRODUCTION

Post-translational modifications (PTMs) direct a variety of biological activities, many of which are critical for sustaining normal cellular function. Phosphorylation is known to modulate enzymatic activity, alter protein folding, and inhibit or promote interactions with other proteins. Abnormalities in phosphorylation pathways have been associated with various diseases, a factor that has stimulated interest in kinases and phosphatases as therapeutic targets and their substrates as potential diagnostic biomarkers.[1,2] The tremendous impact of phosphorylation has inspired the development of extremely powerful mass spectrometry strategies to map the dynamic phosphoproteome and quantify variations in phosphorylation as a function of cell state. The majority of these strategies are based on bottom-up workflows that rely on the ability to identify phosphorylation sites of proteins via MS/MS characterization of diagnostic proteolytic peptides.[3] Due to significant inroads in enrichment methods, chromatographic capabilities, and performance metrics of mass spectrometers, there have been reports of grand scale phosphoproteomics.[4–7] For example, identification of thousands of phosphopeptides, corresponding to 18,055 phosphosites and 4,708 phosphoproteins, was achieved using 1 mg of material via a 2D enrichment/fractionation strategy combining immobilized metal affinity chromatography (IMAC), hydrophilic interaction liquid chromatography (HILIC) and reversed phase LC.[4] Yet even with these advances, there continue to be major efforts to improve depth of coverage, sensitivity of phosphopeptide detection, and confidence in phosphosite localization.

The choice of MS/MS technique can significantly impact the outcome of a PTM mapping experiment, and thus a variety of fragmentation methods have been evaluated for phosphopeptide analysis. Collision induced dissociation (CID) remains popular for peptide sequencing based on the ease of implementation, robust performance, and the

multitude of available data analysis tools.[8] Because CID promotes cleavage at the most labile bonds, phosphorylated peptides readily undergo loss of neutral phosphate groups while more informative backbone fragmentation is suppressed, making it difficult to both sequence the peptide and localize the modification.[9,10] Higher energy collisional dissociation (HCD) is a beam type variation of CID that results in greater energy deposition. This facilitates cleavage of bonds other than those that are the weakest, and the phosphate neutral loss problem is largely mitigated although not entirely overcome.[11] Phosphate neutral loss can be eliminated nearly entirely using electron-based activation methods, including electron transfer dissociation (ETD) and electron capture dissociation (ECD) for which cleavage occurs at the N-Cα bond to form *c* and *z* type ions while PTMs are preserved.[12–16] Both ETD and ECD exhibit a significant charge state dependence, and peptides with insufficient charge density typically do not fragment efficiently. Charge density limitations are particularly problematic for phosphopeptides because the negatively charged phosphate decreases the overall positive charge, thus offsetting the effectiveness of ETD and ECD for phosphopeptide analysis.

Photodissociation has been explored as an alternative to collision- or electron-based dissociation methods. Infrared multiphoton dissociation (IRMPD) has been the most widely adopted, but because fragmentation occurs through the slow heating of vibrational modes, many of the same limitations that are observed for CID of phosphopeptides also apply for IRMPD.[17] At the same time, IRMPD can provide an effective means for screening complex peptide mixtures for phosphorylation based on the superior IR absorbance of phosphopeptides compared to nonphosphorylated peptides, causing the former to dissociate while the later are left intact.[18–21]

In the near IR regime, femtosecond laser-induced ionization/dissociation (fs-LID) uses ultrafast laser pulses to induce electron loss through tunneling followed by peptide

bond cleavage to form *a*, *b*, *c*, *x*, *y*, and *z* ions.[22,23] Because dissociation is initiated by electron abstraction, fs-LID is most effective for singly protonated peptides based on their lower ionization potentials compared to more highly charged ions, a trend that is opposite to CID/HCD and ETD/ECD fragmentation. Along with superior sequence coverage, singly charged phosphopeptides exhibited the greatest phosphate retention on product ions. No phosphate neutral loss was detected from *c*/*z* or *a*/*x* ions with the exception of potentially diagnostic *a*+1-98 ions observed C-terminal to the site of phosphorylation. FsLID has also been applied in negative ion mode for the analysis of multiply deprotonated phosphopeptide anions, where phosphate retaining *a*/*x* ions facilitate sequencing and modification localization.[24] Long activation times and narrow charge state restrictions ultimately have limited the adoption of fsLID for large-scale analysis.

Ultraviolet photodissociation has also been applied for the analysis of phosphopeptides. For phospho-serine containing peptides, 220 nm UVPD resulted in the dominant neutral loss of tyrosine side-chains via homolytic bond cleavage while other fragmentation channels including peptide backbone cleavage and phosphate loss were suppressed.[25] In contrast, 266 nm UVPD yielded mostly *b* and *y* ions with only minor *a* and *x* ions.[26] Likewise, phosphate neutral loss patterns were similar between 266 nm UVPD and CID even for phospho-tyrosine peptides for which the site of phosphorylation coincides with the site of photon absorption.[26] Photodissociation using 157 and 193 nm photons is also widely applicable for proteomics applications because photons are absorbed efficiently by the peptide backbone. High energy fragmentation channels are accessible upon 157 and 193 nm photoexcitation to yield *a*/*x*, *c*/z, and *d*/*v*/*w* side chain loss ions in addition to more conventional *b*/*y* ions.[17,27,28] Both 157 nm and 193 nm UVPD have been used for various proteomics applications [27,29–32], ranging from single

protein analysis to more complex whole cell lysates in both positive and negative modes.[33–36] To date, PTM analysis has been primarily conducted in negative mode based on the improved retention of labile modifications including sulfation, O-glycosylation, and phosphorylation; however, larger scale analysis in the negative mode has been significantly hindered by poor ionization efficiency.[35,37–40] Both 157 nm UVPD and 193 nm UVPD have been used to characterize phosphopeptides.[35,41–44] For example, the fragmentation behavior of singly protonated phosphopeptides with C-terminal arginine residues was explored using 157 nm UVPD, showing that the loss of the phosphate group was time-dependent and varied based on whether the phosphate group was appended to tyrosine, serine, or threonine.[42] 193 nm UVPD of singly protonated phosphopeptides that had N-terminal basic residues resulted in prominent *a*+1-98 ions observed C-terminal to sites of phosphorylation, thus providing a signature for the site of phosphorylation.[41] Product ions that are diagnostic for phosphorylation have also been observed in the 193 nm UVPD spectra of phosphopeptide anions where formation of y-H_3PO_4-NH_3 ions was observed N-terminal to the site of phosphorylation in addition to an array of phosphate retaining sequence ions.[35] Unique phosphorylation sites in the protein TrpM7 were identified using ET-UVPD in which hydrogen rich radical peptide ions generated from electron transfer without dissociation (ETnoD) were activated and dissociated using 193 nm UVPD, resulting in a rich array of product ions for which *z* ions were most abundant.[43] The most recent study compared 193 nm UVPD and HCD for the analysis of phosphopeptides from a HeLa digest and found improved modification retention using UVPD.[44]

The present study expands upon UVPD-based phosphoproteomics via analysis of phosphopeptides enriched from casein proteins in addition to HeLa and HCC70 primary ductal carcinoma cell lysates. UVPD is implemented on an Orbitrap Fusion Tribrid mass

spectrometer, and both polarities are evaluated for phosphopeptide characterization. The overall depth of coverage is compared between UVPD and HCD as well as differences in the identified phosphopeptides such as length, charge state, and number of phosphorylations. The extent of phosphate neutral loss also is also explored.

4.3 EXPERIMENTAL

4.3.1 Cell Culture and Sample Preparation

HeLa and HCC70 stable shRNA scramble cells were maintained at 37 °C and 5% CO_2 and cultured in accordance with ATCC guidelines. HeLa was grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), and HCC70 was grown in RPMI media supplemented with 10% FBS, 2 mM glutamine, 1% antibiotic/antimycotic (10 U/mL penicillin, 10 µg/mL streptomycin, 0.025 µg/mL amphotericin B), and 1 µg/mL puromycin. Cells were lysed in buffer containing 50 mM ammonium bicarbonate (pH 8), 8 M urea, 1 mM sodium orthovanadate, 100 mM sodium fluoride, 10 mM sodium pyrophosphate, 1 µM microcystin-LR, 100 nM calyculin A, complete EDTA-free protease inhibitor mix (Roche), and phosSTOP phosphatase inhibitor mix (Roche). Protein concentration was measured by Bradford assay and lysates were stored at -80°C prior to digestion.

Protein disulfide bonds were reduced by 30 minute incubation with 5 mM dithiothreitol at 56 °C then alkylated by 30 minute incubation with 15 mM iodoacetamide at room temperature in the dark. Lysates were then digested with Lys-C (Promega, Madison, WI) in a 1:200 enzyme to protein ratio for 2-4 hours at 37 °C. The urea concentration was then diluted to 1.5 M by addition of 50 mM Tris HCl, 5 mM $CaCl_2$ and trypsin (Thermo Fisher Scientific, Grand Island, NY) was added in a 1:50 enzyme to

protein ratio for overnight digestion at 37 °C. Digests were desalted using Waters tC18 Sep-Pak cartridges and dried by speed vac. Immobilized metal affinity chromatography (IMAC) was used for phosphopeptide enrichment .[45] Briefly, Ni-NTA magnetic agarose beads (Qiagen, Germantown, MD) were prepared for IMAC by shaking in 40 mM EDTA (pH 7.5) for 1 hour followed by thorough washing with water and shaking in 100 mM $FeCl_3$ for 1 hour. Protein digests were resuspended in IMAC loading buffer composed of 80% ACN containing 0.15% TFA and added to the IMAC beads following the removal of $FeCl_3$ solution. Phosphopeptide binding was carried out over 1 hour of shaking followed by three washes in loading buffer to remove non-phosphorylated peptides. Phosphopeptides were recovered from the beads by vortexing for 1 min and 15 sec in 100 µL of 50% ACN, 0.7% NH_4OH (pH ~11) and the eluent was immediately neutralized in 50 µL of 4% formic acid and dried completely. IMAC enriched HeLa samples from 500 µg of starting material were resuspended in 20 µL of 0.2% formic acid prior to LCMS analysis.

An equimolar mixture of alpha s1 casein, alpha s2 casein, and beta casein from bovine were digested in 50 mM Tris HCl (pH 7.5) with trypsin in a 1:50 enzyme to protein ratio overnight at 37°C. Digested peptides were directly added to a 5x excess (w/w) of ceramic hydroxyapatite type I, 20 µm beads (Bio Rad, Hercules, Ca) suspended in 50 mM Tris HCl. Phosphopeptide binding to HAP occurred for 1 hour at room temperature with shaking. Unmodified peptides were washed away using 50 mM Tris HCl containing 20% ACN. Bound phosphopeptides were recovered from the HAP resin by 15 minutes of shaking in 1 M KH_2PO_4 (pH 7.8). The eluent was split and desalted using both C18 and graphite solid phase extraction (SPE) to avoid the loss of very hydrophilic phosphopeptides. Desalted samples were dried and resuspended in water for LCMS analysis.

4.3.2 LC-MS/MS Analysis

Cell lysate separations were carried out on a Dionex Ultimate 3000 nano LC equipped with water as mobile phase A and acetonitrile (ACN) as mobile phase B, each containing 0.1% formic acid. For LCMS analysis 4 µL of IMAC-enriched HeLa was directly injected onto a 30 cm x 75 µm UPLC column with integrated emitter (New Objective, Woburn, MA), packed in house using 1.8 µm, 120 Å UChrom C18 (nanoLCMS Solutions, Gold River, CA). The column was heated to 60 °C inside a custom built column oven that was fabricated similarly to that described by Coon and coworkers.[46] For HeLa phosphopeptides, separation occurred over an 80 minute gradient in which the percentage of mobile phase B was increased from 2-25% B during the first 65 minutes and further increased from 25-40% B during the final 15 minutes. For HCC phosphopeptides, the same gradient steps were used, but the increase from 2-25% B was carried out more gradually over 118 minutes. Eluting peptides were analyzed using an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific Instruments, Bremen, Germany) equipped with a 193 nm Coherent ExciStar XS excimer laser (Santa Clara, CA) for ultraviolet photodissociation.[47] Orbitrap detection was used for both MS1 and MS2 measurements at resolving powers of 60K and 15K (at *m/z* 200), respectively. Data dependent MS/MS analysis was performed in top speed mode with a 3 s cycle time during which precursors detected within the range of *m/z* 400-1500 were selected for activation in order of abundance. Quadrupole isolation with a 1.6 Da isolation window was used and dynamic exclusion was enabled for 45 s. AGC targets were 2E5 for MS1 and 1E5 for MS2, with 50 ms and 100 ms maximum injection times, respectively. The signal intensity threshold for MS2 was 5E4. HCD was performed using 30% normalized collision energy and for UVPD, 2 pulses at 2.5 mJ were used. For triplicate HCD and triplicate UVPD LCMS runs, one microscan was used for MS^1 and two microscans were

used for MS2. The number of MS2 microscans was decreased to one for the HCC lysate in order to maximize the depth of coverage. Alternating activation experiments were also performed for HeLa in which the same precursor ion was selected for consecutive activation using UVPD and HCD. For these experiments, the MS/MS parameters listed above were used for run one, the MS2 signal intensity threshold was increased to 5E6 for run two, and the maximum MS2 injection time was increased to 300 ms for run three.

Negative mode LCMS analysis of phosphopeptides from alpha and beta casein proteins was carried out on a Thermo Scientific Instruments Velos Pro dual linear ion trap mass spectrometer equipped with a 193 nm Coherent ExciStar XS excimer laser (Santa Clara, CA) as previously described for UVPD.[27,48] A Dionex ultimate 3000 nano LC was used for separations in which mobile phase A was 5 mM ammonium acetate in water and mobile phase B was 5 mM ammonium acetate in 90% methanol. Peptides were loaded for 4 minutes at 5 µL/min onto a 3 cm x 100 µm trap column (New Objective, Woburn, MA) packed with 5 µm Michrom magic C18. The analytical column was 20 cm long x 75 µm OD with an integrated emitter tip (New Objective, Woburn, MA) packed with Waters Xbridge BEH C18 (Milford, MA). Separation occurred over a 45 minute linear gradient from 2-45% B. MS1 survey scans were acquired from *m/z* 400-2000 and data dependant UVPD using a single 2 mJ pulse was carried out on the top five most abundant precursor ions.

4.3.3 Data Analysis

Proteome Discoverer 1.3 was used for positive mode UVPD and HCD analysis of the IMAC enriched cell lysates. For UVPD, a non-fragment filter was applied to MS2 spectra to remove precursor peaks within a 1 Da mass window offset prior to database

search using Sequest and a human database obtained from Uniprot. All product ions including *a, b, c, x, y,* and *z* were used for spectrum matching for UVPD while only a, *b,* and *y* ions were used for HCD. Neutral losses from *a, b,* and *y* ions were considered for both UVPD and HCD. The following additional database search parameters were consistent between UVPD and HCD experiments: 10 ppm precursor mass tolerance; 0.02 Da fragment mass tolerance; N-terminal acetylation, methionine oxidation, and serine/threonine/tyrosine phosphorylation dynamic modifications; carbamidomethyl cysteine static modifications. Percolator was used for PSM validation and phosphoRS 3.0 was used for phosphosite localization. Only rank 1, high confidence PSMs were considered and an isoform probability of 75% was required for phosphosite localization.

Negative mode UVPD results for alpha and beta casein phosphopeptides were interpreted using MassMatrix Xtreme 3.0.10.16, which is programmed to search for *a, x, c, z,* and *y* type product ions. Peptide mass and fragment mass tolerances were ±1.00 Da and ±0.80 Da, respectively, and the minimum pp score was 4 while the minimum pp_{tag} score was 1.

4.4 RESULTS AND DISCUSSION

4.4.1 UVPD and HCD Analysis of Alpha and Beta Casein Phosphopeptides

UVPD and HCD were initially evaluated for phosphopeptide analysis in the positive mode using phosphopeptides generated from trypsin digestion of bovine phosphoproteins alpha S1 casein, alpha S2 casein, and beta casein. Eight singly phosphorylated peptides including DIGpSESTEDQAMDIK, EKVNELpSK, EQLSTpSEENSK, FQpSEEQQQTEDELQD, NMAINPpSK, TVDMEpSTEVFTK, VPQLEIVPNpSAEER, and YKVPQLEIVPNpSAEER as well as three doubly

phosphorylated peptides including DIGpSEpSTEDQAMEDIK, EQLpSTpSEENSK, and EQLpSTpSEENSKK were identified in LC-MS runs by both UVPD and HCD. All peptides were detected in the 2+ charge state and three peptides, FQpSEEQQQTEDELQDK, VPQLEIVPNpSAEER, and YKVPQLEIVPNpSAEER, were detected in both 2+ and 3+ charge states. HCD and UVPD spectra for TVDMEpSTEVFTK are displayed in **Figure 1A-B**.

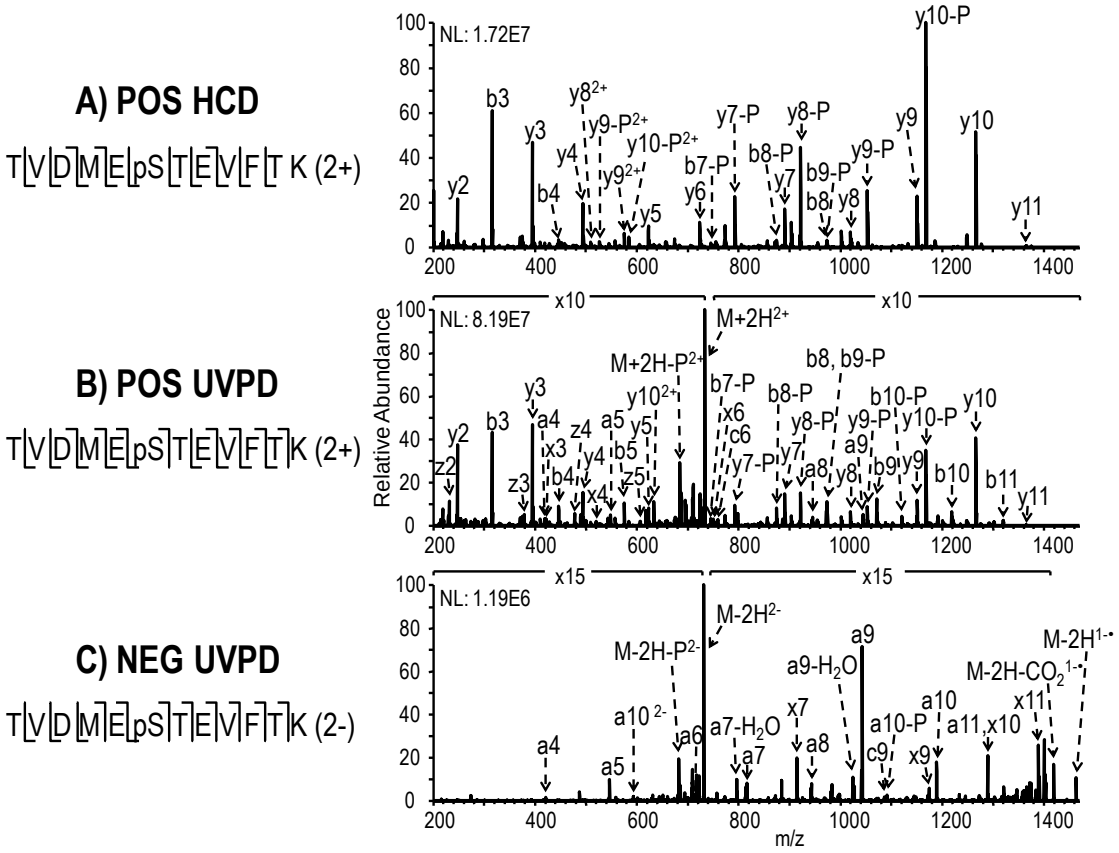


Figure 4.1 MS/MS analysis of doubly charged TVDMEpSTEVFTK (*m/z* 733.8 in positive mode and *m/z* 731.8 in negative mode) from alpha S2 casein using (A) positive mode HCD, (B) positive mode UVPD, and (C) negative UVPD. Two laser pulses of 2.5 mJ were used for UVPD, and 30 NCE was applied for HCD. Each spectrum is the average of five scans.

For both activation methods, the dominant product ions series are *b* and *y* ions; however, UVPD also yields numerous *a*, *x*, *c*, and *z* ions, which contribute to an overall richer MS/MS spectrum. The additional product ions generated by UVPD have the potential to facilitate more confident peptide identifications when using database search programs as previously demonstrated for unmodified peptides.[27] Significantly higher Sequest Xcorr scores for UVPD compared to HCD for seven out of the eleven peptides examined (**Figure 4.2 and 4.3**) are a direct indication of more confident sequence assignments.

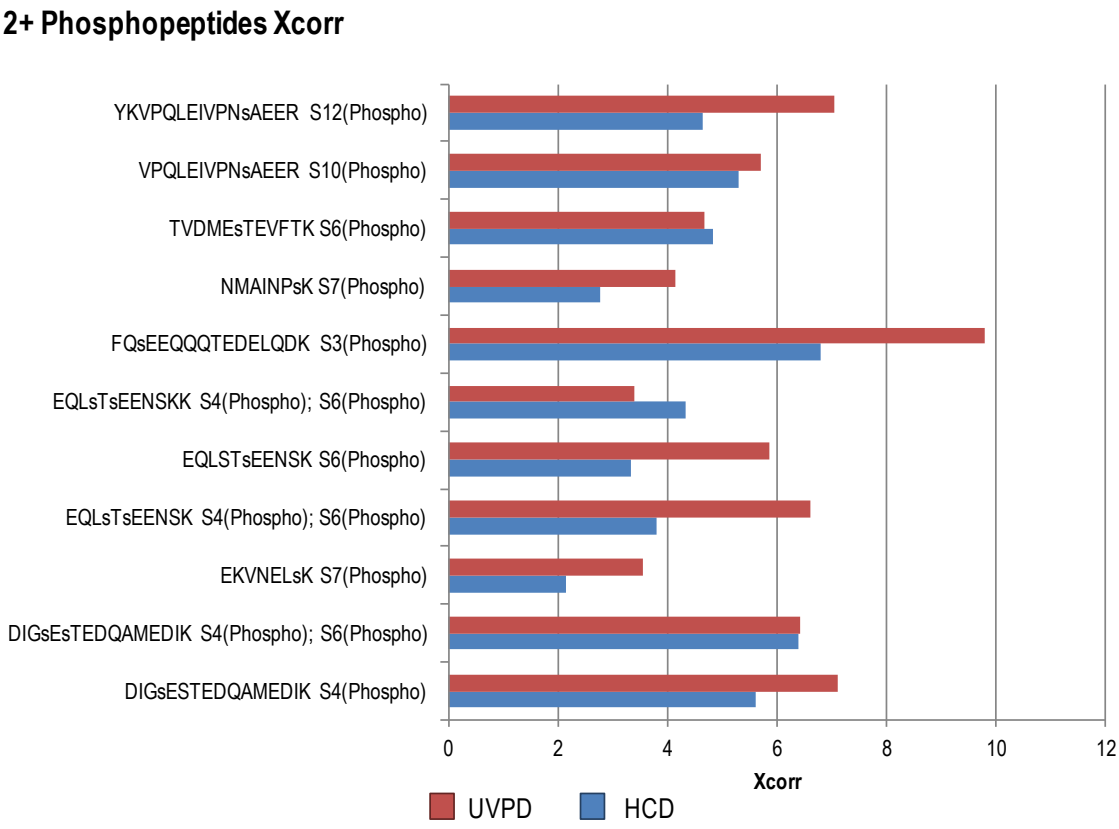


Figure 4.2 Xcorr scores for phosphopeptides in the 2+ charge state identified from alpha and beta casein using UVPD (red) and HCD (blue).

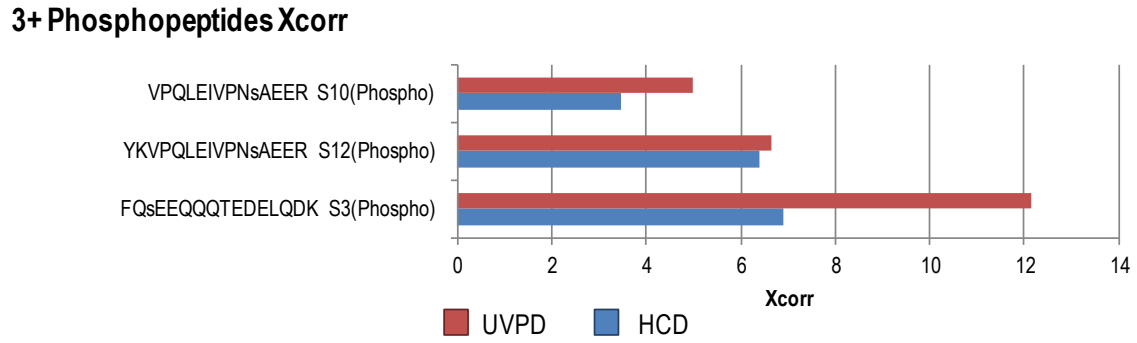


Figure 4.3 Xcorr scores for phosphopeptides in the 2+ charge state identified from alpha and beta casein using UVPD (red) and HCD (blue).

Even small increases in Xcorr can help to distinguish the correct peptide sequence from false matches as demonstrated by the commonly used ΔCn threshold of just 0.1, where ΔCn represents the difference in Xcorr between the highest scoring candidate and each lower scoring candidate.[49]

For peptides that carry PTMs, retention of modifications on product ions is critical in order to facilitate accurate localization within the peptide sequence. For TVDMEpSTEVFTK, phosphate neutral losses from *y7*, *y8*, *y9*, and *y10* are among the most abundant product ions in the HCD spectrum (**Figure 4.1A**), and the abundances of these ions exceed those of their phosphate-retaining counterparts. In contrast, the abundances of the phosphate-retaining *y7*, *y8*, *y9*, and *y10* ions produced upon UVPD of TVDMEpSTEVFTK are greater compared to the same *y*-ions, which undergo phosphate neutral loss (**Figure 4.1B**). Greater phosphate retention using UVPD is consistent with recent reports and is believed to be a result of the fast and high energy deposition upon absorption of 193 nm photons (6.4 eV per photon) compared to lower energy, step-wise collisional activation methods such as CID and HCD.[44] When the abundances of phosphate-retaining product ions were compared to the abundances of phosphate neutral

loss product ions, all eleven casein phosphopeptides exhibited improved phosphate retention when subjected to UVPD compared to HCD. In **Figure 4.4 and 4.5**, the abundance of phosphate-retaining product ions is plotted for each peptide as a percentage of the total abundance of product ions that include the site of phosphorylation. All ions series were considered for UVPD, while *a*, *b*, and *y* ions were considered for HCD. An average increase in phosphate retention of 20% for doubly charged peptides and 25% for triply charged peptides was observed using UVPD.

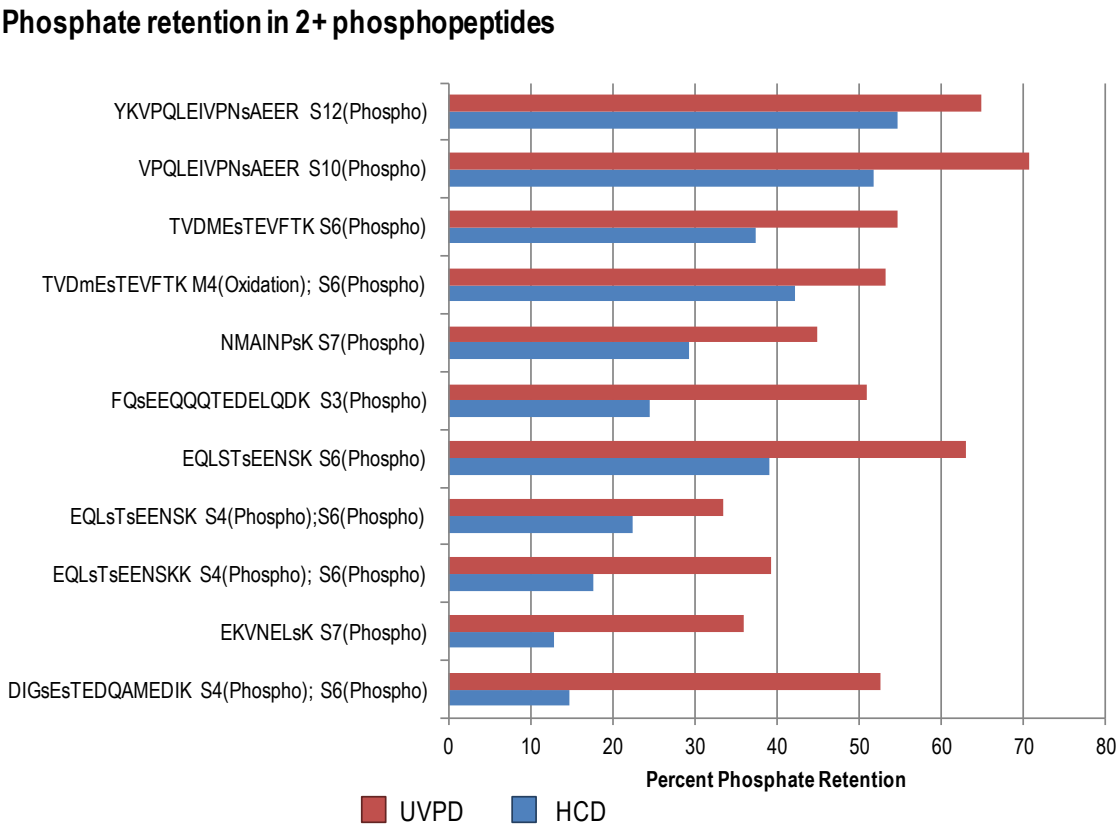


Figure 4.4 Percent phosphate retention on product ions for 2+ phosphopeptides identified from alpha and beta casein using UVPD (red) and HCD (blue).

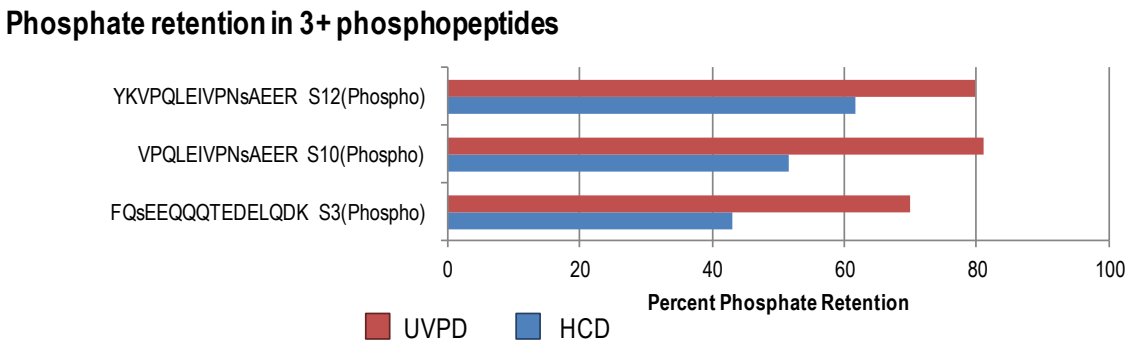


Figure 4.5 Percent phosphate retention on product ions for 2+ phosphopeptides identified from alpha and beta casein using UVPD (red) and HCD (blue).

Photodissociation of deprotonated phosphopeptides yielded significantly different fragmentation patterns compared to the UVPD spectra of protonated phosphopeptides. For UVPD of doubly deprotonated TVDMEpSTEVFTK, *a* and *x* ions accounted for full sequence coverage (**Figure 4.1C**). The phosphate neutral loss behavior also differed, with very few product ions exhibiting phosphate neutral loss for the deprotonated phosphopeptides. This is in stark contrast to both UVPD and HCD of doubly protonated TVDMEpSTEVFTK (**Figure 4.1A-B**) for which nearly every product ion that included the phosphorylated residue experienced some degree of phosphate loss. Despite superior phosphate retention of deprotonated peptides, the ionization efficiency in the negative mode is lower than the positive mode for phosphopeptides and most classes of peptides, making it challenging to adopt negative polarity UVPD-MS for large scale analysis.[50]

Only the most highly phosphorylated peptides with the greatest net negative charge, ionized more efficiently in negative mode compared to positive mode.

These peptides included:

- QMEAEpSIpSpSpSEEIVPNpSVEQK
- NTMEHVpSpSpSEESIIpSQETYK
- ELEELNVPGEIVEpSLpSpSpSEESITR
- RELEELNVPGEIVEpSLpSpSpSEESITR

To aid in the detection of these peptides, hydroxyapatite (HAP) enrichment was used based on the previously demonstrated propensity for isolation of phosphopeptides with higher phosphorylation states as well as acidic sequence motifs.[51,52] Following HAP enrichment, phosphopeptides were desalted using either graphite SPE or C18 SPE to ensure recovery of both hydrophilic and hydrophobic peptides. In **Figure 4.6A-B**, LC-MS base peak chromatograms are displayed for positive mode and negative mode ESI-MS analysis of the same HAP-enriched casein protein digest. Chromatographic peaks shaded in blue represent peptides recovered from C18 SPE and peaks shaded in purple represent peptides recovered using graphite SPE. The most hydrophilic peptides, QMEAEpSIpSpSpSEEIVPNpSVEQK and NTMEHVpSpSpSEESIIpSQETYK, were only detected following desalting on graphite while all other peptides were effectively recovered after desalting on C18. In positive mode (**Figure 4.6A**), singly and doubly phosphorylated peptides were preferentially detected while the most highly phosphorylated peptides from alpha and beta casein were not detected. These peptides, each modified with at least four and up to five negatively charged phosphates, ionized more efficiently in negative mode while at the same time ionization was suppressed for peptides in lower phosphorylation states (**Figure 4.6B**). UVPD was applied for characterization of these multiply phosphorylated peptides. Regardless of the peptide

length and number of phosphorylations, the sequence coverage afforded by UVPD remained high with no significant phosphate neutral losses observed as demonstrated for UVPD of quadruply deprotonated QMEAEpSIpSpSpSEEIVPNpSVEQK shown in **Figure 4.6C**. While UVPD in the negative ion mode has shown merit for phosphopeptide characterization, especially for peptides in high phosphorylation states, improvements in ionization efficiency are required before negative UVPD can be routinely implemented on a larger scale. Thus UVPD analysis in the positive mode will be the focus of the remainder of the study.

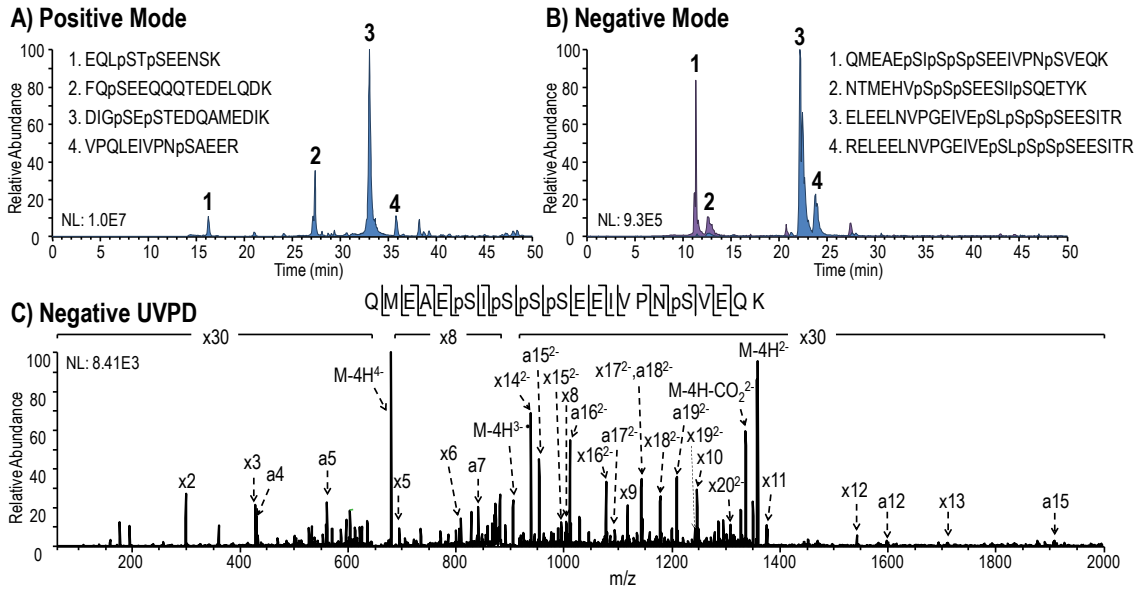
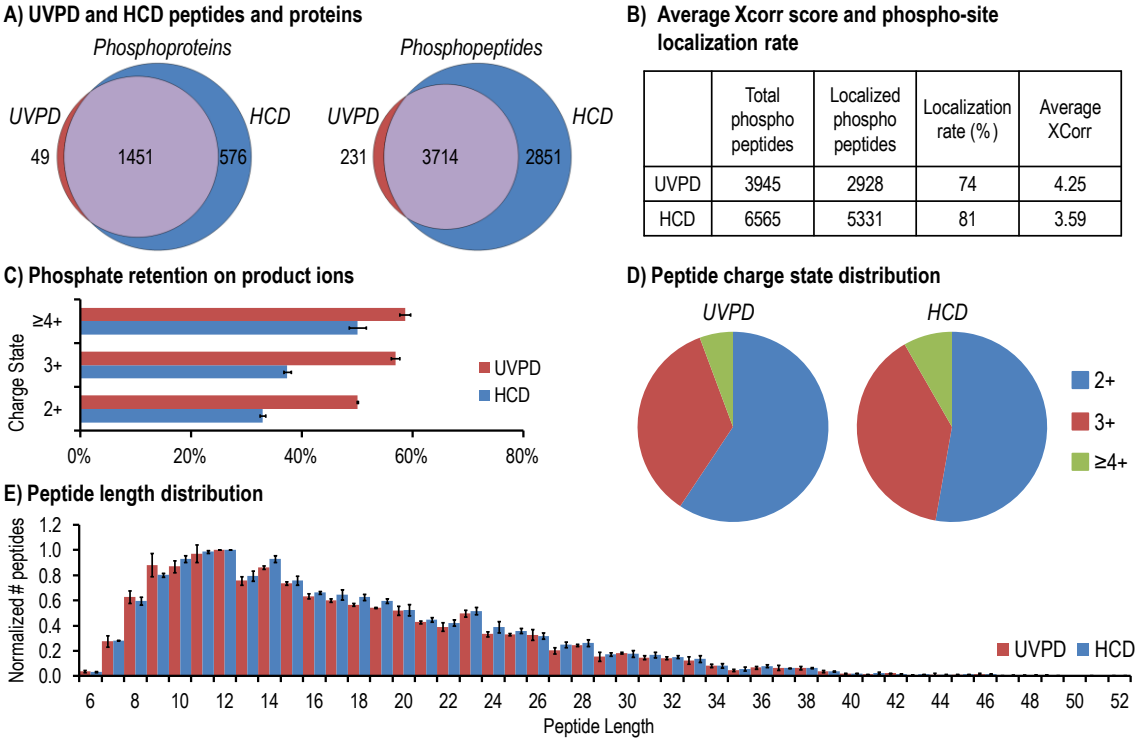


Figure 4.6 Based peak MS1 chromatograms for (A) positive mode and (B) negative LC-MS analysis of alpha S1, alpha S2, and beta casein phosphopeptides enriched on hydroxyapatite beads. Chromatographic peaks shaded in blue are recovered following C18 SPE and those shaded in purple are recovered using graphite SPE. (C) UVPD mass spectrum for phosphopeptide, QMEAEpSIpSpSpSEEIVPNpSVEQK (4-, *m/z* 679.61), from alpha S1 casein using two 2 mJ pulses. Five UVPD spectra were averaged over retention time 10.98-11.96 min.

4.4.2 UVPD and HCD Analysis of HeLa Phosphopeptides

To evaluate the performance of UVPD for phosphopeptide analysis on a more global scale, phosphopeptides were enriched from a HeLa cell lysate using Fe(III)-IMAC. UVPD and HCD analyses of the IMAC-enriched HeLa digests were carried out in triplicate using an Orbitrap Fusion mass spectrometer using high resolution/high accuracy analysis for both precursor ions and fragment ions. This dual high accuracy mode improves the accuracy of peptide identifications, especially for UVPD which uses a greater search space to account for all possible product ion series.



	Total phospho peptides	Localized phospho peptides	Localization rate (%)	Average XCorr
UVPD	3945	2928	74	4.25
HCD	6565	5331	81	3.59

Figure 4.7 Comparison of UVPD and HCD for HeLa phosphoproteome analysis: (A) overlap between UVPD- identified and HCD-identified phosphopeptides and phosphoproteins, (B) phosphorylation site localization performance, (C) phosphate retention on product ions, (D) phosphopeptide charge state distribution, and (E) phosphopeptide length distribution. All results are based on three replicate UVPD and HCD LC-MS runs.

