The Dissertation Committee for Sylvester McCarthy Greer III Certifies that this is the approved version of the following dissertation:

Development of Ultraviolet Photodissociation for High-throughput analysis of Heavily Modified Proteins and Peptides

Committee:

Jennifer Brodbelt, Supervisor

Richard M. Crooks

Livia S. Eberlin

Kyle M. Miller

Emily L. Que

Development of Ultraviolet Photodissociation for High-throughput analysis of Heavily Modified Proteins and Peptides

by

Sylvester McCarthy Greer III

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, 2018

For my son William, loving wife Emma and parents Mac and Brenda

Acknowledgements

I would like to acknowledge Jennifer Brodbelt for her support, patience, impeccable professionalism during my graduate career and her thorough manuscript revisions along the way. I would like to acknowledge Joe Cannon for his guidance and wisdom during my early years as a grad student. The constant camaraderie of other past and present group members, especially the frequent conversations with Tim Cleland, and Dustin Holden, always eased me through the difficult times. I would also like to acknowledge Mike Hollingsworth for the self-confidence to tackle any broken HPLC. My many fine collaborators deserve recognition as well especially Mariel Coradin, Simone Sidoli, Kristina Jonsson-Schmunk, and Marshall Bern whose work and efforts enabled me to complete my graduate studies. Lastly, I owe an enormous debt of gratitude to my family who have sacrificed and supported me throughout my formal education especially these last several years while I pursued my PhD.

Development of Ultraviolet Photodissociation for High-throughput analyses of Heavily Modified Proteins and Peptides

Sylvester McCarthy Greer III, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Jennifer S. Brodbelt

The utility of 193 nm ultraviolet photodissociation (UVPD) is evaluated for high-throughput proteomics applications including: analysis of small peptides in a traditional bottom-up proteomics workflow, analysis of heavily modified larger middle down sized peptides, and heavily modified intact proteins in a top-down proteomics workflow. UVPD uses higher energy ultraviolet photons (193 nm, 6.4 eV per photon), which are absorbed by the backbone to activate and dissociate ions effectively. UVPD dissociation is able to generate extensive backbone fragmentation enabling excellent characterization of peptides and proteins compared to traditional methods. Moreover, UVPD is also less hindered by certain experimental variables such as degree of modification, charge state and even ion polarity. These features are easily capitalized on for proteomics applications especially analysis of post translational modifications (PTM's). Characterization of PTM's is of great interest due to their involvement in several important cellular processes including cell signaling, tumorigenesis and gene expression. The studies covered in this work focus on utilizing the unique capabilities of UVPD to: 1.) characterize underrepresented peptides (acidic peptides and phosphopeptides) in the negative polarity including development of software for the analysis of the data generated, 2.) analyze intact proteins which have

undergone extensive chemical modification and charge state augmentation, and 3.) precisely characterize histone proteins which are heavily modified due to their central role in gene expression and other transcription related functions.

Table of Contents

Chapter 1

Introduction

1.1 INTRODUCTION

Proteomics is the comprehensive study of proteomes including their constituent proteins, protein structure, protein function and protein interactions. Proteomics presents a significant analytical challenge due to the extremely heterogeneous nature and the large dynamic range of proteins in a proteome. Despite these challenges several techniques have been employed such as immunoassay, SDS-PAGE, DNA microarrays, Western blotting, and Edman degradation, however the use of mass spectrometry has recently come to the forefront due to its ability to analyze entire proteomes in short times and its ever-growing capabilities to analyze even the most trace components of the proteome.[1–3]

The first analysis of molecules by mass spectrometry is often credited to J.J. Thomson and his seminal work in the early 20th century. However, mass analysis of biomolecules by mass spectrometry was not fully realized until the era of electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) nearly a century later. These two landmark modes of ionization were developed simultaneously; one mode (ESI) in the lab of John Fenn and the other (MALDI) in the lab of Franz Hillenkamp.[4] Electrospray ionization involves application of a high voltage to a liquid generating an aerosol containing charged analytes which are desolvated and analyzed by the mass

spectrometer.[5] MALDI involves deposition of the analyte in a specific, often acidic, crystalline matrix which is then irradiated by a laser. The ion-containing matrix is then ablated/desorbed from the surface by laser irradiation and desorbed inside the mass spectrometer for mass analysis.[6] These two methods were the first used to routinely ionize large biomolecules such as peptides and proteins. Presently electrospray is the preferred mode for high-throughput protein and peptide analysis due to its natural coupling to high performance liquid chromatography separations.[7,8]

Meanwhile another development, genome sequencing, has enabled modern mass spectrometry based proteomics.[9] Complete genomes have made possible the creation of databases of protein sequences.[10,11] Protein sequences can be generated directly from the genome based on the well-known Central Dogma of translation and transcription. After filtering out non-coding regions and extraneous information a list of proteins in the form of their amino acid sequences is generated, which in turn can be used to interpret mass spectral data based on diagnostic ions. Despite the genome having a finite size, the effective proteome is actually much greater due to the variability of genome expression, point mutations and post-translational modifications of proteins.[12] Presently the characterization of post-translational modifications by mass spectrometry is an ongoing area of development.[13–20] The remainder of this first chapter outlines the methods used to characterize proteins, peptides and their post-translational modifications using mass spectrometry, and a key theme of the doctoral work presented here describes the use of 193 nm ultraviolet photodissociation (UVPD) as a new approach for proteomics.

1.2 High-throughput Analysis of Proteins and Peptides

High-throughput proteomics, analysis of hundreds to thousands of proteins in a single experiment, has become a routine procedure for mass spectrometry due to several developments including: coupling of high performance separation techniques to MS workflows, MS instrumentation development such as novel fragmentation techniques, and development of bioinformatics tools designed for expert analysis of LC-MS data.[4,15,21–26]

Coupling of mass spectrometry to liquid chromatography and other forms of separation before mass analysis allows temporal resolution of hundreds to thousands of compounds.[27–30] Without such separation species which ionize most efficiently would overwhelm less abundant species or those that ionize poorly. Additionally, it is common to encounter species which give rise to signals with overlapping mass-to-charge ratios (m/z), and these can be readily separated based on their differences in sizes or chemical properties prior to mass spectrometry analysis, thus alleviating their *m/z* overlap.[31]

In general, reversed phase LC is the method of choice to separate peptides and proteins. Reversed phase mode separates molecules based on their hydrophobicity which for proteins and peptides this is a proxy for their size and number of hydrophobic residues.[32–34] Other forms of chromatography are often employed for specific applications based on unique molecular characteristics of the sample. For instance, hydrophilic interaction chromatography (HILIC) is well suited for separation of hydrophilic molecules which may not be retained, separable, or well resolved by reversed phase chromatography. HILIC is often the mode of choice for glycosylated and phosphorylated peptides due to the

added hydrophilicity imparted by these modifications.[35–40] HILIC has the added benefit of eluting analytes under a majority organic mobile phase composition, yielding very stable electrospray compared to the often more aqueous mobile phases used for reversed phase applications.[35]

1.2.1 Two dimensional LC-MS

In an effort to combat the complexity of biological samples, integration of additional orthogonal dimensions of separations have gained traction. Traditional peptide and protein chromatography separates these molecules primarily based on a single characteristic (e.g., hydrophobicity, hydrophilicity, size, etc.). Despite the excellent separation power of the reversed phase mode, and other methods such as HILIC, often in samples containing thousands of analytes many will co-elute due to similar hydrophobic character. One solution is to separate the analytes a second time based on an orthogonal chemical characteristic, such that analytes of similar character X have very different character Y. For example, peptides can be separated based on charge using strong cation exchange (SCX) and then separated by hydrophobicity using RPLC. Recent advances have enabled these orthogonal methods to be performed in an integrated online fashion such that the sample is injected onto the LC and both dimensions of separations are performed in an automated fashion using a single column with serial packings after which the analytes are electrosprayed into the MS for analysis.

The most well-known example of 2D LC for proteomics called multidimensional protein identification technology (MUDPIT) was introduced by the Yates lab.[33,41] It uses a

single column packed in serial with two different stationary phases, first with C_{18} RP particles and then with SCX particles to form a SCX-RP system for separation of a peptide mixture prior to analysis by mass spectrometry. The choice of these stationary phases is ideal in that the mobile phases used for each is compatible with the other stationary phase. Peptides are injected onto the SCX column, and a fraction of the absorbed peptides are eluted onto a RP column using a salt step gradient. After washing away salts and buffers, peptides retained on the RP column are eluted from the RP column into the mass spectrometer using a gradient of increasing organic solvent concentration. The RP column is re-equilibrated in order to adsorb another fraction of peptides from the SCX column. This process is repeated for several fractions of increasing salt concentration form SCX column. In its initial implementation by the Yates lab a direct comparison of 1D RP and 2D SCX-RP analysis were compared based on identified proteins from the 80S ribosome from yeast, a model system containing less than 100 total proteins. The 1D approach identified 56 proteins while the 2D approach identified 95 a 70% increase.[42] Subsequently improved implementations of 2D SCX-RPLC was applied to a more complex yeast cell lysate and more than tripled the number of identified proteins compared to 1D technology.[41] 2D-LC methods for large-scale proteomics are now commonplace, as evidenced by adoption of this type of strategy in numerous recent reports.[43–47]

While advancements and integration of chromatography into the MS workflow allowed the introduction of thousands of species, improvements in the capabilities of mass spectrometers allows full advantage to be taken of these high-performance separations.

1.3 ALTERNATIVE APPROACHES TO PROTEOME CHARACTERIZATION

Traditional "bottom-up" proteomics involves enzymatically digesting proteins into small peptides followed by LC-MS analysis. Bioinformatics software is then used to stitch together the original proteins from the identified peptides in a "bottom-up" fashion. Bottom-up proteomics has benefited greatly from improvements in LC efficiency (i.e. nanobore LC), high resolution accurate mass determination (i.e. Orbitrap mass analyzer), and parallelization-capable mass spectrometers (i.e. ion trap-orbitrap hybrid mass spectrometer) enabling more peptide identifications and greater characterization of the proteome than previously possible.[48] Despite the gains in performance, not all peptides are easily interrogated using traditional bottom-up approaches, namely PTM-bearing peptides, very acidic and very basic peptides. Recently much progress has been made in the use of alternative approaches. Three alternative approaches will be introduced and discussed in the next sub-sections: top-down proteomics, middle-down proteomics, and negative polarity proteomics and are also utilized in the following Chapters: top-down (Chapters 5 & 8), middle-down (Chapters 4 & 7), and negative polarity (Chapters 3&6).

1.3.1 Top-down Proteomics

Analysis of intact proteins known as "top-down proteomics" (**Figure 1.1**) presents several technical challenges, some of which have been addressed by the increasing availability of high resolution/high accuracy MS platforms while others such as efficient separations and tandem MS techniques for intact proteins continue to be developed. Separations of intact proteins often suffer lower efficiency due to their size and complex

chemical makeup (i.e acidic, basic, hydrophobic, and hydrophilic residues) giving rise to many poorly controlled analyte-stationary phase interactions and analyte-solvent interactions.[49,50] Peptides generated by bottom-up methods have significantly simplified chemical characteristics, often influenced by the enzyme used to digest them, and separation is controlled predominately by hydrophobic interaction with the stationary phase.[51] For instance, trypsin predominantly generates peptides with a lysine or arginine residue at the C-terminus, making peptides generally more amenable to RPLC separations. Furthermore, peptides contain fewer ionizable sites and generally exist in two or three charge states, whereas intact proteins often exist in more than ten charge states which leads to greater signal dilution and fundamental sensitivity concerns. In essence, the signal of a particular charge state arising from an equal number of molecules of a peptide or a protein may vary several orders of magnitude due to original signal being diluted over many more charge states. This problem is further exacerbated during a tandem MS experiment. Peptides may dissociate into dozens of fragments whereas intact proteins often dissociate into hundreds of possible fragments, again causing signal dilution.[52] Additionally, due to the narrow spacing of the higher charge state species of proteins, a high resolution instrument is required for these analyses; however, this has largely been addressed in recent years by the availability of high resolution Orbitrap mass spectrometers.[48,53] In spite the challenges of analyzing intact proteins, the potential advantages are extremely desirable, achievable only through the top-down approach and are now within reach.[54]

As previously mentioned top-down proteomics offers several distinct advantages. One of the most exploitable advantages is the ability to uniquely characterize the entire protein and any modifications, further referred to as a proteoform.[55] In the case of bottom-up methods the original nature of the proteoform or proteoform must be inferred based on a few representative peptides per protein, thus leaving gaps in sequence coverage and the potential for overlooking PTMs.[56] In contrast, a top-down methodology can utilize the intact mass to identify differing proteins or proteoforms, and use subsequent fragmentation to localize any modifications.[57] One particular case where this is particularly advantageous is in analysis of heavily modified proteins.[58] Proteolysis of a mixture of proteoforms yields peptides with and without modifications, making it nearly impossible to confidently determine whether certain modifications co-exist unless they are located the same peptide.[59] Analysis of a mixture of intact proteoforms yields a list of protein masses which correspond to the protein sequence plus any modifications. Further interrogation of these species by tandem mass spectrometry allows each proteoform to be assigned uniquely (i.e. identity and location of each modification) if the fragmentation pattern is sufficiently detailed.[60] For this and other reasons top-down proteomics has enjoyed a recent surge in exploration. The lynchpin in the process of uniquely identifying all detectable proteoforms in a mixture is the ability of the MS2 technique to provide full characterization, meaning extensive series of fragment ions that allow confirmation of sequence and PTM information. Without identifying and localizing the modifications fully, inferences must be made similar to bottom-up methodologies, and this nullifies much of the clarity imparted

by the top-down method. As described in the previous section, several MS/MS methods have made inroads in the area of PTM characterization of intact proteins, including EThcD, ETD and UVPD which are discussed in more depth in Chapter 8.

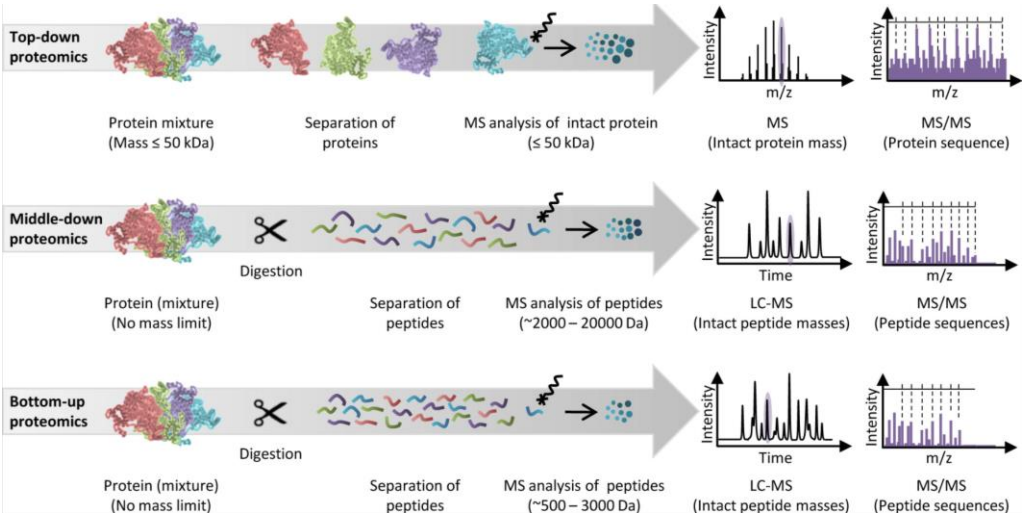


Figure 1.1 Overview of top-down, middle-down, and bottom-up proteomics workflows. (from *J. Proteome Res*., 2013, 12 (3), pp 1067–1077). Briefly, top-down proteomics focuses on analysis of intact minimally processed proteins, middle-down proteomics focuses on analysis of proteins cleaved into fewer large sized peptides, and bottom-up proteomics often utilizes trypsin to generate the largest number of small peptides. All methods are can be coupled to LC for analysis of complex samples. Spectral complexity is the inverse of sample prep (i.e. intact protein generates the richest fragment spectra, while bottom up usually generates the sparsest).

1.3.2 Middle-down Proteomics

Middle-down proteomics (**Figure 1.1**) uses methods common to both top-down and bottom-up analyses. Proteins are cleaved into smaller pieces via enzymatic or chemical digestion similar to bottom-up proteomics.[61] However, the proteolysis is carefully controlled to generate much larger peptides, usually 5-20 kDa in size. The practical methods for generating these peptides has varied greatly and is currently an area of pursuit.

A single method to generate middle-down size peptides remains elusive. Currently endoproteinase GluC is the most commonly used protease to generate larger middle down sized peptides. GluC cleaves C-terminal to glutamic acid.[61] In instances when GluC does not generate the desired peptides, protease AspN is utilized which cleaves N-terminal to aspartic acid.[62] Currently novel proteases are being explored in order to deliver a better solution for generating middle down peptides. For example, the enzyme known as neprosin has shown promise; it is selective for cleavage C-terminal to proline.[63] Proline occurs relatively infrequently compared to other residues, thus making it an ideal proteolytic cleavage point for generating middle down peptides. Chemical digestion using formic acid has also been pursued to generate peptides in this size range.[64] Treatment with formic acid shows specificity towards acidic residues (aspartic acid and glutamic acid); however, it can be relatively non-specific unless the digest conditions are strictly controlled.[64]

Since these peptides are very large, MS analysis often follows a top-down approach using high resolving power, accurate mass instrumentation.[61] Many of the difficulties of top-down analysis are mitigated by focusing on this alternative "middle" size regime. Given the longer stretch of protein backbone present in middle-down sized peptides, combinations of PTMs can often be fully characterized thus overcoming the shortfalls of bottom-up PTM analysis.[65] The middle-down approach for PTM analysis is further discussed in Chapter 7. Middle-down strategies show much promise in the analysis of combinatorial PTMs and is currently an emerging area in mass spectrometry-based proteomics research.

1.3.3 Negative Polarity Proteomics

The vast majority of proteomics experiments are conducted under positive polarity mode where cations are formed and analyzed. However most mass spectrometers can be operated in the negative polarity to generate anions via deprotonation or adduction of halide ions. Negative polarity lends itself naturally to analysis of acidic proteomes and PTMs such as sulfation and phosphorylation which add negative charge to peptides.[66,67] However, analysis of peptide and protein anions via electrospray ionization presents several challenges, namely poor ionization and inadequate fragmentation. Negative mode ionization occurs through deprotonation which is best achieved under basic conditions with analytes which are acidic.[68] However basic spray solvents cause rapid degradation of chromatographic stationary phase and are viewed as incompatible with long-term LC operation.[69] As a result, negative mode is generally practiced under buffered acidic conditions or even using conventional acidic conditions.[70] A common way to cope with poor sensitivity in the negative mode is to maximize the influx of ions by positioning the ESI spray close to the MS inlet. Unfortunately, poor ionization is often exacerbated by corona discharge, a phenomenon which occurs when the ESI spray is close to the mass spectrometer inlet and electrical arcing occurs or the high voltage applied to the ESI needle creates an ionized gas plasma, a factor that occurs more readily in the negative mode owing to the low dielectric strengths of nitrogen and oxygen molecules in the atmosphere.[71] Corona discharge can be mitigated by lowering the ESI spray voltage and the spray solvent is more organic and less aqueous. [67] Despite the greater hurdles associated with negative

ESI, it has been used successfully by utilizing dissociation methods such as UVPD and NETD which generate diagnostic fragment anions and mobile phase additives which promote deprotonation.[69,70] Negative polarity has been used to enhance characterization of the acidic proteome, phosphopeptides and glycopeptides.[69,73,74]

Traditional CID and HCD when performed in the negative mode generate primarily non-diagnostic fragments and neutral losses of water and ammonia, thus proving generally useless for database searching and high throughput applications.[75–77] Non-diagnostic fragmentation can be overcome by using alternative activation methods such as NETD and UVPD and proper experimental conditions discussed further in Chapters 3 and 6, thus opening up areas of study for underrepresented portions of the proteome.[68,70,78] **Figure 7.4** shows the distribution of pI values of the human proteome depicting the large population of acidic proteins and the potential for increasing the depth of proteome coverage with negative mode proteomics studies. Chapters 3 and 6 discuss experimental and data analysis approaches to improve negative polarity proteomics experiments and data interpretation for broader overall proteome characterization.

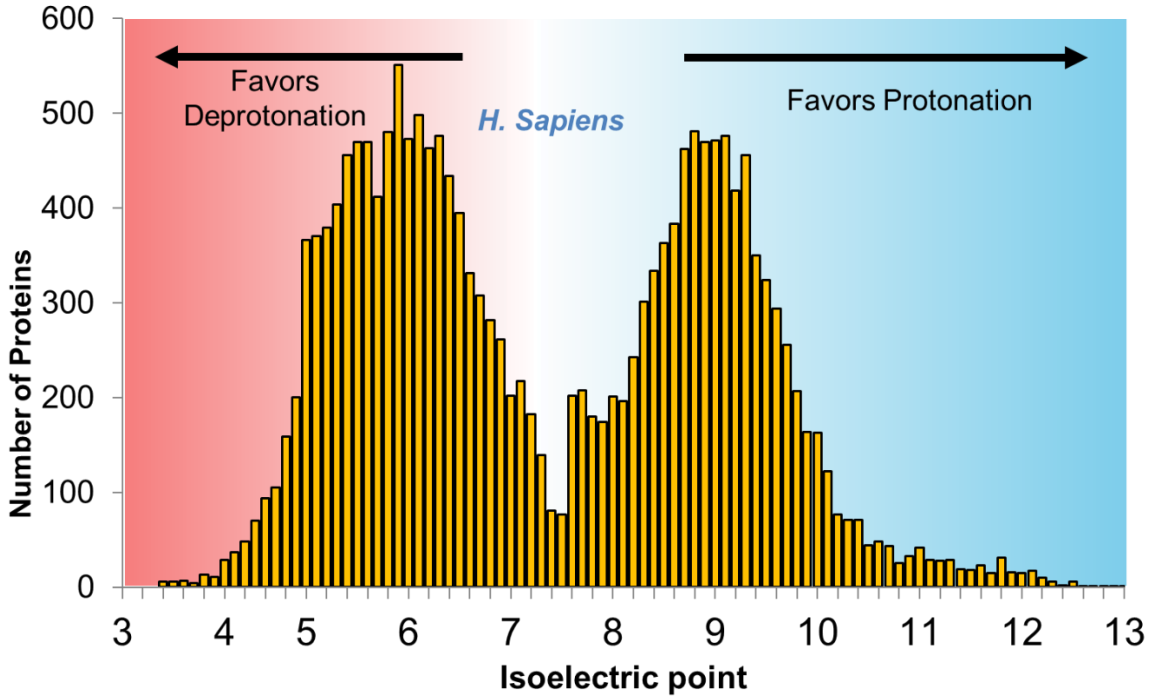


Figure 1.2 Isoelectric point (pI) distribution of proteins in the human proteome showing the large portion of proteins which may benefit from negative polarity proteomics (red region). Protein pI was calculated using an online isoelectric point calculator (http://isoelectric.ovh.org/) and the human proteome from UniProt (ID:9606)

1.4 Protein Backbone Structure and Fragment Nomenclature

The goal of any tandem MS experiment in proteomics is to effect cleavage of the peptide bonds in an extensive, reproducible and predictable way to enable efficient protein sequencing. The position of bond breakage and resulting fragment structures is to a large extent what sets apart the various activation methods.[79] **Figure 1.3** shows a generic tetra-peptide where R represents the amino acid sidechains. Collisional dissociation generates b and y type ions by cleavage of the C-N amide bonds. The amide bond is the weakest bond

and is readily cleaved under CID conditions.[80] Electron-based dissociation generates c and z type ions by cleaving the N-Cα bond. UVPD is unique in that it can cleave Cα-C to generate a and x type ions, while also generating b, c, y and z type ions.[81] The six ion types discussed here all include one terminus of the peptide or protein, either the N- or C-terminus. The N-terminus is present in a, b and c type ions, while the C-terminus is present in the x, y and z type ions. In the case that a peptide or fragment is cleaved more than once or undergoes secondary fragmentation in which it does not contain either the N- or C-terminus, it is termed "internal ion" and is not considered diagnostic. Internal ions are not commonly used in database searches because they are the result of multiple cleavages and do not have a well-defined reference point (i.e. the C- or N-terminus) which allows them to be confidently assigned to particular protein sequence.[82] Moreover, the search space required for assignment of internal ions is orders of magnitude greater than that needed for the far more limited N- and C- terminating fragment ions, thus slowing searches and affording lower statistical confidence in their assignment.[82]

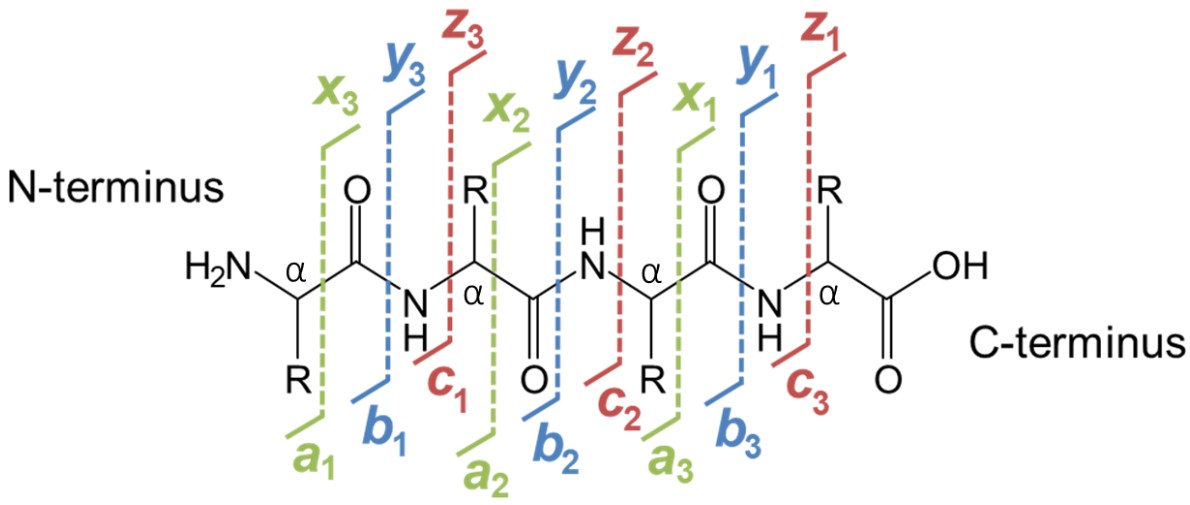


Figure 1.3 Peptide fragment nomenclature mapped onto a tetrapeptide. R represents possible side chain structure attached to the alpha-carbon (α). *a, b* and *c* ions are termed n-terminal ions as they contain the amino terminus of the peptide and conversely *x, y* and *z* ions are termed c-terminal as they contain the carboxyl terminus.

1.5 TANDEM MASS SPECTROMETRY

A tandem mass spectrometer performs mass analysis of analyte ions and then submits selected precursor ions to some form of activation such as collision with gas molecules (CID), interactions with electrons (ETD, ECD etc.), irradiation by photons (UVPD, IRMPD) or other means which produce fragmentation. The resulting fragment ions are mass analyzed. Nearly all of the activation techniques proven feasible for high-throughput analysis of peptides are proteins fall into these three categories: collisional activated dissociation, electron-based dissociation or photodissociation.[83] Despite the variety of techniques, there are several underlying qualities common to all that make them

experimentally useful. One of the essential qualities each method possesses is the ability to generate predictable and reproducible fragment ions, so called "diagnostic ions." **Figure 1.3** depicts the six most common diagnostic ion types. The mass of these ions can be predicted based on the peptide backbone and modifications of side-chains which can in turn be used to generate a list of theoretical masses to identify the peptide or protein being analyzed by database searching. Another key quality of these methods is their ability to be applied to a variety of biomolecules.[84] All three types of activation methods are widely applicable; however, some are much more well suited for certain classes of molecules; the strengths and weaknesses of each activation technique will be described in the subsequent experimental methods chapter. Lastly a quality that can be desirable although not essential is the ability of the technique to be tuned to individual applications.[84–86] Dissociation of peptides and proteins often require significantly different amounts of energy, thus in order to maximize fragment generation of both protein and peptide a tunable dissociation method is required. Many of these techniques can be adjusted based on various factors including the amount of energy deposited in the analyte, duration of exposure to the analyte, or the wavelength of light used, and in some cases multiple parameters can be tuned in concert to achieve the ideal result.[25]

1.5.1 Collisional Induced Dissociation (CID)

Traditionally fragmentation of peptides has been achieved by collisional activation dissociation (CID). CID is the gold standard dissociation technique for biomolecules due to its effective generation of diagnostic *b* and *y* type fragment ions, reproducibility and

ability to be adjusted to suit various analytes.[4] CID has been implemented across the majority of commercial MS platforms (TOF, FT-ICR, quadrupole ion trap, linear ion traps, etc.) Mechanistically CID of peptides has been studied extensively and is governed by the generally accepted proton mobility mechanism stating that cleavage is charge site initiated.[87–89] In practical terms a peptide with de-localized protons will yield the most information rich fragmentation. Protons can become localized around very basic sites such as lysine and arginine side chains, thus in general when a peptide's charge (i.e. extent of protonation) exceeds the number of basic sites rich fragmentation will result upon CID.[90] In ion trapping instruments (the platform used in this dissertation), CID is commonly implemented in two ways. CID can be achieved via ion trapping and subsequent resonance during which the ions are translated radially (CID) within in the trapping region inducing collisions with background gas molecules and resulting in dissociation. Alternatively the ions can be accelerated axially (HCD) resulting in collisions with background gas.[91,92] During this energetic resonance or axial acceleration, the mass-selected precursor ions are accelerated and collide with inert gas molecules like He or N_2. The degree of fragmentation can be adjusted by attenuation of the amplitude of the resonance waveform, (for CID), or attenuation of the acceleration voltage, (for HCD). CID in ion traps has one major drawback known as the low mass cutoff which prevents low mass ions from being detected. LMCO is governed by the following relationship which is related to the RF amplitude applied to the trap.

$$LMCO = \frac{m \cdot q_z}{ejectioncondition@q_z}$$

Where ($\boldsymbol{m}$) is the ion mass and ($\boldsymbol{q_z}$)is the stability factor related to the drive RF (usually 0.25) and the ejection condition at ($\boldsymbol{q_z}$=*0.25*) is 0.908. An in-depth discussion of ion physics in an ion trap can be found in Reference 93.[93]

Certain classes of PTMs are labile and are often lost before backbone dissociation during CID, thus masking their original location on the peptide and making CID a poor choice for analysis of phosphopeptides and glycopeptides. More recently HCD has been adopted as the preferred mode of collisional dissociation due to its improved performance over CID. Energy deposition can be modified by adjusting the accelerating voltage applied to the ions as they travel through the collision cell. HCD has the added benefits of higher energy deposition compared to CID, shorter activation time required compared to CID, and no low mass cutoff. Thus far, neither CID or HCD have shown useful implementations in the negative polarity for high-throughput proteomics applications due to the lack of predictable fragment ion generation and the overwhelming presence of non-diagnostic fragments such as neutral losses (i.e. loss of water and ammonia). [76,94] Very basic and very acidic peptides undergo preferential cleavage by CID at only a few sites in the peptide resulting in poor sequencing of the peptide: proline cleavages dominate basic peptides, while aspartic and glutamic acid cleavages dominate acidic peptides.[95–97] As the importance of PTMs has become well-recognized, several new activation methods have been developed to overcome the CID limitations mentioned above.

1.5.2. Electron-Based Dissociation

Electron-based dissociation techniques such as electron transfer dissociation (ETD) and electron capture dissociation (ECD) are desirable alternatives to CID due to fact that these activation methods do not dislodge labile PTMs yet still allow extensive cleavage of peptides and proteins.[98] ECD was the original electron-based activation method developed for FTICR mass spectrometers, but could not be implemented for quadrupole ion traps (QITs) and linear ion traps (LITs) due to the difficulty of introducing low-energy electrons into a high radiofrequency field such as used in QITs and linear ion traps .[99] ETD had been more widely adopted on several commercial mass spectrometry platforms and has been proven a useful activation technique for high-throughput protein and peptide analyses, therefore the following will focus on ETD.[100] ETD utilizes anions, not free electrons, which are easily manipulated under conditions found in ion trap mass spectrometers, unlike ECD which utilizes electrons. ETD occurs in an ion trapping region where radical anions are co-isolated with peptide/protein cations for a predetermined length of time wherein the two species may interact and promote a transfer of one electron from the reagent anion to the peptide or protein. The process is exothermic, and the excess energy can be dissipated via fragmentation. Other outcomes are also possible, such as simple charge reduction where an electron is transferred but fragmentation does not occur. This phenomenon is known as an "ET-no-D" event.[101] ETD fragmentation is hypothesized to occur via hydrogen migration following the transfer of electron to a protonated site of the cation. The process is believed to be non-ergodic, allowing the PTMs to be retained on the resulting c type and

z type ions which allows the modifications to be localized along the backbone.[99,100,102] Unlike CID, ETD can be achieved under negative polarity conditions by utilizing a radical cation of fluoranthene to abstract an electron from the peptide anion promoting electron rearrangement resulting in extensive backbone cleavage into diagnostic fragments.[78]

The main drawback to ETD is its charge state dependency. ETD is most effective for ions in higher charge states which poses a problem for bottom-up proteomic experiments (**Figure 1**) where the majority of peptides carry only two or three positive charges.[85] For ions in low charge states, this means that the electron transfer reactions are less exothermic, and fragmentation is less efficient. In an effort to surmount this problem, "activated ion" ETD (AI-ETD) uses supplemental energy in the form of gentle collisional activation or photon irradiation to cause supplemental activation of the molecule in order to improve the fragmentation efficiency by alleviating the occurrence of ET-no-D.[101,103,104] Riley and coworkers have shown a 60% improvement in peptide identification using AI-ETD compared to ETD.[105]

1.5.3 Ultraviolet Photodissociation (UVPD)

Ultraviolet photodissociation can be achieved using photons of various wavelengths; the following discussion focuses on 193 nm photons (6.4 eV), which are absorbed by the amide chromophore of the peptide backbone.[106,107] The peptide backbone absorbs the photon promoting the ion into an excited electronic state. Dissociation of the excited state molecule results in a wide array of ion types including all six common diagnostic ions: *a*,*b*,*c*,*x*,*y*,and *z* type ions.[81,108] The excited state ion can also return to the

ground state after internal conversion, and fragmentation results in b/y ions. UVPD is also amenable to negative polarity proteomics experiments because UV photoactivation does not depend on charge.[73,109] Negative UVPD is further discussed in Chapters 3 and 6. Upon UV irradiation peptide anions are dissociated into namely a and x type diagnostic ions. Moreover, UVPD does not rely largely on mobile protons as does CID and is thus charge state independent, unlike both CID and ETD, allowing small low charged and large highly charged peptides to be fragmented in kind.[110]

1.6 Advances in Mass Analyzers

While development of activation techniques enables the creation of more fragment-rich spectra, concurrent improvements in resolving power and mass accuracy of mass analyzers enables better utilization of this information. High resolution/accurate mass (HRAM) spectrometers allow monoisotopic analysis of larger peptides and proteins and resolution of nearly isobaric modifications (e.g. phosphorylation: 79.966, sulfation: 79.957, trimethylation: 42.047, acetylation: 42.011).[48,111] Traditionally high resolving power mass spectrometry has been restricted to FT-ICR instruments which have now achieved resolving powers as high as 2,000,000 using 21 T magnets .[112] However, in 2005 the first commercially available high-resolution mass analyzer not requiring a large high field magnetic became available: the Orbitrap™ mass analyzer.[113] The Orbitrap platform itself has undergone further refinements over the past decade, including reducing the space between the electrodes of the analyzer and more recently improvements in ion injection;

this compression of the ion packet has enabled resolving power of 1,000,000 to be achieved.[113–115]

Feeding on the wide scale availability of high resolving power and mass accuracy, automated data searching programs (so called bioinformatics platforms) have enabled advances in protein discovery and quantification.[23,24,116–120] Generally, database searching programs generate a list of theoretical peptide or protein masses and match these to the masses present in LC-MS data. As the number or proteins or peptides in the sample increases, more ions with near-isobaric *m/z* values are created, causing spectral congestion and requiring both higher resolving power for differentiation and greater mass accuracy for confident assignment. Improvements in mass accuracy from 50-200 ppm for a lower resolution ion trap mass spectrometer to routinely >5 ppm for an Orbitrap mass spectrometer has enabled dramatic reduction in the number of theoretically possible matching species by narrowing the mass window as shown in **Figure 6**. Reducing the number of overlapping species in each resolvable mass bin provides faster data searching and much higher confidence in the matching of theoretical to observed spectra, known as a peptide spectral match (PSM).[121]

1.7 DATA INDEPENDENT ACQUISITION

This dissertation has focused on the traditional mode of tandem mass spectrometry known as data dependent tandem acquisition (DDA) in which the selection of precursor for dissociation is based on a precursor scan and selection criteria specified by the user.

Recently the data independent acquisition (DIA) mode of tandem MS has been gaining traction and proven successful largely due to advances in resolution and accurate mass analysis.[122,123] DIA submits the entire range of precursors to MS2 simultaneously and then reconstructs the precursors from the fragment ions. Alternatively, the *m/z* landscape is divided into smaller mass windows still containing many precursors and submitted to MS2.[124] The latter is a good compromise when the number of precursor signals is too great to resolve their overlapping fragment ions. The most challenging aspect of DIA is data analysis for which software has been developed uniquely to interpret this data in the form of multiplexed fragment ions. Discussion of the DIA data analysis is beyond the scope of this dissertation which exclusively employs data dependent acquisition, however a thorough discussion recently made by Hu et al.[124] Briefly, the fragment ions are monitored over time to generate a chromatogram of the fragments. The fragments are then submitted to a database search which is narrowed-down based on predicted peptide retention times and other inputs such as proteolytic agent, species of origin of the sample (similar to traditional data searching). The fragments are scored as groups in order to determine which fragments are from the same precursor, based how well they match a theoretically generated fragment ion spectrum as a group.

The most popular form of DIA presently is known as Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra (SWATH-MS).[125] By multiplexing precursor fragmentation, compared to one at time in DDA, large gains in protein and peptide IDs can be realized. In a direct comparison of DDA and DIA analysis

of CDK4 affinity purified samples, 5,089 peptides were identified in all three DIA replicates, whereas 2,741 were identified in all three DDA replicates; an 86% increase.[126]

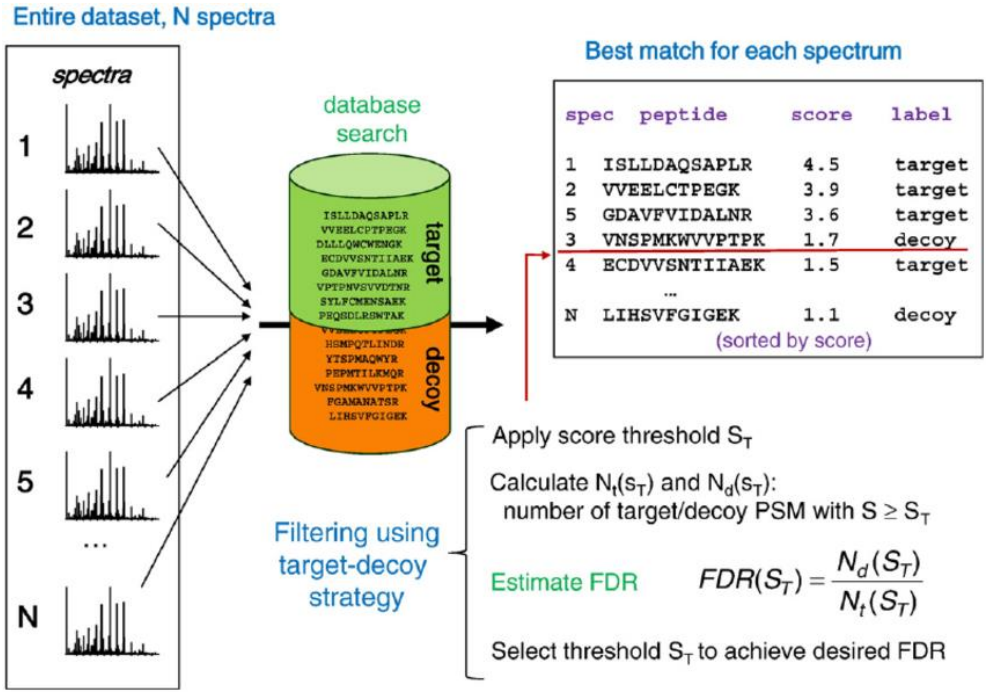


Figure 1.4 Graphical representation of Target Decoy database searching for LC-MS datasets (adapted from Ref. 132). Fragment spectra are compared to true and false (decoy) theoretical spectra based on the user selected proteome. All matches are ranked and then an FDR is calculated based on the frequencies of matches to false and true theoretical spectra

1.8 TARGET-DECOY DATABASE SEARCHING FOR PROTEOMICS DATA

Database searching of mass spectral data is also known as target–decoy searching and is the most popular way to validate PSMs identified by LC-MS.[127] The alternative is known as de novo searching and relies on matching the mass differences between series of fragments and known amino acids. This thesis focuses on target-decoy methods

exclusively; however, a relevant review has been authored by Medzihradszky and Chalkeley.[128]

The earliest software designed to automate LC-MS data analyses, Sequest, scores PSMs based on the correlation between experimental and theoretical fragmentation spectra.[129] In silico, protein sequence databases are digested to yield peptides, for which theoretical product ions are calculated and matched to LC-MS/MS data.[127] The main Sequest score is called XCorr, where a higher value is better.[130] The Xcorr scores is based on the number of ions in common between the experimental and theoretical spectrum. Several other bioinformatic algorithms have been developed since Sequest, and each uses a unique variation of this scoring method.[23,24,116,117,120] Both good and bad PSMs are ranked by score. To filter out bad matches and ultimately report confident results with some threshold score above which only true matches are made needs to be determined. Originally this required manual validation which was time consuming and not standardized across the field.[131] In order to overcome this problem, the experimental data is searched against a database populated by incorrect protein sequences, in addition to the true sequences. The incorrect sequences are generated by reversing the correct sequences, thus the terminology forward and reverse database was adopted for the forward and reversed sequences. [132,133] Reversed sequences are also known as decoys. When all the forward and reverse PSMs are scored and ranked, the resulting scores can be used to discriminate between truly good PSMs (matches to the forward database) and poor PSMs (matches to the decoy database), based on the assumption that high scoring PSMs are the result of a real match between

experimental and theoretical data and low scoring matches are the result of a match made randomly to the decoy theoretical spectra. This enables the calculation of false discovery rate (FDR), shown in **Figure 5**, which is defined as the number of false positives over total positives. A 1% FDR has been adopted by the proteomics community as the universal score boundary.[134–137] PSMs represent individual peptide sequence matches to the mass spectral data; these can be stitched together to build up the parent proteins which are then reported as the final result of an LC-MS experiment.

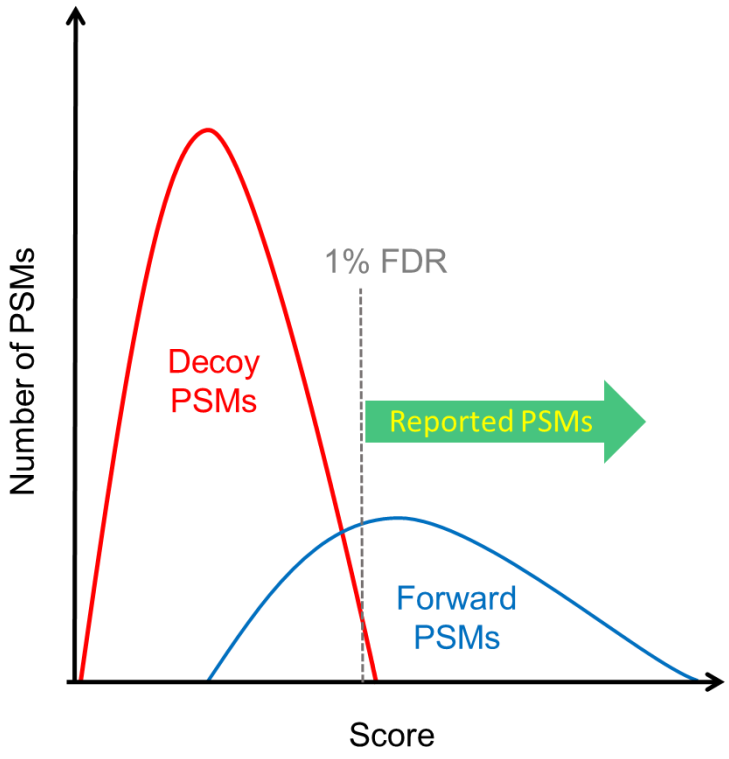


Figure 1.5 Target-decoy PSM distributions illustrating a classical 1% FDR cut off. High quality spectra match well to forward theoretical spectra and thus achieve higher scores, while lower quality or ambiguous spectra may match forward decoy spectra poorly and achieve a low score or match a decoy theoretical spectrum also likely achieving a low score. However, it is accepted that some poor spectra will make high scoring matches to

forward spectra by chance, which has been deemed acceptable at a rate of 1%. Thus, all high score matches are reported and the cutoff for reported spectra is the point at which 1% of the total matched spectra are matches to decoy theoretical spectra.

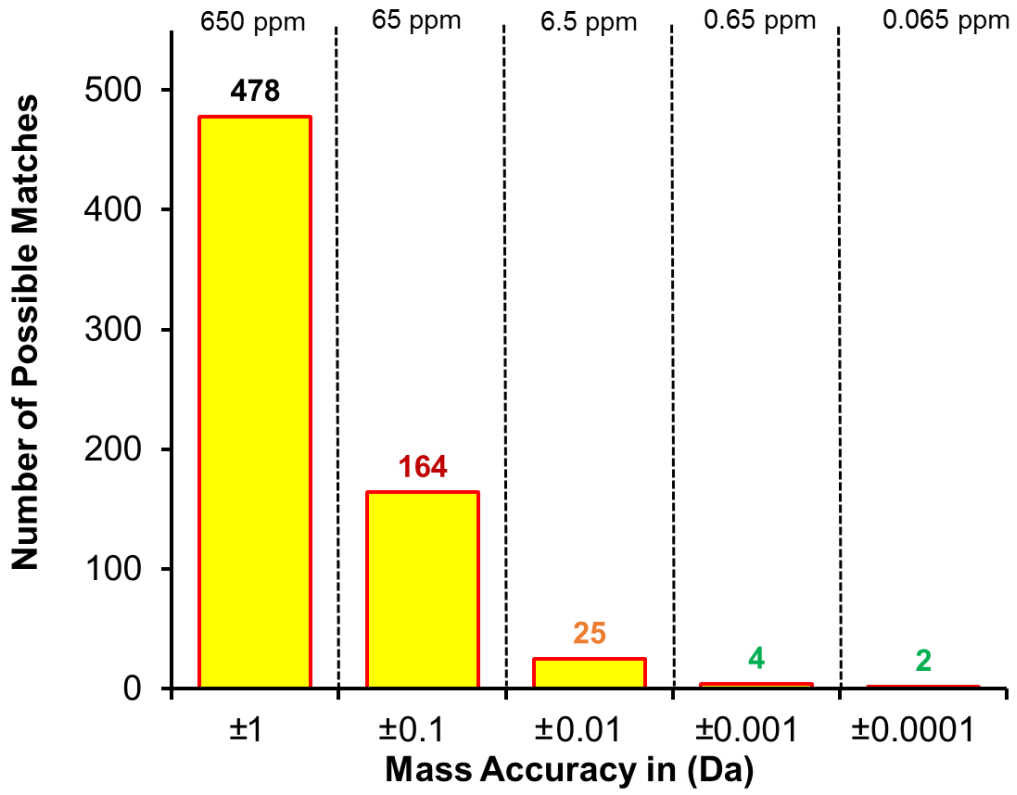


Figure 1.6 Reduction in theoretical peptide candidates achieved by improved mass accuracy as shown by the number of peptide candidates based on increasing mass accuracy for an arbitrary peptide of mass 1160.6575 Da in the human proteome.

1.9 OVERVIEW OF CHAPTERS

This dissertation outlines several applications of UVPD which are aimed at expanding the depth and breadth of proteome coverage. Broadly UVPD can achieve this

aim by expanding sequence coverage and dissociating underrepresented parts of the proteome.

Dissociation of anions has been elusive for traditional dissociation techniques however UVPD provides diagnostic fragmentation of peptide anions. Chapter 3 discusses a simple, efficient derivatization method to enhance ionization of anions resulting in improved UVPD.

In chapter 4 HCD and UVPD dissociation of peptides of various characteristics (length, charge, chromogenicity). HCD excels at dissociating peptides with basic sites at the C-terminus, whereas UVPD exhibits modestly better performance for longer peptides and those with acidic sites near the c-terminus.

UVPD and HCD dissociation are further investigated in chapter 5. Carbamylation is used to probe the role of protonation in mediating the fragmentation of intact proteins. Carbamylation is used to block the sidechain of lysine which dramatically reduces the charge states displayed allowing the same protein to be studied both fully carbamylated and unmodified.

Despite the utility of UVPD there was no software available to analyze both negative and positive mode UVPD LC-MS data. Chapter 6 describes a widely available data analysis software package which is trained to accept both positive and negative mode UVPD. This software is used to analyze LC-MS data from human liver cell lysates using HCD, UVPD and negative UVPD. The complementarity of these three activations,

including enhanced sequence coverage and numbers of protein identifications, is described.

In chapter 7 UVPD is used to analyze heavily modified middle-down size peptides from histone proteins in a high-throughput fashion. UVPD is compared to the current method of choice ETD. Performance metrics such as sequence coverage, modification localization and P-score are evaluated for both methods. UVPD performed comparably to ETD in determination of PTM distributions and total identifications, UVPD was shown to excel in analysis of the most heavily modified forms.

Application of UVPD for analysis of histone is continued in Chapter 8. The feasibility of shotgun UVPD analysis of intact histones is considered. Performance metrics including: number of proteoforms identified, P-score, C-score are compared for HCD, EThcD and UVPD.

1.10 REFERENCES

(1) Aebersold, R.; Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537* (7620), 347–355.

(2) Bensimon, A.; Heck, A. J. R.; Aebersold, R. Mass Spectrometry–Based Proteomics and Network Biology. *Annu. Rev. Biochem.* **2012**, *81* (1), 379–405.

(3) Yates, J. R. The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. *J. Am. Chem. Soc.* **2013**, *135* (5), 1629–1640.

(4) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422* (6928), 198–207.

(5) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization-principles and practice. *Mass Spectrom. Rev.* **1990**, *9* (1), 37–70.

(6) Strupat, K.; Karas, M.; Hillenkamp, F. 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry. *Int. J. Mass Spectrom. Ion Process.* **1991**, *111*, 89–102.

(7) Ikonomou, M. G.; Blades, A. T.; Kebarle, P. Investigations of the electrospray interface for liquid chromatography/mass spectrometry. *Anal. Chem.* **1990**, *62* (9), 957–967.

(8) Ling, V.; Guzzetta, A. W.; Canova-Davis, E.; Stults, J. T.; Hancock, W. S.; Covey, T. R.; Shushan, B. I. Characterization of the tryptic map of recombinant DNA derived tissue plasminogen activator by high-performance liquid chromatography-electrospray ionization mass spectrometry. *Anal. Chem.* **1991**, *63* (24), 2909–2915.

(9) Tyers, M.; Mann, M. From genomics to proteomics. *Nature* **2003**, *422* (6928), 193–197.

(10) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (11), 5011–5015.

(11) Biemann, K. Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods Enzymol.* **1990**, *193*, 455–479.

(12) Olsen, J. V.; Mann, M. Status of Large-scale Analysis of Post-translational Modifications by Mass Spectrometry. *Mol. Cell. Proteomics* **2013**, *12* (12), 3444–3452.

(13) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci.* **2004**, *101* (33), 12130–12135.

(14) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. *Anal. Chem.* **2015**, *87* (8), 4152–4158.

(15) Humphrey, S. J.; Azimifar, S. B.; Mann, M. High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nat. Biotechnol.* **2015**, *33* (9), 990–995.
(16) Li, Y.; Silva, J. C.; Skinner, M. E.; Lombard, D. B. Mass Spectrometry-Based Detection of Protein Acetylation. In *Sirtuins*; Hirschey, M. D., Ed.; Humana Press: Totowa, NJ, 2013; Vol. 1077, pp 81–104.
(17) Zhang, Y.; Song, L.; Liang, W.; Mu, P.; Wang, S.; Lin, Q. Comprehensive profiling of lysine acetylproteome analysis reveals diverse functions of lysine acetylation in common wheat. **2016**, *6*, 1–10.
(18) Carlson, S. M.; Moore, K. E.; Green, E. M.; Martin, G. M.; Gozani, O. Proteome-wide enrichment of proteins modified by lysine methylation. *Nat. Protoc.* **2013**, *9* (1), 37–50.
(19) Guo, A.; Gu, H.; Zhou, J.; Mulhern, D.; Wang, Y.; Lee, K. A.; Yang, V.; Aguiar, M.; Kornhauser, J.; Jia, X.; et al. Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. *Mol. Cell. Proteomics* **2014**, *13* (1), 372–387.
(20) Wang, K.; Dong, M.; Mao, J.; Wang, Y.; Jin, Y.; Ye, M.; Zou, H. Antibody-Free Approach for the Global Analysis of Protein Methylation. *Anal. Chem.* **2016**, *88* (23), 11319–11327.
(21) Altelaar, A. F. M.; Munoz, J.; Heck, A. J. R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **2012**, *14* (1), 35–48.
(22) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The One Hour Yeast Proteome. *Mol. Cell. Proteomics* **2014**, *13* (1), 339–347.
(23) Bern, M.; Kil, Y. J.; Becker, C. Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinforma.* **2012**, *Chapter 13*, Unit13.20.
(24) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10* (4), 1794–1805.
(25) Frese, C. K.; Altelaar, A. F. M.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* **2011**, *10* (5), 2377–2388.
(26) Shen, Y.; Tolić, N.; Xie, F.; Zhao, R.; Purvine, S. O.; Schepmoes, A. A.; Moore, R., J.; Anderson, G. A.; Smith, R. D. Effectiveness of CID, HCD, and ETD with FT MS/MS for Degradomic-Peptidomic Analysis: Comparison of Peptide Identification Methods. *J. Proteome Res.* **2011**, *10* (9), 3929–3943.
(27) Thakur, S. S.; Geiger, T.; Chatterjee, B.; Bandilla, P.; Fröhlich, F.; Cox, J.; Mann, M. Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation. *Mol. Cell. Proteomics* **2011**, *10* (8), M110.003699.
(28) Cristobal, A.; Hennrich, M. L.; Giansanti, P.; Goerdayal, S. S.; Heck, A. J. R.; Mohammed, S. In-house construction of a UHPLC system enabling the

identification of over 4000 protein groups in a single analysis. *The Analyst* **2012**, *137* (15), 3541.

(29) Kay, R. G.; Gregory, B.; Grace, P. B.; Pleasance, S. The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum. *Rapid Commun. Mass Spectrom.* **2007**, *21* (16), 2585–2593.

(30) Jorgenson, J. W. Capillary Liquid Chromatography at Ultrahigh Pressures. *Annu. Rev. Anal. Chem.* **2010**, *3* (1), 129–150.

(31) Michalski, A.; Cox, J.; Mann, M. More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC−MS/MS. *J. Proteome Res.* **2011**, *10* (4), 1785–1793.

(32) Xu, P.; Duong, D. M.; Peng, J. Systematical Optimization of Reverse-Phase Chromatography for Shotgun Proteomics. *J. Proteome Res.* **2009**, *8* (8), 3944–3950.

(33) Ducret, A.; Oostveen, I. V.; Eng, J. K.; Yates, J. R.; Aebersold, R. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci.* **1998**, *7* (3), 706–719.

(34) Mahoney, W. C.; Hermodson, M. A. Separation of large denatured peptides by reverse phase high performance liquid chromatography. Trifluoroacetic acid as a peptide solvent. *J. Biol. Chem.* **1980**, *255* (23), 11199–11203.

(35) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Anal. Bioanal. Chem.* **2008**, *391* (1), 151–159.

(36) McNulty, D. E.; Annan, R. S. Hydrophilic Interaction Chromatography Reduces the Complexity of the Phosphoproteome and Improves Global Phosphopeptide Isolation and Detection. *Mol. Cell. Proteomics* **2008**, *7* (5), 971–980.

(37) Palmisano, G.; Lendal, S. E.; Engholm-Keller, K.; Leth-Larsen, R.; Parker, B. L.; Larsen, M. R. Selective enrichment of sialic acid–containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **2010**, *5* (12), 1974–1982.

(38) Mysling, S.; Palmisano, G.; Højrup, P.; Thaysen-Andersen, M. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Anal. Chem.* **2010**, *82* (13), 5598–5609.

(39) Neue, K.; Mormann, M.; Peter-Katalinić, J.; Pohlentz, G. Elucidation of Glycoprotein Structures by Unspecific Proteolysis and Direct nanoESI Mass Spectrometric Analysis of ZIC-HILIC-Enriched Glycopeptides. *J. Proteome Res.* **2011**, *10* (5), 2248–2260.

(40) Alpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A* **1990**, *499*, 177–196.

(41) Washburn, M. P.; Wolters, D.; Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–247.
(42) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17* (7), 676–682.
(43) Zhu, M.-Z.; Li, N.; Wang, Y.-T.; Liu, N.; Guo, M.-Q.; Sun, B.; Zhou, H.; Liu, L.; Wu, J.-L. Acid/Salt/pH Gradient Improved Resolution and Sensitivity in Proteomics Study Using 2D SCX-RP LC–MS. *J. Proteome Res.* **2017**, *16* (9), 3470–3475.
(44) Zhang, L.; Liu, C.-W.; Zhang, Q. Online 2D-LC-MS/MS Platform for Analysis of Glycated Proteome. *Anal. Chem.* **2017**.
(45) Wang, S.; Shi, X.; Xu, G. Online Three Dimensional Liquid Chromatography/Mass Spectrometry Method for the Separation of Complex Samples. *Anal. Chem.* **2017**, *89* (3), 1433–1438.
(46) Navarro-Reig, M.; Jaumot, J.; Baglai, A.; Vivó-Truyols, G.; Schoenmakers, P. J.; Tauler, R. Untargeted Comprehensive Two-Dimensional Liquid Chromatography Coupled with High-Resolution Mass Spectrometry Analysis of Rice Metabolome Using Multivariate Curve Resolution. *Anal. Chem.* **2017**, *89* (14), 7675–7683.
(47) Li, D.; Jakob, C.; Schmitz, O. Practical considerations in comprehensive two-dimensional liquid chromatography systems (LCxLC) with reversed-phases in both dimensions. *Anal. Bioanal. Chem.* **2015**, *407* (1), 153–167.
(48) Mann, M.; Kelleher, N. L. Precision proteomics: The case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci.* **2008**, *105* (47), 18132–18138.
(49) Capriotti, A. L.; Cavaliere, C.; Foglia, P.; Samperi, R.; Laganà, A. Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics. *J. Chromatogr. A* **2011**, *1218* (49), 8760–8776.
(50) Shen, Y.; Tolić, N.; Piehowski, P. D.; Shukla, A. K.; Kim, S.; Zhao, R.; Qu, Y.; Robinson, E.; Smith, R. D.; Paša-Tolić, L. High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics. *J. Chromatogr. A* **2017**, *1498*, 99–110.
(51) Aguilar, M. I.; Hearn, M. T. High-resolution reversed-phase high-performance liquid chromatography of peptides and proteins. *Methods Enzymol.* **1996**, *270*, 3–26.
(52) Riley, N. M.; Mullen, C.; Weisbrod, C. R.; Sharma, S.; Senko, M. W.; Zabrouskov, V.; Westphall, M. S.; Syka, J. E. P.; Coon, J. J. Enhanced Dissociation of Intact Proteins with High Capacity Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (3), 520–531.
(53) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9* (1), 499–519.
(54) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down proteomics: Facts and perspectives. *Biochem. Biophys. Res. Commun.* **2014**, *445* (4), 683–693.

(55) Smith, L. M.; Kelleher, N. L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla, L.; Kruppa, G.; Zubarev, R.; et al. Proteoform: a single term describing protein complexity. *Nat. Methods* **2013**, *10* (3), 186–187.
(56) Huang, T.; Wang, J.; Yu, W.; He, Z. Protein inference: a review. *Brief. Bioinform.* **2012**, *13* (5), 586–614.
(57) Armirotti, A.; Damonte, G. Achievements and perspectives of top-down proteomics. *PROTEOMICS* **2010**, *10* (20), 3566–3576.
(58) Tian, Z.; Tolić, N.; Zhao, R.; Moore, R. J.; Hengel, S. M.; Robinson, E. W.; Stenoien, D. L.; Wu, S.; Smith, R. D.; Paša-Tolić, L. Enhanced top-down characterization of histone post-translational modifications. *Genome Biol.* **2012**, *13* (10), R86.
(59) Moradian, A.; Kalli, A.; Sweredoski, M. J.; Hess, S. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *PROTEOMICS* **2014**, *14* (4–5), 489–497.
(60) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. The C-Score: A Bayesian Framework to Sharply Improve Proteoform Scoring in High-Throughput Top Down Proteomics. *J. Proteome Res.* **2014**, *13* (7), 3231–3240.
(61) Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W. P.; Mohammed, S.; Heck, A. J. R. Toward an Optimized Workflow for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89* (6), 3318–3325.
(62) Sidoli, S.; Garcia, B. A. Middle-down proteomics: a still unexploited resource for chromatin biology. *Expert Rev. Proteomics* **2017**, *14* (7), 617–626.
(63) Schräder, C. U.; Lee, L.; Rey, M.; Sarpe, V.; Man, P.; Sharma, S.; Zabrouskov, V.; Larsen, B.; Schriemer, D. C. Neprosin, a Selective Prolyl Endoprotease for Bottom-up Proteomics and Histone Mapping. *Mol. Cell. Proteomics* **2017**, *16* (6), 1162–1171.
(64) Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C. High-Throughput Middle-Down Analysis Using an Orbitrap. *J. Proteome Res.* **2010**, *9* (8), 3886–3890.
(65) Sidoli, S.; Schwämmle, V.; Ruminowicz, C.; Hansen, T. A.; Wu, X.; Helin, K.; Jensen, O. N. Middle-down hybrid chromatography/tandem mass spectrometry workflow for characterization of combinatorial post-translational modifications in histones. *PROTEOMICS* **2014**, *14* (19), 2200–2211.
(66) Busman, M.; Schey, K. L.; Oatis, J. E.; Knapp, D. R. Identification of phosphorylation sites in phosphopeptides by positive and negative mode electrospray ionization-tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1996**, *7* (3), 243–249.
(67) Robinson, M. R.; Moore, K. L.; Brodbelt, J. S. Direct Identification of Tyrosine Sulfation by using Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (8), 1461–1471.

(68) Doerr, A. Proteomics: Navigating the negative-mode proteome. *Nat. Methods* **2015**, *12* (9), 808–808.
(69) Riley, N. M.; Rush, M. J. P.; Rose, C. M.; Richards, A. L.; Kwiecien, N. W.; Bailey, D. J.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. The Negative Mode Proteome with Activated Ion Negative Electron Transfer Dissociation (AI-NETD). *Mol. Cell. Proteomics MCP* **2015**, *14* (10), 2644–2660.
(70) 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. High-throughput Database Search and Large-scale Negative Polarity Liquid Chromatography-Tandem Mass Spectrometry with Ultraviolet Photodissociation for Complex Proteomic Samples. *Mol. Cell. Proteomics* **2013**, *12* (9), 2604–2614.
(71) Yamashita, M.; Fenn, J. B. Negative ion production with the electrospray ion source. *J. Phys. Chem.* **1984**, *88* (20), 4671–4675.
(72) McClory, P. J.; Håkansson, K. Corona Discharge Suppression in Negative Ion Mode Nanoelectrospray Ionization via Trifluoroethanol Addition. *Anal. Chem.* **2017**, *89* (19), 10188–10193.
(73) Madsen, J. A.; Ko, B. J.; Xu, H.; Iwashkiw, J. A.; Robotham, S. A.; Shaw, J. B.; Feldman, M. F.; Brodbelt, J. S. Concurrent Automated Sequencing of the Glycan and Peptide Portions of *O* -Linked Glycopeptide Anions by Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (19), 9253–9261.
(74) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. *Anal. Chem.* **2016**, *88* (4), 2303–2310.
(75) 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. Analysis of the Acidic Proteome with Negative Electron-Transfer Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (6), 2875–2882.
(76) Ewing, N. P.; Cassady, C. J. Dissociation of multiply charged negative ions for hirudin (54–65), fibrinopeptide B, and insulin A (oxidized). *J. Am. Soc. Mass Spectrom.* **2001**, *12* (1), 105–116.
(77) Griffiths, W. J.; Wang, Y. Mass spectrometry: from proteomics to metabolomics and lipidomics. *Chem. Soc. Rev.* **2009**, *38* (7), 1882.
(78) Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. Electron transfer dissociation of peptide anions. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (6), 880–882.
(79) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88* (1), 30–51.
(80) Shukla, A. K.; Futrell, J. H. Tandem mass spectrometry: dissociation of ions by collisional activation. *J. Mass Spectrom.* **2000**, *35* (9), 1069–1090.
(81) 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. Complete Protein

Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.
(82) Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L. Analyzing Internal Fragmentation of Electrosprayed Ubiquitin Ions During Beam-Type Collisional Dissociation. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (5), 782–787.
(83) Zhang, Z.; Wu, S.; Stenoien, D. L.; Paša-Tolić, L. High-Throughput Proteomics. *Annu. Rev. Anal. Chem.* **2014**, *7* (1), 427–454.
(84) Brodbelt, J. S. Photodissociation mass spectrometry: new tools for characterization of biological molecules. *Chem Soc Rev* **2014**, *43* (8), 2757–2783.
(85) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell. Proteomics* **2007**, *6* (11), 1942–1951.
(86) Diedrich, J. K.; Pinto, A. F. M.; Yates, J. R. Energy Dependence of HCD on Peptide Fragmentation: Stepped Collisional Energy Finds the Sweet Spot. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (11), 1690–1699.
(87) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000**, *35* (12), 1399–1406.
(88) Bythell, B. J.; Suhai, S.; Somogyi, Á.; Paizs, B. Proton-Driven Amide Bond-Cleavage Pathways of Gas-Phase Peptide Ions Lacking Mobile Protons. *J. Am. Chem. Soc.* **2009**, *131* (39), 14057–14065.
(89) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J. Am. Chem. Soc.* **1996**, *118* (35), 8365–8374.
(90) Laskin, J.; Kong, R. P. W.; Song, T.; Chu, I. K. Effect of the basic residue on the energetics and dynamics of dissociation of phosphopeptides. *Int. J. Mass Spectrom.* **2012**, *330–332*, 295–301.
(91) Douglas, D. J.; Frank, A. J.; Mao, D. Linear ion traps in mass spectrometry. *Mass Spectrom. Rev.* **2005**, *24* (1), 1–29.
(92) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; et al. A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Mol. Cell. Proteomics* **2009**, *8* (12), 2759–2769.
(93) March, R. E.; Todd, J. F. *Quadrupole Ion Trap Mass Spectrometry*; John Wiley & Sons, 2005.
(94) Bowie, J. H.; Brinkworth, C. S.; Dua, S. Collision-induced fragmentations of the (M-H)- parent anions of underivatized peptides: An aid to structure determination and some unusual negative ion cleavages. *Mass Spectrom. Rev.* **2002**, *21* (2), 87–107.
(95) Bokatzian-Johnson, S. S.; Stover, M. L.; Dixon, D. A.; Cassady, C. J. A Comparison of the Effects of Amide and Acid Groups at the C-Terminus on the

Collision-Induced Dissociation of Deprotonated Peptides. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (9), 1544–1557.
(96) Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. Influence of Secondary Structure on the Fragmentation of Protonated Peptides. *J. Am. Chem. Soc.* **1999**, *121* (22), 5142–5154.
(97) Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R. Influence of Basic Residue Content on Fragment Ion Peak Intensities in Low-Energy Collision-Induced Dissociation Spectra of Peptides. *Anal. Chem.* **2004**, *76* (5), 1243–1248.
(98) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci.* **2007**, *104* (7), 2193–2198.
(99) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, Ü. A.; Tsybin, Y. O. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem. Soc. Rev.* **2013**, *42* (12), 5014.
(100) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. *Anal. Chem.* **2017**.
(101) Shaw, J. B.; Kaplan, D. A.; Brodbelt, J. S. Activated Ion Negative Electron Transfer Dissociation of Multiply Charged Peptide Anions. *Anal. Chem.* **2013**, *85* (9), 4721–4728.
(102) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors. *Anal. Chem.* **2007**, *79* (2), 477–485.
(103) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal. Chem.* **2015**, *87* (14), 7109–7116.
(104) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. *J. Proteome Res.* **2017**, *16* (7), 2653–2659.
(105) Ledvina, A. R.; Rose, C. M.; McAlister, G. C.; Syka, J. E. P.; Westphall, M. S.; Griep-Raming, J.; Schwartz, J. C.; Coon, J. J. Activated Ion ETD Performed in a Modified Collision Cell on a Hybrid QLT-Oribtrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (11), 1623–1633.
(106) Reilly, J. P. Ultraviolet photofragmentation of biomolecular ions. *Mass Spectrom. Rev.* **2009**, *28* (3), 425–447.
(107) R. Julian, R. The Mechanism Behind Top-Down UVPD Experiments: Making Sense of Apparent Contradictions. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (9), 1823–1826.

(108) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 nm and its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.
(109) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-nm photodissociation of singly and multiply charged peptide anions for acidic proteome characterization. *PROTEOMICS* **2011**, *11* (7), 1329–1334.
(110) Greer, S. M.; Holden, D. D.; Fellers, R.; Kelleher, N. L.; Brodbelt, J. S. Modulation of Protein Fragmentation Through Carbamylation of Primary Amines. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (8), 1587–1599.
(111) Lesur, A.; Domon, B. Advances in high-resolution accurate mass spectrometry application to targeted proteomics. *PROTEOMICS* **2015**, *15* (5–6), 880–890.
(112) Hendrickson, C. L.; Quinn, J. P.; Kaiser, N. K.; Smith, D. F.; Blakney, G. T.; Chen, T.; Marshall, A. G.; Weisbrod, C. R.; Beu, S. C. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (9), 1626–1632.
(113) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **2015**, *8* (1), 61–80.
(114) Makarov, A.; Denisov, E.; Lange, O. Performance evaluation of a high-field orbitrap mass analyzer. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (8), 1391–1396.
(115) Denisov, E.; Damoc, E.; Lange, O.; Makarov, A. Orbitrap mass spectrometry with resolving powers above 1,000,000. *Int. J. Mass Spectrom.* **2012**, *325–327*, 80–85.
(116) Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra. *J. Proteome Res.* **2014**, *13* (8), 3679–3684.
(117) Brosch, M.; Swamy, S.; Hubbard, T.; Choudhary, J. Comparison of Mascot and X!Tandem Performance for Low and High Accuracy Mass Spectrometry and the Development of an Adjusted Mascot Threshold. *Mol. Cell. Proteomics MCP* **2008**, *7* (5), 962–970.
(118) Branca, R. M. M.; Orre, L. M.; Johansson, H. J.; Granholm, V.; Huss, M.; Pérez-Bercoff, Å.; Forshed, J.; Käll, L.; Lehtiö, J. HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat. Methods* **2014**, *11* (1), 59.
(119) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; et al. Informed-Proteomics: open-source software package for top-down proteomics. *Nat. Methods* **2017**, *14* (9), 909–914.
(120) Granholm, V.; Kim, S.; Navarro, J. C. F.; Sjölund, E.; Smith, R. D.; Käll, L. Fast and Accurate Database Searches with MS-GF+Percolator. *J. Proteome Res.* **2014**, *13* (2), 890–897.
(121) Wenger, C. D.; Coon, J. J. A Proteomics Search Algorithm Specifically Designed for High-Resolution Tandem Mass Spectra. *J. Proteome Res.* **2013**, *12* (3), 1377–1386.

(122) Chapman, J. D.; Goodlett, D. R.; Masselon, C. D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom. Rev.* **2014**, *33* (6), 452–470.
(123) Egertson, J. D.; Kuehn, A.; Merrihew, G. E.; Bateman, N. W.; MacLean, B. X.; Ting, Y. S.; Canterbury, J. D.; Marsh, D. M.; Kellmann, M.; Zabrouskov, V.; et al. Multiplexed MS/MS for improved data-independent acquisition. *Nat. Methods* **2013**, *10* (8), 744–746.
(124) Hu, A.; Noble, W. S.; Wolf-Yadlin, A. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Research* **2016**, *5*, 419.
(125) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012**, *11* (6), O111.016717.
(126) Lambert, J.-P.; Ivosev, G.; Couzens, A. L.; Larsen, B.; Taipale, M.; Lin, Z.-Y.; Zhong, Q.; Lindquist, S.; Vidal, M.; Aebersold, R.; et al. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nat. Methods* **2013**, *10* (12), 1239–1245.
(127) Elias, J. E.; Gygi, S. P. Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. *Methods Mol. Biol. Clifton NJ* **2010**, *604*, 55–71.
(128) Medzihradszky, K. F.; Chalkley, R. J. Lessons in de novo peptide sequencing by tandem mass spectrometry. *Mass Spectrom. Rev.* **2015**, *34* (1), 43–63.
(129) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
(130) Klammer, A. A.; Park, C. Y.; Noble, W. S. Statistical Calibration of the SEQUEST XCorr Function. *J. Proteome Res.* **2009**, *8* (4), 2106–2113.
(131) Carvalho, P. C.; Fischer, J. S. G.; Xu, T.; Cociorva, D.; Balbuena, T. S.; Valente, R. H.; Perales, J.; Yates, J. R.; Barbosa, V. C. Search Engine Processor: filtering and organizing PSMs. *Proteomics* **2012**, *12* (7), 944–949.
(132) Cociorva, D.; L. Tabb, D.; Yates, J. R. Validation of Tandem Mass Spectrometry Database Search Results Using DTASelect. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc., 2002.
(133) Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S. Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. *J. Proteome Res.* **2008**, *7* (1), 29–34.
(134) Choi, H.; Nesvizhskii, A. I. False Discovery Rates and Related Statistical Concepts in Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2008**, *7* (1), 47–50.
(135) Reiter, L.; Claassen, M.; Schrimpf, S. P.; Jovanovic, M.; Schmidt, A.; Buhmann, J. M.; Hengartner, M. O.; Aebersold, R. Protein Identification False Discovery Rates for Very Large Proteomics Data Sets Generated by Tandem Mass Spectrometry. *Mol. Cell. Proteomics* **2009**, *8* (11), 2405–2417.

(136) Aggarwal, S.; Yadav, A. K. False Discovery Rate Estimation in Proteomics. *Methods Mol. Biol. Clifton NJ* **2016**, *1362*, 119–128.
(137) Matthiesen, R. *Mass Spectrometry Data Analysis in Proteomics*; Springer Science & Business Media, 2007.

Chapter 2

Experimental Methods

2.1 Mass Spectrometry

The work in this dissertation involved studying large peptides and proteins requiring high mass accuracy and resolving power for analysis. High resolution mass analyses were performed on a Thermo Fisher Scientific Orbitrap Elite (hybrid linear ion trap/Orbitrap mass spectrometer) or a Thermo Fisher Scientific Fusion Lumos (hybrid quadrupole /linear ion trap/Orbitrap mass spectrometer). The Orbitrap mass analyzer was introduced by Makarov in 2000 and has revolutionized modern proteomic and mass spectrometry workflows.[1] Orbitrap instruments have many advantages compared to traditional ion trap mass spectrometers due to their high mass accuracy, speed, high resolution.[2]

The Orbitrap analyzer essentially consists of an inner rod-like electrode and an outer elliptical shaped hollow cylinder electrode separated by a high vacuum region where the ions are trapped. The Orbitrap operates by trapping ions and allowing them to orbit around the DC only rod-shaped "spindle electrode." Prior to analysis in the Orbitrap assembly, ions are initially held in a curved linear ion trap which accumulates and bunches

the ions. Once a sufficient number of ions are accumulated within the trap, the ion packet is transferred to the Orbitrap via several ion optics by reducing the RF voltages and applying DC gradients to the curved linear ion trap. Ions are measured based on their frequencies as they oscillate along the length of the Orbitrap spindle electrode. The oscillations of the ions being sent back and forth across the center electrode is detected by the split outer electrode which records an image current. The image current can be converted into a mass spectrum because the frequency of the oscillations is proportional to the (*m/z*) of the ion which produced the current. The resolution of the ions is dependent on the number of oscillations recorded within the image current.[3] This simple but sophisticated mass analyzer offers ultra high resolution and high mass accuracy.

2.2 ELECTROSPRAY IONIZATION (ESI)

Electrospray ionization is a soft ionization technique which induces little to no fragmentation during ionization. [4] This process ionizes an analyte dissolved in solvent by applying a potential (500-5000V) on a capillary through which the solution is sprayed. Upon achieving electrospray, the ionized solvent is aerosolized upon exiting the narrow orifice of the capillary. The solvent is evaporated from the plume, leaving charged ions which are detected in the mass spectrometer.[5] Electrospray ionization (ESI), and nanoESI (nESI), were used in the following chapters, depending on the flow rate of sample infusion, and required sensitivity of the measurement.

2.3 LIQUID CHROMATOGRAPHY

For complex proteomic samples such as protein extracts from cell lysates, a separation step is required prior to mass spectrometric analysis.[6] This step reduces signal dilution by separating the complex mixture in time, which reduces the number of species analyzed simultaneously, and increases the detection of many less abundant analytes. Most commonly, peptide/protein mixtures are separated via RP-HPLC, which is then easily coupled online to ESI-MS/MS instrumentation. In order to take advantage of chemical features unique to a certain class of protein or peptide more specialized approaches are used such as hydrophilic interaction liquid chromatography (HILIC). The LC instrumentation and parameters that were utilized in the chapters herein are described in the subsequent sections.

2.3.1 Bottom-Up Proteomics

For the work presented in this dissertation, chromatographic separations were performed using water (A) and acetonitrile (B) mobile phases containing 0.05% acetic acid on an Eksigent Nanoultra 2D Plus nano liquid chromatography system (Redwood,CA). Both trap (35 mm × 0.1 mm) and analytical columns with an integrated emitter (15 cm × 0.075 cm) were packed in house using 3 μm Michrom Magic C18 packing (New Objective, Woburn, MA). Approximately 1 μg of digest was loaded onto the trap column at 2 μl/min for 20 min, then separated with a gradient that changed from 0 to 35% B over the course of 240 minutes. For nanospray, 1.8 kV was applied at a precolumn liquid voltage junction. Alternatively, chromatographic separations were performed using a Dionex RSLC 3000

nanobore LC system with water (A) and acetonitrile (B) as mobile phases, with each containing 0.1% formic acid. The trap (0.075 cm x 3.5 cm) and analytical column (0.075 cm × 20 cm, with integrated emitter) were packed in-house using 3.5 µm XBridge BEH C18 media (Waters, Milford, MA). Approximately 1 µg of digest was loaded onto the trap column at 5 µL/min for 5 min, then separated on the analytical column with a gradient that changed from 0 to 35% B over the course of 240 min at a flow rate of 300 nL min^{-1}. For nanospray, 1.8 kV was applied at a pre-column liquid voltage junction.

2.3.2 Middle-down Proteomics

Histone middle -down sized peptides were resuspended in 2% ACN and separated using a Dionex RSLC 3000 nano-LC system (Thermo Fisher Scientific, San Jose, CA, USA) according to the method of Young *et al.*[7] Approximately 1 µg of peptides was injected onto a 3 cm REPROSIL Gold (3 µm particles, 300 Å pore size, Dr. Maisch Germany) C18 reverse phase trapping column (100 µm i.d.). Peptides were then transferred onto a 20 cm fritted (75 µm i.d.). pulled tip analytical column (New Objective, Woburn, MA) packed in-house with PolyCAT A (Poly LC, Columbia, MD), a weak cation exchange hydrophilic interaction chromatography (WCX-HILIC) media. Peptides were eluted at a flow rate of 300 nL/min using the following gradient: starting at 2% B for 20 minutes, going to 55% B at 23 minutes, then to 90% B at 160 min, and finally to 99% B at 170 min. Mobile phase A was 75% acetonitrile 20 mM propionic acid (pH 6). Mobile phase B was 75% water with formic acid (pH 2.5).

2.3.3 Intact Proteins

Proteins were separated using a Dionex RSLC 3000 nano-liquid chromatograph (Thermo Fisher, San Jose, CA) Approximately 1 µg of proteins were injected onto a 3 cm PLRP reverse phase trapping column (75 µm i.d.) packed with 5 µm particles (1000 Å pore size). Proteins were then eluted onto a 40 cm fritted 75 µm i.d. pulled tip analytical column (New Objective, Woburn MA) packed in-house with PLRP (5 µm particles, 1000 Å pore size) at a flow rate of 300 nL/min using a linear gradient of 2%-50% solvent B (acetonitrile/0.1% formic acid) over 120 minutes. Solvent A was water/0.1% formic acid.

2.4 PEPTIDE AND PROTEIN PREPARATIONS

The model peptides DRVYIHPFHL and WAGGDASGE were obtained from American Peptide Company (Sunnyvale, CA). Bovine serum albumin was obtained from Sigma-Aldrich (St. Louis, MO). *Halobacterium salinarum* was obtained from American Type Culture Collection (ATCC, Manassas, VA). The bacteria were grown in the recommended medium (American Type Culture Collection medium 2185). Cells were suspended in 10 mM Tris-HCL, 10 mM KCl, 1.5 mM $MgCl_2$ at pH 8 to swell and were lysed by dounce homogenization. The whole cell lysate was centrifuged to clarify the soluble lysate and to remove the insoluble pellet.

Both bovine serum albumin and proteins isolated from *H. salinarum* were digested at 37 °C overnight with trypsin. Prior to digestion, proteins were reduced in 5 mM dithiothreitol for 30 min at 55 °C and subsequently alkylated in 10 mM iodoacetamide at room temperature in the dark for 30 min. Alkylation was quenched with a second aliquot of dithiothreitol, thus bringing the final concentration of dithiothreitol to ~10 mM. Trypsin was added to achieve a 1:20 enzyme-to-substrate ratio, and the solution was buffered at pH 8 in 150 mM ammonium bicarbonate. After digestion the sample was dried in a vacuum centrifuge for subsequent derivatization.

Ubiquitin (bovine), cytochrome c (equine), myoglobin (bovine), superoxide dismutase (bovine), lysozyme (galline), carbonic anhydrase (bovine), and urea were obtained from Sigma-Aldrich (St. Louis, MO). *E. coli* 70S ribosome was obtained from New England Biolabs (Ipswich, MA). Solvents were obtained from EMD Millipore (Billerica, MA). Tris(2-carboxyethyl) phosphine (TCEP) was obtained from Thermo-Scientific (Rochford, IL)

HeLa S3 cells were treated for 24 hrs with 10 mM sodium butyrate and harvested. Histones were extracted as previously described. [8] In summary, nuclei were isolated after resuspending the cell pellets in nuclei isolation buffer (250 mM sucrose, 0.2%NP-40, 1 mM $CaCL_2$, 15 mM Tris-HCl pH 7.5, 15 mM NaCl, 60 mM KCL, and 5 mM $MgCl_2$). Nuclei were pelleted and resuspended in 0.4 N H_2SO_4 at a 5:1 ratio (v/v) and incubated for 2 hrs at 4 ºC with shaking. After acid extraction, histones were precipitated with 25% TCA (w/v).

Purified histones (~300 µg) were separated by RP-HPLC as previously described.[8] Briefly, histones were fractionated on a Vydac C18 column (10 mm inner diameter, 250 mm length, 5um particle size). Histones were eluted over a 100 minute gradient from 30% to 60% solvent B at a flow rate of 1 mL/min. Solvent A: 2% triflouroacetic acid and 5% acetonitrile in water, B: 0.19% triflouroacetic acid and 95% acetonitrile in water. The UV detector was adjusted to 214 nm, and fractions for H4, H2A, H2B, H3, H3.3, H3.2 and H3.1 were collected base on their characteristic retention times.[8] Fractions were dried using a SpeedVac concentrator and store at -20 ºC . Finally, the isolated histones H3 and H4 were submitted to GluC digestion for 8 hrs in 50 mM ammonium acetate buffer (pH 4) prior to LC-MS analysis.

2.4.1 Peptide and Protein Carbamylation

Carbamylation was performed as previously reported.[9] Briefly, each sample was split into two aliquots; one for derivatization and one as a control. Each was suspended in 200 mM Tris-HCl in the presence or absence of 8 M urea. Both samples were incubated at 80 °C for 4 h. Samples were desalted using Amicon Ultra 3kDa MWCO spin columns (EMD Millipore; Billerica, MA), then evaporated to dryness and resuspended in solvent to match the LC starting conditions (2% acetonitrile/98% water/0.1% formic acid) or infusion conditions (50% methanol/50% water/1% formic acid.

2.5 DATA ANALYSIS

2.5.1 SEQUEST

The following parameters were used for searching LC-MS RAW peptide files in SEQUEST.[10] Proteome Discoverer version 1.4.1.14 was used. Database searching was performed using the SEQUEST HT. Tryptic enzyme specificity was selected allowing up to 2 missed cleavages. A maximum delta Cn was set to 0.05. A precursor mass tolerance of 7 ppm and fragment mass tolerance of 0.02 Da was used. The following dynamic modifications were used: acetylation of N-termini, deamidation (+0.984) of asparagine, pyroglutamic acid (-17.027 Da) of glutamine, pyroglutamic acid (-18.011 Da) of aspartic acid. Carbamidomethylation (+57.021 Da) of cysteine was treated as a static modification. Result filtering was performed using Percolator with the following parameters: Max Delta Cn: 0.05, target FDR: 0.01 based on q-value.

The above search was modified to include the ptmRS node for improved phosphopeptide searching. Phosphorylation (+79.966 Da) was added as a variable modification on S and T.

2.5.2 MassMatrix

The following parameters were used for searching NUVPD spectra in MassMatrix.[11] Trypsin was selected as the digestion method, and the fragmentation mode was set to UVPD. The following dynamic modifications were selected: acetylation of N-termini, deamidation of asparagine, pyroglutamic acid at glutamic acid, pyroglutamic acid

(-17.027 Da) of glutamine. Iodoacetamide derivatization (carbamidomethyl) of cysteine was set as a fixed modification. The maximum missed cleavage was set to 2, the precursor mass tolerance was set to 7 ppm, and the default fragment mass tolerance of 0.05 Da was used. Minimum score of output was set to 2, minimum pp value and pp2 value was set to 4.3. The minimum pp tag was set to 4.0, and the maximum number of PTMs was set to 4; score of output was set to 2, minimum pp value and pp2 value was set to 4.3. The minimum pp tag was set to 4.0, and the maximum number of PTMs was set to 4.

2.5.3 Byonic

The following parameters were used for searching the LC-MS RAW peptide files in Byonic: add decoys was selected.[12] Cleavage Sites were set to RK and cleavage side was C-terminal. Digestion Specificity was set to Fully Specific and missed cleavages were set to 2. A precursor mass tolerance of 7 ppm was used and fragmentation type: UVPD/HCD/NUVPD were selected where appropriate. A 15 ppm fragment ion tolerance was used. The following modifications were searched for: carbamidomethyl (+57.021464) fixed at cysteine, variable deamidation (+0.984016) of asparagine (common), variable pyroglutamic acid (-17.026549) of glutamine (rare), variable pyroglutamic acid of aspartic acid (-18.010565) (rare), variable acetylation (+42.010565) at protein N-termini (rare). The maximum number of precursors per scan was set to 2, and FDR was set to 1% FDR. Unlike SEQUEST and MassMatrix, Byonic uses a "protein aware FDR". This means that candidate peptides from proteins with many peptides already identified receive a preferential score compared to ones without any protein level evidence.

The above search parameters were modified to include variable phosphorylation (+79.966331) of serine and threonine (rare), when searching for phosphopeptides.

2.5.4 ProSight Lite

ProSight Lite was used to search middle-down peptide and intact protein data files. Prior to analysis in ProSight Lite (Build 1.4.6) several scans were averaged to improve the S/N of fragment ions.[13] The resulting spectra was deconvoluted using the Xtract algorithm available in the Xcalibur Qualbrowser (Thermo Fisher Scientific, San Jose CA) software. Monoisotopic output was selected and the S/N level was set to 3, all other parameters were left to default. The resulting deconvoluted peak list was input into the ProSight Lite software. The canonical H3 or H4 sequences (N-terminal GluC peptide) was imported into ProSight Lite. Monoisotopic input and UVPD or ETD were selected as the fragment type and a 10-ppm tolerance was applied. Choice of PTM location was guided by intact mass, previously reported sites, and primarily the following metrics: P-score, number of matched fragments, and sequence coverage.

2.6 REFERENCES

(1) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **2015**, *8* (1), 61–80.

(2) 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.; et al. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal. Chem.* **2013**, *85* (24), 11710–11714.

(3) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72* (6), 1156–1162.
(4) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization-principles and practice. *Mass Spectrom. Rev.* **1990**, *9* (1), 37–70.
(5) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9.
(6) Karpievitch, Y. V.; Polpitiya, A. D.; Anderson, G. A.; Smith, R. D.; Dabney, A. R. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. *Ann. Appl. Stat.* **2010**, *4* (4), 1797–1823.
(7) Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A. High Throughput Characterization of Combinatorial Histone Codes. *Mol. Cell. Proteomics* **2009**, *8* (10), 2266–2284.
(8) Lin, S.; Garcia, B. A. Chapter One - Examining Histone Posttranslational Modification Patterns by High-Resolution Mass Spectrometry. In *Methods in Enzymology*; Wu, C., Allis, C. D., Eds.; Nucleosomes, Histones & Chromatin Part A; Academic Press, 2012; Vol. 512, pp 3–28.
(9) Angel, P. M.; Orlando, R. Quantitative carbamylation as a stable isotopic labeling method for comparative proteomics. *Rapid Commun. Mass Spectrom. RCM* **2007**, *21* (10), 1623–1634.
(10) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
(11) Xu, H.; Freitas, M. A. A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data. *BMC Bioinformatics* **2007**, *8* (1), 133–143.
(12) Bern, M.; Kil, Y. J.; Becker, C. Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinforma.* **2012**, *Chapter 13*, Unit13.20.
(13) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *PROTEOMICS* **2015**, *15* (7), 1235–1238.

Chapter 3

Improvement of Shotgun Proteomics in the Negative Mode by Carbamylation of Peptides and Ultraviolet Photodissociation Mass Spectrometry

3.1 Overview

Although acidic peptides compose a substantial portion of many proteomes, their less efficient ionization during positive polarity electrospray ionization (ESI) impedes their detection in bottom-up mass spectrometry workflows. We have implemented a derivatization strategy based on carbamylation which converts basic amine sites (Lys, N-termini) to less basic amides for enhanced analysis in the negative mode. Ultraviolet photodissociation (UVPD) is used to analyze the resulting peptide anions, as demonstrated for tryptic peptides from bovine serum albumin and *Halobacterium salinarum* in a high throughput liquid chromatography/tandem mass spectrometry (LC/MS/MS) mode. LC/UVPD-MS of a carbamylated *H. salinarum* digest resulted in 45% more identified peptides and 25% more proteins compared to the unmodified digest analyzed in the negative mode.

3.2 Introduction

With the advent of modern mass spectrometric-based proteomics, hundreds or even thousands of proteins can be identified in a single experiment.[1] Several hurdles still remain in the path to efficient sampling of a complete proteome in high-throughput applications. Chief among these hurdles is identification of underrepresented proteins (based on analysis

of the corresponding peptides created upon proteolysis in the typical bottom-up approach). Underrepresented proteins include those that have fewer copies per cell (low abundance) as well as those for which the proteolytic peptides are undersampled due to a variety of factors. These factors include low protein solubility under the digestion conditions utilized (resulting in ineffective proteolysis and inefficient production of representative peptides), suboptimal peptide size (mass is too large or too small), and peptides being too hydrophilic or hydrophobic resulting in unsatisfactory chromatographic properties and/or poor ionization efficiencies. Moreover, peptides are routinely "missed" due to the stochastic nature of data dependent tandem mass spectrometry (MS). Several strategies have been developed to address undersampling due to low abundance. These approaches include reducing sample complexity by fractionation,[2] enriching low abundance peptides[3] (that could contain a targeted post translational modification, for example), preferential proteolysis and depletion of the most abundant proteins,[4] and immunodepletion of abundant proteins.[5] Fewer studies have reported means to improve the analysis of peptides that ionize poorly upon positive polarity electrospray ionization (ESI) after a conventional low pH liquid chromatography (LC) separation.[6,7] *In silico* digestion of whole proteomes typically result in a bimodal distribution of peptide isoelectric points (p*I*), even when performed with trypsin as the proteolytic agent (which leaves a basic site at both termini).[8] This natural bimodal p*I* distribution of proteolytic peptides (as illustrated in Figure 3.1) from several model proteomes justifies extra effort in targeting the substantial acidic portion of a given peptidome. Although rarely employed in high-throughput

proteomics experiments, negative polarity mass spectrometry provides access to the acidic peptidome which is not well-suited for positive mode analysis. At neutral and slightly basic pH, deprotonation of glutamic and aspartic acid residues promotes the formation of peptide anions which can be readily detected and characterized in the negative mode. In order to achieve the most efficient deprotonation, high pH mobile phases are typically required for LC–MS experiments utilizing negative polarity ESI. Raising the pH of the mobile phase several units above the pK_a of the amino acid side chains results in deprotonation; however, in practice high pH mobile phases are generally incompatible with standard silica based stationary phases and capillaries used in peptide separations.

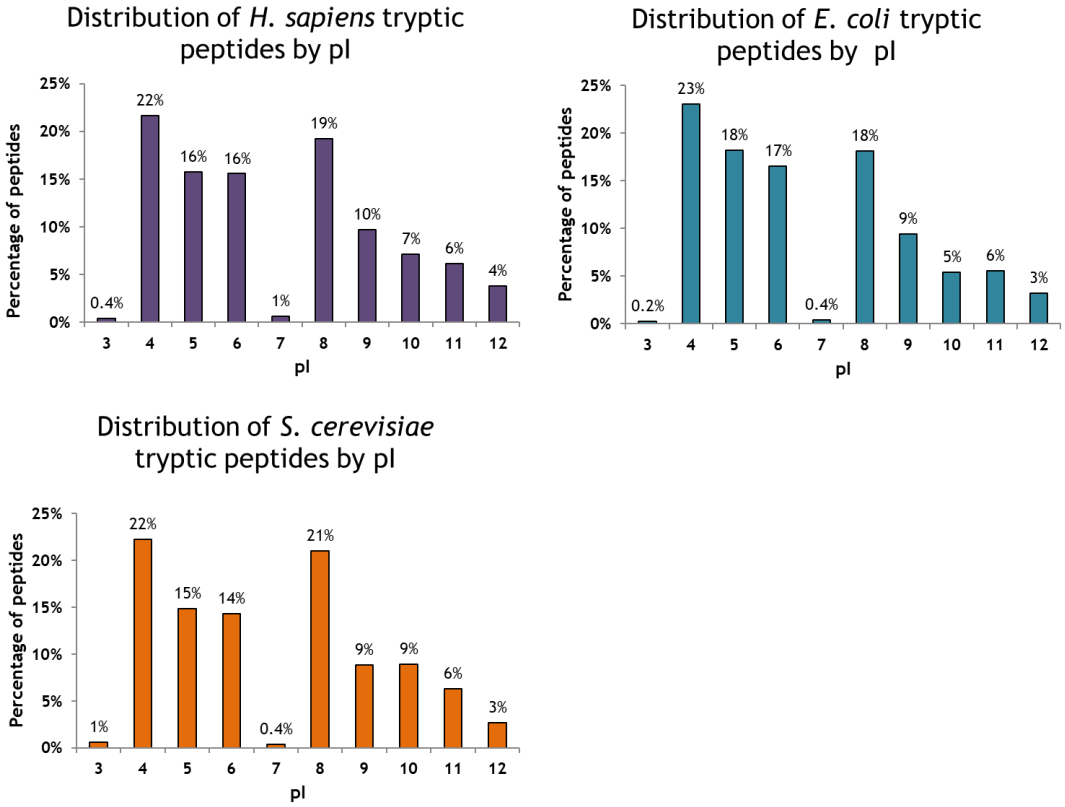


Figure 3.1 pI distribution of in silico generated tryptic peptides from *H. sapiens*, *E.coli*, and *S. cerevisiae* with up to two missed cleavages. Significant portions (52-58%) of these tryptic peptidomes are acidic.

Aside from the high pH required for efficient deprotonation of peptides, the ability to generate informative fragmentation patterns of peptide anions is also challenging. Negative mode analysis suffers from a dearth of options for efficient peptide fragmentation. While positive mode peptide analysis is proficiently accomplished using collision induced dissociation (CID),[9] electron capture or electron transfer dissociation (ECD[10] or ETD,[11] respectively), infrared multiphoton dissociation[12] (IRMPD), or some combination of the above methods, negative mode tandem mass spectrometry (MS/MS) analysis is more

limited. Electron detachment dissociation[13] (EDD), negative mode electron transfer dissociation[14] (NETD), and 193 nm negative mode ultraviolet photodissociation[15] (NUVPD) have been shown to provide diagnostic fragmentation of peptide anions, and the latter two methods have been implemented for the successful analysis of elaborate proteomic mixtures. Kjeldsen optimized EDD for the generation of diagnostic *a* and *x* fragment anions and demonstrated it for LC–MS/MS analysis of a simple single protein digest as well as for phosphopeptide identification from 12 model proteins.[13]Using NETD and a mobile phase around pH 10, Coon and co-workers identified 1412 unique peptides from yeast proteins and showed 45% greater coverage of the acidic yeast GRX1 protein when compared to solely positive mode CID and ETD activation.[6]Despite these positive gains on single proteins, NETD required lengthier (>100 ms) activation times which made it less compatible with high-throughput LC time scales. Madsen et al. reported the identification of over 2000 peptides and 659 proteins upon ultraviolet photodissociation (UVPD) analysis of HeLa cell lysates analyzed in the negative mode.[7] UVPD at 193 nm is successfully implemented using a 2–10 ms activation period (to allow multiple laser pulses) and commonly produces multiple diagnostic ion series; most notably *x*-, *a*-, *b*-, and *z*-type fragments and occasionally *c*- and *y*-type ions. UVPD and NETD have been directly compared[15] for LC–MS analyses of tryptic digests in the negative mode, with the finding that either method, when combined with complementary positive mode CID data, increased sequence coverages and peptide identifications compared to CID alone.[16] UVPD has also proven to be particularly

proficient for analysis of peptides with labile acidic post-translational modifications[17](PTMs), like phosphorylation[18] and sulfation,[19,20] as these PTMs are not lost during UVPD.

Here we introduce a highly efficient means to lower the pK_a of the N-termini and lysine side-chains of peptides by converting the reactive amines to amides. This simple derivatization procedure is readily implemented on complex proteolytic mixtures and results in detection and identification of significantly more peptides in the negative mode by UVPD than obtained for noncarbamylated peptide mixtures, as demonstrated for whole cell lysates of *Halobacterium salinarum*.

3.3 EXPERIMENTAL

3.3.1 Materials

HPLC solvents and buffer components were obtained from Sigma-Aldrich (St. Louis, MO). Proteomics-grade trypsin was obtained from Promega (Madison, WI). All other reagents and solvents were obtained from ThermoFisher Scientific (Fairlawn, NJ). The model peptides DRVYIHPFHL and WAGGDASGE were obtained from American Peptide Company (Sunnyvale, CA). Bovine serum albumin was obtained from Sigma-Aldrich (St. Louis, MO). Halobacterium salinarum was obtained from American Type Culture Collection (ATCC, Manassas, VA). The bacteria were grown in the recommended medium (American Type Culture Collection medium 2185). Cells were suspended in 10 mM Tris-HCL, 10 mM KCl, 1.5 mM $MgCl_2$ at pH 8 to swell and were lysed by dounce

homogenization. The whole cell lysate was centrifuged to clarify the soluble lysate and to remove the insoluble pellet.

3.3.2 Sample Preparation

Both bovine serum albumin and proteins isolated from H. salinarum were digested at 37 °C overnight with trypsin. Prior to digestion, proteins were reduced in 5 mM dithiothreitol for 30 min at 55 °C and subsequently alkylated in 10 mM iodoacetamide at room temperature in the dark for 30 min. Alkylation was quenched with a second aliquot of dithiothreitol, thus bringing the final concentration of dithiothreitol to ~10 mM. Trypsin was added to achieve a 1:20 enzyme-to-substrate ratio, and the solution was buffered at pH 8 in 150 mM ammonium bicarbonate. After digestion the sample was dried in a vacuum centrifuge for subsequent derivatization.

3.3.3 Derivatization

Carbamylation was performed as previously reported.[21] Briefly, each sample was split into two aliquots; one for derivatization and one as a control. Each was resuspended in 200 mM Tris-HCl in the presence or absence of 8 M urea. Both samples were incubated at 80 °C for 4 h. Derivatized peptides were desalted using C18 spin columns (Pierce, Rockford, IL), then evaporated to dryness and resuspended in solvent to match the LC starting conditions (2% acetonitrile/98% water/0.05% acetic acid). The peptides DRVYIHPFHL and WAGGDASGE were derivatized as described above.

3.3.4 LC–MS/MS

The *H. salinarum* samples were analyzed on a Thermo Scientific Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a 193 nm excimer laser (Coherent, Santa Clara, CA) and modified to allow UVPD activation in the HCD cell.[25] Photodissociation was implemented in a manner described previously.[7] Chromatographic separations were performed using water (A) and acetonitrile (B) mobile phases containing 0.05% acetic acid on an Eksigent Nanoultra 2D Plus liquid chromatography system (Redwood, CA). The trap (35 mm × 0.1 mm) and analytical column (with integrated emitter) (15 cm × 0.075 cm) were packed in-house using 3 μm Michrom Magic C18 packing (New Objective, Woburn, MA). Approximately 3 μg of digest was loaded onto the trap column at 2 μL/min for 20 min and separated with a gradient that changed from 0 to 35% B over the course of 240 min at a flow rate of 300 nL min^{-1}. For nanospray, 2.1 kV was applied at a precolumn liquid voltage junction for negative polarity mode, and the tip–inlet distance was carefully adjusted to mitigate the occurrence of corona discharge. Survey and MS/MS scans were acquired by averaging one and three scans, respectively. Automated gain control targets were 1 000 000 for both survey MS and MS^n scan modes. The maximum ion time was 200 ms for MS and MS^n.

All data-dependent nano LC–MS methods on the Orbitrap involved an FT survey scan (m/z 400–2000) at a resolution of 120 000 followed by a series of MS/MS scans on the top 10 most abundant ions from the first survey. The minimum signal required for MS2 selection was 100 000, and the isolation width was fixed at 3 m/z. Dynamic exclusion was

enabled for 30 s with a repeat count of one and a list size of 500 m/z values. For UVPD, three 2-mJ pulses were delivered during an activation period of 6 ms. Product ions from UVPD were detected in the Orbitrap at a resolution of 15 000.

The RAW data files collected on the mass spectrometer were converted to mzXML files by use of MassMatrix data conversion tools (v3.9, http://www.massmatrix.net/download). All data were searched using an in-house MassMatrix Web server (v2.4.2, http://www.massmatrix.net). The search parameters in MassMatrix employed were (i) enzyme, trypsin; (ii) missed cleavage, maximum 2; (iii) modifications, fixed iodoacetamide derivative of cysteine and variable oxidation of methionine (fixed carbamylation of n-term and lysine, when appropriate for modified samples); (iv) precursor ion mass tolerances, 15 ppm for Orbitrap data; (v) product ion mass tolerances, 0.02 Da for UVPD-MS data on Orbitrap; (vi) maximum number of modifications allowed for each peptide, 3; (vii) peptide length, 6–40 amino acid residues; (viii) score thresholds of 5.3 and 1.3 for the pp/pp2 and pp_{tag} scores, respectively. The Halobacterium_sp_nrc1 database was used for Halo data sets. Peptide and protein identifications were both filtered at a 1% false discovery rate. The peptide spectral matches were ranked by confidence and listed in descending order. As the percentage of matches to the decoy database approached one, all spectral matches below that point on the list were discarded.

3.3.5 pK_a Calculation

The change in pK_a between carbamylated and unmodified lysine residues and N-termini were calculated using Marvin (http://www.chemaxon.com/marvin/sketch/index.php), a widely available chemical visualization and property calculation tool (Marvin 14.7.7, 2014). The inverse log of K_a values were calculated using the default parameters.

3.3.6 In Silico Digestion

In silico digests were performed using freely available software (http://omics.pnl.gov/software/protein-digestion-simulator). FASTA files containing the proteomes for H. sapiens, S. cerevisiae, E. coli, and H. salinarum were downloaded from the Swiss-prot database (http://beta.uniprot.org/uniprot/? query=*&fil=reviewed%3Ayes) in their reviewed forms. Tryptic digests were performed allowing up to two missed cleavages.

3.4 Results and Discussion

In order to expand the depth and breadth of coverage in proteomics applications, in particular the ability to analyze intrinsically more acidic peptides which may be less effectively ionized in positive mode, negative mode offers an appealing option. Having previously demonstrated the capabilities of UVPD for analysis of peptides in both the positive and negative modes,[7] we wished to further extend the proteome coverage by

enhancing the range of peptides suitable for analysis in the negative mode. Owing to the often lower efficiency of electrospray ionization in the negative mode (which remains an area of active interest),[22–24] the development of methods to make peptides more amenable to deprotonation is a key objective. In practice, this includes strategies to reduce the basicities of the most basic sites (such as lysines and the N-termini in peptides), thus suppressing protonation and/or increasing the acidities of acidic groups to enhance deprotonation. The high pK_a of the primary amine functional groups is a particularly significant factor, which may strongly influence negative mode ESI efficiencies of peptides. The strategy reported here uses a simple and highly efficient carbamylation reaction which converts primary amines to amide groups, thus decreasing the pK_a values of those functional groups (in particular the lysine side-chains and N-termini). The carbamylation reaction is shown schematically in Figure 3.2, resulting in a mass shift of +43.0058 Da per carbamylation.

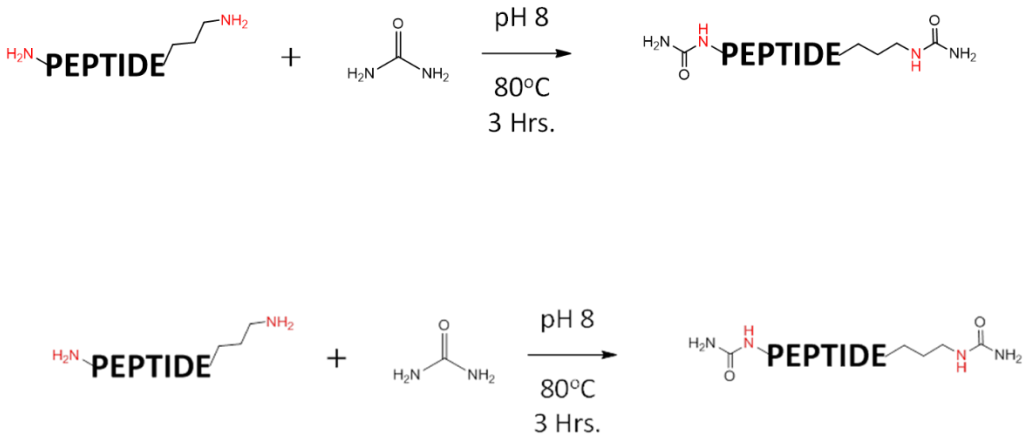


Figure 3.2 Reaction scheme for carbamylation of a peptide bearing a lysine residue

Feasibility experiments were undertaken using model peptides in order to optimize the carbamylation reaction (i.e., minimize the presence of partially reacted species) and cleanup procedure (i.e., minimize sample loss). The carbamylation and C18 spin column cleanup procedure was extremely simple and efficient for individual peptides and, in fact, translated remarkably well to complex mixtures of tryptic peptides, as described later. Reaction efficiencies were estimated based on examination of the abundances of carbamylated and unmodified peptides obtained from extracted ion chromatograms for individual peptides subjected to carbamylation. The ESI mass spectra obtained in the negative mode for one representative unmodified peptide (DRVYIHPFHL) and the same peptide after carbamylation are shown in Figure 3.3. The unmodified peptide is observed primarily as a singly deprotonated species; the carbamylated peptide is observed predominantly as a doubly deprotonated species and its abundance is nearly a factor of 10 greater than that of the unmodified peptide. Examples of the LC traces used to estimate reaction efficiency are shown in Figure 3.4, in which the reaction efficiency of carbamylation was estimated to be 97% for peptide LVNELTEFAK (based on integration of the peak areas for the unmodified and carbamylated peptides). Carbamylation resulted in a modest shift in retention times of peptides because the ionizable amine groups are converted to more hydrophobic amide functionalities. It is estimated that the conversion of the primary amines (N-terminus and lysine side-chains) to amides changes the pK_a of those groups from 9.5 and 10.5, respectively, to an estimated pK_a of −1.7 (Marvin 14.7.7 2014 http://www.chemaxon.com). This is also consistent with the shift in charge state

noted for the DRVYIHPFHL peptide (in Figure 3.3) as well as other peptides upon carbamylation.

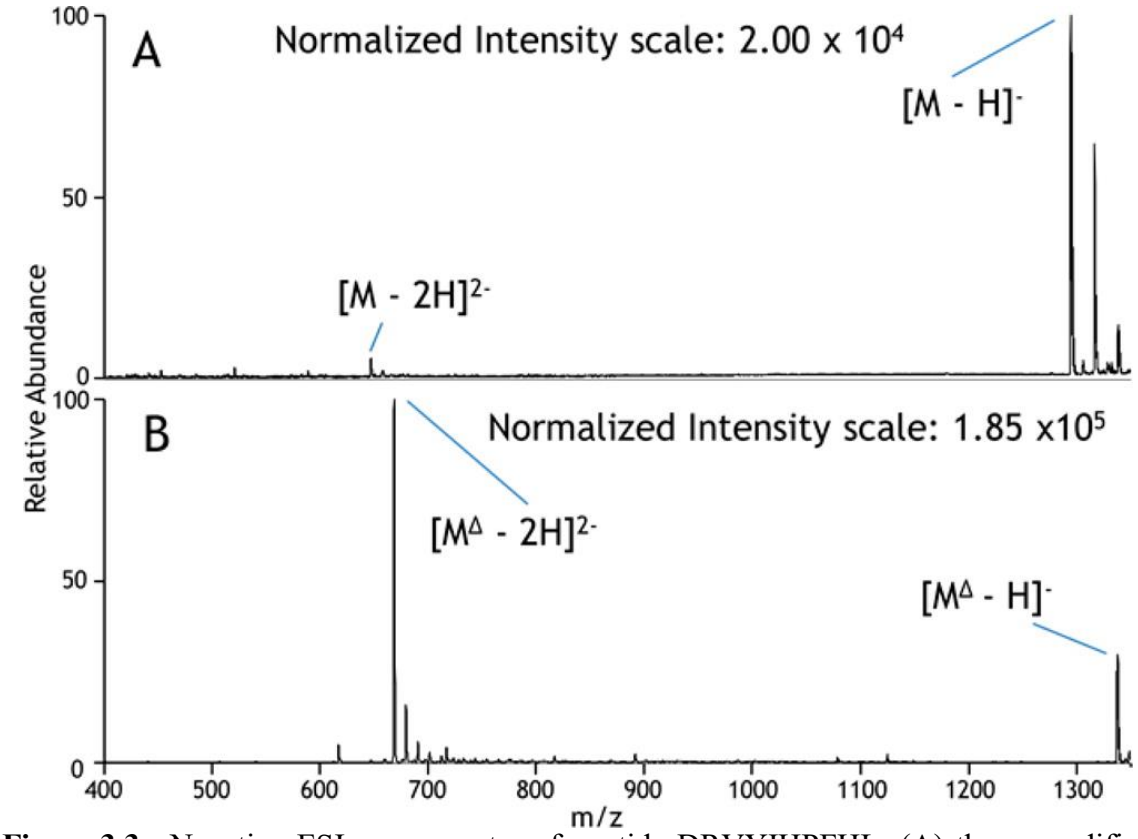


Figure 3.3 Negative ESI mass spectra of peptide DRVYIHPFHL: (A) the unmodified peptide and (B) the carbamylated peptide. Δ represents carbamylation.

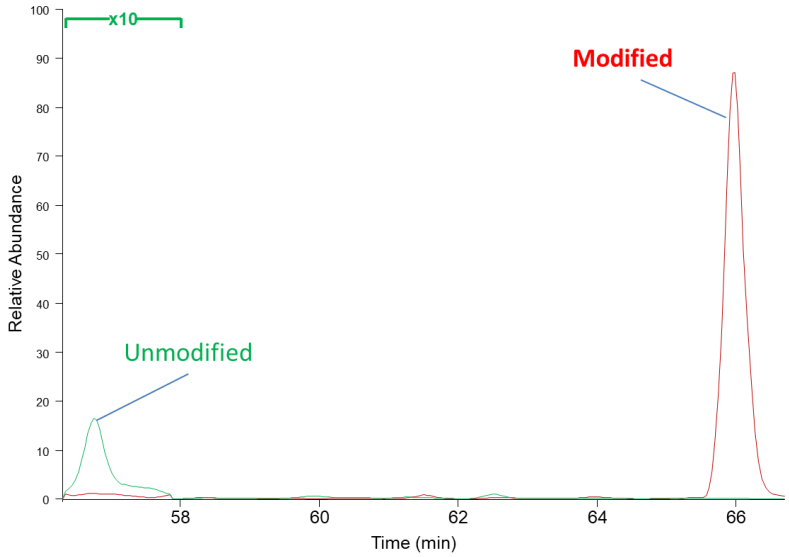


Figure 3.4 Positive mode extracted ion chromatograms of carbamylated and unmodified LVNELTEFAK2+, thus showing the extent of carbamylation of *H. salinarum* tryptic peptides.

In the negative ESI mode, the unmodified peptides typically are detected in low charge states, often as singly deprotonated species of modest abundance, whereas the corresponding carbamylated peptides are observed in higher charge states and with much greater abundances. It is well-known that CID of deprotonated peptides predominantly yields fragment ions resulting from uninformative neutral losses of water and CO_2. In contrast, UVPD of deprotonated peptides results primarily in diagnostic a/x sequence ions in addition to lower abundances of b/y and c/z ions and charge-reduced precursors (i.e., via photoinduced electron detachment).[14] Examples of the rich UVPD mass spectra of an unmodified peptide, GEEVTAEVADGPQSVIFDQAENR, and its carbamylated counterpart are shown in **Figure 3.5**. The relative abundances and types of fragment ions

are similar for both the unmodified and carbamylated peptide, indicating that carbamylation does not suppress or significantly alter the UVPD process.

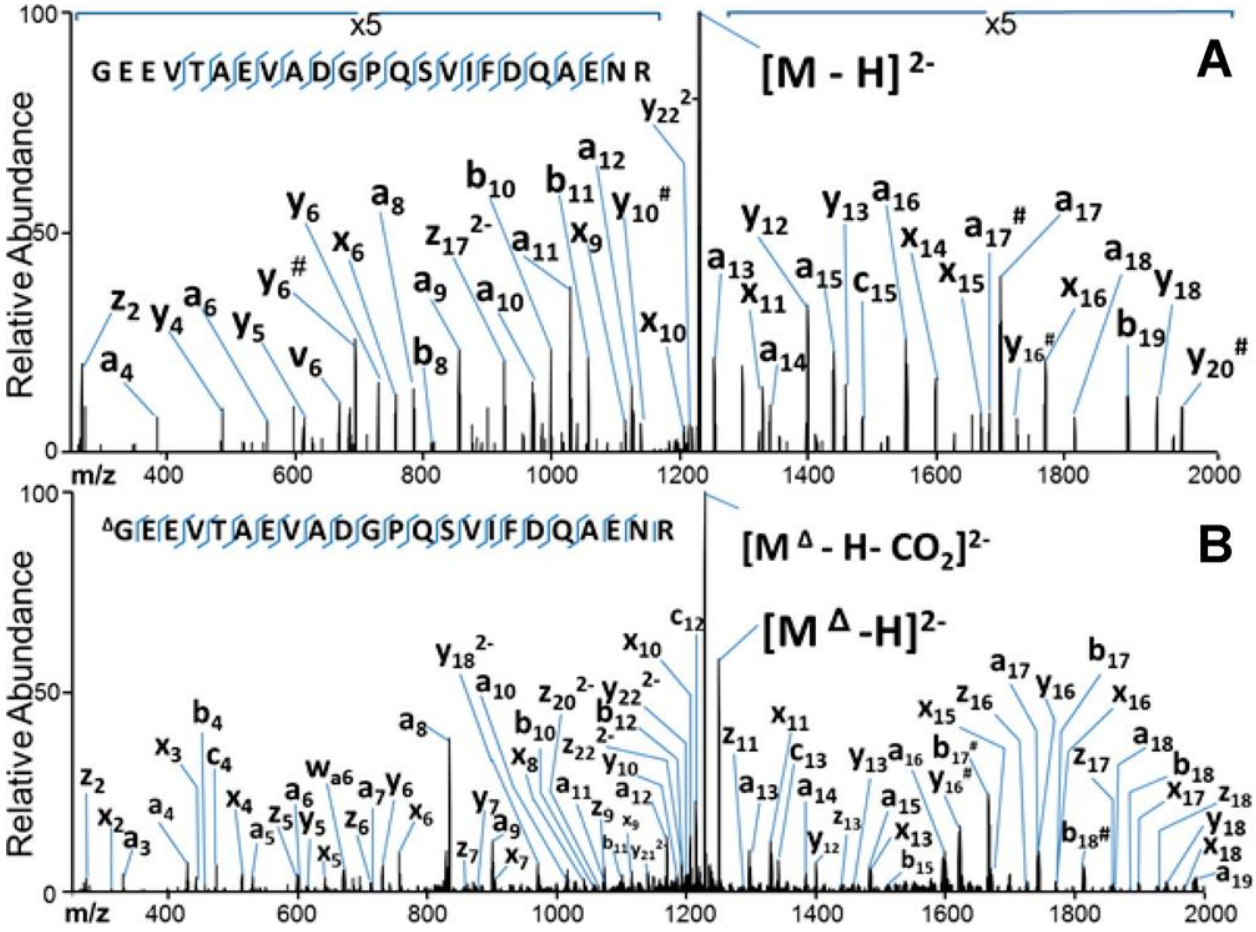


Figure 3.5 UVPD mass spectra of GEEVTAEVADGPQSVIFDQAENR from *H. salinarum*: (A) unmodified (2−) and (B) carbamylated (2−). # indicates the loss of water.

While these initial experiments were important for proving the feasibility of the method, to evaluate the scalability of the carbamylation reaction for more complex mixtures of tryptic peptides, BSA was digested and the resulting peptides were carbamylated.

Extracted ion chromatograms (EIC) were generated to monitor the elution of both unmodified and the corresponding carbamylated peptides. The areas of the extracted ion peaks were used to measure the efficiency of carbamylation according to the following equation where A is chromatographic peak area:

$$\text{reaction efficiency (\%)} = \frac{A_{\text{modified}}}{A_{\text{modified}} + A_{\text{unmodified}}} \times 100$$

In agreement with the reactions of individual model peptides, the reaction efficiencies of the measured BSA tryptic peptides averaged more than 97%. Other derivatization reagents (such as the popular 4-sulfophenyl isothiocyanate) have been used in the past successfully to enhance negative mode ionization, but their success for high-throughput proteomics applications have proven to be subpar due to the low reaction efficiencies for complex multicomponent mixtures.[25] As also noted above, the dominant charge states of the resulting carbamylated tryptic peptides of BSA were typically shifted by one charge (e.g., from 1– to 2−), and the abundances increased by a factor of 7.6 on average relative to the unmodified peptides.

Trypsin was used as the protease of choice in this study because it is the enzyme most commonly used for mass spectrometric-based bottom-up proteomics applications. One advantage of using trypsin for conventional positive mode LC–MS studies is that it generally results in at least two very basic sites (N-terminus and C-terminal K or R residues) which enhances the formation of multiply charged peptide cations that are ideal for CID and database searches. Carbamylation reduces the pK_a values of the lysine side-

chain and N-terminus, thus reducing the basicity of those sites and making them less ionizable in the positive mode and overall making the peptides more amenable to negative mode ESI. By retaining the use of trypsin in the present study, the protein mixtures can be subjected to tryptic digestion, then split into two samples: one for traditional bottom-up/positive mode approach and the other processed in parallel using the carbamylation/negative mode UVPD strategy. This dual positive/negative MS/MS approach should extend the range of peptides (and therefore proteins) identified with confidence. Moreover, the nearly stoichiometric carbamylation reaction efficiencies observed for the model peptides and BSA digest allowed carbamylation of the Lys side-chains and N-termini to be treated as fixed modifications.

A summary of the observed carbamylated peptides and their corresponding peak areas is shown in **Table 3.1** for carbamylated BSA tryptic peptides. In cases where peptides contain multiple primary amines (i.e., one or more lysine side-chains plus the N-terminus), the predominant products were the fully carbamylated species (Figure 3.6). Despite the reaction undertaken in somewhat basic conditions (pH 8), the pK_a of the arginine side-chain is substantially greater (pK_a 12.5) and thus the majority (>99%) of arginine side-chains remained protonated and unreactive.

Extent of Carbamylation of tryptic BSA peptides measured by (+) nano LC/MS/MS chromatographic peak areas										
Singly Carbamylated										
Peptide+modification	**Charge**	**m/z**	**Peak Area**	**Unmodified**	**Charge**	**m/z**	**Peak Area**	**Retention Time (min)**		
								shift from carbamylation	**%Unmod**	**%Mod**
LGEYGFQNALIVR + CARB	2+	761.90	654327139	LGEYGFQNALIVR	2+	740.40	5126023	*+16.25*	<1%	99%
DAFLGSFLYEYSR + CARB	2+	805.88	904112797	DAFLGSFLYEYSR	2+	784.37	4882691	*+24.01*	<1%	99%
LVVSTQTALA + CARB	1+	1045.59	92343433	LVVSTQTALA	1+	1002.58	216255	*+17.20*	<1%	99%
LVVSTQTALA + CARB	2+	523.29	1304867769	LVVSTQTALA	2+	501.79	4033497	*+17.09*	<1%	99%
Doubly Carbamylated										
LVNELTEFAK	2+	582.31	Not detected					Not observed		0
LVNELTEFAK + CARB	2+	603.82	4496449					65.93		1%
LVNELTEFAK + CARB + CARB(K)	2+	625.32	333178921					72.30		99%
Triply Carbamylated										
LKPDPNTLCDEFK	2+	760.37	1284970					42.83		<1%
LKPDPNTLCDEFK + CARB	2+	810.41	247578					48.05		<1%
LKPDPNTLCDEFK + CARB+ CARB(K)	2+	831.90	10378373					49.52		5%
LKPDPNTLCDEFK + CARB+CARB(K) +CARB(K)	2+	853.39	179823475					54.3		94%

Table 3.1 List of carbamylated BSA peptides and the chromatographic peak areas of the carbamylated peptide and its corresponding unmodified peptide. For the case of multiple (double and triple) modifications, each sequential modification is shown. CARB indicates the peptide modification at the N-terminus, and CARB(K) indicates the modification at the lysine side-chain. %Mod represents the estimated percentage of each peptide that is carbamylated (versus remains unreactive), calculated by dividing the peak area of the carbamylated peptide (from the extracted ion chromatogram) by the summed peak areas of the carbamylated and non-carbamylated peptides.

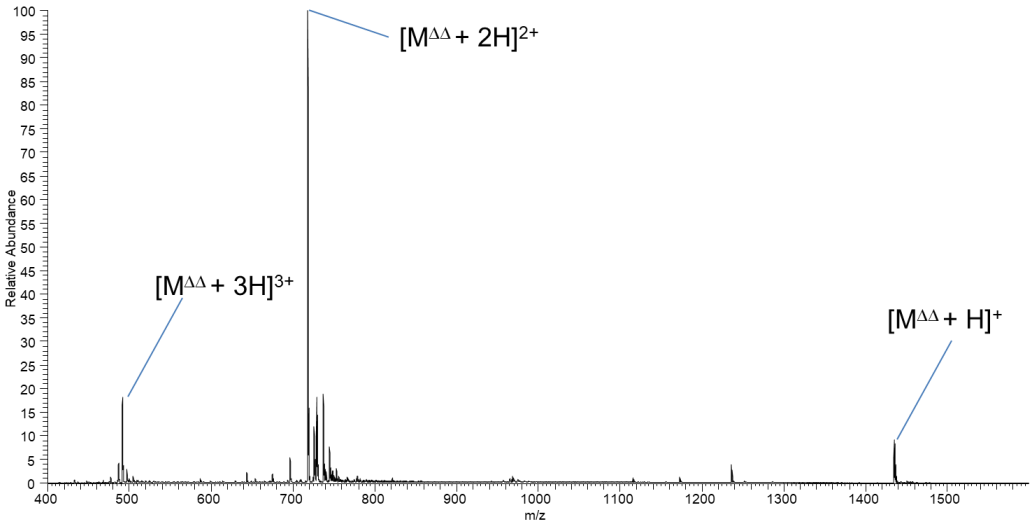


Figure 3.6 Positive mode ESI spectrum of RPKPQQFFGLM (Mr 1347.72 Da) after carbamylation. The major species observed is doubly modified (N-terminus and K), corresponding to a doubly carbamylated species (Mr 1433.72 Da).