Greater depth of coverage was achieved using HCD for which 6,565 phosphopeptides and 2,027 phosphoproteins where identified compared to 3,945 phosphopeptides and 1,500 phosphoproteins for UVPD (**Figure 4.7A-B**), consistent with recent results reported in another study.[44] Not all UVPD identifications were encompassed by the larger HCD dataset as shown in the Venn diagrams in **Figure 4.7A;** 231 phosphopeptides, corresponding to 49 phosphoproteins, were uniquely identified using UVPD.

When the overlap between replicate runs acquired using the same activation mode was evaluated, a significant amount of variation was observed as well. On average 63% of the peptides from each individual run were also identified in the other two replicate runs (**Figure 4.8A**). The overlap in protein identifications was greater with 77% of proteins from each run detected in all three replicates (**Figure 4.8B**).

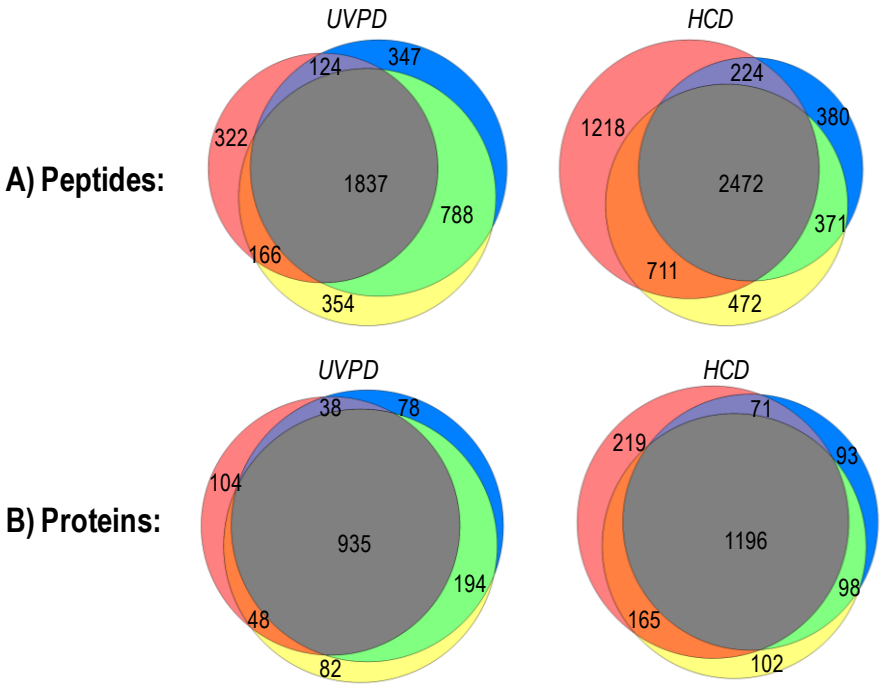


Figure 4.8 Overlap in (A) peptide and (B) protein identifications from IMAC-enriched HeLa analyzed in triplicate using UVPD and HCD.

The variations in peptide and protein identifications between replicate runs are a reflection of the sample complexity, which results in selection of different precursor ions for MS/MS analysis in each LC-MS run. Comparing UVPD and HCD based on the collective results from triplicate analysis improves the likelihood that differences in observed peptide and protein identifications are specifically related to the activation method and not due to random sampling or under-sampling in the MS/MS workflow.

In addition to the number of peptide and protein identifications, other properties of the phosphopeptides identified by UVPD and HCD were explored and compared. Consistent with the results for alpha and beta casein, the average Xcorr score of UVPD-identified peptides (4.3) was greater than the average Xcorr score of HCD-identified peptides (3.6) (**Figure 4.7B**). Despite higher Xcorr scores that are indicative of more extensive fragmentation, phosphate localization using phosphoRS was slightly less successful for UVPD (localization rate of 74%) compared to HCD (81%) based on the lower number of peptides with a phosphoRS isoform probability greater than 75% (**Figure 4.7B**). This result was unexpected given the trend of greater phosphate retention on UVPD-generated product ions that was observed for phosphopeptides from alpha and beta casein. Thus, the percent phosphate retention on UVPD versus HCD product ions was also evaluated for the HeLa dataset. For all peptides in the same charge state and considering only the product ions that contained the site of phosphorylation and therefore could undergo neutral loss, the ion abundances for phosphate-retaining ions and the corresponding phosphate loss ions were summed. Using these values, the percent phosphate retention was calculated by the following equation:

$$\%\ Phosphate\ retention = \frac{\Sigma i(phosphate\ retaining\ ions)}{\Sigma i(phosphate\ retaining\ ions) + \Sigma i(phosphate\ loss\ ions)} \times 100$$

The results for peptides in the 2+, 3+, and ≥4+ charge states are plotted in **Figure 4.7C**. For all charge states, UVPD product ions exhibited greater phosphate retention compared to HCD. The improvement using UVPD compared to HCD is greatest for peptides in lower charge states for which the lower overall proton mobility leads to increased phosphate rearrangement and loss upon collisional activation.[53,54] For doubly charged peptides an 18% increase in phosphate retention was achieved using UVPD relative to HCD, and for a 22% increase was achieved for triply charged peptides. These results suggest that UVPD should perform as well if not better than HCD for phosphorylation site assignment. Consequently, lower UVPD localization rates are likely the result of using non-optimal localization parameters associated with the phophoRS algorithm. With the current photodissociation set-up on the Orbitrap Fusion, UVPD spectra are designated as CID events at zero normalized collision energy. Thus UVPD data is processed by phosphoRS localization software analogous to CID data, and only singly and doubly charged *b* and *y* ions are considered for localization.[55] While *b* and *y* ions do constitute the majority of the UVPD product ion current, other ion types also contribute as shown in the bar graphs in **Figure 4.9A-B** which display the distribution of *a*, *b*, *c*, *x*, *y*, and *z* ions based on the total number of ions (**4.9A**) as well as the abundances of each ion type (**4.9B**). Further, the UVPD product ion distributions appears to be charge state dependent with the ratio of *y* ions to other product ion types decreasing as the peptide charge increases, indicating a shift toward more uniform product ion distribution for peptides in higher charge states. When these and other characteristic UVPD fragmentation trends are incorporated into phosphorylation localization software, the accuracy of phosphorylation site assignment will likely improve and better reflect the improvements in phosphate retention and extent of fragmentation that is observed for UVPD.

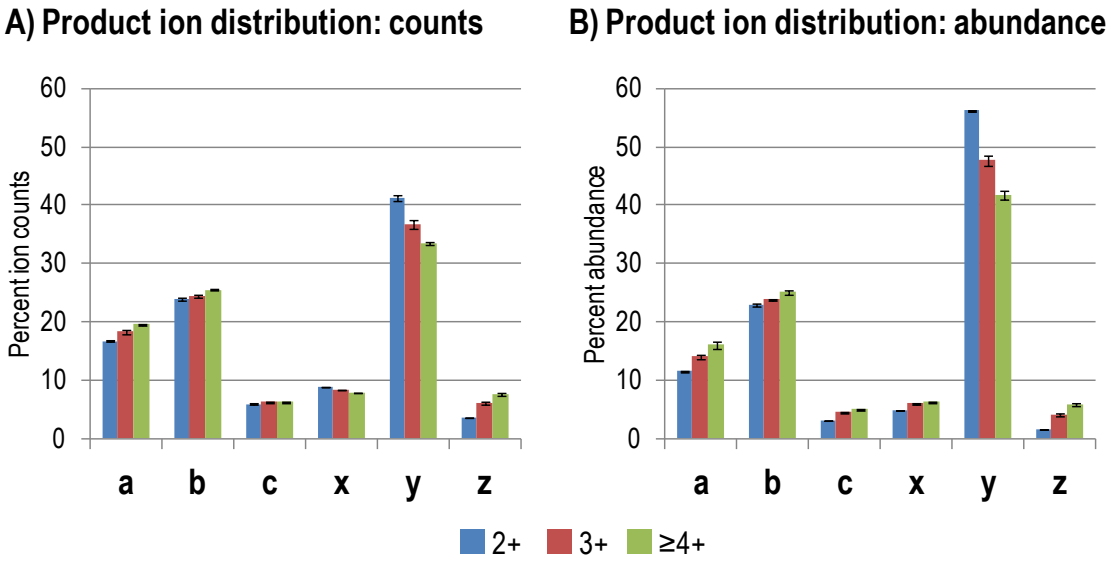


Figure 4.9 Distribution of *a*, *b*, *c*, *x*, *y*, and *z* product ions observed for phosphopeptides from IMAC enriched HeLa displayed as percentages of (A) total ion counts and (B) ion abundances.

In addition to scoring and localization results, several more general characteristics of HeLa phosphopeptides were compared to further distinguish the performance metrics of UVPD and HCD. In **Figure 4.7D**, the distributions of 2+, 3+, and more highly charged (>3+) peptides identified by UVPD and HCD are displayed. Singly charged phosphopeptides are not included in this comparison because only a handful of singly protonated phosphopeptides were identified, thus prohibiting adequate statistical comparisons. While the charge state distributions for UVPD and HCD were similar, a larger percentage of phosphopeptides specifically identified by UVPD were doubly charged compared to HCD, the latter, which identified a larger percentage of more highly charged phosphopeptides. Although subtle, this trend reinforces previous results demonstrating successful UVPD even for peptides with low proton mobility.[33,34] Peptide length distributions were also compared between those phosphopeptides identified by

UVPD versus HCD (**Figure 4.7E**), and UVPD appeared to outperform HCD for the detection of short peptides containing fewer than 12 amino acids. Because shorter peptides are likely to carry fewer charges, the UVPD length distribution correlates well with the UVPD charge state distribution. Finally, differences in the overall phosphorylation state as well as differences in the individual frequencies of phospho-serine, phospho-threonine, and phospho-tyrosine were explored for UVPD and HCD phosphopeptides **(Table 4.1)**. Singly phosphorylated peptides were most commonly observed using both UVPD and HCD for which only 5% and 10% of peptides were detected with more than one phosphorylation, respectively. The distributions of phosphorylation sites found on serine, threonine, and tyrosine residues was almost identical between the two activation methods, in addition to closely agreeing with previously reported distributions for human cell lines, lending additional validity to both datasets.[56]

Phosphorylation distribution in IMAC enriched HeLa

	% Single phospho	% Double phospho	% Multi phospho	% pSer	% pThr	% pTyr
UVPD	95%	5%	<1%	84%	16%	1%
HCD	89%	10%	<1%	85%	15%	1%

Table 4.1 Distribution of phosphorylation observed for IMAC enriched HeLa analyzed by UVPD and HCD. The percentage of singly, doubly, and multiply (3 or greater phosphorylations) phosphorylated peptides are displayed along with the frequency of occurrence on serine, threonine, and tyrosine.

4.4.3 Alternating Activation of HeLa Phosphopeptides

To this point, no significant differences related to UV photoactivation could account for the disparity in the number of peptide and protein identifications compared to HCD. A series of experiments in which HCD and UVPD were implemented in back-to-back alternating scans were carried out to further investigate the basis for the lower number of UVPD identifications.

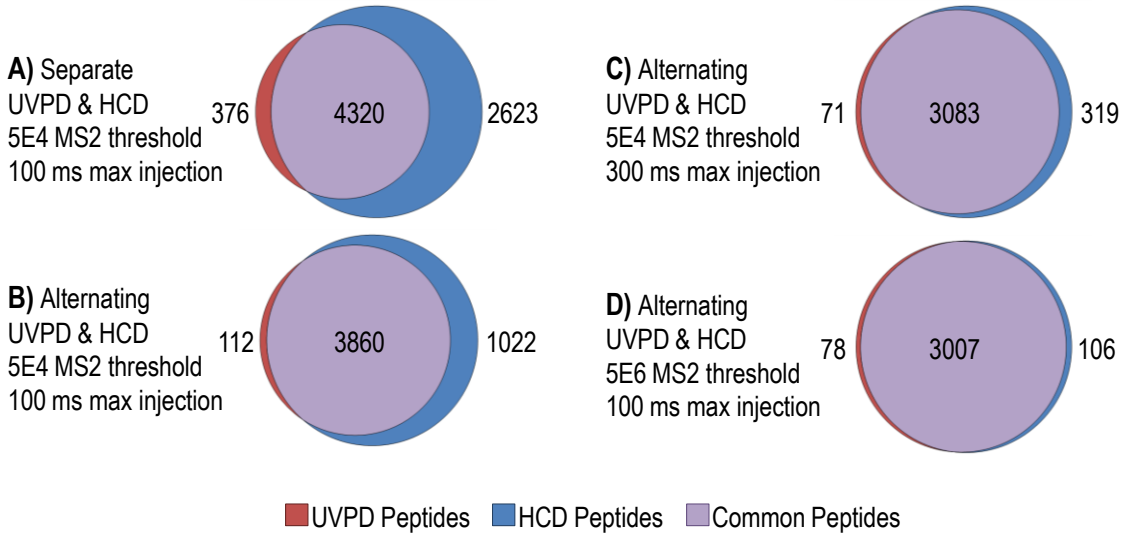


Figure 4.10 Overlap in peptide identifications from IMAC-enriched HeLa analyzed using UVPD and HCD applied in (A) separate LCMS runs or (B-D) alternated for each precursor selected within the same LC-MS run. For alternating UVPD/HCD runs, 100 ms and 300 ms maximum ion injection times were applied with a 5E4 MS2 threshold in B and C, respectively. The threshold for MS2 selection was increased to 5E6 with a 100 ms maximum injection time in D.

First separate UVPD-only and HCD-only analyses were performed to provide a baseline for comparison (**Figure 4.10A**). The number of scans averaged was reduced from two to one for this experiment to minimize the time penalty for UVPD which

requires 2 ms per laser pulse used for activation in contrast to the faster sub-millisecond HCD activation period. In subsequent LC-MS runs, UVPD and HCD were carried out within the same analysis as consecutive, alternating activation events, ensuring identical precursor selection and number of MS/MS acquisitions for each activation method. Based on the data sets acquired using the alternating MS/MS scan mode, the difference in the number of UVPD and HCD peptide identifications dramatically decreased; however, there were still 1022 unique HCD peptides (**Figure 4.10B**) indicating an additional handicap associated with UVPD.

Sensitivity is also a concern for UVPD because UVPD results in a greater array of fragment ions, thus sub-dividing ion current into more channels. The overall lower ion load for a phosphoproteomics sample relative to a complete proteome exacerbates this problem.[44] To evaluate differences in MS/MS sensitivity between UVPD and HCD, two additional alternating LC-MS/MS experiments were conducted in which low abundance precursor ions were either excluded from activation by raising the threshold for MS2 selection, or their abundances were boosted by increasing the maximum ion injection time. When the maximum ion injection time was increased from 100 ms to 300 ms (**Figure 4.10C**), the overlap in peptide identifications substantially improved while the number of unique peptides was reduced to 71 for UVPD and 319 for HCD. The best overlap between activation techniques was achieved when MS/MS activation was restricted to precursor ions with abundances greater than or equal to 5E6 (**Figure 4.10D)**, suggesting that UVPD can approach the same depth of protein identifications when existing sensitivity and speed limitations are further improved.

4.4.4 HCC70 Lysate Analysis

The HCC70 cell line is a primary ductal carcinoma that is classified as triple negative because the cells lack the genes for estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2.[57–59] Because these receptors are the primary targets of current chemotherapies, triple negative breast cancers are more difficult to treat and have a poorer prognosis, thus representing an important area of research.[60] Two HCC70 lysates, one maintained in full serum and the other serum-starved for 6 hrs prior to lysis, were analyzed by UVPD and HCD. Consistent with HeLa results, HCD identified more peptides and proteins than UVPD, but similar trends between serum-fed and serum-starved cells were observed using both activation methods (**Figure 4.11**). For both HCD and UVPD analyses, a greater number of phosphopeptides and phosphoproteins were observed in the serum-starved sample (**Figure 4.11A-B**) for which phosphorylation is expected to be more homogenous based on cell synchronization in G0 phase. Collectively, for both serum-fed and serum-starved cells using UVPD and HCD, 9324 total phosphorylation sites were identified from 13417 phosphopeptides which accounted for 3134 phosphoproteins. Compared to HCD alone, integration of the UVPD and HCD data sets resulted in 207 additional phosphopeptides and 24 additional phosphoproteins (**Figure 4.11C**), thus showing the gains in using UVPD and HCD as complementary methods to extend phosphoprotein analysis.

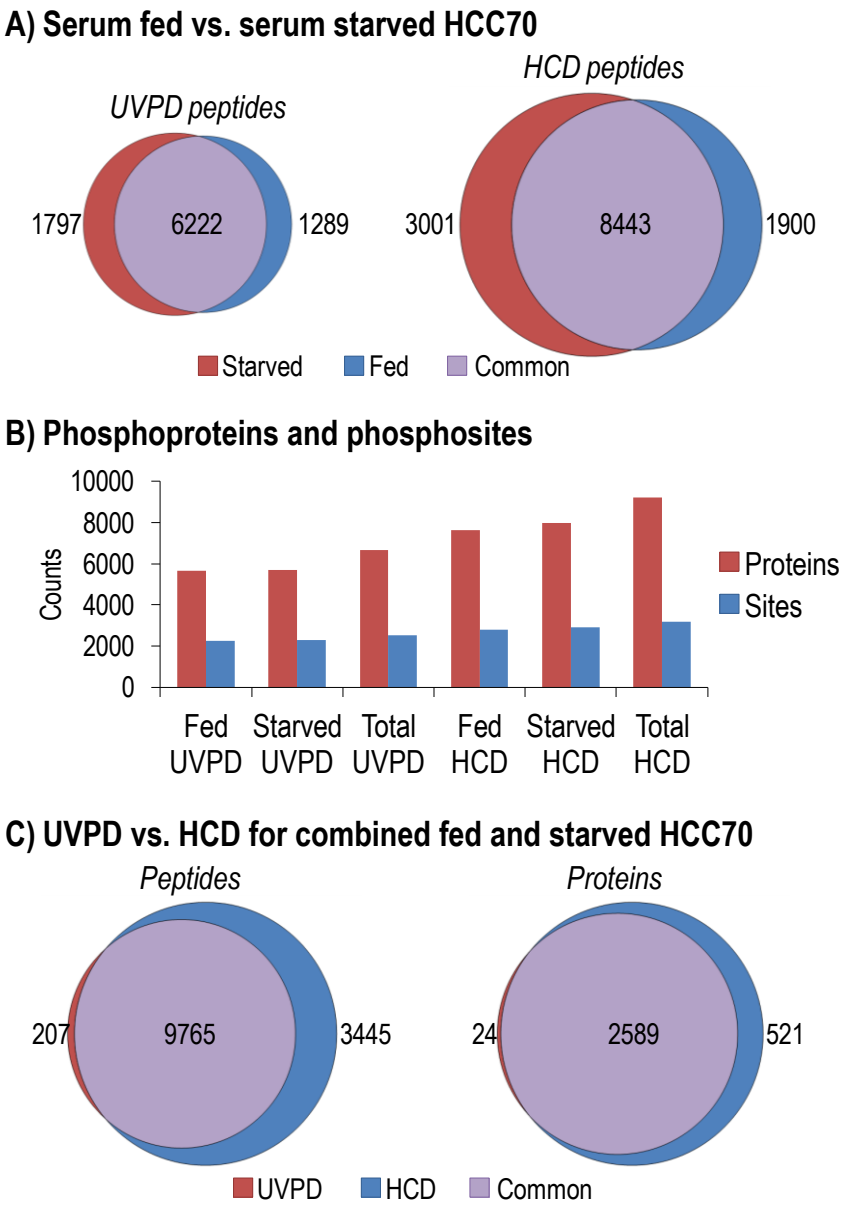


Figure 4.11 HCC70 phosphoproteomics results from UVPD and HCD analysis. (A) Overlap in peptide for cells maintained in full serum (fed) compared to cells that were serum-starved prior to lysis. (B) Number of phosphoproteins and unique phosphosites identified in serum-fed and serum-starved samples using UVPD and HCD. (C) Overlap in total UVPD and HCD peptides and proteins from both serum-fed and serum-starved cells.

4.5 CONCLUSION

Photodissociation at 193 nm offers several promising benefits compared to other activation methods for the analysis of phosphopeptides. The functionality of UVPD for sequencing both peptide cations and anions enabled phosphopeptide characterization across a wider range of phosphorylation states spanning from monophosphorylated peptides up to pentaphosphorylated peptides. The most highly phosphorylated peptides required negative mode analysis for detection, while peptides in lower phosphorylation states were effectively detected in the positive ion mode. Compared to HCD, UVPD facilitated more confident phosphosite determination based on both the greater array of product ions that were generated and the improved phosphate retention on those product ions. For the analysis of phosphopeptides from a HeLa cell lysate, no significant differences in the phosphopeptide populations were observed between UVPD and HCD data sets, but the overall number of identifications was lower for UVPD based on lower speed and sensitivity compared to HCD. While the speed of UVPD is ultimately limited by the repetition rate of the laser, improvements in sensitivity are anticipated with further development and optimization of UVPD-MS instrumentation. In its current state, UVPD remains a complementary technique to HCD for phosphopeptide characterization and can be used strategically to increase the overall depth of coverage for phosphoproteomics experiments.

4.6 REFERENCES

(1) Harsha, H. C.; Pandey, A. *Themat. Issue Oncoproteomics* **2010**, *4* (6), 482–495.
(2) Fang, B.; Haura, E. B.; Smalley, K. S.; Eschrich, S. A.; Koomen, J. M. *Target. Cancer Ther.* **2010**, *80* (5), 739–747.
(3) Riley, N. M.; Coon, J. J. *Anal. Chem.* **2015**.
(4) Zhou, H.; Di Palma, S.; Preisinger, C.; Peng, M.; Polat, A. N.; Heck, A. J. R.; Mohammed, S. *J. Proteome Res.* **2012**, *12* (1), 260–271.
(5) de Graaf, E. L.; Giansanti, P.; Altelaar, A. F. M.; Heck, A. J. R. *Mol. Cell. Proteomics* **2014**, *13* (9), 2426–2434.
(6) Kelstrup, C. D.; Jersie-Christensen, R. R.; Batth, T. S.; Arrey, T. N.; Kuehn, A.; Kellmann, M.; Olsen, J. V. *J. Proteome Res.* **2014**, *13* (12), 6187–6195.
(7) Batth, T. S.; Francavilla, C.; Olsen, J. V. *J. Proteome Res.* **2014**, *13* (12), 6176–6186.
(8) Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (1), 30–51.
(9) Thingholm, T. E.; Jensen, O. N.; Larsen, M. R. *PROTEOMICS* **2009**, *9* (6), 1451–1468.
(10) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. *J. Mass Spectrom.* **2009**, *44* (6), 861–878.
(11) Nagaraj, N.; D'Souza, R. C. J.; Cox, J.; Olsen, J. V.; Mann, M. *J. Proteome Res.* **2010**, *9* (12), 6786–6794.
(12) Wiesner, J.; Premsler, T.; Sickmann, A. *PROTEOMICS* **2008**, *8* (21), 4466–4483.
(13) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. *Rapid Commun. Mass Spectrom.* **2000**, *14* (19), 1793–1800.
(14) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proc. Natl. Acad. Sci.* **2007**, *104* (7), 2199–2204.
(15) Swaney, D. L.; Wenger, C. D.; Thomson, J. A.; Coon, J. J. *Proc. Natl. Acad. Sci.* **2009**, *106* (4), 995–1000.
(16) Sweet, S. M. M.; Bailey, C. M.; Cunningham, D. L.; Heath, J. K.; Cooper, H. J. *Mol. Cell. Proteomics* **2009**, *8* (5), 904–912.
(17) Brodbelt, J. S. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.
(18) Flora, J. W.; Muddiman, D. C. *J. Am. Chem. Soc.* **2002**, *124* (23), 6546–6547.
(19) Flora, J. W.; Muddiman, D. C. *Anal. Chem.* **2001**, *73* (14), 3305–3311.
(20) Crowe, M.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (11), 1581–1592.
(21) Crowe, M. C.; Brodbelt, J. S. *Anal. Chem.* **2005**, *77* (17), 5726–5734.
(22) Kalcic, C. L.; Gunaratne, T. C.; Jones, A. D.; Dantus, M.; Reid, G. E. *J. Am. Chem. Soc.* **2009**, *131* (3), 940–942.
(23) Smith, S. A.; Kalcic, C. L.; Safran, K. A.; Stemmer, P. M.; Dantus, M.; Reid, G. E. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 2031–2040.
(24) Smith, S. A.; Kalcic, C. L.; Cui, L.; Reid, G. E. *Rapid Commun. Mass Spectrom.* **2013**, *27* (24), 2807–2817.

(25) Lemoine, J.; Tabarin, T.; Antoine, R.; Broyer, M.; Dugourd, P. *Rapid Commun. Mass Spectrom.* **2006**, *20* (3), 507–511.
(26) Park, S.; Ahn, W.-K.; Lee, S.; Han, S. Y.; Rhee, B. K.; Oh, H. B. *Rapid Commun. Mass Spectrom.* **2009**, *23* (23), 3609–3620.
(27) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.
(28) Weidong Cui; Thompson, M. S.; Reilly, J. P. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1384–1398.
(29) Zhang, L.; Reilly, J. P. *J. Proteome Res.* **2010**, *9* (6), 3025–3034.
(30) Madsen, J. A.; Xu, H.; Robinson, M. R.; Horton, A. P.; Shaw, J. B.; Giles, D. K.; Kaoud, T. S.; Dalby, K. N.; Trent, M. S.; Brodbelt, J. S. *Mol. Cell. Proteomics* **2013**, *12* (9), 2604–2614.
(31) Greer, S. M.; Cannon, J. R.; Brodbelt, J. S. *Anal. Chem.* **2014**, *86* (24), 12285–12290.
(32) Greer, S. M.; Parker, W. R.; Brodbelt, J. S. *J. Proteome Res.* **2015**, *14* (6), 2626–2632.
(33) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. *Anal. Chem.* **2012**, *84* (5), 2433–2439.
(34) Holden, D. D.; Pruet, J. M.; Brodbelt, J. S. *Int. J. Mass Spectrom.* **2015**, *390*, 81–90.
(35) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. *PROTEOMICS* **2011**, *11* (7), 1329–1334.
(36) Shaw, J.; Madsen, J.; Xu, H.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (10), 1707–1715.
(37) Luo, Y.; Yogesha, S. D.; Cannon, J. R.; Yan, W.; Ellington, A. D.; Brodbelt, J. S.; Zhang, Y. *ACS Chem. Biol.* **2013**, *8* (9), 2042–2052.
(38) Robinson, M.; Moore, K.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (8), 1461–1471.
(39) Madsen, J. A.; Ko, B. J.; Robotham, S. A.; Xu, H.; Horton, A. P.; Iwashkiw, J. A.; Shaw, J. B.; Feldman, M. F.; Brodbelt, J. S. *Anal Chem* **2013**, *85* (19), 9253–9261.
(40) Han, S.-W.; Lee, S.-W.; Bahar, O.; Schwessinger, B.; Robinson, M. R.; Shaw, J. B.; Madsen, J. A.; Brodbelt, J. S.; Ronald, P. C. *Nat Commun* **2012**, *3*, 1153.
(41) Shin, Y.; Moon, J.; Kim, M. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 53–59.
(42) Kim, T.-Y.; Reilly, J. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (12), 2334–2341.
(43) Madsen, J. A.; Cheng, R. R.; Kaoud, T. S.; Dalby, K. N.; Makarov, D. E.; Brodbelt, J. S. *Chem. – Eur. J.* **2012**, *18* (17), 5374–5383.
(44) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. *Anal. Chem.* **2016**, *88* (4), 2303–2310.
(45) Jackson, S. S.; Coughlin, E. E.; Coon, J. J.; Miyamoto, S. *Protein Expr. Purif.* **2013**, *92* (1), 48–53.
(46) Richards, A. L.; Hebert, A. S.; Ulbrich, A.; Bailey, D. J.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. *Nat Protoc.* **2015**, *10* (5), 701–714.

(47) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (1), 1044–1051.
(48) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Anal. Chem.* **2008**, *80* (13), 4807–4819.
(49) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
(50) McAlister, G. C.; Russell, J. D.; Rumachik, N. G.; Hebert, A. S.; Syka, J. E. P.; Geer, L. Y.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J. *Anal. Chem.* **2012**, *84* (6), 2875–2882.
(51) Mamone, G.; Picariello, G.; Ferranti, P.; Addeo, F. *PROTEOMICS* **2010**, *10* (3), 380–393.
(52) Fonslow, B. R.; Niessen, S. M.; Singh, M.; Wong, C. C. L.; Xu, T.; Carvalho, P. C.; Choi, J.; Park, S. K.; Yates, J. R. *J. Proteome Res.* **2012**, *11* (5), 2697–2709.
(53) Cui, L.; Reid, G. E. *PROTEOMICS* **2013**, *13* (6), 964–973.
(54) Lanucara, F.; Chi Hoo Lee, D.; Eyers, C. E. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (2), 214–225.
(55) Taus, T.; Köcher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. *J. Proteome Res.* **2011**, *10* (12), 5354–5362.
(56) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. *Cell* **2006**, *127* (3), 635–648.
(57) Gazdar, A. F.; Kurvari, V.; Virmani, A.; Gollahon, L.; Sakaguchi, M.; Westerfield, M.; Kodagoda, D.; Stasny, V.; Cunningham, H. T.; Wistuba, I. I.; Tomlinson, G.; Tonk, V.; Ashfaq, R.; Leitch, A. M.; Minna, J. D.; Shay, J. W. *Int. J. Cancer* **1998**, *78* (6), 766–774.
(58) Kim, M. S.; Kim, T.; Kong, S.-Y.; Kwon, S.; Bae, C. Y.; Choi, J.; Kim, C. H.; Lee, E. S.; Park, J.-K. *PLoS ONE* **2010**, *5* (5), e10441.
(59) Kim, M. S.; Kwon, S.; Kim, T.; Lee, E. S.; Park, J.-K. *Biomaterials* **2011**, *32* (5), 1396–1403.
(60) J, C. K.; V, G. S.; Stanley, L. *Breast Dis.* **2010**, No. 1,2, 35–48.

Chapter 5

Phosphorylation Mapping the C-terminal Domain of RNA Polymerase II using Ultraviolet Photodissociation Mass Spectrometry

5.1 Overview

Phosphorylation in the C-terminal domain (CTD) of RNA polymerase II plays a critical role in regulating transcription. A novel method for phosphorylation mapping in wild type CTD from yeast (*Saccharomyces cerevisiae*) and fruit fly (*Drosophila melanogaster)* is presented which combines alternative proteolysis using proteinase K or chymotrypsin and ultraviolet photodissociation tandem mass spectrometry to attain residue resolved phosphorylation information. Two kinases, Erk2 and TFIIH, were used to phosphorylated the CTD. For yeast CTD, single phosphorylations on Ser2 or Ser5 in the consensus heptad, YSPTSPS, were observed across all experimental conditions. Phosphorylation also occurred preferentially at Ser2 and Ser5 in the fruit fly CTD, and the phosphorylation marks in divergent heptads revealed the significance of Tyr and Pro residues for phosphorylation to occur.

5.2 Introduction

The "CTD Code" refers to the collection of post-translational modifications that are reversibly added to the C-terminal domain (CTD) of RNA polymerase II (RNAP II), a eukaryotic protein complex involved in transcription, as well as their associated protein factors.[1–5] A species-specific number of repeats of the consensus amino acid heptad, Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7, make up the CTD.[6] Phosphorylation may occur at Tyr1, Ser2, Thr4, Ser5, and Ser7, leading to many potential phosphorylation patterns

which act to coordinate the recruitment of protein factors that influence transcription at different stages and promote RNA processing, gene regulation, and accurate initiation and termination of transcription.[7] Historically, phosphorylation in the CTD has been studied using specific monoclonal antibodies against modifications of the consensus heptad, but these methods suffer from several inherent limitations. Because antibodies are raised against consensus CTD, heptads that diverge from the consensus sequence will not be recognized which precludes confident phosphorylation characterization in these regions. The similarity of phosphate accepting motifs (i.e. YS_2P vs. TS_5P vs. PS_7Y) also presents the opportunity for potential antibody cross-reactivity across multiple phosphorylation sites. Finally, antibodies cannot identify the location of phosphorylation marks in the context of the full length CTD sequence.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) can provide more detailed information for peptide and protein sequencing; however, MS/MS methods for phosphorylation analysis are hindered by several shortcomings. Phosphate neutral loss is the dominant product of collision induced dissociation (CID) methods which promote cleavage of the most labile bonds.[8,9] To address this problem, a beam type CID configuration termed higher energy collisional dissociation (HCD) has been adopted to impart higher energy to peptide ions, thus improving backbone fragmentation.[10] While HCD offers an improved method for phosphopeptide characterization, not all peptides fragment equally well and those in low charge states frequently display limited sequence coverage and/or loss of the modifications which impedes their localization.

Electron transfer dissociation (ETD) and electron capture dissociation (ECD) are peptide sequencing methods that are considered complementary to collision based methods, although fragmentation efficiency remains largely dictated by peptide charge state.[11] Peptide cations must have at least two charges to avoid neutralization upon

electron capture or transfer, and three or more charges are needed for optimal dissociation. For phosphopeptides that meet these criteria, unparalleled modification retention is observed on both precursor and product ions leading to highly confident phosphorylation site assignment.[12–16] For this reason, ETD and ECD remain attractive alternatives to CID and HCD for phosphopeptide analysis despite any inherent difficulties for achieving sufficient positive charge density on peptides that carry acidic modifications.

Ultraviolet photodissociation (UVPD) using 193 nm photons is an alternative to collision and electron based activation methods that offers several advantages for phosphorylation mapping in the CTD. Charge state bias is largely overcome using UVPD, and high sequence coverage has been demonstrated even for singly charged precursor ions.[17–21] The high energy deposition (6.4 eV) that is achieved upon absorption of a 193 nm photon permits access to fragmentation pathways that are not available using traditional methods leading to the formation of *a*, *b*, *c*, *x*, *y*, and *z* ions which account for cleavage of each bond in the peptide backbone.[22] The greater number of product ions obtained using UVPD increases the confidence of peptide sequencing results while also improving the ability to pinpoint the sites of potential modifications. Another merit of UVPD is the ability to generate diagnostic product ions from peptide anions.[23–26] Although peptide analysis is generally undertaken in the positive ion mode based on greater sensitivity and number of applicable MS/MS techniques, the negative mode offers unique benefits for certain types of peptides. This is especially true for characterization of labile PTMs including phosphorylation, sulfation, and O-glycosylation, all of which exhibit superior retention using negative mode UVPD.[23,27–30] Alternating between positive and negative electrospray ionization modes is easily done, even within a single LCMS run, thus further increasing the versatility of UVPD-MS.

The unique structure of the CTD adds another layer of complexity to an already challenging phosphoproteomics problem. A significant portion of the CTD is comprised of serine, threonine, and tyrosine residues while lysine and arginine, basic residues that facilitate trypsin proteolysis and subsequent protonation of peptides, represent only a minor component of the protein sequence, if present at all. To determine the correct phosphorylation pattern when there are many putative sites for modification, complete sequence coverage is critical but difficult to achieve for CTD peptides that lack the basic character required to fragment well using conventional activation methods. To circumvent the lack of tryptic cleavage sites in wild type CTD, Lys and Arg were incorporated into yeast[31] and human[32] CTDs to permit digestion and facilitate phosphorylation localization. CID analysis of the mutant constructs revealed that phosphorylation is evenly dispersed throughout the length of CTD; significantly more abundant on Ser2 and Ser5 compared to other sites; and significantly less dense than once thought, with the most heptads only accepting a single phosphate.[31,32] In the present study, alternative proteolysis using proteinase K and chymotrypsin was employed to digest wild type CTD from yeast (*Saccharomyces cerevisiae*) and fruit fly (*Drosophila melanogaster).* UVPD was used to its full advantage in both positive and negative modes to elucidate the phosphorylation patterns following treatment with two kinases, Erk2 and TFIIH.

5.3 EXPERIMENTAL

5.3.1 Materials

Sequencing grade chymotrypsin was obtained from Promega (Madison, WI) and MS grade Pierce trypsin was obtained from Thermo Fisher Scientific (Grand Island, NY). LC-MS grade solvents were obtained from EMD Millipore (Temecula, CA). Other reagents were obtained from Sigma (St. Louis, MO). Integrafrit columns (360 µm O.D. x 100 µm I.D.) and picofrit columns (360 µm O.D. x 75 µm I.D. x 30 µm emitter tip I.D.) were purchased from New Objective (Woburn, MA).

5.3.2 Sample Preparation

Yeast GST-CTD samples were prepared for bottom-up analysis using a two step proteolysis method. First, overnight digestion at 37 °C with trypsin was carried out using a 1:50 enzyme to substrate ratio to cleave within the GST portion of the protein while leaving the Lys-free/Arg-free 26mer yeast CTD intact. The resulting digest was passed through a 10 KDa molecular weight cutoff (MWCO) filter to both remove tryptic GST peptides and buffer exchange the retained 26mer into 50 mM Tris HCl containing 10 mM $CaCl_2$ (pH 8) in preparation for subsequent proteinase K digestion. Proteinase K was added in a 1:100 ratio and digestion proceeded overnight at 37 °C. Samples were diluted to 1 µM in 0.2% formic acid for LC-MS analysis.

Fruit fly GST-CTD samples were reduced for 30 minutes at 55 °C using 5 mM dithiothreitol followed by alkylation of reduced cysteines for 30 minutes at room temperature in the dark using 15 mM iodoacetamide. Samples were then diluted into 100 mM Tris HCl containing 10 mM $CaCl_2$ (pH 8) and digested overnight at room temperature with chymotrypsin using a 1:50 enzyme to substrate ratio. Digests were

quenched by the addition of 0.5% trifluoroacetic acid and desalted on C18 spin columns. Samples were resuspended to 1 µM in 0.1% formic acid for bottom-up LC-MS analysis.

For top down analysis, fruit fly CTD5 constructs were buffer exchanged into 0.1% formic acid using a 7 KDa MWCO Zeba size exclusion spin column (Thermo Fisher Scientific). Samples were concentrated to a volume of 1 mg/mL prior to analysis.

5.3.3 MS, LC, and UVPD

Bottom-up analysis of the yeast CTD was performed on a Velos Pro dual linear ion trap mass spectrometer (Thermo Fisher Scientific, San Jose, Ca) equipped with a Coherent ExciStar XS excimer laser (Santa Clara, Ca) operated at 193 nm and 500 Hz as previously described for UVPD.[18,33] Two pulses at 2 mJ were used for photodissociation. Separations were carried out on a Dionex Ultimate 3000 nano liquid chromatograph configured for preconcentration. Integrafrit trap columns were packed to 3.5 cm using 5 µm Michrom Magic C18 while picofrit analytical columns were packed to 20 cm using 3.5 µm Waters Xbridge BEH C18 (Milford, MA). Mobile phase A was water and B was acetonitrile, each containing 0.1% formic acid. Peptides were loaded onto the trap column for 5 minutes in aqueous solvent containing 2% acetonitrile and 0.1% formic acid at a flow rate of 5 µL/min. Separations occurred over a 20 minute linear gradient in which the percent B was increased from 2-15% during the first 15 minutes and further increased to 35% during the last 5 minutes. The flow rate was maintained at 0.3 µL/min during the separation. A top seven data dependent acquisition method was first used to identify the main phosphorylated species. A targeted analysis followed in which m/z 818, corresponding to the singly phosphorylated heptad peptide, was continually selected for UVPD activation (between MS^1 acquisitions that occurred after every five MS/MS

events) in order to better resolve partially co-eluting phospho-isomers. The resulting UVPD spectra were manually interpreted.

Fruit fly CTD peptides were analyzed by LC-MS in both positive and negative modes. Negative mode analysis was performed on the Velos Pro mass spectrometer and Dionex nano LC equipped with C18 columns as described above. To facilitate the formation of peptide anions, methanol was used in place of acetonitrile in mobile phase B, and 0.1% trifluoroethanol (TFE) was added to all mobile phases in place of formic acid. The loading solvent consisted of 98% water, 2% methanol, and 0.1% TFE. Following sample loading at 5 µL/min for 3 minutes, a 50 minute linear gradient from 2-90% B at a flow rate of 0.25 µL/min was used for separations. MS^1 spectra were acquired from m/z 400-2000 and the top eight most abundant ions were selected for UVPD using a single 2 mJ pulse. Dynamic exclusion was enabled with an exclusion duration of 8.00 seconds. MassMatrix database search engine was used to interpret the negative mode UVPD spectra.

An Orbitrap Fusion Tribrid mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) equipped with a Coherent ExciStar XS excimer laser operated at 193 nm was used for positive mode LC-MS analysis of the fruit fly CTD. The Fusion mass spectrometer was modified for UVPD as described earlier.[34] Nano LC conditions were analogous to those described for separations of the yeast CTD, except that peptides were loaded directly onto the C18 analytical column and separated over 60 minutes using a gradient from 2-40% B. Photoactivation was achieved using 2 pulses at 2 mJ in a 3 ms top speed data-dependent method. All data was acquired in the orbitrap where MS^1 and MS^2 spectra were collected at resolving powers of 60K and 15K (at *m/z* 200), respectively. Data analysis was performed using Proteome Discoverer 2.0.

Top down LC-MS analysis of intact CTD5 constructs was carried out on the Orbitrap Fusion Tribrid mass spectrometer. The nano LC was set up with a 3 cm preconcentration column and a 25 cm analytical column containing PLRP-S resin (5 µm, 1000 Å) and operated under acidic conditions as described for positive mode bottom-up analysis. Each construct was preconcentrated online followed by separation using a fast ramp from 2-23% B over 5 minutes followed by a shallower gradient from 23-50 %B over 25 minutes. All MS^1 data was collected at a resolving power of 240K at *m/z* 200. To improve spectral signal-to-noise prior to deconvolution, the maximum number of informative spectra were averaged together and subsequently Xtracted at S/N threshold of 3 to obtained the deconvolved mass of each construct.

5.3.4 Data Analysis

Database search was used to interpret the results from both negative and positive mode UVPD analysis of the fruit fly CTD. Regardless of the program used, all data was searched against a forward and reverse FASTA database containing only fruit fly and yeast GST-CTD sequences. Phosphorylation of serine, threonine, and tyrosine was set as a variable modification in all searches and carbamidomethyl was a fixed modification of cysteines in only the positive mode searches (reduction and alkylation was not carried out prior to negative mode analysis). MassMatrix Xtreme 3.0.10.16, which is programmed to search for *a*, *x*, *c*, *z*, and *y* type product ions, was used to interpret the negative UVPD results. Peptide mass and fragment mass tolerances were ±1.00 Da and ±0.80 Da, respectively, and the minimum pp score was 6.0 while the minimum pp_{tag} score was 3.0. All sites of phosphorylation reported by MassMatrix were manually verified due to the lack of companion PTM localization software for negative mode fragmentation results.

Positive UVPD data was analyzed in Proteome Discoverer 2.0 using Sequest HT database search and ptmRS site localization software. Prior to database search, a non-fragment filter was applied to remove precursor peaks from MS/MS spectra within a 1 Da window offset. The precursor mass tolerance was 10 ppm, and the fragment mass tolerance was 0.02 Da. All possible product ions including *a*, *b*, *c*, *x*, *y*, and *z* ions were considered for spectrum matching. PSMs were validated using a fixed value PSM validator which filters matches based on a maximum Delta Cn of 0.05. Strict and relaxed target FDR settings were 0.01 and 0.05 respectively for both PSMs and peptides. Phosphorylation site localization was achieved using ptmRS operating in PhosphoRS mode. Only sites with greater than 99% isoform confidence probability were considered localized without further manual inspection.

5.4 RESULTS AND DISCUSSION

5.4.1 Yeast CTD Analysis

Phosphorylation mapping of the C-terminal domain of RNA polymerase II was initially carried out for the highly consensus yeast CTD based on using a GST-tagged version of recombinant CTD. The unique sequence of the yeast GST-CTD required the development of a novel proteolytic workflow in order to facilitate characterization of the phosphorylation pattern of CTD via bottom-up LC-MS/MS analysis (**Figure 5.1A**). The first step involved removal and separation of the CTD from the GST tag used for protein purification. Digestion with trypsin provided an efficient means for this aim based on the lack of trypsin cleavage sites in the CTD compared to the GST tag. Passing the digest through a 10 KDa molecular weight cut off filter following overnight incubation of GST-CTD with trypsin allowed retention and efficient recovery of the large CTD peptide,

whereas the GST peptides were washed through the filter. Proteinase K, which exhibits broad cleavage specify, was subsequently used to digest the purified CTD, resulting in the formation of two peptides with sequences YSPTSPS and SPSYSPT. Each peptide constituted a complete heptad repeat and thus the sample complexity for the entire GST-CTD was effectively reduced to two peptides (including all potential phosphoforms) generated following kinase treatment.

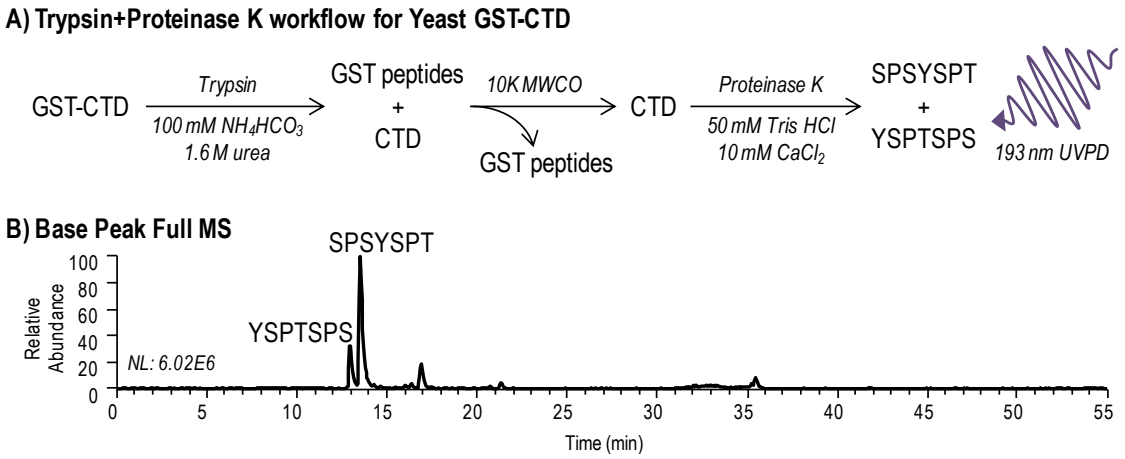


Figure 5.1 Analytical workflow for yeast GST-CTD (A) and the resulting LC-MS base peak full MS chromatogram (B)

The clean chromatographic traces that are obtained for the yeast CTD following trypsin and proteinase K digestion are shown in **Figure 5.1B**, where the heptad peptides are the dominant species in the base peak full MS chromatogram Ultraviolet photodissociation at 193 nm and CID were evaluated for sequencing the heptads generated by proteinase K digestion. Using UVPD more extensive fragmentation including the production of *a* and *x*-type sequence ions was achieved (**Figure 5.2A-B**) while the abundance of less informative water loss ions was decreased relative to CID (**Figure 5.2C-D**). Additionally, improved phosphate retention on product ions has been

demonstrated for UVPD, making it ultimately better suited for CTD phosphorylation analysis following kinase treatment.[35]

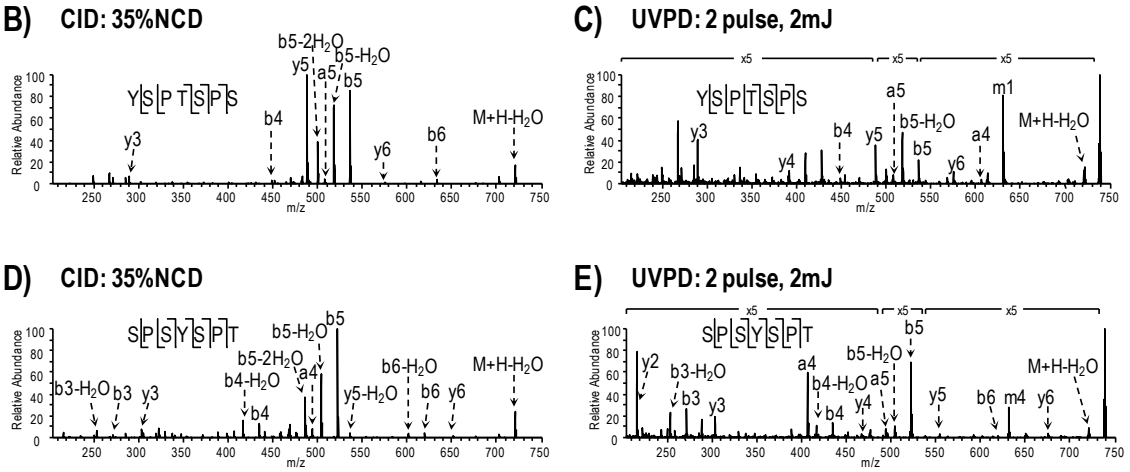


Figure 5.2 UVPD and CID analysis in positive ion mode for singly charged heptad peptides YSPTSPS (top, A and C), SPSYSPT (bottom, B and D). Tyr side chain losses generated by UVPD are denoted m1 and m4 for YSPTSPS and SPSYSPT, respectively.

Two kinases were used to phosphorylate the yeast CTD including TFIIH and Erk2. TFIIH is a multi-protein complex that is necessary for the phosphorylation of Ser 5 of the CTD during the initiation of RNAP II transcription.[36,37] Erk2 phosphorylation of poised RNAP II has been shown to occur at developmentally important genes.[38] When the yeast GST-CTD was treated with TFIIH kinase, a singly charged ion of m/z 818 corresponding to the mass of the consensus heptad plus one phosphorylation was observed in the LC-MS chromatogram (**Figure 5.3A**) in addition to the two previously detected unmodified heptad peptides of m/z 738 corresponding to YSPTSPS and SPSYSPT (**Figure 5.1B)**. The MS/MS spectra acquired during the elution of the phospho-heptad showed distinctive variations at different elution time-points, thus revealing the presence of two isomers (**Figure 5.3C-D**).

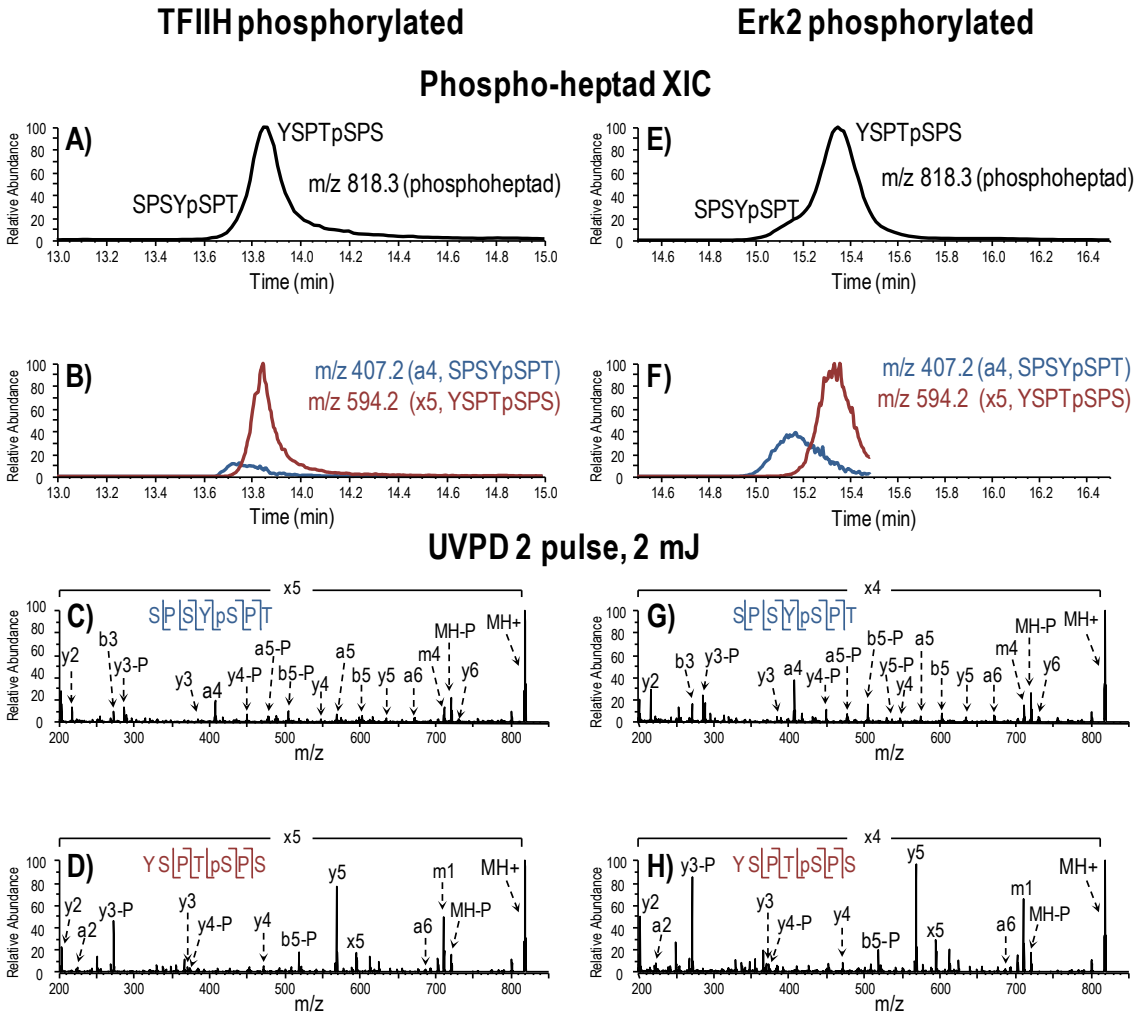


Figure 5.3 LC-MS/UVPD analysis of TFIIH and Erk2 treated yeast GST-CTD digested with trypsin and proteinase K. Two singly phosphorylated heptads, m/z 818.3, are partially resolved in the base peak MS1 chromatogram (A,E). In subsequent LC-MS analysis, m/z 818.3 was targeted for activation during the course of elution and extracted ion chromatograms (XICs) for distinguishing product ions, a4 from SPSYpSPT and x5 from YSPTpSPS, were generated to track the isomeric peptides (B,F). UVPD using two 2 mJ pulses was used to sequence the heptad peptides and localize the sites of phosphorylation (C-D,G-H). Ions that have undergone phosphate neutral loss are denoted with "-P". Tyr side chain losses generated by UVPD are denoted m1 and m4 for YSPTSPS and SPSYSPT, respectively.

Targeted LC-UVPD-MS runs allowed better characterization of the two isomers. Two abundant UVPD product ions which were unique to the early (m/z 407, a4 from SPSYpSPT) and late (m/z 594, x5 from YSPTpSPS) portions of the elution profile were identified. Extracted ion chromatograms (**Figure 5.3B**) revealed that the two ions matched to different heptad peptides with phosphorylation at different positions as defined by the consensus sequence (YSPTSPS). The UVPD mass spectra confirmed the sequences as SPSYpSPT with phosphorylation on Ser2 (**Figure 5.3C**) and YSPTpSPS with phosphorylation on Ser5 (**Figure 5.3D**). In addition to the unique *a4* (SPSYpSPT) and *x5* (YSPTpSPS) ions, other diagnostic ions that confidently differentiated each of the two phosphopeptides were *y2* and *a2* and the presence or absence of *y6*.

The LC-UVPD-MS results obtained for yeast GST-CTD treated with Erk2 kinase mirrored those observed following TFIIH reaction in terms of the detection and differentiation of the same two phosphorylated heptads (**Figure 5.3E-F**). Again, phosphorylation was confirmed at Ser2 and Ser5 upon Erk2 treatment of CTD based on the UVPD mass spectra (**Figure 5.3G-H**), but never at both sites within the same heptad thus providing further evidence that phosphorylation in the CTD is non-saturating and may only occur once per heptad.[31,32] Phosphorylation was not detected at Ser7, Thr4, or Tyr1, meaning that these species were either not present or not abundant enough to be detected without phospho-enrichment.

5.4.2 Fruit Fly CTD Analysis

The high degree of divergence from the consensus sequence within the fruit fly CTD (only 2 out of 45 heptads adhere to the consensus sequence) allows

phosphorylations to be mapped not only within individual heptads but also in the greater context of the protein sequence. Moreover, the role that specific neighboring residues play in modulating phosphorylation within each heptad can be explored. In order to fully characterize the fruit fly CTD and its phosphorylation pattern, a different digestion protocol was implemented to account for the deviations from the consensus sequence. The presence of lysine and arginine residues throughout the fruit fly CTD sequence precluded the use of trypsin for selective digestion and removal of the GST-tag as was undertaken for the yeast CTD. In addition, the low frequency of cleavage sites throughout the fruit fly CTD prohibited tryptic digestion into appropriately sized peptides for bottom up LC-MS analysis, and alternative proteases were evaluated. Proteinase K was rejected for its poorly defined cleavage specificity and tendency to cleave at multiple sites within each heptad as demonstrated for the yeast CTD in which cleavage occurred C-terminal to both Thr4 and Ser7 rather than as expected C-terminal to tyrosine.[39] Deviation from the consensus sequence makes the cleavage specificity of proteinase K even more difficult to predict within the fruit fly CTD. More importantly, because the fruit fly CTD is composed of mostly unique heptads instead of many repeating consensus units, competing cleavage channels would lower the abundance of each individual peptide while also increasing the sample complexity of an already more complicated digest.

As an alternative approach, chymotrypsin, which cleaves more consistently C-terminal to aromatic residues and to a lesser extent Met, Leu, and His, was ultimately used to achieve a more optimal degree of proteolysis within the fruit fly CTD. Missed cleavages were common and the majority of peptides spanned two or more heptads. To account for the possibility of di- and tri-phosphorylated species arising from phosphorylation in neighboring heptads of multi-heptad long peptides, the fruit fly CTD was analyzed by UVPD in both positive and negative ion modes (i.e. analysis of

protonated and deprotonated phosphopeptides, respectively). The negative mode UVPD-MS analysis ensured the detection of multiply phosphorylated peptides which ionize more readily as anions instead of cations. Additionally, several different fruit fly CTD constructs with varying degrees of truncation relative to the full length CTD protein were prepared in order to target specific regions of the CTD and thereby decrease the overall complexity in each LC-MS run. The sequences of the various CTD constructs are summarized in **Table 5.1**. CTD2 and CTD3 included the N-terminal region of the protein from heptad 1-16 and 1-25, respectively, whereas CTD4 included the C-terminal region from heptad 26-45. CTD5 covered an interior region of the protein spanning heptads 16-24. In comparison to the full length fruit fly CTD, the majority of peptide identifications from the truncated constructs provided only redundant sequence and phosphorylation site information. CTD4 was the only exception yielding a single unique phosphopeptide, SPApSPKYSPTSPL, which was identified in both positive and negative modes.

Full length *Drosophila* GST-CTD

MHHHHHHSSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVK
LTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCH
KTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGG
GDHPPKSSSLEVLFQGPGSGMSPSYSPTSPNYTASSPGGASPNYSPSSPNYSPTSPLYASPRYASTTPNFNP
QSTGYSPSSSGYSPTSPVYSPTVQFQSSPSFAGSGSNIYSPGNAYSPSSSNYSPNSPSYSPTSPSYLPSSPSY
SPTSPCYSPTSPSYSPTSPNYTPVTPSYSPTSPNYSASPQYSPASPAYSQTGVKYSPTSPTYSPPSPSYDGSP
GSPQYTPGSPQYSPASPKYSPTSPLYSPSSPQHSPSNQYSPTGSTYSATSPRYSPNMSIYSPSSTKYSPTSPT
YTPTARNYSPTSPMYSPTAPSHYSPTSPAYSPSSPTFEESED

CTD2: Heptads 1-16

MHHHHHHSSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVK
LTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCH
KTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGG
GDHPPKSSSLEVLFQGPGSGMSPSYSPTSPNYTASSPGGASPNYSPSSPNYSPTSPLYASPRYASTTPNFNP
QSTGYSPSSSGYSPTSPVYSPTVQFQSSPSFAGSGSNIYSPGNAYSPSSSNY

CTD3: Heptads 1-25

MHHHHHHSSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVK
LTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCH
KTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGG
GDHPPKSSSLEVLFQGPGSGMSPSYSPTSPNYTASSPGGASPNYSPSSPNYSPTSPLYASPRYASTTPNFNP
QSTGYSPSSSGYSPTSPVYSPTVQFQSSPSFAGSGSNIYSPGNAYSPSSSNYSPNSPSYSPTSPSYSPSSPSY
SPTSPCYSPTSPSYSPTSPNYTPVTPSYSPTSPNYSASP

CTD4: Heptads 26-45

MHHHHHHSSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVK
LTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCH
KTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGG
GDHPPKSSSLEVLFQGPGSGMQYSPASPAYSQTGVKYSPTSPTYSPPSPSYDGSPGSPQYTPGSPQYSPAS
PKYSPTSPLYSPSSPQHSPSNQYSPTGSTYSATSPRYSPNMSIYSPSSTKYSPTSPTYTPTARNYSPTSPMYS
PTAPSHYSPTSPAYSPSSPTFEESED

CTD5: Heptads 16-24

MHHHHHHSSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVK
LTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCH
KTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGG
GDHPPKSSSLEVLFQGPGSGMYSPSSSNYSPNSPSYSPTSPSYSPSSPSYSPTSPCYSPTSPSYSPTSPNYTP
VTPSYSPTSPN

Table 5.1 Full length fruit fly CTD and constructs CTD2, CTD3, CTD4, and CTD5. Each sequence includes the N-terminal GST-tag

Positive mode UVPD analysis provided the best overall sequence coverage of the fruit fly CTD. Following treatment with Erk2, 22 phosphopeptides were identified accounting for 20 unique phosphosites from 20 individual heptads (**Table 5.2A**). Phosphorylation of two additional heptads was also detected but without confident localization to one specific serine, threonine, or tyrosine residue (**Table 5.2B**).

Table 2A: CTD peptides with localized phosphosites from positive mode UVPD analysis

Sequence	Heptad # (Phosphosite in Heptad)	Construct
TASSPGGASPNYSPSSPNYSPT**pS**PLY	6 (S5)	Full
YA**pS**PRYASTTPNFNPQSTGY	7 (S5)	Full
ASPRYAST**pT**PNFNPQSTGY	8 (T5)	Full
SPT**pS**PVY**pS**PTVQF	11 (S5), 12 (S2)	Full
QS**pS**PSFAGSGSNIY	13 (S5)	Full
QSSPSFAGSGSNIY**pS**PGNAY	15 (S2)	Full
SPGNAY**pS**PSSSNY	16 (S2)	Full
SASPQYSPA**pS**PAYSQTGVKY	26 (S5)	Full
SPA**pS**PAYSQTGVKY	26 (S5)	Full
SPA**pS**PKYSPTSPL	32 (S5)	CTD4
YSPSSPQHSPSNQY**pS**PTGSTY	36 (S2)	Full
SAT**pS**PRYSPNMSIYSPSSTKY	37 (S5)	Full
SATSPRY**pS**PNMSIY	38 (S2)	Full, CTD4
pSPSSTKY	39 (S2)	Full
SPT**pS**PTY**pT**PTARNY	40 (S5), 41 (T2)	Full
SPTSPTY**pT**PTARNY	41 (T2)	Full, CTD4
SPT**pS**PMYSPTAPSHY	42 (S5)	Full, CTD4
SPT**pS**PMY**pS**PTAPSHY	42 (S5), 43 (S2)	Full
SPT**pS**PAYSPS**pS**PTFEESED	44 (S5), 45 (S5)	Full

Table 2B: CTD peptides with ambiguous phosphosites from positive mode UVPD analysis

Sequence	Heptad # (Phosphosite in Heptad)	Construct
TASSPGGASPNY**p(SPSS)**PNYSPT**pS**PLY	5 (S2/S4/S5), 6 (S5)	Full
p(SPTS)PTYSPPSPSY	28 (S2/T4/S5)	CTD4
p(SPTS)PAYSPSSPTFEESED	44 (S2/T4/S5)	Full, CTD4

Table 5.2 Phosphopeptides with localized (A) and ambiguous (B) modification sites identified from Erk2 treated *Drosophila* CTD using positive mode UVPD. Phosphorylated residues are preceded by a lowercase "p" in the peptide sequence. The sites of localized phosphorylations are shown in red while ambiguous phosphorylations are shown in purple with all potential sites for a single phosphorylation to reside grouped together by parenthesis. The heptad number, phosphorylation site within each heptad, and the *Drosophila* CTD construct that each peptide was identified in are listed in columns to the right. Full indicates the full length *Drosophila* CTD, while CTD4 includes only heptads 26-45.

Fewer overall phosphopeptides were identified using negative UVPD (**Table 5.3A-B**); however, one phosphosite from heptad 5 that could not be pinpointed by positive UVPD

was confidently localized by negative UVPD analysis. All other site assignments agreed based on the UVPD spectra acquired for the protonated and deprotonated phosphopeptides, and the absence of more highly phosphorylated peptides in the negative mode UVPD dataset further supported the hypothesis that phosphorylation only occurs at one site per heptad.

Table 3A: CTD Peptides with localized phosphosites from negative mode UVPD analysis

Sequence	Heptad # (Phosphosite in Heptad)	Construct
SPS**pS**PNYSPT**pS**PLY	5 (S5), 6 (S5)	Full
SPSSPNYSPT**pS**PLY	6 (S5)	Full
SPT**pS**PVY**pS**PTVQF	11 (S5), 12 (S2)	Full
SPTSPVY**pS**PTVQF	12(S2)	Full
QS**pS**PSFAGSGSNIY	13 (S5)	Full
QSSPSFAGSGSNIY**pS**PGNAY	15 (S2)	Full
SPA**pS**PAYSQTGVKY	26 (S5)	Full, CTD4
SPA**pS**PKYSPTSPL	32 (S5)	CTD4
SIY**pS**PSSTKY	39 (S2)	CTD4
SPTSPMY**pS**PTAPSHY	43 (S2)	Full, CTD4
SPT**pS**PAYSPS**pS**PTFEESED	44 (S5), 45 (S5)	CTD4
SPTSPAYSPS**pS**PTFEESED	45 (S5)	Full, CTD4

Table 3B: CTD Peptides with ambiguous phosphosites from negative mode UVPD analysis

Sequence	Heptad # (Phosphosite in Heptad)	Construct
AS**p(TT)**PNFNPQSTGY	8 (T4/T5)	Full
SP**p(TS)**PVYSPTVQF	11 (T4/S5)	Full
SA**p(TS)**PRY**pS**PNMSIY	37 (T4/S5), 38 (S2)	CTD4
SP**p(TS)**PAYSPSSPTFEESED	44 (T4/S5)	Full, CTD4

Table 5.3 Phosphopeptides with localized (A) and ambiguous (B) modification sites identified from Erk2 treated *Drosophila* CTD using negative mode UVPD. Phosphorylated residues are preceded by a lowercase "p" in the peptide sequence. The sites of localized phosphorylations are shown in red while ambiguous phosphorylations are shown in purple with all potential sites for a single phosphorylation to reside grouped together by parenthesis. The heptad number, phosphorylation site within each heptad, and the drosophila CTD construct that each peptide was identified in are listed in columns to the right. Full indicates the full length *Drosophila* CTD, while CTD4 includes only heptads 26-45.

Additional phosphorylation trends are revealed upon examination of all the phosphorylation sites in the context of the entire protein sequence as shown in **Figure 5.4A**. Displaying the CTD as a series of vertically aligned heptads serves to further map the position of phosphorylation within each heptad while also efficiently showing the sites in each heptad that have deviated from the consensus sequence. Heptads shaded in grey, including heptads 17-24 and 30-31, were not detected by LC-MS in the positive or negative modes. The lack of sequence coverage in these regions was also observed prior to phosphorylation of CTD with Erk2, indicating that the result is more likely a function of the CTD protein structure and not due to the addition of phosphorylation. The remaining CTD heptads were effectively characterized by analysis of peptides created upon digestion with chymotrypsin, and heptads that lack any color coding distinctions were detected without phosphorylation. Single residues highlighted in green represent localized phosphorylation sites, while sets of residues highlighted in gold within the same heptad represent ambiguous phosphorylation where all putative sites for a single modification are highlighted. Based on the pattern of phosphorylation observed in the fruit fly CTD (**Figure 5.4A**), several rules governing phosphorylation by Erk2 are proposed. First, there is a clear preference for phosphorylation on Ser5 but only when followed directly by Pro in the six position. One of the many peptides exhibiting this trend is SPTpSPVYSPTVQ for which both positive (**Figure 5.4B**) and negative (**Figure 5.4C**) UVPD mass spectra display extensive sequence coverage and confident phosphate localization.

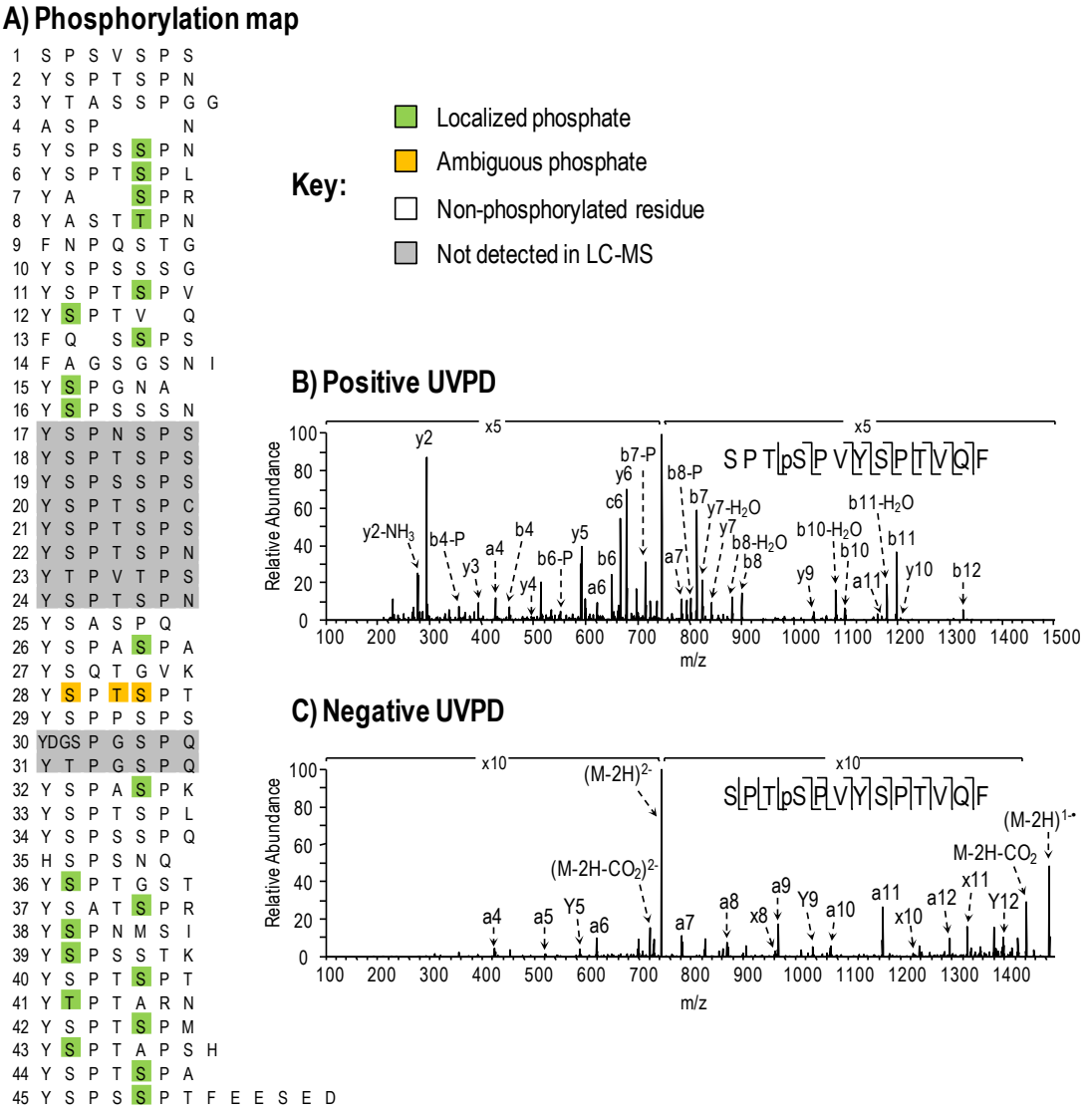


Figure 5.4 Phosphorylations identified in *Drosophila* CTD following treatment with ERK2 (A), where sites highlighted in green and gold indicate localized and non-localized sites, respectively. Regions of the protein highlighted in grey were not detected in the Erk2 treated or control CTD samples. The phosphorylation map is the composite of sites identified using positive mode and negative mode LC UVPD-MS. Representative UVPD mass spectra from positive mode (B) and negative mode (C) analysis is shown for the chymotryptic peptide SPTpSPVYSPTVQF which covers heptads 11 and 12. In both polarities, the doubly charged ions of m/z 745.3 and 743.3 for positive and negative modes, respectively, were activated using 2 pulses at 2 mJ. Ions that are detected following phosphate neutral loss are denoted by "-P".

For heptads in which Ser5 is not conserved or is not followed by Pro, Ser2 may be phosphorylated provided that the Ser-Pro motif is satisfied. Additionally, if Thr takes the place of Ser in the two or five position beside a neighboring Pro, then phosphorylation may occur on Thr as demonstrated in heptads 8 and 41. The phosphorylation pattern does not appear to be influenced by the presence or absence of Thr4 and Ser7. The four and seven positions in the heptad both exhibit significant divergence from the consensus sequence and no correlation could be made between residue identity and the phosphorylation behavior of Erk2 within the corresponding heptad. Finally, the presence of an aromatic residue in the 1 position appeared to dictate phosphorylation, and no phosphorylation sites were detected in heptads that were not initiated by Tyr or Phe.

5.4.3 Intact Mass Analysis of CTD5

To further explore the significance of Tyr for kinase recognition, two position 1 sequence variants were prepared from CTD5 and analyzed by western blot and top down mass spectrometry before and after reaction with Erk2. Results for the wild type CTD5 are shown in **Figure 5.5A**, and phosphorylation is clearly observed following kinase treatment based on the appearance of characteristic mass shifts in both the western blot and mass spectrometry data. Up to two phosphoforms were detected, which are clearly identified by consecutive mass additions of ~80 Da in the deconvoluted mass spectrum. When each Tyr residue of CTD5 was mutated to Ala (Tyr to Ala), no mass shifts were observed between the control and Erk2-treated samples (**Figure 5.5B**), indicating a lack of kinase recognition in the absence of Tyr. In the second sequence variant Tyr was mutated to the more structurally similar Phe (Tyr to Phe), and in doing so an extensive array of mass-shifted bands appeared in the western blot after the reaction with Erk2

(**Figure 5.5C**). Although the quality of the corresponding mass spectrum of the intact protein(s) was marginal due to spectra complexity and no conclusions regarding the exact phosphorylation state could be made, a definite shift towards higher mass was observed for the Erk2-treated construct compared to the control. Clearly, Phe also plays a role in directing phosphorylation within the CTD, but more experiments are required to fully understand the evolutionary significance.

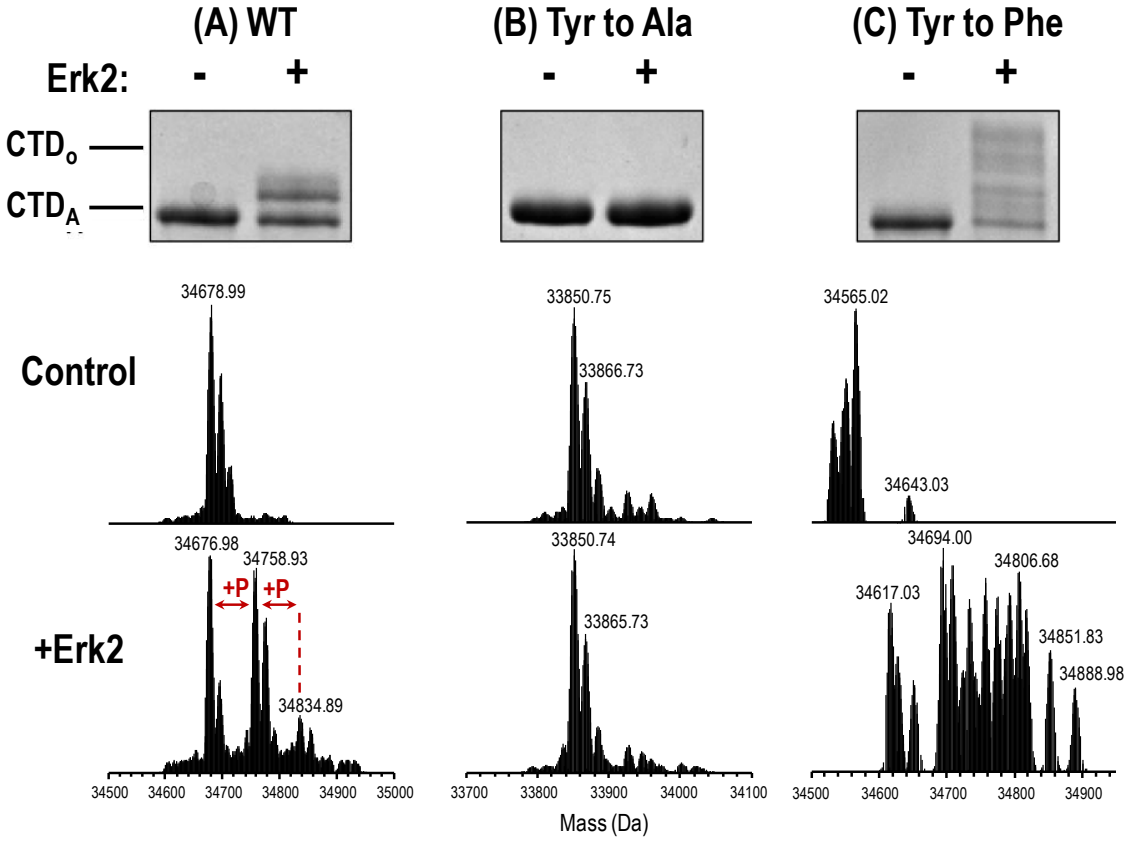


Figure 5.5 Western blot and top down mass spectrometry results for CTD5 before and after treatment with Erk2. Three different constructs were analyzed including (A) wild type (WT) where Tyr was present at position one in each heptad , (B) Tyr to Ala where Tyr was mutated to Ala, and (C) Tyr to Phe where Tyr was mutated to Phe.

5.5 CONCLUSIONS

Phosphate mapping in wild type yeast and fruit fly CTDs was successfully demonstrated using LC-MS and 193 nm UVPD. Alternative proteolysis was first established, using proteinase K for yeast CTD and chymotrypsin for fruit fly CTD, to generate suitably sized peptides for bottom-up analysis without the need for mutation within the CTD sequence. Erk2 and TFIIH were used to phosphorylate the yeast CTD, and both kinases modified the CTD at the same sites, Ser2 or Ser5, in the consensus heptad. Only one phosphorylation was observed per heptad which provides added evidence that phosphorylation in the CTD is non-saturating. In the fruit fly CTD, Erk2 preferentially phosphorylated the same Ser2 and Ser5 marks, and again no instances of multiple phosphorylations within a single heptad were observed. The high degree of divergence in the fruit fly CTD was key for determining the influence of different amino acids for kinase recognition and phosphorylation. Notably, Tyr was needed in the 1 position and Pro in the +1 position relative to Ser5 or Ser2 in order for phosphorylation to occur. Based on the detailed, residue resolved phosphorylation assignment afforded by UVPD-MS, this analytical strategy will likely be extended to future CTD studies.

5.6 REFERENCES

(1) Steinmetz, E. J.; Warren, C. L.; Kuehner, J. N.; Panbehi, B.; Ansari, A. Z.; Brow, D. A. *Mol. Cell* **2006**, *24* (5), 735–746.
(2) Koch, F.; Jourquin, F.; Ferrier, P.; Andrau, J.-C. *Trends Biochem. Sci.* **2008**, *33* (6), 265–273.
(3) Koch, F.; Fenouil, R.; Gut, M.; Cauchy, P.; Albert, T. K.; Zacarias-Cabeza, J.; Spicuglia, S.; de la Chapelle, A. L.; Heidemann, M.; Hintermair, C.; Eick, D.; Gut, I.; Ferrier, P.; Andrau, J.-C. *Nat. Struct. Mol. Biol.* **2011**, *18* (8), 956–963.
(4) Jeronimo, C.; Bataille, A. R.; Robert, F. *Chem. Rev.* **2013**, *113* (11), 8491–8522.
(5) Buratowski, S. *Nat. Struct. Biol.* **2003**, *10* (9), 679–680.
(6) Chapman, R. D.; Heidemann, M.; Hintermair, C.; Eick, D. *Trends Genet. TIG* **2008**, *24* (6), 289–296.
(7) Eick, D.; Geyer, M. *Chem. Rev.* **2013**, *113* (11), 8456–8490.
(8) Palumbo, A. M.; Smith, S. A.; Kalcic, C. L.; Dantus, M.; Stemmer, P. M.; Reid, G. E. *Mass Spectrom. Rev.* **2011**, *30* (4), 600–625.
(9) Brown, R.; Stuart, S.; Houel, S.; Ahn, N.; Old, W. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (7), 1128–1142.
(10) Nagaraj, N.; D'Souza, R. C. J.; Cox, J.; Olsen, J. V.; Mann, M. *J. Proteome Res.* **2010**, *9* (12), 6786–6794.
(11) Wiesner, J.; Premsler, T.; Sickmann, A. *PROTEOMICS* **2008**, *8* (21), 4466–4483.
(12) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. *Rapid Commun. Mass Spectrom.* **2000**, *14* (19), 1793–1800.
(13) Sweet, S. M. M.; Bailey, C. M.; Cunningham, D. L.; Heath, J. K.; Cooper, H. J. *Mol. Cell. Proteomics* **2009**, *8* (5), 904–912.
(14) Shi, S. D.-H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W. *Anal. Chem.* **2001**, *73* (1), 19–22.
(15) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proc. Natl. Acad. Sci.* **2007**, *104* (7), 2199–2204.
(16) Creese, A.; Cooper, H. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (9), 1263–1274.
(17) Thompson, M.; Cui, W.; Reilly, J. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (8), 1439–1452 – 1452.
(18) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.
(19) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. *Anal. Chem.* **2012**, *84* (5), 2433–2439.
(20) Holden, D. D.; Pruet, J. M.; Brodbelt, J. S. *Int. J. Mass Spectrom.*
(21) Greer, S. M.; Parker, W. R.; Brodbelt, J. S. *J. Proteome Res.* **2015**, *14* (6), 2626–2632.
(22) Brodbelt, J. S. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.
(23) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. *PROTEOMICS* **2011**, *11* (7), 1329–1334.