To evaluate the carbamylation/UVPD strategy for a larger array of peptides, the method was applied to the analysis of the H. salinarum proteome. Many of the proteins in the H. salinarumproteome are naturally acidic, thus resulting in a large distribution of tryptic peptides possessing lower than average pI values (average 5.8) (Figure 3.7). In fact, over 65% of the predicted tryptic peptides are expected to have pI values below 7. Carbamylation was used to further reduce the average peptide pI values by decreasing the pK_a values of the N-termini and lysine residues. After tryptic digestion of the proteins extracted from the H. salinarum lysate, the peptides were incubated and carbamylated in 8 M urea and desalted. Upon comparison of the chromatograms obtained from the carbamylated and noncarbamylated tryptic peptides, the carbamylated peptides were

retained on the column for between 5 and 15 min longer than their underivatized counterparts, and the degree of the retention time shift scaled with the number of carbamylated sites. This increase in hydrophobicity agrees with the findings mentioned earlier for the model peptides and BSA peptides and is consistent with replacement of the ionizable primary amines by the more hydrophobic amide moieties. Despite the increased retention times, most carbamylated peptides eluted when the mobile phase composition contained less than 35% (v/v) acetonitrile.

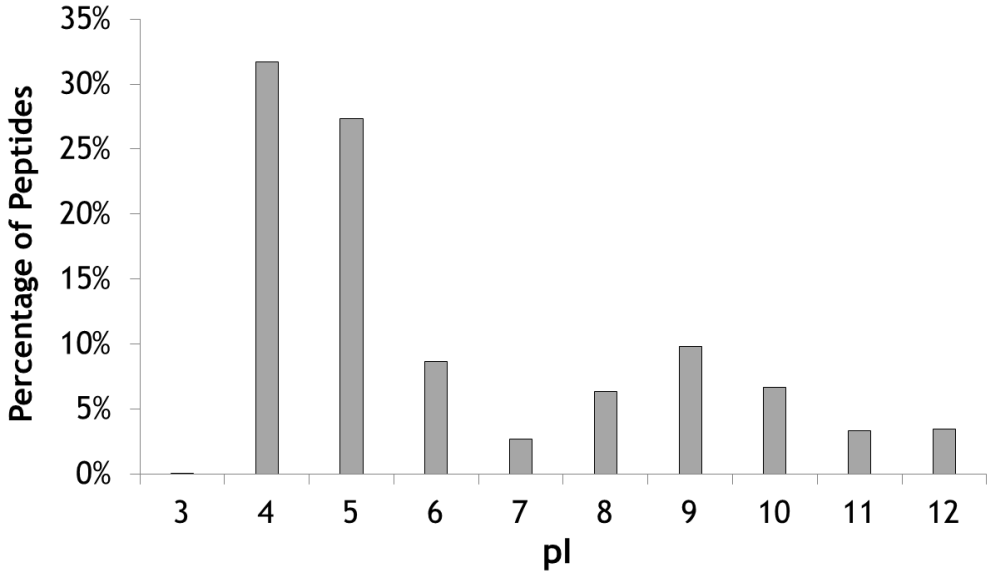


Figure 3.7 pI distribution of in silico generated tryptic peptides from H. salinarum with up to two missed cleavages.

An example of an LC trace for a carbamylated tryptic digest of H. salinarum and a representative UVPD mass spectrum for one peptide ($^{\Delta}$DNVAAIIIGSR carbamylated at its N-terminus) is shown in Figure 3.8. The UVPD mass spectrum is dominated by a/x ions

with lower abundances of z ions, and the sequence coverage for this peptide is very high (as is also the case for many of the other carbamylated tryptic peptides). Inspection of the charge state distributions of those peptides identified by UVPD for H. salinarum demonstrates a shift in average charge state for the carbamylated peptides, as summarized in Figure 3.9. Among the 789 carbamylated peptides identified by UVPD, a larger portion was detected as 3– and 2– charge states, whereas more were detected as 2– and 1– charge states among the 549 peptides identified for the noncarbamylated digest. More importantly, the average peptide abundances were higher for the carbamylated digest than the unmodified digest by a factor of 2.4, thus confirming the signal enhancement obtained by reducing the net pI values of the peptides upon carbamylation.

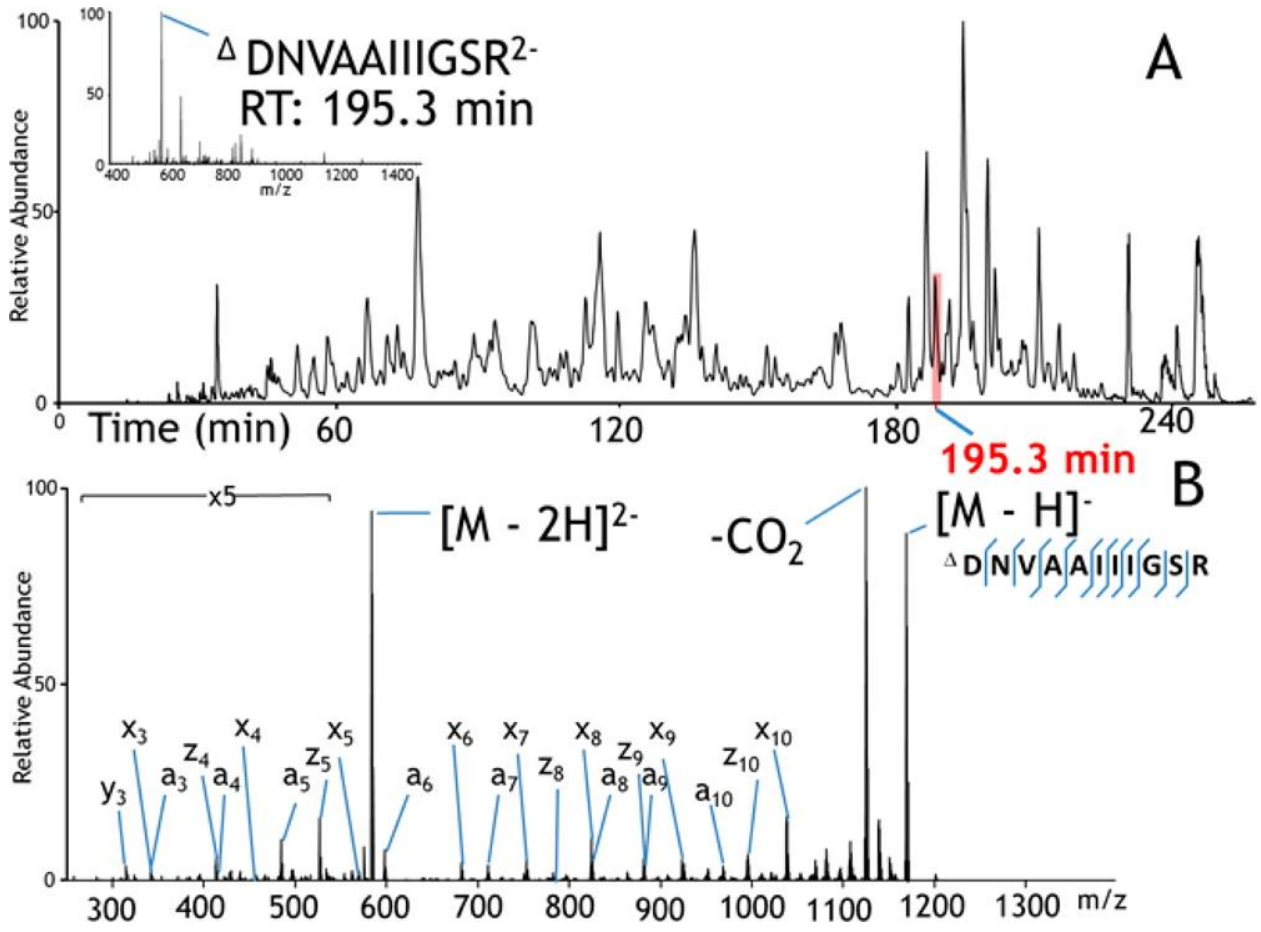


Figure 3.8 (A) Negative mode LC–MS trace (total ion chromatogram) of *H. salinarum* tryptic peptides. Inset: ESI mass spectrum acquired at 195.3 min. (B) UVPD mass spectrum of carbamylated peptide DNVAAIIIGSR (2−) eluting at 195.3 min.

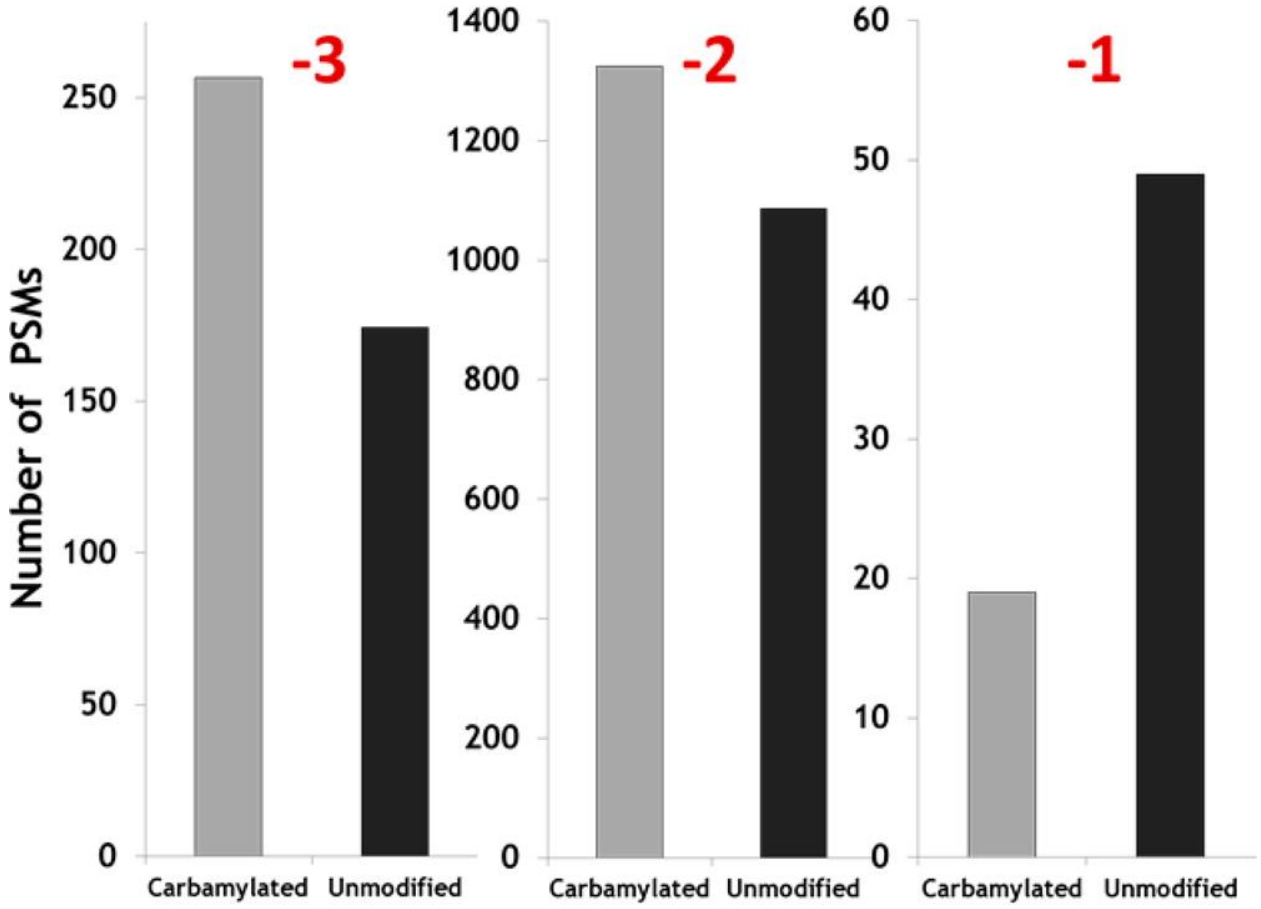


Figure 3.9 Distributions of charge states for carbamylated (light bars) and unmodified (dark bars) tryptic peptide spectral matches (PSMs) found for the digest of *H. salinarum.*

The UVPD fragmentation patterns of the more highly charged peptides give better peptide sequence coverage. For carbamylated peptides, 1.7 diagnostic ions per residue on average were generated for peptides in the 3– charge state, 1.3 diagnostic ions per residue for peptides in the 2– charge state, and 0.6 diagnostic ions per residue in the 1– charge state. The similarities in the average number and types of fragment ions for carbamylated versus

noncarbamylated peptides offers assurance that the carbamylation reaction does not suppress or significantly alter the rich UVPD patterns generated for peptide anions. With respect to the types of fragment ions, the distributions are nearly identical for both the carbamylated and unmodified peptides: averaging 50% a/x ions, 30% b/y ions, and 20% c/z ions (Figure 3.10)

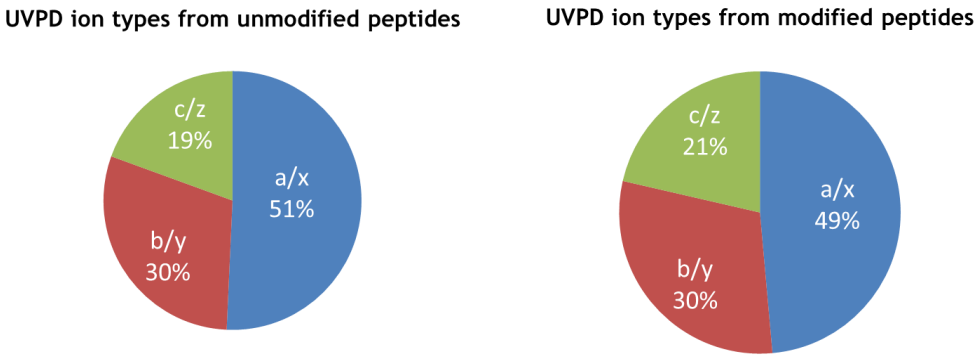


Figure 3.10 Distribution of ion types resulting from UVPD of unmodified and carbamylated tryptic peptide anions from *H. salinarum*

The average number and standard deviation of peptides and proteins identified from the H. salinarum proteome were calculated from triplicate negative mode LC–MS UVPD analyses of the carbamylated and the unmodified tryptic digests, as summarized in Figures Figures 3.11 (histograms) and and Figure 3.12 (Venn diagrams). Combining three runs led to the identification of 1086 peptides for the carbamylated digest compared to 747 for the unmodified digest. Similarly, at the protein level, 430 proteins were identified based on the peptides found in the carbamylated digests compared to 348 proteins for the unmodified

digests. Upon combining the results of three runs, 682 peptides were found uniquely for the carbamylated digest, 343 peptides were found exclusively for the unmodified digest, and surprisingly only 404 were found for both digests. At the protein level, 156 proteins were uniquely identified from the results of the carbamylated tryptic digest, whereas 74 were exclusively found for the unmodified digest, and 274 proteins were identified in both cases. The results show 45% more peptide identifications and 25% more protein identifications after carbamylation when compared to the unmodified digest. With respect to the charge states of the peptides that were identified, on average 267 carbamylated peptide spectral matches (PSMs) were found in the 3– charge state compared to 174 PSMs for unmodified peptides, 1325 carbamylated PSMs were found in the 2– charge state compared to 1086 PSMs for unmodified peptides, and 19 carbamylated PSMs were found in the 1– charge state compared to 49 PSMs for unmodified peptides.

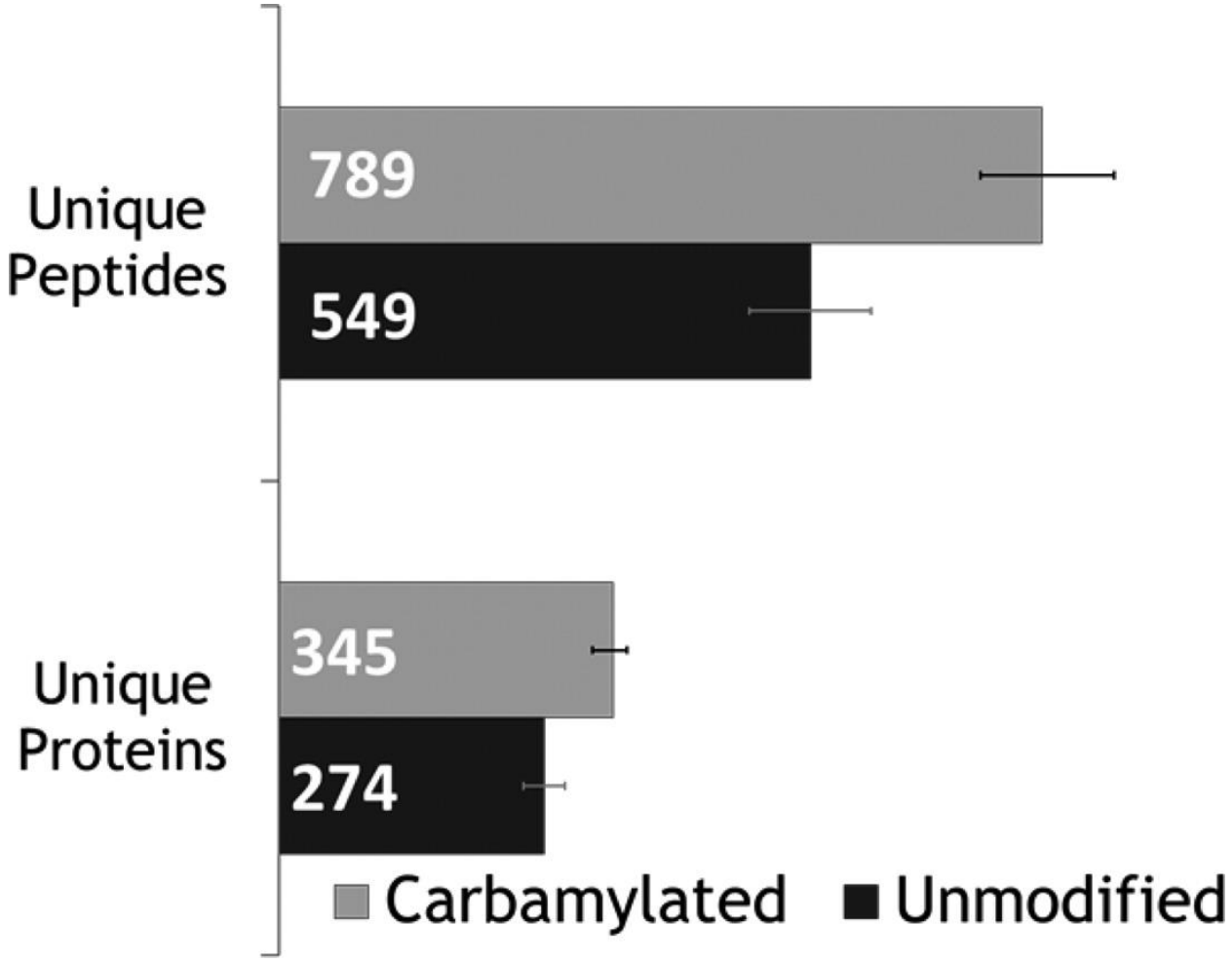


Figure 3.11 Average number of peptide and protein identifications from negative LC/UVPD-MS analyses of carbamylated and unmodified *H. salinarum* tryptic peptides (in triplicate).

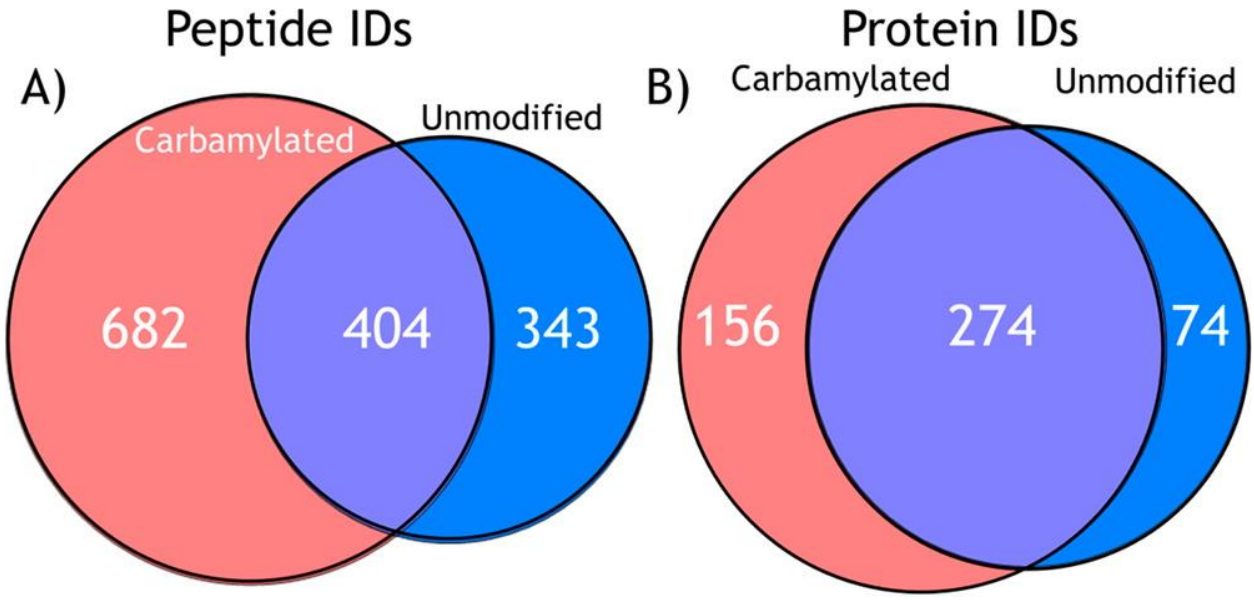


Figure 3.12 Combined number of (A) unique peptide and (B) protein identifications from LC/UVPD-MS analyses of carbamylated and unmodified tryptic peptides from H. salinarum in the negative mode in triplicate.

Average number of peptide and protein identifications from negative LC/UVPD-MS analyses of carbamylated and unmodified H. salinarum tryptic peptides (in triplicate).

Interestingly a reasonably large number of peptides (343) were identified only from analysis of the unmodified digest. Given the stochastic nature of data dependent acquisition and bias toward the most abundant peptide precursor ions,[26] many peptides are not selected for fragmentation in a routine mass spectrometry proteomics experiment.[27] Closer inspection of those peptides identified only for the unmodified tryptic digests show larger peptides on average than ones commonly identified for the carbamylated digests (Figure 3.13). On the basis of the retention times of these peptides (Figure 3.14), they are also more hydrophobic (with longer elution times), and thus carbamylation of these large peptides

would be expected to further increase their hydrophobicities and further delay elution. This may account in part for why this set of peptides was not identified for the corresponding carbamylated digest.

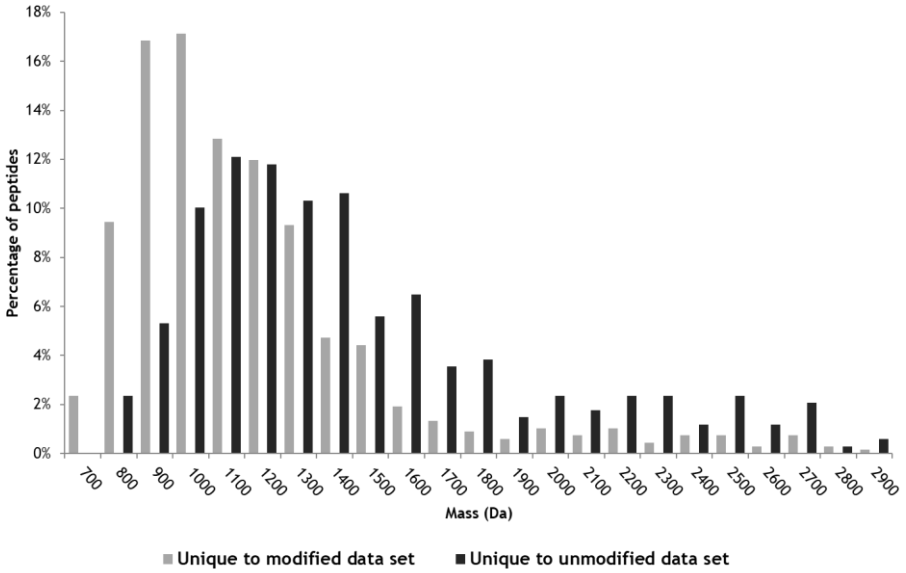


Figure 3.13 Mass distribution of peptides identified uniquely in the carbamylated peptide data set (light bars) and peptides identified uniquely in the unmodified peptide data set (dark bars) for UVPD of a tryptic digest of *H. salinarum.*

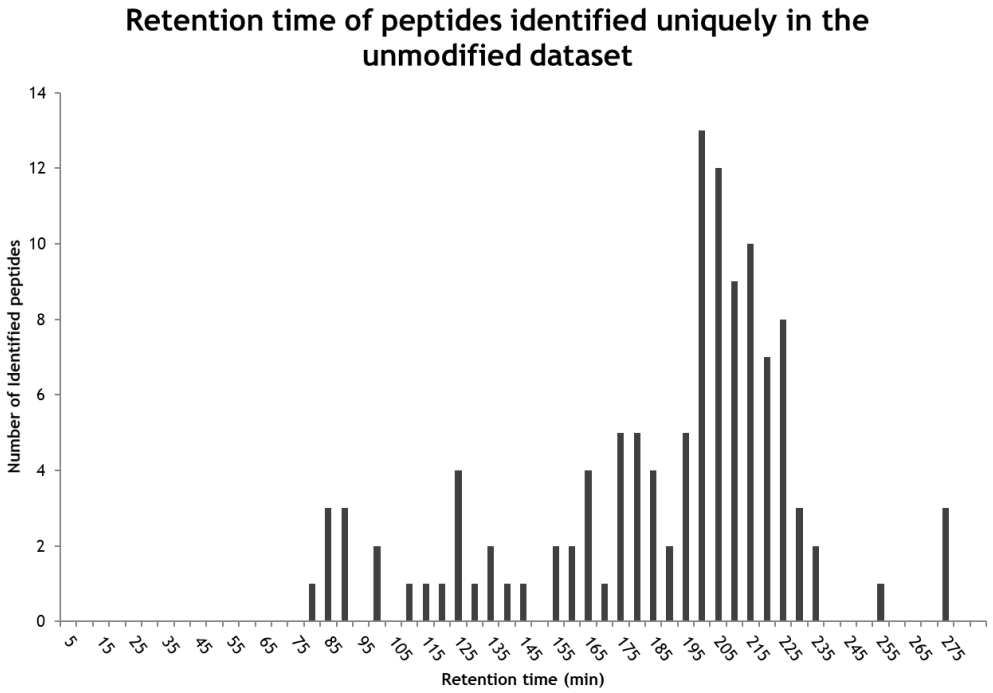


Figure 3.14 Number of peptides identified uniquely in the unmodified peptide data set sorted by elution time for a tryptic digest of *H. salinarum.*

3.5 CONCLUSION

Carbamylation of lysine residues and N-termini was utilized to enhance the ionization of peptides by negative polarity ESI and improve the sensitivity of negative mode LC/UVPD-MS analyses. Results show a significant enhancement in negative mode ionization of carbamylated peptides compared to unmodified peptides, consistent with the significant decrease in pK_a upon carbamylation of primary amines. Carbamylation of tryptic digests resulted in 45% more peptide identifications and 25% more protein identifications compared to that obtained for the unmodified digests, confirming the enhancement in

sensitivity in the negative mode. The improvement in peptide identification metrics also arises from a shift to higher charge states, thus yielding more efficient UVPD. The carbamylation method could be combined with other proteases, like LysC, to ensure multiple modifications of each peptide.

3.6 REFERENCES

(1) Bensimon, A.; Heck, A. J. R.; Aebersold, R. Annu. Rev. Biochem. 2012, 81, 379−405.

(2) Gunaratne, J.; Schmidt, A.; Quandt, A.; Neo, S. P.; Sarac, O. S.; Gracia, T.; Loguercio, S.; Ahrne, E.; Xia, R. L. H.; Tan, K. H.; Lossner, C.; Bahler, J.; Beyer, A.; Blackstock, W.; Aebersold, R. Mol. Cell. Proteomics 2013, 12, 1741−1751.

(3) Gygi, S.; Villen, J.´ Nat. Protoc. 2008, 3, 1630−1638.

(4) Fonslow, B. R.; Stein, B. D.; Webb, K. J.; Xu, T.; Choi, J.; Park, S. L.; Yates, J. R., Jr. Nat. Methods 2013, 10, 54−56.

(5) Kuhn, E.; Whiteaker, J.; Mani, D. R.; Jackson, A.; Lei, Z.; Pope, M.; Smith, D.; Rivera, K.; Anderson, N. L.; Skates, S. J.; Pearson, T. W.; Paulovich, A. G.; Carr, S. A. Mol. Cell. Proteomics 2012, 9, 184− 196.

(6) 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, 2875−2882.

(7) 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, 2604−2614.

(8) Schwartz, R.; Ting, C. S.; King, J. Genome Res. 2001, 11, 703− 709.

(9) Wells, J. M.; McLuckey, S. A. In Methods in Enzymology; Elsevier: Amsterdam, The Netherlands, 2005; Vol. 402, pp 148−185.

(10) Cooper, H. J.; Hakansson, K.; Marshall, A. G.˚ Mass Spectrom. Rev. 2005, 24, 201−222.

(11) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Proc. Natl. Acad. Sci. U.S.A. 2004, 101, 9528−9533.

(12) Gardner, M. A.; Ledvina, A. R.; Smith, S.; Madsen, J.; Schwartz, G. C.; Stafford, G. C.; Coon, J. J.; Brodbelt, J. S. Anal. Chem. 2009, 81, 8109−8118.

(13) Kjeldsen, F.; Hørning, O. B.; Jensen, S. S.; Giessing, A. M. B.; Jensen, O. N. J. Am. Soc. Mass Spectrom. 2008, 19, 1156−1162.

(14) Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. J. Am. Soc. Mass. Spectrom. 2005, 16, 880−882.
(15) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. Proteomics 2011, 11, 1329−1334.
(16) Shaw, J. B.; Madsen, J. A.; Xu, H.; Brodbelt, J. S. J. Am. Soc. Mass Spectrom. 2012, 23, 1707−1715.
(17) Henderson, J. C.; Fage, C. D.; Cannon, J. R.; Brodbelt, J. S.; Keatinge-Clay, A. T.; Trent, M. S. ACS Chem. Biol. 2014, 9, 2382− 2392.
(18) Luo, Y.; Yogesha, S. D.; Cannon, J. R.; Yan, W.; Brodbelt, J. S.; Zhang, Y. ACS Chem. Biol. 2013, 8, 2042−2052.
(19) Han, S. W.; Lee, S. W.; Bahar, O.; Schwessinger, B.; Robinson, M. R.; Shaw, J. B.; Madsen, J. A.; Brodbelt, J. S. Nat. Commun. 2012, 3, 1153.
(20) Robinson, M.; Moore, K.; Brodbelt, J. S. J. Am. Soc. Mass Spectrom. 2014, 25, 1461−71.
(21) Angel, P. M.; Orlando, R. Rapid Commun. Mass Spectrom. 2007, 10, 1623−1634.
(22) Allen, S. J.; Schwartz, A. M.; Bush, M. F. Anal. Chem. 2013, 85, 12055−12061.
(23) Douglass, K. A.; Venter, A. R. Anal. Chem. 2013, 85, 8212− 8218.
(24) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Anal. Chem. 2013, 85, 2−9.
(25) Vasicek, L. A.; Ledvina, A. R.; Shaw, J. B.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J.; Brodbelt, J. S. J. Am. Soc. Mass Spectrom. 2011, 22, 1105−1108.
(26) Liu, H.; Sadygov, R. G.; Yates, J. R. Anal. Chem. 2004, 76, 4193−4201.
(27) Michalski, A.; Cox, J.; Mann, M. J. Proteome Res. 2011, 10, 1785−1793. Analytical Chemistry Article 12290 dx.doi.org/10.1021/ac5035314 | A

Chapter 4

Impact of Protease on Ultraviolet Photodissociation Mass Spectrometry for Bottom-up Proteomics

4.1 Overview

Recent mass spectrometric studies have reported enhanced proteome coverage by employing multiple proteases or by using multiple or alternative activation methods such as electron transfer dissociation in combination with collisional activated dissociation (CAD). In this study the use of 193 nm ultraviolet photodissociation for analysis of thousands of *Halobacterium salinarum* peptides generated by four proteases (trypsin, LysC, GluC and chymotrypsin) was evaluated in comparison to higher energy CAD (HCD). Proteins digested by trypsin resulted in greater sequence coverage for HCD over UVPD. LysC digestion resulted in similar sequence coverages for UVPD and HCD; however, for proteins digested by GluC and chymotrypsin 5-10% more sequence coverage on average was achieved by UVPD. HCD resulted in more peptide identifications (at 1% false discovery rate) for trypsin (4356 peptides by HCD versus 3907 peptides by UVPD), whereas UVPD identified greater numbers of peptides for LysC digests (1033 peptides by UVPD versus 844 HCD), chymotrypsin digests (3219 peptides for UVPD versus 2921 for HCD) and GluC digests (2834 peptides for UVPD and 2393 for HCD) and correspondingly greater numbers of proteins.

4.2 INTRODUCTION

Bottom-up mass spectrometric methods have become the mainstream approach for high throughput proteomics, including both qualitative and quantitative applications.[1–5] The tremendous success is due in part to the ability to generate extensive arrays of characteristic peptides upon proteolytic digestion of proteins, thus facilitating highly effective database searches based on MS/MS spectra of peptides. Trypsin has conventionally been the protease of choice for bottom-up proteomics due to the desirable characteristics of the resulting peptides.[1,3,6] Tryptic peptides are terminated by residues possessing basic side-chains (Lys or Arg), thus affording sites that protonate readily and yielding efficient ionization in the positive mode. Moreover, given the frequency of tryptic cleavage sites many of the resulting peptides are predicted to be small (< 7 residues) and thus less amenable to effective sequencing by MS/MS.[7,8] Recently groups have shown the merits of utilizing multiple proteases in conjunction with collisional activated dissociation (CAD) in a bottom up proteomics workflow.[8–10] The use of multiple proteases in series or parallel on a common sample has been adopted to increase the breadth of proteome coverage by taking advantage of the differential specificities of the various proteases employed and the different characteristics of the resulting peptides (i.e. size, hydrophobicity, charged sites etc).[8–12]

Introducing variation in peptide character (i.e. size, amino acid composition, and location, type and frequency of ionizable sites) via multiple proteases may result in non-ideal peptides for CAD. For example, peptides bearing internal basic sites may result in

production of unassignable fragment ions upon MS/MS. This shortcoming can be addressed by use of other activation methods. For instance, ETD has been shown to provide more extensive sequence coverage of large peptides with the added advantage of labile modification retention.[13] However, at the same time electron-based activation methods are biased towards peptides in higher charge states and with greater sequence lengths. Recently an approach combining both ETD and CAD in conjunction with multiple protease digestion of a HeLa cell lysate was reported, resulting in substantial improvement in peptide backbone fragmentation and more robust peptide identification.[14]

In recent years ultraviolet photodissociation (UVPD) has proven to be suitable for a broad range of proteomics applications and conveniently offers many of the desirable characteristics of both CAD and ETD.[15–27] In particular, activation of protonated peptides by 193 nm UVPD yields *a, b, c, x, y,* and *z* fragment ions, a "blend" of both CAD- and ETD-type fragments. The present study integrates the use of multiple proteases to create orthogonal sets of peptides and UVPD for peptide characterization with the goal of uncovering fundamental insight into the effects of peptide size, charge state and amino acid composition on photoactivated fragmentation. For instance, the peptide backbone (i.e. amide functionality) serves as a chromophore for UV absorption,[28] thus peptides of various lengths, such as those generated by LysC versus trypsin, may exhibit varying degrees of fragmentation. Moreover, it is known that the presence of aromatic residues enhances UV cross-sections, and thus peptides bearing aromatic residues, such as those generated by chymotrypsin, may display enhanced UVPD.[29] In general, coupling the versatility of

UVPD with the potential benefits of using multiple proteases afford a compelling opportunity to further extend the depth of proteome sequence coverage. To date UVPD remains uncharacterized across multiple proteases for bottom-up analyses. In this study we evaluate the 193 nm UVPD fragmentation of thousands of peptides arising from multiple proteases, including trypsin, chymotrypsin, GluC and LysC.

4.3 EXPERIMENTAL METHODS

4.3.1 Materials

HPLC solvents were obtained from EMD Millipore (Temecula, CA), and buffer components were obtained from Sigma-Aldrich (St. Louis, MO). Proteomics-grade trypsin, r-LysC, GluC, and chymotrypsin were obtained from Promega (Madison, WI). All other reagents and solvents were obtained from Thermo Fisher Scientific (Fairlawn, NJ). *Halobacterium salinarum* was obtained from American Type Culture Collection (ATCC, Manassas, VA).

4.3.2 Sample Preparation

H. salinarum was grown in the recommended media (American Type Culture Collection media 2185). Cells were suspended in 10 mM Tris-HCl, 10 mM KCl, 1.5 mM $MgCl_2$ at pH 8 to swell and were lysed by dounce homogenization. The whole cell lysate was centrifuged to clarify the soluble lysate and to remove the insoluble pellet. Proteins isolated from *H. salinarum* were digested with various enzymes according to the following

procedures. Prior to digestion, proteins were sequentially reduced in 5 mM dithiothreitol (DTT) for 30 minutes at 55 °C and alkylated in 10 mM iodoacetamide at room temperature in the dark for 30 minutes. Alkylation was quenched with a second aliquot of DTT, bringing the final concentration of DTT to ~10 mM. After alkylation each sample was split into three separate aliquots for triplicate analysis. The same process was followed for each protease (chymotrypsin, GluC, LysC, and trypsin) prior to digestion.

Proteases were added in a 1:50 enzyme-to-substrate ratio, and the solution was buffered at pH 8 in 150 mM ammonium bicarbonate for trypsin and GluC digests (using 25 mM Tris-HCl and 1 mM EDTA pH 8.5 for LysC digests, and 100 mM Tris-HCl and 10 mM $CaCl_2$ pH 8.0 for chymotrypsin digests). Digestion proceeded for 18 hr at 37 ºC for trypsin, LysC and GluC or 12 hr at 25ºC for chymotrypsin. After digestion all samples were quenched with 1% formic acid and cleaned over a spin cartridge loaded with C18 resin (Pierce Biotechnology) prior to LC-MS analysis.

4.3.3 Liquid Chromatography and Mass Spectrometry

The *H. salinarum* digests were analyzed on a Thermo Scientific Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a 193 nm excimer laser (Coherent, Santa Clara, CA) and modified to allow UVPD in the HCD cell.[30] Chromatographic separations were performed using water (A) and acetonitrile (B) mobile phases containing 0.05% acetic acid on an Eksigent Nanoultra 2D Plus nano liquid chromatography system (Redwood,CA). Trap (35 mm × 0.1 mm) and analytical column with an integrated emitter (15 cm × 0.075 cm) were packed in house using 3 μm Michrom

Magic C18 packing (New Objective, Woburn, MA). Approximately 1 µg of digest was loaded onto the trap column at 2 µl/min for 20 min, then separated with a gradient that changed from 0 to 35% B over the course of 240 minutes. For nanospray, 1.8 kV was applied at a precolumn liquid voltage junction. MS^1 and MS^2 scan were single scans without averaging. Automated gain control targets were 1,000,000 for both survey MS and MSn scan modes. The maximum ion time was 100 ms for MS and MSn.

While low energy collision induced dissociation (CID) is the most popular benchmark fragmentation method for high-throughput bottom-up proteomics experiments, for the purposes of the present comparison higher energy collision dissociation (HCD) was chosen as the comparative fragmentation method to 193 nm UVPD. Both HCD and UVPD are implemented in the HCD cell of the modified Thermo Orbitrap Elite mass spectrometer. Moreover neither HCD nor UVPD suffer from the low mass cut-off which plagues conventional CID in ion trap instruments. Additionally both methods generate, to different extents, ion types not traditionally seen in low energy CID such as *a, c, x* and *z* type ions.

All data-dependent nano LC-MS methods on the Orbitrap involved an FT MS^1 scan (*m/z* 400–2000) at a resolution of 120,000 followed by a series of MS^2 scans on the top ten most abundant ions from the MS^1 scan. All MS^1 and MS^2 scans were comprised of a single scan; no averaging was performed. The minimum signal required for MS2 selection was 10,000, and the isolation width was fixed at 3 *m/z*. Dynamic exclusion was enabled for 30 s with a repeat count of one and a list size of 500 *m/z* values. For UVPD, normalized collision energy of 1.0 was used to transfer ions into the HCD cell, following which two 2

mJ pulses were delivered during an activation period of 4 msec. Product ions from UVPD were detected in the Orbitrap at a resolution of 15,000. For HCD, normalized collision energy of 35 was used to activate precursors during a 0.1 ms period. Product ions were detected in the Orbitrap at R = 15,000. A single MS2 scan was collected for UVPD and HCD fragmentation. For MS/MS spectra collected at R = 15,000, the signal is sampled for 100 msec in the Orbitrap analyzer, and thus the Orbitrap analysis time is the rate limiting step (not the UVPD or HCD steps). Data was analyzed using Proteome Discoverer v 1.3 (Thermo Fisher Scientific) running the SEQUEST search algorithm. Peptides were identified at a 1% false discovery rate. Raw MS/MS data was searched against the *Halobacterium salinarum* (strain ATCC 700922 / JCM 11081 / NRC) proteome containing 2426 protein sequences which can be found online at www.uniprot.org under Proteome ID UP000000554. The protein list was downloaded as a FASTA file and compiled into a SEQUEST protein database.

4.4 RESULTS AND DISCUSSION

For this systematic evaluation of the impact of the protease and activation method on bottom-up mass spectrometric strategies, *Halobacterium salinarum* was used as a model proteome. *H. salinarum* is a more acidic proteome than the more commonly explored *Saccharomyces cerevisiae* (yeast) or *E. coli*, thus allowing assessment of bottom-up methods for proteomes with lower pKa values. The proteins extracted from the *H. salinarum* lysate were subjected to proteolysis via chymotrypsin, GluC, trypsin, or LysC,

prior to chromatographic separation and electrospray ionization of the eluting peptides and tandem mass spectrometric characterization by HCD or UVPD. The numbers, charge states, and average masses of the peptides identified were determined as a function of protease and activation method, as well as sequence coverages and numbers of proteins identified. Examples of the MS/MS spectra obtained for two peptides, DIHPTAIIK (2+) from the trypsin digest and HDGAPAIDGIDDTIISDDTARY (3+) from the chymotrypsin digest, are shown in **Figure 4.1**. As well established, the HCD spectra are dominated by diagnostic *b/y* ions, and the UVPD spectra show a more diverse array of product ions, including *a/x, b/y* and *c/z* ions.

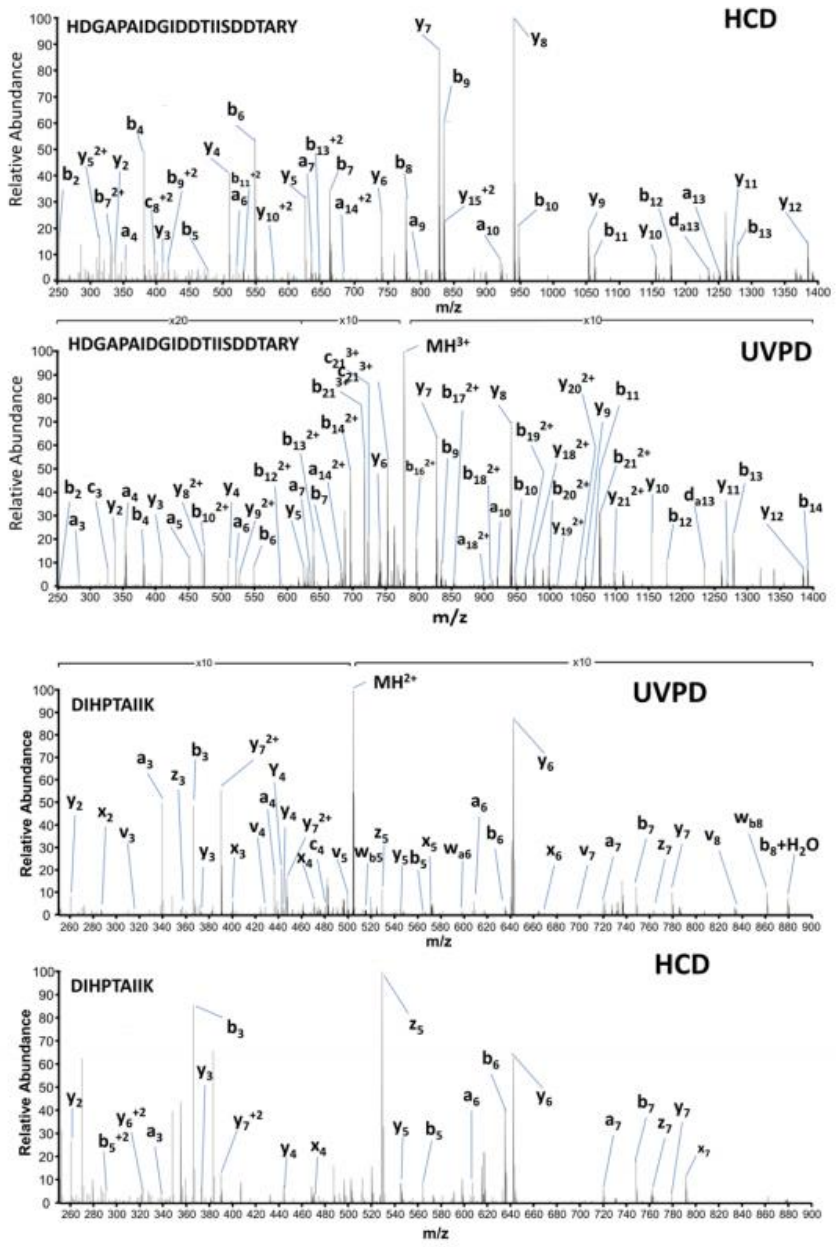


Figure 4.1 Comparative UVPD and HCD fragment ion spectra for a Chymotryptic peptide (HDGAPAIDGIDDTIISDDTARY, 3+) (top) and tryptic peptide (DIHPTAIIK, 2+) (bottom)

The average numbers of peptides identified upon HCD and UVPD analysis of the digests generated from the four proteases are shown in **Figure 4.2**, along with the corresponding number of proteins identified from the matched peptides. As expected, the greatest number of peptides were identified from the tryptic digest (4356 by HCD and 3905 by UVPD), followed by the chymotryptic digest, then the GluC digest, then the LysC digest. Interestingly, although HCD outperformed UVPD for the tryptic digest, with identification of more than 10% additional peptides, UVPD yielded more peptide identifications than HCD for both the GluC and chymotrypsin digests: 2834 peptides (UVPD) versus 2393 peptides (HCD) for GluC, a 20% increase, and 3219 peptides (UVPD) versus 2921 peptides (HCD) for chymotrypsin, a 10% enhancement. The results obtained for the LysC digest were similar for both HCD and UVPD (1061 versus 1031 peptides, respectively).

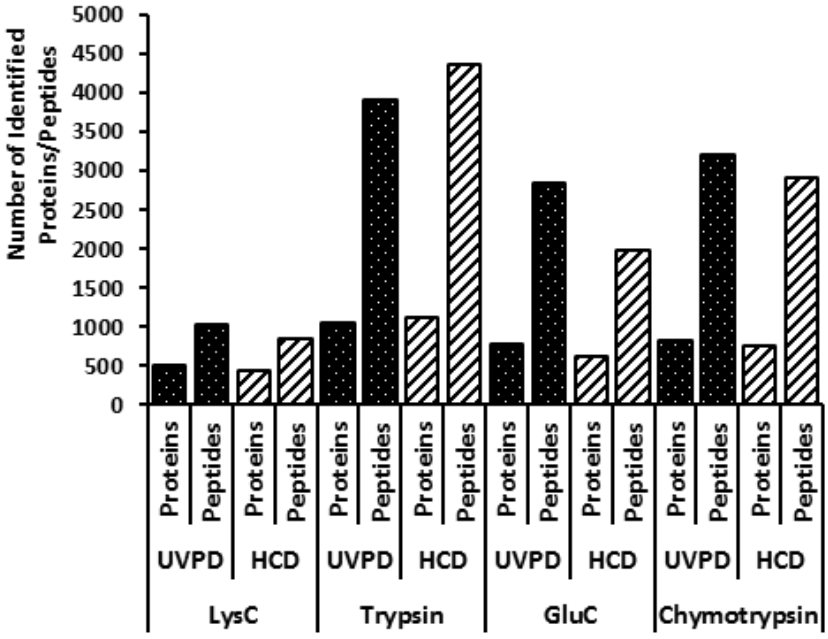


Figure 4.2 Average number of identified peptides and proteins by UVPD and HCD

Choice of protease for a bottom-up workflow has a direct impact on the sizes of peptides generated, and this size effect is further reflected by the activation method used to interrogate the peptides. The average sizes, calculated in terms of mass, of the *H. salinarum* peptides identified by HCD and UVPD are summarized in **Figure 4.3** for each of the four proteases. The LysC peptides are considerably larger (>30% on average) than those generated by the other three proteases. The greater peptide size is likely due to the low frequency of cleavage sites (only lysine) for LysC relative to the other proteases which hydrolyze at more than one type of amino acid. The peptides successfully identified by HCD were found to be typically 5-15% larger than those identified by UVPD, with the average peptide size upon UVPD found to be 1436 Da for the chymotryptic digest (1547

Da for HCD) , 1418 Da for the GluC digest (1597 Da for HCD), 1647 Da for the tryptic digest (1636 Da for HCD), and 2061 Da for the LysC digest (2194 Da for HCD).

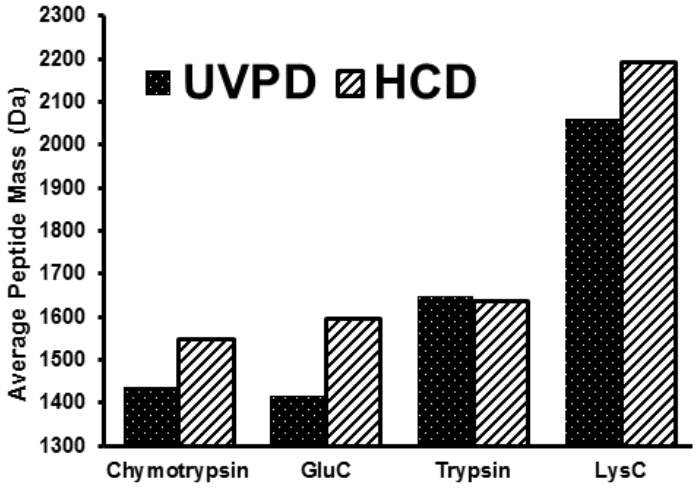


Figure 4.3 Average mass of peptides identified by UVPD and HCD for four proteases.

The distributions of peptide sizes identified upon UVPD of each digest are shown in more detail in **Figure 4.4**. The average masses and mass distributions of the peptides identified for the chymotrypsin and GluC digests are similar, peaking around 1400-1450 Da. The average peptide mass is higher for the tryptic digest (closer to 1600 Da) with a broader distribution, and this trend is further exaggerated for the LysC digest, with the average mass shifted closer to 2100 Da. The total number of peptides identified by UVPD was lowest for the LysC digest as evidenced by the histogram in **Figure 4.4.**

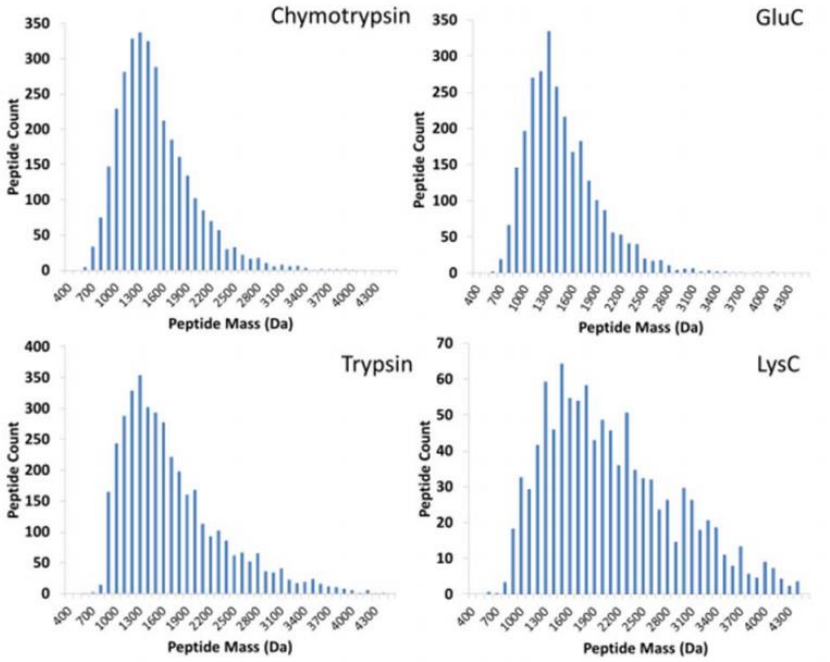


Figure 4.4 Histograms showing distributions of peptide sizes (based on mass) identified by UVPD from chymotrypsin (top left), GluC (top right), trypsin (bottom left), and LysC (bottom right).

The difference in peptide sizes noted in **Figure 4.3** may arise in part from the charge states of the peptides that lead to the most informative and confident MS/MS spectra by UVPD versus HCD. The distributions of the charge states (1+ to 5+) of the peptides identified for each of the four digests are illustrated in **Figure 4.5**. A greater proportion of lower charged peptides were identified by UVPD for all proteases. This observation is most notable for the chymotrypsin digest for which over 16% of UVPD spectral matches are in the 1+ charge state, whereas only 4% of HCD spectral matches are in the 1+ charge state. Poor HCD fragmentation is anticipated for singly charged precursors due to the lack of mobile protons needed to facilitate the charge-mediated fragmentation pathways

commonly promoted by collisional activation.[31] This shortcoming is further exacerbated by the particularly acidic nature of the peptides produced upon GluC digestion. UVPD does not appear to be as heavily dependent on charge state as HCD given the larger proportion of less highly charged peptides (singly and doubly charged peptides) identified by UVPD relative to those identified by HCD. In addition, chymotrypsin, which cleaves at the carboxyl side of Y, F, W and L residues, as well as causing some other non-specific cleavages, results in smaller peptides (**Figure 4.4**) which consequently tend to be less highly charged than their larger LysC counterparts. UVPD proved well suited for identifying the smaller peptides in low charge states or larger peptides with few basic sites compared to HCD.

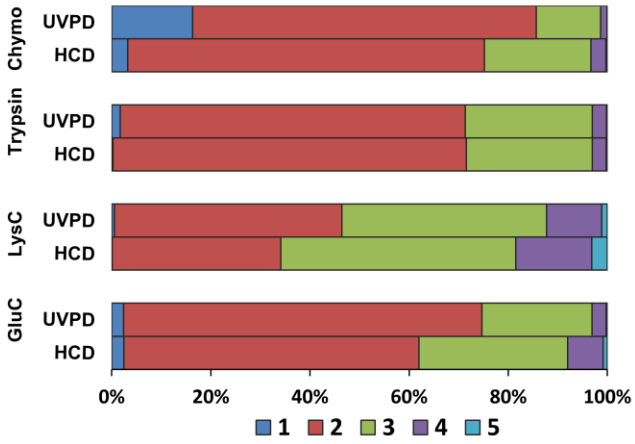


Figure 4.5 Charge state distribution of peptide identified by UVPD and HCD for four proteases

On average peptides from the LysC digest were identified in the highest charge states compared to peptides produced from the other proteases (**Figure 4.5**). Due to the

creation of larger peptides upon LysC digestion as evidenced by the larger average sizes of LysC peptides identified by UVPD and HCD (2194 Da and 2093 Da average mass for peptides detected by UVPD and HCD, respectively), the likelihood of having more charge bearing residues and populating higher charge states is greater.

The distributions of charge states of peptides identified by HCD and UVPD were nearly identical when generated by proteolysis with trypsin. The majority of peptides identified were doubly charged as expected for trypsin which cleaves after basic lysine or arginine residues (e.g. leaving at least two sites for protonation: one basic residue and the N-terminus). In contrast, the GluC peptides identified by UVPD were somewhat more likely to be found in the 2+ charge state than those identified by HCD (75% and 62%, respectively, **Figure 4.5**). Despite the formation of similar sized peptides for GluC and trypsin, the fact that GluC yields peptides terminated by acidic residues accounts for the difference in charges states when compared to the peptide pool produced by trypsin, (**Figures 4.4** and **4.5**).

The total number of proteins identified (**Figure 4.1**) and the distribution of sequence coverages for those proteins (**Figure 4.6**) provide two other metrics that allow comparison of the performance of UVPD and HCD for the analysis of the four proteolytic digests. As expected based on the numbers of identified peptides discussed above, the numbers of identified proteins follow a parallel trend, with UVPD identifying the greatest number of proteins for the GluC and chymotrypsin digests. HCD outperformed UVPD for the trypsin digest (1127 proteins identified for HCD versus 1061 proteins for UVPD, or

6% more for HCD), but UVPD outperformed HCD for the LysC, GluC, and chymotryptic digests in terms of the number of identified proteins (LysC: 513 UVPD vs 430 HCD, GluC: 773 UVPD vs 727 HCD, chymotrypsin: 829 UVPD vs 763 HCD) (**Figure 4.2**).

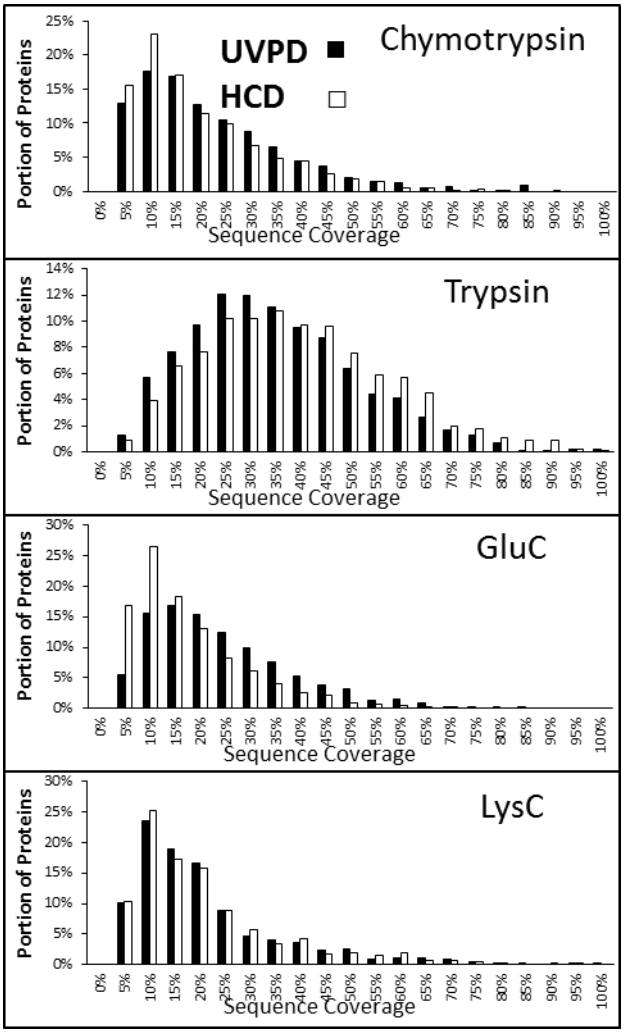


Figure 4.6 Histograms of protein sequence coverages by UVPD and HCD for four proteases

Furthermore, proteolysis with trypsin yielded the best sequence coverages for both UVPD and HCD (**Figure 4.6**). The histograms of protein sequence coverage showed an average coverage of 35-40% per protein identified for the tryptic digest and significantly more proteins identified with greater than 50% sequence coverage compared to the analogous findings for the LysC, GluC, and chymotrypsin digests. The distributions of GluC, LysC and chymotrypsin sequence coverages all had average values in the range of 15-20% and displayed significantly fewer proteins with coverages greater than 50% (compared to the trypsin results). These trends are not surprising given the widespread adoption and refinement of trypsin protocols which have made trypsin the gold standard for mass spectrometry-based bottom-up proteomic workflows. Trypsin exhibits excellent fidelity (i.e. very little nonspecific cleavage) whereas the other proteases, especially chymotrypsin, routinely promote nonspecific cleavages. Non-specific cleavage may result in peptides which are too small for optimal detection (i.e. < 400 Da) or which may not be properly processed in database searches. In this study several low specificity cleavages (M and L residues) were discovered when processing the chymotrypsin data.

With respect to the sequence coverages obtained upon UVPD versus HCD, although the shapes of the distributions were similar (**Figure 4.6**), there was a consistent increase in the portion of proteins identified with higher sequences coverages for UVPD compared to HCD for the chymotrypsin and GluC digests, whereas HCD outperformed UVPD for the trypsin digest. The sequence coverages for the LysC digest were nearly

indistinguishable by UVPD and HCD. These results suggest that UVPD may be the preferred activation method when GluC or chymotrypsin is used.

Figure 4.7 shows in a Venn diagram format the total number of proteins which were identified in common and uniquely by UVPD and HCD. For all four digests, the majority of the same proteins were identified by both UVPD and HCD; however, there was a notable subset of proteins which were uniquely identified by UVPD or HCD. For the GluC, chymotrypsin, and LysC digests, 20-25% of unique proteins were contributed by UVPD, compared to 7-12% of proteins uniquely identified by HCD. This trend reflects in part the greater total number of proteins identified by UVPD for the GluC, chymotrypsin and LysC digests (**Figure 4.2**). HCD identified more unique proteins than UVPD only for the trypsin digest. In general, HCD and UVPD exhibited good complementarity, generally increasing the total number of identified proteins by 10% or more.