(24) Shaw, J.; Madsen, J.; Xu, H.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (10), 1707–1715.
(25) Madsen, J. A.; Xu, H.; Robinson, M. R.; Horton, A. P.; Shaw, J. B.; Giles, D. K.; Kaoud, T. S.; Dalby, K. N.; Trent, M. S.; Brodbelt, J. S. *Mol. Cell. Proteomics* **2013**, *12* (9), 2604–2614.
(26) Greer, S. M.; Cannon, J. R.; Brodbelt, J. S. *Anal. Chem.* **2014**, *86* (24), 12285–12290.
(27) Luo, Y.; Yogesha, S. D.; Cannon, J. R.; Yan, W.; Ellington, A. D.; Brodbelt, J. S.; Zhang, Y. *ACS Chem. Biol.* **2013**, *8* (9), 2042–2052.
(28) Robinson, M.; Moore, K.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (8), 1461–1471.
(29) Madsen, J. A.; Ko, B. J.; Robotham, S. A.; Xu, H.; Horton, A. P.; Iwashkiw, J. A.; Shaw, J. B.; Feldman, M. F.; Brodbelt, J. S. *Anal Chem* **2013**, *85* (19), 9253–9261.
(30) Han, S.-W.; Lee, S.-W.; Bahar, O.; Schwessinger, B.; Robinson, M. R.; Shaw, J. B.; Madsen, J. A.; Brodbelt, J. S.; Ronald, P. C. *Nat Commun* **2012**, *3*, 1153.
(31) Suh, H.; Ficarro, S. B.; Kang, U.-B.; Chun, Y.; Marto, J. A.; Buratowski, S. *Mol. Cell* **2016**, *61* (2), 297–304.
(32) Schüller, R.; Forné, I.; Straub, T.; Schreieck, A.; Texier, Y.; Shah, N.; Decker, T.-M.; Cramer, P.; Imhof, A.; Eick, D. *Mol. Cell* **2016**, *61* (2), 305–314.
(33) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Anal. Chem.* **2008**, *80* (13), 4807–4819.
(34) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. *Anal. Chem.* **2015**.
(35) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. *Anal. Chem.* **2016**.
(36) Mayfield, J. E.; Burkholder, N. T.; Zhang, Y. J. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2016**, *1864* (4), 372–387.
(37) Jeronimo, C.; Robert, F. *Nat. Struct. Mol. Biol.* **2014**, *21* (5), 449–455.
(38) Tee, W.-W.; Shen, S. S.; Oksuz, O.; Narendra, V.; Reinberg, D. *Cell 156* (4), 678–690.
(39) Pähler, A.; Banerjee, A.; Dattagupta, J. K.; Fujiwara, T.; Lindner, K.; Pal, G. P.; Suck, D.; Weber, G.; Saenger, W. *EMBO J.* **1984**, *3* (6), 1311–1314.

Chapter 6

Direct Identification of Tyrosine Sulfation by using Ultraviolet Photodissociation Mass Spectrometry[2]

6.1 OVERVIEW

Sulfation is a common post-translational modification of tyrosine residues in eukaryotes; however, detection using traditional liquid chromatography-mass spectrometry (LC-MS) methods is challenging based on poor ionization efficiency in the positive ion mode and facile neutral loss upon collisional activation. In the present study, 193 nm ultraviolet photodissociation (UVPD) is applied to sulfopeptide anions to generate diagnostic sequence ions which do not undergo appreciable neutral loss of sulfate even using higher energy photoirradiation parameters. At the same time, neutral loss of sulfate is observed from the precursor and charge reduced precursor ions, a spectral feature that is useful for differentiating tyrosine sulfation from the nominally isobaric tyrosine phosphorylation. LC-MS detection limits for UVPD analysis in the negative mode were determined to be around 100 fmol for three sulfated peptides, caerulein, cionin, and leu-enkephalin. The LC-UVPD-MS method was applied for analysis of bovine fibrinogen, and its key sulfated peptide was confidently identified.

---

[2] Robinson, M. R.; Moore, K. L.; Brodbelt, J. S. Direct Identification of Tyrosine Sulfation by using Ultraviolet Photodissociation Mass Spectrometry. *Journal of the American Society for Mass Spectrometry.* **2014**, *25*, 1461–1471.
KLM donated bovine fibrinogen and reviewed the manuscript prior to publication. JSB provided mentorship and reviewed the manuscript.

6.2 INTRODUCTION

The comprehensive identification of protein post-translational modifications (PTMs) continues to be an important goal of proteomics research in order to gain a better understanding of biological systems, especially in the context of how PTMs influence protein structure and function.[1] Despite advancements in analytical technology, particularly in mass spectrometry (MS), PTM mapping remains a challenging task based on the diverse array of PTMs, their low abundance and lability, and their unique chemical properties, thus driving the development of new techniques to aid in characterization. O-sulfation, first discovered in 1954 on bovine fibrinogen, is a primary modification of tyrosine with the potential for sulfate addition on up to an estimated 1% of all tyrosine residues of the total protein in an organism.[2–4] Modification is limited to secretory and transmembrane proteins that have traversed the trans-Golgi network where two membrane-bound tyrosylprotein sulfotransferase enzymes (TPST1 and TPST2) catalyze the transfer of sulfate from adenosine 3'-phosphate 5'-phosphosulfate (PAPS) to the tyrosine phenol.[5–11] The primary function of tyrosine sulfation is the modulation of protein-protein interactions in the extracellular region.[12–15] More specifically, sulfation has been shown to play a profound role in numerous physiological and pathological processes, including hormonal regulation, hemostasis, inflammation and viral entry into host cells.[16,17] However, other role(s) for tyrosine sulfation in protein function may exist.

Despite the biological significance of tyrosine sulfation, the sulfoproteome remains largely unexplored due to the analytical challenges associated with characterization using mass spectrometry. Several properties of sulfated peptides, including an often very acidic amino acid sequence and the labile sulfo-ester bond, present major handicaps for conventional positive mode MS analysis.[18,19] Traditionally soft ionization techniques such as electrospray ionization (ESI) and matrix-assisted laser

desorption ionization (MALDI) result in partial or complete loss of the modification in the positive mode. Sulfopeptides that remain intact during ionization and the first stage of mass analysis undergo the predominant neutral loss of sulfate upon collisional induced dissociation (CID) and any product ions observed exhibit loss of modification.[20,21] Electron-based activation (ETD and ECD) also promote sulfate loss from product ions;[22] however, modification retention has been observed for highly basic sulfopeptides likely due to formation of a salt bridge between the acidic sulfo-moiety and arginine side chains.[23] For more acidic peptides, gas-phase adduction using metal cations or guanidinium groups has been used to generate stabilizing salt bridges making sulfation site localization possible upon ECD.[24–26] An alternative strategy for site localization in the positive mode takes advantage of the lability of sulfate in a subtractive-based identification method. In this method free tyrosine residues are acetylated prior to MS analysis so that any unmodified tyrosine residues detected must necessarily originate from sulfate loss in the mass spectrometer.[27,28] While effective, these techniques rely on quantitative reaction of unmodified tyrosine and require more front-end sample processing.

Mass spectrometry analysis in the negative mode can provide a more direct approach for the detection of tyrosine sulfation based on the greater stability of sulfopeptides as gas-phase anions. The consistent detection of intact deprotonated sulfopeptides upon ESI is a significant advantage compared to the prevalent decomposition of protonated sulfopeptides during ESI; however, there remains a need for improved MS/MS characterization. The primary fragmentation pathway for CID of peptide anions is neutral loss of sulfate and while this information is useful for confirming the presence of sulfation, the lack of peptide backbone fragments is an impediment.[29–31] Alternative activations methods including metastable atom-activated dissociation (MAD),[32] which

uses a beam of high kinetic energy helium atoms for ion activation, and negative ion electron capture dissociation (niECD),[33] have shown promise for tyrosine sulfation mapping. Both techniques provide a high level of peptide sequence coverage without significant losses of the sulfate modifications. Specifically, MAD results in the formation of a diverse array of fragment ions, including *a, b, c, x, y* and *z,* and niECD favors formation of *c* and *z* fragment anions. Long activation times and/or extensive spectral averaging are required for optimal results for these two methods, limiting the compatibility of these methods with online LCMS methods.

Ultraviolet photodissociation (UVPD) at 193 nm is a fast activation method that has shown merits for peptide sequencing in both the positive and negative modes,[34–40] including analysis of peptides decorated with acidic PTMs.[38,41,42] Upon UVPD activation, peptide anions dissociate into predominantly *a* and *x* type ions with other ion series including *c, z,* and *y* observed less frequently. The unique UVPD fragmentation behavior of peptide anions has been incorporated into a database search engine (MassMatrix) to effectively streamline data interpretation and make possible the analysis of more complex proteomic samples.[39] Like electron based activation methods, UVPD is a fragmentation technique that does not promote neutral loss of post-translational modifications, one of the considerable disadvantages of CID. For example, in the UVPD analysis of deprotonated phosphorylated peptides, backbone cleavage remains the primary fragmentation pathway, thus allowing both the peptide sequence and the site of modification to be determined.[38,40] Similar promising results have recently been obtained for a sulfated peptide derived from the Ax21 protein in the gram negative bacterium *Xanthomonas oryzae* pv. *oryzae.*[42] This first successful identification of tyrosine sulfation using UVPD has prompted a more in-depth investigation of UVPD analysis of deprotonated sulfopeptides. In the present study, several figures of merit which are

relevant for sulfate mapping using UVPD are evaluated. These include: the overall peptide sequence coverage, changes in modification retention as a function of the laser settings, MS1 and MS2 sensitivity, and the compatibility with online LCMS analysis.

6.3 EXPERIMENTAL

6.3.1 Materials

LCMS grade solvents and mobile phase additives were obtained from Fisher Scientific (Fairlawn, NJ). Other reagents were obtained from Sigma (St. Louis, Mo). Peptides GlpQDsYTGWMDF-NH2 (caerulein), RDsYTGWNleDF-NH2 (Thr28,Nle31-cholecystokinin-33 sulfated), Ac-DpYVPML-NH2, RRLIEDAEpYAARG-NH2, Ac-IpYGEF-NH2 (P60c-src Substrate II, phosphorylated), and TSTEPQpYQPGENL (Pp60c-src 521-553) were purchased from American Peptide Company (Sunnyvale, Ca); GDFEEIPEEsYLQ (hirudin fragment 54-65) and NsYsYGWMDF-NH2 (cionin) were purchased from Sigma; and sYGGFL (leucine-enkephalin sulfated) was purchased from Phoenix Pharmaceuticals (Burlingame, Ca). Bovine fibrinogen was obtained from Calbiochem. The protein was reduced with 5 mM dithiothreitol (DTT) at 55 °C for 45 minute. Iodoacetamide was then added to 15 mM, and alkylation proceeded for 45 minutes in the dark. Additional DTT was added to quench the alkylation. Trypsin (Promega, Madison, WI) was added in a 1:20 enzyme: protein ratio, and digestion occurred overnight at 37 °C.

6.3.2 MS, LC, and UVPD

All experiments were conducted on a Thermo Fisher Scientific Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a

Coherent ExciStar XS excimer laser (Santa Clara, Ca) operated at 193 nm and 500 Hz as previously described.[43,44] For direct infusion at 4 µL/min, 10 µM peptide solutions were prepared in water containing 25% methanol and either 0.1% formic acid, 0.1% ammonium hydroxide, or 5 mM ammonium acetate for analysis in the positive, negative, or both ion modes, respectively. For positive mode analysis, the ESI source was operated at 3.5 kV with a sheath gas flow of 4 units and auxiliary and sweep gases both zero. Orbitrap MS1 and MS2 automatic gain control (AGC) targets were 1,000,000 and 50,000, respectively. For negative ion mode experiments, the heated ESI source (HESI) was used and the source parameters were tuned before each analysis in order to optimize the spray stability. Optimum HESI temperatures ranged from 40-60°C with source voltage 2.3-3 kV. A high sheath gas flow between 25-40 units improved desolvation, while lower amounts of auxiliary gas and sweep gas were needed (both were operated between 0-5 units). Orbitrap AGC targets were 1,000,000 for MS1 and 100,000 for MS2 in negative mode. In both polarities CID and HCD were performed using 35% normalized collision energy with activation times of 10 and 0.1 ms, respectively. To optimize the UVPD conditions, a variety of energies including 1, 2, 3, 4, and 5 mJ were used with the number of laser pulses ranging from 1 to 6. Ultimately, 3 pulses at 2 mJ were selected for LCMS experiments.

Liquid chromatography was performed on a Dionex Ultimate 3000 capillary LC operated at a flow rate of 4 µL/min. Mobile phase A was 5 mM ammonium acetate in water, and mobile phase B was 5 mM ammonium acetate in 90% methanol, 10% water. Peptide solutions were prepared in 100% mobile phase A for separation on a 3 x 150 mm Agilent Zorbax Extend-C18 column with 3.5 µm particle size (Santa Clara, CA). Separations of an equimolar mixture of caerulein, cionin, and leu-enkephalin were accomplished using a linear gradient that increased from 25% B to 60% B over 15

minutes. Doubly deprotonated caerulein (m/z 674.7162) and cionin (m/z 625.6651) and singly deprotonated leu-enkephalin (m/z 634.2178) were targeted for negative UVPD using a precursor mass list with associated 10 ppm m/z tolerance. Dynamic exclusion was disabled and MS1 (400-2000 m/z) and MS2 scans, both collected in the Orbitrap at resolution 15000, were alternated over the course of the LCMS run.

For negative UVPD analysis of tryptic bovine fibrinogen, about 5 μg (≈14.7 pmol) of protein digest was injected, and the percent B was increased linearly from 2% to 35% over 45 minutes. A top ten data dependent scan program was used in which the first scan was a negative full FTMS survey scan over m/z 400-2000 at resolution 120,000 followed by 10 UVPD events on the ten most abundant ions from scan event 1. Dynamic exclusion was enabled for 25 seconds with a single repeat count. For UVPD at 15000 resolution, the isolation width was set to 3 Da, the HCD normalized collision energy was 1% and the activation time was 6 ms in order to generate 3 laser pulses. After the initial analysis of fibrinogen, an additional segment from time 16.85-18.40 min was included to continually target a sulfated peptide of interest, GlpFPTDsYDEGQDDRPK (fibrinopeptide B), for UVPD. During this segment the m/z of interest, 935.35160, was isolated using an increased width of 6 Da and dissociated using 3 laser pulses each at 2 mJ.

6.3.3 Database Search

Results from LCMS analysis of fibrinogen were interpreted using the MassMatrix database search algorithm [45–48]. The experimental data was searched against the fibrinogen bovine FASTA and a reversed decoy database. Trypsin (no P rule) was specified for digestion while a maximum of 3 missed cleavages. The minimum peptide

length was 5 amino acids while the maximum length was 75 amino acids. The peptide mass tolerance was 20 ppm and the fragment mass tolerance was ±0.02 Da. MassMatrix uses three independent statistical scores including, pp, pp_2, and pp_{tag}, to evaluate quality of peptide-spectrum matches, and the minimum output for each score was defined as 5.0, 5.0, and 1.3 respectively. Iodoacetamide derivatization of cysteine was a fixed modification, while sulfated tyrosine and pyroglutamate from glutamate were variable modifications.

6.4 RESULTS AND DISCUSSION

6.4.1 UVPD, CID, and HCD of Sulfopeptides

Four singly sulfated peptides including caerulein (GlpQDsYTGWMDF-NH_2), cholecystokinin (RDsYTGWLDF-NH_2), hirudin (GDFEEIPEEsYLQ), and leu-enkephalin (sYGGFL), and one doubly sulfated peptide, cionin (NsYsYGWMDF-NH_2) were analyzed in both positive and negative modes and characterized using CID, HCD and UVPD. The spectral quality of MS survey scans in the positive mode was generally poor, even for cholecystokinin, the most basic of the five peptides analyzed and thus the most amenable to positive mode ionization. **Figure 6.1A-B** show full MS results for cholecystokinin electrosprayed in both acidic (0.1% formic acid containing) and neutral (5 mM ammonium acetate containing) solutions.

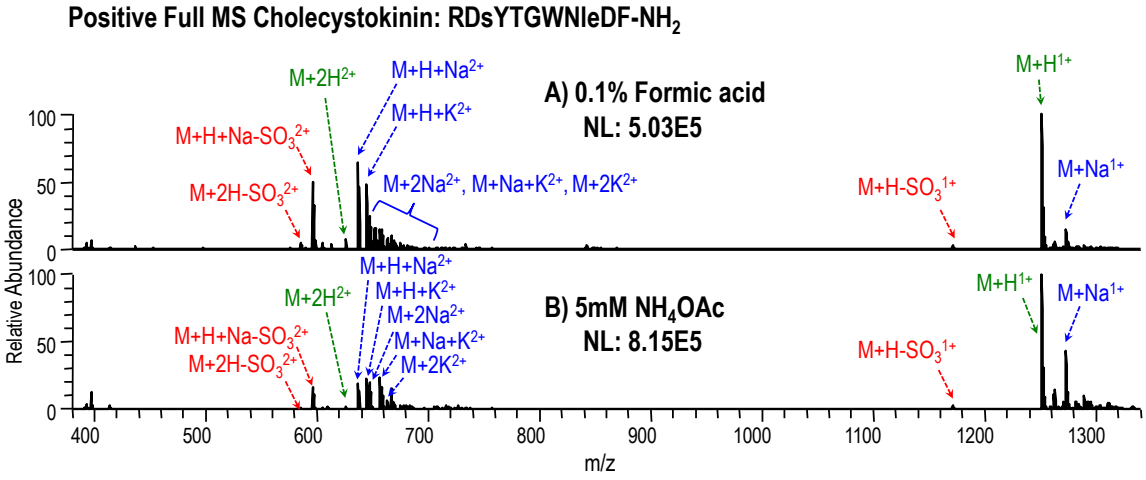


Figure 6.1 Positive mode ESI of 10 µM sulfated cholecystokinin, RDsYTGWNleDF-NH_2 in 25% methanol with the following solvent additives: 0.1% formic acid (A), 5mM ammonium acetate (B). Each spectrum is the average of 20 FTMS scans. Cholecystokinin ions, M+H and M+2H, are annotated in green while various salt adducts of these ions are annotated in blue. Ions that exhibit neutral loss of SO3 are annotated in red.

Under both conditions, singly protonated cholecystokinin was the most abundant charge state. A doubly protonated ion was also observed but in low abundance relative to an extensive array of sodium and potassium adducts. Further, spontaneous loss of SO_3 was observed during ESI for all charge states, with neutral loss ions appearing about 40% abundant relative to 2+ sulfated precursor ions and about 3% abundant relative to 1+ sulfated precursor ions.

Upon CID of 1+ cholecystokinin, loss of SO_3 was the predominant product ion observed with b_6 through b_8 minus SO_3 ions observed only after magnifying the spectrum 50 times (**Figure 6.2A**). The higher energy deposition of HCD facilitated extensive backbone fragmentation of 1+ cholecystokinin (**Figure 6.2B**); however, all product ions lacked the sulfate modification. UVPD is an even higher energy process, with a single

193 nm photon having an energy of 6.4 eV. As a result, fragmentation can occur at each bond along the peptide backbone to generate *a*, *b*, *c*, *x*, *y*, and *z* ions. UVPD performed similarly to HCD (**Figure 6.2C**), producing an array of backbone fragments including *a*, *b*, and *c* ions, none of which retained the SO_3 modification, allowing the amino acid sequence to be determined while the actual site of sulfation could not be pinpointed.

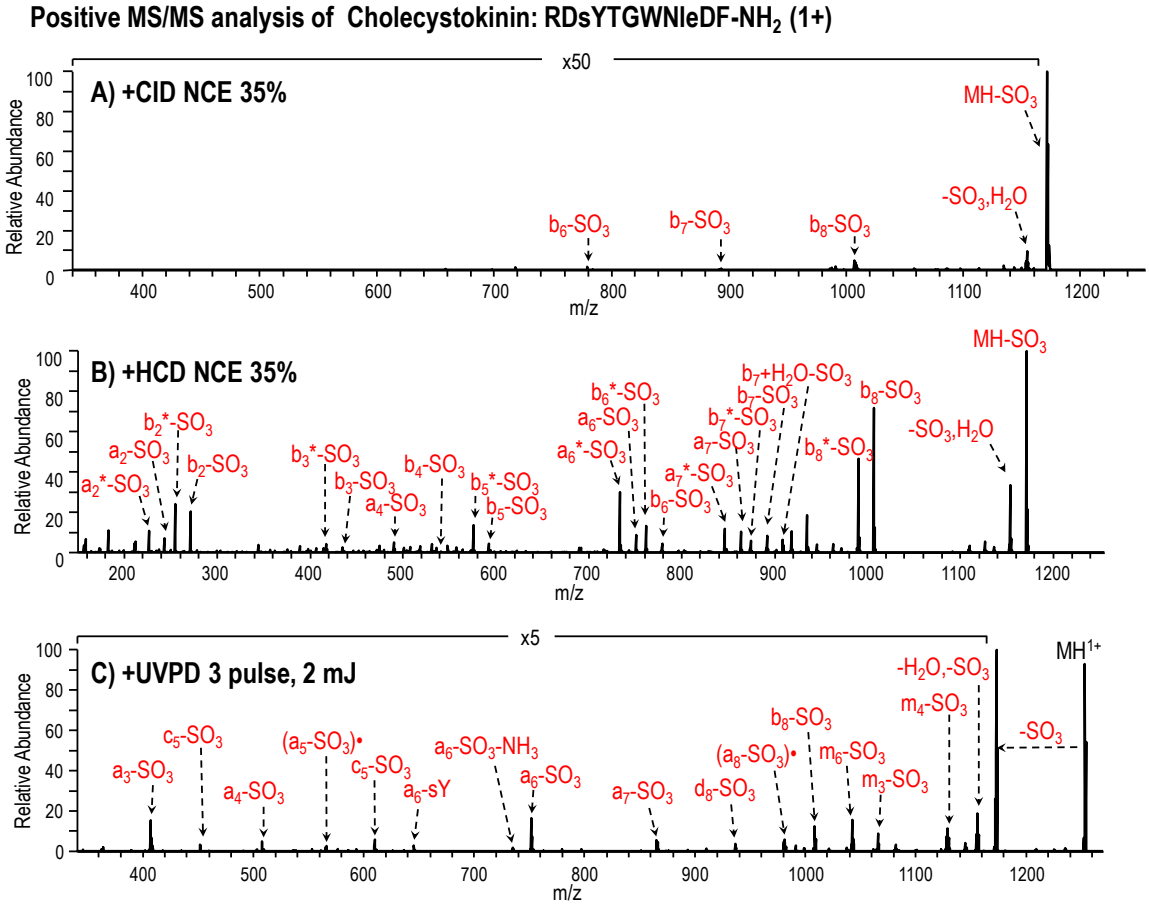


Figure 6.2 MS/MS fragmentation behavior of 1+ cholecystokinin upon (A) CID, (B) HCD, each at 35% NCE, and (C) UVPD using three 2 mJ laser pulses Neutral loss of the entire sulfo-tyrosine side chain ($CH_2C_6H_4HSO_4$) was observed upon UVPD and annotated as "-sY". In all spectra, neutral loss of SO_3 or sY are annotated in red. Neutral loss of NH_3 is indicated by an asterisk (*).

Although doubly protonated cholecystokinin was also present in the MS1 spectrum, this charge state was more difficult to analyze based on instability during the ion isolation stage prior to MS/MS activation. **Figure 6.3** shows a drastic decrease in ion abundance as well as a shift from the normal isotope distribution of 2+ cholecystokinin using an isolation width of 2 without applying normalized collision energy.

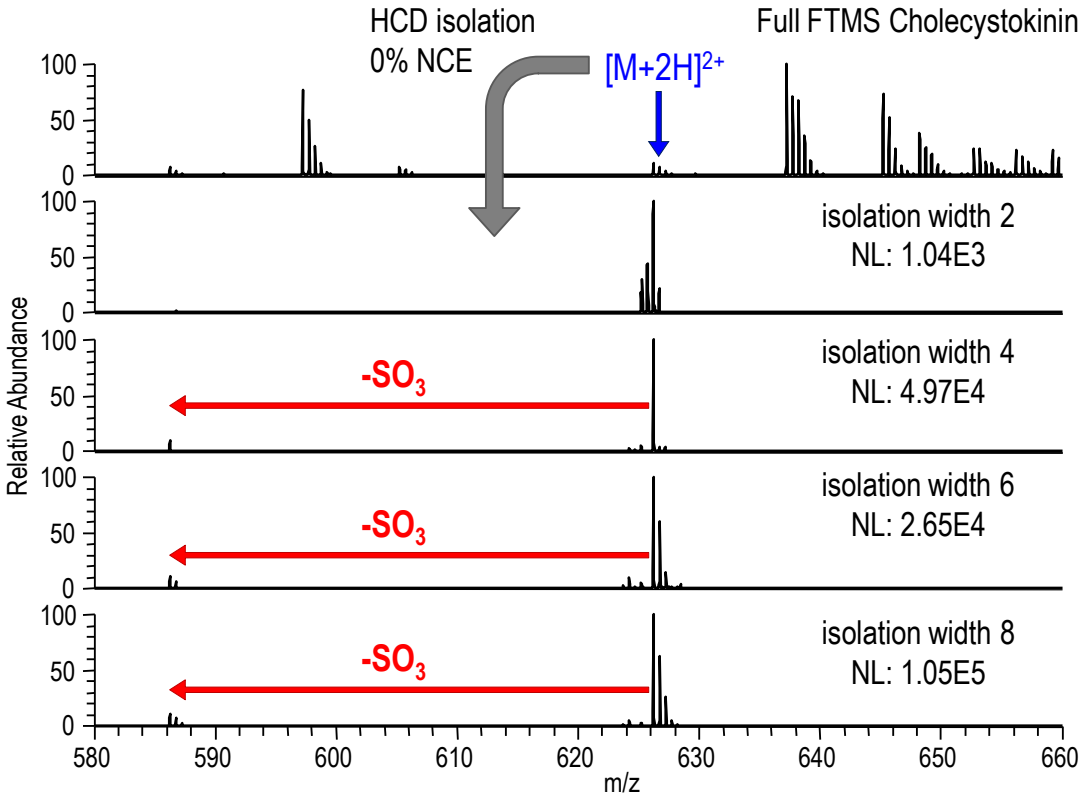


Figure 6.3 FT isolation of 2+ sulfated cholecystokinin at various isolation widths

Subsequent isolations at widths 4, 6, and 8 were performed while maintaining 0% normalized collision energy to determine the optimal isolation width. At isolation widths 6 and 8, the isotope distribution returned to normal and the ion abundance was on scale with that observed in the full mass spectrum prior to isolation. Despite these

improvements at increased isolation widths, SO_3 loss is observed following ion isolation further demonstrating the lability of the sulfoester bond. Activation using CID, HCD, and UVPD (**Figure 6.4A-C**) resulted in spectra similar to those obtained for 1+ cholecystokinin. Upon ETD (**Figure 6.4D**), three sulfated c ions, c_6-c_8, were generated accounting for 38% sequence coverage.

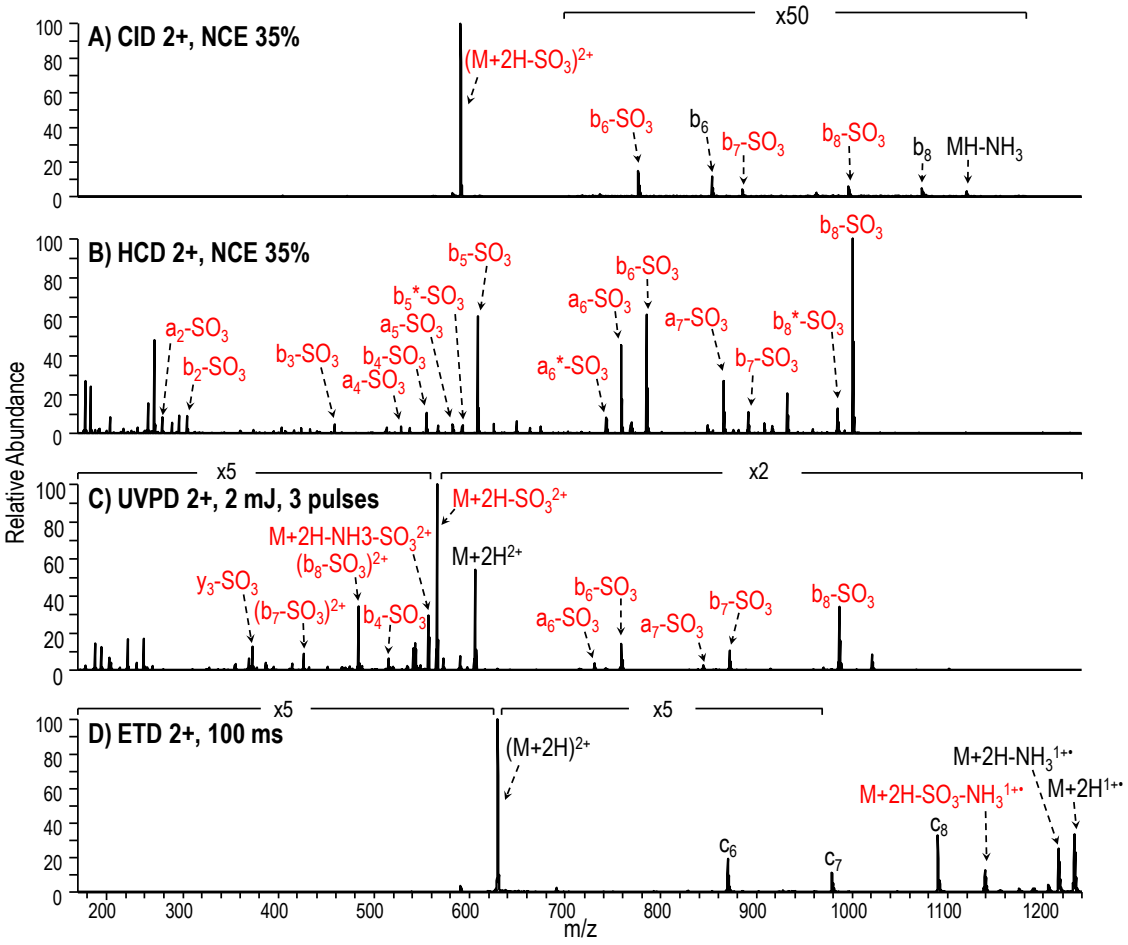


Figure 6.4 MS/MS fragmentation behavior of 2+ cholecystokinin upon (A) CID, (B) HCD, each at 35% NCE, and (C) UVPD using three 2 mJ laser pulses, and (D) ETD using activation time 100 ms. Neutral losses of SO_3 are annotated in red.

Under negative electrospray conditions, sulfopeptides ionized more readily and exhibited no loss of SO_3. **Figure 6.5** shows full MS results for cholecystokinin in pH neutral solution containing 5 mM ammonium acetate and basic solution containing 0.1% ammonium hydroxide. Electrospray using both solutions yielded an abundant 2- sulfated precursor ion with minimal salt adducts.

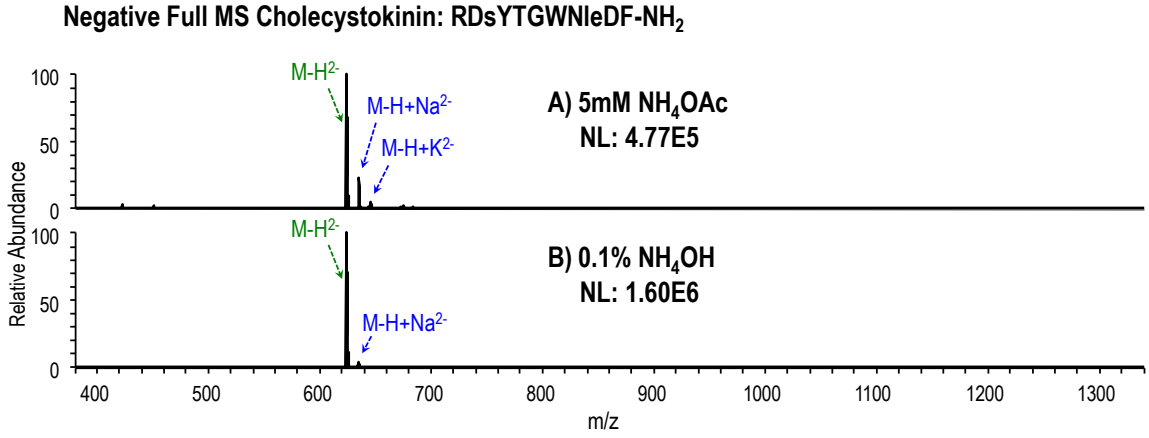
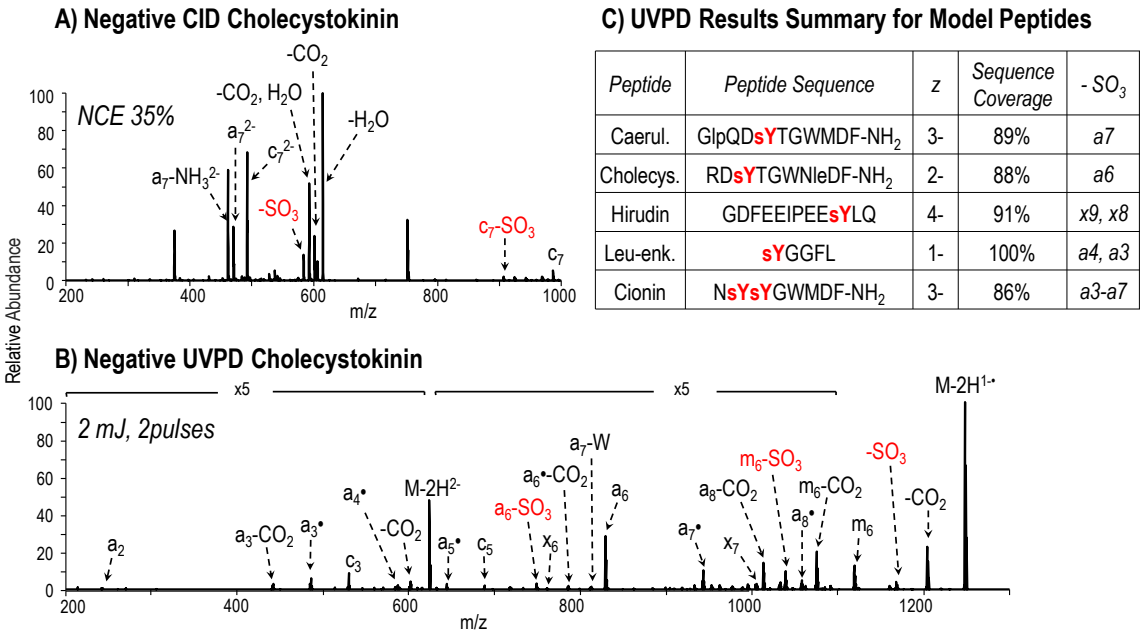


Figure 6.5 Negative ESI of 10 µM sulfated cholecystokinin, RDsYTGWNleDF-NH_2. in 25% methanol with the following solvent additives: 5 mM ammonium acetate (A), 0.1% ammonium hydroxide (B). Each spectrum is the average of 20 FTMS scans. Cholecystokinin M-2H ions are annotated in green. Sodium and potassium adducts to the precursor ion are annotated in blue.

The CID spectrum for doubly deprotonated cholecystokinin (**Figure 6.6A)** was dominated by the neutral losses of H_2O, CO_2, and SO_3, with water loss being the preferred fragmentation pathway. A few product ions were derived from cleavage of the peptide backbone, including a_7 and c_7. While these fragment ions retained the SO_3 modification, alone they did not provide sufficient information to sequence the peptide.



C) UVPD Results Summary for Model Peptides

Peptide	*Peptide Sequence*	*z*	*Sequence Coverage*	*- SO_3*
Caerul.	GlpQD**sY**TGWMDF-NH_2	3-	89%	*a7*
Cholecys.	RD**sY**TGWNleDF-NH_2	2-	88%	*a6*
Hirudin	GDFEEIPEE**sY**LQ	4-	91%	*x9, x8*
Leu-enk.	**sY**GGFL	1-	100%	*a4, a3*
Cionin	N**sY**s**Y**GWMDF-NH_2	3-	86%	*a3-a7*

Figure 6.6 MS/MS for 2- cholecystokinin (RDsYTGWNleDF-NH_2), m/z 624.24, using (A) CID at normalized collision energy (NCE) 35% and (B) UVPD using 3 pulses at 2 mJ. "-W" denotes side chain loss from tryptophan (C_9H_7N, 129 Da). Products ion that have lost SO_3 are annotated in red font. UVPD data for the other model sulfopeptides is summarized in part (C) with sequence coverage listed for the most abundant charge state of each peptide. Product ions from which SO_3 loss is observed are also listed.

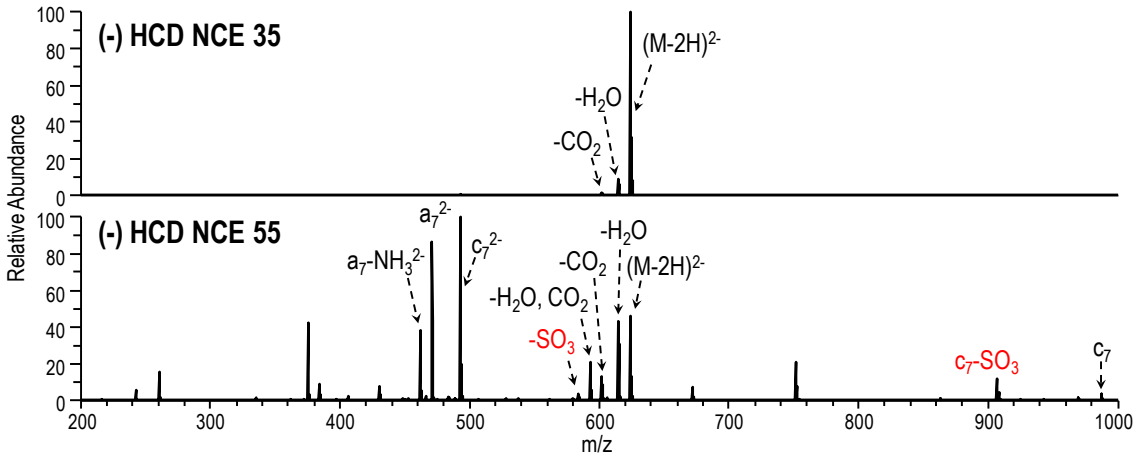


Figure 6.7 HCD fragmentation of doubly deprotonated cholecystokinin, RDsYTGWLDF-NH_2(2-), m/z 624.24 using NCE 35% (top) and NCE 55% (bottom). Sulfate loss ions are annotated in red.

HCD results for deprotonated cholecystokinin (**Figure 6.7**) mirrored those obtained using CID although more energy was required to obtain the same fragmentation results. In contrast, UVPD of deprotonated cholecystokinin (**Figure 6.6B)** provided nearly complete sequence coverage afforded by the broad series of *a* and *x* ions with only a single missed cleavage between the first and second amino acid residues (no a_1 or x_8 ions). Also, significant SO_3 loss was only observed for one product ion, a_6, for which the analogous intact sulfated product ion was also observed and in greater abundance relative to the corresponding sulfate loss ion. Nearly full sequence coverage was likewise obtained upon UVPD of deprotonated caerulein, hirudin, and leu-enkephalin analyzed in 3-, 4-, and 1- charge states, respectively. These results highlight the lack of charge state dependence on UVPD performance, thus making ion abundance the most important factor in choosing a particular charge state for dissociation. Some loss of SO_3 from product ions was also observed for each peptide (listed in **Figure 6.6C**), but again the modified form of each ion was always detected. Interestingly, the UVPD results for di-sulfated cionin were very different from the UVPD results for the singly sulfated peptides. While an extensive series of *a* ions was generated, each ion with the exception of a_2 also underwent neutral loss of one out of the two SO_3 modifications (**Figure 6.8**).

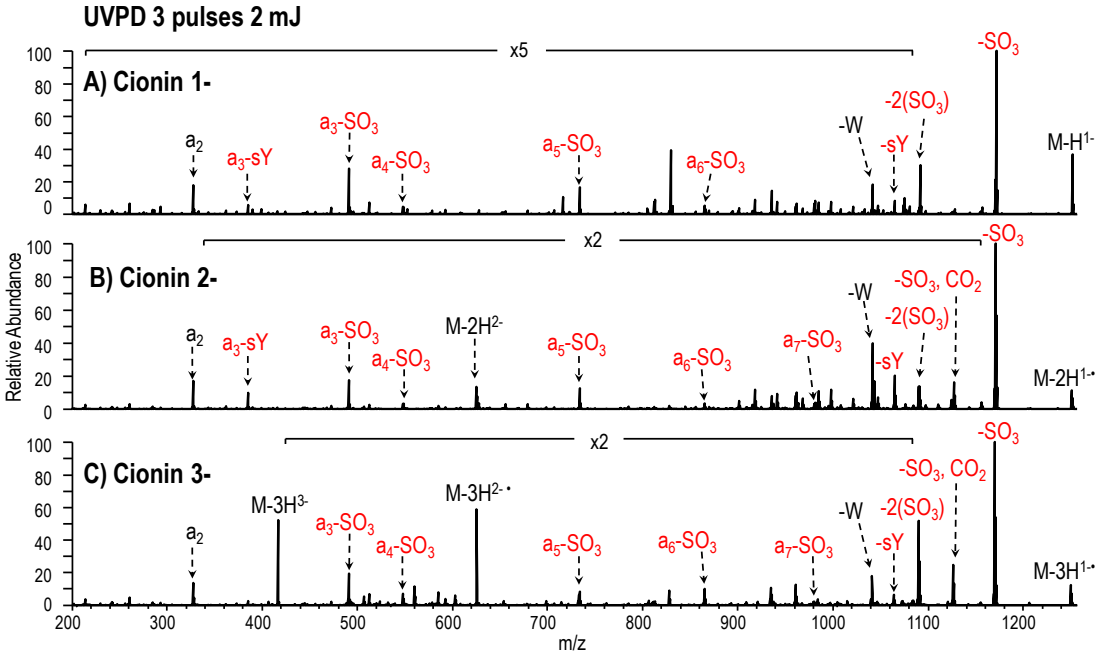


Figure 6.8 UVPD of doubly sulfated cionin in1- (A), 2- (B), and 3- (C) charge states.

This result may be rationalized based on the side-by-side positioning of the sulfotyrosine residues in the peptide sequence and the ensuing instability caused by simultaneous deprotonation of both of the sulfate moieties, thus causing proton driven loss of one sulfate.[33,49]

In addition to formation of diagnostic sequence ions, charge reduction of the deprotonated precursor ion via electron photodetachment [50] and concomitant loss of CO_2 and SO_3 from these ions were also dominant fragmentation pathways upon UVPD. Several amino acid side chains also proved to be labile upon UVPD including the tryptophan side chains (C_9H_7N, 129 Da) and glutamic acid side chains ($C_3H_4O_2$, 72 Da), which were observed as abundant neutral losses from the precursor and charge reduced radical ions. These side chain losses have also been reported previously following

negative electron transfer dissociation (NETD) and 266 nm UVPD of proton deficient radical cations.[51,52] Combinations of CO_2, SO_3, and peptide side chain neutral losses were also commonly observed upon UVPD.

6.4.2 Differentiating sY from pY

After thorough analysis of UVPD fragmentation of deprotonated sulfated peptides, UVPD of deprotonated phospho-tyrosine containing peptides was examined to determine if the two modifications could be distinguished. Differentiating sulfotyrosine from phosphotyrosine is challenging based on the nearly isobaric nature of the modifications, with sulfation adding 79.956 Da and phosphorylation adding 79.966 Da. Other studies have approached the problem by exploiting differences in the neutral loss characteristics of phosphotyrosine- and sulfotyrosine-containing peptide anions upon CID.[30] Based on the UVPD fragmentation behavior of four deprotonated phosphotyrosine-containing peptides including TSTEPQpYQPGENL, Ac-DpYVPML-NH_2, RRLIEDAEpYAARG, and Ac-IpYGEF-NH_2, it appears that a similar method monitoring sulfo and phospho neutral losses can be used for UVPD. An example of the comparative UVPD spectra obtained for a deprotonated sulfopeptide and phosphopeptide is shown in **Figure 6.9** (for sulfated leu-enkephalin and phosphorylated P60-src substrate IIAc-IpYGEF-NH_2).

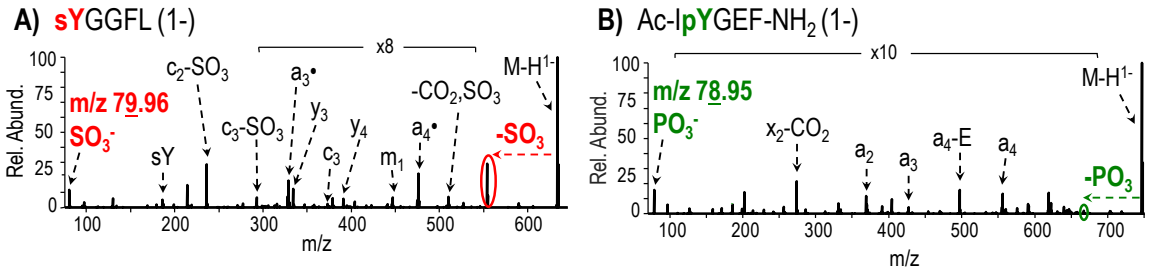


Figure 6.9 Negative UVPD mass spectra of singly deprotonated peptides: **(A)** sulfated leu-enkephalin and **(B)** phosphorylated P60c-src substrate II Ac-IpYGEF-NH_2 using 3 laser pulses at 2 mJ.

Table 6.1 displays the average relative percentage of loss of SO_3 and PO_3 from all the singly sulfated and phosphorylated peptides in the 1-, 2-, and 3- charge states (doubly sulfated cionin is excluded). For all phosphopeptides in all charge states, the percentage neutral loss was below 1%. The sulfopeptides, in contrast, showed a much greater extent of neutral loss with 23%, 27%, and 66% loss from 1-, 2-, and 3- precursor and charge-reduced ions.

Average Neutral Loss Percentage

	1-	2-	3-
Sulfated:	23.4	26.7	65.8
Phosphorylated:	0.1	0.2	0.6

Table 6.1 Average percentage of the neutral loss product observed from the selected precursor and charge reduced precursor ions. Cionin is excluded from the sulfated peptide averages.

For doubly and triply deprotonated precursor ions, the greatest neutral loss of the modification was observed from the 1- charge reduced ion derived from one electron photodetachment from 2- precursors and from two electron photodetachment products from 3- precursors. Finally, distinct reporter ions of m/z 79.96 and m/z 78.95 were

detected in the UVPD mass spectra for deprotonated sulfotyrosine and phosphotyrosine peptides, corresponding to SO_3^- and PO_3^- ions, respectively. A one Dalton difference is easily distinguished using current mass spectrometry instrumentation; however, for detection in an Orbitrap mass analyzer (which UVPD necessitates since it is undertaken in the HCD cell of the instrument) the m/z of the precursor ion must be low enough such that the reporter ions fall within 1/20 of the precursor m/z. The UVPD spectra of deprotonated peptides sYGGFL (leu-enkephalin) and Ac-IpYGEF-NH_2 in **Figure 6.9** showcase both the detection of low mass reporter ions and the significant difference in the degree of SO_3 and PO_3 neutral loss from the precursor ion, thus allowing phosphorylation and sulfation to be readily differentiated.

6.4.3 Laser Parameter Optimization

Photodissociation offers a high degree of tunability in terms of the laser energy and the number of pulses applied for MS/MS activation, enabling the selection of different UVPD parameters to suit particular applications. For sulfopeptide analysis, the optimal UVPD settings should maximize the abundance of sequence ions while minimizing the confounding neutral loss of SO_3 from these ions. Also, the neutral loss of SO_3 from the precursor ion and the charge-reduced precursor ion would ideally be prominent, as these ions provide further evidence for the presence of sulfation on the peptide and help to differentiate sulfotyrosine from phosphotyrosine. To establish the laser conditions that best meet these criteria, sulfopeptide analysis was undertaken using a matrix of 30 different laser parameters derived from combinations of 1, 2, 3, 4, 5, or 6 pulses at energies of 1, 2, 3, 4, or 5 mJ.

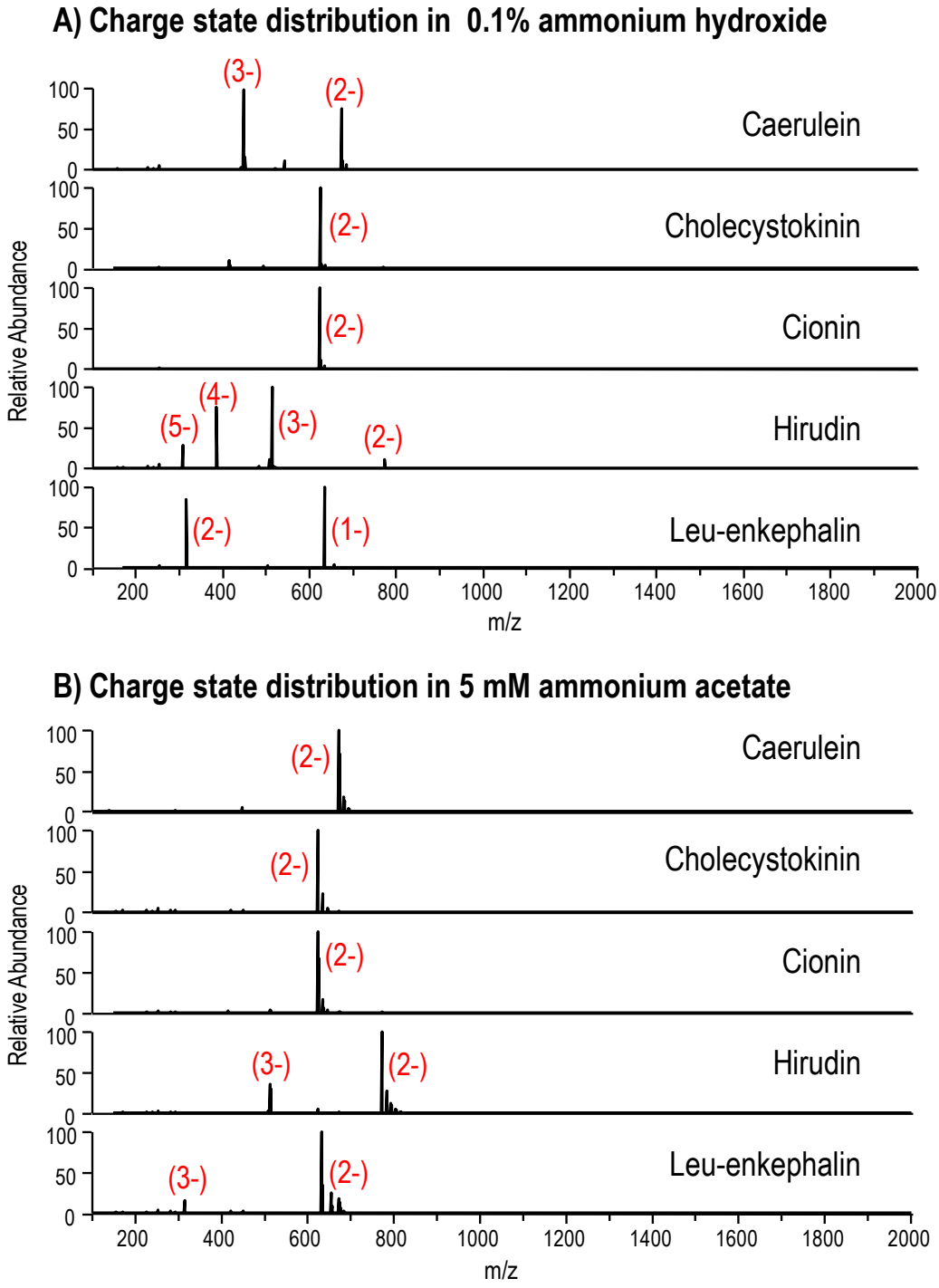


Figure 6.10 Charge state distribution of sulfopeptides caerulein, cholecystokinin, cionin, hirudin, and leu-enkephalin in (A) basic (0.1% ammonium hydroxide) and (B) neutral (5mM ammonium acetate) solutions.

Various precursor charge states were accessed using ammonium hydroxide containing solutions (**Figure 6.10A**) including 1-, 2-, 3- for caerulein and hirudin and 1-, 2- for cholecystokinin and leu-enkephalin were analyzed in order to assess the impact of charge state on UVPD fragmentation. Hirudin, the most acidic peptide, was also observed in 4- and 5- charge states; however, UVPD data from these precursors was not included for the laser parameter optimization because of the lack of supporting data from other peptides that produce ions in charge states beyond 3-. Additionally, when peptides were analyzed in 5 mM ammonium acetate buffers which are more analogous to those used for LC-MS, charge states did not exceed 3- (**Figure 6.10B**).

For each peptide, laser settings were evaluated based on the abundance of the neutral loss of sulfate from the precursor and/or charge-reduced precursor as well as the abundance of three singly deprotonated *a* or *x* products, all containing tyrosine to allow both sulfated and nonsulfated forms of each ion to be monitored. Specifically these were a_4, a_7, a_8 (m/z 568.13, 912.28, 1043.32) from caerulein; a_3, a_6, a_7 (m/z 485.15, 829.29, 942.38) from cholecystokinin; a_8, a_9, a_{10} (m/z 1124.41, 1253.45, 1259.44) from hirudin; and a_2, a_3, a_4 (m/z 271.04, 328.06, 475.13) from leu-enkephalin. To determine which laser conditions promoted the most efficient generation of informative sequence ions, the abundance of sulfated product ions were summed and charted as a function of laser conditions. Absolute abundances were then normalized to 100% for each peptide prior to averaging the values for all peptides of the same precursor charge state in order to avoid biasing the results towards the most abundant product ions arising from a single peptide. **Figure 6.11A** shows the results for 1-, 2-, and 3- peptides as heat maps in which the color red represents the greatest abundance of product ions and blue represents the lowest abundance of product ions.

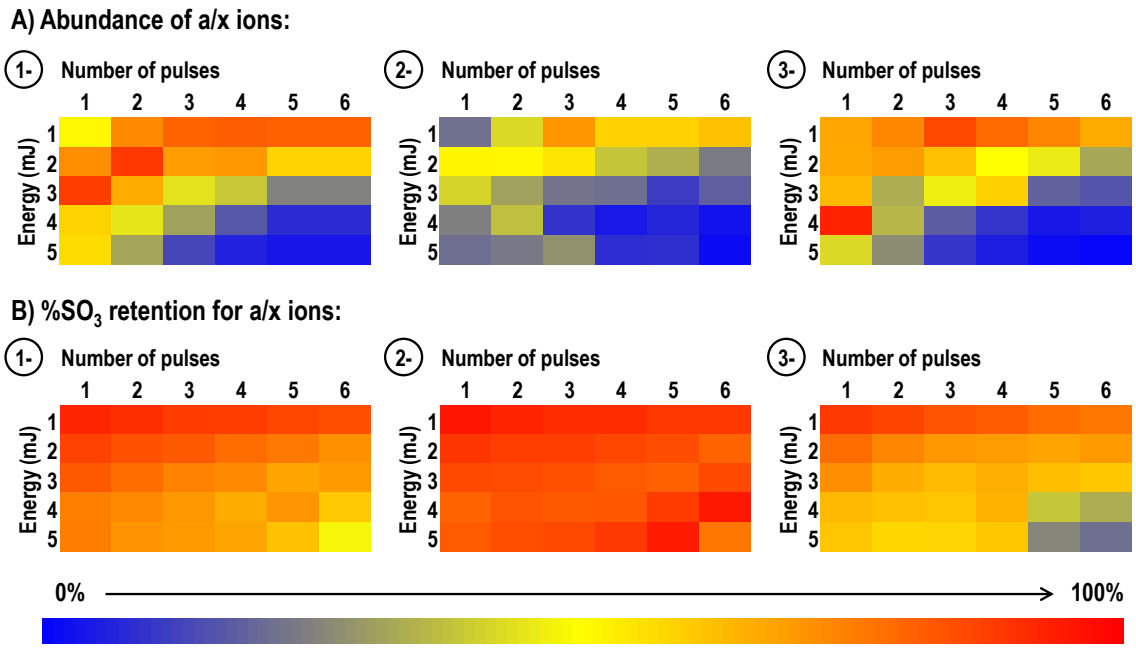


Figure 6.11 Optimization of abundances of *a/x* ions and minimization of sulfate loss: For each peptide the abundances of three singly charged Y-containing *a/x* product ions were monitored in both their sulfated and non-sulfated forms. Specifically these were *a4*, *a7*, *a8* from caerulein; *a3*, *a6*, *a7* from cholecystokinin; *x8*, *x9*, *a10* from hirudin; and *a2*, *a3*, *a4* from leu-enkephalin. **(A)** The summed abundances of sulfated product ions (normalized to 100%) is charted as a function of both laser energy (1-5 mJ) and number of pulses (1-6). **(B)** Percent sulfate retention is determined for each set of laser conditions based on the total abundance of the sulfated products ions divided by the total abundance of both sulfated and non-sulfated (neutral loss) product ions. Heat maps for 1- and 2-precursor ions (left and middle, respectively) contain data from all four peptides, whereas heat maps for 3- precursor ions (right) contain data only from hirudin and caerulein. The deepest red shade indicates the most optimum outcome.

While there is no single combination of laser energy and pulse number that outperformed all others, there is a clear trend of greater *a/x* ion abundance at lower energies and fewer laser pulses. This trend was consistent for 1-, 2-, and 3- precursor ions. These results align well with previous UVPD studies that favored the use of

minimal laser pulses and energy for fragmentation of deprotonated peptides.[38–40,42] When the abundances of SO_3 neutral loss ions from precursor and charge-reduced precursor ions were tabulated, normalized, and averaged across peptides (**Figure 6.12**), the results showed the same pattern that was observed for *a*/*x* sequence ions, with the SO_3 neutral loss ions increasing in abundance as fewer laser pulses at lower energy were used for activation.

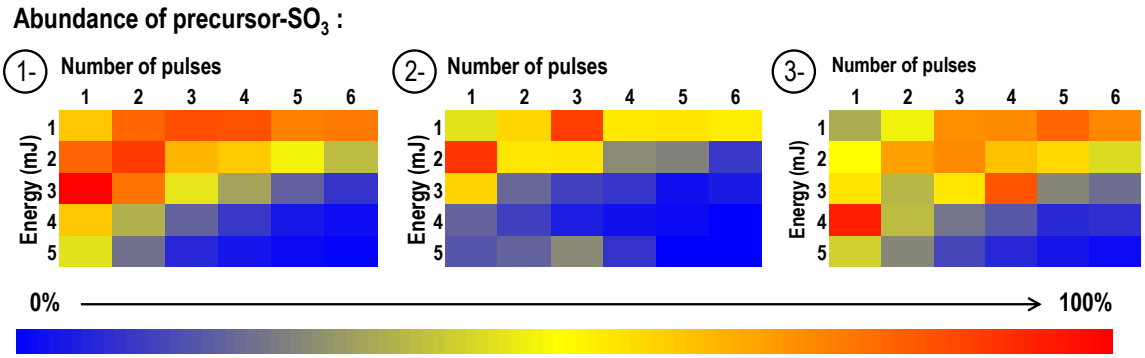


Figure 6.12 Laser parameter optimization evaluated based on the degree of SO_3 loss from the precursor or charge reduced precursor ions. The abundances of neutral loss ions were tracked across all laser conditions for each peptide and normalized to 100%, which is represented as the color red in the heat maps (0% is blue in color). Heat maps for 1- and 2- precursor ions (top and middle, respectively) contain data from all four peptides, whereas heat maps for 3- precursor ions (bottom) contain data only from hirudin and caerulein. The deepest red shade indicates the most optimum outcome.

While the neutral loss of SO_3 from precursor and charge-reduced precursor ions is a useful marker for peptide sulfation, loss of SO_3 from product ions is an undesirable outcome of UVPD that can prevent localization of the sulfation sites. Thus, in addition to monitoring the sulfated product ion abundance, it is also critical to note the abundances of non-sulfated product ions so that laser conditions which promote this neutral loss can

be identified and avoided. Changes in sulfate retention are presented as percentages based on the abundance of the three sulfated *a*/*x* ions chosen for each peptide (listed above) and the abundance of the same *a*/*x* ions found 79 Da lower in mass after sulfate neutral loss using the following equation:

$$\%SO3\ retention = \frac{abundance\ sulfated\ ions}{total\ abundance\ sulfated\ and\ nonsulfated\ ions} \times 100$$

The percentage SO_3 retention for peptides of the same charge state is averaged and displayed as heat maps in **Figure 6.11B**, with red representing 100% sulfate retention and blue 0%. Interestingly, the decrease in the abundance of sulfated products observed at high energy using many pulses does not correspond with a significant increase in the abundance of non-sulfated, neutral loss ions. Instead the percentage of SO_3 retention remains relatively constant across all laser conditions and only under the most energetic dissociation conditions (4-5 mJ, 5-6 pulse) does the extent of SO_3 loss greatly increase, showing that UVPD is a robust method for sulfation mapping. Based on these analyses, the optimum UVPD parameters were 2-3 laser pulses at 2-3 mJ.

6.4.4 MS1 and MS2 Limit of Detection

Another important performance metric to consider for negative UVPD analysis of sulfopeptides is the absolute limit of detection because of the generally low abundance of modified peptides in biological samples. Detection limits are especially pertinent for negative mode LC-MS workflows because of the lower flux of precursor ions generated compared to using the positive ESI mode.[53–56] Several LC mobile phase additives have been reported to improve ESI efficiency in the negative mode including acetic acid (pH

4) and ammonium hydroxide (pH 11);[57,58] however, prolonged exposure to acidic solutions may result in hydrolysis of sulfate and thus should be avoided in favor of higher pH alternatives.[59] At the same time, high pH solutions were damaging to the LC system and ultimately abandoned in favor of more robust separations at neutral pH using 5 mM ammonium acetate-containing mobile phases. Additionally methanol was used instead of acetonitrile as the organic mobile phase constituent to further boost the ESI efficiency.[57]

For the LOD determination, seven solutions containing 100, 250, 500, 750, 1000, 1250, and 1500 femtomoles each of caerulein, cionin, and leu-enkephalin were analyzed by LCMS. A precursor mass list was used to target each peptide in its most abundant charge state (1- for leu-enkephalin and 2- for caerulein and cionin), and MS1 and MS2 scans were alternated to ensure an even distribution of MS survey spectra and UVPD mass spectra across each eluting peak. MS extracted ion chromatograms (XIC) were generated for each peptide, and peak areas were plotted against the femtomole amount injected to create a linear calibration curve (**Figure 6.13A-B**). Each data point on the calibration curve has an associated error bar showing the standard deviation in the peak area for four replicate runs, with high reproducibility observed for measurements made at 500 femtomoles and greater (≤3%RSD) and slightly lower reproducibility (≤20%RSD) at the most dilute concentrations. Assuming the limit of detection was equal to three times the standard deviation of the peak area, divided by the slope of the calibration curve, then the resulting MS1 LODs for caerulein, cionin, and leu-enkephalin were determined to be 113 fmol, 182 fmol, and 96 fmol, respectively.

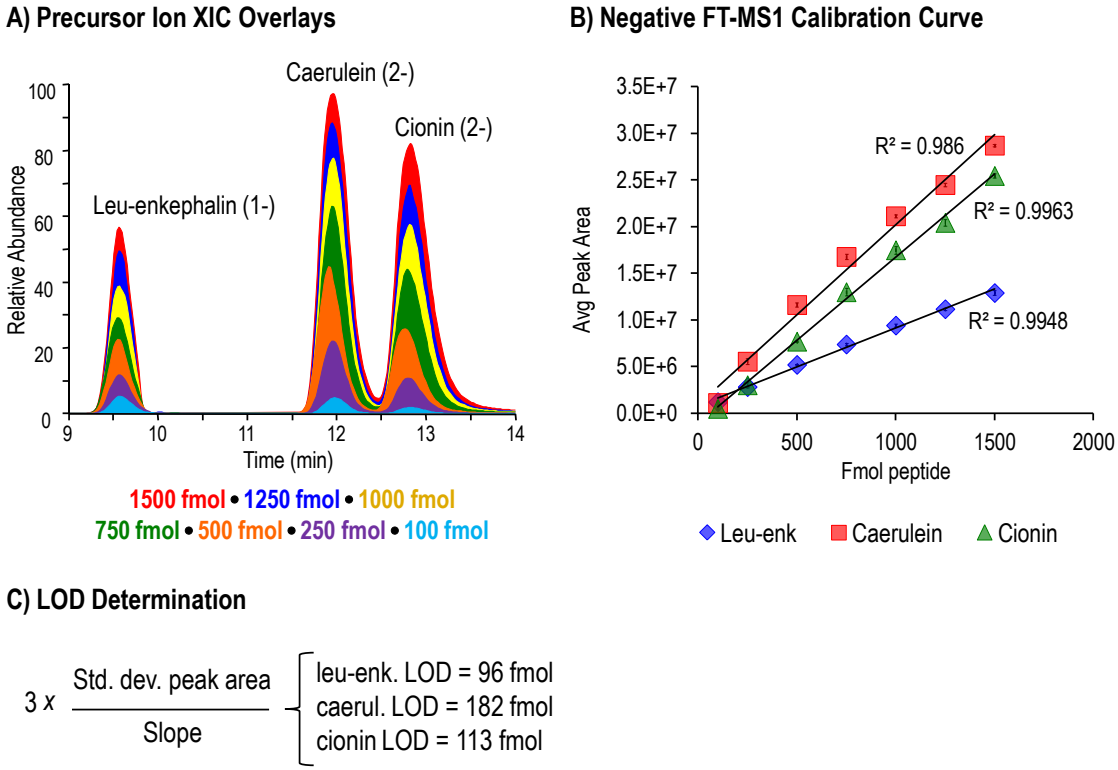


Figure 6.13 MS1 LOD for leu-enkephalin (1-), caerulein (2-), and cionin (2-) **A)** XIC for injections of 1500, 1250, 1000, 750, 500, and 250 fmol of peptide. **B)** The area under the curve was plotted as a function of femtomoles of peptide injected to generate a calibration curve for each peptide. **C)** MS1 LOD was determined for each peptide using the calibration curve slope and the standard deviation of the average peak area.

To evaluate the MS2 LODs, XICs were generated based on the sum of the three most abundant *a*-type product ions for each peptide including: a_4, a_7, a_8 for caerulein; a_2, a_3, a_4 for leu-enkephalin; and a_2, a_3-SO_3, a_5-SO_3 for cionin. XICs and calibration curves for each peptide are displayed in **Figure 6.14A-B**. When the calibration curve was used to calculate MS2 LOD, the detection limits for each peptide were lower than those obtained for MS1, a result which is logically unsound because an ion cannot be selected for MS2 activation if it is not first detected by MS1.

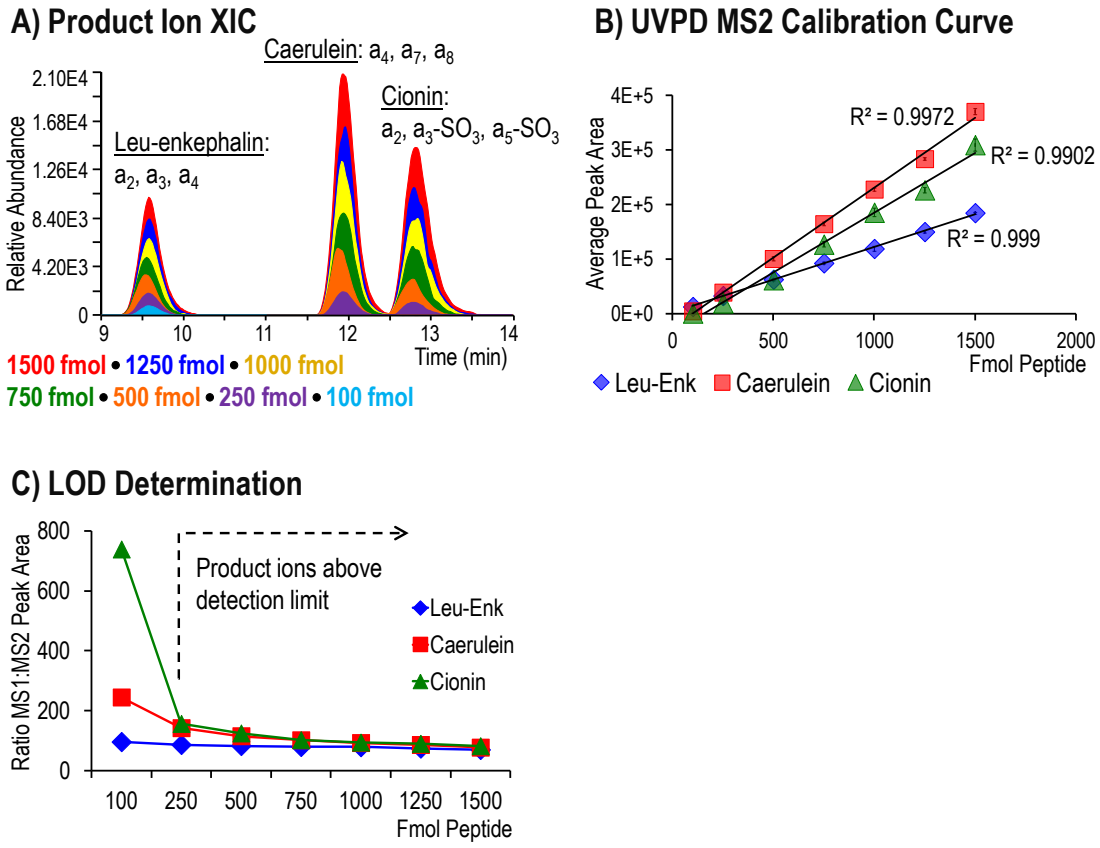


Figure 6.14 MS2 LOD determined for leu-enkephalin (1-), caerulein (2-), and cionin (2-) using three 2 mJ pulses for photodissociation. **(A)** XIC representing the summed abundance of three product ions from each peptide were generated for each dilution and overlaid. **(B)** The area under the curve was plotted as a function of femtomoles of peptide injected to generate a calibration curve for each peptide. **(C)** Precursor ion peak area (MS1) was divided by product ion abundance (MS2) and plotted against the number of femtomoles injected.

Following this result, a more empirical measure of MS2 LOD was employed based on the characteristic decrease in product ion abundance relative to precursor ion abundance upon UVPD. This was systematically accomplished using the MS1 peak areas from the XICs in **Figure 6.13A** and dividing each by the corresponding MS2 peak areas

from **Figure 6.14A**. **Figure 6.14C** shows the plot of MS1:MS2 ratio (factor decrease in abundance from MS1 to MS2) as a function of the number of femtomoles injected. In the range of 1500-250 fmol, the MS1:MS2 peak area ratio for each peptide is about 100 which corresponds to a 2 order of magnitude difference in precursor and product ion abundance. For caerulein and cionin, an increase in the MS1:MS2 ratio is observed at 100 femtomoles, indicating a larger than characteristic drop in the product ion abundance following UVPD and suggesting that the detection limit has been exceeded. Leu-enkephalin did not exhibit the same increase and thus the limit of detection is expected to be lower than 100 fmol. Inspection of the MS/MS spectra at 250 and 100 femtomoles for each peptide (Supplemental figure 12) confirmed the LOD results of 250 fmol for caerulein and cionin and near 100 fmol for leu-enkephalin.

Because molar detection limits will be specific for different peptides though, a more universal measure for the MS2 detection limit could be defined as the lowest precursor ion signal that can undergo UVPD and produce product ions which can be distinguished from background noise. Since, UVPD product ions are generally about two orders of magnitude lower in abundance compared to their precursor ion abundances, precursor ion abundances should be at least on the order of 10,000 units in order to detect the most abundant UVPD sequence ions and on the order of at least 100,000 units for the best quality UVPD data. Understanding the lower detection limits of UVPD can be useful for data dependent LCMS runs in which a threshold signal for MS2 can be defined, therefore making it possible to avoid activating ions that will not provide meaningful data. Also noteworthy is the abundance of the neutral loss of SO_3 from precursor and charge reduced ions which is typically greater compared to the abundance of sequence ions. For 100 fmol injections of caerulein and cionin (**Figure 6.15**), the SO_3 neutral loss ion was observed despite a lack of corresponding sequence ions. Detection of these

sulfate loss ions can provide evidence of peptide sulfation and enable specific ions to be targeted in subsequent analyses to obtain improved MS/MS results.

The MS1 and MS2 detection limits reported for the LC-UVPD-MS strategy are suitable for the analysis of isolated sulfopeptides in which several micrograms of total digest are routinely injected for LC-MS analysis. For a complex biological sample, enrichment of sulfated peptides would be necessary based on their low stoichiometric abundances relative to unmodified proteins in the sample. Various enrichment methods have been reported including ones that utilize immobilized metal affinity chromatography, weak anion exchange, and anti-sulfotyrosine monoclonal antibodies and we are currently integrating a robust enrichment method with the sensitive UVPD-MS approach[60–62]. Additionally, the UVPD method is not designed for direct peptide quantification per se, although this could be achieved using peptide standards to generate a calibration curve or through the use of various label or label-free methods for relative quantification between samples.

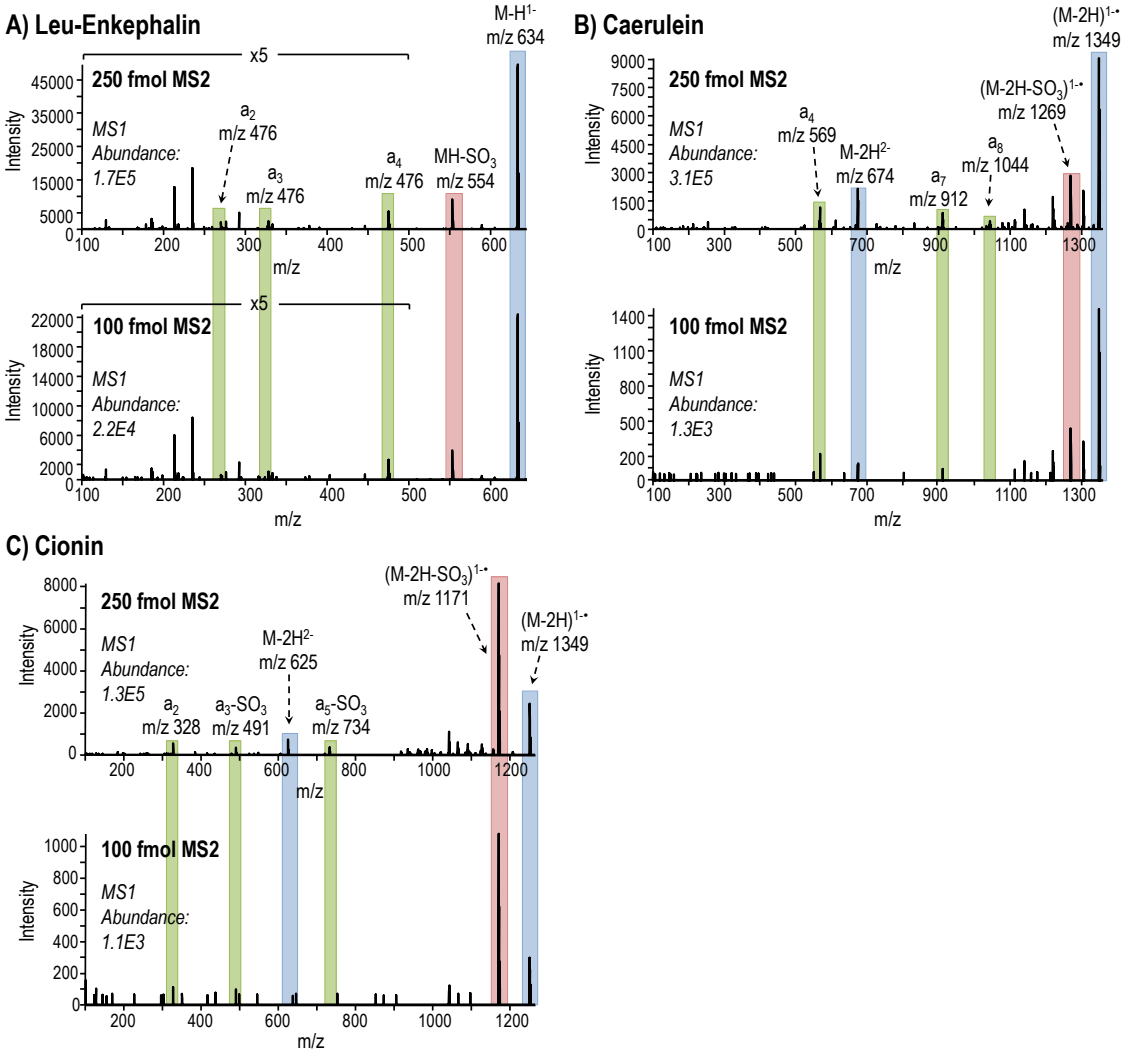


Figure 6.15 UVPD for (A) leu-enkephalin, (B) caerulein, and (C) cionin in 100 and 250 femtomole injections of the equimolar peptide mixture. In each set of spectra the three *a* ions used for MS2 LOD are highlighted in green, the precursor and charge reduced precursor and highlighted in blue, and the neutral loss of SO_3 from the precursor is highlighted in red.

6.4.5 Analysis of Bovine Fibrinogen

Bovine fibrinogen is a heterohexameric protein containing two sets of three non-identical alpha (615 residues, 67 kDa), beta (468 residues, 53 kDa), and gamma (444

residues, 50 kDa) chains. A single sulfo-modification is expected at tyrosine 6 on the beta chain. Trypsin-digested fibrinogen was analyzed by LCMS-UVPD in the negative mode using 3 pulses at 2 mJ based on the parameter optimization undertaken for the sulfopeptides. The base peak chromatogram is shown in red in **Figure 6.16A** with an XIC for the sulfopeptide of interest overlaid in green.

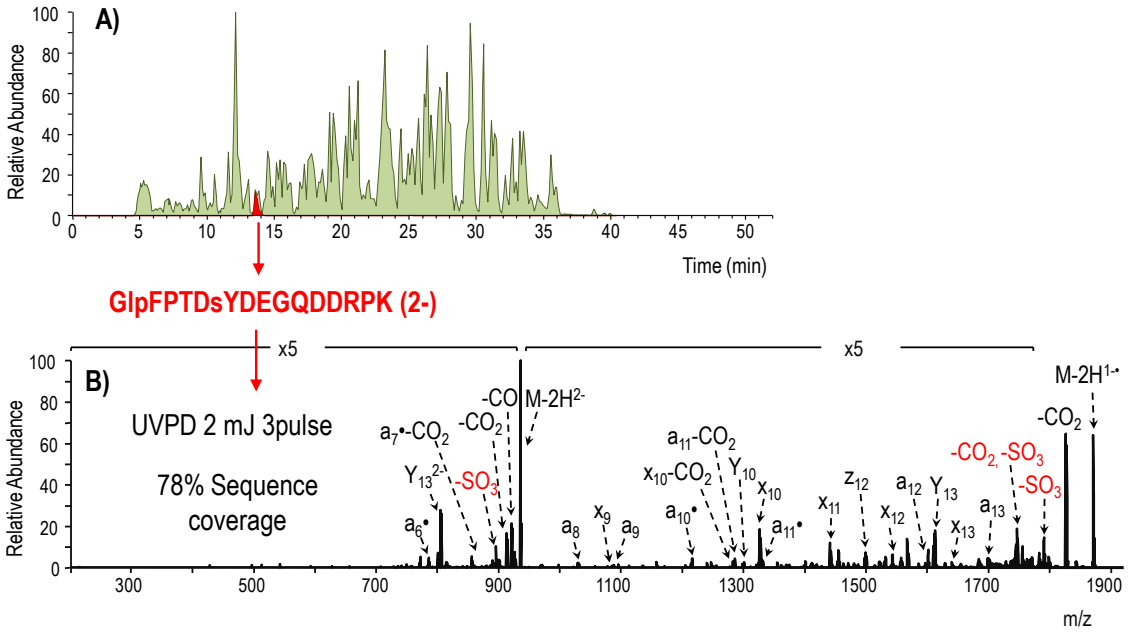


Figure 6.16 (A) Base peak chromatogram of tryptic bovine fibrinogen (red) with extracted ion chromatogram for sulfopeptide GlpFPTDsYDEGQDDRPK (2-) from the fibrinogen beta chain (green) overlaid. B) Negative UVPD mass spectrum (three pulses at 2 mJ) of GlpFPTDsYDEGQDDRPK (2-) from the average of 18 MS/MS scans acquired over 12 seconds.