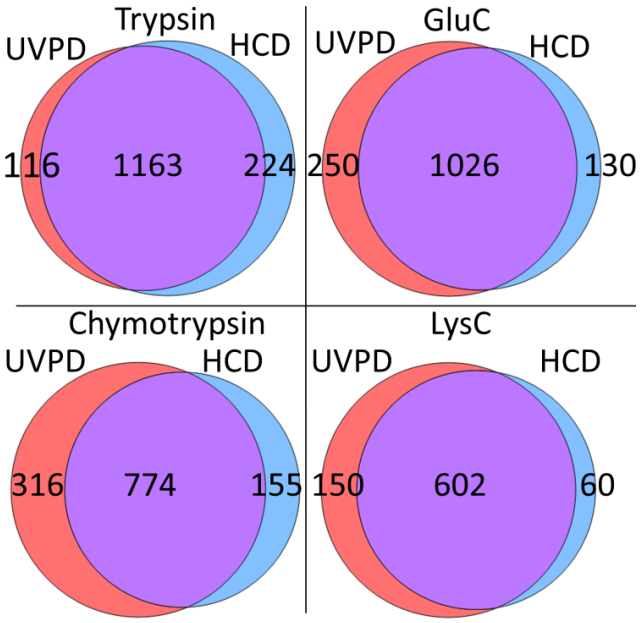


Figure 4.7 Venn diagram showing number of unique proteins identified by UVPD and HCD

Coon et al. showed a considerable increase (average of 31%) in *S. cerevisiae* protein identification upon inclusion of unique protein identifications from combined CAD and ETD results of multiple proteases including LysN, GluC, trypsin, ArgC, and LysC over any single protease.[8] Following on this prior observation, combining the unique protein identifications of all proteases identified by HCD and UVPD for all four proteolytic digests in the present study yielded a total of 1986 uniquely identified proteins (**Figure 4.8**) from *H. salinarum,* a 45% average increase in identifications over the number obtained by analyzing the peptides generated from any single protease, using a single activation

method. This result reiterates the utility of using multiple proteases and complementary fragmentation methods.

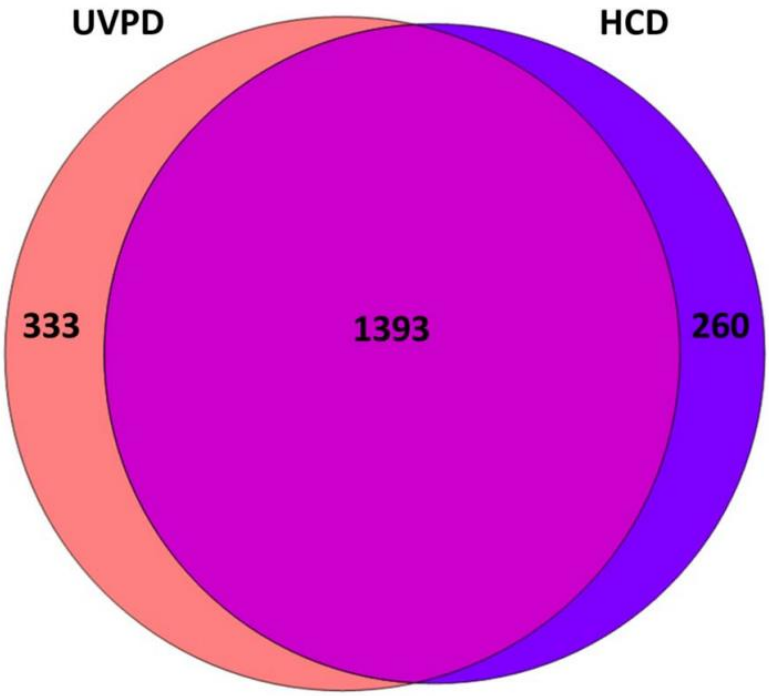


Figure 4.8 Venn diagram of total unique proteins identified by UVPD and HCD across all proteases

One of the most notable characteristics of 193 nm UVPD is the great diversity of fragment ion types. The distributions of fragment ion types (*a, b, c, x, y, z* produced via back-bone cleavages and *d,v* and *w* ions from side-chain losses in conjunction with backbone cleavages) produced by UVPD and HCD of peptides from the four digests are shown in **Figure 4.9**. With respect to the comparison of the ion types produced by UVPD versus HCD, the relative portion of the N-terminal *a, b* and *c* ions was remarkably

consistent, whereas the portions of C-terminal *x,y*, and *z* ions showed much greater discrepancies between UVPD and HCD. The contributions from *x* and *z* ions were more significant for UVPD, and *y* ions were far more dominant for HCD. In addition, various side-chain loss ions (*d,v,w*) were produced upon UVPD but not by HCD. The variation in the distributions of ion types is attributed to the different mechanisms of UVPD and HCD, HCD pathways are typically charge-mediated processes that result in cleavage of the most labile amide backbone bonds along with a few preferential cleavages (such as occurring adjacent to proline). In contrast, dissociation directly from excited electronic states may occur after UV photoabsorption, thus allowing access to pathways not active for HCD. Interestingly, the distribution of fragment ion types varied even more dramatically based on the protease used to create the peptides. As expected, for the peptides identified in the LysC and trypsin digests by UVPD and HCD, there was a much greater portion of C-terminal fragment ions than N-terminal fragment ions, an outcome consistent with the placement of a basic Arg or Lys residue at the C-terminus of those peptides. In contrast, N-terminal fragment ions were favored for the peptides identified from the GluC and chymotrypsin digests. The peptides from the GluC and chymotrypsin digests have the standard N-terminus primary amine as a consistent basic site, and may have Arg or Lys residues throughout the sequence but not restricted to the C-terminus. These characteristics favor formation of N-terminal fragment ions. Several *d, v,* and *w* ions were observed upon UVPD of selected peptides from the LysC, chymotrypsin and GluC digests. These ions have proven useful for differentiation of leucine and isoleucine in peptides.[32]

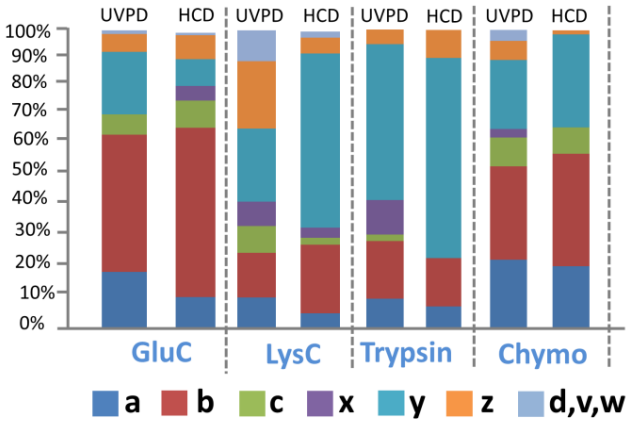


Figure 4.9 Distribution of UVPD and HCD fragment ions for different proteases used for Halobacterium cell lysate. The abundances of the various fragment ion types were compiled for the 25 most confidently identified peptides (based on XCorr scores) from each digest.

4.5 CONCLUSION

Following LC/MS-MS analysis of thousands of *H. salinarum* peptides generated by multiple proteases, UVPD and HCD showed several distinctions. UVPD was able to identify peptides in significantly lower charge states across all samples, supporting the lower charge state dependence of UVPD and lower reliance on mobile protons for generating diagnostic fragment ions. At the same time a greater portion of smaller peptides were identified by UVPD (10% smaller by mass on average than those identified by HCD). HCD out-performed UVPD for identification of tryptic peptides (11% more identifications than by UVPD), whereas similar numbers of proteins and peptides were identified by HCD and UVPD for the LysC digests. For those peptides which were not terminated by basic residues (e.g. those from GluC and chymotrypsin digests), significantly more were

identified by UVPD over HCD (20% more identifications by UVPD on average). More modest gains in total protein sequence coverage were found based on UVPD of the GluC or chymotryptic digests (increases of 10% and 5%, respectively) compared to the sequence coverages obtained by HCD.

4.6 REFERENCES

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

(2) Cox, J.; Mann, M. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. *Annu. Rev. Biochem.* **2011**, *80*, 273–299.

(3) Zhang, Z.; Wu, S.; Stenoien, D. L.; Paša-Tolić, L. High-Throughput Proteomics. *Annu. Rev. Anal. Chem.* **2014**, *7*, 427–454.

(4) Mann, M.; Kulak, N. A.; Nagaraj, N.; Cox, J. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. *Mol. Cell* **2013**, *49*, 583–590.

(5) Yates, J. R. The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. *J. Am. Chem. Soc.* **2013**, *135*, 1629–1640.

(6) Vandermarliere, E.; Mueller, M.; Martens, L. Getting intimate with trypsin, the leading protease in proteomics: TRYPSIN IN PROTEOMICS. *Mass Spectrom. Rev.* **2013**, 000–000.

(7) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. *Anal. Chem.* **2014**, *86*, 3005–3012.

(8) Swaney, D. L.; Wenger, C. D.; Coon, J. J. Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2010**, *9*, 1323–1329.

(9) Wiśniewski, J. R.; Mann, M. Consecutive Proteolytic Digestion in an Enzyme Reactor Increases Depth of Proteomic and Phosphoproteomic Analysis. *Anal. Chem.* **2012**, *84*, 2631–2637.

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

(11) McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R. Comparison of three directly coupled HPLC MS/MS strategies for identification

of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int. J. Mass Spectrom.* **2002**, *219*, 245–251.
(12) Cheung, W. C.; Beausoleil, S. A.; Zhang, X.; Sato, S.; Schieferl, S. M.; Wieler, J. S.; Beaudet, J. G.; Ramenani, R. K.; Popova, L.; Comb, M. J.; et al. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat. Biotechnol.* **2012**, *30*, 447–452.
(13) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc. Natl. Acad. Sci.* **2007**, *104*, 2193–2198.
(14) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2012**, *84*, 9668–9673.
(15) 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. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135*, 12646–12651.
(16) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-nm photodissociation of singly and multiply charged peptide anions for acidic proteome characterization. *PROTEOMICS* **2011**, *11*, 1329–1334.
(17) Yeh, G. K.; Sun, Q.; Meneses, C.; Julian, R. R. Rapid peptide fragmentation without electrons, collisions, infrared radiation, or native chromophores. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 385–393.
(18) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. *Angew. Chem. Int. Ed.* **2009**, *48*, 7130–7137.
(19) Hendricks, N. G.; Lareau, N. M.; Stow, S. M.; McLean, J. A.; Julian, R. R. Bond-Specific Dissociation Following Excitation Energy Transfer for Distance Constraint Determination in the Gas Phase. *J. Am. Chem. Soc.* **2014**, *136*, 13363–13370.
(20) Feketeová, L.; Khairallah, G. N.; Brunet, C.; Lemoine, J.; Antoine, R.; Dugourd, P.; O'Hair, R. A. J. Fragmentation of the tryptophan cluster [Trp9-2H]2- induced by different activation methods. *Rapid Commun. Mass Spectrom. RCM* **2010**, *24*, 3255–3260.
(21) Aravind, G.; Klærke, B.; Rajput, J.; Toker, Y.; Andersen, L. H.; Bochenkova, A. V.; Antoine, R.; Lemoine, J.; Racaud, A.; Dugourd, P. Photodissociation pathways and lifetimes of protonated peptides and their dimers. *J. Chem. Phys.* **2012**, *136*, 014307.
(22) Girod, M.; Sanader, Z.; Vojkovic, M.; Antoine, R.; MacAleese, L.; Lemoine, J.; Bonacic-Koutecky, V.; Dugourd, P. UV Photodissociation of Proline-containing

Peptide Ions: Insights from Molecular Dynamics. *J. Am. Soc. Mass Spectrom.* **2014**.
(23) Moon, J. H.; Yoon, S. H.; Bae, Y. J.; Kim, M. S. Dissociation kinetics of singly protonated leucine enkephalin investigated by time-resolved photodissociation tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1151–1158.
(24) Shin, Y. S.; Moon, J. H.; Kim, M. S. Observation of phosphorylation site-specific dissociation of singly protonated phosphopeptides. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 53–59.
(25) Yoon, S. H.; Moon, J. H.; Kim, M. S. Dissociation mechanisms and implication for the presence of multiple conformations for peptide ions with arginine at the C-terminus: time-resolved photodissociation study. *J. Mass Spectrom. JMS* **2010**, *45*, 806–814.
(26) Yoon, S. H.; Moon, J. H.; Chung, Y. J.; Kim, M. S. Influence of basic residues on dissociation kinetics and dynamics of singly protonated peptides: time-resolved photodissociation study. *J. Mass Spectrom. JMS* **2009**, *44*, 1532–1537.
(27) 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. Tyrosine sulfation in a Gram-negative bacterium. *Nat. Commun.* **2012**, *3*, 1153.
(28) Guan, Z.; Kelleher, N. L.; O'Connor, P. B.; Aaserud, D. J.; Little, D. P.; McLafferty, F. W. 193 nm photodissociation of larger multiply-charged biomolecules. *Int. J. Mass Spectrom. Ion Process.* **1996**, *157-158*, 357–364.
(29) Vasicek, L.; Brodbelt, J. S. Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores. *Anal. Chem.* **2010**, *82*, 9441–9446.
(30) Vasicek, L. A.; Ledvina, A. R.; Shaw, J.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J.; Brodbelt, J. S. Implementing Photodissociation in an Orbitrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 1105–1108.
(31) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000**, *35*, 1399–1406.
(32) Papayannopoulos, I. A. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.* **1995**, *14*, 49–73.

Chapter 5

Modulation of Protein Fragmentation Through Carbamylation of Primary Amines

5.1 Overview

We evaluate the impact of carbamylation of the primary amines of the side-chains of lysines and the N-termini on the fragmentation of intact protein ions and the chromatographic properties of a mixture of *E. coli* ribosomal proteins. The fragmentation patterns of the six unmodified and carbamylated proteins obtained by higher energy collision dissociation (HCD) and ultraviolet photodissociation (UVPD) were compared. Carbamylation significantly reduced the total number of protons retained by the protein owing to the conversion of basic primary amines to non-basic carbamates. Carbamylation caused a significant negative impact on fragmentation of the protein by HCD (i.e. reduced sequence coverage and fewer diagnostic fragment ions) consistent with the mobile proton model which correlates peptide fragmentation with charge distribution and the opportunity for charge-directed pathways. In addition, fragmentation was enhanced near the N- and C-termini upon HCD of carbamylated proteins. For LCMS/MS analysis of *E. coli* ribosomal proteins, the retention times increased by 16 minutes on average upon carbamylation, an outcome attributed to the increased hydrophobicity of the proteins after carbamylation. As noted for both the six model proteins and the ribosomal proteins, carbamylation had relatively little impact on the distribution or types of fragment ions product by UVPD,

supporting the proposition that the mechanism of UVPD for intact proteins does not reflect the mobile proton model.

5.2 INTRODUCTION

Improved chromatographic methods coupled with high performance mass analyzers and increasingly sophisticated informatics have facilitated the efficient separation, analysis, and identification of intact proteins in the gas phase, thus inspiring great interest in top-down strategies for proteomics.[1–3] While measurement of the accurate mass of a protein is a crucial first step, complete characterization of a proteoform (*i.e.* a unique molecular form of a protein including its mutations and specific post-translational modifications) requires much more information about the sequence, as well as the identity, number and position of modifications.[4] There are several established methods to activate and dissociate intact proteins; collisionally activated dissociation (CAD[5]) and beam-type higher energy collision dissociation (HCD[6,7]) and electron-based methods (most commonly electron transfer dissociation (ETD [8,9]), have been used for the most significant high throughput top-down studies. Ultraviolet photodissociation (UVPD) is the newest activation method that has been developed for the analysis of intact proteins.[10–16] The absorption of high energy photons (typically 6.4 eV per 193 nm photon) results in extensive backbone cleavages that result in formation of *a,b,c,x,y,* and *z* ions. UVPD affords high sequence coverage, the ability to map sites of post-translational modifications, and has shown promise for top-down LC-MS applications.[16]

Top-down methods have not reached the widespread adoption of bottom-up methods for high throughput proteomics, in part owing to less effective activation methods for intact proteins.[1] In the context of activation of proteins, performance metrics tend to decrease with increasing mass, and charge state plays a major role.[17,18] Upon electrospray ionization, protein ions are generated in a wide array of charge states, thus making it especially important to more extensively evaluate and understand the impact of charge state on the fragmentation of proteins.[5,19] There have been several systematic studies of protein dissociation using collisional activation, including ones that have examined the influence of charge state and other factors on fragmentation pathways.[19–23] CAD of intact proteins depends on proton mobility for fragmentation, as also well-recognized for peptide fragmentation induced by collisional activation.[24] Protons are typically sequestered at the more basic sites (Arg, Lys, His, N-terminus), but these protons can be mobilized via addition of energy to the ion.[25] For proteins in higher charge states, the additional protons associated with less basic sites along the backbone facilitate *b/y* fragmentation pathways.[24,26,27] As similarly noted for peptides, cleavages are preferentially enhanced at acidic residues for lower charge states (i.e. absence of mobile protons). McLuckey and co-workers have shown that CAD of intact proteins in low charge states (i.e. ones typified by low proton mobility) results in enhanced cleavage at glutamic acid and aspartic acid residues and a reduction in other diagnostic backbone *b/y* fragments compared to fragmentation of higher charge states having a greater number of mobile protons.[20] In another CAD study of intact proteins, Agar *et al.* reported enhanced cleavages adjacent to

glycine, lysine, glutamine and N-terminal to serine, tyrosine, isoleucine, leucine and proline, none of which are prominent in typical CAD spectra of tryptic peptides.[23]The irregular distribution of basic residues in proteins may give rise to these uncommon cleavage pathways. Not surprisingly, CAD spectra of tryptic peptides with missed cleavages (i.e. peptides having a basic Arg residue other than at the C-terminus) often display these same dissociation pathways.[23]

The production of diagnostic *c/z* sequence ions upon ETD of intact proteins is also highly dependent on the charge density and charge state. ETD of proteins in low charge states result in far fewer fragments and reduced sequence coverage compared to ones in higher charge states owing to the propensity for non-dissociative charge reduction associated with electron attachment.[28,29] The ability of ETD to map post-translational modifications remains a particularly compelling advantage which balances the sub-par performance for proteins in lower charge states.[30,31] In contrast to collisional and electron-based activation, UVPD has shown less dependence on charge state, and many of the fragmentation processes do not require mobile protons.[32]

The number and locations of charge sites can be further modulated by addition of supercharging agents to the solutions or by derivatization to convert specific functional groups to more or less basic ones.[33] Williams and Iavarone studied the impact of supercharging on fragmentation of intact proteins. Supercharging was achieved via addition of *m*-nitrobenzyl alcohol to the solution prior to ESI.[34] Collisionally activated fragmentation of the supercharged states yielded a small number of highly abundant

fragment ions clustered around narrow stretches of the protein backbone compared to the more widespread fragmentation observed upon CAD of intermediate charge states.[35] The Smith group manipulated the charge states of intact proteins via chemical derivatization of acidic sites and addition of basic moieties or ones with fixed charges.[36] Capping acidic sites with neutral moieties shifted the charge states very little during ESI, suggesting that carboxylic acid side-chains played a relatively minor role in determining the charge states of proteins upon ESI. Addition of basic and fixed charge moieties had a significant impact on the charge state of intact denatured proteins, suggesting that the number of basic sites modulated the range of charge states adopted by the proteins.[36] Guanidination increases the basicity of lysine residues and promotes proton sequestration (reducing proton mobility), and was used as a means to probe the influence of proton mobility on CAD of ubiquitin.[37] For the 10+ charge state of ubiquitin, the resulting fragmentation of the guanidinated protein occurred largely C-terminal to aspartic acid in a charge remote fashion as predicted by the mobile proton theory. For the 10+ charge state of non-guanidinated ubiquitin, non-specific amide bond cleavages to produce traditional *b/y* ions and enhanced cleavage N-terminal to proline were observed. This contrast in fragmentation behavior that arose from guanidination demonstrated that reduction in proton mobility restricted the non-specific fragmentation pathways.[37]

In the present study we directly evaluate the dependence of UVPD and HCD on proton mobility and charge state for several proteins. In order to affect both protein charge state and proton mobility, we employ a highly efficient carbamylation reaction which

converts the basic primary amines of lysine sidechains and the N-terminus to less basic carbamates. Not only does this reduce the average charge state adopted by a given protein upon electrospray ionization, it removes sites of proton sequestration which alters proton mobility. The impact of charge state and proton mobility on HCD and UVPD of six proteins and a mixture of ribosomal proteins were investigated. Comparisons of sequence coverage, distributions of sequence ions, and fragment ion type are reported for multiple charge states of unmodified and carbamylated proteins.

5.3 EXPERIMENTAL

5.3.1 Materials

Ubiquitin (bovine), cytochrome c (equine), myoglobin (bovine), superoxide dismutase (bovine), lysozyme (galline), carbonic anhydrase (bovine), and urea were obtained from Sigma-Aldrich (St. Louis, MO). *E. coli* 70S ribosome was obtained from New England Biolabs (Ipswich, MA). Solvents were obtained from EMD Millipore (Billerica, MA). Tris(2-carboxyethyl) phosphine (TCEP) was obtained from Thermo-Scientific (Rochford, IL)

5.3.2 Carbamylation

Carbamylation was performed as previously reported.[33] Briefly, each sample was split into two aliquots; one for derivatization and one as a control. Each was suspended in 200 mM Tris-HCl in the presence or absence of 8 M urea. Both samples were incubated at

80 °C for 4 h. Samples were desalted using Amicon Ultra 3kDa MWCO spin columns (EMD Millipore; Billerica, MA), then evaporated to dryness and resuspended in solvent to match the LC starting conditions (2% acetonitrile/98% water/0.1% formic acid) or infusion conditions (50% methanol/50% water/1% formic acid).

5.3.3 Separation

Proteins were separated using a Dionex RSLC 3000 nano-liquid chromatograph (Thermo Fisher, San Jose, CA) Approximately 1 µg of proteins were injected onto a 3 cm PLRP reverse phase trapping column (75 µm i.d.) packed with 5 µm particles (1000 Å pore size). Proteins were then eluted onto a 40 cm fritted 75 µm i.d. pulled tip analytical column (New Objective, Woburn MA) packed in-house with PLRP (5 µm particles, 1000 Å pore size) at a flow rate of 300 nL/min using a linear gradient of 2%-50% solvent B (acetonitrile/0.1% formic acid) over 120 minutes. Solvent A was water/0.1% formic acid.

5.3.4 Mass Spectrometry

Proteins for infusion were suspended in a solution of water, acetonitrile and formic acid (49.5/49.5/1) at a final concentration of 10 µM. For proteins having known disulfide bonds, a 20X molar excess of TCEP was added to the solution prior to infusion. The proteins were either infused directly at 3 µL/min using a HESI II Source (Thermo Fisher Scientific, San Jose CA) or introduced by nano LC ESI into an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose CA) customized for implementation of UVPD as described previously.[38]

Spectra were analyzed in the Orbitrap mass analyzer at a resolving power of 120,000 at m/z 200, using Intact Protein Mode. 250 scans were collected and averaged for infusion experiments. LC-MS data was collected in a top speed (7 s cycle) data-dependent manner where each MS1 consisted of 4 μscans (AGC target of 1.0E+05, max injection time of 100 ms) and MS2 consisted of 6 μscans (AGC target of 5.5E05, max injection time of 250 ms). Precursor ions were filtered according to intact protein monoisotopic precursor selection, thus focusing on proteins with charge states greater than 5+. Spray voltage was set to 1.8 kV. MS2 isolation width was set to 5 m/z using the quadrupole for mass filtering. Precursors selected more than five times in 120 s were excluded from MS2 selection for 120 s. HCD normalized collision energy (NCE) was optimized (10-30NCE) per charge state for infusion experiments and set to 20NCE for LC-MS experiments. UVPD performed in the high pressure cell of the dual linear ion trap was achieved via a single 5 ns laser pulse from a Coherent ExciStar XS 500 (Santa Clara, CA) 193 nm excimer laser. Laser power was set to 1.0 mJ for both infusion and LC-MS experiments.

5.3.5 Data Analysis

High resolution intact protein fragmentation spectra were deconvoluted using the Xtract algorithm enabled in Thermo XCalibur Qual Browser (Thermo Fisher Scientific, San Jose, CA.). The deconvoluted data was further processed via Prosight Lite build 1.3.5744.1622 (http://prosightlite.northwestern.edu/) to generate sequence coverage maps and confirm the degree of carbamylation based on the presence of fragment ions within a 10 ppm tolerance [39]. Only fragment ions that contain the N-terminal or C-terminal residue

of the sequence are searched and identified; internal ions are not identified. Protein backbone cleavage maps were generated using msProduct (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct) outputs to assign cleavage position, fragment intensity and ion type. These results were further processed and represented graphically using Microsoft Excel (Microsoft, Redmond WA.)

LC-MS data was processed using ProSight PC 4.0. The protein sequence database was populated using the Uniprot reviewed *E. coli* K12 database (accessed October 2016). To enable analysis of the MS/MS spectra of carbamylated proteins, a custom sequence database was created which treated all lysine residues as carbamylated and N-termini as carbamylated or acetylated. The results were filtered at the proteoform level using a P-Score cutoff of 1.0E-05.

5.4 RESULTS & DISCUSSION

This study focuses on evaluating the impact of modifying charge sites (lysines, N-termini) of intact proteins on the outcome and metrics of HCD and UVPD. Six model proteins and a ribosomal protein mixture were introduced using infusion or via nanoLC, respectively, then subjected to HCD and UVPD. In particular, this work aims to compare the fragmentation of carbamylated and unmodified proteins to evaluate the influence of mobile protons and charge state upon HCD and UVPD. Carbamylation of the primary amines of lysines and the N-terminal amine converts them to non-basic groups (**Figure 5.1**), leaving arginine residues as the most basic sites, followed by histidines. Carbamylation of proteins changes not only the number of protons retained by each protein

during ESI but also the localization and mobility of protons, as evidenced by the sometimes significant variations in fragmentation observed in the MS/MS spectra generated upon HCD, as described in more detail below. Six model proteins were selected to span an array of molecular sizes and have a range of number, locations and distributions of basic Lys/Arg sites. For example, lysozyme, cytochrome c and myoglobin have similar total numbers of highly basic sites (20, 18, and 22, respectively), but the number of arginine residues varies considerably (only two for cytochrome c and myoglobin, but 11 for lysozyme). Since fragmentation near the termini is dominant upon collisional activation of intact proteins, proteins were selected with lysines near the N-terminus (ubiquitin, cytochrome c, lysozyme, superoxide dismutase) or near the C-terminus (cytochrome c, superoxide dismutase, carbonic anhydrase) or with arginines in those segments. A list of the six protein sequences is provided in **Figure 5.2**.

5.4.1 Effect of carbamylation on charge state distribution of intact proteins

Prior to evaluating the variations in fragmentation patterns upon carbamylation, first the distributions of charge states of unmodified versus carbamylated proteins were examined. The charge states adopted by a protein upon ESI are influenced by several factors, including protein size, number of basic and acidic residues, solvent composition and solvent additives, among others. In this study proteins were sprayed using conventional denaturing conditions prior to or after carbamylation of all lysines and N-termini. For example, ubiquitin (8.5 kDa) contains seven lysine residues and was modified a total of eight times upon carbamylation with nearly 100% efficiency, indicating highly efficient

carbamylation of all seven primary amines of the lysine side-chains plus the N-terminus **(Figure 5.3**). **Figure 5.3** shows the ESI mass spectra of unmodified and carbamylated ubiquitin, myoglobin, and carbonic anhydrase (and the same pairs of ESI mass spectra are shown in **Figure 5.4** for cytochrome c, lysozyme, and superoxide dismutase). For each of the six proteins, the shift in the charge state distributions after carbamylation is dramatic. The range of charge states for unmodified ubiquitin is +6 to +13, whereas it is +4 to +9 after carbamylation. Myoglobin (16.9 kDa) has 19 lysines and displays charge states from +10 to +24 prior to carbamylation and +7 to +16 after carbamylation. Carbonic anhydrase (29 kDa, 18 Lys) also shows a significant shift in charge state distribution upon carbamylation, ranging from +14 to +35 prior to carbamylation and from +10 to +28 after carbamylation. The same shift in charge states also occurs for cytochrome C, lysozyme, and superoxide dismutase (**Figure 5.4**). The significantly lower basicity of carbamylated groups in comparison to primary amines accounts for the notable reduction in charge states of the proteins.

For the subsequent MS/MS experiments described in the next sections, several charge states were selected for HCD and UVPD. Typically one charge state higher than the median charge state and one charge state lower than the median were selected for MS/MS analysis, as well as one charge state that "overlapped" between each unmodified and carbamylated protein. Owing to the incredibly rich MS/MS spectra of intact often containing more than 100 fragment ions, displaying numerous annotated spectra is cumbersome and thus an alternative graphical representation was used for this study. In

order to show the distribution of cleavages of the backbone, the relative fragment ion abundances originating from cleavages at each backbone position were plotted as histograms spanning the protein sequence. The abundances of all N-terminal (*a,b,c* ions) and C-terminal (*x,y,z* ions) corresponding to each inter-residue position were summed, and the two sums were stacked and placed at their appropriate inter-residue cleavage site along the protein backbone.

5.4.2 Ubiquitin

Collisional activation of the 10+ and 12+ charge states of ubiquitin (**Figure 5.6**) resulted in remarkably similar fragmentation patterns, exhibiting significant cleavage C-terminal to three glutamic acid residues (Glu16, Glu18, Glu18), as well as N-terminal to proline (P19). This pattern shows that fragmentation of ubiquitin is dominated by preferential pathways (e.g. adjacent to acidic residues and proline); similar behavior has been reported previously [15,[20],[23]. There were also many non-specific cleavages across much of the backbone, yielding numerous low abundance *b* and *y* ions and resulting in similar total sequence coverages for both charge states (84% for 10+ and 89% for 12+). Collisional activation of the 10+ charge state of carbamylated ubiquitin (**Figure 5.7**) resulted in a similar preference towards proline-mediated cleavage; however, the dominant cleavage occurred N-terminal to a different proline residue: Pro37 (IleI36/Pro37) instead of Pro19 and HCD resulted in somewhat lower sequence coverage (72%) compared to the 10+ charge state of the unmodified protein.

HCD of one representative low charge state (6+) of unmodified ubiquitin displayed extensive cleavage across the entire backbone (**Figure 5.6**), along with enhanced cleavage N-terminal to P19 and N-terminal to P37 and a number of enhanced cleavages C-terminal to acidic residues (Asp32, Glu34, Asp39, Asp52, and Asp57). HCD of the corresponding 6+ charge state of carbamylated ubiquitin showed prominent fragmentation channels N-terminal to Pro19 and Pro37 as well as an array of nonspecific backbone cleavages spanning residues Ile3 to Glu18 (**Figure 5.7**). The fragmentation pattern of carbamylated ubiquitin (6+) upon HCD most closely resembled the fragmentation of the 10+ charge state of unmodified ubiquitin. With respect to this similarity in fragmentation behavior, it appears that the lower charge state of carbamylated ubiquitin (6+) was in part compensated by the greater mobility of protons upon carbamylation of the lysine side-chains. Interestingly, the HCD fragmentation pattern of the 4+ charge state of carbamylated ubiquitin did not exhibit the preferential Pro cleavage observed for all of the other charge states, and instead cleavages C-terminal to acidic residues were exceptionally prominent (Asp21, Asp32, Asp39, Asp52, Asp58, Glu64). The exaggerated enhancement of cleavages C-terminal to acidic residues has been noted previously for low charge states of unmodified proteins,[15,40–42] and it is echoed for the very low charge state (4+) of carbamylated ubiquitin in the present study. In general, the fragmentation patterns of each of the three representative charge states of carbamylated ubiquitin (4+, 6+, 10+) displayed significant differences in the locations and sites of preferential cleavages, and the total

sequence coverages (56%-72%) were notably lower than the coverages obtained for the unmodified protein (83%-89%) upon HCD (see **Table 5.1**).

Unlike the variations in fragmentation patterns observed upon HCD of unmodified ubiquitin, the fragmentation patterns generated upon UVPD are nearly independent of charge state and only modest differences are noted among the backbone cleavage histograms of the 6+, 10+, and 12+ charge states (**Figure 5.8**). Closer inspection of the histograms reveals subtle variations in the relative portions of C- and N-terminal product ions, but overall the fragmentation is considerably more uniform across the backbone upon UVPD compared to HCD. This trend is also reflected in the consistently high sequence coverage obtained from UVPD of ubiquitin irrespective of charge state (99% - 100%, **Table 1**), an outcome which also holds true for carbamylated ubiquitin (87% - 99%). For all charge states, UVPD of carbamylated ubiquitin showed suppressed backbone fragmentation in the stretch from Pro19 to Ile36 compared to the unmodified protein, along with significantly lower abundances of N-terminal fragment ions across the entire sequence (**Figure 5.9**). However, in contrast to HCD of carbamylated ubiquitin, which favored preferential cleavages at proline and acidic residues, UVPD of carbamylated ubiquitin largely exhibited non-specific backbone fragmentation akin to the pattern observed upon UVPD of unmodified ubiquitin.

5.4.3 Cytochrome c

Upon ESI, cytochrome c retains the heme group (+616.191 Da) bound at C14 and C17, and this heme group is incorporated in assignment of all fragment ions that encompass

those cysteine residues. HCD of the 12+, 14+ and 16+ charge states of cytochrome c (**Figure 5.10**) resulted in a dominant cleavage N-terminal to proline (P77). Backbone cleavages adjacent to two lysine residues (K26 and K28) were also prominent, and HCD promoted a wide variety of non-specific backbone cleavages to produce ample series of *b/y* ions. For each of these charge states, HCD resulted in high sequence coverage (72% - 88%, **Table 1**) as evidenced by relatively broad fragmentation along the backbone, aside from a few stretches of little or no fragmentation (Cys14 to Thr19, Phe36 to Gln42). HCD of carbamylated cytochrome c (**Figure 5.11**) also resulted in enhanced backbone cleavage N-terminal to Pro77; however, aside from consistent fragment ions near the C- and N-termini, the MS/MS spectra obtained for the 6+, 8+ and 10+ charge states exhibited large stretches lacking any fragmentation and the overall sequence coverages were significantly lower for carbamylated cytochrome c (50% - 70%).

In contrast to the fragmentation patterns observed upon HCD of cytochrome c, UVPD (**Figure 5.12** does not result in a dominant cleavage N-terminal to Pro77, and rather backbone cleavage N-terminal to Pro30 and Pro45 were more prominent pathways. Moreover, UVPD resulted in non-specific fragmentation across nearly the entire backbone that yielded high sequence coverages for all charge states (87% to 90%). The heme-binding domain covering the stretch from Cys14 though Cys17 remain resistant to fragmentation by both HCD and UVPD, confirming the stabilization of this region owing to the thioether bonds between the heme and two cysteine residues.

UVPD of carbamylated cytochrome c resulted in extensive non-specific fragmentation across the backbone (**Figure 5.13**). At the lowest charges state (6+) carbamylated cytochrome c exhibited preferential cleavage N-terminal to Pro77; however, as charge state increased (8+ and 10+) this cleavage became less prominent and the cleavages N-terminal to Pro30 and Pro45 were enhanced at higher charge states. As noted earlier for the ubiquitin analysis, carbamylation had a much lower impact on the fragmentation of cytochrome c, regardless of charge state, compared to the far more striking impact on HCD in which backbone fragmentation throughout the protein was significantly curbed. The sequence coverage afforded by UVPD was 84% to 91% for the various charge states of carbamylated cytochrome c (**Table 1**), nearly identical to the range obtained for the unmodified protein and well above that obtained upon HCD.

5.4.4 Lysozyme

Lysozyme, which naturally has four disulfide bonds, was reduced prior to MS/MS analysis to mitigate the well-known suppression of fragmentation caused by disulfide bonds in proteins. Upon ESI, lysozyme produced ions in charge states that spanned 9+ to 20+ (**Figure 5.4**). HCD of lysozyme in the 12+, 15+, and 18+ charge states resulted in approximately 50% sequence coverage (**Table 1)**, with significant gaps at the N- and C-termini as well as the mid-section of the protein. For the two lower charge states (12+ and 15+), some C-terminal and N-terminal ions were observed across the entire backbone, whereas only N-terminal *b* ions were observed exclusively for the first half of the sequence and only C-terminal *y* ions were observed for the last half of the sequence (**Figures 5.4** and

5.14). This trend for selective formation of N-terminal and C-terminal product ions was further enhanced for the 18+ charge state and fragmentation was preferentially clustered around a few residues (e.g. Asn27/Trp28, Ile88/Thr89).

Lysozyme has six lysine residues and therefore was carbamylated at seven positions (N-terminus, Lys1, Lys13, Lys33, Lys96, Lys97, Lys116). After carbamylation the charge states ranged from 7+ to 14+. HCD of carbamylated lysozyme (**Figure 5.5** and **5.15**) resulted in markedly sparser fragmentation dominated by cleavages consolidated in the stretch from Asn26 to Phe33 (NWVCAAKF) and Lys96 to Asn103 (KKIVSDGN). These two regions are the longest stretches of the protein sequence which contain no basic Arg or His residues and only contain carbamylated Lys residues. The limited fragmentation observed for the carbamylated protein resulted in low sequence coverages (<30% **Table 1**) for all three charge states. Similar to the behavior observed upon HCD of the higher charge states of the unmodified protein, predominantly N-terminal *b* type ions were produced for N-terminal half of the protein and C-terminal *y* ions for the C-terminal half of the protein. Four of the most dominant products generated upon HCD of unmodified lysozyme arose from cleavage C-terminal to aspartic acid. Only one of these cleavages remained prominent after carbamylation.

As illustrated in **Figure 5.5** and **5.16**, UVPD of lysozyme (12+, 15+, and 18+ charge states) yielded nonspecific cleavages across nearly the protein backbone independent of charge state and resulted in far greater sequence coverages (averaging 85%, **Table 1**) than observed for HCD (averaging 54%). Interestingly the number of

complementary pairs of N-terminal and C-terminal fragment ions decreased as charge state increased, and in the 18+ charge state the majority of C-terminal fragment ions were restricted to the C-terminal half of the protein just as the majority of N-terminal fragment ions were restricted to the N-terminal half of the protein. This result was similar to the segregation of C-terminal and N-terminal fragment ions noted above for HCD of lysozyme, albeit with a far greater total number of backbone cleavage sites observed upon UVPD.

The sequence coverage obtained upon UVPD of carbamylated lysozyme (**Figure 5.5** and **5.17** for 8+, 10+, 12+) decreased (averaging 62%) compared to the coverage obtained upon UVPD of unmodified lysozyme (averaging 85%), just like was noted upon HCD. However, there was still significantly greater fragmentation throughout the protein upon UVPD than observed upon HCD. The prominent fragmentation across the Asn27-Phe34 stretch was also observed upon UVPD, like HCD. The UVPD fragmentation trends showed less dependence on charge state for both the unmodified and carbamylated protein compared to HCD, again attesting to the reduced impact of mobile protons on modulation of fragmentation.

5.4.5 Superoxide dismutase

Upon ESI, superoxide dismutase (SOD) produced ions in charge states ranging from 10+ to 23+ after reduction of disulfide bonds. HCD of SOD in the 12+, 16+, and 20+ charge states resulted in nonspecific cleavage (**Figure 5.18)** of the backbone to yield approximately 50% sequence coverage (**Table 1**). The most extensive fragmentation was concentrated in the regions spanning Val27 to Gly35 and Gly91 to Tyr108. SOD exhibited

C-terminal and N-terminal product ion segregation that increased with charge state such that only N-terminal *b* ions are dominant for the first half of the protein and C-terminal *y* ions are dominant for the second half of the protein.

SOD underwent efficient carbamylation at all ten lysine residues (the acetylated N-terminus was not modified) and the dominant charge state shifted from 20+ for the unmodified protein to 14+ after carbamylation. Similar to the trend noted for the other proteins, the sequence coverage obtained upon HCD of carbamylated SOD decreased significantly relative to that of unmodified SOD, even for the same charge state (e.g. 35% coverage for carbamylated SOD (12+) and 55% coverage for unmodified SOD (12+)) (**Table 1**). HCD of carbamylated SOD decreased as the charge state decreased (**Figure 5.19**), and overall was sparse compared to HCD of the unmodified protein. For the 12+ charge state, fragmentation was prominent only near N-terminus and the region spanning Gly92 to Asp99 with minor contributions near the C-terminus. For the 14+ and 16+ charge states, HCD resulted in some selective fragmentation in the middle region of the sequence, including stretches from Val27 to Thr37 and Asn84 to Pro100.

UVPD of both unmodified SOD (12+, 16+ and 20+) and carbamylated SOD (12+, 14+, 16+) resulted in numerous nonspecific cleavages (**Figures 5.20, 5.21**) across the protein backbone and was virtually independent of charge state, yielding sequence coverages that averaged 73% (**Table 1**). Both N-terminal and C-terminal products were generated throughout the protein. Enhancement of cleavages adjacent to proline residues (Pro13, Pro64, Pro72, Pro100, Pro121) was observed for both unmodified and

carbamylated SOD. What is perhaps most remarkable about the fragmentation of SOD is the lack of overlap among the regions of enhanced fragmentation for HCD versus UVPD and for the carbamylated versus unmodified protein. For example, the region of greatest backbone fragmentation spanned residues Val92 to Gly106 for HCD of unmodified SOD (16+), residues Val92 to Pro100 for HCD of carbamylated SOD (16+), residues His19 to Thr86 for UVPD of unmodified SOD (16+), and residues Phe43 to Pro100 for UVPD of carbamylated SOD (12+). This comparison highlights the complementary nature of HCD and UVPD, as well as the impact of carbamylation on fragmentation.

5.4.6 Myoglobin

Myoglobin produced ions in charge states ranging from 9+ to 24+; after carbamylation the charge states ranged from 7+ to 16+ and the clean spectra confirmed that carbamylation occurred with near 100% efficiency at 20 sites (19 lysines plus the N-terminus). HCD of myoglobin in the 12+, 16+, and 20+ charge states resulted in nonspecific cleavage across much of the backbone (**Figures 5.6** and **5.22**) and yielded an average of 50% sequence coverage (**Table 1**). Several prominent preferential cleavages were observed, such as N-terminal to Pro120, and C-terminal to both Leu2 and Glu6. The trend of significant *b* and *y* ion segregation (i.e. *b* ions preferentially observed for the N-terminal half of the protein and *y* ions dominating for the C-terminal half) was again noted, especially for the higher charge states. After carbamylation, there was a significant decrease in fragmentation of myoglobin upon HCD as evidenced by sequence coverages that averaged only 20% (**Table 1**). The few sequence ions that were observed were

clustered near the first 20 residues of the N-terminus (*b* ions) or C-terminus (mostly *y* ions), leaving the mid-section of the protein unsequenced (**Figure 5.23**). Myoglobin contains a large number of K residues in the region of the protein devoid of fragments (15 Lys out of 19 total Lys), suggesting that heavy modification of the internal portion of the protein disrupts the formation of fragment ions by HCD. Myoglobin contains only two highly basic arginine residues (Arg31, Arg139) in its entire sequence, meaning that the 11 His residues serve as the other basic sites mostly likely to be protonated during ESI. Interestingly, the regions which do yield fragment ions do not contain His residues, suggesting that histidine may sequester protons or otherwise hinder fragmentation of carbamylated proteins.

UVPD of myoglobin resulted in mainly nonspecific cleavages which resulted in high sequence coverage (up to 94% for the 16+ charge state and averaging 92% for all of the charge states examined, **Table 1**). Cleavage N-terminal to Pro120 was observed upon UVPD, just as it was prominent upon HCD, but cleavage C-terminal to Phe33, Phe46, and Phe48 were of similar relative abundance to that of the P120 cleavage (**Figure 5.24**), suggesting that fragmentation may be enhanced adjacent to aromatic residues which are known to have high UV photoabsorption cross-sections. UVPD of carbamylated myoglobin results in numerous nonspecific cleavage along the backbone **Figure 5.25**), yielding sequence coverages from 74% to 83% for the 8+, 10+, and 12+ charge states (**Table 1**). Although this level of sequence coverage is lower than the average 92% coverage observed for unmodified myoglobin upon UVPD, it is three to four times greater than observed upon HCD. Several prominent backbone cleavages were noted C-terminal

to Phe33 and N-terminal to several Pro residues which were identical to those observed upon UVPD of unmodified myoglobin. In general, myoglobin showed the greatest difference in overall fragmentation between HCD of the carbamylated and unmodified protein and most similarity for UVPD of the unmodified and modified protein. This finding suggests that for proteins which are arginine poor, modification of lysine residues severely hinders production of *b* and *y* type ions upon collisional activation but has little impact on the performance or outcome of UVPD.

5.4.7 Carbonic anhydrase

Upon ESI, carbonic anhydrase produced ions in charge states ranging from 14+ to 35+ **(Figure 5.3**). HCD resulted in rather sparse fragmentation of all charge states (**Figure 5.26**), yielding sequence coverages averaging only 35%. Several dominant cleavages occurred at Pro residues, and the majority of fragments entailed backbone cleavages within 75 residues of either the C- or N-termini. The highest charge state examined by HCD (34+) displayed a shift towards fragmentation at only the C-terminus region. Carbamylation of carbonic anhydrase was efficient, resulting in modification of all 18 Lys residues plus the C-terminus and shifting the charge states to 10+ to 28+ upon ESI (**Figure 5.3**). HCD of carbamylated carbonic anhydrase resulted in preferential cleavages similar to that observed for unmodified carbonic anhydrase, albeit with even lower sequence coverage (averaging 20%) and with virtually no fragmentation along the N-terminal half of the protein (**Figure 5.27**).

UVPD of carbonic anhydrase resulted in a larger degree of non-specific cleavages (**Figure 5.28**), yielding sequence coverages that averaged 63% (**Table 1**). Similar to HCD, several cleavages adjacent to Pro residues were moderately enhanced. UVPD of carbamylated carbonic resulted in an average sequence coverage of 32% (**Table 1**). Similar to what was observed for HCD, a large stretch of the protein remained unsequenced upon UVPD of the carbamylated protein (**Figure 5.29**), although the suppression of fragmentation was less dramatic than noted for HCD. The UVPD fragmentation patterns of carbamylated carbonic anhydrase (16+, 22+, 29+) most closely resembled the UVPD fragmentation pattern of unmodified carbonic anhydrase in the 34+ charge state, supporting the observation that fragmentation patterns of carbamylated proteins are most similar to the ones obtained for the highest charge states of the unmodified counterparts.

5.4.8 Effect of carbamylation on sequence coverage

Sequence coverage for all six proteins studied was lowest for carbamylated proteins activated by HCD and highest for unmodified proteins activated by UVPD (**Table 1** with sequence maps shown in **Figures 5.30-5.35** for each of the six proteins). In general, carbamylation caused reduced sequence coverage regardless of activation method, and the reduction in sequence coverage for the carbamylated proteins was notably more precipitous for HCD than UVPD. UVPD routinely achieved higher coverage and consistently outperformed HCD for both carbamylated and unmodified proteins. Myoglobin, the most lysine-rich protein, displayed the largest disparity between HCD and UVPD of the carbamylated species (77% coverage for UVPD and only 22% for HCD over three charge

states), an outcome also mirrored for the unmodified proteins (92% coverage for UVPD and only 54% for HCD on average). The relatively modest reduction in sequence coverage for UVPD of carbamylated proteins relative to unmodified proteins recapitulates the lack of significant dependence on proton mobility for the UVPD process relative to HCD. Based on analysis of the series of fragment ion maps, HCD of carbamylated proteins enhances terminal-mediated fragmentation (i.e. enhancement of smaller *a,b,c* ions near the N-terminal portion of the protein, *x,y,z* ions near the C-terminal portion, and sparse fragmentation in the mid-section) which could explain the reduced sequence coverage as protein size increases.

5.4.9 Effect of carbamylation on cleavage preferences

Under certain circumstances, collisional activation of proteins generates dominant fragment ions resulting from preferential cleavages, typically ones directly related to specific amino acids.[15,23] In general, this phenomena is most prominent upon activation of the highest and lowest charge states and is most frequently manifested in preferential cleavages N-terminal to Pro and C-terminal to Asp and Glu residues.[43–45] To assess the overall impact of carbamylation on these preferential cleavages, the distributions of key categories of fragment ions of ubiquitin (**Figure 5.36**), cytochrome c (**Figure 5.37**), and were compiled owing to the relatively extensive sequence coverage of these three proteins regardless of carbamylation or activation method.

HCD of unmodified ubiquitin showed an increasing degree of preferential N-terminal proline cleavage with increasing charge state and a significant portion of

preferential cleavages C-terminal to acidic residues (Glu and Asp) (**Figure 5.36),** similar to previous observations by Reid and McLuckey.[20] Upon carbamylation of ubiquitin, N-terminal proline cleavage was significantly suppressed for the +4 charge state, whereas cleavage C-terminal to Asp and Glu were greatly enhanced, suggesting that charge remote fragmentation dominates at this low charge state generated upon carbamylation (**Figure 5.36**). The 10+ charge state of carbamylated ubiquitin displayed a degree of N-terminal Pro cleavage upon HCD similar to that observed for the 12+ of unmodified ubiquitin and a reduced percentage of C-terminal Asp and Glu cleavages. UVPD of both unmodified and carbamylated ubiquitin resulted in lower portions of preferential cleavages and significantly more contributions from non-specific cleavages of the entire backbone.

HCD resulted in similar portions of preferential N-terminal Pro and C-terminal Glu and Asp cleavages of cytochrome c regardless of charge state, and these distributions changed only modestly upon carbamylation (**Figure 5.37**). In particular, Pro-specific fragmentation was enhanced after carbamylation for the 8+ charge state of cytochrome c. The contribution of preferential cleavages diminished for HCD of the 10+ charge state of carbamylated cytochrome c, but this distribution was likely skewed owing to the notable enhancement of non-specific cleavages at the C-terminus end of the protein. UVPD of cytochrome c displayed more dominant fragmentation C-terminal to Phe than HCD, reiterating that the aromatic chromophore played a role in directing site-specific fragmentation by UVPD. Interestingly, UVPD of both unmodified and carbamylated cytochrome c resulting in enhanced N-terminal Pro cleavage and somewhat suppressed C-

terminal Asp and Glu cleavages compared to HCD. In addition, the sites of the preferential cleavages varied for HCD relative to UVPD. Cleavage adjacent to Pro44 and Pro76 were favored for HCD, whereas UVPD displayed enhanced N-terminal cleavage at two additional proline residues: Pro30 and Pro71. Thus, UVPD did not show any particular discrimination of Pro residues, whereas the preferential Pro cleavage was selective upon HCD, suggesting that specific sequence motifs or charge site locations influenced the proline cleavages of cytochrome c upon HCD.

Lysozyme contains only two proline residues with both located in the middle section of the protein, a region where fragmentation is typically suppressed for top-down MS/MS methods. Upon HCD, both unmodified and carbamylated lysozyme displayed little Pro-selective cleavage, and instead preferential cleavages adjacent to Glu and Asp were more prominent (**Figure 5.14**). Upon carbamylation a slight increase in C-terminal Asp and Glu cleavages was observed for the 8+ charge state upon HCD. The more extensive and non-selective fragmentation across the backbone by UVPD generated higher sequence coverage and resulted in ample N-terminal Pro and C-terminal Glu and Asp cleavages. Upon carbamylation, the degree of Pro cleavage decreased significantly upon UVPD, as well as a decrease in cleavage C-terminal to Phe.

5.4.10 LC-MS of carbamylated E.coli ribosomal proteins

Carbamylation not only causes a significant change in the charge states of proteins and their fragmentation patterns but also alters the chromatographic properties of proteins. To examine the potential impact of carbamylation on a top-down LCMS/MS workflow,

the *E.coli* ribosome containing 56 proteins was used as a benchmark mixture. **Figure 5.38** shows the base peak LCMS traces for a mixture of ribosomal proteins prior to and after carbamylation. For the mixture subjected to carbamylation, the composition of eluting proteins was checked throughout the LC run, and there was no evidence for non-carbamylated forms. In essence, carbamylation proceeded with near 100% efficiency. As specific examples of eluting proteins, **Figures** 5.**39** and **5.40** show extracted ion chromatograms and corresponding MS1 spectra of ribosomal protein L24 from the 50S subunit prior to and after carbamylation. The significant reduction in charge state upon carbamylation is evident in **Figures** 5.**39** and **5.40** for which the most abundant charge state shifts from 16+ (unmodified) to 11+ (carbamylated) for 50S L24. The retention time changes by over 36 minutes upon carbamylation, thus reflecting the increase in hydrophobicity after modification of the 16 lysines and N-terminus of this 11.2 kDa (103 residues) protein. In general, carbamylation increases the retention times of the ribosomal proteins by an average of 16 ± 7 minutes, an increase that is modulated by the number of lysine residues per protein (e.g. the hydrophobicity and concomitant change in retention time scales with the number of carbamylated groups). The sequence maps are shown in the lower half of **Figures** 5.**39** and **5.40**, and the following sequence coverages were obtained: 62% for HCD and 80% for UVPD of the L24 protein (16+) and 27% for HCD and 71% for UVPD of the carbamylated L24 protein (10+), again exhibiting a significant decrease in coverage upon HCD of the carbamylated protein and a much smaller decrease in coverage upon UVPD.

Table 5.2 summarizes LC-MS results comparing HCD and UVPD for the collective set of ribosomal proteins, with specific performance results (including molecular weight, retention time, and sequence coverage) for a subset of six individual proteins shown in **Table 5.3**. Carbamylation significantly reduced the number of ribosomal proteins and matched fragment ions identified by HCD but had a far more modest impact on the UVPD results. In fact, carbamylation resulted in a greater than 50% reduction in total protein identifications for HCD (from 53 to 24 proteins), with a similar trend for the number of proteoforms characterized (180 to 75 proteoforms), number of ribosomal proteins identified (42 to 22 proteins), and average number of fragments matched (23 to 13 matching fragments per protein). For the subset of proteins analyzed in detail in **Table 5.3**, the sequence coverage averaged 46% for the six unmodified proteins but plunged to 22% for the same carbamylated proteins. UVPD generated nearly 50% more matched fragments on average compared to HCD, regardless of carbamylation, indicating better protein characterization and sequence coverage (**Table 5.2**). Carbamylation caused little change in the number of proteins (46 versus 48) and proteoforms (134 versus 134) identified by UVPD. For the same subset of proteins reported in **Table 5.3**, the sequence coverage averaged 70% for the six unmodified proteins by UVPD and decreased to 60% for the carbamylated proteins.

5.5 CONCLUSION

Carbamylation offers a highly efficient means to modify the lysine sidechains and N-terminus of proteins, occurring with nearly 100% efficiency. Carbamylation of intact

protein molecules causes (i) reduction of the observed charge states upon ESI, (ii) modulation of HCD and UVPD fragmentation, and (iii) increases in LC retention times owing to the greater hydrophobicity after conversion of primary amines to carbamylated groups. Overall carbamylation significantly decreased the sequence coverage produced by HCD and resulted in a much more modest impact on UVPD. Fragmentation was preferentially enhanced near the N- and C-termini for HCD of carbamylated proteins (typically resulting in smaller fragment ions) with a concomitant reduction in fragmentation in the mid-sections (larger fragment ions). Although this apparent decrease in fragmentation in the mid-section of the protein could suggest that the larger fragments were converted to smaller N- and C-terminal fragment ions by secondary fragmentation pathways, there was no evidence for this premised based on examination of the fragmentation patterns as the HCD collision energy was varied. Carbamylation had relatively little influence on the distribution or types of fragment ions generated upon UVPD, recapitulating the premise that the mechanism of UVPD is not highly dependent on mobile protons. MS/MS analysis of intact proteins remains a formidable challenge, and carbamylation offers a convenient way to modulate charge states and resolve overlapping charge state distributions as has been previously shown for gas-phase proton transfer reactions.[15]

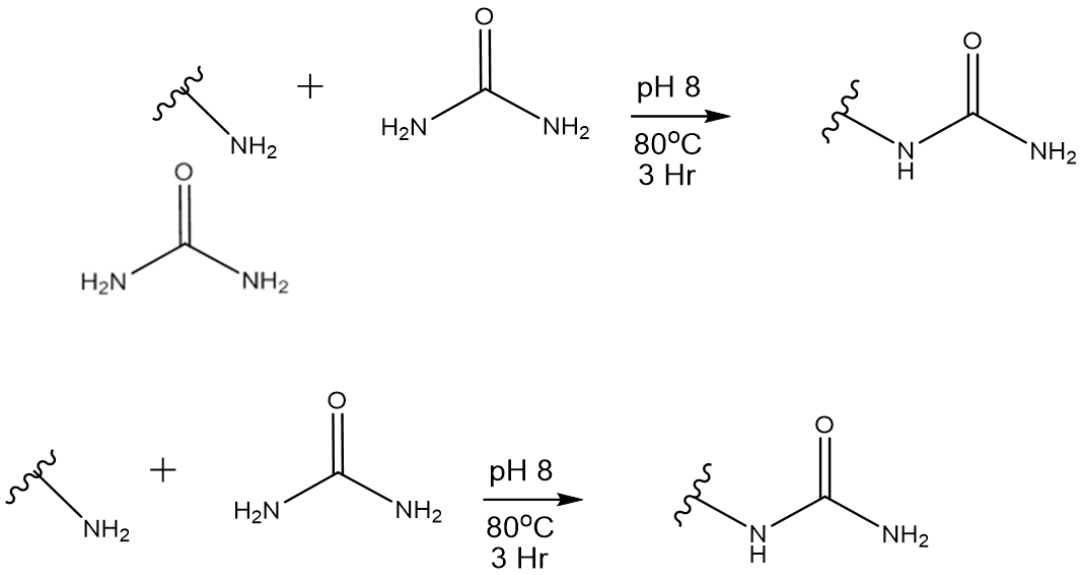


Figure 5.1 Carbamylation of primary amines

Ubiquitin (8.5 kDa, 76 aa, 2R/7K/11H)
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG

Cytochrome C (11.7 kDa, 105 aa, 2R/18K/3H)
GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFYTYDANKNKGITWKEETLMETYLENPKKYIPGTK
MIFAGIKKKTEREDLIAYLKKATNE

Lysozyme (15.5 kDa, 129 aa, 11R/6K/1H)
KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCNDGRTPGSRNLCNIP
CSALLSSDITASVNCAKKIVSDGNGMNAQVAWRNRCKGTDVQAQIRGCRL

Superoxide Dismutase (15.5 kDa, 151 aa, 4R/10K/8H)
ATKAVCVLKGDGPVQGTIHFEAKGDTVVVTGSITGLTEGDHGFHVHQFGDNTQGCTSAGPHFNPLSKKHGGPKDEERHVG
DLGNVTADKNGVAIVDIVDPLISLSGEYSIIGRTMVVHEKPDDLGRGGNEESTKTGNAGSRLACGVIGIAK

Myoglobin (16.9 kDa, 153 aa, 2R/19K/11H)
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKGH
HEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG

Carbonic Anhydrase (29.8 kDa, 259 aa, 9R/18K/1H)
SHHWGYGKHNGPEHWHKDFPIANGERQSPVDIDTKAVVQDPALKPLALVYGEATSRRMVNNGHSFNVEYDDSQDKAVLK
DGPLTGTYRLVQFHFHWGSSDDQGSEHTVDRKKYAAELHLVHWNTKYGDFGTAAQQPDGLAVVGVFLKVGDANPALQKV
LDALDSIKTKGKSTDFPNFDPGSLLPNVLDYWTYPGSLTTPPLLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLAN
WRPAQPLKNRQVRGFPK

Figure 5.2 The sequence of each protein is shown along with the molecular weight, the number of residues, and the composition of basic residues (Arg, Lys, His). Lysine residues are highlighted in bold red font; arginines are shaded in blue; histidines are shaded in yellow.

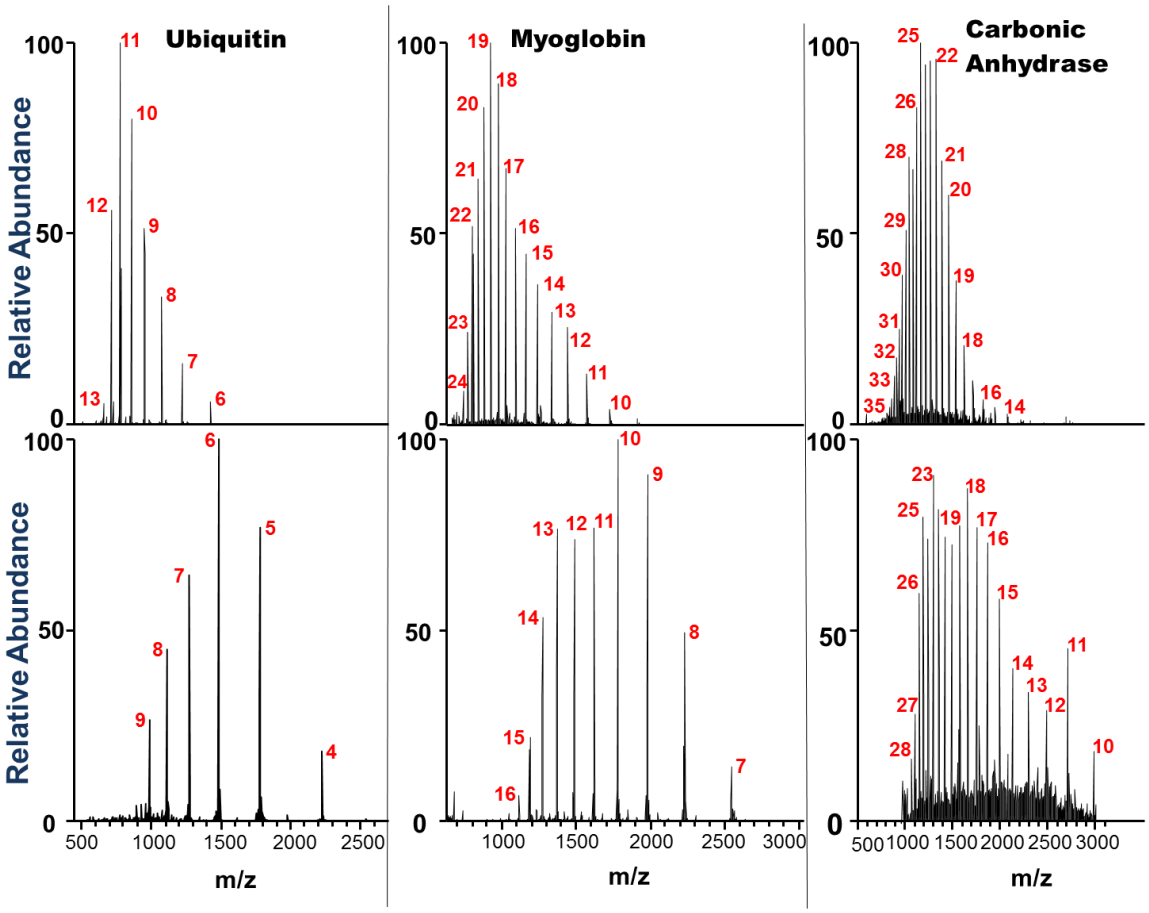


Figure 5.3 a,b,c) ESI-MS of unmodified proteins. d,e,f) ESI MS of the analogous carbamylated proteins.