The sequence coverage of the alpha, beta, and gamma chains obtained using UVPD was about 38%, 40%, and 19% respectively. The sulfated peptide, GlpFPTDsYDEGQDDRPK, was identified based on accurate precursor mass and the diagnostic 80 Da neutral loss ion; however, the quality of the MS/MS spectrum was poor

due to a low precursor ion signal (E4 range). To increase the abundance of the sulfated peptide, a greater amount of sample was injected and the isolation width was increased from 3 to 6. Additionally, the data dependent LCMS program was revised to include a targeted segment in which UVPD was performed on GlpFPTDsYDEGQDDRPK over the course of the peptide elution to allow MS/MS spectral averaging. The targeted run yielded an extensive series of *a/x* ions, including a_6- a_{14} and x_9 - x_{14}, facilitated identification of the peptide. Because all of these product ions (with the exception of x_9) contain the sulfation and there are cleavages at both the N-terminal and C-terminal sides of the sulfo-tyrosine, the site of sulfation was unambiguously assigned.

6.5 CONCLUSION

The present report demonstrates the suitability of 193 nm UVPD-MS for the analysis of sulfopeptide anions. For mono-sulfated peptides caerulein, cholecystokinin, leu-enkephalin, and hirudin, UVPD resulted in almost full sequence coverage across all charge states analyzed. Di-sulfated cionin also generated an extensive array of product ions; however, each ion only retained one of the two sulfate moieties. Systematic evaluation of laser parameters confirmed the stability of SO_3 across various energies and numbers of pulses. This allowed the selection of laser parameters that maximized the abundance of sequence ions and SO_3 loss ions which are useful for distinguishing between the nominally isobaric tyrosine sulfation and phosphorylation modifications. LC-UVPD-MS detection limits near 100 fmol were obtained for sulfated peptides, caerulein, cionin, and leu-enkephalin. The LC-UVPD-MS method was applied for analysis of bovine fibrinogen and the expected sulfated peptide, GlpFPTDsYDEGQDDRPK, was confidently identified. We anticipate that UVPD-LC-

MS could be used for global sulfation analysis following enrichment of sulfated peptides from a biological sample.

6.6 REFERENCES

(1) Mann, M.; Jensen, O. N. *Nat Biotech* **2003**, *21* (3), 255–261.
(2) Bettelheim, F. R. *J. Am. Chem. Soc.* **1954**, *76* (10), 2838–2839.
(3) Huttner, W. B. *Nature* **1982**, *299*, 273–276.
(4) Baeuerle, P. A.; Huttner, W. B. *Journal of Biological Chemistry* **1985**, *260* (10), 6434–6439.
(5) Roland Beisswanger; Corbeil, D.; Vannier, C.; Thiele, C.; Dohrmann, U.; Roland Kellner; Ashman, K.; Christof Niehrs; Wieland B. Huttner. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95* (19), 11134–11139.
(6) Ouyang, Y.-B.; Lane, W. S.; Moore, K. L. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95* (6), 2896–2901.
(7) Ouyang, Y.-B.; Moore, K. L. *Journal of Biological Chemistry* **1998**, *273* (38), 24770–24774.
(8) Seibert, C.; Cadene, M.; Sanfiz, A.; Chait, B. T.; Sakmar, T. P. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99* (17), 11031–11036.
(9) Danan, L. M.; Yu, Z.; Hoffhines, A. J.; Moore, K. L.; Leary, J. A. *Journal of the American Society for Mass Spectrometry* **2008**, *19* (10), 1459–1466.
(10) Danan, L. M.; Yu, Z.; Ludden, P. J.; Jia, W.; Moore, K. L.; Leary, J. A. *Journal of the American Society for Mass Spectrometry* **2010**, *21* (9), 1633–1642.
(11) Moore, K. L. *The Journal of Biological Chemistry* **2003**, *278* (27), 24243–24246.
(12) Huttner, W. B. *Annual Review of Physiology* **1988**, *50*, 363–376.
(13) Stone, M. J.; Chuang, S.; Hou, X.; Shoham, M.; Zhu, J. Z. *New Biotechnology* **2009**, *25* (5), 299–317.
(14) Hille, A.; Huttner, W. B. *European Journal of Biochemistry* **1990**, *188* (3), 587–596.
(15) Kehoe, J. W.; Bertozzi, C. R. *Chemistry & Biology* **2000**, *7* (3), R57–R61.
(16) Hortin, G. *Blood* **1990**, *76* (5), 946–952.
(17) Farzan, M.; Mirzabekov, T.; Kolchinsky, P.; Wyatt, R.; Cayabyab, M.; Gerard, N. P.; Gerard, C.; Sodroski, J.; Choe, H. *Cell* **1999**, *96* (5), 667–676.
(18) Monigatti, F.; Hekking, B.; Steen, H. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2006**, *1764* (12), 1904–1913.
(19) Seibert, C.; Sakmar, T. P. *Biopolymers* **2008**, *90* (3), 459–477.
(20) Nemeth-Cawley, J. F.; Karnik, S.; Rouse, J. C. *J. Mass Spectrom.* **2001**, *36* (12), 1301–1311.
(21) Wolfender, J.-L.; Chu, F.; Ball, H.; Wolfender, F.; Fainzilber, M.; Baldwin, M. A.; Burlingame, A. L. *J. Mass Spectrom.* **1999**, *34* (4), 447–454.
(22) Medzihradszky, K. F.; Guan, S.; Maltby, D. A.; Burlingame, A. L. *Journal of the American Society for Mass Spectrometry* **2007**, *18* (9), 1617–1624.

(23) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2006**, *1764* (12), 1811–1822.
(24) Yagami, T.; Kitagawa, K.; Aida, C.; Fujiwara, H.; Futaki, S. *The Journal of Peptide Research* **2000**, *56* (4), 239–249.
(25) Liu, H.; Håkansson, K. *Anal. Chem.* **2006**, *78* (21), 7570–7576.
(26) Cantel, S.; Brunel, L.; Ohara, K.; Enjalbal, C.; Martinez, J.; Vasseur, J.-J.; Smietana, M. *Proteomics* **2012**, *12* (14), 2247–2257.
(27) Yu, Y.; Hoffhines, A. J.; Moore, K. L.; Leary, J. A. *Nat Meth* **2007**, *4* (7), 583–588.
(28) Kim, J.-S.; Song, S.-U.; Kim, H.-J. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (11), 1916–1925.
(29) Drake, S. K.; Hortin, G. L. *The International Journal of Biochemistry & Cell Biology* **2010**, *42* (1), 174–179.
(30) Edelson-Averbukh, M.; Shevchenko, A.; Pipkorn, R.; Lehmann, W. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (12), 2256–2268.
(31) Gibson, B. W.; Cohen, P. In *Methods in Enzymology*; James A. McCloskey, Ed.; Academic Press, 1990; Vol. Volume 193, pp 480–501.
(32) Cook, S.; Jackson, G. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (6), 1088–1099.
(33) Hersberger, K. E.; Håkansson, K. *Anal. Chem.* **2012**, *84* (15), 6370–6377.
(34) Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (2), 197–206.
(35) Ly, T.; Julian, R. R. *Angewandte Chemie International Edition* **2009**, *48* (39), 7130–7137.
(36) Reilly, J. P. *Mass Spectrom. Rev.* **2009**, *28* (3), 425–447.
(37) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *Journal of Proteome Research* **2010**, *9* (8), 4205–4214.
(38) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. *Proteomics* **2011**, *11* (7), 1329–1334.
(39) Madsen, J. A.; Xu, H.; Robinson, M. R.; Horton, A. P.; Shaw, J. B.; Giles, D. K.; Kaoud, T. S.; Dalby, K. N.; Trent, M. S.; Brodbelt, J. S. *Molecular & Cellular Proteomics* **2013**, *12* (9), 2604–2614.
(40) Shaw, J.; Madsen, J.; Xu, H.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (10), 1707–1715.
(41) Madsen, J. A.; Ko, B. J.; Robotham, S. A.; Xu, H.; Horton, A. P.; Iwashkiw, J. A.; Shaw, J. B.; Feldman, M. F.; Brodbelt, J. S. *Anal. Chem.* **2013**, *85* (19), 9253–9261.
(42) Han, S.-W.; Lee, S.-W.; Bahar, O.; Schwessinger, B.; Robinson, M. R.; Shaw, J. B.; Madsen, J. A.; Brodbelt, J. S.; Ronald, P. C. *Nat Commun* **2012**, *3*, 1153.
(43) Vasicek, L.; Ledvina, A.; Shaw, J.; Griep-Raming, J.; Westphall, M.; Coon, J.; Brodbelt, J. *Journal of The American Society for Mass Spectrometry* **2011**, *22* (6), 1105–1108 – 1108.
(44) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.

(45) Xu, H.; Freitas, M. *BMC Bioinformatics* **2007**, *8* (1), 133.
(46) Xu, H.; Yang, L.; Freitas, M. *BMC Bioinformatics* **2008**, *9* (1), 347.
(47) Xu, H.; Freitas, M. A. *Proteomics* **2009**, *9* (6), 1548–1555.
(48) Xu, H.; Freitas, M. A. *J. Proteome Res.* **2008**, *7* (7), 2605–2615.
(49) Yagami, T.; Kitagawa, K.; Aida, C.; Fujiwara, H.; Futaki, S. *The Journal of Peptide Research* **2000**, *56* (4), 239–249.
(50) Antoine, R.; Joly, L.; Tabarin, T.; Broyer, M.; Dugourd, P.; Lemoine, J. *Rapid Commun. Mass Spectrom.* **2007**, *21* (2), 265–268.
(51) Rumachik, N.; McAlister, G.; Russell, J.; Bailey, D.; Wenger, C.; Coon, J. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (4), 718–727.
(52) Sun, Q.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R. *J. Proteome Res.* **2008**, *8* (2), 958–966.
(53) Straub, R. F.; Voyksner, R. D. *Journal of the American Society for Mass Spectrometry* **1993**, *4* (7), 578–587.
(54) Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88* (20), 4671–4675.
(55) Hiraoka, K.; Kudaka, I. *Rapid Commun. Mass Spectrom.* **1992**, *6* (4), 265–268.
(56) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20* (6), 362–387.
(57) Zhang, X.; Clausen, M. R.; Zhao, X.; Zheng, H.; Bertram, H. C. *Anal. Chem.* **2012**, *84* (18), 7785–7792.
(58) McAlister, G. C.; Russell, J. D.; Rumachik, N. G.; Hebert, A. S.; Syka, J. E. P.; Geer, L. Y.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J. *Anal. Chem.* **2012**, *84* (6), 2875–2882.
(59) Balsved, D.; Bundgaard, J. R.; Sen, J. W. *Analytical Biochemistry* **2007**, *363* (1), 70–76.
(60) Balderrama, G. D.; Meneses, E. P.; Orihuela, L. H.; Hernández, O. V.; Franco, R. C.; Robles, V. P.; Batista, C. V. F. *Rapid Commun. Mass Spectrom.* **2011**, *25* (8), 1017–1027.
(61) Amano, Y.; Shinohara, H.; Sakagami, Y.; Matsubayashi, Y. *Analytical Biochemistry* **2005**, *346* (1), 124–131.
(62) Hoffhines, A. J.; Damoc, E.; Bridges, K. G.; Leary, J. A.; Moore, K. L. *Journal of Biological Chemistry* **2006**, *281* (49), 37877–37887.

Chapter 7

Integrating Weak Anion Exchange and Ultraviolet Photodissociation Mass Spectrometry with Strategic Modulation of Peptide Basicity for the Enrichment of Sulfopeptides

7.1 Overview

Tyrosine sulfation is an important post-translational modification, but remains difficult to detect in biological samples because of the lack of effective enrichment methods. In the present study, weak anion exchange (WAX) is evaluated for the enrichment of model sulfopeptides that have been modified via carbamylation to convert all primary amines to less basic carbamates. Liquid chromatography tandem mass spectrometry (LC-MS/MS) in the negative ion mode was used to analyze the eluent from WAX, and ultraviolet photodissociation was applied for peptide sequencing. The decrease in basicity enhanced the binding of carbamylated sulfopeptides to WAX relative to a mixture of non-sulfated peptides from bovine serum albumin. The ultimate potential for sulfopeptide enrichment by two step wash and elute was evaluated, and three washes at 200 mM NH_4Cl coupled with four elutions at 5 M NH_4Cl was found to be most effective for isolating the sulfopeptides of interest.

7.2 Introduction:

Tyrosine sulfation is a post-translational modification (PTM) of secretory and transmembrane proteins that is responsible for modulating extracellular protein-protein interactions to initiate a variety of physiological and pathogenic responses.[1–3] The

important role of tyrosine sulfation in the normal mammalian lifecycle has been showcased by studies of knockdown mice which suffer abnormalities in growth, development and fertility or premature death in the absence of one or both tyrosylprotein sulfotransferase enzymes (TPST-1 and TPST-2) that are responsible for sulfation.[4–6] A growing number of tyrosine-sulfated proteins have been identified, many of which are known to function as coagulation factors[7–10], adhesion molecules[11], or G-protein coupled receptors (GPCR).[12,13] Because GPCRs and more specifically chemokine receptors play an important role in leukocyte recruitment and adhesion during the inflammatory response, these proteins represent potential therapeutic targets for the treatment of autoimmune disorders such as multiple sclerosis and rheumatoid arthreitis.[14] Infectious diseases may also proceed through chemokine receptors as demonstrated by the AIDs and malaria pathogens, HIV and *p. vivax*, which invade target cells through interaction with the sulfo-tyrosine containing portions of the chemokine receptors CCR5 and DARC.[15,16] Clearly tyrosine sulfation is a critical modification of proteins, but additional effort is needed to determine the true scope of this PTM and further define its role in biological systems. Attaining a broader understanding of sulfation has been impeded in part by inadequate methods to identify sulfoproteins.

Liquid chromatography tandem mass spectrometry (LC-MS/MS) is the premier analytical technique for global PTM analysis based on the ability to accurately identify, localize, and quantify modifications in great detail without sacrificing throughput. Many PTMs such as phosphorylation,[17] acetylation,[18] and glycosylation,[19] are routinely profiled in biological systems using LC-MS/MS, but to date no mass spectrometry-based global sulfation analysis has been reported due to a lack of appropriate methods for sulfopeptide characterization. The highly labile sulfo-ester bond and the acidic sequence motifs that often surround the site of sulfation present a mismatch for conventional methods that

employ positive mode ionization coupled with collision induced dissociation (CID).[20,21] Using these methods, sulfopeptides experience partial or complete modification loss during ionization and the first stage of mass analysis, and any remaining sulfated ions are effectively stripped of SO_3 upon CID.[22,23] Electron transfer and electron capture dissociation (ETD and ECD) also largely fail to maintain sulfations despite a proven record of preserving other labile PTMs, further underscoring the need for new LC-MS/MS methodologies.[24] One strategy for improved sulfopeptide analysis embraces the lability of sulfate in a subtractive-based identification scheme in which free tyrosine residues are acetylated prior to MS analysis so that any unmodified tyrosine residues detected must be derived from sulfate loss in the mass spectrometer.[25,26] These techniques vastly improve sulfate site localization using conventional LC-MS/MS methods; however, incomplete reaction of unmodified tyrosine can lead to false positive sulfopeptide identifications. Direct identification of intact sulfo-tyrosine peptides is an attractive alternative but requires that the mass spectrometer be operated in negative ion mode to improve ionization and maximize modification stability. Because collision based activation methods are ineffective for sequencing peptide anions, a number of alternative activation techniques have been developed such as metastable atom-activated dissociation (MAD),[27] negative ion electron capture dissociation (niECD),[28] and ultraviolet photodissociation (UVPD).[29–31] Each technique provides a high level of peptide sequence coverage, with diagnostic *c*/*z* backbone cleavage ions observed for NiECD and all ion series (*a/x*, *b/y*, *c/z*) observed for UVPD and MAD, without significant losses of the sulfate modifications. Both MAD and niECD have limited compatibility with online chromatography because long activation times and/or extensive spectral averaging is required to achieve optimal results. In contrast, UVPD is a fast

activation method that is routinely applied in LC-MS analysis, making it the best choice for larger scale applications that require front-end separations.[32–34]

Although significant inroads have been made for improving the analysis of sulfopeptides using mass spectrometry, the potential for analysis on a global scale remains limited by the lack of effective methods for sulfopeptide enrichment from complex biological matrices. Despite the chemical and structural similarities between sulfation and phosphorylation, the majority of enrichment techniques that have been developed for phosphopeptides are not directly applicable to sulfopeptides. These include strong cation exchange chromatography (SCX) in which peptides are separated by differences in net charge,[35–37] immobilized metal affinity chromatography (IMAC) in which metal cations coordinate phosphate groups by affinity ,[38–40] and titanium dioxide chromatography (TiO_2) which proceeds through Lewis acid-base interactions.[41–43] For optimal selectivity, each of the above methods must be carried out near pH 2 to ensure the protonation of carboxylic acids and thus reduce unwanted interactions between acidic unmodified peptides and the enrichment media. Since prolonged exposure to very acidic conditions has been shown to promote hydrolysis of sulfo-tyrosine, alternative enrichment strategies must be devised to specifically target sulfopeptides.[44]

Ga(III)-IMAC enrichment was successfully applied at a more moderate pH of 3 for the enrichment of sulfopeptides from a complex peptide mixture isolated from the skin secretions of the *P. dacnicolor* frog.[45] Four sulfopeptides were structurally characterized for the first time, however several other non-sulfated peptides were co-enriched indicating a lack of specificity. Greater specificity for sulfopeptide enrichment is possible using novel anti-sulfotyrosine antibodies developed using phage display technology.[46,47] The PSG2 antibody in particular binds with high affinity and incredible specificity to sulfotyrosine resides, even discriminating against phosphotyrosine residues

and sulfated glycans.[47] Binding occurs independently of sequence context, making the anti-sulfotyrosine antibodies widely applicable for the enrichment of sulfopeptides from complex biological samples. Despite these superior performance metrics, antibodies have not been broadly adopted for sulfopeptide enrichment likely due to their high cost and susceptibility to denaturation. Molecularly imprinted polymers (MIPs) are a class of synthetic polymer receptors that combine the specificity of antibodies with the high thermal, chemical, and stress tolerance of polymers to offer a robust, reproducible, and low cost alternative to biological receptors such as antibodies.[48–50] By designing MIPs with arrangements of urea based hydrogen-bond donors to complement the desired sulfo-tyrosine acceptors, specificity was achieved for tyrosine sulfated peptides.[50] For a simple three peptide mixture, efficient separation of unmodified, phosphorylated, and sulfated peptides was demonstrated using MIPs, but this technique has yet to be evaluated on a larger scale.

Weak anion exchange chromatography exploits the strong affinity of sulfate for anion exchange relative to other peptide anion moieties such as carboxylates without the need for acidic solutions. Upon removal of basic C-terminal lysine and arginine residues in tryptic peptides using carboxypeptidase B, sulfopeptide binding to WAX was further enhanced.[51] WAX enrichment has also been demonstrated for carboxypeptidase B treated sulfated glycopeptides as well as heparin sulfate molecules.[52–54] Chemical derivatization is another option for modulating peptide charge and the carbamylation reaction provides an efficient means for neutralizing positive charges on primary amines.[55–57] In the present study, weak anion exchange chromatography is applied to sulfopeptides that have been modified to convert primary amines to less basic carbamates for improved enrichment prior to MS/MS using ultraviolet photodissociation.

7.3 EXPERIMENTAL

7.3.1 Materials

Sulfated peptides GlpQDsYTGWMDF-NH2 (caerulein) and RDsYTGWNleDF-NH2 (Thr28,Nle31-cholecystokinin-33 sulfated) were purchased from American Peptide Company (Sunnyvale, CA); GDFEEIPEEsYLQ (hirudin fragment 54-65) and NsYsYGWMDF-NH2 (cionin) were purchased from Sigma; and sYGGFL (leucine-enkephalin sulfated) was purchased from Phoenix Pharmaceuticals (Burlingame, CA). LCMS grade solvents were obtained from EMD Millipore (Temecula, CA), and mobile phase additives and other reagents were also obtained from either Sigma or Fisher Scientific (Fairlawn, NJ).

7.3.2 Protein digestion, carbamylation, and weak anion exchange

Bovine serum albumin (BSA) was reduced with 5 mM dithiothreitol (DTT) at 55 °C for 30 minutes followed by alkylation with 15 mM iodoacetamide (IAM) proceeding for 30 min at room temperature in the dark. An additional aliquot of 5 mM DTT was added to quench the alkylation reaction. MS grade trypsin (Thermo Fisher Scientific, Grand Island, NY)) was added in a 1:50 enzyme to protein ratio, and digestion occurred overnight at 37 °C. After proteolysis, model sulfopeptides were spiked into solution with the BSA peptides and the resulting mixture was immediately subject to either weak anion exchange chromatography (WAX) or carbamylation reaction. For carbamylation, urea was added directly to digest solutions to a concentration of ~8 M and the samples were incubated at 80 °C for 4 hours. Both unmodified and carbamylated peptide mixtures were diluted to 500 µL in 50 mM ammonium chloride in preparation for weak anion exchange.

Weak anion exchange (WAX) columns for offline fractionation and enrichment of sulfopeptides were prepared as follows. A 50 mg/mL slurry of Diethylaminoethyl (DEAE)-Sephadex A-25 (Sigma, St. Louis, MO) resin was prepared in 50 mM ammonium acetate and allowed to swell overnight at room temperature. Fritted columns were loaded with 400 µL of the slurry accounting for 20 µg of DEAE-Sephadex, and the resin bed was secured by placement of a second frit on top. Flow through the WAX columns was driven by gravity. Columns were conditioned with several milliliters of 50 mM ammonium chloride loading buffer. Samples were loaded in 50 mM ammonium chloride and the flow-through, unbound fraction was saved. For fractionation 500 µL aliquots of increasing ammonium chloride concentration (200 mM, 400 mM, 600 mM, 800 mM, 1 M, 2 M, and 4 M) were passed through the WAX column in succession and collected separately. For the enrichment of sulfopeptides, washing and elution steps were evaluated using various volumes of different salt concentrations to determine the optimal conditions for maximum specificity. Prior to LC-MS analysis, all WAX fractions were desalted using C18 stage tips that were fabricated using Empore C18 extraction disks (3M, Minneapolis, MN) according to published protocols.[58]

7.3.3 LC, MS, and UVPD

Liquid chromatography was performed on a Dionex Ultimate 3000 capillary LC system operated at a flow rate of 4 µL/min. Mobile phase A was 5 mM ammonium acetate in water that was adjusted to pH ~8 using ammonium hydroxide. Mobile phase B was 5 mM ammonium acetate in 85% acetonitrile with an equal volume of ammonium hydroxide relative to mobile phase A. WAX fractions were reconstituted in 95% mobile phase A to match starting gradient conditions. Separations of were carried on a 3 x 150

mm Agilent Zorbax Extend-C18 column with 3.5 μm particle size (Santa Clara, CA) using a linear gradient in which the percentage of mobile phase B was increased from 5-45% over 45 min.

Mass spectrometry analysis was carried out using a Velos Pro dual linear ion trap mass spectrometer (Thermo Scientific, San Jose, CA) equipped for photodissociation at 193 nm as previously described.[59] All experiments were conducted in the negative ion mode using a spray voltage of 4 kV. For negative UVPD analysis of unmodified and carbamylated WAX fractions, a top eight data dependent scan program was used in which the first scan was a negative full MS survey scan over *m/z* 400-2000, followed by eight UVPD events on the eight most abundant ions from scan event 1. Dynamic exclusion was enabled for 10 seconds with a single repeat count. For UVPD, the CID normalized collision energy was set to zero, the activation q was decreased to 0.1, the isolation width was 3 Da, and the activation time was 4 ms (allowing 2 laser pulses per scan). AGC targets for MS1 and MS2 were 30000 ions and 10000 ions, respectively.

7.4 RESULTS AND DISCUSSION

7.4.1 Carbamylation of Sulfopeptides

With the establishment of appropriate MS/MS tools for sulfopeptide characterization, the next step towards global sulfation analysis is the development of effective enrichment methodologies. Weak anion exchange (WAX) was targeted to meet this aim because it can be performed under conditions of neutral or basic pH to preserve sulfate modifications without compromising the overall selectivity for sulfopeptides. The selectivity of ion exchange is derived from various factors that modulate Coulombin interactions

including the number and location of charges on the molecule and the ion exchanger as well as the charge density. Ions with a higher charge and smaller solvated radius bind more strongly, and it has been shown that sulfate has the strongest affinity for WAX followed by chloride, phosphate, acetate, and hyrdoxide.[60] This general trend of affinity is expected to shift in the context of peptides for which multiple charged moieties contribute to the overall binding characteristics in anion exchange. In sulfopeptides the presence of multiple negatively charged (deprotonated) aspartate and glutamate residues can further influence binding relative to nonsulfated peptides which typically do not exhibit the same acidic sequence motifs. At the same time, positively charged (protonated) lysine and arginine residues diminish the binding affinity through repulsive interactions with like positive charges on the anion exchanger.[61] For the present study, in order to enhance the interaction between sulfopeptides and the WAX resin, carbamylation was used to convert the highly basic lysine residues and N-termini primary amines to less basic carbamates, thus effectively removing those ionizable sites from interaction with the WAX resin. The carbamylation reaction is an incredibly simple, efficient, and cost effective method for the derivatization of primary amines but commonly is performed at high temperatures over an extended period of time (80°C for 4 hrs) to proceed to completion.[55] To ensure that these rather harsh reaction conditions would not promote hydrolysis of the labile sulfate modification, carbamylation was first applied to several model sulfopeptides including doubly sulfated cionin and singly sulfated cholecystokinin, Leu-enkephalin, and hirudin. Each of these peptides contains one N-terminal primary amine that is expected to undergo carbamylation. LC-MS/MS in the negative ion mode using 193 nm

UVPD was used to confirm the integrity of the sulfopeptides following carbamylation and also evaluate the reaction efficiency (**Figure 7.1A-D**).

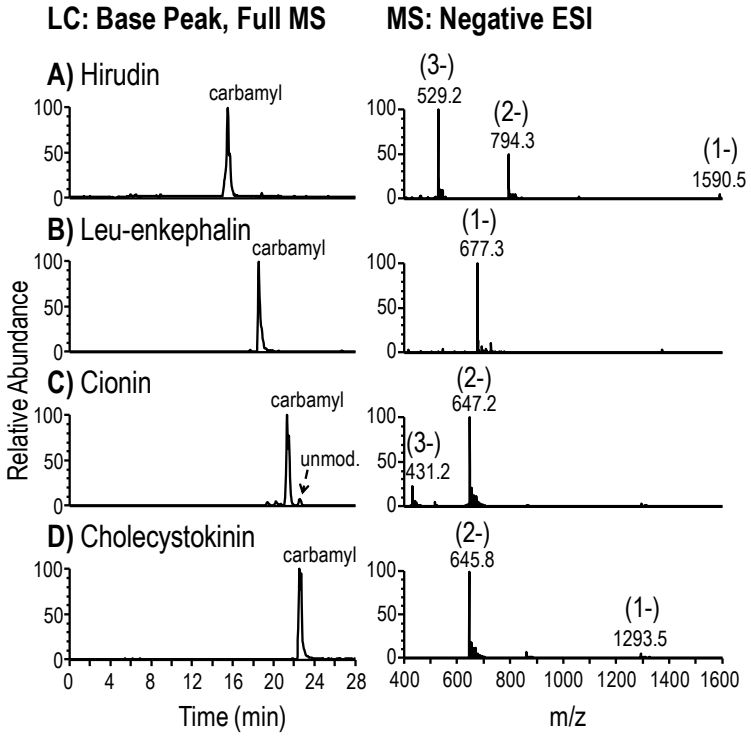


Figure 7.1 Carbamylation of sulfopeptides (A) hirudin, (B) leu-enkephalin , (C) cionin, and (D) cholecystokinin demonstrated in base peak LC-MS chromatograms and negative ESI spectra

For each peptide, a single dominant peak was observed in the LC-MS chromatogram (**Figure 7.1, left panel**) which was identified by mass spectrometry as the fully carbamylated peptide with retention of sulfation (**Figure 7.1, right panel**). Some evidence for desulfation was observed by monitoring the *m/z* values of the corresponding non-sulfated peptides, but these species were several orders of magnitude lower in abundance compared to the sulfated

carbamylation reaction products, as shown in **Figure 7.2** showing the base peak MS1 extracted ion chromatograms for sulfated and desulfated leu-enkephalin.

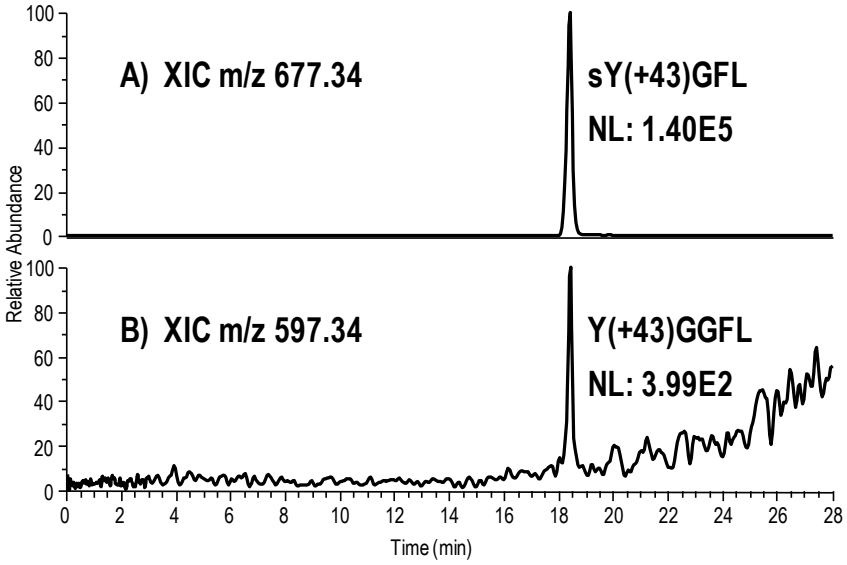


Figure 7.2 Extracted ion chromatograms showing carbamylated leu-enkephalin with (A) retained sulfate and (B) following sulfate loss.

A small amount of unreacted (i.e. non-carbamylated) peptide was detected for doubly sulfated cionin (**Figure 7.1C**), whereas all singly sulfated peptides (**Figure 7.1A-B, D**) were fully carbamylated. Incomplete carbamylation of cionin may be attributed to an increased probability of salt bridge formation between the positively charged peptide N-terminus and one of the two available negatively charged sulfates, thus blocking the α-amine and impeding carbamylation.[62] Carbamylation did not have a profound effect on the ionization efficiency of the sulfopeptides in the negative mode; however, the charge state distribution was shifted towards more negatively charged ions following reaction as shown in **Figure 7.3**. The UVPD fragmentation was not adversely affected by carbamylation because a mobile proton, as typically required for fragmentation of

protonated peptides upon collisional activation, is not required for dissociation of peptides energized by absorption of UV photons.

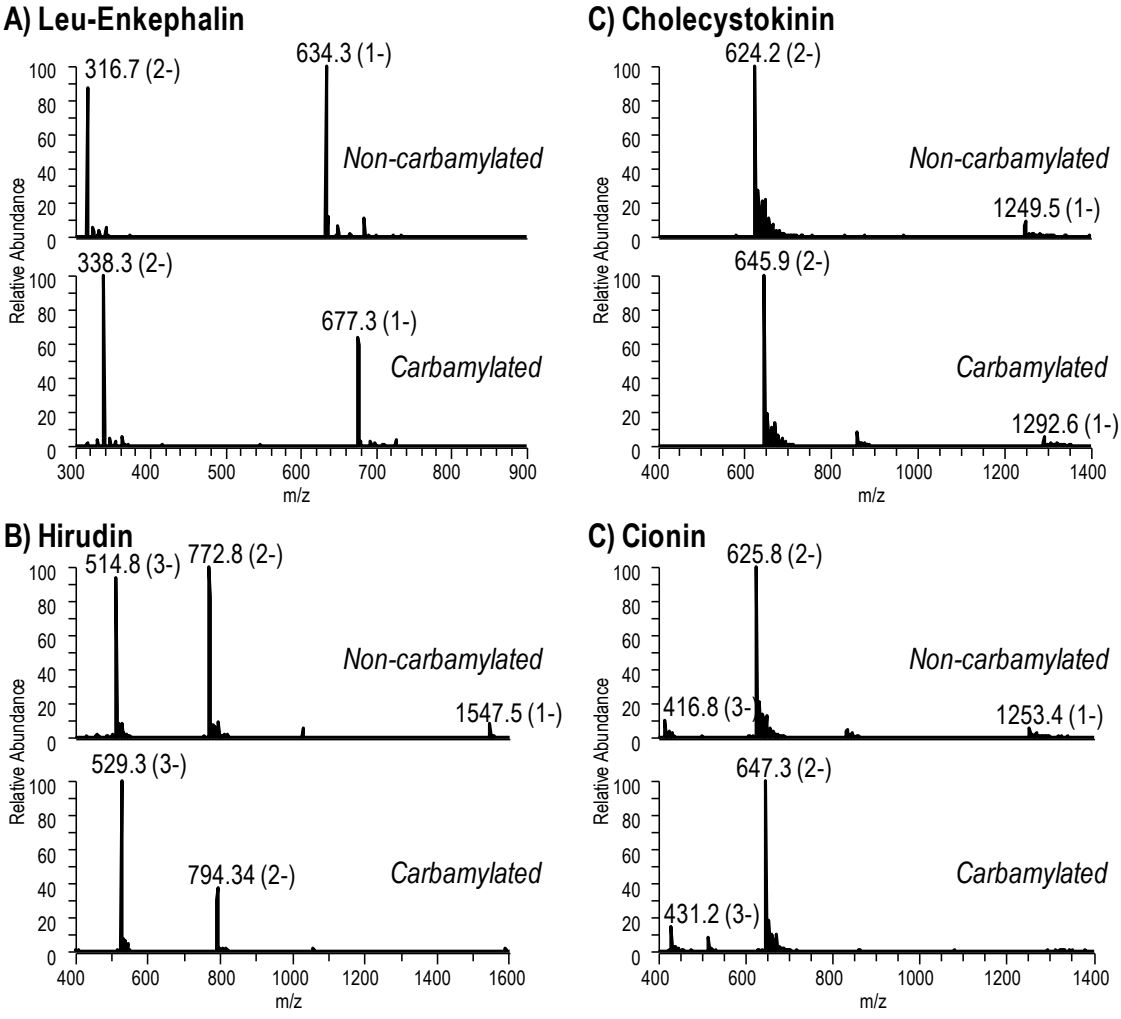


Figure 7.3 Negative ESI mass spectra for sulfopeptides before and after carbamylation

Figure 7.3 shows the UVPD spectra for doubly deprotonated hirudin acquired before and after carbamylation. The same fragmentation pattern, composed of mostly *a* and *x* type

ions, was observed for both peptides with the exception of product ions containing the site of carbamylation which were mass shifted by 43 Da in the spectrum of the modified peptide.

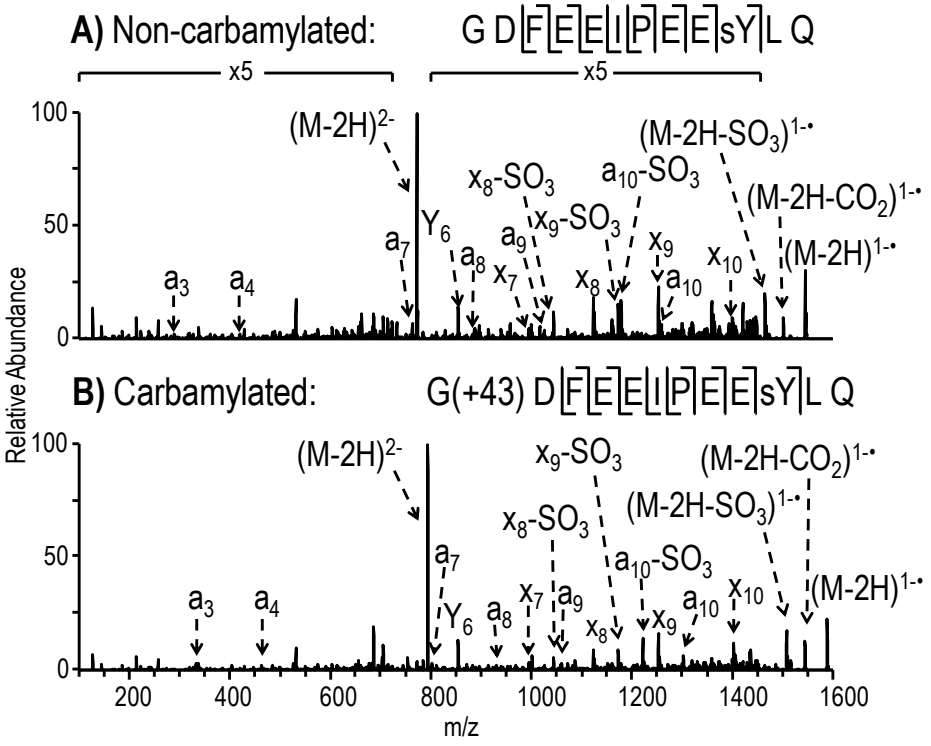


Figure 7.4 UVPD using 2 pulses at 2 mJ for (A) non-carbamylated and (B) carbamylated cholecystokinin in the 2- charge state at *m/z* 772.8 and 794.3, respectively. The carbamylation site is designated as +43 at the N-terminus of the peptide sequence.

7.4.2 Weak Anion Exchanges of Sulfopeptides

To determine the impact of carbamylation on enhancing sulfopeptide binding in weak anion exchange, a mixture of non-carbamylated sulfopeptides was first fractionated by

WAX to provide a baseline for subsequent comparisons. A mixture of cionin, cholecystokinin, Leu-enkephalin, hirudin, and caerulein (a mono-sulfated peptide that does not contain any primary amines) was loaded onto a WAX column and eluted by sequential salt steps of increasing ammonium acetate concentration ranging from 50 mM up to 4 M. LC-MS/MS was used to track the sulfopeptides in the resulting WAX fractions, and elution profiles were constructed based on chromatographic peak areas for each peptide detected in each fraction. No sulfopeptides were collected in the unbound fraction and each mono-sulfated peptide was observed in multiple subsequent WAX fractions to yield a Gaussian-like elution profile (**Figure 7.5**).

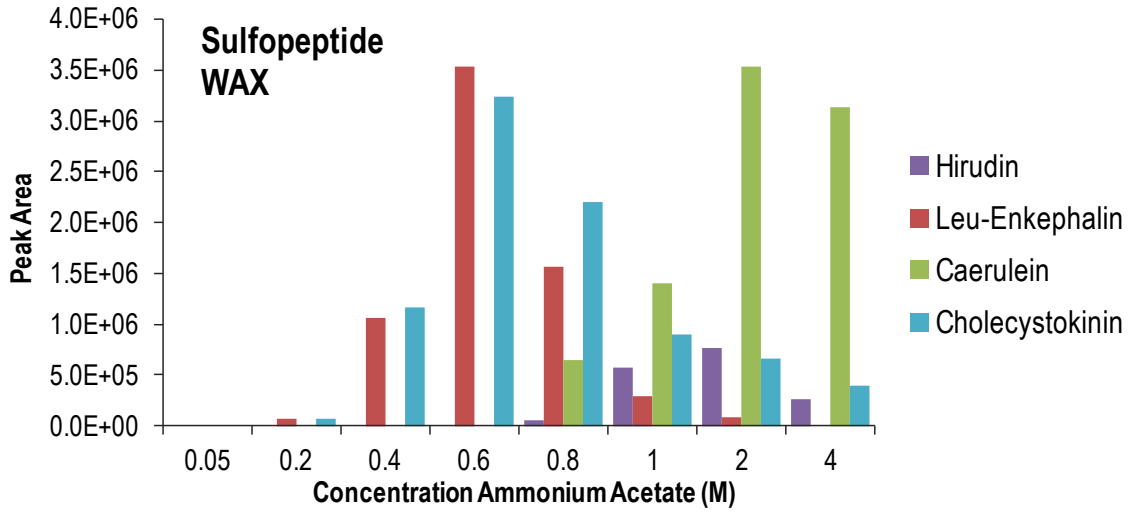


Figure 7.5 WAX elution profile of sulfopeptides using ammonium acetate eluents

Di-sulfated cionin was not recovered in any fraction despite using a high concentration of ammonium acetate (4 M) in the final elution, an outcome indicative of very strong binding to

the WAX resin. To improve the recovery of cionin and preemptively account for peptides in even higher sulfation states which may be present in biological samples, ammonium chloride (NH_4Cl) was used in place of ammonium acetate to introduce a stronger competing anion (Cl^-). Doing so greatly improved the recovery of cionin without the need to increase the salt concentration beyond 4 M, and thus NH_4Cl was used for all subsequent experiments.

With general conditions established for sulfopeptide elution from WAX, the five sulfated peptides were spiked into a mixture of tryptic peptides from BSA and fractionated by WAX both before and after carbamylation to evaluate (1) the shift in sulfopeptide retention after the neutralization of primary amines, and (2) the potential for sulfopeptide separation from a complex matrix of unmodified peptides. Including the 0.05 M NH_4Cl unbound fractions, seven additional WAX fractions were collected for each run by elution with 0.2 M, 0.4 M, 0.6 M, 0.8 M, 1 M, 2 M, and 4 M salt solutions. LC-MS analysis of the WAX fractions collected prior to carbamylation revealed differences in the elution profiles for each sulfopeptide that appeared to be linked to variations in the number of sulfate modifications as well as the number of basic sites (**Figure 7.6A-E,** blue bars). The elution order, in which Leu-enkephalin elutes first (**Figure 7.6A**) followed by cholecystokinin (**Figure 7.6B**), hirudin (**Figure 7.6C**), caerulein (**Figure 7.6E**), and lastly cionin (**Figure 7.6D**), was governed by a combination of attractive and repulsive interactions occurring between the WAX anion exchanger and the acidic and basic sites of the peptides, respectively, with more influence attributed to the overall peptide basicity.

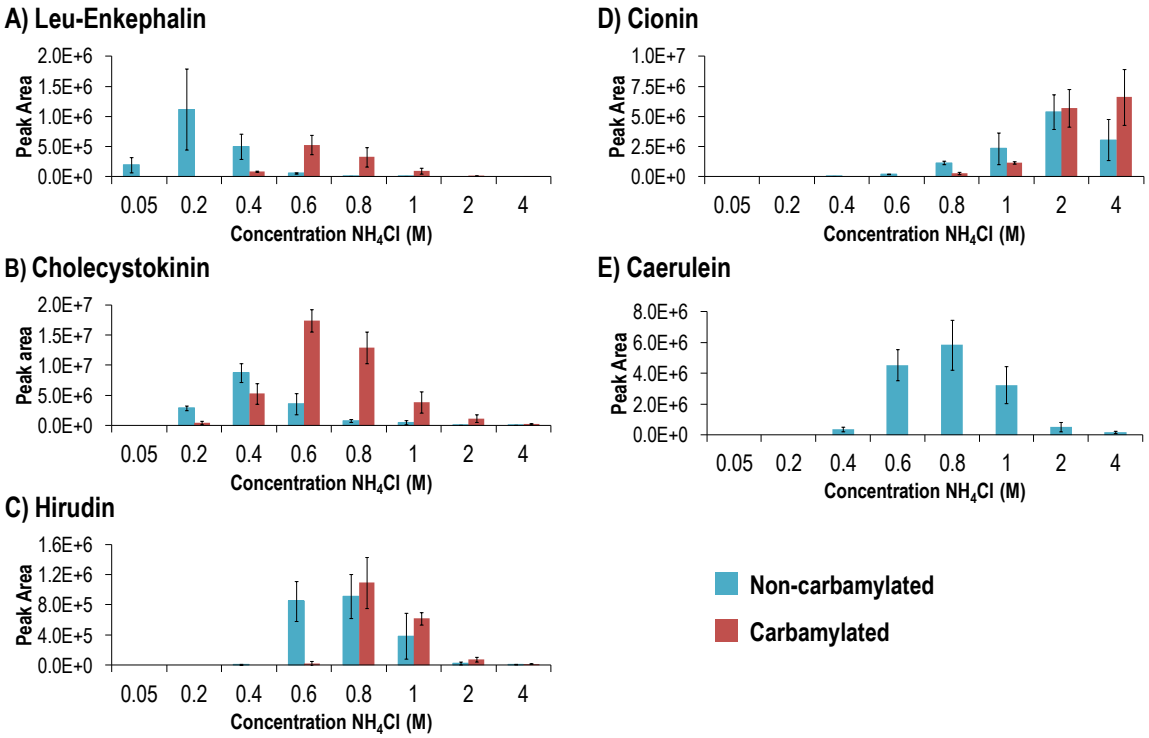


Figure 7.6 Sulfopeptide elution profiles from WAX using ammonium chloride eluents before and after carbamylation

This trend in the elution order is most apparent when comparing the elution of cholecystokinin (RDsYTGWNleDF-NH_2, **Figure 7.6B**) and caerulein (GlpQDsYTGWMDF-NH_2, **Figure 7.6E**) whose amino acid sequences are identical with the exception of an N-terminal arginine on the former and an N-terminal pyroglutamate/glutamine combination on the latter. These differences contribute to a decrease of up to two positive charges on caerulein and account for its significantly longer retention on WAX. In fact, caerulein exhibits the strongest binding of all mono-sulfated peptides despite having an uncharged amidated C-terminus and three fewer aspartic and glutamic acid residues compared to the most acidic peptide, hirudin. For peptides of equally basic character, such as Leu-enkephalin and hirudin,

a greater number of acidic residues lead to longer WAX retention. As expected, sulfo-tyrosine residues were more effective than glutamate or aspartate for strengthening the overall peptide binding to WAX, and the retention of doubly sulfated cionin exceeded the retention of singly sulfated hirudin despite hirudin having seven acidic sites compared to only four in cionin.

When the carbamylated sulfopeptide mixture was fractionated by WAX, the desired shift to greater retention was observed for all peptides (**Figure 7.6A-D,** red bars). Doubly sulfated cionin remained the most highly retained peptide, and the elution order of the mono-sulfated peptides was dictated by the number of acidic sites. Carbamylated hirudin replaced caerulein as the most strongly retained mono-sulfated peptide due to the greater number of aspartic and glutamic acid residues compared to caerulein. Cholecystokinin was the only peptide that retained basic sites after carbamylation, and thus was expected to take the place of Leu-enkephalin as the first peptide to elute. While the switch in elution order was not pronounced, the salt concentrations required for the elution of cholecystokinin and leu-enkephalin were essentially the same within the precision of the measurement, so the general trend of decreasing sulfopeptide retention as a function of basicity seems to hold true. The order of elution otherwise remained unchanged between underivatized and carbamylated sulfopeptides.. To further quantify the increase in WAX retention following carbamylation, the average salt concentration required for peptide elution before and after derivatization was determined by weighted average based on peptide peak areas across the WAX fractions (**Table 7.1**).

Average NH_4Cl concentration for elution (M)			
Peptide	**Non-carbamylated**	**Carbamylated**	**Delta**
Leu-enkephalin	0.26 (±0.03)	0.71 (±0.05)	0.44
Cholecystokinin	0.45 (±0.02)	0.73 (±0.05)	0.27
Hirudin	0.78 (±0.01)	0.94 (±0.04)	0.16
Cionin	2.14 (±0.13)	2.84 (±0.08)	0.70

Table 7.1 Average NH_4Cl concentration for sulfopeptides elution from WAX

Comparing the average NH_4Cl concentrations required for peptide elution revealed absolute molar increases of 0.44 M, 0.27 M, 0.16 M, and 0.70 M after carbamylation for Leu-enkephalin, cholecystokinin, hirudin, and cionin, respectively. In terms of the percent increase in salt concentration required for elution of the carbamylated peptides relative to elution of the underivatized peptides, the greatest shifts in WAX retention were observed for Leu-enkephalin, at 173%, and cholecystokinin, at 62%. Leu-enkephalin is the smallest of all the sulfopeptides with just five amino acids in its sequence, and thus single residue changes have a greater impact on the overall peptide retention characteristics on WAX. Carbamylation of Leu-enkephalin resulted in a complete loss of basicity which likely explains the very large shift in retention relative to underivatized Leu-enkephalin. All basic sites on hirudin and cionin were similarly converted to non-ionizable sites upon carbamylation, but these peptides experienced much lower percent increases in NH_4Cl concentration required for elution, likely owing to the significant acidity of these peptides which translated to fairly strong retention on WAX prior to carbamylation.

The efficiency of weak anion exchange for separating sulfopeptides from non-sulfated BSA peptides was next evaluated before and after carbamylation reaction. To visualize the separation, extracted ion chromatograms (XIC) showing the sulfopeptides of interest (red) were overlaid on the base peak MS1 chromatogram (blue) for each WAX fraction, as shown in **Figure 7.7**. For the non-carbamylated sample (**Figure 7.7,** left panel), the bulk of the BSA peptides eluted in the 0.05 M unbound and 0.2 M NH_4Cl fractions. Nearly all of the Leu-enkephalin loaded onto the WAX column co-eluted with the unmodified peptides in the 0.2 M fraction, and a significant amount of cholecystokinin also eluted in this fraction. Hirudin, caerulein, and cionin were effectively isolated in later fractions. After carbamylation, all peptides, including the non-sulfated BSA peptides which also underwent carbamylation at lysine residues and N-termini, were retained more strongly on the WAX column (**Figure 7.7,** right panel). Importantly, Leu-enkephalin and cholecystokinin which were both poorly separated from the unmodified peptides prior to carbamylation were much more effectively separated after reaction.

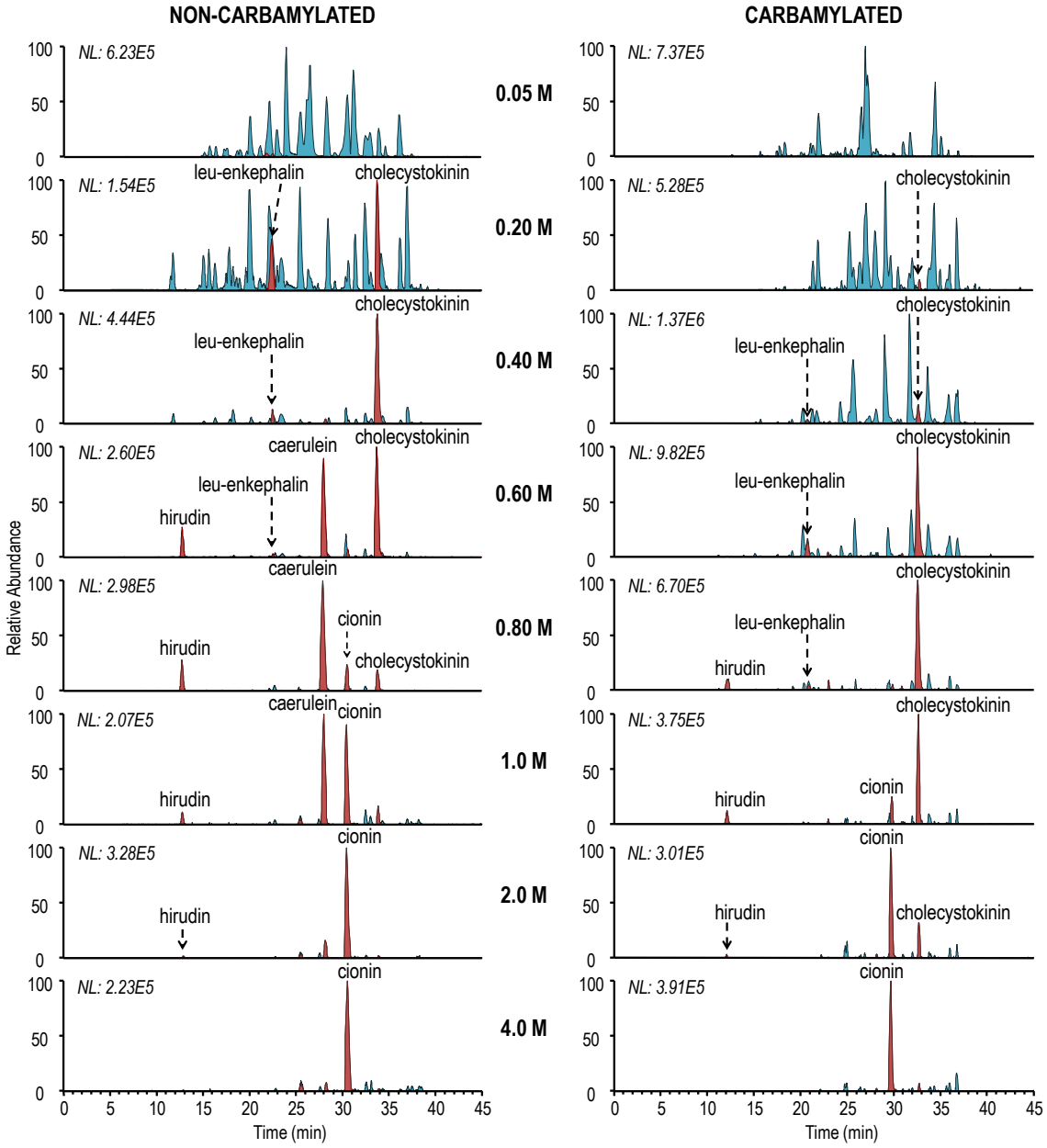


Figure 7.7 Base peak full MS chromatograms (blue) showing WAX fractions collected for a mixture of tryptic BSA and model sulfopeptides before (left) and after (right) carbamylation. XICs for sulfopeptides (red) are overlaid in each chromatogram.

7.4.3 Washing and Elution Optimization

To streamline the carbamylation/WAX strategy for the enrichment of sulfopeptides, washing and elution conditions were optimized. Ideally, washes would be performed using a single solution of NH_4Cl at a concentration which allows retention of carbamylated sulfopeptides while removing carbamylated non-sulfated peptides from the WAX column. For elution, conditions must be established that maximize sulfopeptide recovery. Initially, three 500 µL washes using 400 mM NH_4Cl and three 500 µL elutions using 4 M NH_4Cl were tested to separate sulfated and BSA peptides. Each wash and elution solution was collected separately and analyzed by LC-MS to evaluate each phase of washing and elution, with the ultimate goal of pooling all elution fractions together to maximize the signal in LC-MS. The chromatographic peak area of each sulfopeptide was tracked and then summed across all washes and all elutions to determine the percentage that was detected in the wash (blue bars) or elution (red bars) steps as shown in **Figure 7.8A**. Despite choosing a concentration of NH_4Cl that was well below the average NH_4Cl concentration required for the elution of each carbamylated sulfopeptide (**Table 7.1**), greater than 80% of cholecystokinin and Leu-enkephalin and about 50% of caerulein was prematurely eluted from WAX in the washing steps. The elution was also found to be sub-optimal based on the significant percentage of cionin and hirudin that was detected in the final elution relative to the first and second elutions which suggested that more peptides remained on the WAX column (**Figure 7.8B**). Two additional sets of conditions were tested which decreased the concentration of NH_4Cl used in the washes and increased the amount of NH_4Cl used for elution. **Figure 7.8C-D** shows the

results obtained using three washes with 300 mM NH_4Cl combined with four elutions using 5 M NH_4Cl. **Figure 7.8E-F** shows the results from three washes with 200 mM NH_4Cl and four elutions using 4 M NH_4Cl.

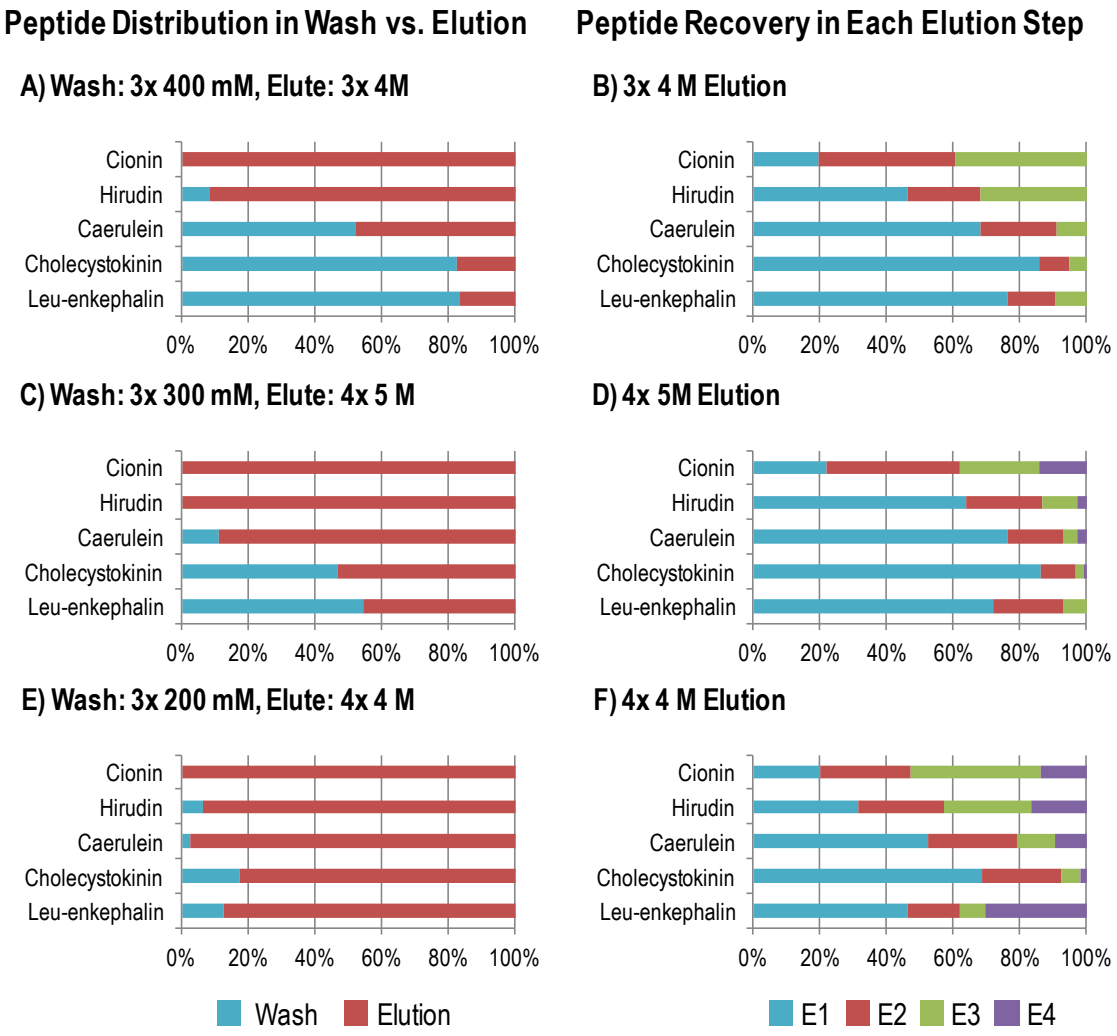


Figure 7.8 Wash and elution optimization for sulfopeptide enrichment using WAX. E1, E2, E3, and E4 refer to sequential elutions, where E1 is the first elution step.

By washing with 200 mM NH_4Cl, the best retention of carbamylated sulfopeptides was achieved while still removing the majority of the BSA background peptides. Four elutions using 5 M NH_4Cl appeared to optimize the peptide recovery based on the low percentage of each peptide in the final elution relative to the total peptide recovered in all elutions. In comparison to the initial results from WAX fractionation, the increase in NH_4Cl concentration required for the final elution was expected given the non-Gaussian elution profile of cionin in the WAX fractions in which the maximum peak area was detected in the final, 4 M NH_4Cl fraction (**Figure 7.6D**). The lower than expected retention of sulfopeptides during the wash optimization is more difficult to rationalize. Perhaps when multiple washes are incorporated, the elution effect is additive which requires the use of lower overall salt concentration to retain the sulfopeptides of interest.

7.5 CONCLUSION

The enrichment of model sulfopeptides by weak anion exchange following carbamylation reaction was demonstrated. Carbamylation efficiently converted all primary amines to less basic carbamates without promoting loss of the labile sulfate moiety or altering the UVPD fragmentation efficiency. In WAX fractionation using NH_4Cl as the eluent, sulfopeptide binding was enhanced following carbamylation. Carbamylated peptides eluted at an average concentration of NH_4Cl ranging from 0.7 M for the weakest retained peptide, leu-enkephalin, to 2.8 M for the most strongly retained and most highly sulfated peptide, cionin. Non-sulfated, carbamylated BSA peptides eluted in earlier WAX fractions, thus

demonstrating the potential of the WAX/carbamylation strategy for sulfopeptide enrichment. Multiple washes and elutions at static salt thresholds were evaluated for sulfopeptides enrichment from BSA in place of stepwise fractionation by increasing salt concentration. Sulfopeptides were optimally isolated using three washes at 200 mM NH_4Cl together with four elutions at 5 M NH_4Cl. These results provide a basis for sulfopeptide enrichment using carbamylation and WAX which may be applied for larger scale analysis of sulfo-tyrosine modifications in biological samples.

7.6 REFERENCES

(1) Moore, K. L. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106* (35), 14741–14742.
(2) Stone, M. J.; Chuang, S.; Hou, X.; Shoham, M.; Zhu, J. Z. *New Biotechnology* **2009**, *25* (5), 299–317.
(3) Yuh-Shyong Yang; Chen-Chu Wang; Bo-Han Chen; You-Hua Hou; Kuo-Sheng Hung; Yi-Chih Mao. *Molecules* **2015**, *20* (2), 2138–2164.
(4) Ouyang, Y.-B.; Crawley, J. T. B.; Aston, C. E.; Moore, K. L. *J. Biol. Chem.* **2002**, *277* (26), 23781–23787.
(5) Borghei, A.; Ouyang, Y.-B.; Westmuckett, A. D.; Marcello, M. R.; Landel, C. P.; Evans, J. P.; Moore, K. L. *J. Biol. Chem.* **2006**, *281* (14), 9423–9431.
(6) Westmuckett, A. D.; Hoffhines, A. J.; Borghei, A.; Moore, K. L. *General and Comparative Endocrinology* **2008**, *156* (1), 145–153.
(7) Bettelheim, F. R. *J. Am. Chem. Soc.* **1954**, *76* (10), 2838–2839.
(8) Hortin, G. *Blood* **1990**, *76* (5), 946–952.
(9) Michnick, D. A.; Pittman, D. D.; Wise, R. J.; Kaufman, R. J. *Journal of Biological Chemistry* **1994**, *269* (31), 20095–20102.
(10) Arruda, V. R. *Blood* **2001**, *97* (1), 130–138.
(11) Sako, D.; Comess, K. M.; Barone, K. M.; Camphausen, R. T.; Cumming, D. A.; Shaw, G. D. *Cell* **1995**, *83* (2), 323–331.
(12) Bonomi, M.; Busnelli, M.; Persani, L.; Vassart, G.; Costagliola, S. *Mol. Endocrinol.* **2006**, *20* (12), 3351–3363.
(13) Ludeman, J. P.; Stone, M. J. *British Journal of Pharmacology* **2014**, *171* (5), 1167–1179.
(14) Hsu, W.; Rosenquist, G. L.; Ansari, A. A.; Gershwin, M. E. *Autoimmunity Reviews* **2005**, *4* (7), 429–435.
(15) Farzan, M.; Mirzabekov, T.; Kolchinsky, P.; Wyatt, R.; Cayabyab, M.; Gerard, N. P.; Gerard, C.; Sodroski, J.; Choe, H. *Cell* **1999**, *96* (5), 667–676.
(16) Choe, H.; Moore, M. J.; Owens, C. M.; Wright, P. L.; Vasilieva, N.; Li, W.; Singh, A. P.; Shakri, R.; Chitnis, C. E.; Farzan, M. *Mol. Microbiol.* **2005**, *55* (5), 1413–1422.
(17) Riley, N. M.; Coon, J. J. *Anal. Chem.* **2015**.
(18) Choudhary, C.; Weinert, B. T.; Nishida, Y.; Verdin, E.; Mann, M. *Nature Reviews Molecular Cell Biology* **2014**, *15* (8), 536–550.
(19) Budnik, B. A.; Lee, R. S.; Steen, J. A. J. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2006**, *1764* (12), 1870–1880.
(20) Monigatti, F.; Hekking, B.; Steen, H. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2006**, *1764* (12), 1904–1913.
(21) Seibert, C.; Sakmar, T. P. *Biopolymers* **2008**, *90* (3), 459–477.
(22) Nemeth-Cawley, J. F.; Karnik, S.; Rouse, J. C. *J. Mass Spectrom.* **2001**, *36* (12), 1301–1311.

(23) Wolfender, J.-L.; Chu, F.; Ball, H.; Wolfender, F.; Fainzilber, M.; Baldwin, M. A.; Burlingame, A. L. *J. Mass Spectrom.* **1999**, *34* (4), 447–454.
(24) Medzihradszky, K. F.; Guan, S.; Maltby, D. A.; Burlingame, A. L. *Journal of the American Society for Mass Spectrometry* **2007**, *18* (9), 1617–1624.
(25) Yu, Y.; Hoffhines, A. J.; Moore, K. L.; Leary, J. A. *Nat Meth* **2007**, *4* (7), 583–588.
(26) Kim, J.-S.; Song, S.-U.; Kim, H.-J. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (11), 1916–1925.
(27) Cook, S.; Jackson, G. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (6), 1088–1099.
(28) Hersberger, K. E.; Håkansson, K. *Anal. Chem.* **2012**, *84* (15), 6370–6377.
(29) Robinson, M. R.; Moore, K. L.; Brodbelt, J. S. *Journal of The American Society for Mass Spectrometry* **2014**, *25* (8), 1461–1471.
(30) Han, S.-W.; Lee, S.-W.; Bahar, O.; Schwessinger, B.; Robinson, M. R.; Shaw, J. B.; Madsen, J. A.; Brodbelt, J. S.; Ronald, P. C. *Nat Commun* **2012**, *3*, 1153.
(31) Pruitt, R. N.; Schwessinger, B.; Joe, A.; Thomas, N.; Liu, F.; Albert, M.; Robinson, M. R.; Chan, L. J. G.; Luu, D. D.; Chen, H.; Bahar, O.; Daudi, A.; De Vleesschauwer, D.; Caddell, D.; Zhang, W.; Zhao, X.; Li, X.; Heazlewood, J. L.; Ruan, D.; Majumder, D.; Chern, M.; Kalbacher, H.; Midha, S.; Patil, P. B.; Sonti, R. V.; Petzold, C. J.; Liu, C. C.; Brodbelt, J. S.; Felix, G.; Ronald, P. C. *Science Advances* **2015**, *1* (6), e1500245–e1500245.
(32) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *Journal of Proteome Research* **2010**, *9* (8), 4205–4214.
(33) Madsen, J. A.; Xu, H.; Robinson, M. R.; Horton, A. P.; Shaw, J. B.; Giles, D. K.; Kaoud, T. S.; Dalby, K. N.; Trent, M. S.; Brodbelt, J. S. *Molecular & Cellular Proteomics* **2013**, *12* (9), 2604–2614.
(34) Greer, S. M.; Parker, W. R.; Brodbelt, J. S. *J. Proteome Res.* **2015**, *14* (6), 2626–2632.
(35) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *Proceedings of the National Academy of Sciences* **2004**, *101* (33), 12130–12135.
(36) Lim, K. B.; Kassel, D. B. *Analytical Biochemistry* **2006**, *354* (2), 213–219.
(37) Hennrich, M. L.; van den Toorn, H. W. P.; Groenewold, V.; Heck, A. J. R.; Mohammed, S. *Analytical Chemistry* **2012**, *84* (4), 1804–1808.
(38) Thingholm, T. E.; Jensen, O. N. In *Phospho-Proteomics*; Graauw, M. de, Ed.; Walker, J. M., Series Ed.; Humana Press: Totowa, NJ, 2009; Vol. 527, pp 47–56.
(39) Posewitz, M. C.; Tempst, P. *Analytical Chemistry* **1999**, *71* (14), 2883–2892.
(40) Zhou, H.; Low, T. Y.; Hennrich, M. L.; van der Toorn, H.; Schwend, T.; Zou, H.; Mohammed, S.; Heck, A. J. R. *Molecular & Cellular Proteomics* **2011**, *10* (10), M110.006452–M110.006452.
(41) Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R. *Analytical Chemistry* **2004**, *76* (14), 3935–3943.
(42) Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jørgensen, T. J. D. *Mol. Cell Proteomics* **2005**, *4* (7), 873–886.

(43) Thingholm, T. E.; Jørgensen, T. J. D.; Jensen, O. N.; Larsen, M. R. *Nat Protoc* **2006**, *1* (4), 1929–1935.
(44) Balsved, D.; Bundgaard, J. R.; Sen, J. W. *Analytical Biochemistry* **2007**, *363* (1), 70–76.
(45) Balderrama, G. D.; Meneses, E. P.; Orihuela, L. H.; Hernández, O. V.; Franco, R. C.; Robles, V. P.; Batista, C. V. F. *Rapid Commun. Mass Spectrom.* **2011**, *25* (8), 1017–1027.
(46) Kehoe, J. W.; Velappan, N.; Walbolt, M.; Rasmussen, J.; King, D.; Lou, J.; Knopp, K.; Pavlik, P.; Marks, J. D.; Bertozzi, C. R.; Bradbury, A. R. M. *Molecular & Cellular Proteomics* **2006**, *5* (12), 2350–2363.
(47) Hoffhines, A. J.; Damoc, E.; Bridges, K. G.; Leary, J. A.; Moore, K. L. *Journal of Biological Chemistry* **2006**, *281* (49), 37877–37887.
(48) Andersson, L. I. *Journal of Chromatography B: Biomedical Sciences and Applications* **2000**, *745* (1), 3–13.
(49) Helling, S.; Shinde, S.; Brosseron, F.; Schnabel, A.; Müller, T.; Meyer, H. E.; Marcus, K.; Sellergren, B. *Analytical Chemistry* **2011**, *83* (5), 1862–1865.
(50) Shinde, S.; Bunschoten, A.; Kruijtzer, J. A. W.; Liskamp, R. M. J.; Sellergren, B. *Angewandte Chemie International Edition* **2012**, *51* (33), 8326–8329.
(51) Amano, Y.; Shinohara, H.; Sakagami, Y.; Matsubayashi, Y. *Analytical Biochemistry* **2005**, *346* (1), 124–131.
(52) Toyoda, M.; Narimatsu, H.; Kameyama, A. *Analytical Chemistry* **2009**, *81* (15), 6140–6147.
(53) Wei, W.; Miller, R. L.; Leary, J. A. *Analytical Chemistry* **2013**, *85* (12), 5917–5923.
(54) Sabol, J. K.; Wei, W.; López-Hoyos, M.; Seo, Y.; Andaya, A.; Leary, J. A. *Matrix Biology* **2014**, *40*, 54–61.
(55) Angel, P. M.; Orlando, R. *Rapid Communications in Mass Spectrometry* **2007**, *21* (10), 1623–1634.
(56) Greer, S. M.; Cannon, J. R.; Brodbelt, J. S. *Anal. Chem.* **2014**, *86* (24), 12285–12290.
(57) Robotham, S. A.; Horton, A. P.; Cannon, J. R.; Cotham, V. C.; Marcotte, E. M.; Brodbelt, J. S. *Anal. Chem.* **2016**.
(58) Rappsilber, J.; Mann, M.; Ishihama, Y. *Nat Protoc* **2007**, *2* (8), 1896–1906.
(59) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Analytical Chemistry* **2008**, *80* (13), 4807–4819.
(60) Subramonian, S.; Clifford, D. *Journal of Solution Chemistry* **1989**, *18* (6), 529–543.
(61) Amano, Y.; Shinohara, H.; Sakagami, Y.; Matsubayashi, Y. *Analytical Biochemistry* **2005**, *346* (1), 124–131.
(62) Yagami, T.; Kitagawa, K.; Aida, C.; Fujiwara, H.; Futaki, S. *The Journal of Peptide Research* **2000**, *56* (4), 239–249.

Chapter 8

Conclusions

Significant advances have been made in the field of proteomics, yet the analysis of post-translational modifications remains a challenging task. Hundreds of PTMs exist that are responsible for maintaining normal cellular health and function; thus effective PTM characterization is necessary to attain a comprehensive view of the proteome. Abnormalities in PTM expression have been linked to a number of diseases, including cancer and autoimmune disorders, which has spurred the development of methods for increasingly deeper PTM profiling to aid in the discovery of potentially useful biomarkers or therapeutic targets. The low abundance and often transient lifetime of PTMs in biological systems present a significant analytical challenge. Advancements in sample handling and enrichment methods have helped to mitigate these problems. In contrast to the myriad of PTM specific techniques used for protein processing prior to mass spectrometry analysis, tandem MS is still largely restricted to conventional collision induced dissociation methods. This one-size-fits-all approach to peptide sequencing is problematic for PTM analysis based on the labile chemistries of many important PTMs, including phosphorylation and sulfation. The research in this dissertation sought to address the need for alternative activation methods by developing ultraviolet photodissociation for the characterization of PTMs.

In chapter 3, a *de novo* sequencing method was developed which biased fragmentation for the generation of N-terminal ions. This was accomplished by applying Lys-N proteolysis and imidazolinylation of the resulting N-terminal lysine ε-amine prior to LC-MS-UVPD. The resulting mass spectra were composed primarily of *a*, *b*, and *c* "golden triplet" ions while C-terminal ions were much sparser owing to the effective

capture of protons at the N-terminus. For cleavage at the same residue position, the mass shift between *a*, *b*, and *c* ions is static, and thus these ions can be easily distinguished and used for *de novo* sequencing.

De novo sequencing methods offer significant flexibility for PTM discovery but lack the sensitivity of more focused analysis that is directed towards a single or small number of PTMs. Chapters 4-7 sought to develop UVPD-MS methods specifically for the analysis of two labile PTMs, phosphorylation and sulfation.

In chapter 4 UVPD was applied for the analysis of phosphopeptides, and its utility for the analysis of peptide anions as well as cations was explored. Negative mode provided superior characterization of highly phosphorylated peptides from alpha and beta casein and also exhibited the greatest phosphate retention on product ions, but the overall sensitivity was low. Positive mode UVPD provided much greater sensitivity compared to negative mode for peptides in lower phosphorylation states, and phosphate neutral loss was lessened but not entirely eliminated. For phosphoproteomic analysis of HeLa and HCC70 cell lysates in positive ion mode, more peptides and proteins were identified using HCD compared to UVPD, but phosphate was better retained on UVPD product ions, suggesting that UVPD can be applied to improve phospho-site localization.

Following the success of photodissociation for large scale phosphopeptide analysis is chapter 4, UVPD was next applied in chapter 5 to more specifically map phosphorylation sites within the C-terminal domain (CTD) of RNA polymerase II. Because phosphorylation extends along the entire length of the CTD, very high sequence coverage at the protein level from bottom-up LC-MS/MS is required for complete characterization of all phosphorylation sites. To this end, alternative proteolysis using proteinase K or chymotrypsin was used to effectively digest the unconventional CTD, which consists of a species specific number of repeats of the consensus amino acid

sequence YSPTSPS. UVPD analysis of the resulting CTD peptides provided ample peptide level sequence coverage to allow phosphate localization despite the presence of many phosphate accepting sites within each heptad. Phosphorylation on Ser2 and Ser5 was favored in CTDs from two different species, yeast and fruit fly, and using two different kinases, TFIIH and ERK2.

UVPD was next evaluated for the analysis of tyrosine sulfation in chapter 6. Unlike phosphorylation, which may be characterized reasonably well using conventional methods such as CID and HCD, sulfation cannot be directly identified using these methods based on the extreme lability of the sulfo-ester. In fact, analysis in the negative ion mode is required to stabilize sulfate PTMs. For this reason, UVPD was a natural fit based on superior sequencing capabilities for peptide anions. Upon negative UVPD analysis, a rich array of sulfate retaining *a* and *x* type ions was produced which allowed peptide sequencing as well as sulfate localization. As further proof of concept, the negative mode LC-UVPD-MS strategy was successfully applied to identify the sulfated peptide of interest from bovine fibrinogen, a large 340 kDa heterohexamer plasma protein.

No mass spectrometry based global sulfation analysis has been reported primarily due to the lack of effective MS based characterization tools. Following the successful application of negative mode UVPD-MS for sulfopeptide analysis in chapter 6, efforts were next aimed at extending the method for sulfopeptide characterization on a more global scale in chapter 7. To do this, an enrichment method was devised which used weak anion exchange chromatography (WAX) and strategic modulation of peptide basicity. Carbamylation was used to neutralize all primary amines to less basic carbamates in sulfopeptides without knocking off the labile sulfate modification. Doing so improved sulfopeptide retention on WAX by the removal of interfering positive charges. The shift

to longer retention was more significant for sulfopeptides compared to carbamylated non-sulfated peptides which allowed their separation from a complex background of BSA.

While the research presented in this dissertation showcases the utility of UVPD for PTM analysis, several avenues for continued development remain. Although UVPD provides richer fragmentation in both the positive and negative ion modes, the fragmentation efficiency is lower compared to CID and HCD. Thus sensitivity is a significant limitation, especially for the analysis of PTMs which are already present in low abundance in biological samples. This problem may be addressed in a number of different ways. The attachment of aromatic chromophores has been demonstrated for improving the UVPD efficiency in positive mode, and applying similar derivatization to peptides carrying PTMs might provide the needed boost in sensitivity.[1] Alternatively, as UVPD becomes more widely adopted, new developments in instrument design will almost certainly improve UVPD performance for peptide sequencing.

In negative ion mode, the sensitivity problem is further exacerbated by low ionization efficiencies especially during the high aqueous phase of gradient LC separations. While a variety of different mobile phase systems were tested over the course of this dissertation, other alternative mobile phase additives may prove more effective for improving ionization. Chemical modification is another option, and carbamylation has been used to neutralize peptide positive charges and improve the ionization of peptide anions.[2] Aside from modifying the LC mobile phases or the analyte of interest, changes to the ionization source and other front end components of the mass spectrometer may be used to improve ionization efficiency.[3] Improvements in sensitivity would be invaluable for improving global PTM analysis in the negative mode, and will be necessary in order to scale up the WAX sulfopeptide enrichment method described in chapter 7 to a biologically relevant sample.

Finally, although the research presented in this dissertation was directed for the analysis of phosphorylation and sulfation, it is conceivable that UVPD-MS may be applied for the analysis of other PTMs with the greatest advantage expected for those that are labile and thus poorly characterized using conventional methods. Examples of other labile PTMs include s-nitrosylation,[4] arginine phosphorylation,[5] and histidine phosphorylation,[6] among others, which provides a rich landscape for the continued development of UVPD-MS.

8.1 REFERENCES

(1) Vasicek, L.; Brodbelt, J. S. *Analytical Chemistry* **2010**, *82* (22), 9441–9446.
(2) Greer, S. M.; Cannon, J. R.; Brodbelt, J. S. *Anal. Chem.* **2014**, *86* (24), 12285–12290.
(3) Tang, K.; Page, J. S.; Marginean, I.; Kelly, R. T.; Smith, R. D. *Journal of The American Society for Mass Spectrometry* **2011**, *22* (8), 1318–1325.
(4) Torta, F.; Elviri, L.; Bachi, A. In *Methods in Enzymology*; Elsevier, 2010; Vol. 473, pp 265–280.
(5) Schmidt, A.; Trentini, D. B.; Spiess, S.; Fuhrmann, J.; Ammerer, G.; Mechtler, K.; Clausen, T. *Molecular & Cellular Proteomics* **2014**, *13* (2), 537–550.
(6) Lapek, J. D.; Tombline, G.; Friedman, A. E. *Journal of Proteome Research* **2011**, *10* (2), 751–755.

References

(1) Domon, B.; Aebersold, R. *Science* **2006**, *312* (5771), 212–217.
(2) Cox, J.; Mann, M. *Annu. Rev. Biochem.* **2011**, *80* (1), 273–299.
(3) Bensimon, A.; Heck, A. J. R.; Aebersold, R. *Annu. Rev. Biochem.* **2012**, *81* (1), 379–405.
(4) Zubarev, R. A. *PROTEOMICS* **2013**, *13* (5), 723–726.
(5) Karve, T. M.; Cheema, A. K. *J. Amino Acids* **2011**, *2011*, 1–13.
(6) Mann, M.; Jensen, O. N. *Nat Biotech* **2003**, *21* (3), 255–261.
(7) Nørregaard Jensen, O. *Curr. Opin. Chem. Biol.* **2004**, *8* (1), 33–41.
(8) Reinders, J.; Lewandrowski, U.; Moebius, J.; Wagner, Y.; Sickmann, A. *PROTEOMICS* **2004**, *4* (12), 3686–3703.
(9) Silva, A. M. N.; Vitorino, R.; Domingues, M. R. M.; Spickett, C. M.; Domingues, P. *Free Radic. Biol. Med.* **2013**, *65*, 925–941.
(10) Olsen, J. V.; Mann, M. *Mol. Cell. Proteomics* **2013**, *12* (12), 3444–3452.
(11) Walsh, C. *Posttranslational modification of proteins: expanding nature's inventory*; Roberts and Co. Publishers: Englewood, Colo, 2006.
(12) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. *Sci. Rep.* **2011**, *1*.
(13) Jensen, O. N. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (6), 391–403.
(14) Webb, K.; Bennett, E. J. *Nat. Methods* **2013**, *10* (7), 620–621.
(15) Huang, Y.; Xu, B.; Zhou, X.; Li, Y.; Lu, M.; Jiang, R.; Li, T. *Mol. Cell. Proteomics* **2015**, *14* (3), 761–770.
(16) Gajadhar, A. S.; White, F. M. *Curr. Opin. Biotechnol.* **2014**, *28*, 83–87.
(17) Wang, G. G.; Allis, C. D.; Chi, P. *Trends Mol. Med.* **2007**, *13* (9), 363–372.
(18) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. *Cell* **2006**, *127* (3), 635–648.
(19) Moore, K. L. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (35), 14741–14742.
(20) Monigatti, F.; Hekking, B.; Steen, H. *Posttranslational Modif. Proteomics* **2006**, *1764* (12), 1904–1913.
(21) Riley, N. M.; Coon, J. J. *Anal. Chem.* **2015**.
(22) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
(23) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20* (18), 3551–3567.
(24) Xu, H.; Freitas, M. A. *PROTEOMICS* **2009**, *9* (6), 1548–1555.
(25) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. *J. Proteome Res.* **2004**, *3* (5), 958–964.
(26) Craig, R.; Cortens, J. P.; Beavis, R. C. *J. Proteome Res.* **2004**, *3* (6), 1234–1242.
(27) Bern, M.; Kil, Y. J.; Becker, C. In *Current Protocols in Bioinformatics*; Baxevanis, A. D., Petsko, G. A., Stein, L. D., Stormo, G. D., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2012.