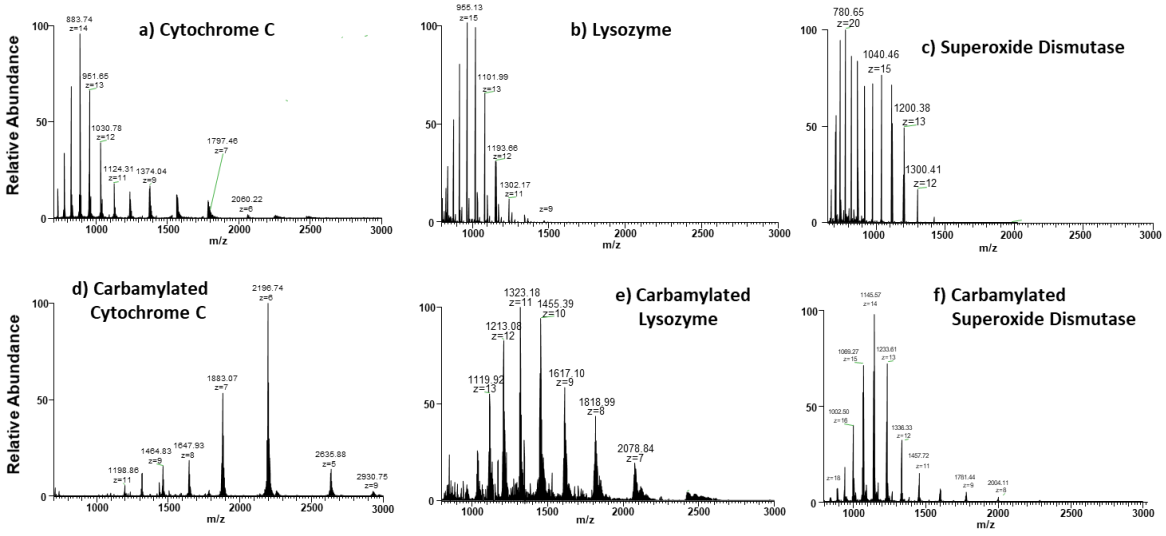


Figure 5.4 a,b,c) ESI-MS of unmodified proteins. d,e,f) ESI MS of the analogous carbamylated proteins

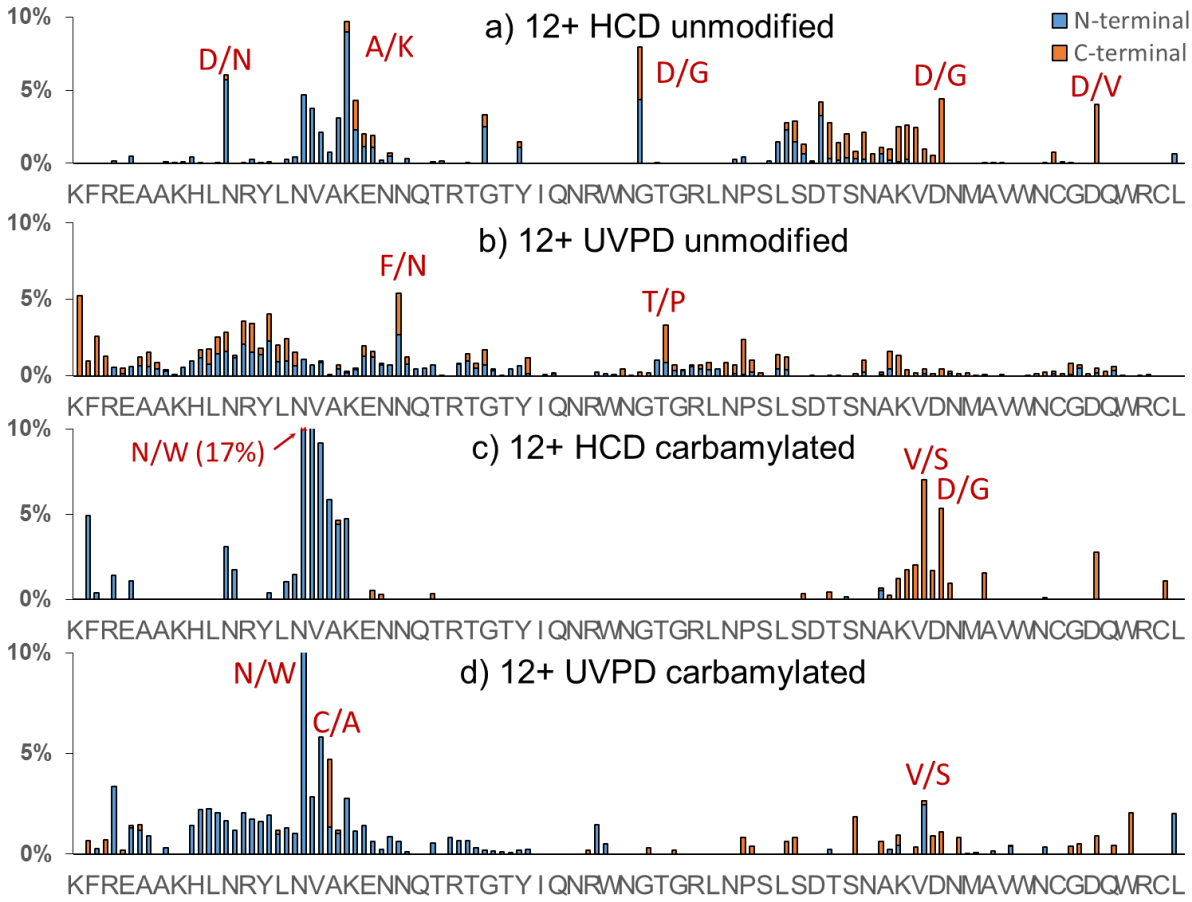


Figure 5.5. Backbone cleavage histograms of lysozyme (12+): a) HCD of unmodified lysozyme, b) UVPD of unmodified lysozyme, c) HCD of carbamylated lysozyme, and d) UVPD of carbamylated lysozyme. Some of the most enhanced backbone cleavage sites are labelled. The sequence of lysozyme is shown along the x-axis, with every other residue omitted. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars.

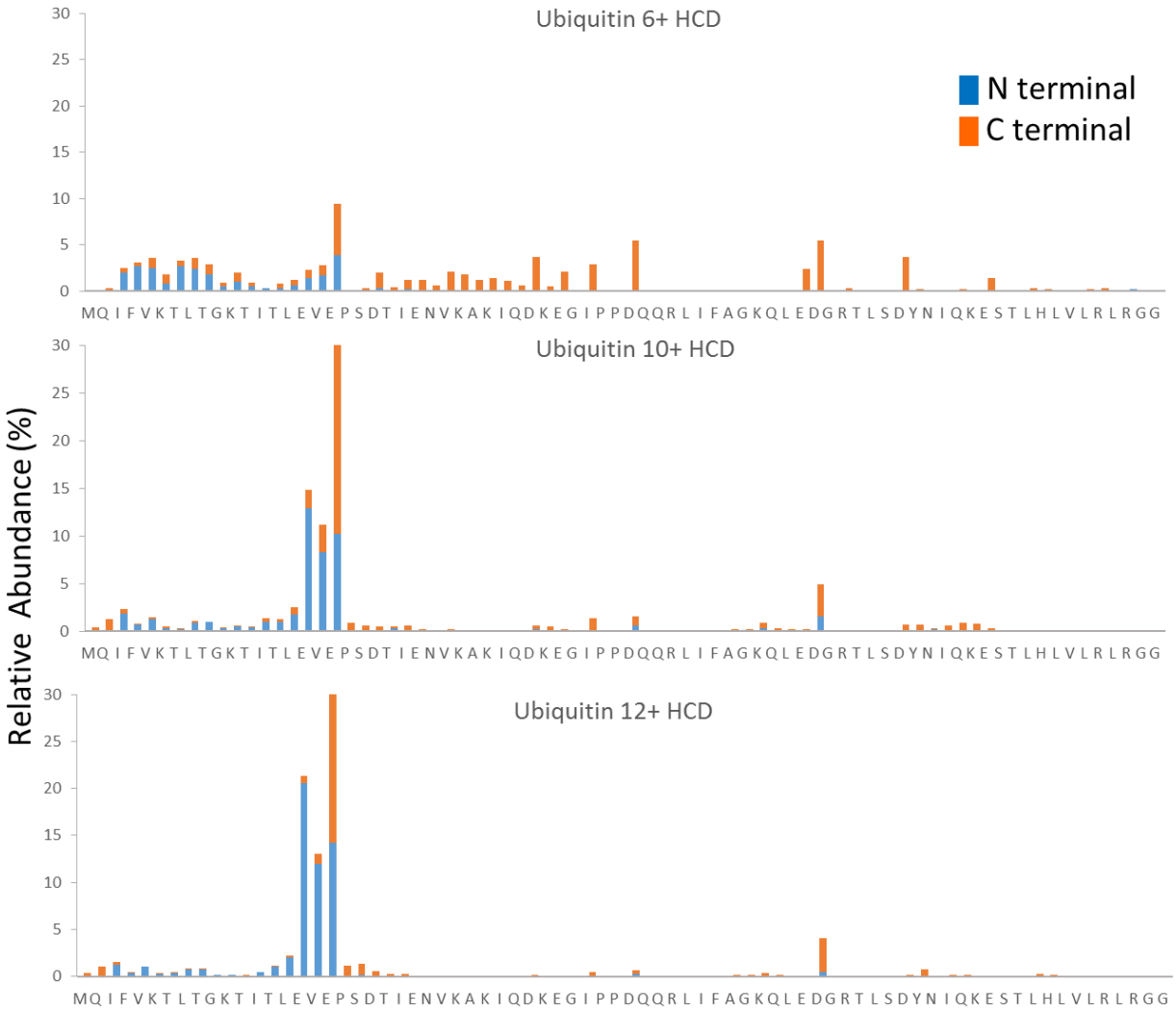


Figure 5.6 Backbone cleavage histograms of ubiquitin by HCD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 6+: 30NCE, 10+: 20 NCE, 12+: 16 NCE

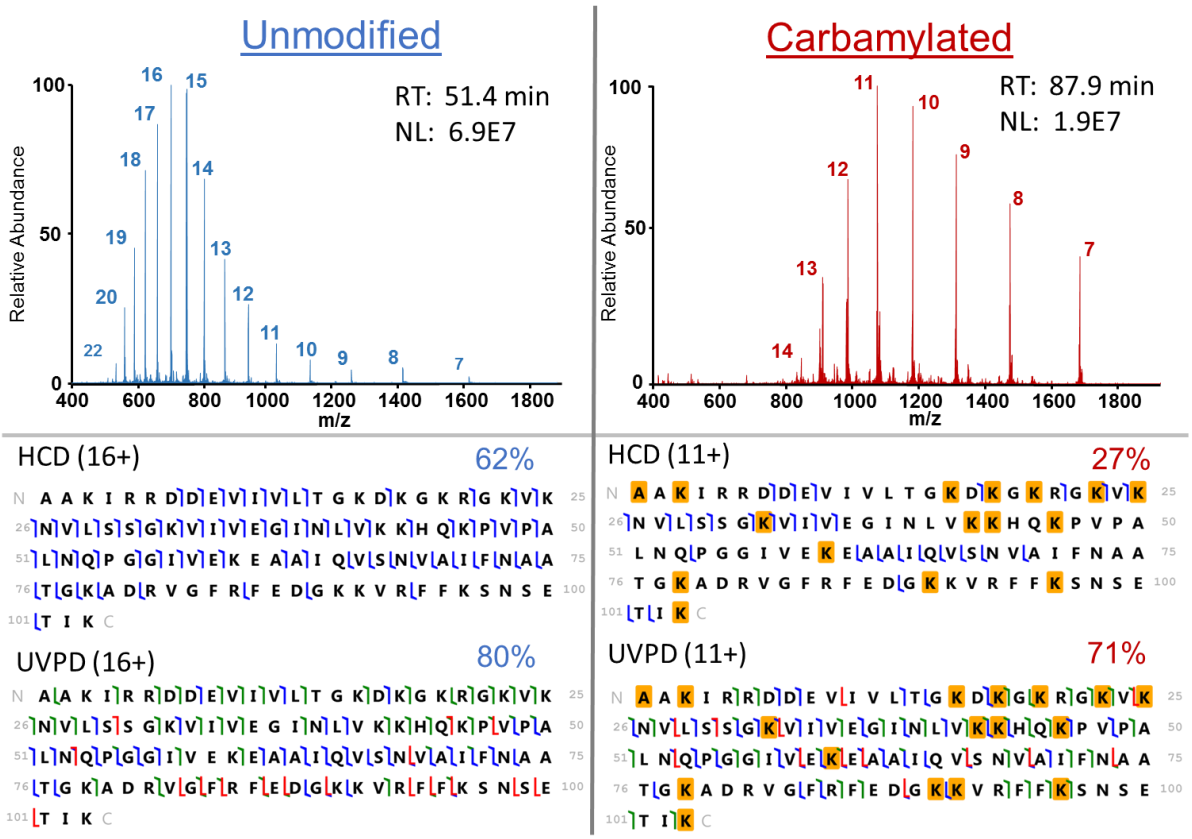


Figure 5.7 Top: ESI mass spectra of unmodified and carbamylated ribosomal protein 50S L24 from *E.coli*. Bottom: Sequence maps annotated for HCD fragmentation (upper sequence maps) and UVPD (lower sequence maps). The types of fragment ions generated from backbone cleavages are color coded as follows: a/x green; b/y blue; c/z red. Gold-highlighted boxes denote sites of carbamylation.

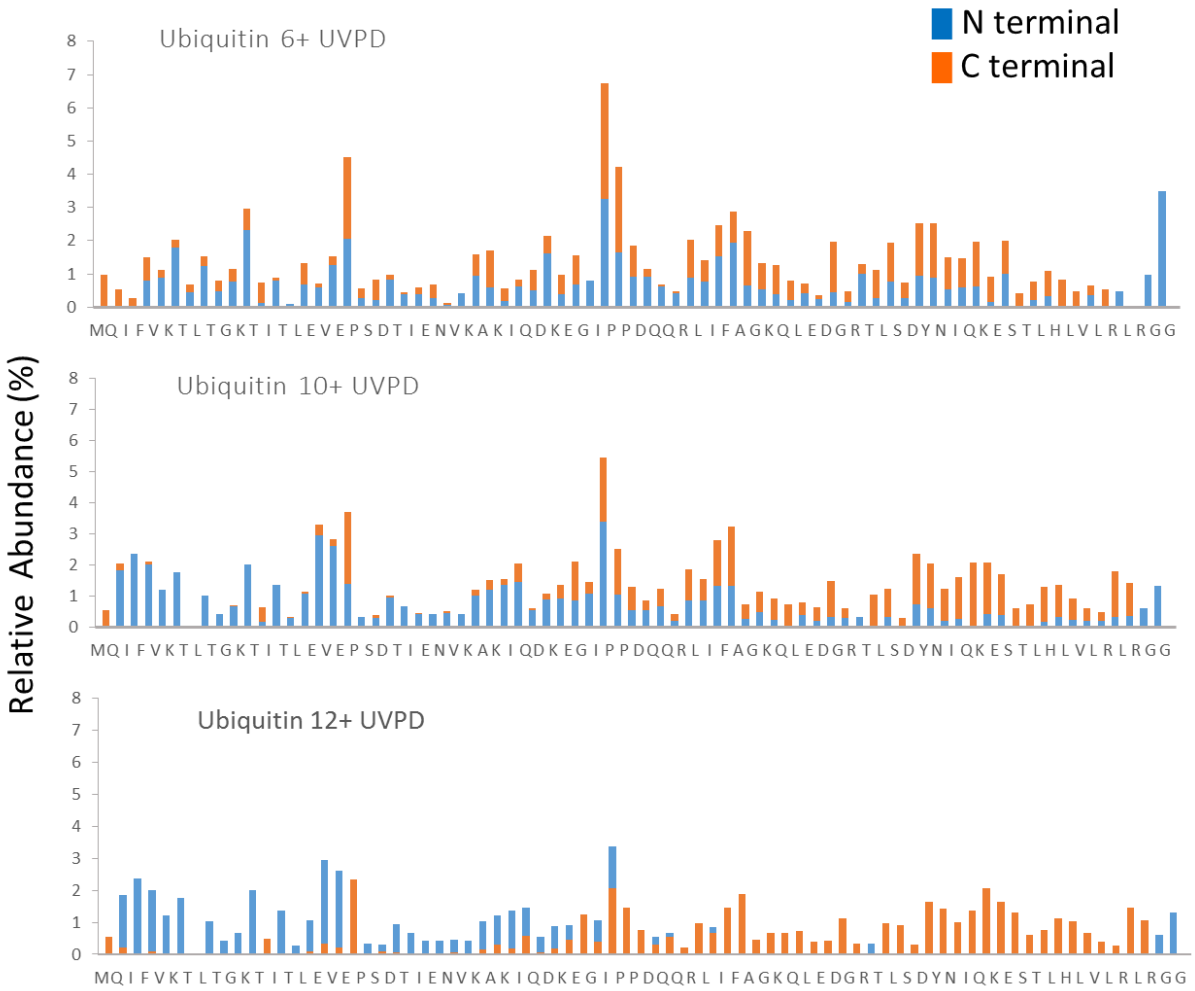


Figure 5.8 Backbone cleavage histograms of ubiquitin by UVPD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. UVPD was set to 1 pulse 1 mJ.

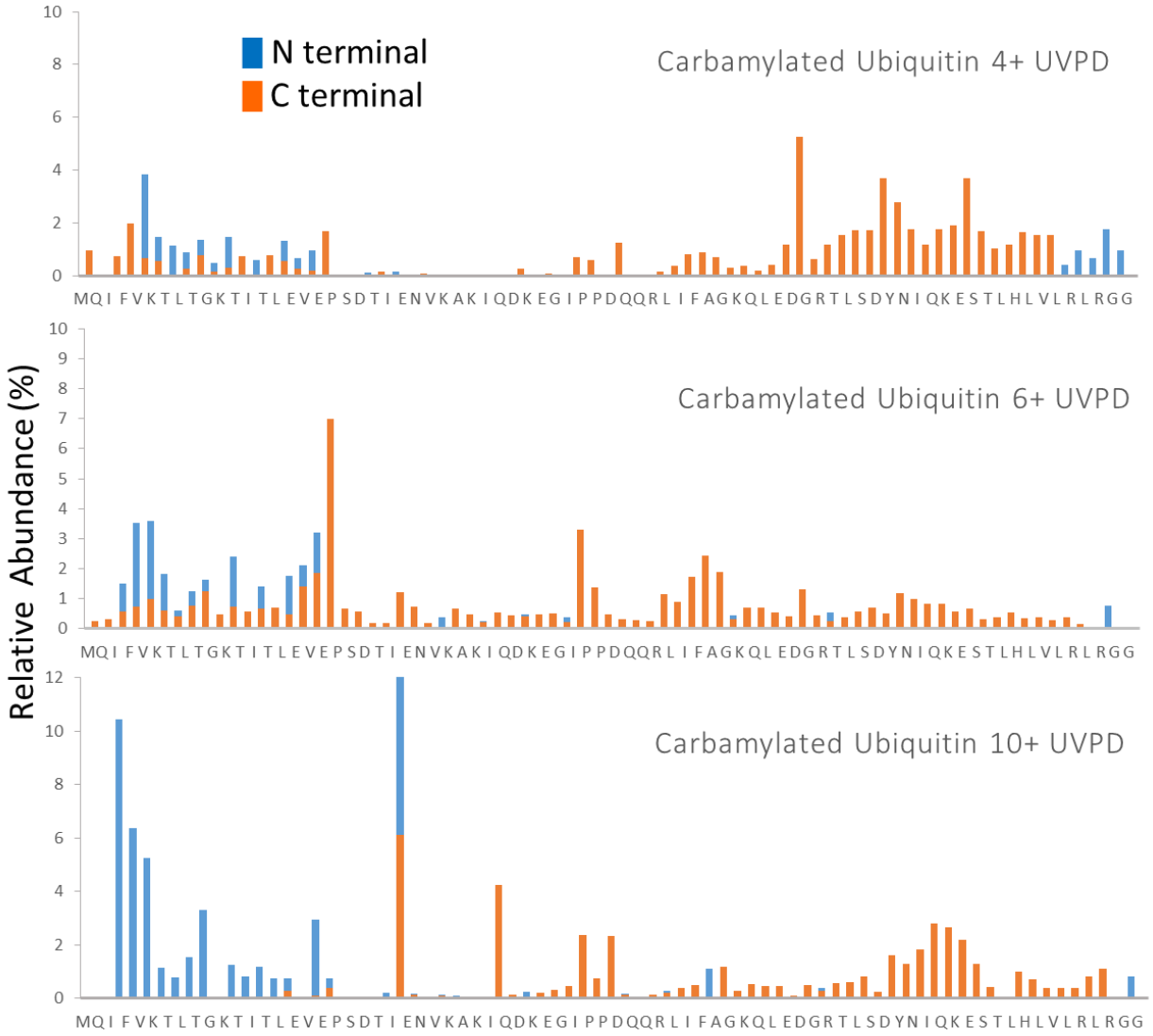


Figure 5.9 Backbone cleavage histograms of carbamylated ubiquitin by UVPD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. UVPD was set to 1 pulse 1 mJ.

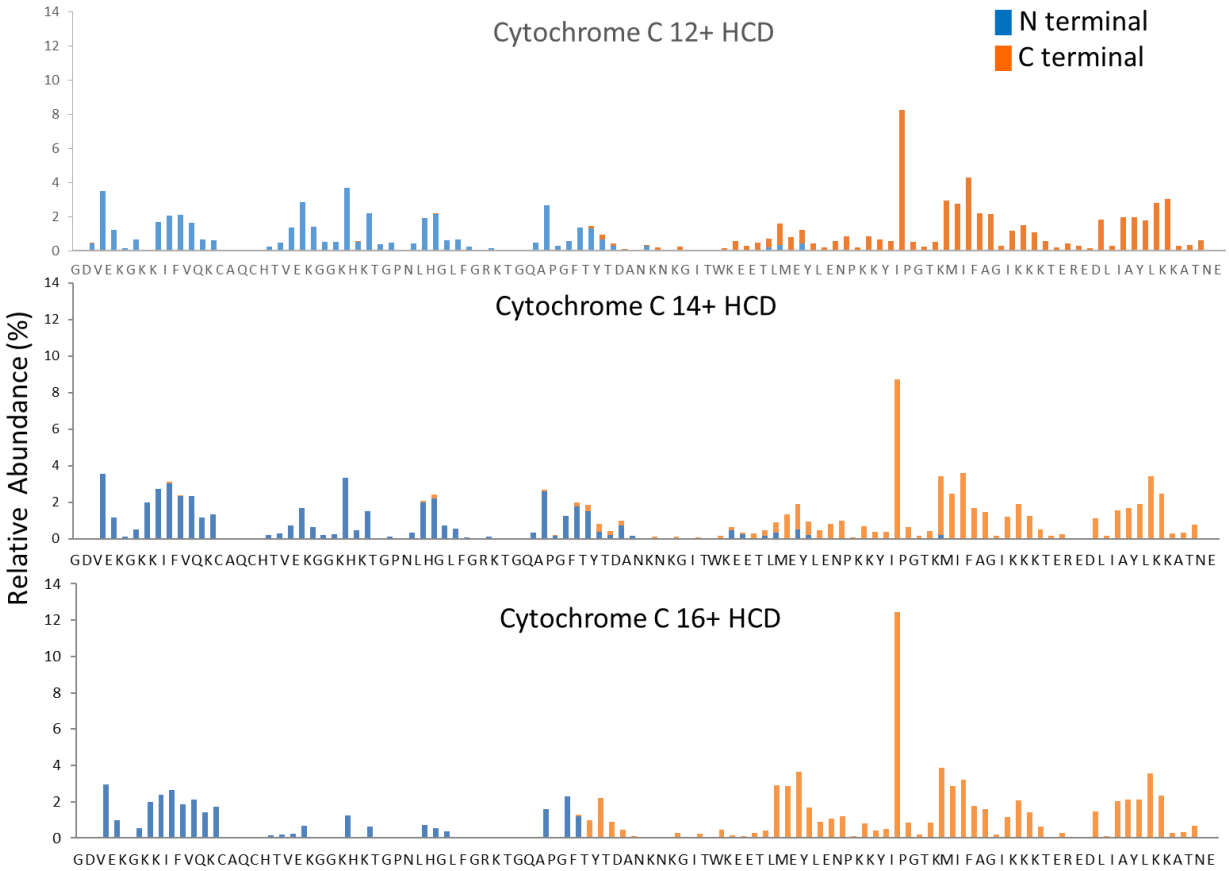


Figure 5.10 Backbone cleavage histograms of cytochrome C by HCD. Some of the most enhanced backbone cleavage sites are labelled. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 12+: 29NCE, 14+: 25 NCE, 16+: 25 NCE

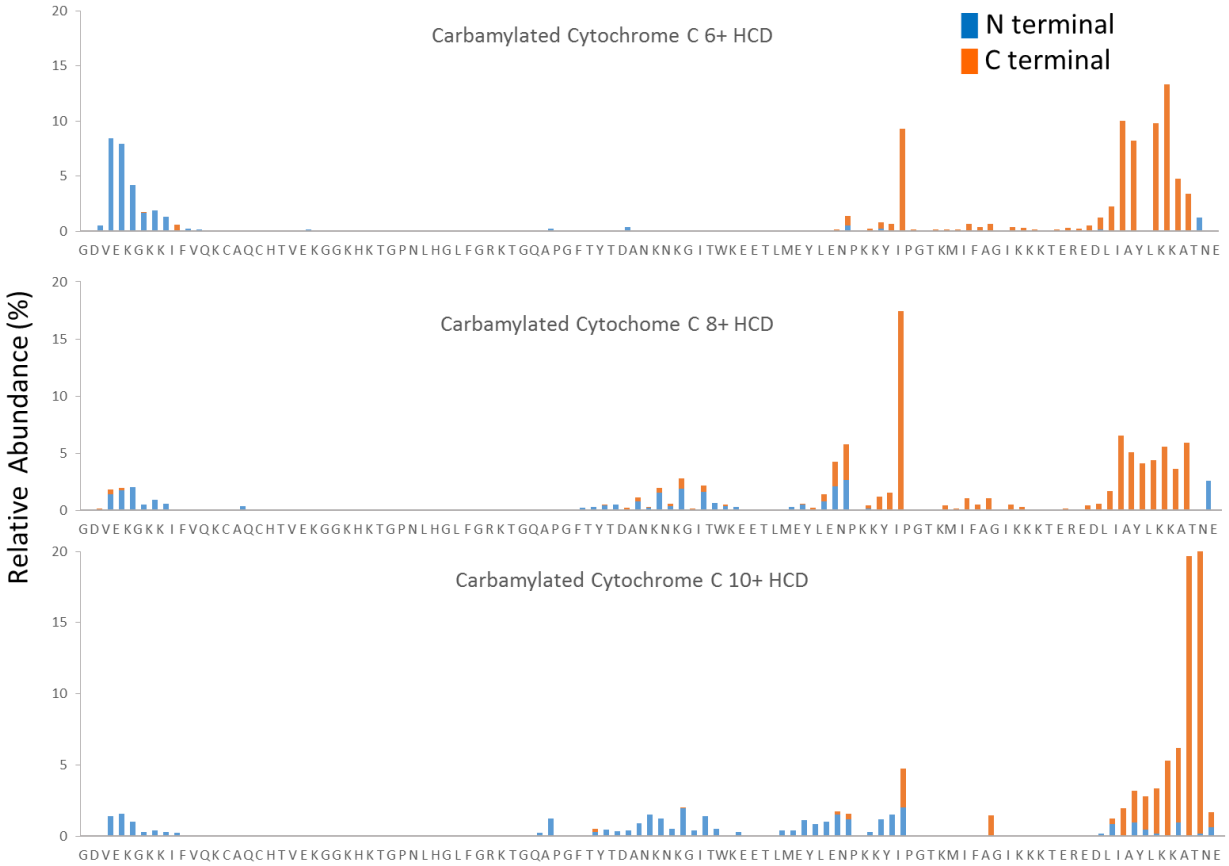


Figure 5.11 Backbone cleavage histograms of carbamylated cytochrome C by HCD. Some of the most enhanced backbone cleavage sites are labelled. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 6+: 23NCE, 8+: 18 NCE, 10+: 18 NCE

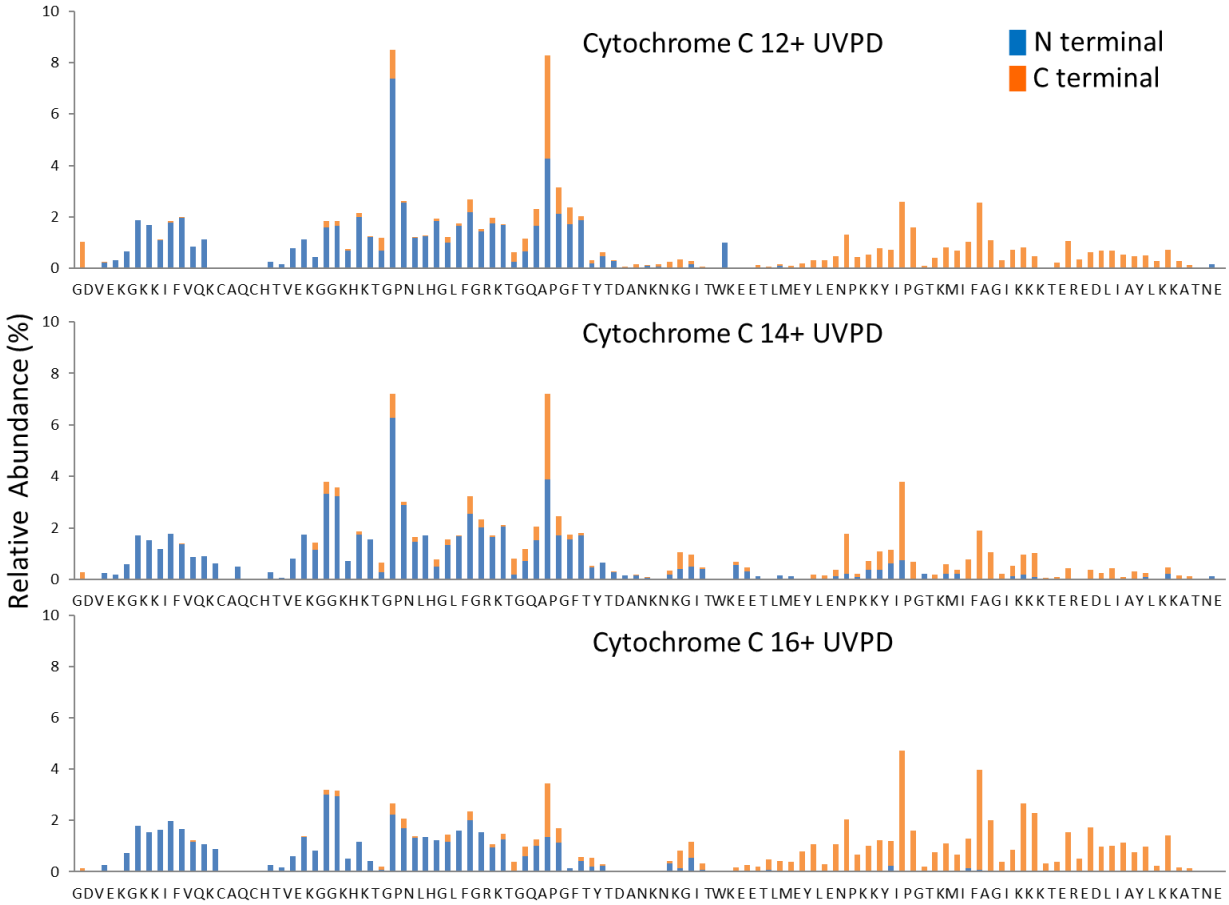


Figure 5.12 Backbone cleavage histograms of cytochrome C by UVPD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. UVPD was set to 1 pulse 1 mJ.

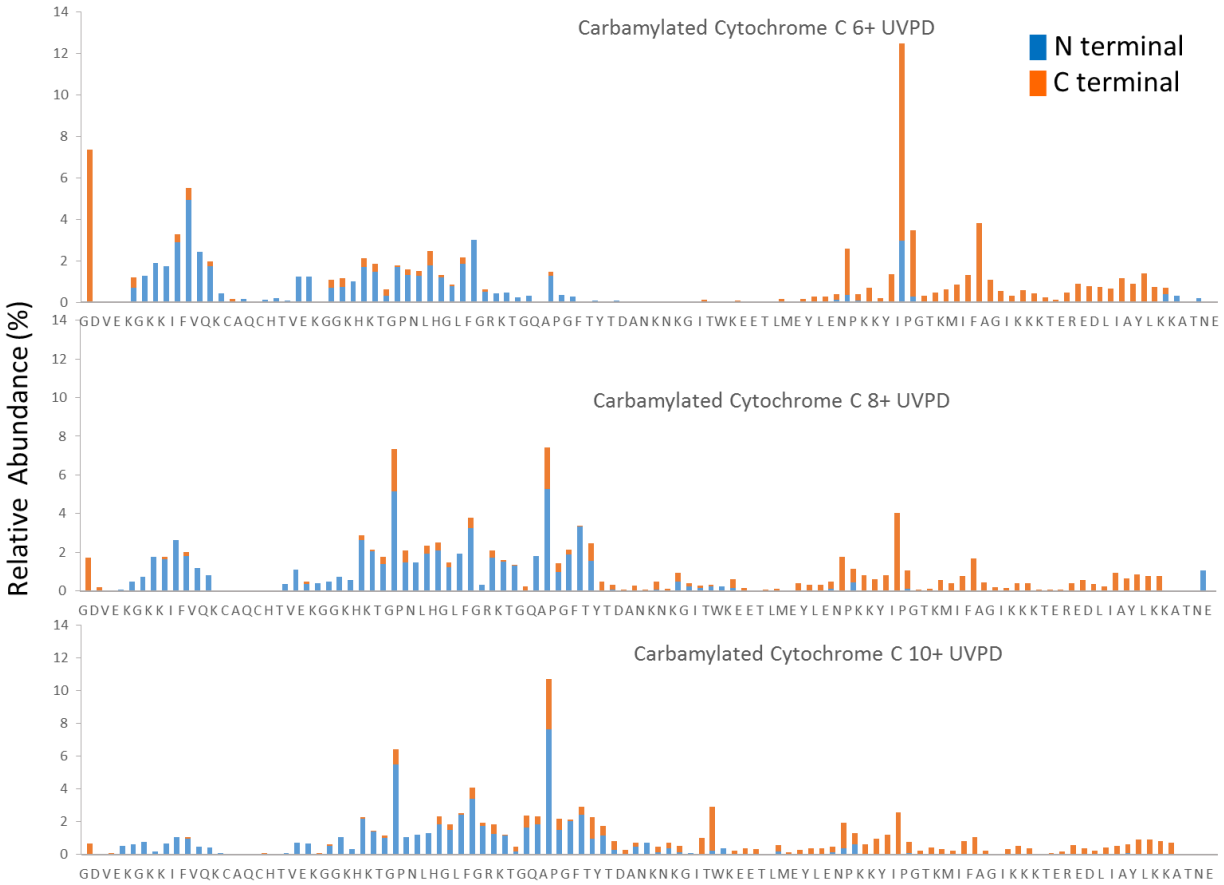


Figure 5.13 Backbone cleavage histograms of carbamylated cytochrome C by UVPD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. UVPD was set to 1 pulse 1 mJ.

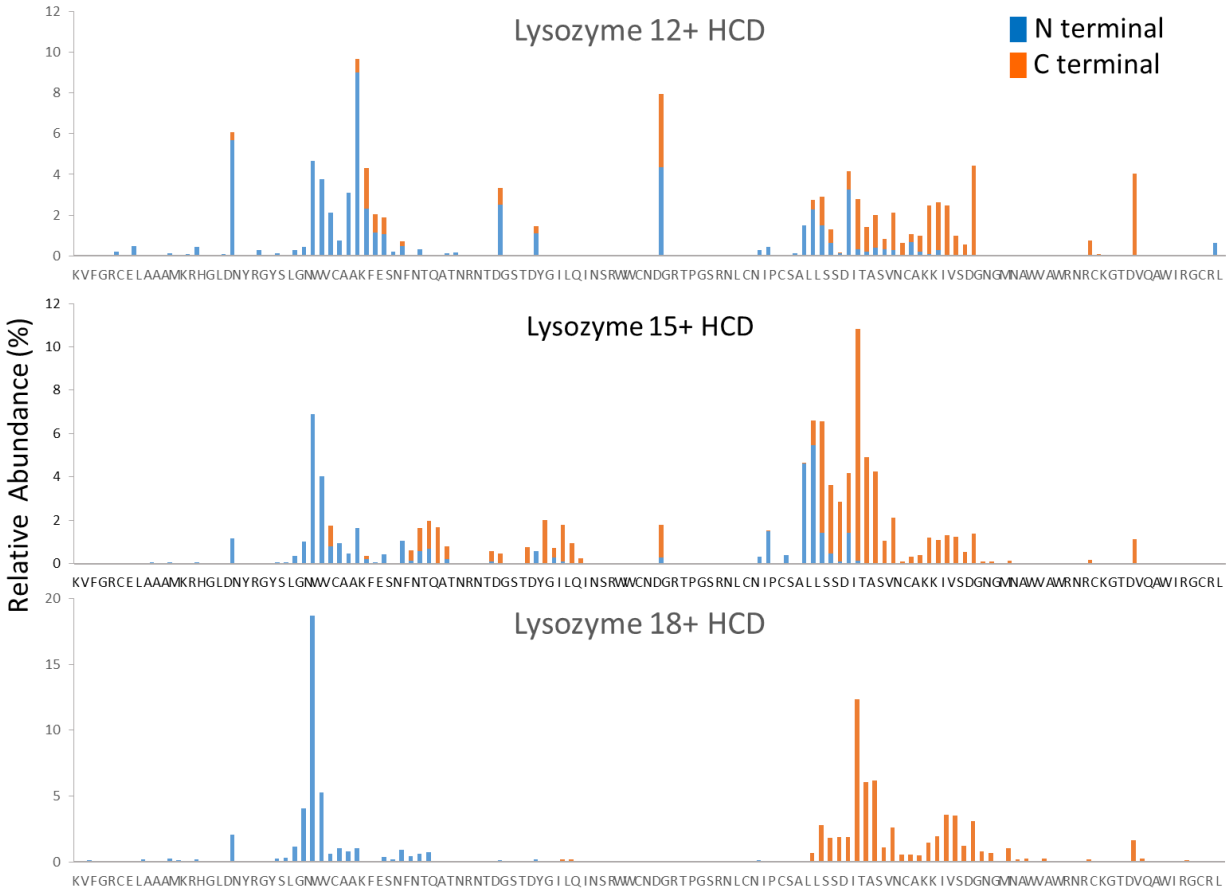


Figure 5.14 Backbone cleavage histograms of lysozyme by HCD. Some of the most enhanced backbone cleavage sites are labelled. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 12+: 25 NCE, 15+: 20 NCE, 18+: 20 NCE

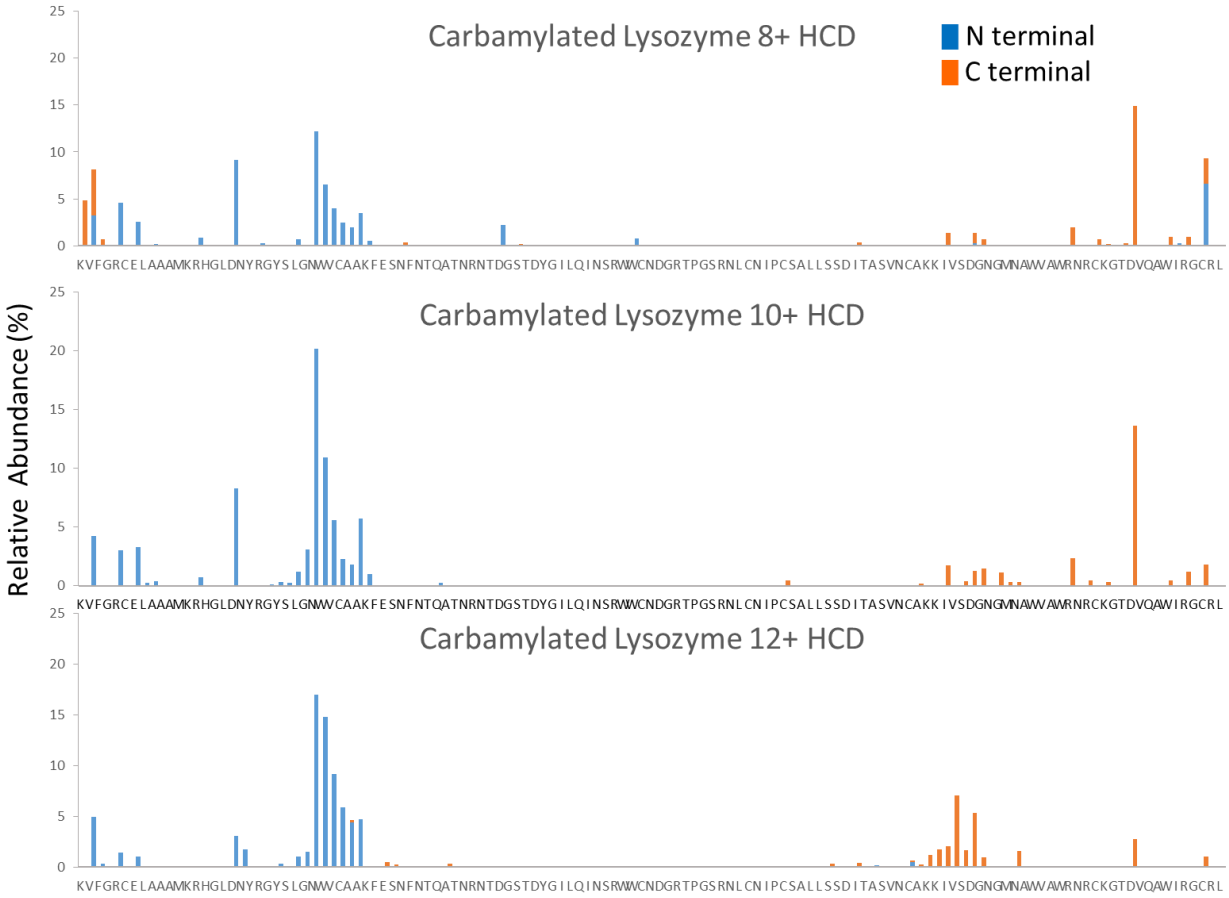


Figure 5.15 Backbone cleavage histograms of carbamylated lysozyme by HCD. Some of the most enhanced backbone cleavage sites are labelled. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 8+: 30NCE, 10+: 25 NCE, 12+: 25 NCE

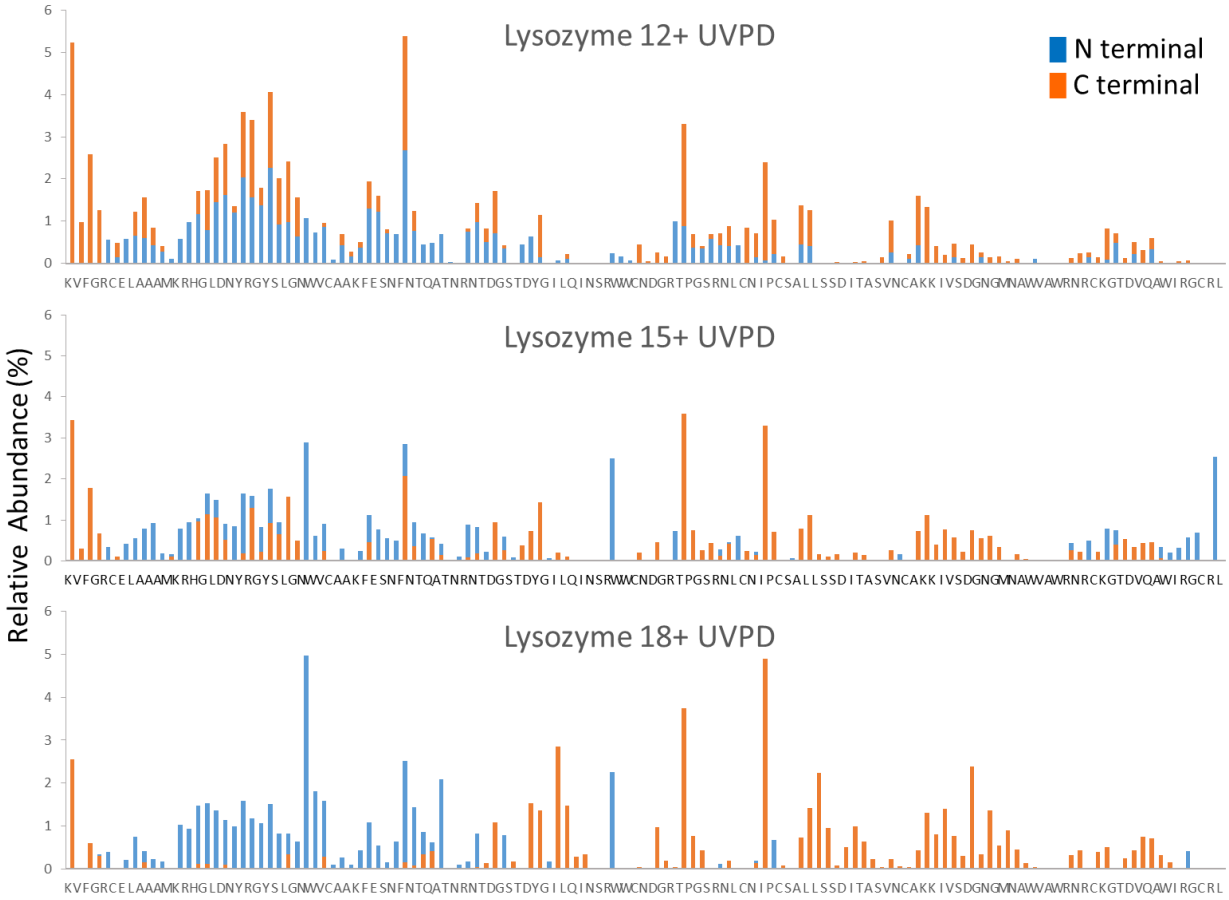


Figure 5.16 Backbone cleavage histograms of lysozyme by UVPD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. UVPD was set to 1 pulse 1 mJ.

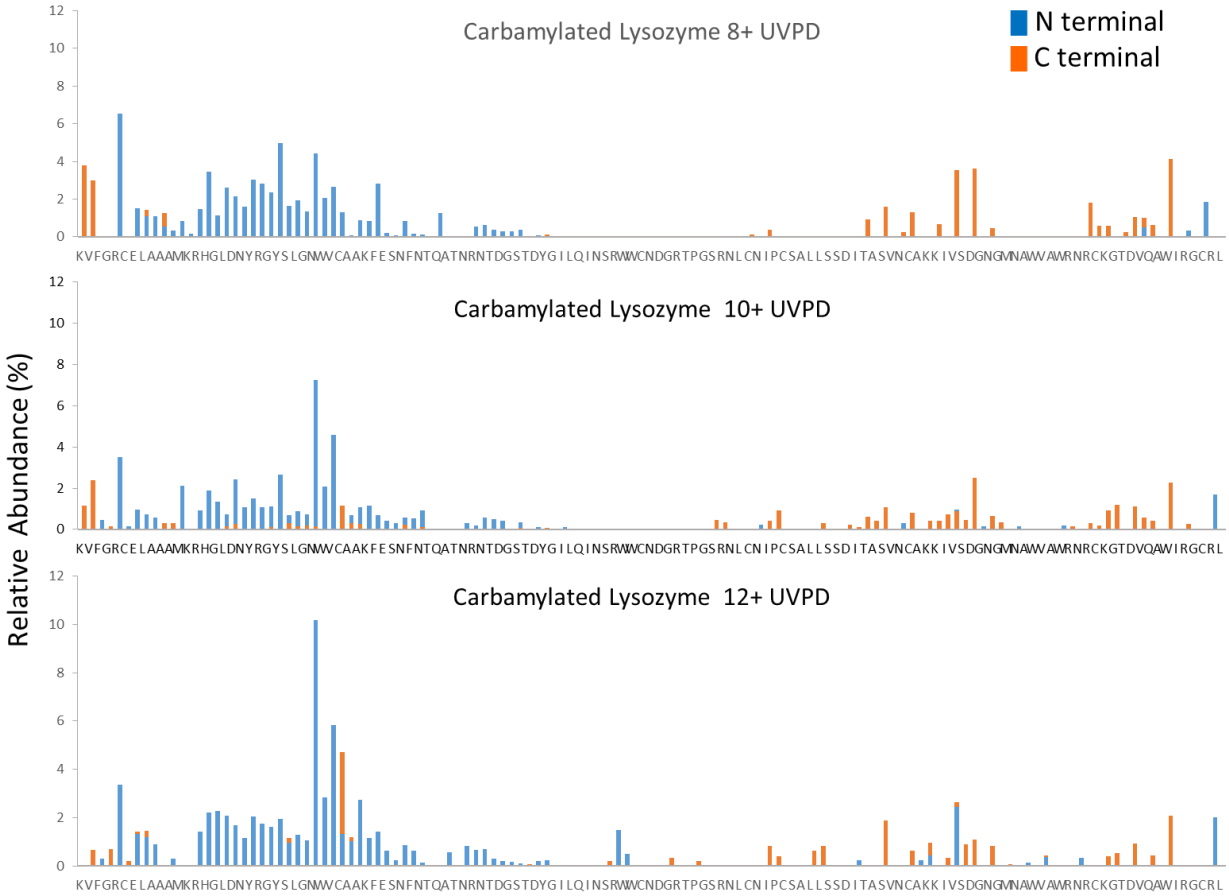


Figure 5.17 Backbone cleavage histograms of carbamylated lysozyme by UVPD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. UVPD was set to 1 pulse 1 mJ.

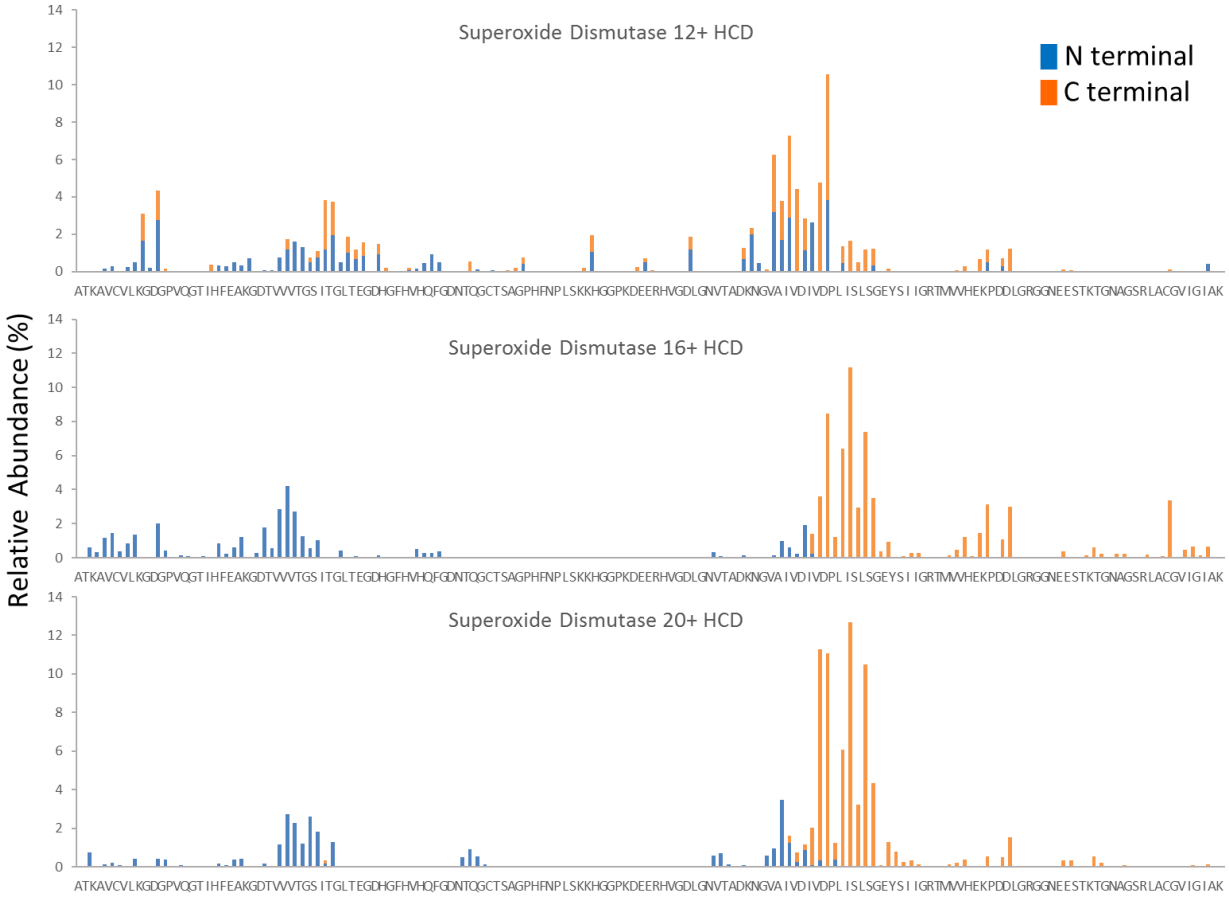


Figure 5.18 Backbone cleavage histograms of superoxide dismutase by HCD. Some of the most enhanced backbone cleavage sites are labelled. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 12+: 20NCE, 16+: 17 NCE, 20+: 10 NCE

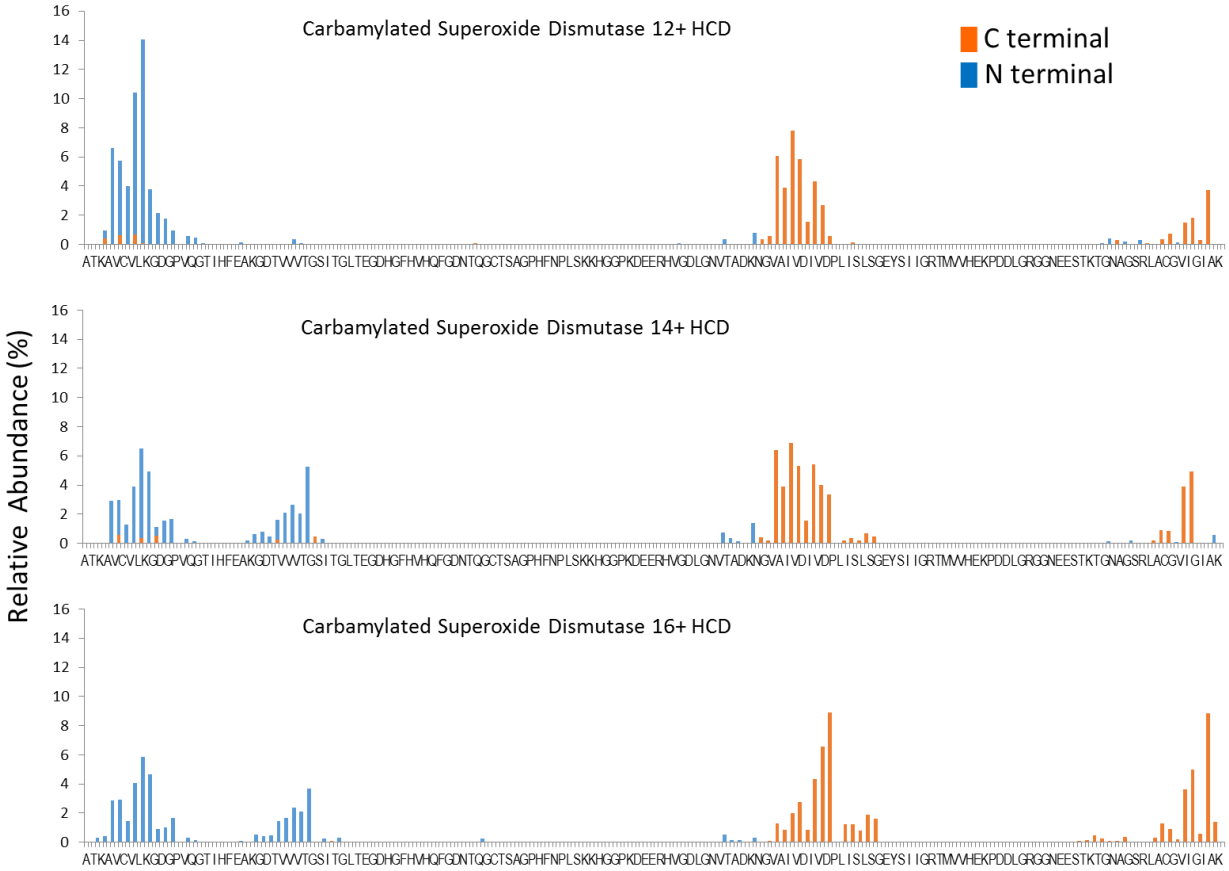


Figure 5.19 Backbone cleavage histograms of carbamylated superoxide dismutase by HCD. Some of the most enhanced backbone cleavage sites are labelled. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 12+: 20 NCE, 14+: 15 NCE, 16+: 12 NCE

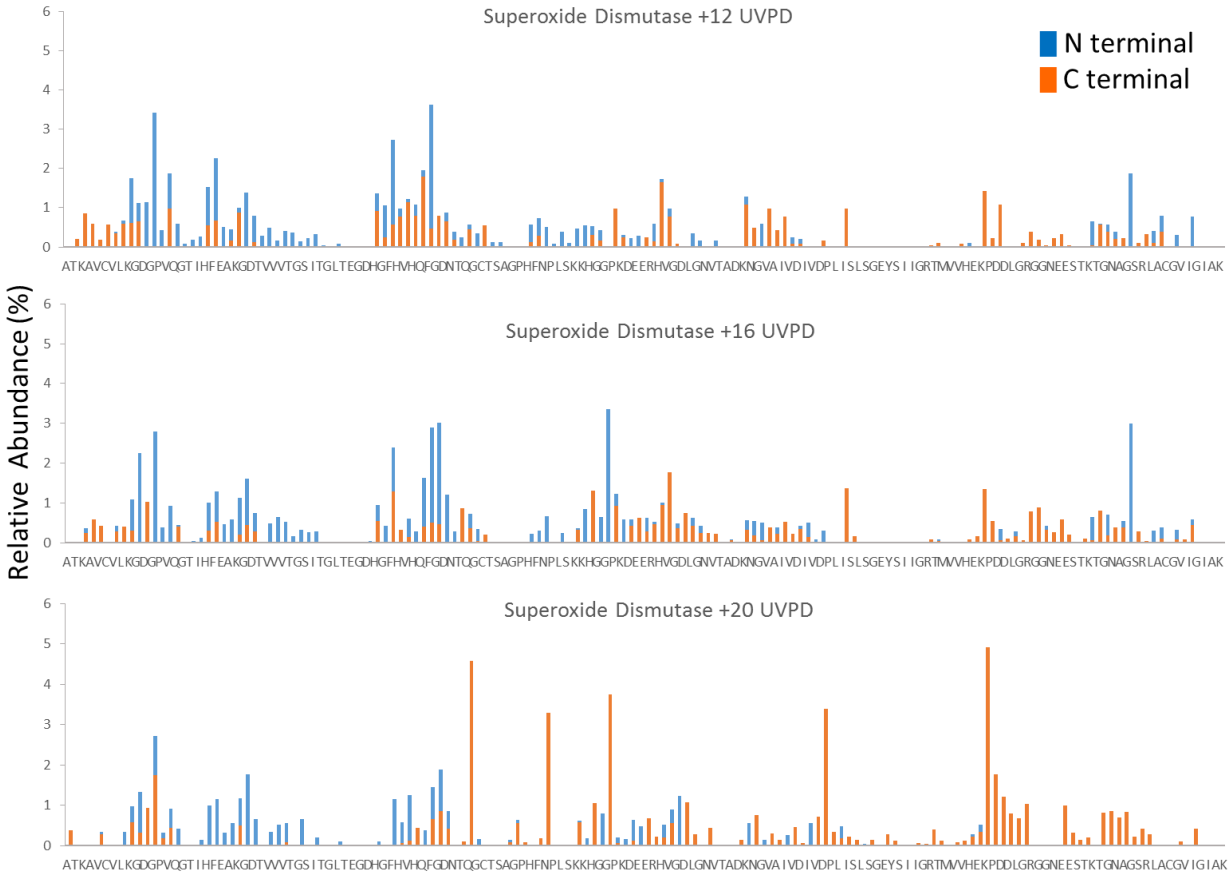


Figure 5.20 Backbone cleavage histograms of superoxide dismutase by UVPD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars.