(28) Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26* (12), 1367–1372.
(29) Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. *J. Proteome Res.* **2014**, *13* (8), 3679–3684.
(30) Yates, J. R. *J. Am. Chem. Soc.* **2013**, *135* (5), 1629–1640.
(31) Richards, A. L.; Merrill, A. E.; Coon, J. J. *Curr. Opin. Chem. Biol.* **2015**, *24*, 11–17.
(32) Mayne, J.; Ning, Z.; Zhang, X.; Starr, A. E.; Chen, R.; Deeke, S.; Chiang, C.-K.; Xu, B.; Wen, M.; Cheng, K.; Seebun, D.; Star, A.; Moore, J. I.; Figeys, D. *Anal. Chem.* **2016**, *88* (1), 95–121.
(33) Hardman, M.; Makarov, A. A. *Anal. Chem.* **2003**, *75* (7), 1699–1705.
(34) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *J. Mass Spectrom.* **2005**, *40* (4), 430–443.
(35) Mann, M.; Kelleher, N. L. *Proc. Natl. Acad. Sci.* **2008**, *105* (47), 18132–18138.
(36) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. *Mol. Cell. Proteomics* **2014**, *13* (12), 3698–3708.
(37) Senko, M. W.; Remes, P. M.; Canterbury, J. D.; Mathur, R.; Song, Q.; Eliuk, S. M.; Mullen, C.; Earley, L.; Hardman, M.; Blethrow, J. D.; Bui, H.; Specht, A.; Lange, O.; Denisov, E.; Makarov, A.; Horning, S.; Zabrouskov, V. *Anal. Chem.* **2013**, *85* (24), 11710–11714.
(38) Beck, S.; Michalski, A.; Raether, O.; Lubeck, M.; Kaspar, S.; Goedecke, N.; Baessmann, C.; Hornburg, D.; Meier, F.; Paron, I.; Kulak, N. A.; Cox, J.; Mann, M. *Mol. Cell. Proteomics* **2015**, *14* (7), 2014–2029.
(39) Richards, A. L.; Hebert, A. S.; Ulbrich, A.; Bailey, D. J.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. *Nat Protoc.* **2015**, *10* (5), 701–714.
(40) Tyanova, S.; Albrechtsen, R.; Kronqvist, P.; Cox, J.; Mann, M.; Geiger, T. *Nat. Commun.* **2016**, *7*, 10259.
(41) Thingholm, T. E.; Jensen, O. N.; Larsen, M. R. *PROTEOMICS* **2009**, *9* (6), 1451–1468.
(42) Macek, B.; Mann, M.; Olsen, J. V. *Annu. Rev. Pharmacol. Toxicol.* **2009**, *49* (1), 199–221.
(43) Ongay, S.; Boichenko, A.; Govorukhina, N.; Bischoff, R. *J. Sep. Sci.* **2012**, *35* (18), 2341–2372.
(44) Thaysen-Andersen, M.; Packer, N. H. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2014**, *1844* (9), 1437–1452.
(45) Zhang, K.; Tian, S.; Fan, E. *The Analyst* **2013**, *138* (6), 1628.
(46) Mertins, P.; Qiao, J. W.; Patel, J.; Udeshi, N. D.; Clauser, K. R.; Mani, D. R.; Burgess, M. W.; Gillette, M. A.; Jaffe, J. D.; Carr, S. A. *Nat. Methods* **2013**, *10* (7), 634–637.
(47) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. *Rapid Commun. Mass Spectrom.* **2000**, *14* (19), 1793–1800.
(48) Sweet, S. M. M.; Bailey, C. M.; Cunningham, D. L.; Heath, J. K.; Cooper, H. J. *Mol. Cell. Proteomics* **2009**, *8* (5), 904–912.

(49) Shi, S. D.-H.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W. *Anal. Chem.* **2001**, *73* (1), 19–22.
(50) Molina, H.; Horn, D. M.; Tang, N.; Mathivanan, S.; Pandey, A. *Proc. Natl. Acad. Sci.* **2007**, *104* (7), 2199–2204.
(51) Creese, A.; Cooper, H. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (9), 1263–1274.
(52) Hogan, J. M.; Pitteri, S. J.; Chrisman, P. A.; McLuckey, S. A. *J. Proteome Res.* **2005**, *4* (2), 628–632.
(53) Zhu, Z.; Su, X.; Clark, D. F.; Go, E. P.; Desaire, H. *Anal. Chem.* **2013**, *85* (17), 8403–8411.
(54) Villén, J.; Beausoleil, S. A.; Gygi, S. P. *PROTEOMICS* **2008**, *8* (21), 4444–4452.
(55) Ulintz, P. J.; Yocum, A. K.; Bodenmiller, B.; Aebersold, R.; Andrews, P. C.; Nesvizhskii, A. I. *J. Proteome Res.* **2009**, *8* (2), 887–899.
(56) Saba, J.; Dutta, S.; Hemenway, E.; Viner, R. *Int. J. Proteomics* **2012**, *2012*, 1–7.
(57) Wu, S.-W.; Pu, T.-H.; Viner, R.; Khoo, K.-H. *Anal. Chem.* **2014**, *86* (11), 5478–5486.
(58) Beausoleil, S. A.; Villén, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. *Nat. Biotechnol.* **2006**, *24* (10), 1285–1292.
(59) Bailey, C. M.; Sweet, S. M. M.; Cunningham, D. L.; Zeller, M.; Heath, J. K.; Cooper, H. J. *J. Proteome Res.* **2009**, *8* (4), 1965–1971.
(60) Taus, T.; Köcher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. *J. Proteome Res.* **2011**, *10* (12), 5354–5362.
(61) Fermin, D.; Walmsley, S. J.; Gingras, A.-C.; Choi, H.; Nesvizhskii, A. I. *Mol. Cell. Proteomics* **2013**, *12* (11), 3409–3419.
(62) Dallas, D. C.; Martin, W. F.; Hua, S.; German, J. B. *Brief. Bioinform.* **2013**, *14* (3), 361–374.
(63) Monigatti, F.; Gasteiger, E.; Bairoch, A.; Jung, E. *Bioinformatics* **2002**, *18* (5), 769–770.
(64) Huang, S.-Y.; Shi, S.-P.; Qiu, J.-D.; Sun, X.-Y.; Suo, S.-B.; Liang, R.-P. *Anal. Biochem.* **2012**, *428* (1), 16–23.
(65) Ma, B.; Johnson, R. *Mol. Cell. Proteomics* **2012**, *11* (2), O111.014902–O111.014902.
(66) Hughes, C.; Ma, B.; Lajoie, G. A. In *Proteome Bioinformatics*; Hubbard, S. J., Jones, A. R., Eds.; Humana Press: Totowa, NJ, 2010; Vol. 604, pp 105–121.
(67) Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. *PROTEOMICS* **2010**, *10* (4), 634–649.
(68) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342.
(69) Frank, A.; Pevzner, P. *Anal. Chem.* **2005**, *77* (4), 964–973.
(70) Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J. M. *Anal. Chem.* **2005**, *77* (22), 7265–7273.
(71) Mo, L.; Dutta, D.; Wan, Y.; Chen, T. *Anal. Chem.* **2007**, *79* (13), 4870–4878.

(72) Pan, C.; Park, B. H.; McDonald, W. H.; Carey, P. A.; Banfield, J. F.; VerBerkmoes, N. C.; Hettich, R. L.; Samatova, N. F. *BMC Bioinformatics* **2010**, *11* (1), 118.
(73) Song, Y.; Yu, M. *Inf. Process. Lett.* **2015**, *115* (2), 377–381.
(74) Chi, H.; Chen, H.; He, K.; Wu, L.; Yang, B.; Sun, R.-X.; Liu, J.; Zeng, W.-F.; Song, C.-Q.; He, S.-M.; Dong, M.-Q. *J. Proteome Res.* **2013**, *12* (2), 615–625.
(75) Münchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem.* **2000**, *72* (17), 4047–4057.
(76) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. *Anal. Chem.* **2005**, *77* (19), 6300–6309.
(77) Hennrich, M.; Mohammed, S.; Altelaar, A.; Heck, A. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 1957–1965 – 1965.
(78) Cagney, G.; Emili, A. *Nat Biotech* **2002**, *20* (2), 163–170.
(79) Robotham, S. A.; Kluwe, C.; Cannon, J. R.; Ellington, A.; Brodbelt, J. S. *Anal. Chem.* **2013**, *85* (20), 9832–9838.
(80) Miyashita, M.; Hanai, Y.; Awane, H.; Yoshikawa, T.; Miyagawa, H. *Rapid Commun. Mass Spectrom.* **2011**, *25* (9), 1130–1140.
(81) Madsen, J. A.; Brodbelt, J. S. *Anal. Chem.* **2009**, *81* (9), 3645–3653.
(82) Roepstorff, P. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601.
(83) Johnson, R. S.; Martin, S. A.; Biemann, K.; Stults, J. T.; Watson, J. T. *Anal. Chem.* **1987**, *59* (21), 2621–2625.
(84) Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (1), 30–51.
(85) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci.* **2004**, *101* (26), 9528–9533.
(86) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.
(87) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72* (3), 563–573.
(88) Zubarev, R. A. *Curr. Opin. Biotechnol.* **2004**, *15* (1), 12–16.
(89) Brodbelt, J. S. *Chem. Soc. Rev.* **2014**, *43* (8), 2757–2783.
(90) Johnson, R. S.; Martin, S. A.; Biemann, K. *Int. J. Mass Spectrom. Ion Process.* **1988**, *86*, 137–154.
(91) Weidong Cui; Thompson, M. S.; Reilly, J. P. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (8), 1384–1398.
(92) McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **1992**, *3* (6), 599–614.
(93) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118* (35), 8365–8374.
(94) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35* (12), 1399–1406.
(95) Gu, C.; Tsaprailis, G.; Breci, L.; Wysocki, V. H. *Anal Chem* **2000**, *72* (23), 5804–5813.
(96) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. *Anal. Chem.* **2005**, *77* (18), 5800–5813.

(97) Breci, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H. *Anal. Chem.* **2003**, *75* (9), 1963–1971.
(98) Vaisar, T.; Urban, J. *J. Mass Spectrom.* **1996**, *31* (10), 1185–1187.
(99) Palumbo, A. M.; Smith, S. A.; Kalcic, C. L.; Dantus, M.; Stemmer, P. M.; Reid, G. E. *Mass Spectrom. Rev.* **2011**, *30* (4), 600–625.
(100) Brown, R.; Stuart, S.; Houel, S.; Ahn, N.; Old, W. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (7), 1128–1142.
(101) Nemeth-Cawley, J. F.; Karnik, S.; Rouse, J. C. *J. Mass Spectrom.* **2001**, *36* (12), 1301–1311.
(102) Wolfender, J.-L.; Chu, F.; Ball, H.; Wolfender, F.; Fainzilber, M.; Baldwin, M. A.; Burlingame, A. L. *J. Mass Spectrom.* **1999**, *34* (4), 447–454.
(103) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nat. Methods* **2007**, *4* (9), 709–712.
(104) McAlister, G. C.; Phanstiel, D. H.; Brumbaugh, J.; Westphall, M. S.; Coon, J. J. *Mol. Cell. Proteomics* **2011**, *10* (5), O111.009456–O111.009456.
(105) Henriksen, P.; Wagner, S. A.; Weinert, B. T.; Sharma, S.; Bacinskaja, G.; Rehman, M.; Juffer, A. H.; Walther, T. C.; Lisby, M.; Choudhary, C. *Mol. Cell. Proteomics* **2012**, *11* (11), 1510–1522.
(106) Nagaraj, N.; D'Souza, R. C. J.; Cox, J.; Olsen, J. V.; Mann, M. *J. Proteome Res.* **2010**, *9* (12), 6786–6794.
(107) Madsen, J. A.; Xu, H.; Robinson, M. R.; Horton, A. P.; Shaw, J. B.; Giles, D. K.; Kaoud, T. S.; Dalby, K. N.; Trent, M. S.; Brodbelt, J. S. *Mol. Cell. Proteomics* **2013**, *12* (9), 2604–2614.
(108) Waugh, R. J.; Bowie, J. H. *Rapid Commun. Mass Spectrom.* **1994**, *8* (2), 169–173.
(109) Han, S.-W.; Lee, S.-W.; Bahar, O.; Schwessinger, B.; Robinson, M. R.; Shaw, J. B.; Madsen, J. A.; Brodbelt, J. S.; Ronald, P. C. *Nat Commun* **2012**, *3*, 1153.
(110) Pruitt, R. N.; Schwessinger, B.; Joe, A.; Thomas, N.; Liu, F.; Albert, M.; Robinson, M. R.; Chan, L. J. G.; Luu, D. D.; Chen, H.; Bahar, O.; Daudi, A.; De Vleesschauwer, D.; Caddell, D.; Zhang, W.; Zhao, X.; Li, X.; Heazlewood, J. L.; Ruan, D.; Majumder, D.; Chern, M.; Kalbacher, H.; Midha, S.; Patil, P. B.; Sonti, R. V.; Petzold, C. J.; Liu, C. C.; Brodbelt, J. S.; Felix, G.; Ronald, P. C. *Sci. Adv.* **2015**, *1* (6), e1500245–e1500245.
(111) Madsen, J. A.; Ko, B. J.; Robotham, S. A.; Xu, H.; Horton, A. P.; Iwashkiw, J. A.; Shaw, J. B.; Feldman, M. F.; Brodbelt, J. S. *Anal Chem* **2013**, *85* (19), 9253–9261.
(112) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. *PROTEOMICS* **2011**, *11* (7), 1329–1334.
(113) Vasicek, L.; Ledvina, A.; Shaw, J.; Griep-Raming, J.; Westphall, M.; Coon, J.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (6), 1105–1108 – 1108.
(114) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. *Anal. Chem.* **2016**, *88* (4), 2303–2310.
(115) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (1), 1044–1051.

(116) Shaw, J. B.; Robinson, E. W.; Paša-Tolić, L. *Anal. Chem.* **2016**, *88* (6), 3019–3023.
(117) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. *Science* **1989**, *246* (4926), 64–71.
(118) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (6), 659–669.
(119) Makarov, A. *Anal. Chem.* **2000**, *72* (6), 1156–1162.
(120) Gardner, M. W.; Vasicek, L. A.; Shabbir, S.; Anslyn, E. V.; Brodbelt, J. S. *Anal. Chem.* **2008**, *80* (13), 4807–4819.
(121) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.
(122) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.
(123) Jackson, S. S.; Coughlin, E. E.; Coon, J. J.; Miyamoto, S. *Protein Expr. Purif.* **2013**, *92* (1), 48–53.
(124) Rappsilber, J.; Mann, M.; Ishihama, Y. *Nat. Protoc.* **2007**, *2* (8), 1896–1906.
(125) Xu, H.; Freitas, M. *BMC Bioinformatics* **2007**, *8* (1), 133.
(126) Xu, H.; Yang, L.; Freitas, M. *BMC Bioinformatics* **2008**, *9* (1), 347.
(127) Xu, H.; Freitas, M. A. *PROTEOMICS* **2009**, *9* (6), 1548–1555.
(128) Xu, H.; Freitas, M. A. *J. Proteome Res.* **2008**, *7* (7), 2605–2615.
(129) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21* (3), 255.
(130) Aguiar, M.; Haas, W.; Beausoleil, S. A.; Rush, J.; Gygi, S. P. *J Proteome Res* **2010**, *9* (6), 3103–3107.
(131) Taylor, J. A.; Johnson, R. S. *Anal. Chem.* **2001**, *73* (11), 2594–2604.
(132) Zhang, Z. *Anal. Chem.* **2004**, *76* (21), 6374–6383.
(133) Bern, M.; Goldberg, D. *J. Comput. Biol.* **2006**, *13* (2), 364–378.
(134) Qin, J.; Chait, B. T. *J Am Chem Soc* **1995**, *117* (19), 5411–5412.
(135) Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. *J Am Chem Soc* **1999**, *121* (22), 5142–5154.
(136) Reilly, J. P. *Mass Spectrom. Rev.* **2009**, *28* (3), 425–447.
(137) Brodbelt, J. S.; Wilson, J. J. *Mass Spectrom. Rev.* **2009**, *28* (3), 390–424.
(138) Ly, T.; Julian, R. R. *Angew. Chem. Int. Ed.* **2009**, *48* (39), 7130–7137.
(139) Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (2), 197–206.
(140) Brancia, F. L.; Montgomery, H.; Tanaka, K.; Kumashiro, S. *Anal. Chem.* **2004**, *76* (10), 2748–2755.
(141) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. *Proc. Natl. Acad. Sci.* **2000**, *97* (19), 10313–10317.
(142) Zhang, L.; Reilly, J. P. *Anal. Chem.* **2010**, *82* (3), 898–908.
(143) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proc. Natl. Acad. Sci.* **1999**, *96* (13), 7131–7136.
(144) Vasicek, L.; Wilson, J.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 377–384 – 384.

(145) Hohmann, L.; Sherwood, C.; Eastham, A.; Peterson, A.; Eng, J. K.; Eddes, J. S.; Shteynberg, D.; Martin, D. B. *J. Proteome Res.* **2009**, *8* (3), 1415–1422.
(146) Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. *Mol. Cell. Proteomics* **2009**, *8* (4), 650–660.
(147) Hennrich, M. L.; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J. R.; Mohammed, S. *Anal. Chem.* **2009**, *81* (18), 7814–7822.
(148) Carabetta, V.; Li, T.; Shakya, A.; Greco, T.; Cristea, I. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (6), 1050–1060 – 1060.
(149) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Anal Chem* **2006**, *79* (2), 477–485.
(150) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35* (12), 1399–1406.
(151) Zhang, L.; Reilly, J. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (7), 1378–1390 – 1390.
(152) Bringans, S.; Kendrick, T. S.; Lui, J.; Lipscombe, R. *Rapid Commun. Mass Spectrom.* **2008**, *22* (21), 3450–3454.
(153) Harsha, H. C.; Pandey, A. *Themat. Issue Oncoproteomics* **2010**, *4* (6), 482–495.
(154) Fang, B.; Haura, E. B.; Smalley, K. S.; Eschrich, S. A.; Koomen, J. M. *Target. Cancer Ther.* **2010**, *80* (5), 739–747.
(155) Zhou, H.; Di Palma, S.; Preisinger, C.; Peng, M.; Polat, A. N.; Heck, A. J. R.; Mohammed, S. *J. Proteome Res.* **2012**, *12* (1), 260–271.
(156) de Graaf, E. L.; Giansanti, P.; Altelaar, A. F. M.; Heck, A. J. R. *Mol. Cell. Proteomics* **2014**, *13* (9), 2426–2434.
(157) Kelstrup, C. D.; Jersie-Christensen, R. R.; Batth, T. S.; Arrey, T. N.; Kuehn, A.; Kellmann, M.; Olsen, J. V. *J. Proteome Res.* **2014**, *13* (12), 6187–6195.
(158) Batth, T. S.; Francavilla, C.; Olsen, J. V. *J. Proteome Res.* **2014**, *13* (12), 6176–6186.
(159) Brodbelt, J. S. *Anal. Chem.* **2016**, *88* (1), 30–51.
(160) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. *J. Mass Spectrom.* **2009**, *44* (6), 861–878.
(161) Wiesner, J.; Premsler, T.; Sickmann, A. *PROTEOMICS* **2008**, *8* (21), 4466–4483.
(162) Swaney, D. L.; Wenger, C. D.; Thomson, J. A.; Coon, J. J. *Proc. Natl. Acad. Sci.* **2009**, *106* (4), 995–1000.
(163) Flora, J. W.; Muddiman, D. C. *J. Am. Chem. Soc.* **2002**, *124* (23), 6546–6547.
(164) Flora, J. W.; Muddiman, D. C. *Anal. Chem.* **2001**, *73* (14), 3305–3311.
(165) Crowe, M.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (11), 1581–1592.
(166) Crowe, M. C.; Brodbelt, J. S. *Anal. Chem.* **2005**, *77* (17), 5726–5734.
(167) Kalcic, C. L.; Gunaratne, T. C.; Jones, A. D.; Dantus, M.; Reid, G. E. *J. Am. Chem. Soc.* **2009**, *131* (3), 940–942.
(168) Smith, S. A.; Kalcic, C. L.; Safran, K. A.; Stemmer, P. M.; Dantus, M.; Reid, G. E. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (12), 2031–2040.
(169) Smith, S. A.; Kalcic, C. L.; Cui, L.; Reid, G. E. *Rapid Commun. Mass Spectrom.* **2013**, *27* (24), 2807–2817.

(170) Lemoine, J.; Tabarin, T.; Antoine, R.; Broyer, M.; Dugourd, P. *Rapid Commun. Mass Spectrom.* **2006**, *20* (3), 507–511.
(171) Park, S.; Ahn, W.-K.; Lee, S.; Han, S. Y.; Rhee, B. K.; Oh, H. B. *Rapid Commun. Mass Spectrom.* **2009**, *23* (23), 3609–3620.
(172) Zhang, L.; Reilly, J. P. *J. Proteome Res.* **2010**, *9* (6), 3025–3034.
(173) Greer, S. M.; Cannon, J. R.; Brodbelt, J. S. *Anal. Chem.* **2014**, *86* (24), 12285–12290.
(174) Greer, S. M.; Parker, W. R.; Brodbelt, J. S. *J. Proteome Res.* **2015**, *14* (6), 2626–2632.
(175) Robinson, M. R.; Madsen, J. A.; Brodbelt, J. S. *Anal. Chem.* **2012**, *84* (5), 2433–2439.
(176) Holden, D. D.; Pruet, J. M.; Brodbelt, J. S. *Int. J. Mass Spectrom.* **2015**, *390*, 81–90.
(177) Shaw, J.; Madsen, J.; Xu, H.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (10), 1707–1715.
(178) Luo, Y.; Yogesha, S. D.; Cannon, J. R.; Yan, W.; Ellington, A. D.; Brodbelt, J. S.; Zhang, Y. *ACS Chem. Biol.* **2013**, *8* (9), 2042–2052.
(179) Robinson, M.; Moore, K.; Brodbelt, J. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (8), 1461–1471.
(180) Shin, Y.; Moon, J.; Kim, M. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (1), 53–59.
(181) Kim, T.-Y.; Reilly, J. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (12), 2334–2341.
(182) Madsen, J. A.; Cheng, R. R.; Kaoud, T. S.; Dalby, K. N.; Makarov, D. E.; Brodbelt, J. S. *Chem. – Eur. J.* **2012**, *18* (17), 5374–5383.
(183) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
(184) McAlister, G. C.; Russell, J. D.; Rumachik, N. G.; Hebert, A. S.; Syka, J. E. P.; Geer, L. Y.; Westphall, M. S.; Pagliarini, D. J.; Coon, J. J. *Anal. Chem.* **2012**, *84* (6), 2875–2882.
(185) Mamone, G.; Picariello, G.; Ferranti, P.; Addeo, F. *PROTEOMICS* **2010**, *10* (3), 380–393.
(186) Fonslow, B. R.; Niessen, S. M.; Singh, M.; Wong, C. C. L.; Xu, T.; Carvalho, P. C.; Choi, J.; Park, S. K.; Yates, J. R. *J. Proteome Res.* **2012**, *11* (5), 2697–2709.
(187) Cui, L.; Reid, G. E. *PROTEOMICS* **2013**, *13* (6), 964–973.
(188) Lanucara, F.; Chi Hoo Lee, D.; Eyers, C. E. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (2), 214–225.
(189) Gazdar, A. F.; Kurvari, V.; Virmani, A.; Gollahon, L.; Sakaguchi, M.; Westerfield, M.; Kodagoda, D.; Stasny, V.; Cunningham, H. T.; Wistuba, I. I.; Tomlinson, G.; Tonk, V.; Ashfaq, R.; Leitch, A. M.; Minna, J. D.; Shay, J. W. *Int. J. Cancer* **1998**, *78* (6), 766–774.
(190) Kim, M. S.; Kim, T.; Kong, S.-Y.; Kwon, S.; Bae, C. Y.; Choi, J.; Kim, C. H.; Lee, E. S.; Park, J.-K. *PLoS ONE* **2010**, *5* (5), e10441.
(191) Kim, M. S.; Kwon, S.; Kim, T.; Lee, E. S.; Park, J.-K. *Biomaterials* **2011**, *32* (5), 1396–1403.

(192) J, C. K.; V, G. S.; Stanley, L. *Breast Dis.* **2010**, No. 1,2, 35–48.
(193) Steinmetz, E. J.; Warren, C. L.; Kuehner, J. N.; Panbehi, B.; Ansari, A. Z.; Brow, D. A. *Mol. Cell* **2006**, *24* (5), 735–746.
(194) Koch, F.; Jourquin, F.; Ferrier, P.; Andrau, J.-C. *Trends Biochem. Sci.* **2008**, *33* (6), 265–273.
(195) Koch, F.; Fenouil, R.; Gut, M.; Cauchy, P.; Albert, T. K.; Zacarias-Cabeza, J.; Spicuglia, S.; de la Chapelle, A. L.; Heidemann, M.; Hintermair, C.; Eick, D.; Gut, I.; Ferrier, P.; Andrau, J.-C. *Nat. Struct. Mol. Biol.* **2011**, *18* (8), 956–963.
(196) Jeronimo, C.; Bataille, A. R.; Robert, F. *Chem. Rev.* **2013**, *113* (11), 8491–8522.
(197) Buratowski, S. *Nat. Struct. Biol.* **2003**, *10* (9), 679–680.
(198) Chapman, R. D.; Heidemann, M.; Hintermair, C.; Eick, D. *Trends Genet. TIG* **2008**, *24* (6), 289–296.
(199) Eick, D.; Geyer, M. *Chem. Rev.* **2013**, *113* (11), 8456–8490.
(200) Thompson, M.; Cui, W.; Reilly, J. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (8), 1439–1452 – 1452.
(201) Holden, D. D.; Pruet, J. M.; Brodbelt, J. S. *Int. J. Mass Spectrom.*
(202) Suh, H.; Ficarro, S. B.; Kang, U.-B.; Chun, Y.; Marto, J. A.; Buratowski, S. *Mol. Cell* **2016**, *61* (2), 297–304.
(203) Schüller, R.; Forné, I.; Straub, T.; Schreieck, A.; Texier, Y.; Shah, N.; Decker, T.-M.; Cramer, P.; Imhof, A.; Eick, D. *Mol. Cell* **2016**, *61* (2), 305–314.
(204) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. *Anal. Chem.* **2015**.
(205) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. *Anal. Chem.* **2016**.
(206) Mayfield, J. E.; Burkholder, N. T.; Zhang, Y. J. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2016**, *1864* (4), 372–387.
(207) Jeronimo, C.; Robert, F. *Nat. Struct. Mol. Biol.* **2014**, *21* (5), 449–455.
(208) Tee, W.-W.; Shen, S. S.; Oksuz, O.; Narendra, V.; Reinberg, D. *Cell 156* (4), 678–690.
(209) Pähler, A.; Banerjee, A.; Dattagupta, J. K.; Fujiwara, T.; Lindner, K.; Pal, G. P.; Suck, D.; Weber, G.; Saenger, W. *EMBO J.* **1984**, *3* (6), 1311–1314.
(210) Bettelheim, F. R. *J. Am. Chem. Soc.* **1954**, *76* (10), 2838–2839.
(211) Huttner, W. B. *Nature* **1982**, *299*, 273–276.
(212) Baeuerle, P. A.; Huttner, W. B. *J. Biol. Chem.* **1985**, *260* (10), 6434–6439.
(213) Roland Beisswanger; Corbeil, D.; Vannier, C.; Thiele, C.; Dohrmann, U.; Roland Kellner; Ashman, K.; Christof Niehrs; Wieland B. Huttner. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (19), 11134–11139.
(214) Ouyang, Y.-B.; Lane, W. S.; Moore, K. L. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (6), 2896–2901.
(215) Ouyang, Y.-B.; Moore, K. L. *J. Biol. Chem.* **1998**, *273* (38), 24770–24774.
(216) Seibert, C.; Cadene, M.; Sanfiz, A.; Chait, B. T.; Sakmar, T. P. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (17), 11031–11036.
(217) Danan, L. M.; Yu, Z.; Hoffhines, A. J.; Moore, K. L.; Leary, J. A. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (10), 1459–1466.

(218) Danan, L. M.; Yu, Z.; Ludden, P. J.; Jia, W.; Moore, K. L.; Leary, J. A. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (9), 1633–1642.
(219) Moore, K. L. *J. Biol. Chem.* **2003**, *278* (27), 24243–24246.
(220) Huttner, W. B. *Annu. Rev. Physiol.* **1988**, *50*, 363–376.
(221) Stone, M. J.; Chuang, S.; Hou, X.; Shoham, M.; Zhu, J. Z. *Spec. Issue Biotechnol. Annu. Rev. 2009* **2009**, *25* (5), 299–317.
(222) Hille, A.; Huttner, W. B. *Eur. J. Biochem.* **1990**, *188* (3), 587–596.
(223) Kehoe, J. W.; Bertozzi, C. R. *Chem. Biol.* **2000**, *7* (3), R57–R61.
(224) Hortin, G. *Blood* **1990**, *76* (5), 946–952.
(225) Farzan, M.; Mirzabekov, T.; Kolchinsky, P.; Wyatt, R.; Cayabyab, M.; Gerard, N. P.; Gerard, C.; Sodroski, J.; Choe, H. *Cell* **1999**, *96* (5), 667–676.
(226) Seibert, C.; Sakmar, T. P. *Pept. Sci.* **2008**, *90* (3), 459–477.
(227) Medzihradszky, K. F.; Guan, S.; Maltby, D. A.; Burlingame, A. L. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (9), 1617–1624.
(228) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. *Posttranslational Modif. Proteomics* **2006**, *1764* (12), 1811–1822.
(229) Yagami, T.; Kitagawa, K.; Aida, C.; Fujiwara, H.; Futaki, S. *J. Pept. Res.* **2000**, *56* (4), 239–249.
(230) Liu, H.; Håkansson, K. *Anal. Chem.* **2006**, *78* (21), 7570–7576.
(231) Cantel, S.; Brunel, L.; Ohara, K.; Enjalbal, C.; Martinez, J.; Vasseur, J.-J.; Smietana, M. *PROTEOMICS* **2012**, *12* (14), 2247–2257.
(232) Yu, Y.; Hoffhines, A. J.; Moore, K. L.; Leary, J. A. *Nat Meth* **2007**, *4* (7), 583–588.
(233) Kim, J.-S.; Song, S.-U.; Kim, H.-J. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (11), 1916–1925.
(234) Drake, S. K.; Hortin, G. L. *Int. J. Biochem. Cell Biol.* **2010**, *42* (1), 174–179.
(235) Edelson-Averbukh, M.; Shevchenko, A.; Pipkorn, R.; Lehmann, W. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (12), 2256–2268.
(236) Gibson, B. W.; Cohen, P. In *Methods in Enzymology*; James A. McCloskey, Ed.; Academic Press, 1990; Vol. Volume 193, pp 480–501.
(237) Cook, S.; Jackson, G. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (6), 1088–1099.
(238) Hersberger, K. E.; Håkansson, K. *Anal. Chem.* **2012**, *84* (15), 6370–6377.
(239) Yagami, T.; Kitagawa, K.; Aida, C.; Fujiwara, H.; Futaki, S. *J. Pept. Res.* **2000**, *56* (4), 239–249.
(240) Antoine, R.; Joly, L.; Tabarin, T.; Broyer, M.; Dugourd, P.; Lemoine, J. *Rapid Commun. Mass Spectrom.* **2007**, *21* (2), 265–268.
(241) Rumachik, N.; McAlister, G.; Russell, J.; Bailey, D.; Wenger, C.; Coon, J. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (4), 718–727.
(242) Sun, Q.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R. *J. Proteome Res.* **2008**, *8* (2), 958–966.
(243) Straub, R. F.; Voyksner, R. D. *J. Am. Soc. Mass Spectrom.* **1993**, *4* (7), 578–587.
(244) Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88* (20), 4671–4675.
(245) Hiraoka, K.; Kudaka, I. *Rapid Commun. Mass Spectrom.* **1992**, *6* (4), 265–268.

(246) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20* (6), 362–387.
(247) Zhang, X.; Clausen, M. R.; Zhao, X.; Zheng, H.; Bertram, H. C. *Anal. Chem.* **2012**, *84* (18), 7785–7792.
(248) Balsved, D.; Bundgaard, J. R.; Sen, J. W. *Anal. Biochem.* **2007**, *363* (1), 70–76.
(249) Balderrama, G. D.; Meneses, E. P.; Orihuela, L. H.; Hernández, O. V.; Franco, R. C.; Robles, V. P.; Batista, C. V. F. *Rapid Commun. Mass Spectrom.* **2011**, *25* (8), 1017–1027.
(250) Amano, Y.; Shinohara, H.; Sakagami, Y.; Matsubayashi, Y. *Anal. Biochem.* **2005**, *346* (1), 124–131.
(251) Hoffhines, A. J.; Damoc, E.; Bridges, K. G.; Leary, J. A.; Moore, K. L. *J. Biol. Chem.* **2006**, *281* (49), 37877–37887.
(252) Yuh-Shyong Yang; Chen-Chu Wang; Bo-Han Chen; You-Hua Hou; Kuo-Sheng Hung; Yi-Chih Mao. *Molecules* **2015**, *20* (2), 2138–2164.
(253) Ouyang, Y.-B.; Crawley, J. T. B.; Aston, C. E.; Moore, K. L. *J. Biol. Chem.* **2002**, *277* (26), 23781–23787.
(254) Borghei, A.; Ouyang, Y.-B.; Westmuckett, A. D.; Marcello, M. R.; Landel, C. P.; Evans, J. P.; Moore, K. L. *J. Biol. Chem.* **2006**, *281* (14), 9423–9431.
(255) Westmuckett, A. D.; Hoffhines, A. J.; Borghei, A.; Moore, K. L. *Gen. Comp. Endocrinol.* **2008**, *156* (1), 145–153.
(256) Michnick, D. A.; Pittman, D. D.; Wise, R. J.; Kaufman, R. J. *J. Biol. Chem.* **1994**, *269* (31), 20095–20102.
(257) Arruda, V. R. *Blood* **2001**, *97* (1), 130–138.
(258) Sako, D.; Comess, K. M.; Barone, K. M.; Camphausen, R. T.; Cumming, D. A.; Shaw, G. D. *Cell* **1995**, *83* (2), 323–331.
(259) Bonomi, M.; Busnelli, M.; Persani, L.; Vassart, G.; Costagliola, S. *Mol. Endocrinol. Baltim. Md* **2006**, *20* (12), 3351–3363.
(260) Ludeman, J. P.; Stone, M. J. *Br. J. Pharmacol.* **2014**, *171* (5), 1167–1179.
(261) Hsu, W.; Rosenquist, G. L.; Ansari, A. A.; Gershwin, M. E. *Autoimmun. Rev.* **2005**, *4* (7), 429–435.
(262) Choe, H.; Moore, M. J.; Owens, C. M.; Wright, P. L.; Vasilieva, N.; Li, W.; Singh, A. P.; Shakri, R.; Chitnis, C. E.; Farzan, M. *Mol. Microbiol.* **2005**, *55* (5), 1413–1422.
(263) Choudhary, C.; Weinert, B. T.; Nishida, Y.; Verdin, E.; Mann, M. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (8), 536–550.
(264) Budnik, B. A.; Lee, R. S.; Steen, J. A. J. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2006**, *1764* (12), 1870–1880.
(265) Robinson, M. R.; Moore, K. L.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (8), 1461–1471.
(266) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *Proc. Natl. Acad. Sci.* **2004**, *101* (33), 12130–12135.
(267) Lim, K. B.; Kassel, D. B. *Anal. Biochem.* **2006**, *354* (2), 213–219.

(268) Hennrich, M. L.; van den Toorn, H. W. P.; Groenewold, V.; Heck, A. J. R.; Mohammed, S. *Anal. Chem.* **2012**, *84* (4), 1804–1808.
(269) Thingholm, T. E.; Jensen, O. N. In *Phospho-Proteomics*; Graauw, M. de, Ed.; Walker, J. M., Series Ed.; Humana Press: Totowa, NJ, 2009; Vol. 527, pp 47–56.
(270) Posewitz, M. C.; Tempst, P. *Anal. Chem.* **1999**, *71* (14), 2883–2892.
(271) Zhou, H.; Low, T. Y.; Hennrich, M. L.; van der Toorn, H.; Schwend, T.; Zou, H.; Mohammed, S.; Heck, A. J. R. *Mol. Cell. Proteomics* **2011**, *10* (10), M110.006452–M110.006452.
(272) Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R. *Anal. Chem.* **2004**, *76* (14), 3935–3943.
(273) Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jørgensen, T. J. D. *Mol. Cell. Proteomics MCP* **2005**, *4* (7), 873–886.
(274) Thingholm, T. E.; Jørgensen, T. J. D.; Jensen, O. N.; Larsen, M. R. *Nat. Protoc.* **2006**, *1* (4), 1929–1935.
(275) Kehoe, J. W.; Velappan, N.; Walbolt, M.; Rasmussen, J.; King, D.; Lou, J.; Knopp, K.; Pavlik, P.; Marks, J. D.; Bertozzi, C. R.; Bradbury, A. R. M. *Mol. Cell. Proteomics* **2006**, *5* (12), 2350–2363.
(276) Andersson, L. I. *J. Chromatogr. B. Biomed. Sci. App.* **2000**, *745* (1), 3–13.
(277) Helling, S.; Shinde, S.; Brosseron, F.; Schnabel, A.; Müller, T.; Meyer, H. E.; Marcus, K.; Sellergren, B. *Anal. Chem.* **2011**, *83* (5), 1862–1865.
(278) Shinde, S.; Bunschoten, A.; Kruijtzer, J. A. W.; Liskamp, R. M. J.; Sellergren, B. *Angew. Chem. Int. Ed.* **2012**, *51* (33), 8326–8329.
(279) Toyoda, M.; Narimatsu, H.; Kameyama, A. *Anal. Chem.* **2009**, *81* (15), 6140–6147.
(280) Wei, W.; Miller, R. L.; Leary, J. A. *Anal. Chem.* **2013**, *85* (12), 5917–5923.
(281) Sabol, J. K.; Wei, W.; López-Hoyos, M.; Seo, Y.; Andaya, A.; Leary, J. A. *Matrix Biol.* **2014**, *40*, 54–61.
(282) Angel, P. M.; Orlando, R. *Rapid Commun. Mass Spectrom.* **2007**, *21* (10), 1623–1634.
(283) Greer, S. M.; Cannon, J. R.; Brodbelt, J. S. *Anal. Chem.* **2014**, *86* (24), 12285–12290.
(284) Robotham, S. A.; Horton, A. P.; Cannon, J. R.; Cotham, V. C.; Marcotte, E. M.; Brodbelt, J. S. *Anal. Chem.* **2016**.
(285) Subramonian, S.; Clifford, D. *J. Solut. Chem.* **1989**, *18* (6), 529–543.
(286) Amano, Y.; Shinohara, H.; Sakagami, Y.; Matsubayashi, Y. *Anal. Biochem.* **2005**, *346* (1), 124–131.
(287) Vasicek, L.; Brodbelt, J. S. *Anal. Chem.* **2010**, *82* (22), 9441–9446.
(288) Tang, K.; Page, J. S.; Marginean, I.; Kelly, R. T.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (8), 1318–1325.
(289) Torta, F.; Elviri, L.; Bachi, A. In *Methods in Enzymology*; Elsevier, 2010; Vol. 473, pp 265–280.
(290) Schmidt, A.; Trentini, D. B.; Spiess, S.; Fuhrmann, J.; Ammerer, G.; Mechtler, K.; Clausen, T. *Mol. Cell. Proteomics* **2014**, *13* (2), 537–550.

(291) Lapek, J. D.; Tombline, G.; Friedman, A. E. *J. Proteome Res.* **2011**, *10* (2), 751–755.

Vita

Michelle Renee Robinson grew up in Pennsylvania with parents Jim and Joanne and siblings Chris and Shannon. After graduating from Chambersburg Area Senior High School in the spring of 2006, she attended West Virginia University. She graduated summa cum laude with dual Bachelor of Science degrees in chemistry and forensic and investigative science in May of 2010. During her time at WVU she was involved in researching analytical techniques for the discrimination of nail polish samples under the direction of Dr. Patrick Buzzini. Following graduation she moved to Austin, Texas to start her graduate studies under the guidance of Dr. Jennifer S. Brodbelt. Her current research aims to improve the characterization of protein post-translational modifications using mass spectrometry.

Permanent email address: mrrobinson@utexas.edu

This dissertation was typed by the author.