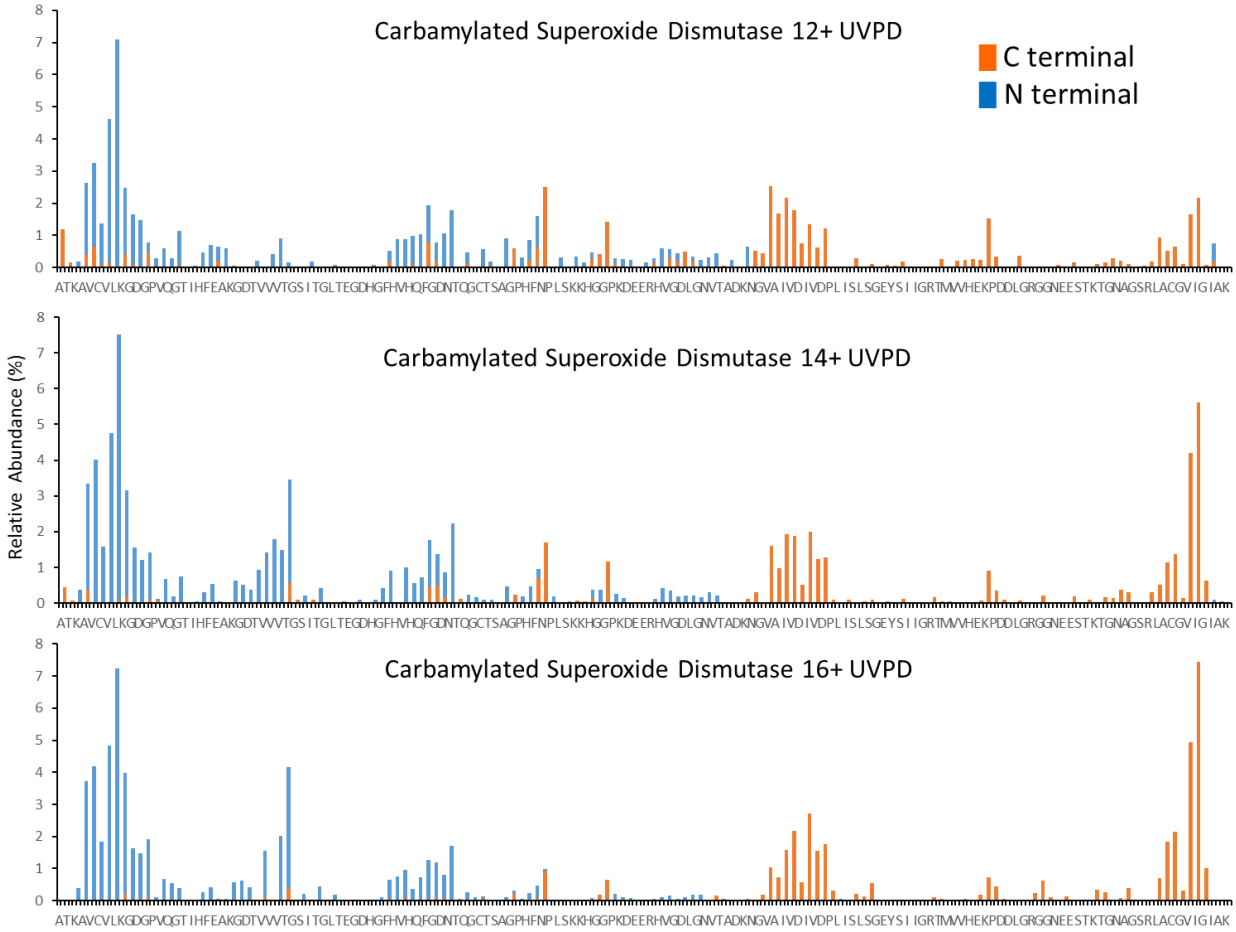


Figure 5.21 Backbone cleavage histograms of carbamylated superoxide dismutase by UVPD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars.

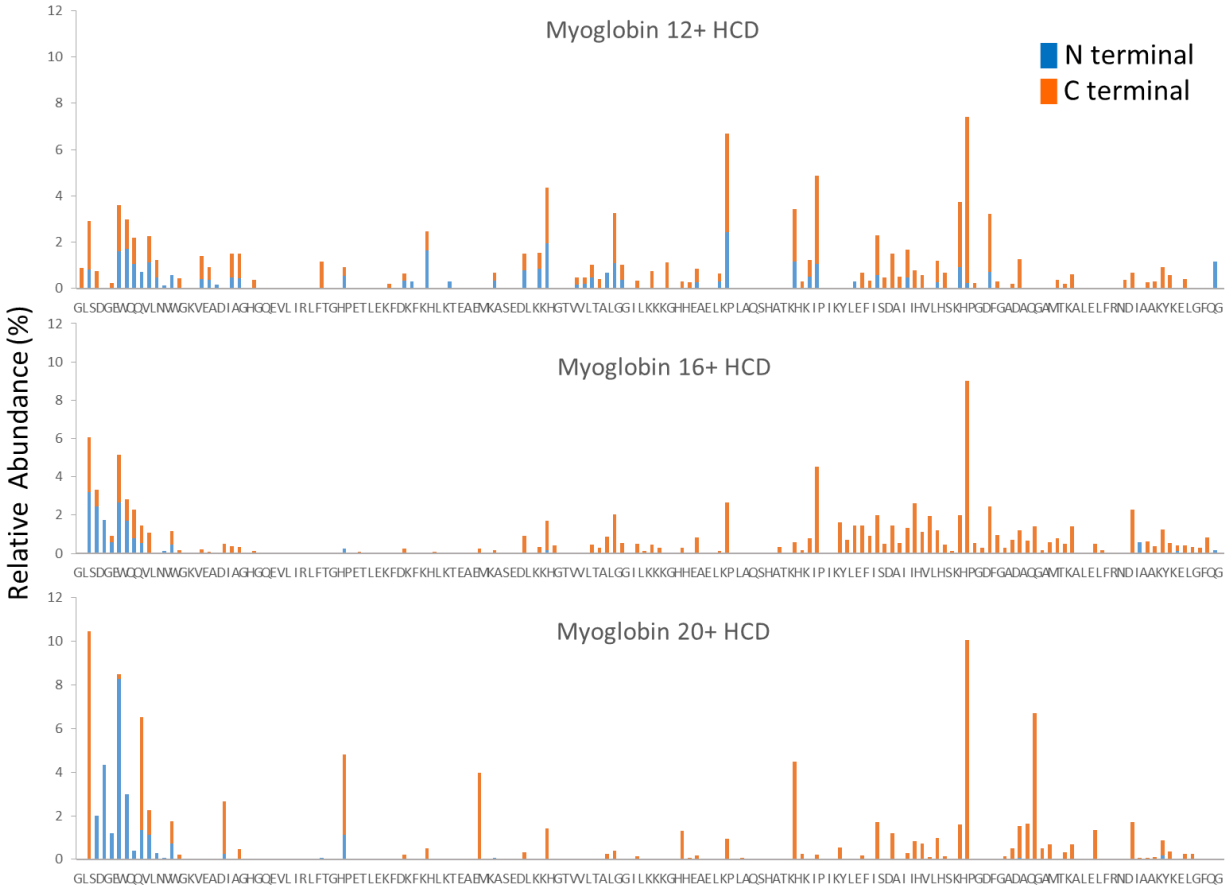


Figure 5.22 Backbone cleavage histograms of myoglobin by HCD. Some of the most enhanced backbone cleavage sites are labelled. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 12+: 20NCE, 16+: 20 NCE, 20+: 17 NCE

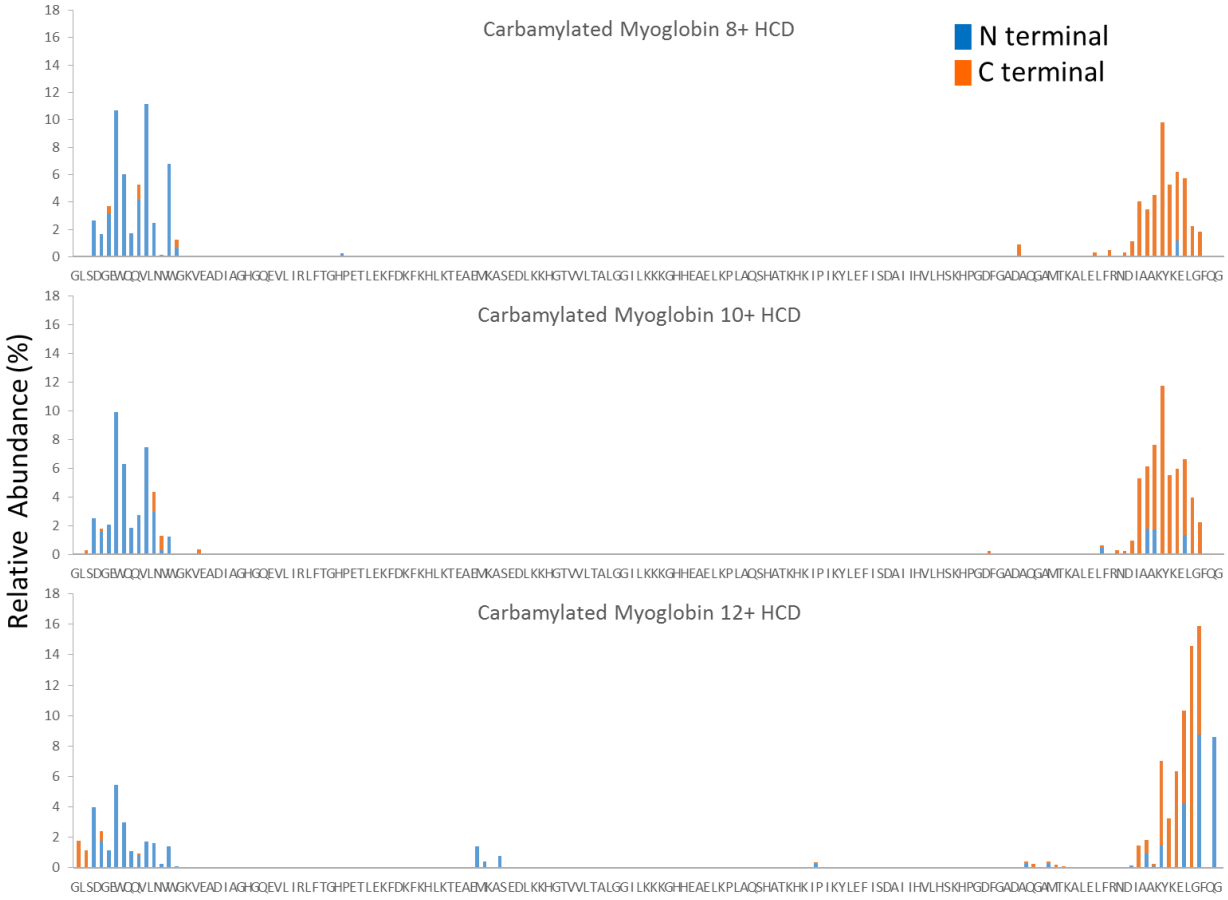


Figure 5.23 Backbone cleavage histograms of carbamylated myoglobin by HCD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 8+: 25NCE, 10+: 20 NCE, 12+: 20 NCE

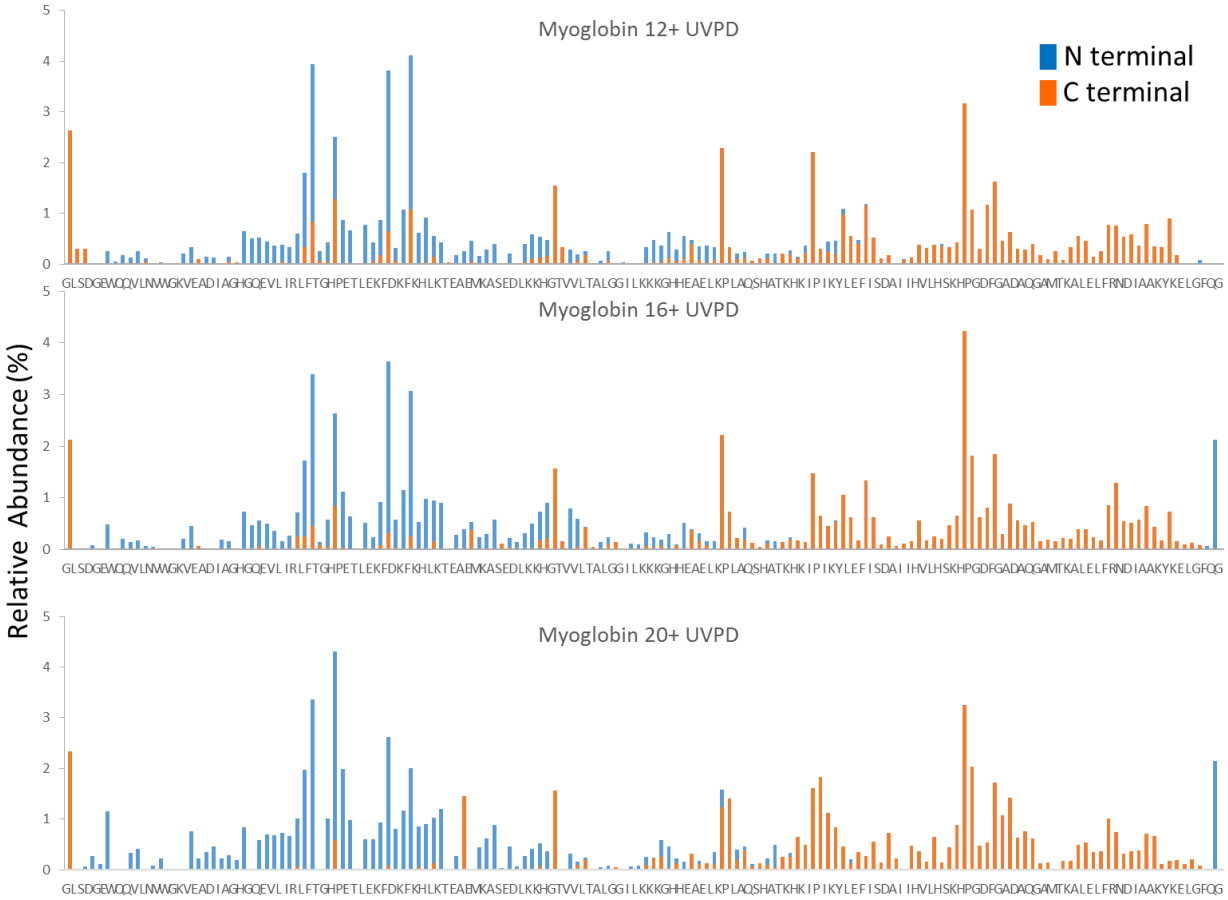


Figure 5.24 Backbone cleavage histograms of myoglobin by UVPD. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. UVPD was set to 1 pulse 1 mJ.

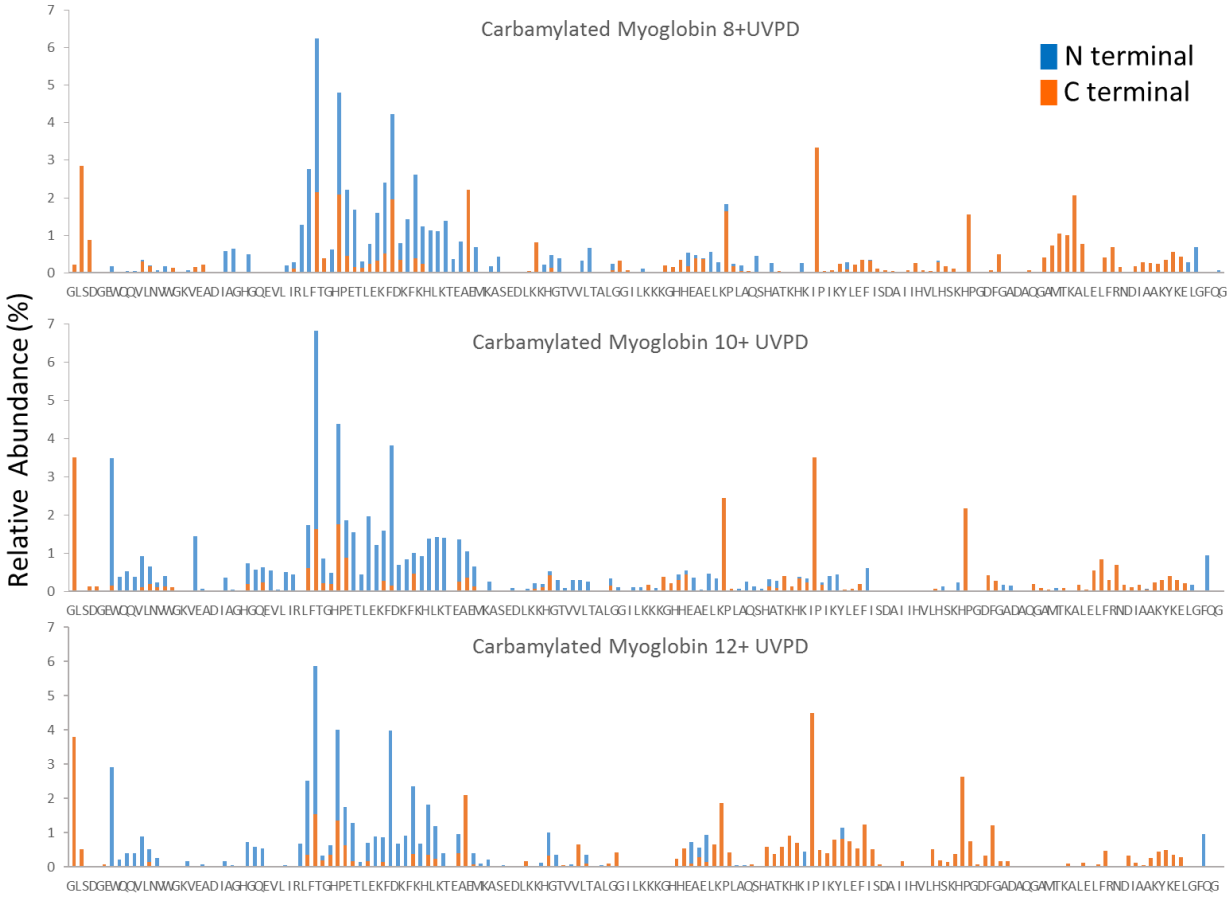


Figure 5.25 Backbone cleavage histograms of carbamylated myoglobin by UVPD. Some of the most enhanced backbone cleavage sites are labelled. The sequence of the protein is shown along the x-axis. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. UVPD was set to 1 pulse 1 mJ.

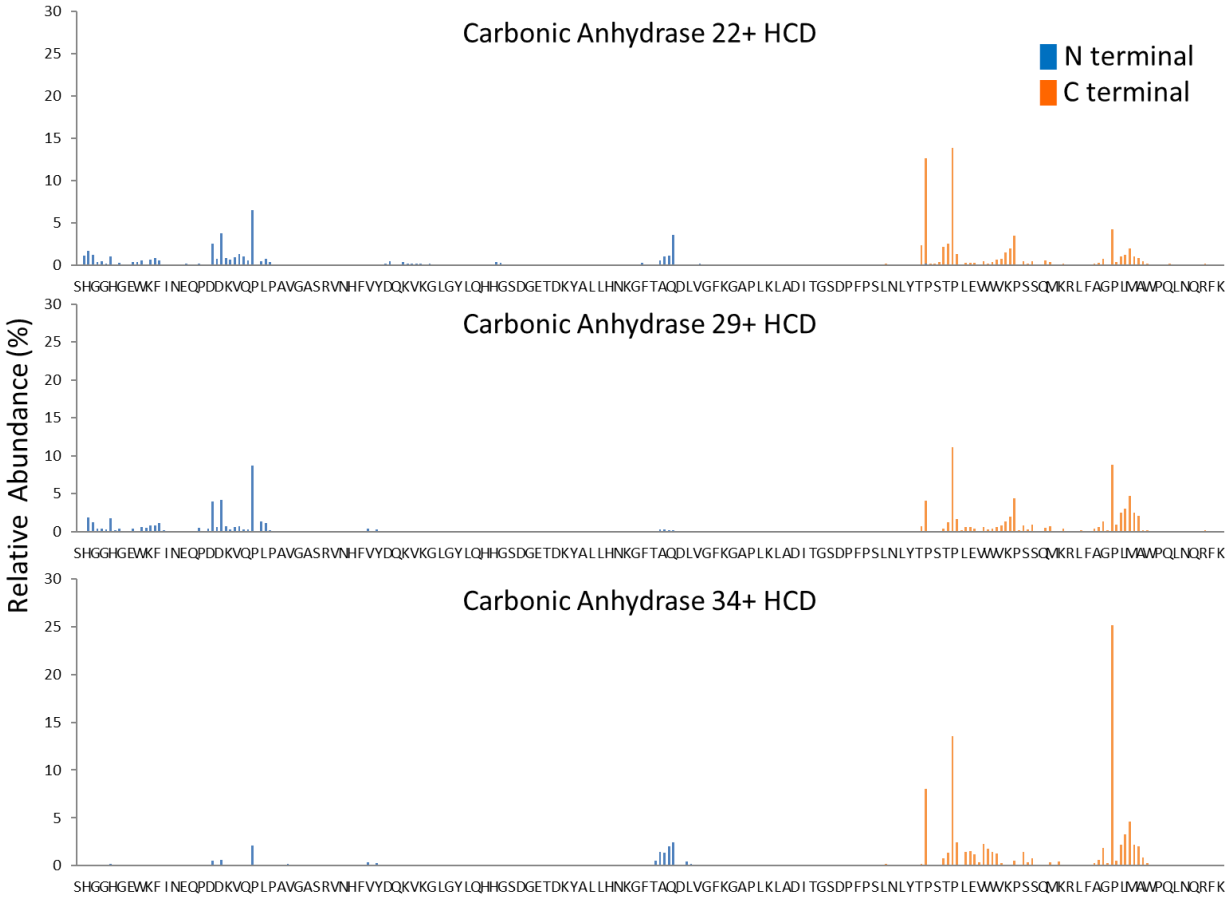


Figure 5.26 Backbone cleavage histograms of carbonic anhydrase by HCD. Some of the most enhanced backbone cleavage sites are labelled. The sequence of the protein is shown along the x-axis with every other residue omitted. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 22+: 15NCE, 29+: 15 NCE, 34+: 10 NCE

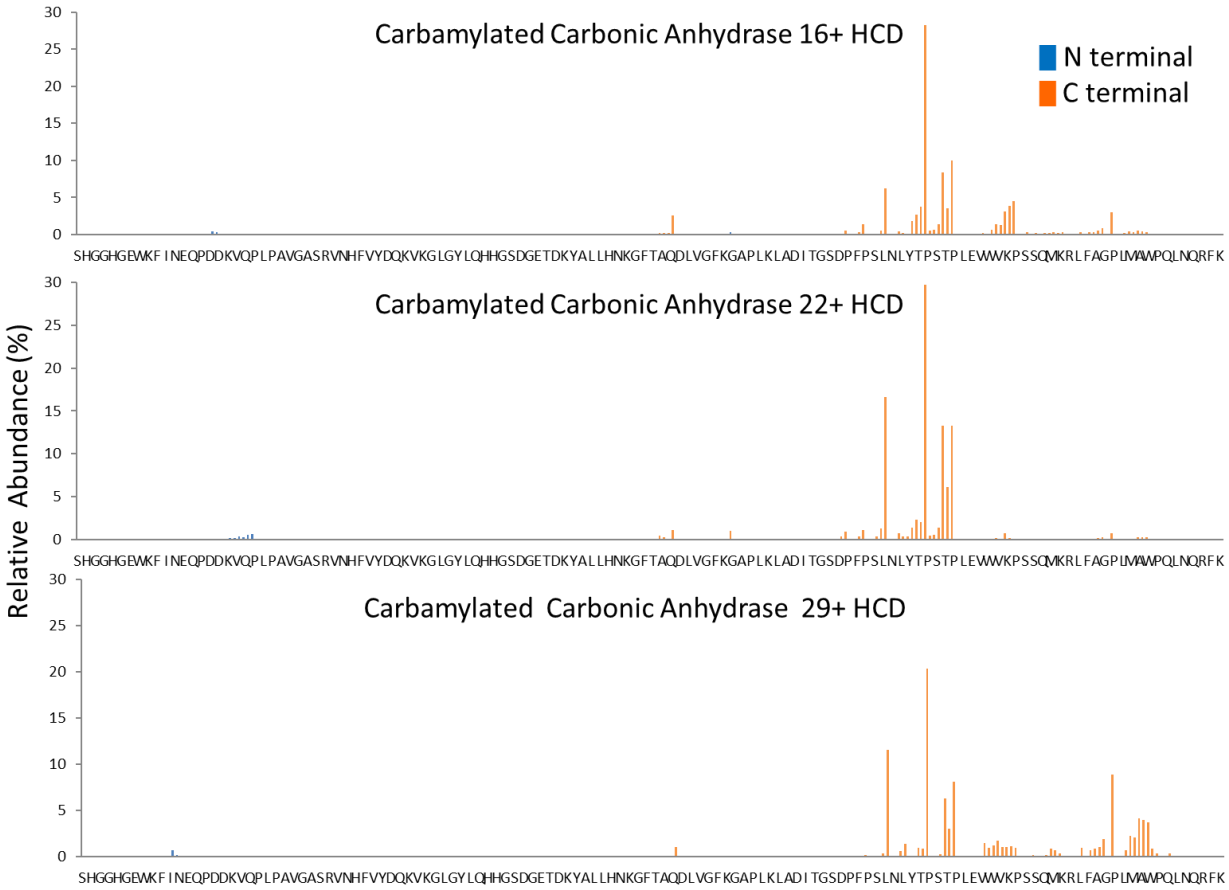


Figure 5.27 Backbone cleavage histograms of carbamylated carbonic anhydrase by HCD. Some of the most enhanced backbone cleavage sites are labelled. The sequence of the protein is shown along the x-axis with every other residue omitted. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. HCD NCE was optimized as follows 16+: 15NCE, 22+: 10 NCE, 29+: 10 NCE

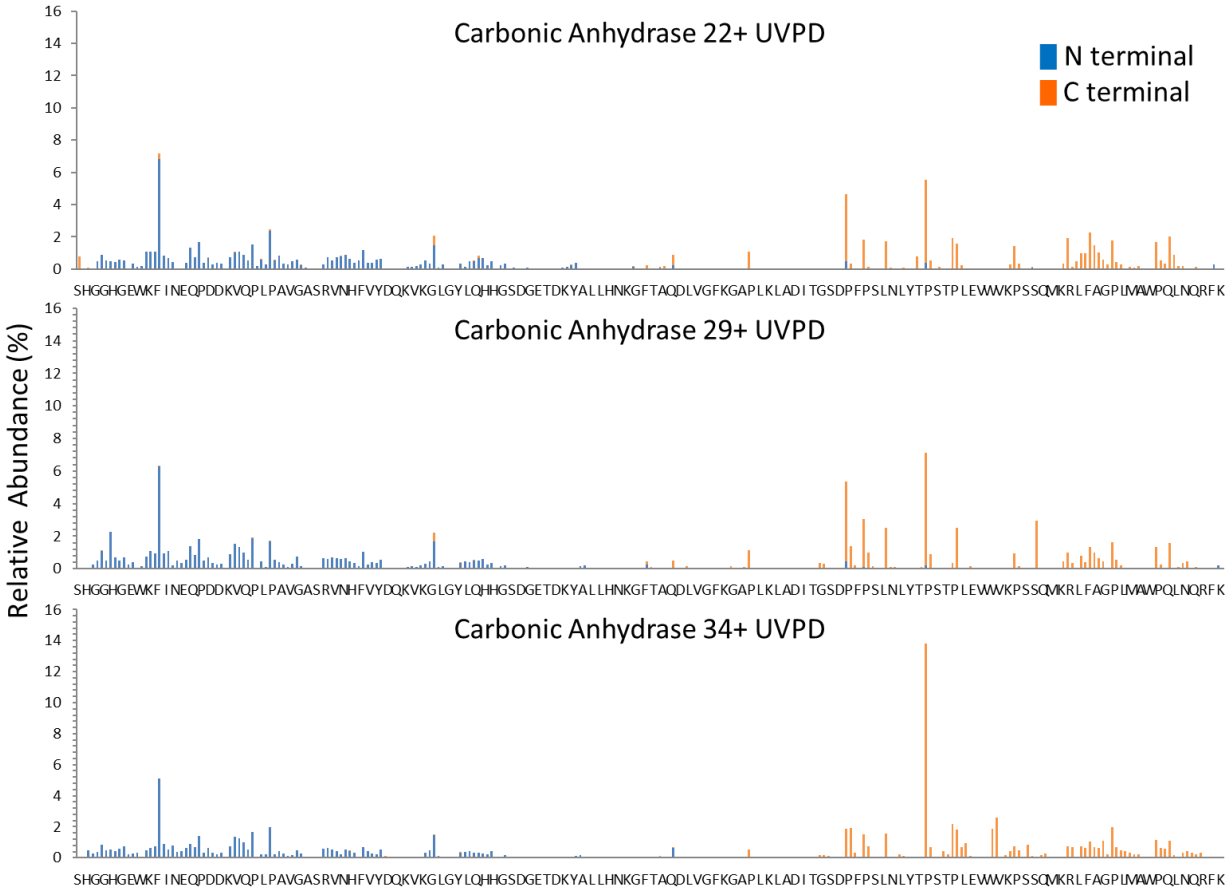


Figure 5.28 Backbone cleavage histograms of carbonic anhydrase by UVPD. The sequence of the protein is shown along the x-axis with every other residue omitted. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. UVPD was set to 1 pulse 1 mJ.

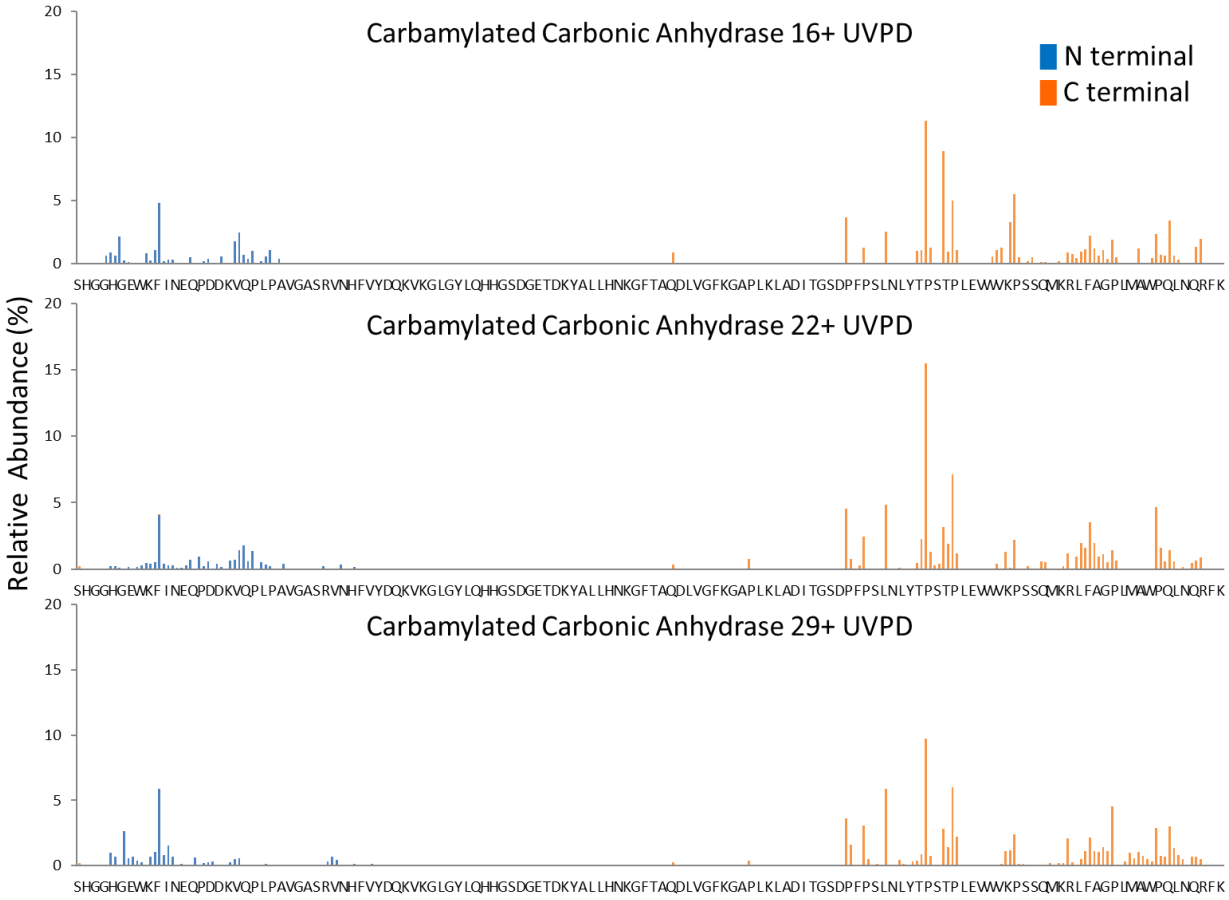


Figure 5.29 Backbone cleavage histograms of carbamylated carbonic anhydrase by UVPD. The sequence of the protein is shown along the x-axis with every other residue omitted. All N-terminal sequence ions are shown as blue bars; all C-terminal sequence ions are shown as orange bars. UVPD was set to 1 pulse 1 mJ.

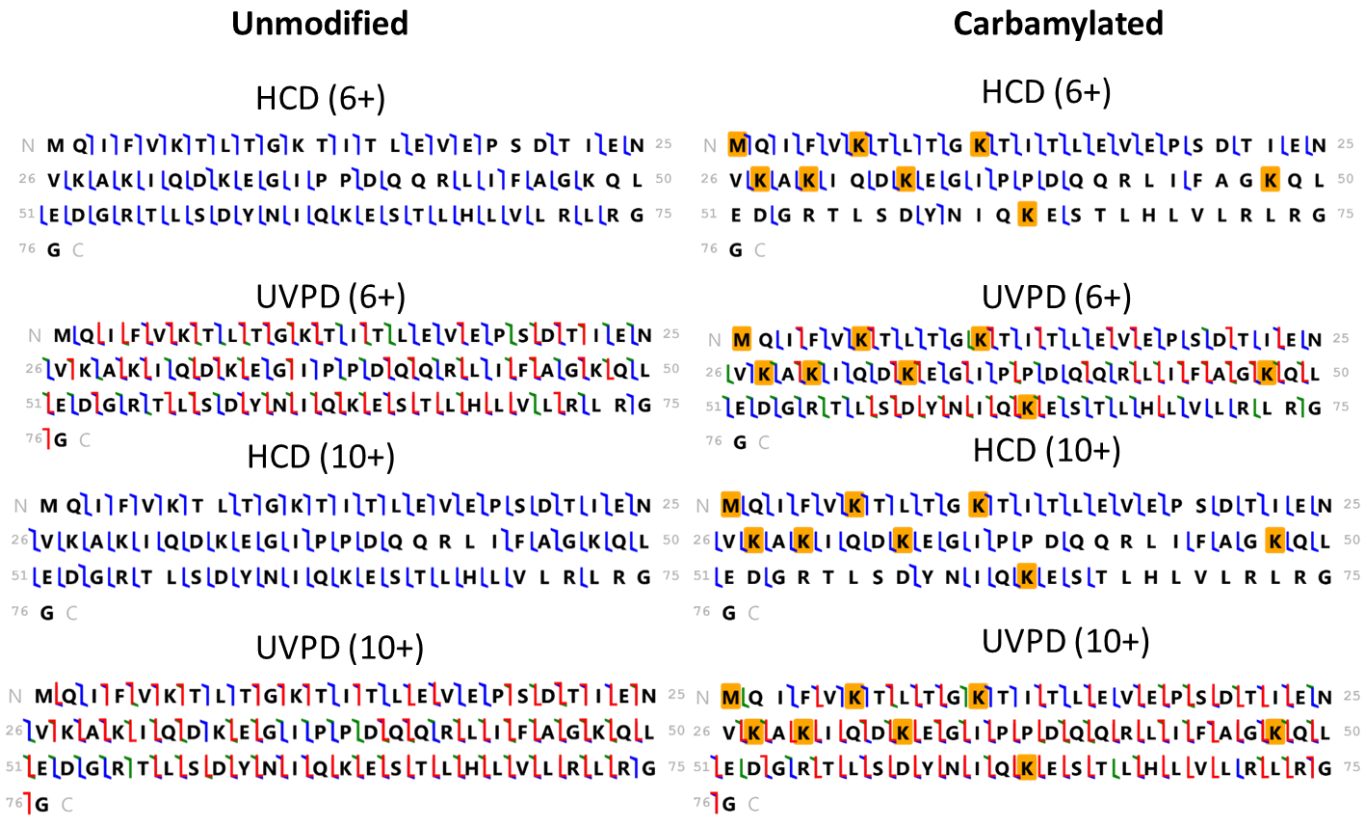


Figure 5.30 Representative sequence maps for unmodified and carbamylated ubiquitin. The types of fragment ions generated from backbone cleavages are color coded as follows: a/x green; b/y blue; c/z red. Gold-shaded residues denote sites of carbamylation.

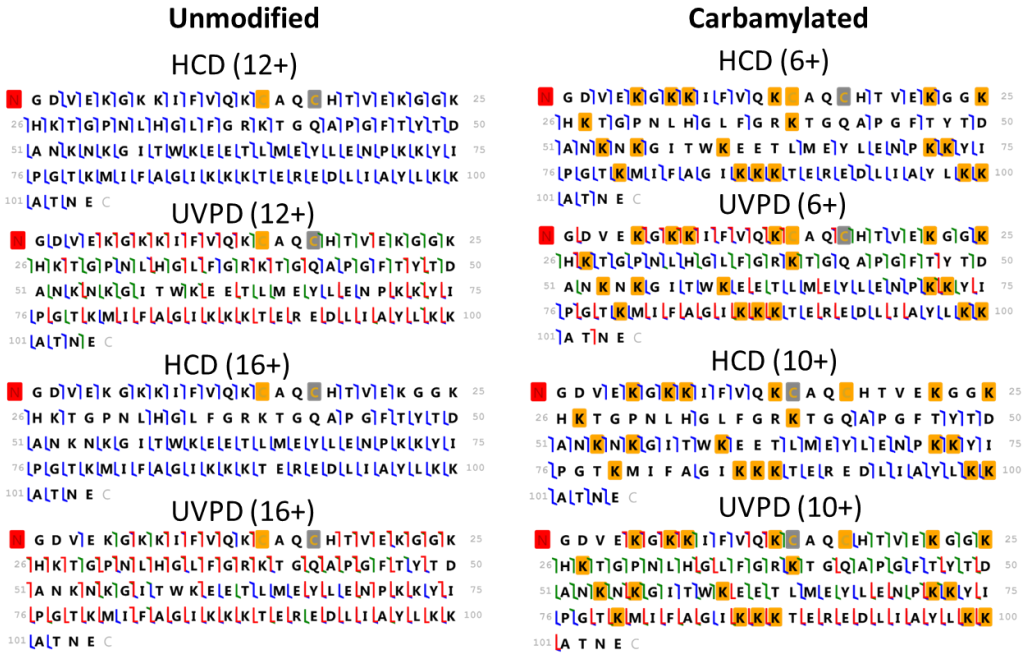


Figure 5.31 Representative sequence maps for unmodified and carbamylated cytochrome c. The types of fragment ions generated from backbone cleavages are color coded as follows: a/x green; b/y blue; c/z red. Gold-shaded boxes denote sites of carbamylation. Red-shaded residues denote acetylated N-termini. Gold-shaded and gray-shaded residues indicated the location of the heme group.

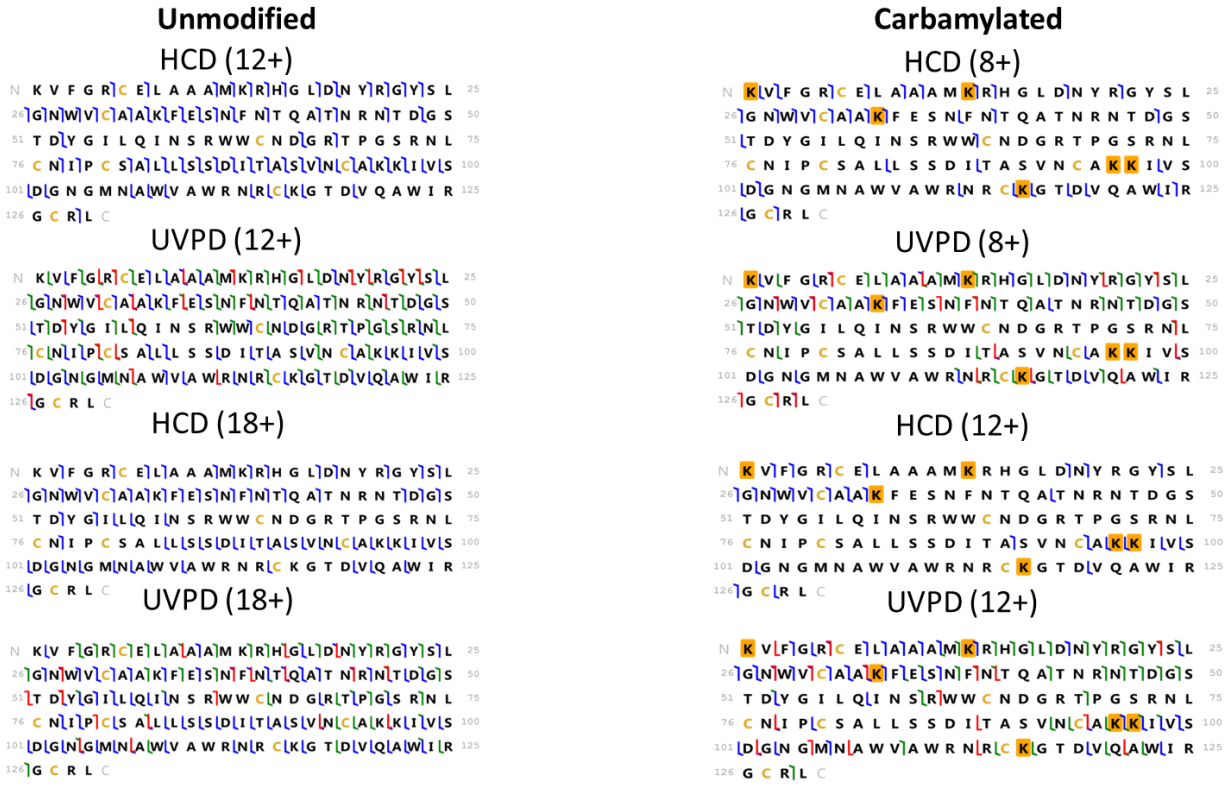


Figure 5.32 Representative sequence maps for unmodified and carbamylated lysozyme. The types of fragment ions generated from backbone cleavages are color coded as follows: a/x green; b/y blue; c/z red. Gold-shaded residues denote sites of carbamylation.

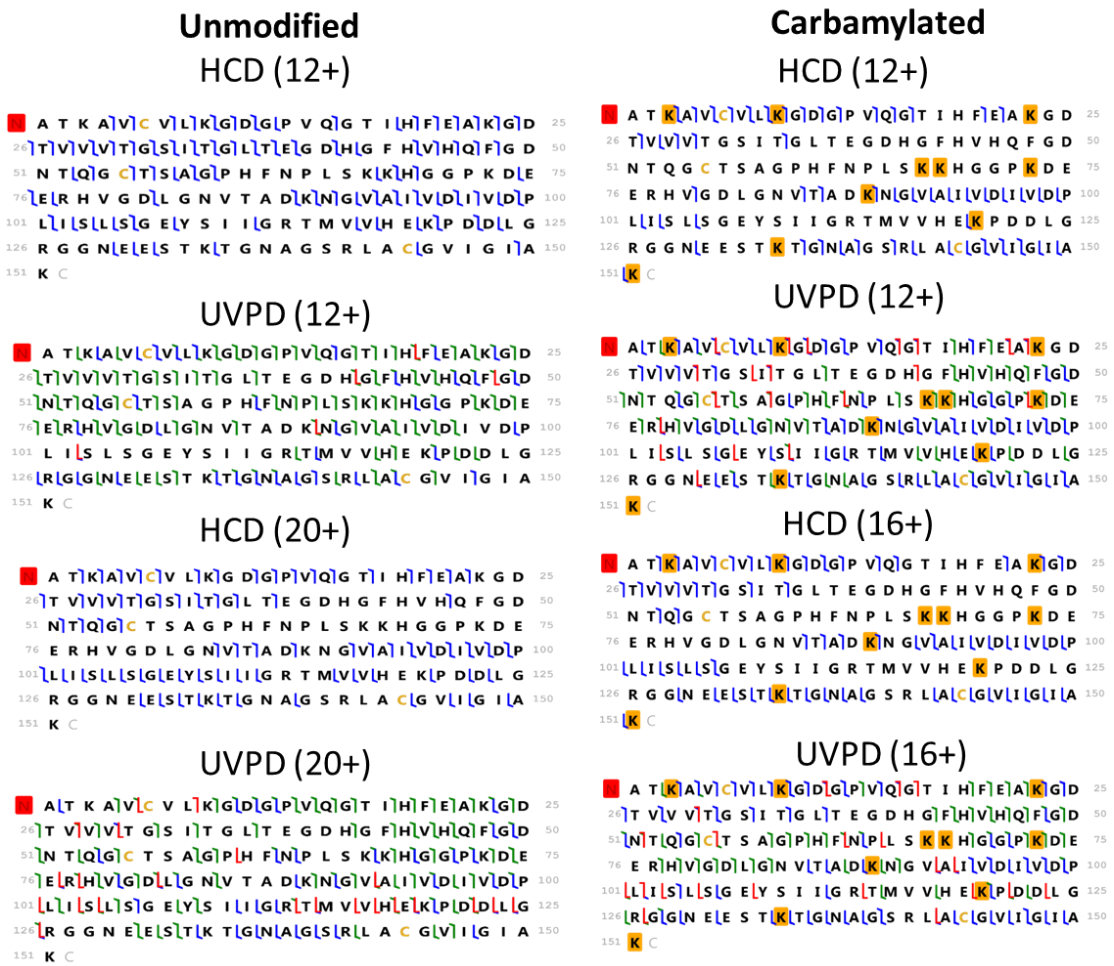


Figure 5.33 Representative sequence maps for unmodified and carbamylated superoxide dismutase. The types of fragment ions generated from backbone cleavages are color coded as follows: a/x green; b/y blue; c/z red. Gold-shaded residues denote sites of carbamylation. Red-shaded residues indicate acetylated N-termini.

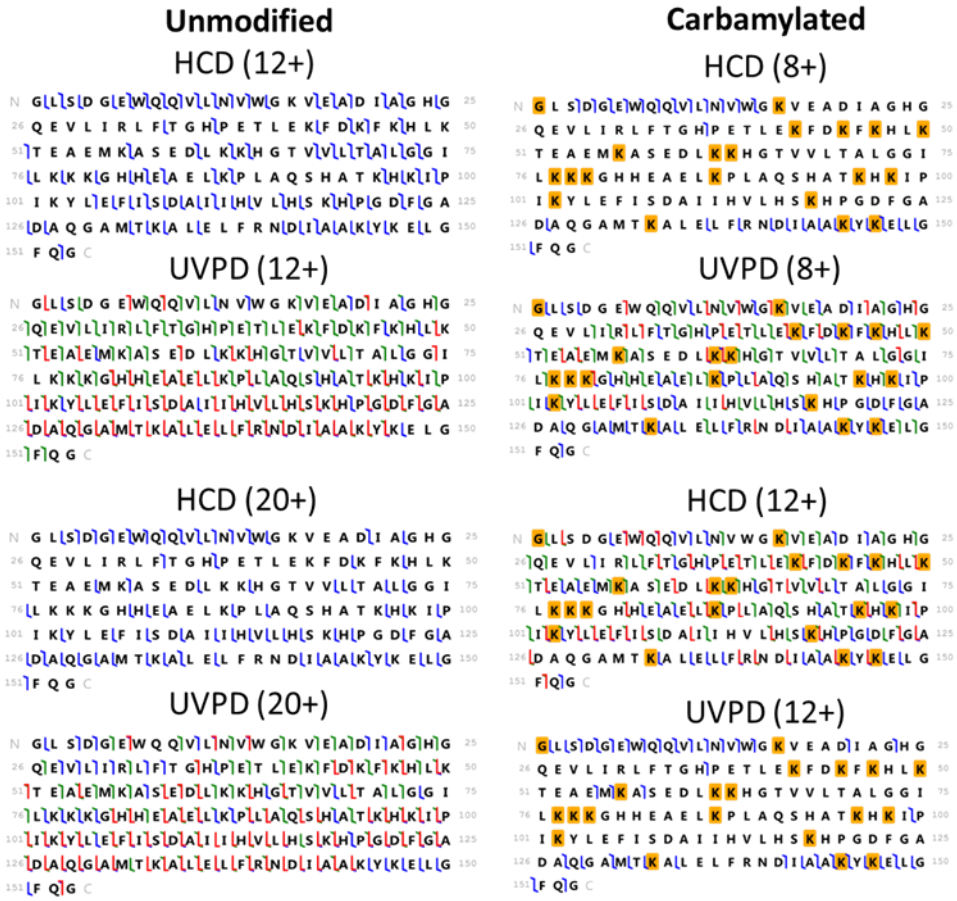


Figure 5.34 Representative sequence maps for unmodified and carbamylated myoglobin. The types of fragment ions generated from backbone cleavages are color coded as follows: a/x green; b/y blue; c/z red. Gold-shaded residues denote sites of carbamylation.

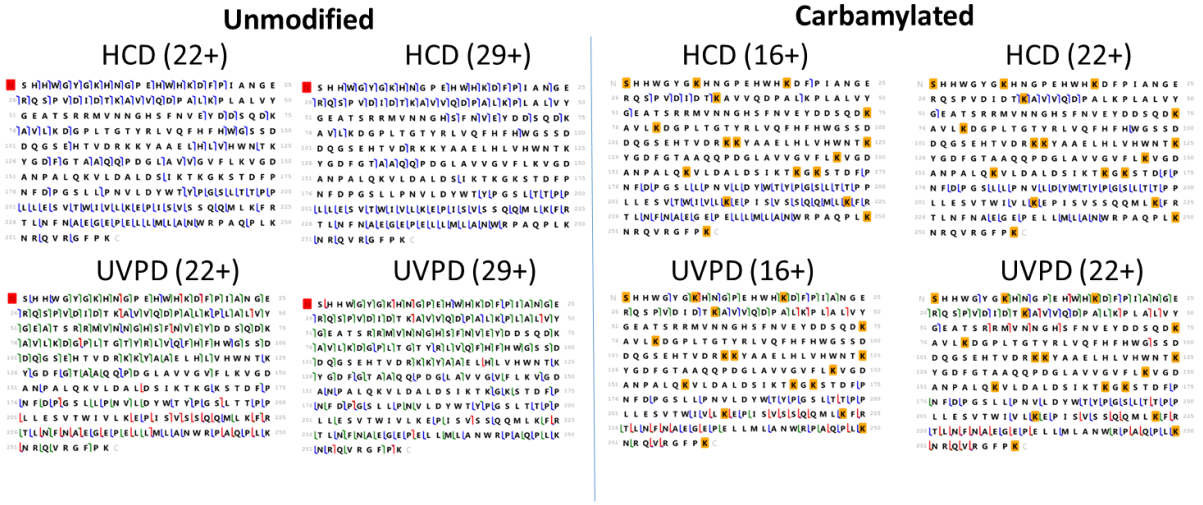


Figure 5.35 Representative sequence maps for unmodified and carbamylated carbonic anhydrase. The types of fragment ions generated from backbone cleavages are color coded as follows: a/x green; b/y blue; c/z red. Gold-shaded residues denote sites of carbamylation. Red-shaded residues indicated acetylated N-termini.

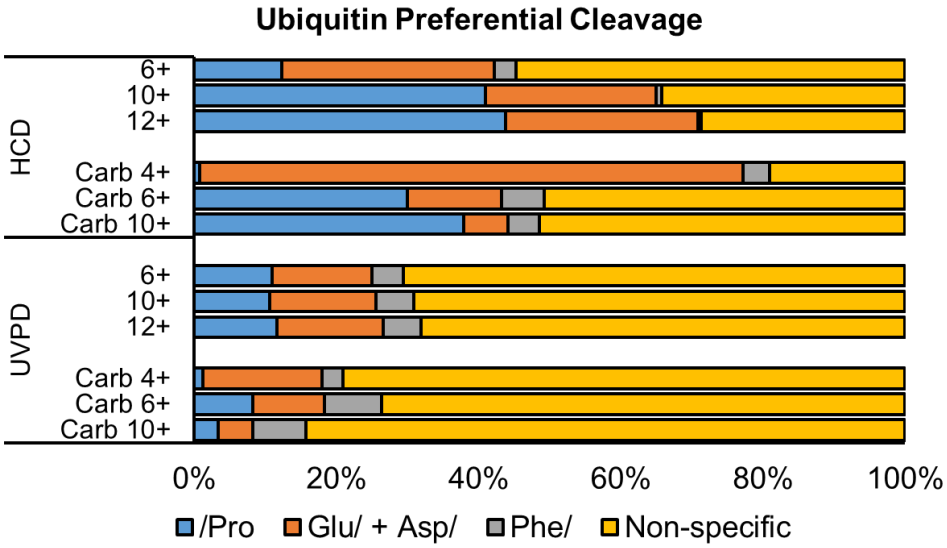


Figure 5.36 Distribution of fragment ions generated by HCD and UVPD categorized as preferential cleavages (N-terminal to proline, C-terminal to glutamic and aspartic acids, C-terminal to phenylalanine) and all non-specific pathways (other N-terminal and C-terminal cleavages) for unmodified and carbamylated ubiquitin.

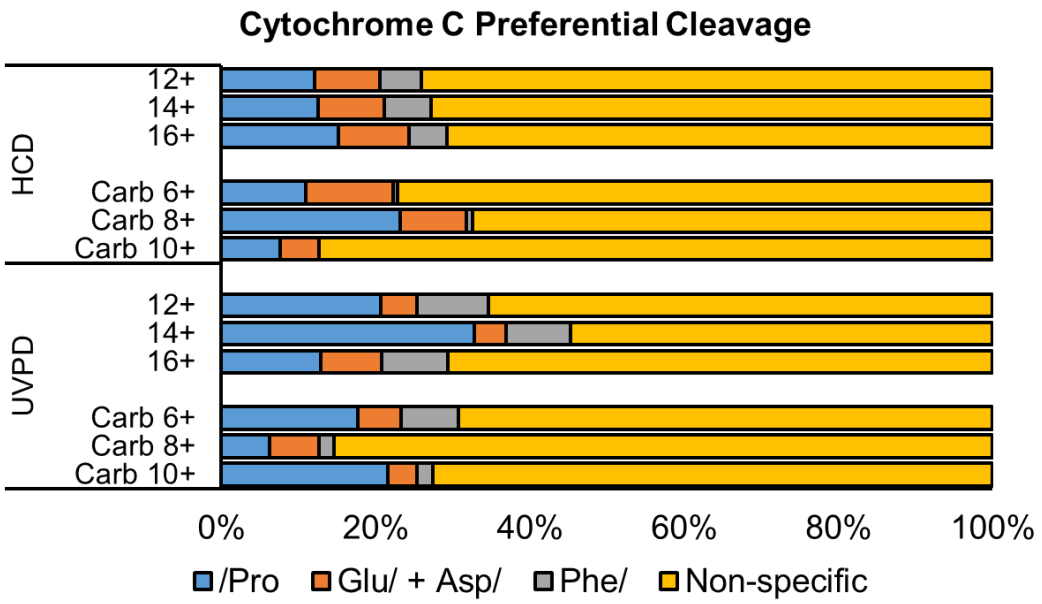


Figure 5.37 Distribution of fragment ions generated by HCD and UVPD categorized as preferential cleavages (N-terminal to proline, C-terminal to glutamic and aspartic acids, C-terminal to phenylalanine) and all non-specific pathways (other N-terminal and C-terminal cleavages) for unmodified and carbamylated cytochrome c.

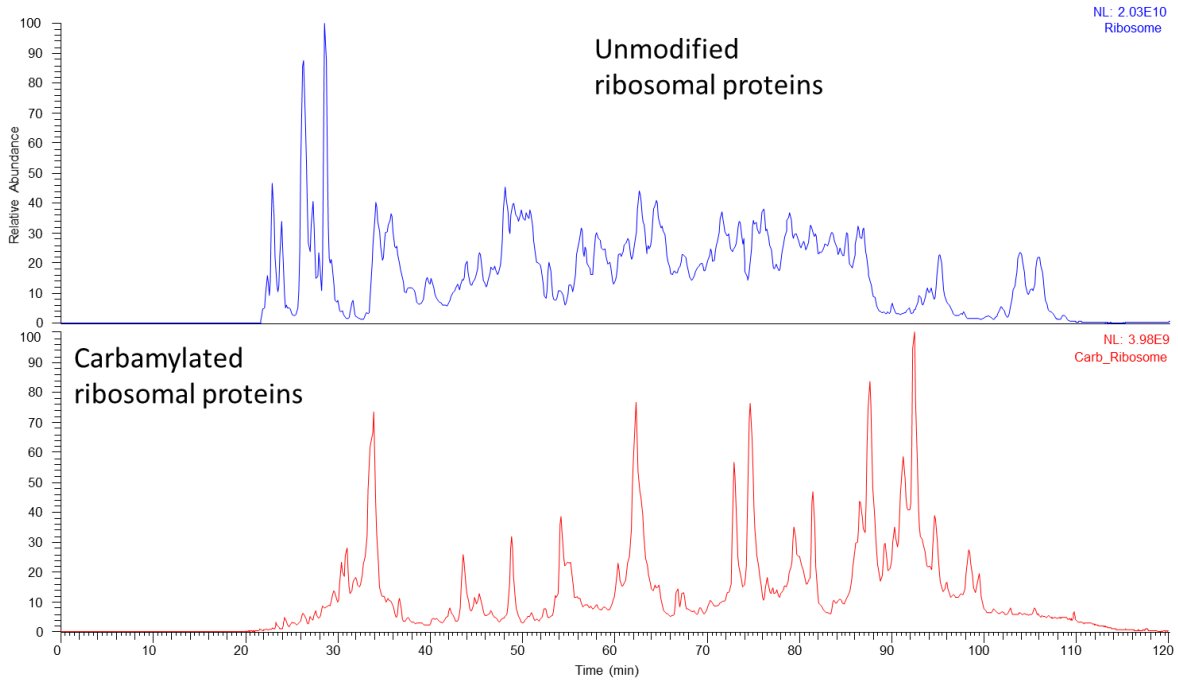


Figure 5.38 LC-MS trace of unmodified (top) and carbamylated (bottom) *E. coli* ribosomal proteins.

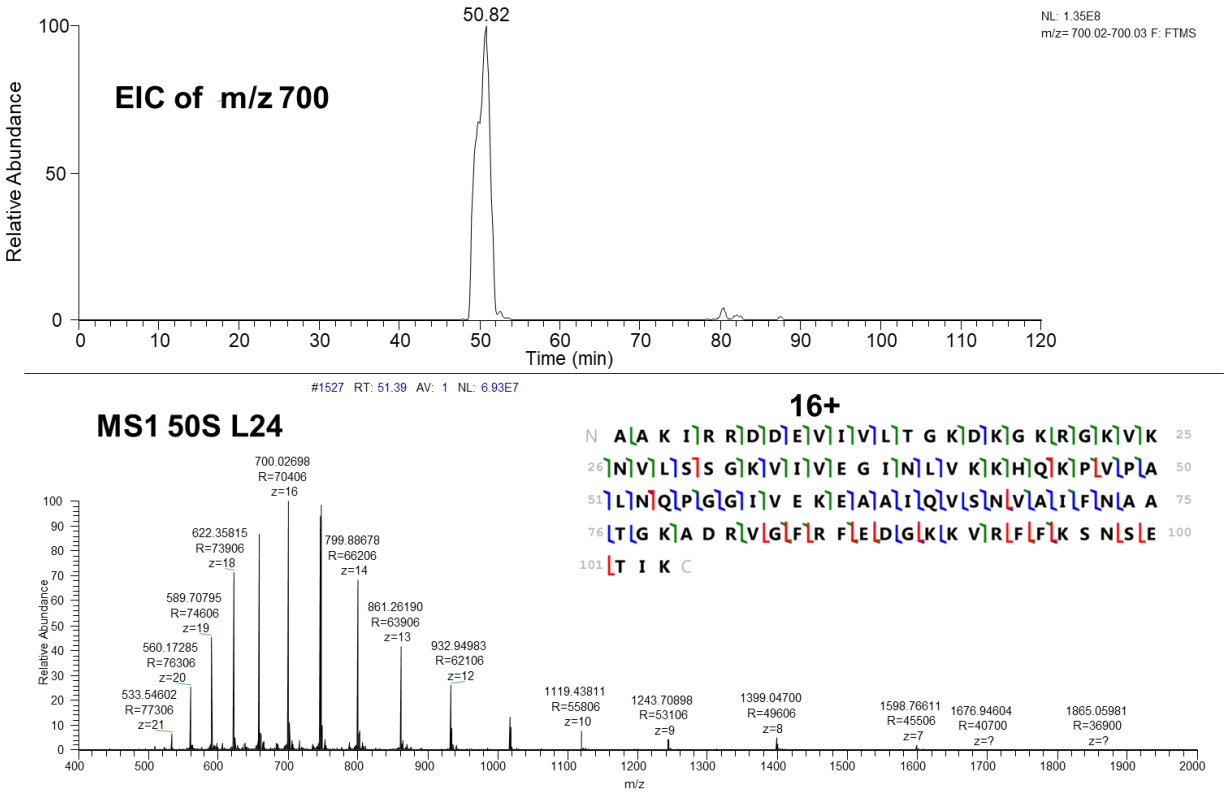


Figure 5.39 Top: Extracted ion chromatogram (EIC) of 50S L24 ribosomal protein showing elution at ~51 minutes. Bottom: MS1 spectrum shows charge states ranging from 7+ to 21+ and the corresponding UVPD sequence coverage map (16+).

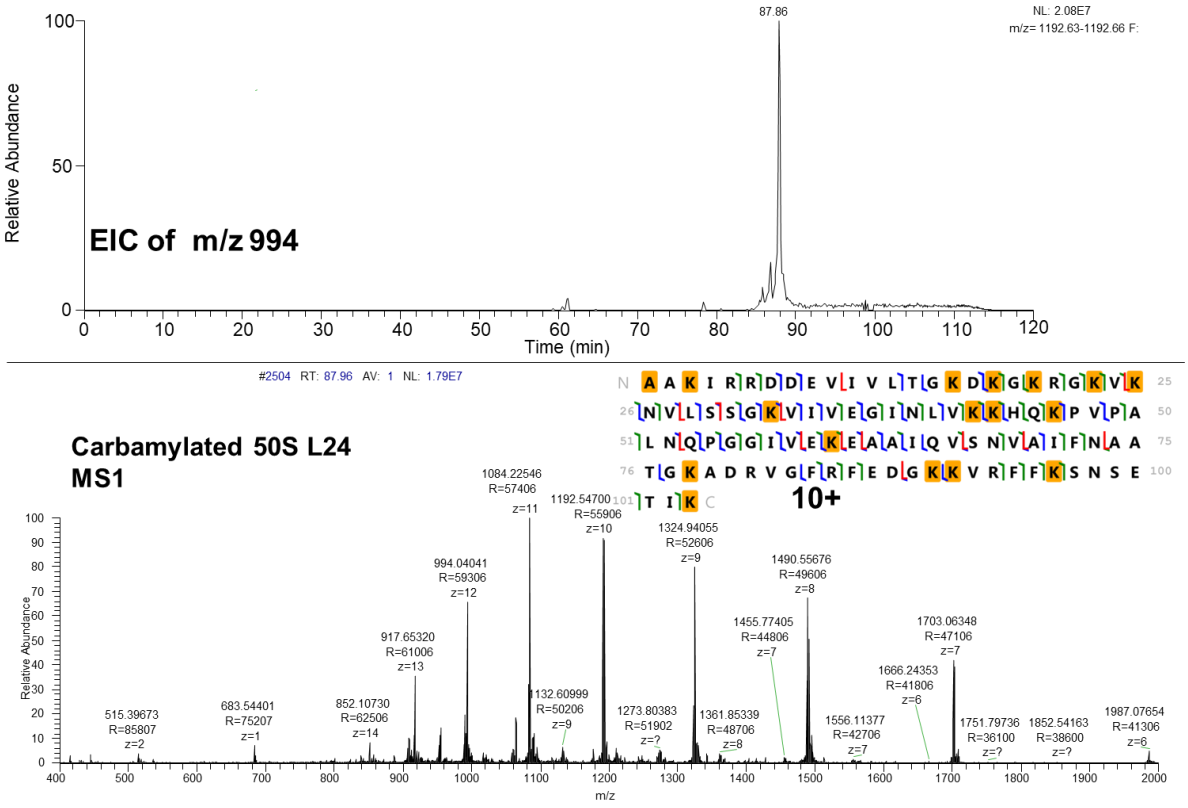


Figure 5.40 Top: Extracted ion chromatogram (EIC) of carbamylated 50S L24 ribosomal protein showing elution at ~88 minutes. Bottom: MS1 mass spectrum showing charge states ranging from 6+ to 13+ and the corresponding UVPD sequence coverage map (10+).

Ubiquitin	**4+**	**6+**	**10+**	**12+**		
Unmod. HCD		83%	84%	89%		
Carb. HCD	64%	56%	72%			
Unmod. UVPD		99%	100%	100%		
Carb. UVPD	87%	97%	99%			
Cytochrome C	**6+**	**8+**	**10+**	**12+**	**14+**	**16+**
Unmod. HCD				88%	83%	72%
Carb. HCD	63%	70%	50%			
Unmod. UVPD				89%	90%	87%
Carb. UVPD	84%	90%	91%			
Lysozyme	**8+**	**10+**	**12+**	**15+**	**18+**	
Unmod. HCD			52%	55%	54%	
Carb. HCD	28%	28%	27%			
Unmod. UVPD			85%	85%	85%	
Carb. UVPD	59%	67%	59%			
SOD	**12+**	**14+**	**16+**	**20+**		
Unmod. HCD	55%		54%	50%		
Carb. HCD	35%	39%	43%			
Unmod. UVPD	73%		77%	71%		
Carb. UVPD	77%	78%	75%			
Myoglobin	**8+**	**10+**	**12+**	**16+**	**20+**	
Unmod. HCD			52%	64%	45%	
Carb. HCD	18%	21%	24%			
Unmod. UVPD			90%	94%	91%	
Carb. UVPD	74%	83%	74%			
Carbonic anhydrase	**16+**	**22+**	**29+**	**34+**		
Unmod. HCD		43%	38%	26%		
Carb. HCD	24%	17%	19%			
Unmod. UVPD		65%	60%	66%		
Carb. UVPD	28%	34%	34%			

Table 5.1 Sequence coverages obtained for various charge states of unmodified and carbamylated proteins by HCD and UVPD.

	HCD		UVPD	
	Unmodified	Carbamylated	Unmodified	Carbamylated
Proteins*	53 ± 1	24 ± 1	46 ± 2**	48 ± 1**
Proteoforms	180 ± 2	75 ± 3	134 ± 1	134 ± 11
Ribosomal proteins	42 ± 1	22 ± 0	40 ± 2	32 ± 4
Number of matching fragments	23 ± 12	13 ± 5	45 ± 25	38 ± 18
*includes proteins identified in sample not associated with *E. coli* ribosomal subunits ** not significantly different ($p=0.19$)				

Table 5.2 LC-MS metrics for UVPD and HCD of a mixture of *E. coli* ribosomal proteins.

Protein			HCD Unmodified				HCD Carbamylated			
Name	M.W (kDa)	Accession	R.T. (mn)	Coverage	Peak Area	FWHM (min)	R.T. (min)	Coverage	Peak Area	FWHM (min)
50S L34	6.3	P0A7N4	26.0	*62%*	6.6E+08	0.3	35.7	*29%*	1.1E+08	0.5
50S L36	4.3	P0A7Q6	27.3	*49%*	2.6E+09	0.3	48.1	*27%*	4.2E+08	0.7
50S L24	11.2	P60624	51.8	*62%*	4.2E+09	0.5	87.2	*27%*	2.0E+09	0.3
30S S19	10.3	P0A7U3	60.3	*58%*	1.7E+09	0.7	86.8	*22%*	4.1E+08	0.4
50S L18	12.8	P0C018	78.9	*27%*	1.3E+09	0.8	90.4	*13%*	2.3E+09	0.7
50S L15	14.9	P02413	83.8	*21%*	6.6E+08	1.6	93.2	*14%*	7.7E+07	0.4

Protein			UVPD Unmodified				UVPD Carbamylated			
Name	M.W (kDa)	Accession	R.T. (mn)	Coverage	Peak Area	FWHM (min)	R.T. (min)	Coverage	Peak Area	FWHM (min)
50S L34	6.3	P0A7N4	26.2	*87%*	9.4E+09	0.6	33.4	*84%*	4.4E+09	0.8
50S L36	4.3	P0A7Q6	27.4	*97%*	1.0E+10	0.6	45.7	*54%*	1.6E+08	0.5
50S L24	11.2	P60624	50.8	*80%*	6.8E+09	0.7	87.9	*71%*	2.1E+09	0.3
30S S19	10.3	P0A7U3	56.3	*87%*	7.3E+09	1.3	87.2	*48%*	2.2E+08	1.7
50S L18	12.8	P0C018	75.9	*36%*	1.1E+09	0.8	91.1	*63%*	1.3E+09	0.7
50S L15	14.9	P02413	79.1	*36%*	2.2E+09	1.1	94.8	*43%*	8.5E+07	0.5

Table 5.3 Comparison of HCD and UVPD results for selected ribosomal proteins.

5.6 REFERENCES

(1) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9* (1), 499–519.

(2) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. *Anal. Chem.* **2014**, *86* (6), 3005–3012.

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

(4) Smith, L. M.; Kelleher, N. L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla, L.; Kruppa, G.; Zubarev, R.; et al. Proteoform: a single term describing protein complexity. *Nat. Methods* **2013**, *10* (3), 186–187.

(5) Senko, M. W.; Speir, J. P.; McLafferty, F. W. Collisional Activation of Large Multiply Charged Ions Using Fourier Transform Mass Spectrometry. *Anal. Chem.* **1994**, *66* (18), 2801–2808.

(6) Zhou, M.; Paša-Tolić, L.; Stenoien, D. L. Profiling of Histone Post-Translational Modifications in Mouse Brain with High-Resolution Top-Down Mass Spectrometry. *J. Proteome Res.* **2017**, *16* (2), 599–608.

(7) Fornelli, L.; Durbin, K. R.; Fellers, R. T.; Early, B. P.; Greer, J. B.; LeDuc, R. D.; Compton, P. D.; Kelleher, N. L. Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J. Proteome Res.* **2017**, *16* (2), 609–618.

(8) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci.* **2004**, *101* (26), 9528–9533.

(9) Anderson, L. C.; Karch, K. R.; Ugrin, S. A.; Coradin, M.; English, A. M.; Sidoli, S.; Shabanowitz, J.; Garcia, B. A.; Hunt, D. F. Analyses of Histone Proteoforms Using Front-end Electron Transfer Dissociation-enabled Orbitrap Instruments. *Mol. Cell. Proteomics* **2016**, *15* (3), 975–988.

(10) Cammarata, M. B.; Brodbelt, J. S. Characterization of Intra- and Intermolecular Protein Crosslinking by Top Down Ultraviolet Photodissociation Mass Spectrometry. *ChemistrySelect* **2016**, *1* (3), 590–593.

(11) Cannon, J. R.; Kluwe, C.; Ellington, A.; Brodbelt, J. S. Characterization of green fluorescent proteins by 193 nm ultraviolet photodissociation mass spectrometry. *PROTEOMICS* **2014**, *14* (10), 1165–1173.

(12) Cannon, J. R.; Martinez-Fonts, K.; Robotham, S. A.; Matouschek, A.; Brodbelt, J. S. Top-Down 193-nm Ultraviolet Photodissociation Mass Spectrometry for Simultaneous Determination of Polyubiquitin Chain Length and Topology. *Anal. Chem.* **2015**, *87* (3), 1812–1820.

(13) Cammarata, M.; Lin, K.-Y.; Pruet, J.; Liu, H.; Brodbelt, J. Probing the Unfolding of Myoglobin and Domain C of PARP-1 with Covalent Labeling and Top-Down

Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2014**, *86* (5), 2534–2542.
(14) 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. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.
(15) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. Integration of Ultraviolet Photodissociation with Proton Transfer Reactions and Ion Parking for Analysis of Intact Proteins. *Anal. Chem.* **2016**, *88* (1), 1008–1016.
(16) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. *Anal. Chem.* **2014**, *86* (4), 2185–2192.
(17) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. The C-Score: A Bayesian Framework to Sharply Improve Proteoform Scoring in High-Throughput Top Down Proteomics. *J. Proteome Res.* **2014**, *13* (7), 3231–3240.
(18) Armirotti, A.; Damonte, G. Achievements and perspectives of top-down proteomics. *PROTEOMICS* **2010**, *10* (20), 3566–3576.
(19) Chanthamontri, C.; Liu, J.; McLuckey, S. A. Charge state dependent fragmentation of gaseous α-synuclein cations via ion trap and beam-type collisional activation. *Int. J. Mass Spectrom.* **2009**, *283* (1–3), 9–16.
(20) Reid, G. E.; Wu, J.; Chrisman, P. A.; Wells, J. M.; McLuckey, S. A. Charge-State-Dependent Sequence Analysis of Protonated Ubiquitin Ions via Ion Trap Tandem Mass Spectrometry. *Anal. Chem.* **2001**, *73* (14), 3274–3281.
(21) Amunugama, R.; Hogan, J. M.; Newton, K. A.; McLuckey, S. A. Whole Protein Dissociation in a Quadrupole Ion Trap: Identification of an a Priori Unknown Modified Protein. *Anal. Chem.* **2004**, *76* (3), 720–727.
(22) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; Elsevier, 2005; Vol. 402, pp 148–185.
(23) Cobb, J. S.; Easterling, M. L.; Agar, J. N. Structural characterization of intact proteins is enhanced by prevalent fragmentation pathways rarely observed for peptides. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (6), 949–959.
(24) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000**, *35* (12), 1399–1406.
(25) Bythell, B. J.; Suhai, S.; Somogyi, Á.; Paizs, B. Proton-Driven Amide Bond-Cleavage Pathways of Gas-Phase Peptide Ions Lacking Mobile Protons. *J. Am. Chem. Soc.* **2009**, *131* (39), 14057–14065.

(26) Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R. Influence of Basic Residue Content on Fragment Ion Peak Intensities in Low-Energy Collision-Induced Dissociation Spectra of Peptides. *Anal. Chem.* **2004**, *76* (5), 1243–1248.
(27) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J. Am. Chem. Soc.* **1996**, *118* (35), 8365–8374.
(28) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal. Chem.* **2015**, *87* (14), 7109–7116.
(29) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, Ü. A.; Tsybin, Y. O. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem. Soc. Rev.* **2013**, *42* (12), 5014.
(30) Pesavento, J. J.; Kim, Y.-B.; Taylor, G. K.; Kelleher, N. L. Shotgun Annotation of Histone Modifications: A New Approach for Streamlined Characterization of Proteins by Top Down Mass Spectrometry. *J. Am. Chem. Soc.* **2004**, *126* (11), 3386–3387.
(31) Karch, K. R.; DeNizio, J. E.; Black, B. E.; Garcia, B. A. Identification and interrogation of combinatorial histone modifications. *Front. Genet.* **2013**, *4*.
(32) Morrison, L. J.; Rosenberg, J. A.; Singleton, J. P.; Brodbelt, J. S. Statistical Examination of the a and a + 1 Fragment Ions from 193 nm Ultraviolet Photodissociation Reveals Local Hydrogen Bonding Interactions. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (9), 1443–1453.
(33) Greer, S. M.; Cannon, J. R.; Brodbelt, J. S. Improvement of Shotgun Proteomics in the Negative Mode by Carbamylation of Peptides and Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2014**, *86* (24), 12285–12290.
(34) Iavarone, A. T.; Jurchen, J. C.; Williams, E. R. Supercharged Protein and Peptide Ions Formed by Electrospray Ionization. *Anal. Chem.* **2001**, *73* (7), 1455–1460.
(35) Iavarone, A. T.; Williams, E. R. Collisionally Activated Dissociation of Supercharged Proteins Formed by Electrospray Ionization. *Anal. Chem.* **2003**, *75* (17), 4525–4533.
(36) Krusemark, C. J.; Frey, B. L.; Belshaw, P. J.; Smith, L. M. Modifying the charge state distribution of proteins in electrospray ionization mass spectrometry by chemical derivatization. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (9), 1617–1625.
(37) Pitteri, S. J.; Reid, G. E.; McLuckey, S. A. Affecting Proton Mobility in Activated Peptide and Whole Protein Ions via Lysine Guanidination. *J. Proteome Res.* **2004**, *3* (1), 46–54.
(38) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2016**, *88* (1), 1044–1051.

(39) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *PROTEOMICS* **2015**, *15* (7), 1235–1238.
(40) Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. Influence of Secondary Structure on the Fragmentation of Protonated Peptides. *J. Am. Chem. Soc.* **1999**, *121* (22), 5142–5154.
(41) Qin, J.; Chait, B. T. Preferential Fragmentation of Protonated Gas-Phase Peptide Ions Adjacent to Acidic Amino Acid Residues. *J. Am. Chem. Soc.* **1995**, *117* (19), 5411–5412.
(42) Jockusch, R. A.; Schnier, P. D.; Price, W. D.; Strittmatter, E. F.; Demirev, P. A.; Williams, E. R. Effects of Charge State on Fragmentation Pathways, Dynamics, and Activation Energies of Ubiquitin Ions Measured by Blackbody Infrared Radiative Dissociation. *Anal. Chem.* **1997**, *69* (6), 1119–1126.
(43) Breci, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H. Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra. *Anal. Chem.* **2003**, *75* (9), 1963–1971.
(44) Newton, K. A.; Pitteri, S. J.; Laskowski, M.; McLuckey, S. A. Effects of Single Amino Acid Substitution on the Collision-Induced Dissociation of Intact Protein Ions: Turkey Ovomucoid Third Domain. *J. Proteome Res.* **2004**, *3* (5), 1033–1041.
(45) Paizs, B.; Suhai, S. Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **2005**, *24* (4), 508–548.

Chapter 6

Extending Proteome Coverage by Combining MS/MS Methods and a Modified Bioinformatics Platform adapted for Database Searching of Positive and Negative Polarity 193 nm Ultraviolet Photodissociation Mass Spectra

6.1 OVERVIEW

To extend proteome coverage obtained from bottom-up mass spectrometry approaches, three complementary ion activation methods, higher energy collision dissociation (HCD), ultraviolet photodissociation (UVPD), and negative mode UVPD (NUVPD), are used to interrogate the tryptic peptides in a human hepatocyte lysate using a high performance Orbitrap™ mass spectrometer. The utility of combining results from multiple activation techniques (HCD+UVPD+NUVPD) is analyzed for total depth and breadth of proteome coverage. This study also benchmarks a new version of the Byonic algorithm which has been customized for database searches of UVPD and NUVPD data. Searches utilizing the customized algorithm resulted in over 50% more peptide identifications for UVPD and NUVPD tryptic peptide datasets compared to other search algorithms. Inclusion of UVPD and NUVPD spectra resulted in over 600 additional protein identifications relative to HCD alone.

6.2 INTRODUCTION

Ongoing advances in mass spectrometry (MS)-based proteomics have resulted in the identification of an ever-increasing number of proteins and proteoforms in cell lysates and other complex biological samples.[1–4] This increase arises from several technological improvements. New sample preparation methods target scarce post-translational modifications (PTMs) or enrich specific classes of molecules, thus improving the depth or breadth of sample analysis.[5–9] High performance instrumentation has been designed specifically for proteomics applications with greater throughput at the forefront, such as implementation of parallelization methods.[10] In addition, software that exploits the increasing resolving power and mass accuracy metrics of new mass spectrometers and that accommodates a growing array of novel ion activation methods has been developed.[11–17] Until recently these advancements have primarily been adapted for positive mode ionization (e.g. formation and analysis of protonated peptides), owing to three reasons. First, electrospray ionization typically generates greater signal intensities in the positive mode because signal suppression arising from corona discharge under negative polarity conditions is a common occurrence.[18,19] Second, liquid chromatography methods usually use an acidic mobile phase modifier to improve LC peak shape which naturally enhances protonation of eluting peptides.[20,21] Third, protonated peptides generally exhibit more informative fragmentation patterns upon collisional activation than do deprotonated peptides, thus facilitating effective database searches.[22–24] These three factors have contributed to the pervasiveness of positive mode proteomics and account for the

prevailing success of bottom-up strategies. In contrast, far less effort has been devoted to the negative ionization mode, which means there are avenues of opportunity for extending proteomic analysis. For example, the negative mode is well suited for acidic peptides, ones that might be overlooked in the positive mode owing to their naturally low abundances (such as for phosphopeptides), spectral congestion from co-eluting peptides, or poor ionization due to high pKa values. Compounding the experimental difficulties of LC-MS methods in the negative mode is the dearth of bioinformatics software adapted to analyze the MS/MS spectra of deprotonated peptides. The often uninformative nature of MS/MS spectra generated by collision induced dissociation (CID) of deprotonated peptides, spectra which are often dominated by small neutral losses, has played a prominent role in this context.[25]

Despite these hurdles, some recent effort has been directed at exploring the negative ionization mode for bottom-up proteomics, and significant inroads have been reported.[13,26–32] Primarily two alternative ion activation methods for peptide anions are under development for high throughput workflows: negative electron transfer dissociation (NETD)[28–30,32] and negative ultraviolet photodissociation (NUVPD).[13,26,33,34] The performance of NETD has been significantly improved by incorporation of a supplemental activation step to increase the fragmentation efficiency.[31,32] This enhanced NETD method is termed activated ion NETD (AI-NETD) and has shown promising results.[31,32,35] In the most impressive AI-NETD proteomics study to date, Riley *et al.* identified over 8600 unique peptides and over 1300 proteins in a 90-min analysis of human embryonic stem cell

lysate using AI-NETD.[32] UVPD uses photon absorption, rather than electron transfer, to activate and dissociate peptides. Madsen *et al.* identified over 3600 peptides and over 800 proteins from a cell lysate of *Halobacterium salinarum* by NUVPD.[13] The hallmark of both these novel activation methods is their ability to generate diagnostic fragment ions for negatively charged peptides in a high throughput LC-MS/MS mode. Both AI-NETD and NUVPD tend to generate *a* and *x* ions; either as radical or non-radical forms. The development of these methods has spurred adaptation and customization of bioinformatics platforms to support data analysis. Originally OMSSA was modified by the Coon group for database searches of NETD and AI-NETD spectra,[30] whereas MassMatrix was modified by the Brodbelt group for analysis of NUVPD data.[13] More recently the bioinformatics platform Byonic was modified and optimized for negative polarity spectral matching and was shown to outperform previous software for analysis of NETD data acquired in a high throughput LC-MS manner, culminating in greater than 50% more peptide identifications in a tryptic digest of yeast to their in-house customized OMSSA software.[12]

In this study, we implemented UVPD and NUVPD on a high performance Orbitrap mass spectrometer and demonstrated the complementarity of using negative and positive polarity analyses to broaden proteome coverage. The Byonic database search algorithm was adapted for interrogating UVPD spectra in both positive and negative ion modes to facilitate database searches, yielding the greatest number of peptide spectral matches (PSMs) and identified peptides and proteins in a negative polarity LC-UVPD-MS

experiment reported to date. Peptide and protein metrics from NUVPD datasets analyzed by both Byonic and MassMatrix were compared, and the performance of Byonic was compared to Proteome Discoverer/SEQUEST for assessment of UVPD and HCD data in the positive mode.

6.3 MATERIALS & METHODS

6.3.1 Preparation of human hepatocyte lysate

HC-04 cells (MRA-975, ATCC Manassas, VA), human hepatocytes, were maintained in Dulbecco's modified Eagle's medium/Ham's F12 50/50 mix (Gibco Life Technologies, Grand Island, NY) supplemented with 10% heat-inactivated fetal bovine serum (Gibco Life Technologies) and 2 mM l-glutamine (Gibco Life Technologies). Cells were washed twice with PBS and then scraped from the culture surface with 10 ml of ice-cold PBS and centrifuged at 2,000 *g* at 4°C for 5 minutes. The supernatant was then removed and replaced with 1 ml of ice-cold urea lysis buffer (50 mM Tris, pH 8, 8 M urea, 150 mM NaCl, 1mM EDTA) supplemented with 2 μg/ml of aprotinin (MD Biomedicals, Solon, OH), 10 μg/ml of leupeptin (Roche Diagnostics, Indianapolis, IN), 1mM iodoacetamide (Sigma-Aldrich, St. Louis, MO), 50 μM PR-619 (Lifesensors, Malvern, PA) and 1 mM phenylmethanesulfonyl fluoride (PMSF; Sigma-Aldrich). Cells were lysed by extrusion through a 20-gauge needle (Becton Dickinson, Franklin Lakes, NJ) attached to a 1-ml syringe (Becton Dickinson) 25 times and placed on ice for 40 minutes. The lysate was

then cleared by centrifugation at 14,000 *g* for 20 minutes and final protein concentration was determined using the DC protein assay (Bio-Rad, Hercules, CA).

6.3.2 Materials

HPLC solvents were obtained from EMD Millipore (Temecula, CA), and buffer components were obtained from Sigma-Aldrich (St. Louis, MO). Proteomics-grade trypsin was obtained from Promega (Madison, WI). All other reagents and solvents were obtained from Thermo Fisher Scientific (Fairlawn, NJ).

6.3.3 Protein digestion

Proteins isolated from the liver cell lysate were sequentially reduced in 5 mM dithiothreitol (DTT) for 30 min at 55 °C then alkylated in 10 mM iodoacetamide at room temperature in the dark for 30 min. Alkylation was quenched with a second aliquot of DTT, bringing the final concentration of DTT to ∼10 mM. After alkylation, the sample was split into three separate aliquots for triplicate analysis.

Trypsin was added in a 1:50 protease-to-protein ratio, and the solution was buffered at pH 8 in 100 mM ammonium bicarbonate. Digestion proceeded for 18 h at 37 °C. After digestion, all samples were quenched with 1% formic acid and cleaned over a spin cartridge loaded with C18 resin (Pierce Biotechnology) prior to LC–MS analysis.

6.3.4 LC-MS

Chromatographic separations were performed using a Dionex RSLC 3000 nanobore LC system with water (A) and acetonitrile (B) as mobile phases, with each containing 0.1% formic acid. The trap (0.075 cm x 3.5 cm) and analytical column (0.075 cm × 20 cm, with integrated emitter) were packed in-house using 3.5 μm XBridge BEH C18 media (Waters, Milford, MA). Approximately 1 μg of digest was loaded onto the trap column at 5 μL/min for 5 min, then separated on the analytical column with a gradient that changed from 0 to 35% B over the course of 240 min at a flow rate of 300 nL min^{-1}. For nanospray, 1.8 kV was applied at a precolumn liquid voltage junction.

The hepatocyte digests were analyzed on a Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific Instruments) equipped with a 193 nm excimer laser (Coherent, Santa Clara, CA) and modified to allow UVPD in the linear ion traps as previously described.[36] UVPD was performed in the low-pressure trap using two pulses (2.5 mJ, 5 ns per pulse). HCD was performed in the ion routing multipole at 35% NCE during a 0.1 ms period.

The Orbitrap mass spectrometer was operated using the following parameters regardless of fragmentation type: top speed (3 sec cycle), MS1 at 60 000 resolving power, 1 μscan per spectrum, 1×10^5 AGC target, 15 V source fragmentation, peptide monoisotopic precursor selection; and MS2 at 15 000 resolving power, 1 μscan per spectrum, and 1×10^5 AGC target. The minimum signal required for MS2 selection was 400,000, and the isolation width was fixed at 3 *m*/*z*. Dynamic exclusion was enabled for

30 s with a repeat count of 3 and exclusion duration of 60 sec. In essence, this means that any *m/z* value may be selected and subjected to MS/MS three times within a 30 s interval, then excluded from selection for the next 60 s.

6.3.5 Data analysis

Detailed descriptions of the data analysis workflow and parameters are provided in the supplemental material. Briefly the RAW data files were imported directly into Proteome Discoverer (SEQUEST+ Percolator) or Byonic for processing; data files were converted to mzXML files by use of the MassMatrix data conversion tools for analysis by MassMatrix. Results from all three informatics platforms were filtered to 1% FDR. Data was searched against the human proteome (UP000005640) downloaded from Uniprot (02/14/2017). Positive polarity HCD and UVPD LC-MS data files were analyzed using Byonic and SEQUEST+Percolator. Negative polarity UVPD data were analyzed using Byonic and MassMatrix.

6.4 RESULTS AND DISCUSSION

The goal of this study is to evaluate the utility of combining the results of three MS/MS activation techniques for the analysis of a tryptic cell lysate. The three activation techniques considered were HCD for positively charged peptides and UVPD for both positively and negatively charged peptides. For this effort a widely available proteomics search engine, Byonic, was customized to further improve the informatics tools available for UVPD spectra. Herein we describe how the use of UVPD for both positively and

negatively charged peptides can uniquely complement conventional HCD of positively charged peptides. At the same time, we report the performance enhancements offered by a search algorithm customized for UVPD data, thus significantly extending the breadth of proteome coverage obtained from UVPD data based on greater numbers of identified peptides and proteins and greater sequence coverage of proteins identified. We report results based on analyses of tryptic digests of human liver cell lysate. Baseline performance was established using MassMatrix (previously the only tool available for large scale analysis of proteomics datasets acquired in the negative mode by UVPD) and SEQUEST+Percolator for conventional MS/MS data acquired in the positive mode.

Byonic has been shown to be more sensitive than Sequest, Mascot, PEAKS, MaxQuant, and MS Amanda.[37] Enhanced sensitivity has been achieved through a series of refinements.[38] These refinements include: (1) use of peak ranks (that is, most intense, second most intense, and so forth) rather than presence/absence, (2) sequence-based peak intensity prediction, implemented as multiplicative weights (for example, a large weight for collisional fragmentation on the N-terminal side of proline), (3) reduced score for larger *m/z* errors, (4) "2D" FDR estimation (that is, simultaneous control of both protein and peptide FDRs), and, most relevant here, (5) peak prediction based upon fragmentation method. For positive-mode UVPD, Byonic scores a- and a-dot (radical) ions, x-ions, b-ions, and y-ions. Sequence-based adjustments of weights, for example, up-weighting on the N-terminal side of proline, are as in HCD. For negative mode UVPD, Byonic scores negative a-dot (radical) and x-ions, along with neutral losses of carbon dioxide, with

roughly equal weights for all predicted ions. In comparison to Byonic, Sequest does not use the peptide sequence to predict theoretical fragment ion abundances, and the statistical basis for considering fragment ion abundances is less sophisticated. In addition, the quantization of *m/z* errors (categorization based on full value, half value, zero) is less refined in Sequest. A closer examination of the nuances of MassMatrix based on manual inspection of high-scoring decoy matches indicates that unfragmented precursor ions may be mistaken for product ions, and there appears to be less consistency of scoring based on specific ion types. These factors, which are considered or addressed in the Byonic algorithm, may contribute to the different outcomes of Byonic relative to Sequest and MassMatrix.

6.4.1 Complementarity of UVPD and HCD

Figure 6.1 shows representative HCD, UVPD and NUVPD mass spectra for one peptide analyzed in both positive and negative modes (HCD (3+), UVPD (3+) and NUVPD (2-) spectra of LNDGHFMPVLGFGTYAPPEVPR, a tryptic peptide originating from protein P42330). HCD generates the expected *b/y* fragment ions for the protonated peptide, whereas UVPD generates predominantly *b/y* ions, some *a* ions, and to a lesser extent some *c, x* and *z* ions. Neither UVPD nor NUVPD generates more than 50% of the total possible *y* or *x* ions that contain the C-terminus; however, when combined nearly 70% of the possible *x/y*-type ions are confirmed. HCD of this peptide resulted in 62% coverage, UVPD resulted in 70% sequence coverage; and NUVPD afforded 70% sequence coverage; the combination yielded a total of 77% sequence coverage. Other examples of HCD,

UVPD, and NUVPD spectra are shown in **Figures 6.2** and **.** Both HCD and UVPD provide high sequence coverage for many peptides in the positive mode, and the sizes and charge states of peptides identified by each method are similar, as discussed later. The contrast in diagnostic quality of the MS/MS spectra is much more notable for peptides analyzed in the negative mode. In this case, HCD mainly results in neutral losses and internal fragment ions which are less useful for peptide sequencing, whereas UVPD produces characteristic *a/x* ions that afford high sequence coverage (**Figure 6.1)** Examination of the other MS/MS spectra in **Figures 6.2** and **6.3** provides more evidence of the complementary nature of HCD and UVPD, especially when both positive and negative modes are utilized. In general, the series of backbone cleavages for individual peptides varies for each activation method, and in fact the use of more than one activation mode may fill in the gaps.

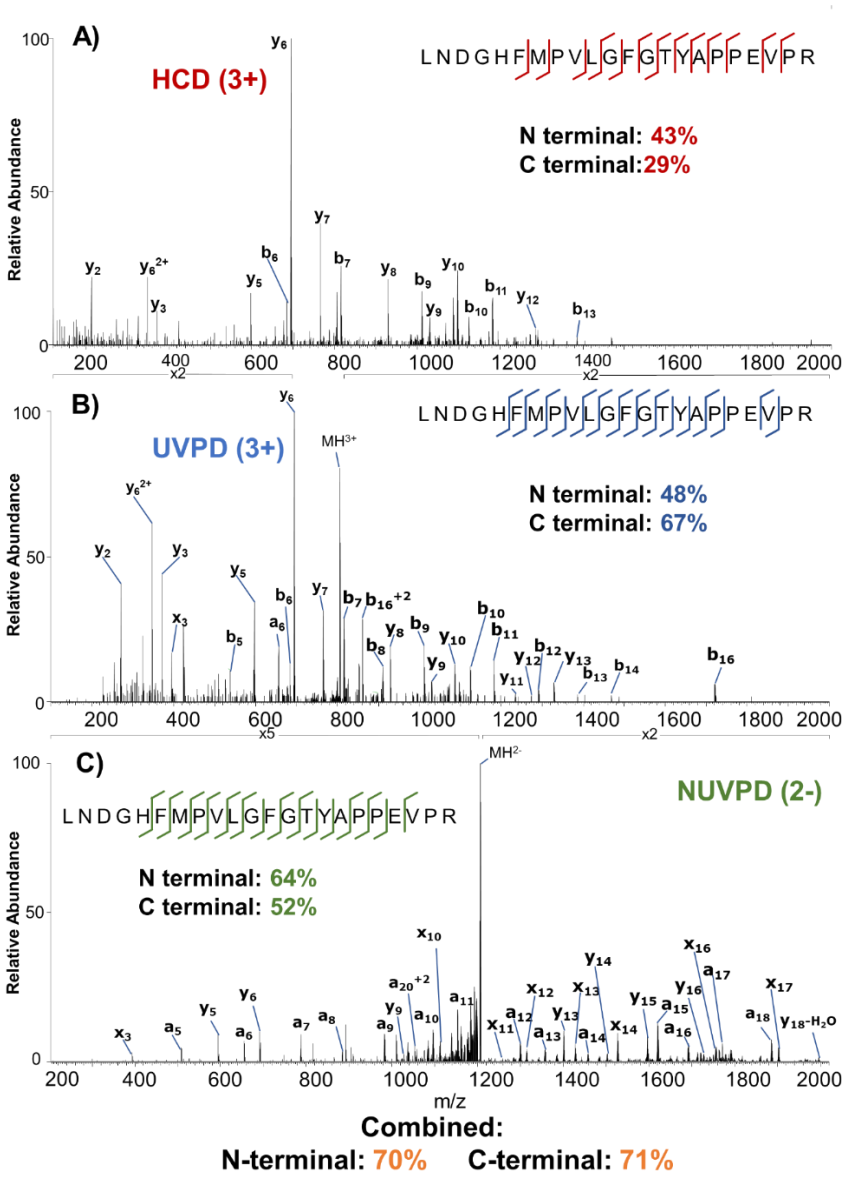


Figure 6.1 A) HCD (3+), B) UVPD (3+), and C) NUVPD (2-) spectra of tryptic peptide LNDGHFMPVLGFGTYAPPEVPR showing complementary information from each technique. Combining all methods yields 70% coverage by a and b type ions and 70% coverage by x and y type ions.

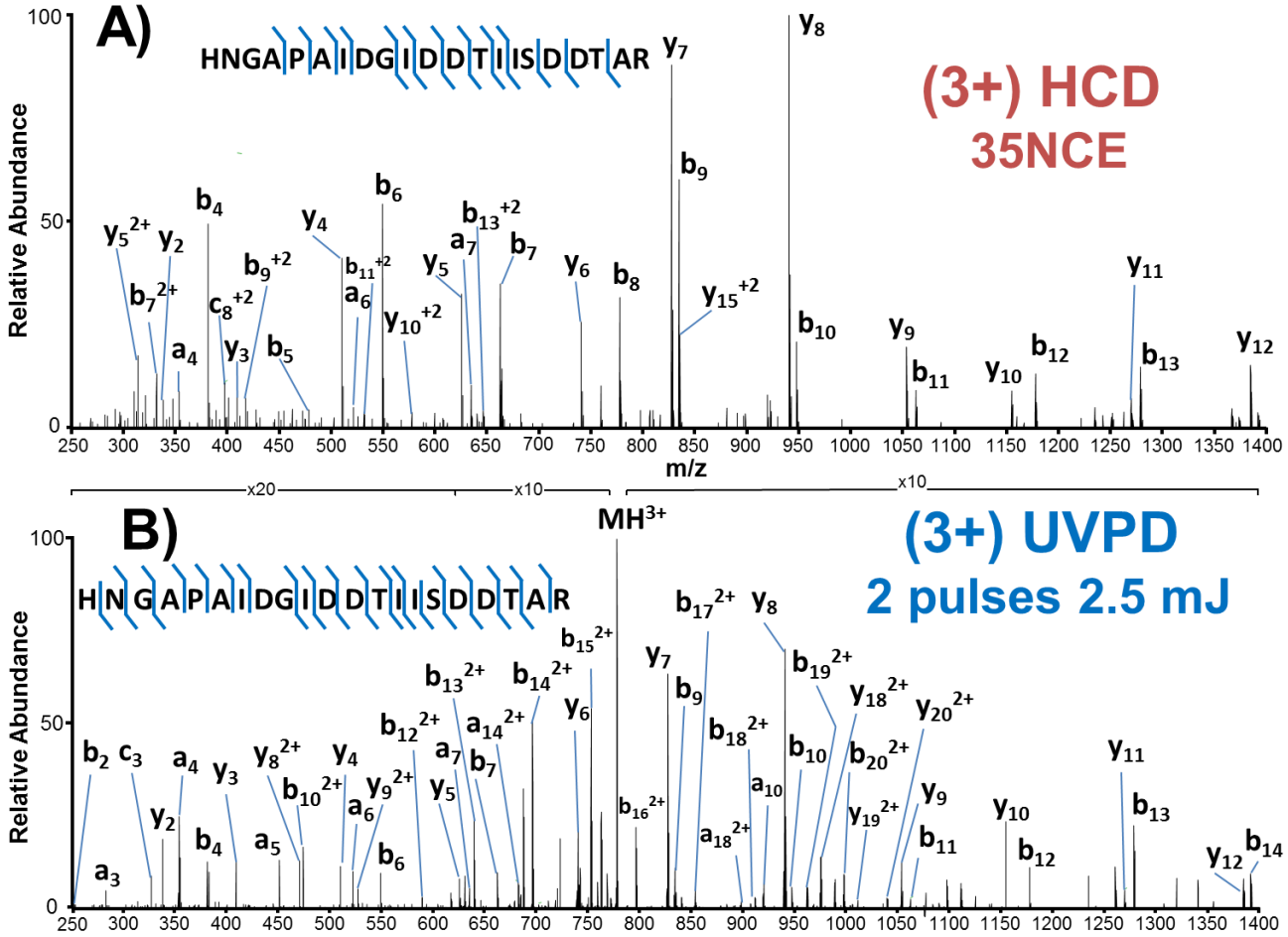


Figure 6.2 *MS/MS spectra generated by A) HCD (35 NCE) and B) UVPD (2 pulses, 2.5 mJ per pulse) for peptide HNGAPAIDGIDDTIISDDTAR (3+) from a tryptic digest of proteins extracted from human hepatocytes.*

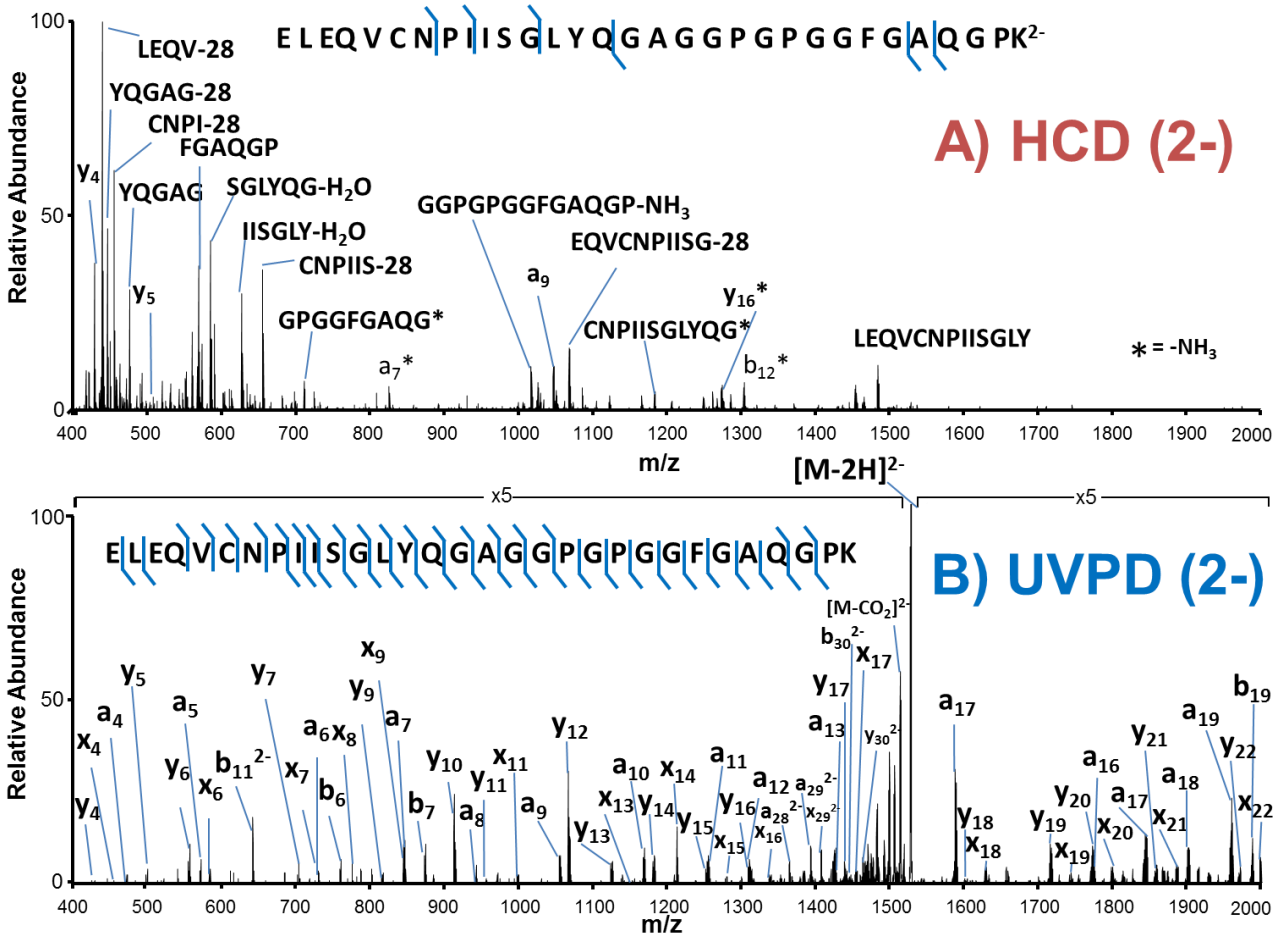


Figure 6.3 Negative polarity MS/MS spectra generated by A) HCD (35 NCE) and B) UVPD (2 pulses, 2.5 mJ per pulse) for peptide ELEQVCNPIISGLYQGAGGPGPGGFGAQGPK (2-) from a tryptic digest of proteins extracted from human hepatocytes.

The average number of peptides and proteins identified using Byonic versus SEQUEST for positive-mode UVPD runs and versus MassMatrix for NUVPD runs, based on triplicate runs is shown in **Figure 6.4**. Customization of Byonic's algorithm contributed only a small amount to its greater sensitivity for analysis of positive-mode UVPD spectra, as the differences between Byonic's UVPD and HCD fragment ion predictions for

unmodified peptides are not large: no immonium ions in UVPD mass spectra, more abundant a- and a-dot (radical) ions in UVPD mass spectra, and some modest differences in favored neutral losses (for example, carbon dioxide rather than water and ammonia loss). Customization of Byonic for negative-mode mass spectrometry was required to obtain any results at all for activation methods such as NETD or NUVPD. The difference in scoring an NETD spectrum and a NUVPD spectrum is not large: an NUVPD spectrum is scored much like an NETD spectrum of a precursor with one more negative charge, that is, Byonic scores an NUVPD spectrum of a precursor with charge 2- much as it scores an NETD spectrum of a precursor with charge 3-.[11] Construction of Venn diagrams allows evaluation of the overlap in the peptide and protein identifications at 1% FDR for each MS/MS mode, as illustrated in **Figure 6.5**.

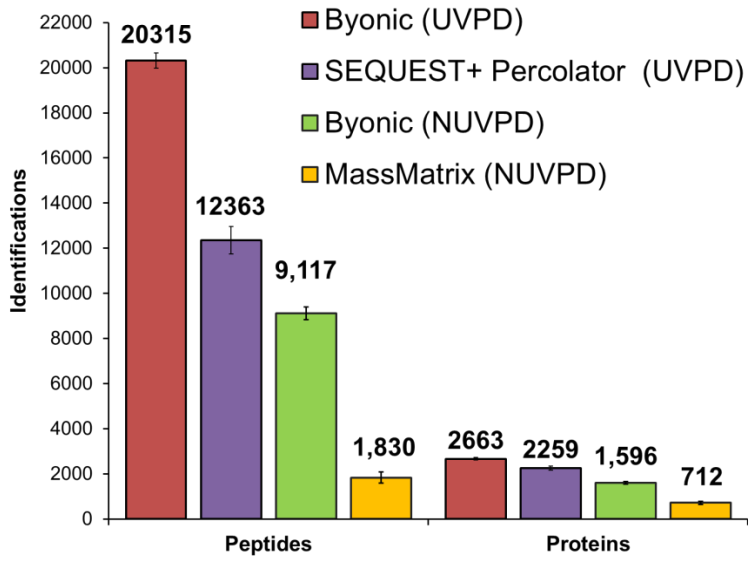


Figure 6.4 Number of tryptic peptide and protein identifications based on Byonic, MassMatrix and SEQUEST searches of UVPD and NUVPD spectra.

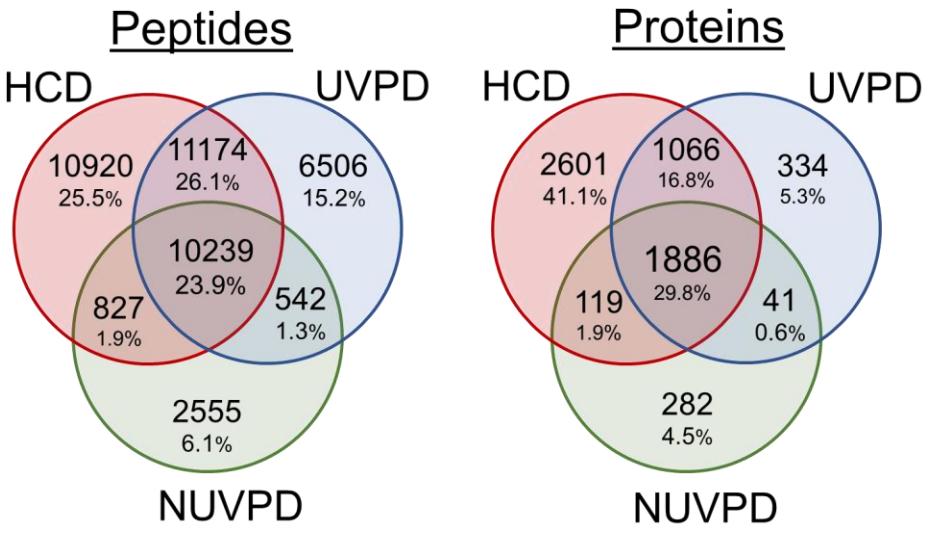


Figure 6.5 Venn diagrams illustrating the number of peptides (left) and proteins (right) identified based on HCD in the positive mode and UVPD in both the positive and negative modes.

By combining the results of three runs (and eliminating peptides identified more than once), HCD identified over 33,000 peptides, and UVPD identified over 28,000 peptides in the positive mode and over 14,000 peptides in the negative mode. Most significantly, over 6,500 peptides were uniquely identified by UVPD in the positive mode and over 2,500 peptides were uniquely identified by UVPD in the negative mode. These additional peptide identifications meant that UVPD (combining positive and negative modes) identified over 600 additional proteins not found by HCD alone. Overlap of unique peptides from the triplicate runs of each method accounted for around one third of all identified peptides, identification by two or more runs accounted for approximately two thirds of all identified peptides across methods (**Figure 6.6**). Overall the greatest number of peptides was identified based on the HCD method, likely owing to the greater signal to noise of HCD

mass spectra compared to UVPD mass spectra, thus resulting in informative, interpretable spectra from lower abundance precursor ions.

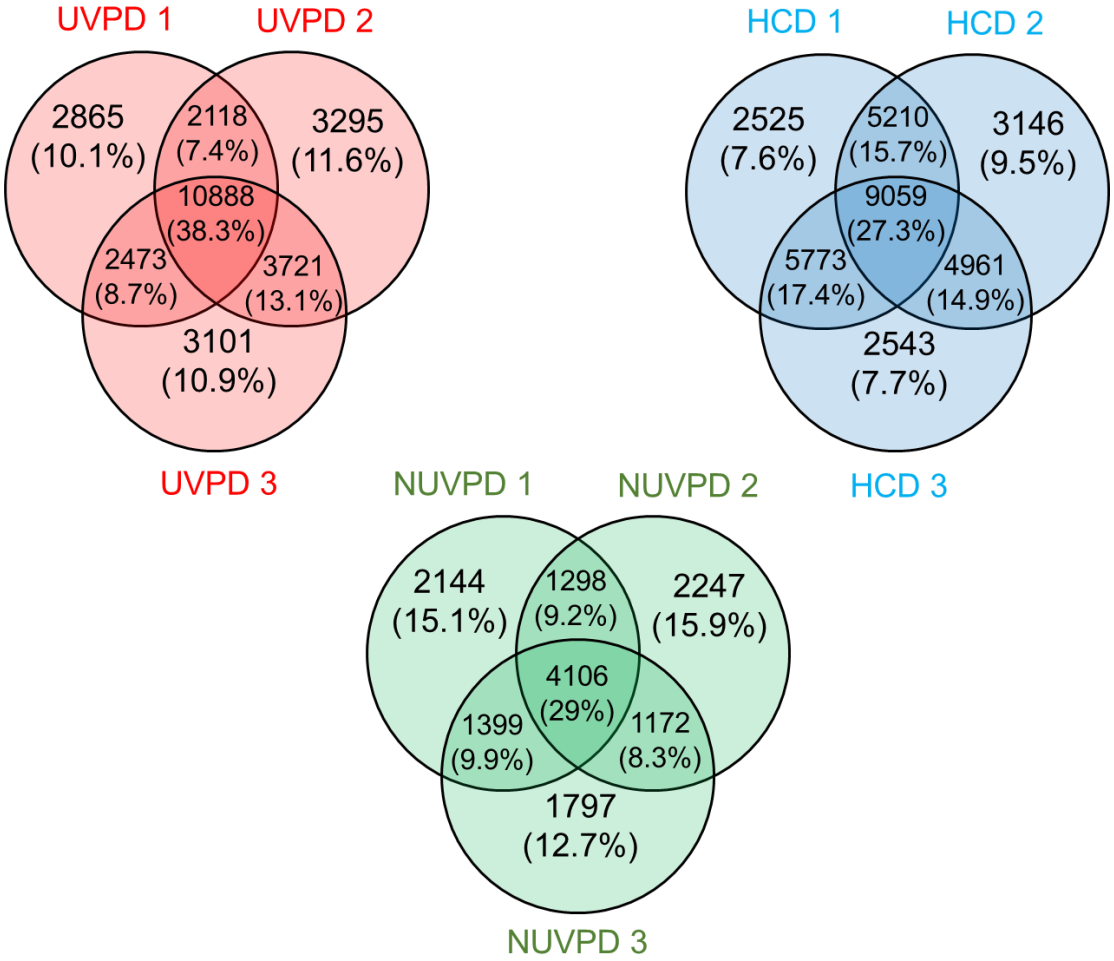


Figure 6.6 Overlap of peptides from HCD, UVPD and NUVPD replicates from tryptic digests of proteins extracted from human hepatocytes.

The high degree of overlapping protein and peptide identifications shows the complementary nature of UVPD, NUVPD and HCD, and at the same time the unique identifications for each activation mode extend the breadth of total proteins identified.

Furthermore, the overlap of peptides identified by SEQUEST+Percolator and Byonic from the UVPD dataset is shown in **Figure 6.7**.

Byonic identified over 98% of the all peptides identified by UVPD. SEQUEST+Percolator uniquely identified ~1.5% (457 peptides.)

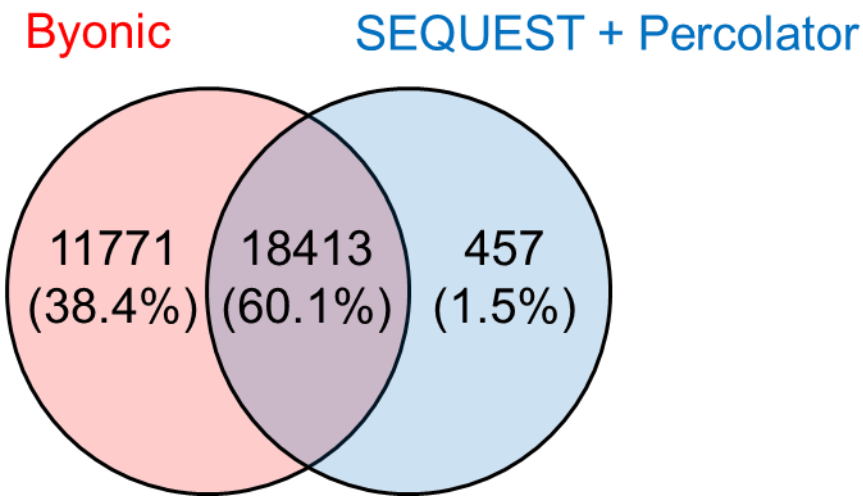


Figure 6.7 Overlap of UVPD peptides identified by Byonic and SEQUEST+Percolator from tryptic digests of proteins extracted from human hepatocytes.

Figure 6.8 shows a sequence map for protein P42330 (an aldo-keto reductase) which was identified based on peptide spectral matches from HCD, UVPD and NUVPD from the human liver lysate. Individually, no single activation method achieved over 52% sequence coverage. In fact, the regions identify by each method were largely orthogonal. For example, only 5% of the sequence overlapped for the peptides identified by HCD and NUVPD, thus illustrating the potential of NUVPD to characterize peptides that might not

be well covered by HCD alone. Combining the MS/MS spectra from all three activation methods resulted in total sequence coverage of 70% of the protein.

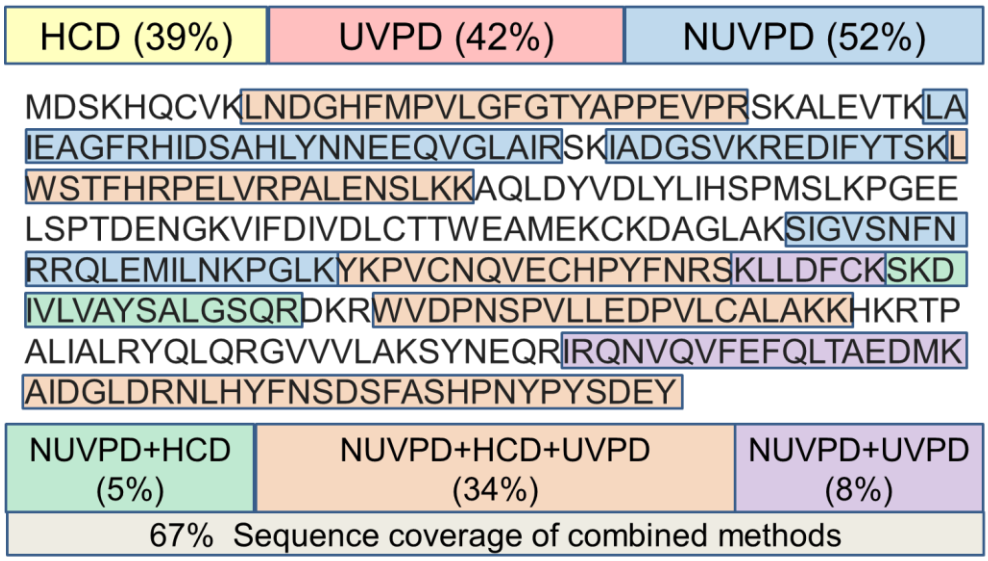


Figure 6.8 Sequence coverage of P42330 (36853 Da) resulting from HCD (yellow), UVPD (pink) and NUVPD (blue). Peptides sequenced by all three methods are highlighted in orange, those sequenced by UVPD and NUVPD in purple, and those sequenced by NUVPD and HCD are shown in green.

The distribution of peptide sizes and charge states for all peptide spectral matches is displayed in histogram format in **Figure 6.9**. The y-axis is normalized to the total population of unique peptides from each search. The population of unique peptide masses identified by Byonic for HCD and UVPD appear as skewed-right distributions with average peptide masses of 2008 Da and 2128 Da respectively. However, the distribution of NUVPD identified peptide masses was more symmetric with a larger spread and an average mass of 2095 Da. With respect to the charge states, the portions of 2+, 3+, and 4+ peptides identified by UVPD and HCD in the positive mode are similar, with nearly equal

percentages of doubly and triply protonated peptides. In the negative mode, there is a significant preference for identification of doubly deprotonated peptides by NUVPD. The shift to lower charge states for the peptides identified by NUVPD is reasonable given that the acidic mobile phase most commonly used for reversed phase liquid chromatography suppresses extensive deprotonation of peptides. The size distribution of NUVPD identified peptides is skewed towards larger peptides compared to UVPD and HCD. We speculate that the apparent shift to larger peptides for NUVPD may be due to two factors. One factor is the potential low efficiency of deprotonation of small tryptic peptides which is dominated by their terminal basic sites. Second, longer peptides often have greater hydrophobicities and elute later in the LC gradient. Ionization in the negative mode is less efficient in the early, highly aqueous portion of the gradient, and ionization efficiency increases as the organic content of the mobile phase increases. Thus, the apparent increase in size of peptides for NUVPD may be related to the improved electrospray conditions in the latter part of the gradient.

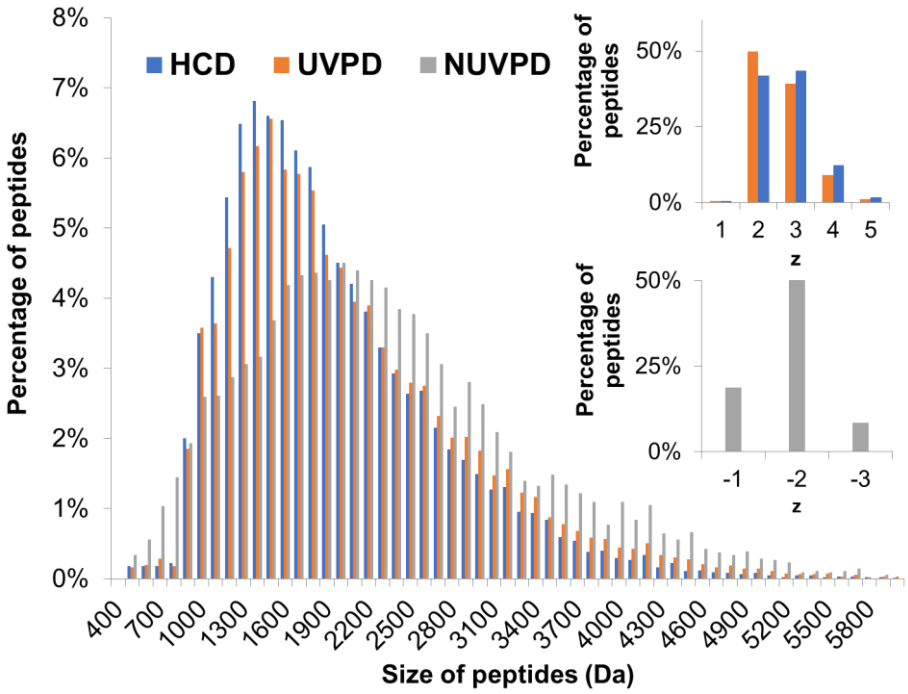


Figure 6.9 Histogram showing size of peptides identified by HCD, UVPD and NUVPD using Byonic. The inset histograms display the charge states of the identified peptides in the positive and negative modes.

The isoelectric point (pI) distribution of peptides identified by UVPD and NUVPD was plotted in **Figure 6.10**.

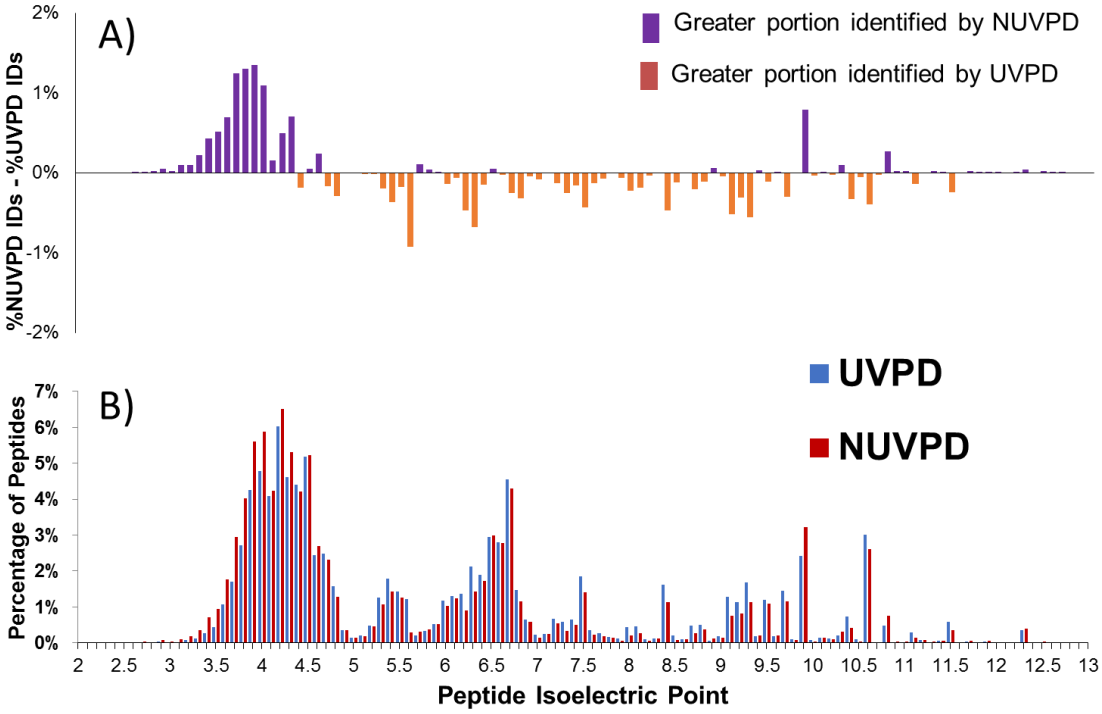


Figure 6.10 A) Portions of peptides identified by NUVPD versus UVPD from tryptic digests of proteins extracted from human hepatocytes. The peptides are ordered based on isoelectric point. B) Distribution of isoelectric point of peptides identified by UVPD and NUVPD.

A larger proportion of the identified peptides from NUVPD analysis populated the pI 3-5 range while the range of 5-11 was generally dominated by a larger proportion identified by UVPD suggested NUVPD is well suited to study particularly acidic peptides and proteins.

6.4.2 Comparison of Forward and Reverse PSMs Ranking Between Platforms

To further investigate the sensitivity/specificity tradeoff, receiver operator characteristic (ROC) curves were plotted for forward and decoy PSMs derived from

searches using Byonic, SEQUEST+Percolator and MassMatrix (the latter only for NUVPD). A ROC curve displays the number of forward PSMs (correct hits) as a function of the reverse PSMs (false hits). These plots display the ability of a method to discriminate between true and false PSMs at a given threshold (or number of false PSMs). An ideal method would generate an ROC curve which lies along the top left of the graph (immediate plateau) suggesting complete discrimination between decoy and forward PSMs; conversely a poor method would show an ROC curve lying along the line y=x, suggesting that for each forward PSM a reverse PSM occurs. For these comparisons, no cutoff was used for the false discovery rate (FDR). All PSMs were filtered to a single best PSM per spectrum (i.e. the best PSM from any given spectrum was considered unique to that spectrum, and any other PSMs to the same spectrum were not plotted). The PSMs were ranked by q-value for SEQUEST-Percolator data, Log Prob for Byonic and pp-tag for MassMatrix. The metrics q-value, Log Prob, and pp-tag are used to establish the quality of a match and also to establish FDR for a set of PSMs. The resulting ROC curves represent the number of PSMs matching to the decoy database that must be tolerated to achieve a given number of forward PSMs.

ROC curves for PSMs determined using the Byonic and SEQUEST+Percolator search algorithms based on the UVPD mass spectra of the tryptic human liver peptides acquired in the positive mode are shown in **Figure 6.11**. The ROC curves show the full sensitivity/ specificity tradeoff for both Byonic and SEQUEST + Percolator. For example, with Byonic, approximately 41,000 PSMs with only a few decoy PSMs (too few to be

observed in the plot), about 49,000 PSMs with 50 decoy PSMs, or over 52,000 PSMs with 1000 decoy PSMs, are obtained. For construction of this plot, all top-ranking PSMs from the search engines were retrieved, rather than allowing the software to threshold at 1% FDR. In this particular case, it is apparent that 1% FDR may be a worse choice than 0.1% FDR for Byonic, because 1% FDR increases sensitivity only slightly while degrading specificity significantly. The ROC curve from the Byonic data processing is shifted higher and to the left, suggesting greater specificity and sensitivity in assignment of PSMs, mainly due to generic scoring improvements such as rank-based intensities and *m/z*-error scoring, rather than SEQUEST's *m/z* binning. The SEQUEST+Percolator ROC curve plateaus at 46,533 forward PSMs, whereas the Byonic ROC curve plateaus at 52,658 forward PSMs, yielding a 13% improvement in sensitivity. At an estimated PSM level FDR of 1%, 44,179 forward PSMs clear the threshold based on the use of the SEQUEST+Percolator search method, whereas 51,859 PSMs clear the threshold based on the Byonic search. It is not exactly clear why MassMatrix gives much lower performance than Byonic, but manual inspection of high-scoring decoy matches shows that MassMatrix sometimes mistakes unfragmented precursor ion for a product ion, and can achieve a high score from a mix of ion types and hydrogen atom transfers rather than from a consistent set of ion types.

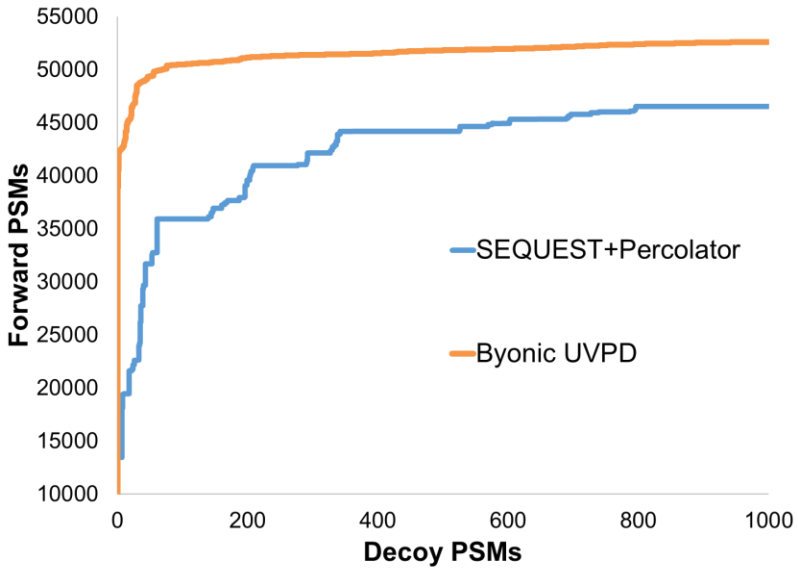


Figure 6.11 ROC plot for UVPD data searched by Byonic (Orange) and SEQUEST (blue).

Figure 6.12 shows the ROC curves for PSMs identified from the NUVPD data set assigned by MassMatrix and Byonic. The ROC curve based on the Byonic output plateaus significantly higher at 31,972 compared to MassMatrix at 12,395, a 157% improvement in sensitivity. Interestingly, the slope for both curves is similar, suggesting that they have similar specificity towards forward PSMs. Re-analysis of a HeLa/NUVPD dataset reported in Madsen *et al.*[13] using the optimized Byonic algorithm resulted in a nearly 70% increase in protein and over 85% increase in peptide identifications, confirming the practical gains realized by this upgraded NUVPD spectral search algorithm (additional details are provided in the supplemental material).

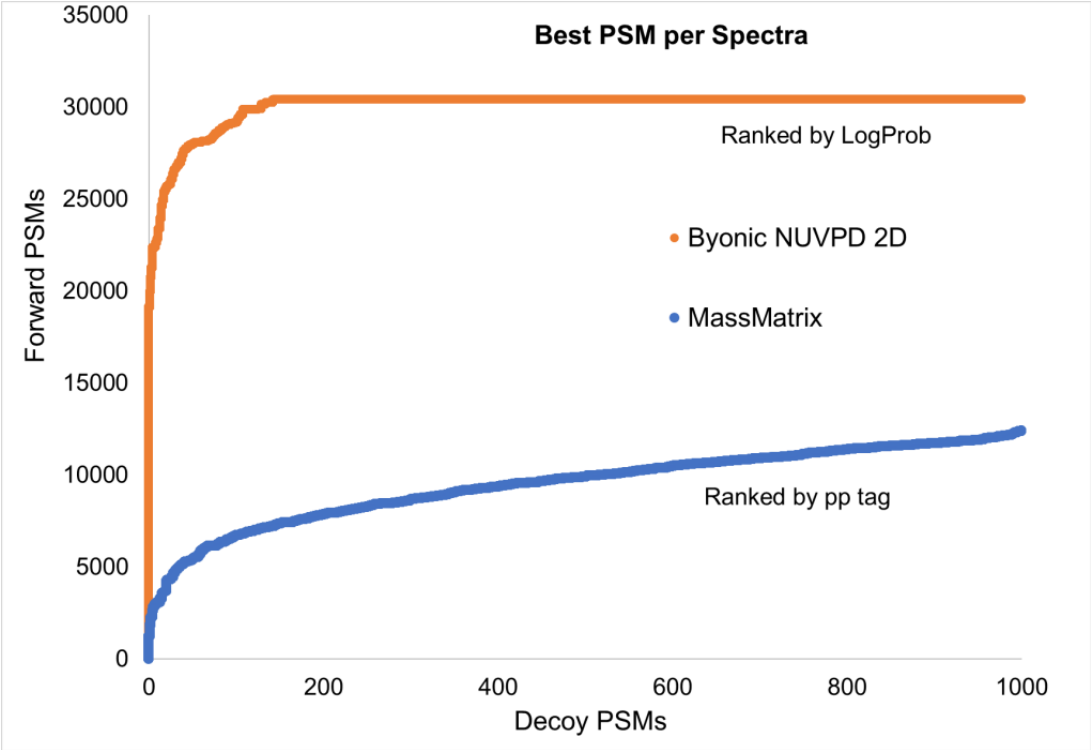


Figure 6.12. ROC plot for NUVPD data analyzed by MassMatrix (blue) and Byonic (orange) from tryptic digests of proteins extracted from human hepatocytes.

6.4.3 Protein Oblivious and Protein Aware

One of the unique features of the Byonic algorithm is its default use of a two-dimensional, protein aware false discovery rate (2D FDR) which bolsters peptide identifications by enhancing the scores of peptides from proteins which are confidently identified. We have compared the UVPD and NUVPD results with and without utilization of the 2D FDR feature. The use of 2D FDR did not impact protein identification for the UVPD dataset. Using a traditional protein oblivious 1% FDR (1D FDR or just FDR), on average 2660

proteins were identified, whereas 2663 proteins were identified when the 2D FDR (protein aware FDR) at 1% FDR was employed. Greater gains were observed for the number of unique peptide identification when using the 2D FDR method. A 1D search identified 17534 peptides on average, whereas a 2D search identified 20315 unique peptides for a gain of 14%. Compared to the corresponding SEQUEST+Percolator search which yielded 12363 peptide identifications, the 2D FDR method in Byonic yielded a 64% gain, 35% of which are attributed to unique peptides assigned based on the use of a 2D FDR. A comparable gain in peptide identifications was noted for NUVPD data. The number of protein IDs was unchanged between 1D and 2D results for the NUVPD searches. However, incorporation of the 2D FDR resulted in 13% more peptide identifications at 1% FDR. These results are reflected by ROC analysis of the 1D and 2D searches (**Figures 6.13** and **6.14**). The selectivity is nearly the same for both 1D and 2D ROC curves, but the onset of the plateau (sensitivity) is about 10% higher for the 2D analysis.

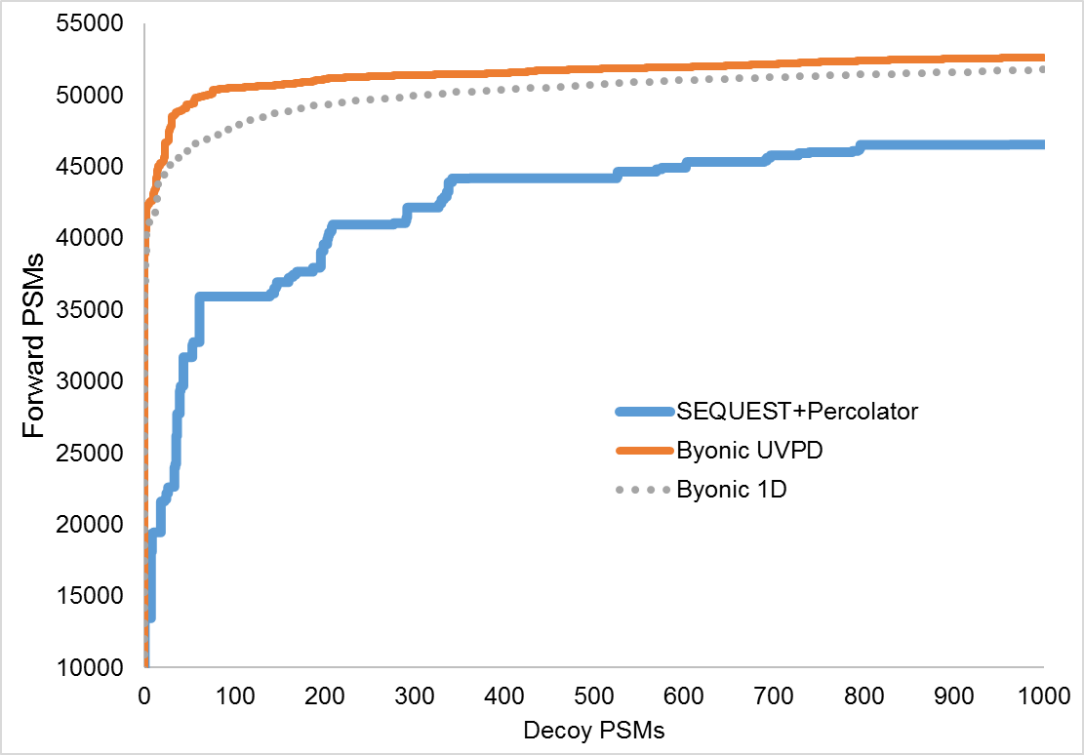


Figure 6.13 ROC plot comparing 1D and 2D FDR results for UVPD data from tryptic digests of proteins extracted from human hepatocytes. The graphs indicate that the majority of improvement is based on the search algorithm used, not inclusion of 2D FDR.

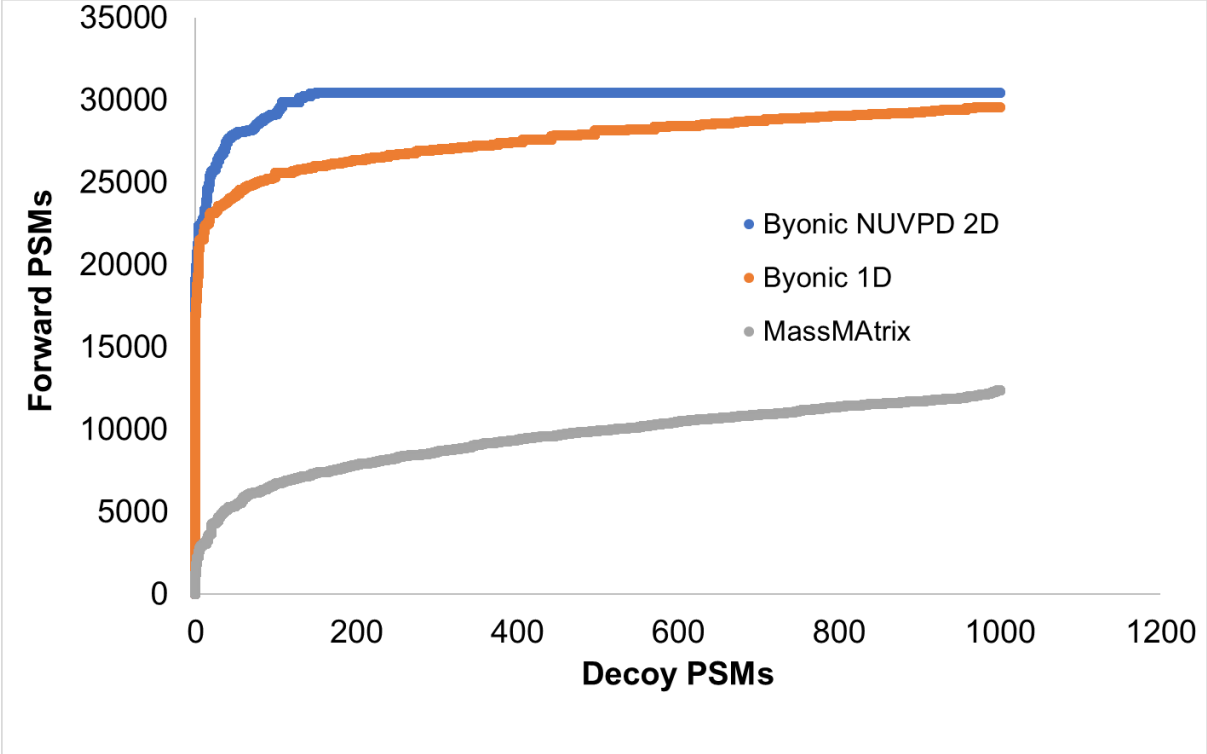


Figure 6.14 ROC plot comparing 1D and 2D FDR results for NUVPD data from tryptic digests of proteins extracted from human hepatocytes.The graphs indicate that the majority of the improvement is based on the search algorithm used, not inclusion of 2D FDR.

6.4.4 Increased protein sequence coverage

Given the increase in the number of peptides identified by Byonic compared to SEQUEST and MassMatrix, protein sequence coverage would be expected to change as well. **Figures 6.15**, **6.16** and **6.17** show the distribution of protein sequence coverages for HCD, UVPD, and NUVPD, respectively, reported by Byonic (blue), SEQUEST (red) and MassMatrix (green). The distribution of protein sequence coverages are shifted to higher values when using Byonic, likely owing to the use of the protein aware scoring to include

peptides that would have otherwise been filtered out conventionally.[30]

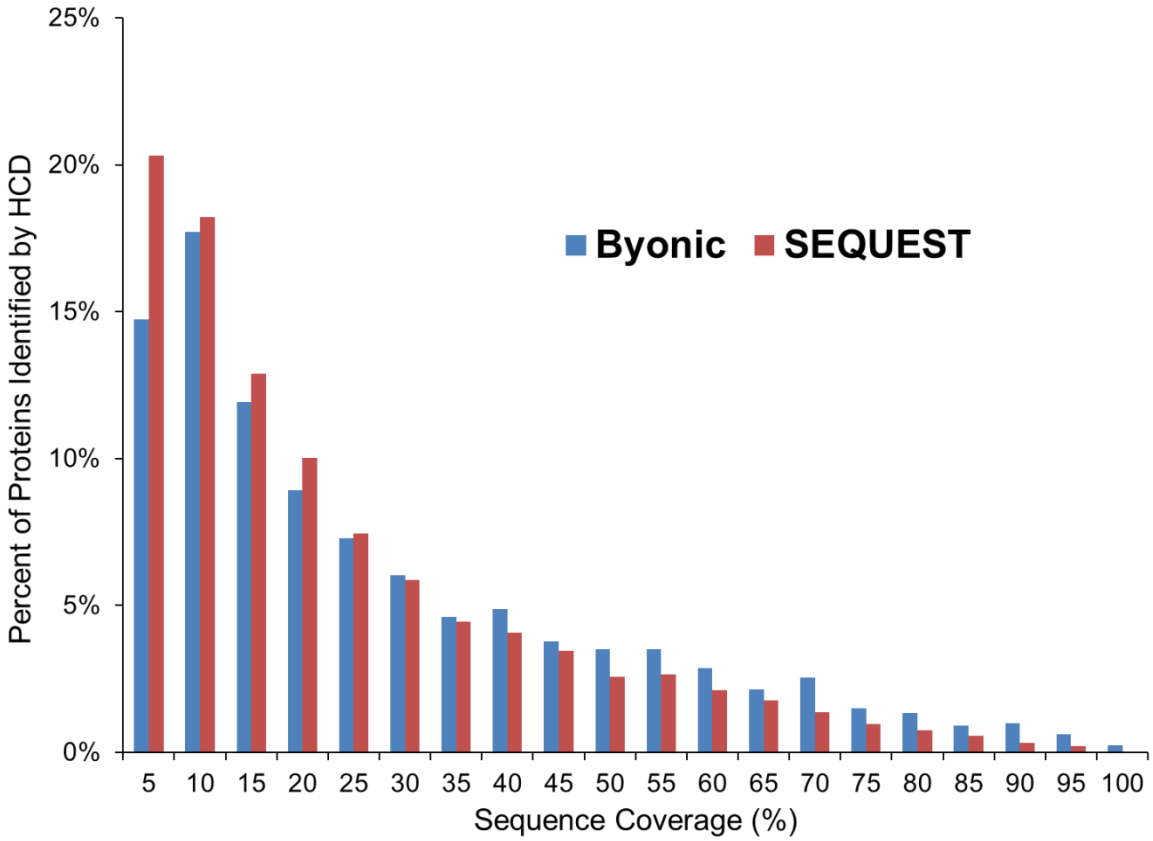


Fig 6.15 Distribution of sequence coverages obtained for proteins by HCD reported by Byonic (blue) and SEQUEST (red) from tryptic digests of proteins extracted from human hepatocytes.

.

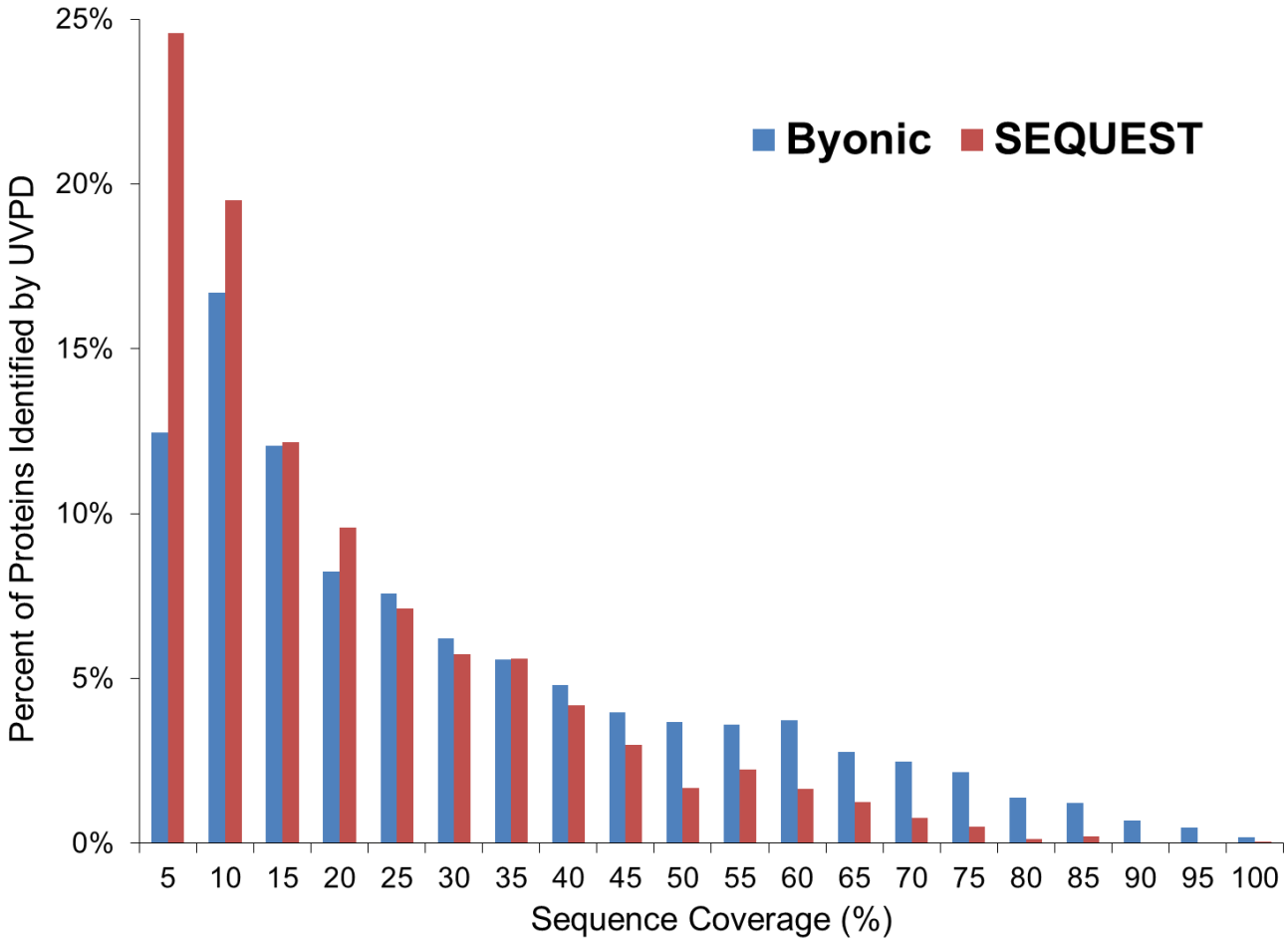


Fig 6.16 Distribution of sequence coverages obtained for proteins by UVPD reported by Byonic (blue) and SEQUEST (red) from tryptic digests of proteins extracted from human hepatocytes.

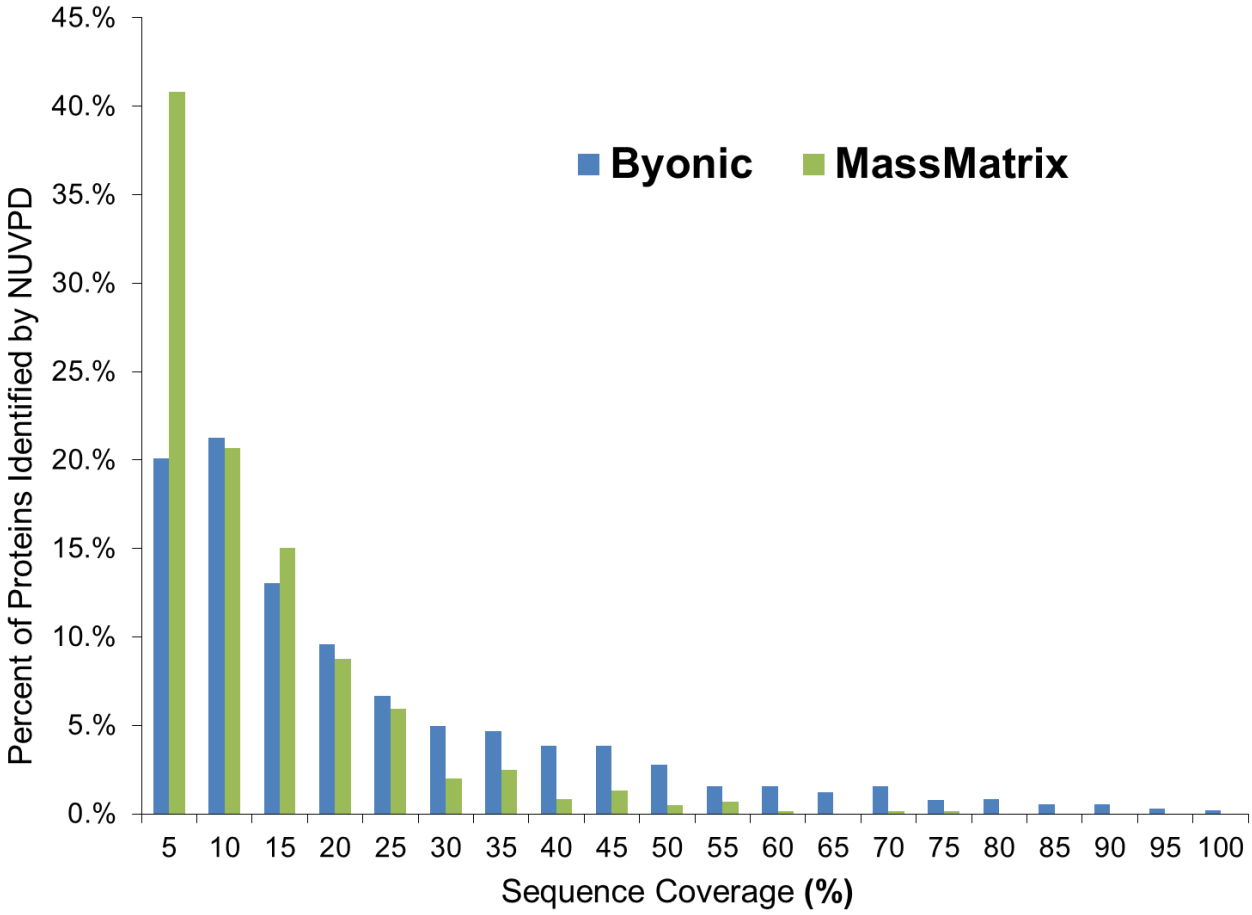


Fig 6.17 Distribution of sequence coverages obtained for proteins by NUVPD reported by Byonic (blue) and MassMatrix (green) from tryptic digests of proteins extracted from human hepatocytes.

The average protein sequence coverage reported for HCD was 25.1% for Byonic compared to 16.3% for SEQUEST in **Figure 6.15**. The average protein sequence coverage reported for UVPD was 27.3% for Byonic compared to 18.9% using SEQUEST in **Figure 6.16**. The average protein sequence coverage reported for NUVPD was 20.5% for Byonic compared to 13.6% for MassMatrix in **Figure 6.17**. Overall Byonic improved protein sequence by more than 6% across all activation methods. A comparison of protein

sequence coverage distributions for HCD, UVPD, and NUVPD data sets using the same search algorithm (Byonic) is shown in **Figure 6.18.**

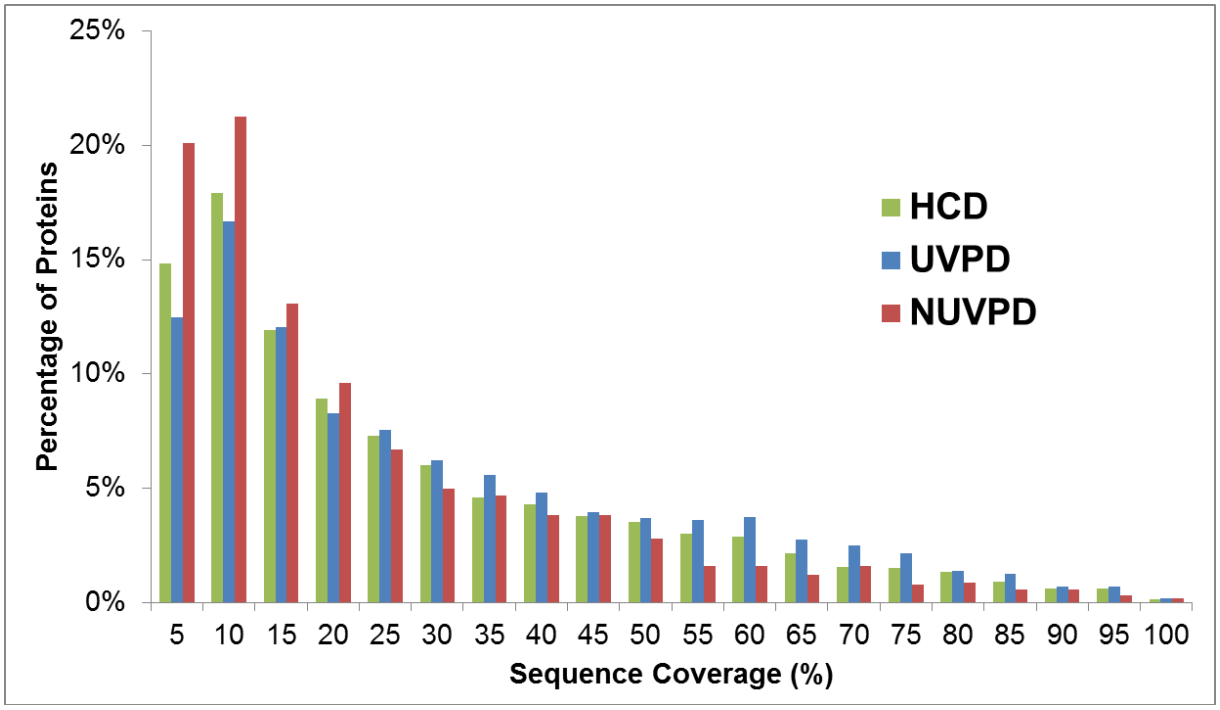


Fig 6.18 Distribution of sequence coverages obtained for HCD, UVPD, and NUVPD from tryptic digests of proteins extracted from human hepatocytes.

6.4.5 Identification of post-translational modifications: phosphopeptides

It has been previously shown that UVPD affords a special ability to characterize post-translational modifications (PTMs) owing to the fact that these modifications are not labile upon photoactivation, and the high sequence coverage afforded by UVPD allows confident pinpointing of the sites of PTMs.[33,34,39–42] An example of one phosphopeptide identified in the liver cell lysate using all three MS/MS strategies is shown in **Figure**

6.19. All three methods yielded good sequence coverage of the peptide: 92% coverage for HCD (3+), 83% coverage for UVPD (3+), and 100% coverage for NUVPD (2-). Despite similar sequence coverages, the retention of the phosphate group varied significantly among the three activation modes. As shown in **Figure 6.19,** the dominant products generated by HCD did not retain the phosphate group, an outcome particularly notable for the *y* ion series. UVPD of the same 2+ peptide showed only moderate phosphate loss. In general, UVPD generated a nearly 50/50 distribution of products that retained or lost the phosphate group. Further optimization of the laser energy and pulse number could improve phosphate retention. UVPD of the doubly deprotonated peptide generated a clean series of fragments retaining the phosphate moiety, displaying less than 5% phosphate loss. No enrichment was performed on these samples nor were they treated during growth or lysis for retention of PTMs, and a more targeted approach would be essential in order to maximize the identification of phosphopeptides and other modified peptides in complex cell lysates akin to the ones analyzed in the present study. However, this initial result recapitulates the promising combined strategy of HCD, UVPD, and NUVPD to enhance sequence coverages and localization of PTMs.

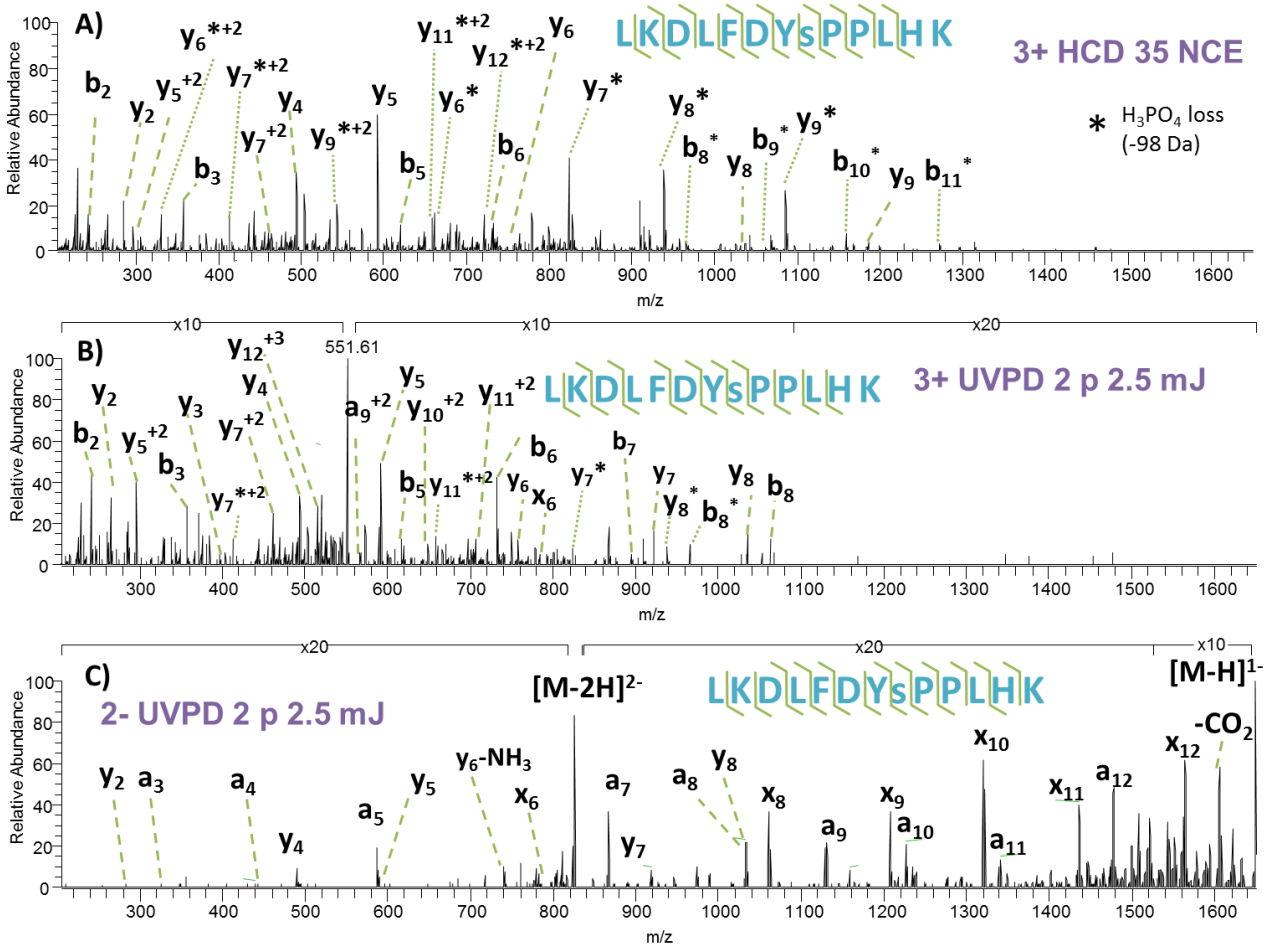


Figure 6.19 A) HCD, B) UVPD, and C) NUVPD of phosphorylated LKDLFDYSPPLHK from a tryptic digest of proteins extracted from human hepatocytes, showing varying degrees of phosphate retention.

6.5 CONCLUSIONS

Analyses of tens of thousands of HCD, UVPD, and NUVPD spectra revealed the complementary nature of traditional collisional activation (HCD) and novel photon-based activation (UVPD and NUVPD) methods. Combining these methods increased the total number of proteins identified by 12% and peptides identified by 29%. There are numerous instances for which a single method alone sequenced a limited portion of a protein, whereas

combining the results of multiple activation methods afforded nearly complete coverage. Furthermore, traditional collisional activation does not provide adequate fragmentation for confident peptide assignments in the negative polarity mode, resulting in limited identification of the most acidic portions of the proteome. A modified Byonic platform optimized for analysis of UVPD spectra returned a 64% improvement in the number of identified peptides relative to the more commonly employed database search algorithm (SEQUEST) and 18% more identified proteins compared to other available database search programs.

6.6 REFERENCES

(1) Bensimon, A.; Heck, A. J. R.; Aebersold, R. Mass Spectrometry–Based Proteomics and Network Biology. *Annu. Rev. Biochem.* **2012**, *81* (1), 379–405.

(2) Cox, J.; Mann, M. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. *Annu. Rev. Biochem.* **2011**, *80* (1), 273–299.

(3) Mann, M.; Kulak, N. A.; Nagaraj, N.; Cox, J. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. *Mol. Cell* **2013**, *49* (4), 583–590.

(4) Yates, J. R. The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. *J. Am. Chem. Soc.* **2013**, *135* (5), 1629–1640.

(5) Sharma, K.; D'Souza, R. J.; Tyanova, S.; Schaab, C.; Wisniewski, J.; Cox, J.; Mann, M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Rep.* **2014**, *8* (5), 1583–1594.

(6) Villén, J.; Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* **2008**, *3* (10), 1630–1638.

(7) Carlson, S. M.; Moore, K. E.; Green, E. M.; Martin, G. M.; Gozani, O. Proteome-wide enrichment of proteins modified by lysine methylation. *Nat. Protoc.* **2013**, *9* (1), 37–50.

(8) Li, Y.; Silva, J. C.; Skinner, M. E.; Lombard, D. B. Mass Spectrometry-Based Detection of Protein Acetylation. In *Sirtuins*; Hirschey, M. D., Ed.; Humana Press: Totowa, NJ, 2013; Vol. 1077, pp 81–104.

(9) Zhang, Y.; Song, L.; Liang, W.; Mu, P.; Wang, S.; Lin, Q. Comprehensive profiling of lysine acetylproteome analysis reveals diverse functions of lysine acetylation in common wheat. **2016**, *6*, 1–10.

(10) 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.; et al. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal. Chem.* **2013**, *85* (24), 11710–11714.
(11) Bern, M. W.; Kil, Y. J. Two-dimensional target decoy strategy for shotgun proteomics. *J. Proteome Res.* **2011**, *10* (12), 5296–5301.
(12) Riley, N. M.; Bern, M.; Westphall, M. S.; Coon, J. J. Full-Featured Search Algorithm for Negative Electron-Transfer Dissociation. *J. Proteome Res.* **2016**, *15* (8), 2768–2776.
(13) 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. High-throughput Database Search and Large-scale Negative Polarity Liquid Chromatography-Tandem Mass Spectrometry with Ultraviolet Photodissociation for Complex Proteomic Samples. *Mol. Cell. Proteomics* **2013**, *12* (9), 2604–2614.
(14) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: An open-source MS/MS sequence database search tool. *PROTEOMICS* **2013**, *13* (1), 22–24.
(15) Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra. *J. Proteome Res.* **2014**, *13* (8), 3679–3684.
(16) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10* (4), 1794–1805.
(17) Chi, H.; He, K.; Yang, B.; Chen, Z.; Sun, R.-X.; Fan, S.-B.; Zhang, K.; Liu, C.; Yuan, Z.-F.; Wang, Q.-H.; et al. pFind–Alioth: A novel unrestricted database search algorithm to improve the interpretation of high-resolution MS/MS data. *J. Proteomics* **2015**, *125*, 89–97.
(18) Yamashita, M.; Fenn, J. B. Negative ion production with the electrospray ion source. *J. Phys. Chem.* **1984**, *88* (20), 4671–4675.
(19) Straub, R. F.; Voyksner, R. D. Negative ion formation in electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1993**, *4* (7), 578–587.
(20) Garcia, M. The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography–electrospray mass spectrometry. *J. Chromatogr. B* **2005**, *825* (2), 111–123.
(21) McCalley, D. V. Effect of buffer on peak shape of peptides in reversed-phase high performance liquid chromatography. *J. Chromatogr. A* **2004**, *1038* (1–2), 77–84.
(22) Busman, M.; Schey, K. L.; Oatis, J. E.; Knapp, D. R. Identification of phosphorylation sites in phosphopeptides by positive and negative mode electrospray ionization-tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1996**, *7* (3), 243–249.
(23) Bokatzian-Johnson, S. S.; Stover, M. L.; Dixon, D. A.; Cassady, C. J. A Comparison of the Effects of Amide and Acid Groups at the C-Terminus on the

Collision-Induced Dissociation of Deprotonated Peptides. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (9), 1544–1557.
(24) Bowie, J. H.; Brinkworth, C. S.; Dua, S. Collision-induced fragmentations of the (M-H)- parent anions of underivatized peptides: An aid to structure determination and some unusual negative ion cleavages. *Mass Spectrom. Rev.* **2002**, *21* (2), 87–107.
(25) Tran, T. T. N.; Brinkworth, C. S.; Bowie, J. H. The identification of disulfides in ricin D using proteolytic cleavage followed by negative-ion nano-electrospray ionization mass spectrometry of the peptide fragments: Identification of disulfides in ricin by negative-ion MS. *Rapid Commun. Mass Spectrom.* **2015**, *29* (2), 182–190.
(26) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-nm photodissociation of singly and multiply charged peptide anions for acidic proteome characterization. *PROTEOMICS* **2011**, *11* (7), 1329–1334.
(27) Shaw, J. B.; Madsen, J. A.; Xu, H.; Brodbelt, J. S. Systematic Comparison of Ultraviolet Photodissociation and Electron Transfer Dissociation for Peptide Anion Characterization. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (10), 1707–1715.
(28) Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. Electron transfer dissociation of peptide anions. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (6), 880–882.
(29) Huzarska, M.; Ugalde, I.; Kaplan, D. A.; Hartmer, R.; Easterling, M. L.; Polfer, N. C. Negative Electron Transfer Dissociation of Deprotonated Phosphopeptide Anions: Choice of Radical Cation Reagent and Competition between Electron and Proton Transfer. *Anal. Chem.* **2010**, *82* (7), 2873–2878.
(30) 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. Analysis of the Acidic Proteome with Negative Electron-Transfer Dissociation Mass Spectrometry. *Anal. Chem.* **2012**, *84* (6), 2875–2882.
(31) Shaw, J. B.; Kaplan, D. A.; Brodbelt, J. S. Activated Ion Negative Electron Transfer Dissociation of Multiply Charged Peptide Anions. *Anal. Chem.* **2013**, *85* (9), 4721–4728.
(32) Riley, N. M.; Rush, M. J. P.; Rose, C. M.; Richards, A. L.; Kwiecien, N. W.; Bailey, D. J.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. The Negative Mode Proteome with Activated Ion Negative Electron Transfer Dissociation (AI-NETD). *Mol. Cell. Proteomics MCP* **2015**, *14* (10), 2644–2660.
(33) Robinson, M. R.; Taliaferro, J. M.; Dalby, K. N.; Brodbelt, J. S. 193 nm Ultraviolet Photodissociation Mass Spectrometry for Phosphopeptide Characterization in the Positive and Negative Ion Modes. *J. Proteome Res.* **2016**, *15* (8), 2739–2748.
(34) Greer, S. M.; Cannon, J. R.; Brodbelt, J. S. Improvement of Shotgun Proteomics in the Negative Mode by Carbamylation of Peptides and Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2014**, *86* (24), 12285–12290.

(35) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal. Chem.* **2015**, *87* (14), 7109–7116.
(36) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2016**, *88* (1), 1044–1051.
(37) Chalkley, R. J.; Bandeira, N.; Chambers, M. C.; Clauser, K. R.; Cottrell, J. S.; Deutsch, E. W.; Kapp, E. A.; Lam, H. H. N.; McDonald, W. H.; Neubert, T. A.; et al. Proteome Informatics Research Group (iPRG)_2012: A Study on Detecting Modified Peptides in a Complex Mixture. *Mol. Cell. Proteomics MCP* **2014**, *13* (1), 360–371.
(38) Bern, M.; Cai, Y.; Goldberg, D. Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal. Chem.* **2007**, *79* (4), 1393–1400.
(39) Madsen, J. A.; Ko, B. J.; Xu, H.; Iwashkiw, J. A.; Robotham, S. A.; Shaw, J. B.; Feldman, M. F.; Brodbelt, J. S. Concurrent Automated Sequencing of the Glycan and Peptide Portions of *O* -Linked Glycopeptide Anions by Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85* (19), 9253–9261.
(40) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. *Anal. Chem.* **2015**, *87* (8), 4152–4158.
(41) Robinson, M. R.; Moore, K. L.; Brodbelt, J. S. Direct Identification of Tyrosine Sulfation by using Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (8), 1461–1471.
(42) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. *Anal. Chem.* **2016**, *88* (4), 2303–2310.

Chapter 7

Extensive Characterization of Heavily Modified Histone tails by 193 nm Ultraviolet Photodissociation

7.1 Overview

The characterization of proteins bearing several post-translational modifications (PTMs) remains challenging for the traditional mass spectrometry (MS) based proteomics workflow, i.e. short peptide analysis (4-20 residues) via bottom-up MS. This is due to the lability of PTMs upon collisional activation dissociation and the difficulty of mapping combinatorial patterns of PTMs. There are also hurdles associated with top-down MS approaches related to limited data analysis options for heavily modified proteoforms and less efficient separation methods for intact modified proteins, together leading to poor quantification. These shortcomings have accelerated interest in middle-down MS methods that focus on analysis of large peptides generated by certain enzymes (e.g. GluC, LysC, AspN), limited digestion, or chemical cleavage (e.g., formic acid). Mapping multiple PTMs simultaneously requires the ability to obtain extensive sequence coverage to allow confident localization of the modifications. Ultraviolet photodissociation (UVPD) has been shown to generate high sequence coverage for peptides and proteins compared to traditional MS/MS methods. Histones are an ideal system to test the ability of UVPD to characterize multiple modifications as the combinations of PTMs are the underpinning of the biological significance of histones and at the same time create an imposing challenge for characterization. The present study focuses on determining the feasibility of UVPD for

identification and localization of PTMs on histones by UVPD and comparison to a popular alternative, electron-transfer dissociation (ETD), via a high throughput middle-down LC-MS/MS strategy. In total, over 300 modified forms were identified, and the distributions of PTMs were quantified between these two methods. Results showed that both UVPD and ETD results efficiently quantified significant differences of PTM abundance when comparing control HeLa cell culture and treatment with deacetylase inhibitors. Additional ion types generated by UVPD proved essential for extensive characterization of the most heavily modified forms (> 5 PTMs).

7.2 INTRODUCTION

Post-translational modifications (PTMs) of proteins are implicated in an ever-expanding number of crucial biological processes, ranging from gene expression to tumorigenesis to cell death.[1–4] Not only the type of modification but also the number, sites, and pattern, collectively known as combinatorial modifications, create an elaborate diversity of protein structure and function. Even minor variations in the distribution of PTMs can significantly influence the outcomes of myriad of cellular processes. Key examples of landmark proteins in which combinatorial modifications have been found to be essential for triggering and regulating downstream effects include p53,[5–7] histones (chromatin structural units),[8,9] and the C-terminal domain of RNA polymerase.[10,11] Hundreds of types of PTMs are now recognized to contribute to the coding of protein function,[12,13] and the interplay between different PTMs has created an enormous need for methodologies that can characterize PTMs

and allow the cross-talk arising from combinatorial modifications to be deciphered. Significant effort has focused on improving analytical tools, particularly advanced mass spectrometry (MS), to characterize PTMs.[14,15] Routine characterization of proteins which are modified at multiple residues remains a challenge,[16–18] owing to the dynamic nature of PTMs, their low abundances and the large variation in stoichiometries.[19] Other techniques are available to identify PTMs, such as modification-specific antibodies,[20] and gel electrophoresis (limited applicability).[21] However, these methods cannot identify unknown modifications nor co-existing combinatorial PTMs.

MS offers special attributes in the realm of high-throughput PTM analysis.[22–24] In recent years, advances in MS analysis of PTMs have facilitated identification of even greater numbers and types of PTMs in a single experiment.[25–27] Chief among the innovations enabling PTM analysis by MS are the development of selective enrichment methods[25,28,29] and the introduction of new ion activation methods which reduce the loss of labile PTMs[16,30–33] or enhance localization of modifications via greater sequence coverage.[34] For example, collisional activation methods (CID, HCD) result in preferential cleavage of labile modifications such as phosphorylation and sulfation, thus impeding the ability to localize the sites of these modifications.[30,35] Implementation of electron-based activation methods, such as electron captured dissociation (ECD)[37] and electron transfer dissociation (ETD)[38], and 193 nm ultraviolet photodissociation (UVPD) alleviates the loss of these labile modifications, enabling their identification and localization for high throughput applications.[36–38] Another new option for MS/MS analysis is EThcD which is a hybrid method combining higher energy collisional

dissociation (HCD) and ETD, to more efficiently generate two series of diagnostic fragment ions (*b/y, c/z* type ions).[39] Moreover, activated ion electron dissociation (AI-ETD) is a hybrid method which uses concurrent infrared laser irradiation during ETD to counteract the charge state dependency inherent to electron activation and overcome "ET-no-D" events which limit sequence coverage.[40] UVPD is an alternative to electron- or collisional-based or hybrid methods. UVPD stands out among these methods in that it uses photons for ion excitation and affords retention of labile modifications.[30,36,41] UVPD typically generates a larger array of ion types than other activation methods, including *a*, *a+1*, *b*, *c*, *x*, *x+1*, *y*, *y-1*, and *z* type ions.

The development of alternative ion activation methods that allow retention and thus localization of PTMs has been particularly important for the characterization of some of the most heavily modified classes of protein, such as histones.[19,42] Histones act as a structural scaffold for packaging of DNA as it wraps into chromatin. Importantly, the N-terminal and C-terminal regions of histones extend beyond the coils of DNA and act as key coding substrates for modification in a way that modulates DNA-protein interactions. Acetylated N-terminal tails promote loose histone-DNA association that guide interactions with translation factors, whereas N-terminal methylation hinders DNA translation.[43] The extent of modifications along the N-terminal histone tails (e.g. first 50 residues) can be quite complex with many possible co-existing modification patterns with different biological ramifications. This complex relationship between modification states and biological outcomes has been termed the "histone code."[8]

When mapping histone PTMs by tandem mass spectrometry (MS/MS), there are two notable hurdles that limit the success of traditional bottom-up MS/MS methods. Deciphering the contextual network of modifications among heterogeneous mixtures of histones is virtually insurmountable based on analysis of small proteolytic peptides. In addition, the prevalence of lysine and arginine residues of the N-terminal tails result in small tryptic peptides, often ones that are poorly separated by reversed phase chromatographic methods and for which combinatorial patterns are lost owing to the short lengths of the peptides.[42] In an effort to directly characterize combinatorial PTMs, several groups have developed methods to characterize large peptides (middle-down) or intact proteins (top-down), enabling observation of all PTMs in a key region of a protein or in the entire protein sequence.[42,44–46] Top-down workflows analyze even heavily modified proteins as intact species, thus offering an unsurpassed opportunity for mapping all PTMs. However, top-down analysis has greater technical challenges with respect to effective separation of intact proteins, ion activation methods that perform adequately for large ions, and bioinformatics needed to interpret very complicated spectra of proteins.[44] The bioinformatics issue becomes particularly challenging when analyzing heavily modified proteins, such as histones, owing to the exponential increase in number of potential modification sites of intact proteins.[47] Top-down analysis of histones has been extensively developed and successfully evaluated by several groups, including in a high throughput format for complex mixtures of histones.[48,49] The most expansive top-down characterization of histones has been achieved by isolating and analyzing individual families of histones.[49] Despite the advantages of direct analysis of intact proteins, the need for excellent

ion activation methods and the limited software for assignment and confident scoring of proteoforms with multiple PTMs have impeded widespread successful implementation of large scale LC-MS analyses of intact proteins.

Middle-down strategies offer an intermediate compromise between top-down and bottom-up methodologies, typically achieved via enzymatic or chemical procedures which limit the extent of protein digestion, thus producing peptides that are typically much larger than those generated in conventional bottom-up workflows.[19,45] MS/MS analysis of middle-down sized peptides has the added advantage of having fewer fragmentation channels in which to distribute ion current compared to analysis of intact proteins, thus affording better S/N in the resulting spectra.[40] Furthermore, database searches of middle-down sized peptides is accommodated by the multitude of robust informatics platforms currently available for analysis of bottom-up sized peptides.[18,42,46] A majority of PTMs found on histones exist on the first 50-60 amino acids, and this N-terminal stretch may be covered by a single long peptide generated by using GluC or AspN to cleave the histones. The canonical analysis of histone PTMs by tandem mass spectrometry (MS/MS) adopts chemical derivatization of lysine residues to limit trypsin proteolytic cleavage, necessary owing to the large content of lysine and arginine residues on histone sequences. One of the most used approaches was developed by Hunt *et al*., who introduced the derivatization of lysine and N-terminal amines with propionic anhydride.[50] Derivatization blocked the ε-amino groups of unmodified and monomethyl lysine residues, meaning that conventional trypsin proteolysis occurred only C-terminal to arginine residues instead of at both arginine and lysine residues, ultimately resulting

in longer peptides. Moreover, N-terminal derivatization increased peptide hydrophobicity and thus retention on reversed phase media, affording better chromatographic separation.[51] Recently a selective protease, neprosin, has also been successfully utilized for middle-down analysis of histones.[52] Neprosin cleaves C-terminal to proline providing 3-4 kDa size peptides, thus offering a promising option for histone characterization.[52]

Despite the isolation of the most heavily modified region of the histone, the issue of separating hundreds of modified species has remained challenging. The Garcia lab pioneered the use of a mixed bed weak cation exchange hydrophilic interaction liquid chromatography (WCX-HILIC) resin to enable separation of N-terminal histone peptides based on number of modifications.[17] Significant strides have been made to further develop this method for robust usage, including the introduction of software for filtering results obtained from canonical database searches and quantification.[53] For example, false positive modification localization is a common problem encountered when analyzing MS/MS spectra of heavily modified peptides. The problem arises when one of two modifiable sites is modified, and upon MS/MS no backbone cleavage occurs between them to unambiguously assign the location of the modification. Often both possible sites will be reported despite one being a false positive.[18] Recently developed software (e.g., Histone Coder and isoScale) has addressed this issue by ensuring that each reported modified site was confirmed by the presence of fragment ions that unambiguously localize modifications, thus allowing curation of false positives and quantification of more than 700 combinatorial histone marks.[18]

Here, we report the use of UVPD for characterization of middle-down sized histone peptides. We used the canonical middle-down MS workflow, including GluC proteolysis followed by histone tail separation with WCX-HILIC coupled online to MS.[17,42] We have previously shown that UVPD results in extensive fragmentation of proteins and peptides and does not cause loss of labile PTMs.[36,44,54] In addition, the performance of UVPD is not strongly dependent on the size of the peptide nor charge state, thus making UVPD well-positioned for the analysis of histones.[55] In this study, UVPD performance is benchmarked with attention to number of backbone cleavages, PTM site localization, and characterization of combinatorial PTMs. Having previously evaluated 193 nm UVPD for characterization of modifications on intact histone proteins,[56] the advantages discussed above regarding the middle-down approach merited further investigation in order to evaluate the applicability of UVPD for characterization of PTMs.

7.3 Materials and Methods

7.3.1 Materials

GluC was purchased from Promega (Madison, WI.) All other reagents and solvents were purchased from Millipore Sigma (St. Louis MI) unless otherwise noted.

7.3.2 HeLa cell preparation

HeLa S3 cells were treated for 24 hrs with 10 mM sodium butyrate and harvested. Histones were extracted as previously described.[82] Briefly, nuclei were isolated after

suspending the cell pellets in nuclei isolation buffer (250 mM sucrose, 0.2%NP-40, 1 mM $CaCL_2$, 15 mM Tris-HCl pH 7.5, 15 mM NaCl, 60 mM KCL, and 5 mM $MgCl_2$). Nuclei were then pelleted and resuspended in 0.4 N H_2SO_4 at a 5:1 ratio (v/v) and incubated for 2 hrs at 4 ^oC with shaking. After acid extraction, histones were precipitated with 25% TCA (w/v). Purified histones (~300 µg) were separated by RP-HPLC as previously described.[82] Briefly, histones were fractionated on a Vydac C18 column (10 mm inner diameter, 250 mm length, 5 um particle size). Histones were eluted over a 100 minute gradient from 30% to 60% solvent B at a flow rate of 1 mL/min. Solvent A consisted of 2% trifluoroacetic acid and 5% acetonitrile in water. Solvent B contained 0.19% triflouroacetic acid and 95% acetonitrile in water. The UV detector was adjusted to 214 nm, and fractions for H4, H2A, H2B, H3, H3.3, H3.2 and H3.1 were collected based on their characteristic retention times [82]. Fractions were dried using a SpeedVac concentrator and stored at -20 ^oC . Finally, the isolated histones H3 and H4 were submitted to GluC digestion (20:1, w/w – Histone to GluC) for 8 hrs in 50 mM ammonium acetate buffer (pH 4) prior to LC-MS analysis.

7.3.3 Liquid chromatography mass spectrometry (LC-MS)

The GluC peptides were resuspended in 2% ACN and separated using a Dionex RSLC 3000 nano-LC system (Thermo Fisher Scientific, San Jose, CA, USA). Approximately 1 µg of digest was loaded onto a 3 cm trapping column (100 µm i.d.) packed in-house with REPROSIL Gold (C18,3 µm particles, 300 Å pore size, Dr. Maisch, Germany). Peptides were then transferred onto a 20 cm fritted (75 µm i.d.) pulled tip analytical column (New Objective, Woburn, MA) packed in-house with (PolyCAT A, 3

μm 1500 Å pore size) a weak cation exchange hydrophilic interaction chromatography (WCX-HILIC) media (Poly LC, Columbia, MD). Peptides were eluted at a flow rate of 300 nL/min using the following gradient: starting at 2% B for 20 minutes, going to 55% B at 23 minutes, then to 90% B at 160 min, and finally to 99% B at 170 min. Mobile phase A was 75% acetonitrile with 20 mM propionic acid (pH 6). Mobile phase B was 75% water with formic acid (pH 2.5).

The nanoLC system was coupled to a Fusion Lumos Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) modified for 193 nm UVPD, as previously described.[83] UVPD was performed in the high-pressure linear ion trap using two pulses (2.5 mJ) from a 193 nm Excistar XS excimer laser (Coherent, Santa Clara, CA). ETD was performed in the high cell of the linear ion trap with 30 ms reaction time for ETD (2×10^5 reagent AGC,) based on optimized conditions reported by Sidoli *et al*.[24] The mass spectrometer was run using the following parameters regardless of fragmentation type: MS1 at 60000 resolving power, 2 μscans averaged per spectrum, 1×10^6 AGC target and MS2 at 30000 resolving power running a top 8 data dependent method. 10 μscans were averaged per MS2 spectrum (1×10^6 AGC target). Only precursor ions in the 8+ charge-state were selected for activation. An example chromatogram and precursor MS scan is shown in **Figure 7.1.**

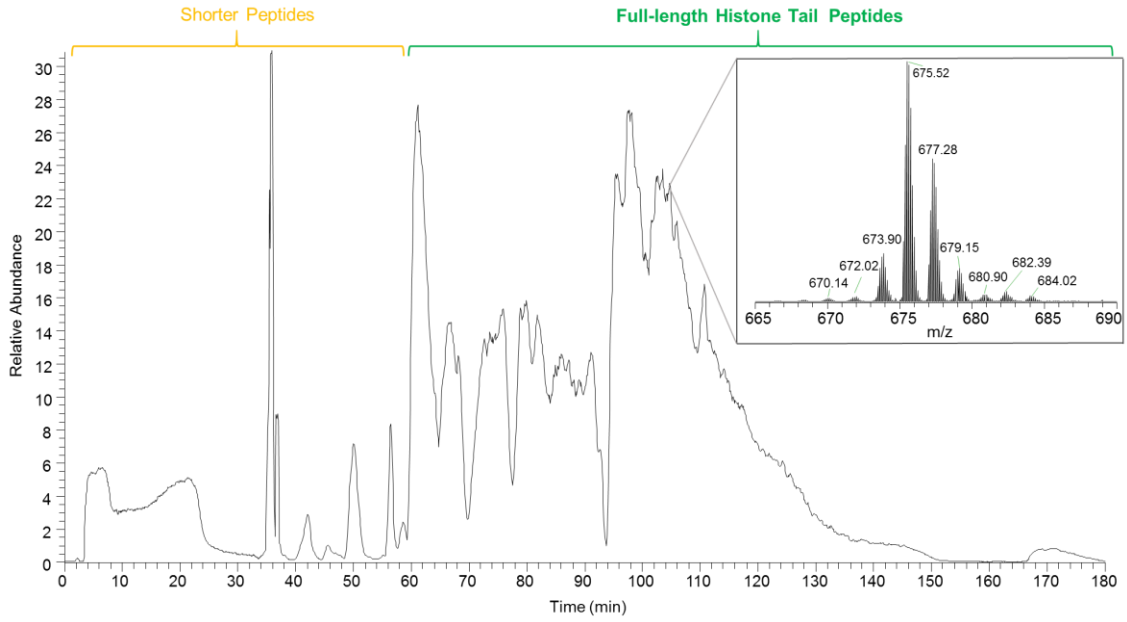


Figure 7.1 Example chromatogram and precursor ion scan (inset) depicting typical nano WCX-HILIC separation of Histone tails. The inset shows the 8+ charge state of several H4 modification sates. The shorter,less modified, peptides tend to elute in the first 60 minutes. The full length modified H4 n-terminal tails eluted from 60-180 minutes.

7.3.4 Data Analysis

LC-MS data was searched using MASCOT (v 12) database searching software. Peptide specificity was set to c-terminal to E. For ETD c and z ions were selected and for UVPD a, b, c, x, y, and z ions were selected. The MASCOT output was further processes using the isoScale slim software package to remove any ambiquously localized modifications.

Prior to analysis in ProSight Lite (Build 1.4.6) 5 scans were averaged to improve the S/N of fragment ions. The resulting spectra was deconvoluted using the Xtract algorithm available in the Xcalibur Qualbrowser (Thermo Fisher Scientific, San Jose CA)

software. Monoisotopic output was selected and the S/N level was set to 3, all other parameters were left to default. The resulting deconvoluted peak list was input into the ProSight Lite software. The canonical H3 or H4 sequences (N-terminal GluC peptide) was imported into ProSight Lite. Monoisotopic input and UVPD or ETD were selected as the fragment type and a 10-ppm tolerance was applied. Choice of PTM location was guided by intact mass, previously reported sites, and primarily the following metrics: P-score, number of matched fragments, and sequence coverage.

7.4 Results and Discussion

The performance of each of the MS/MS methods was evaluated based on peptide level metrics including the number of unique peptide forms (including modifications) identified, sequence coverage, and number and position of diagnostic fragment ions; especially modification-localizing ions.

7.4.1 UVPD Optimization

The energy of a single 193 nm photon (6.4 eV) is sufficient to dissociate most peptides; however, other considerations such as photon flux and number of pulses affects the total energy deposition and potential for secondary dissociation.[59] The photoabsorption cross-section scales with the size of the peptide or protein as the amides serve as the chromophores for 193 nm photoabsorption. A related consideration is the possibility of excessive energy deposition from absorption of multiple photons which can cause secondary fragmentation of ions in a manner that leads to production of un-assignable

internal ions or overly small, uninformative sequence ions. Thus, UVPD parameters were optimized for an ideal 5.6 kDa middle-down sized histone peptide originated from acH4K20me2 (52 residues of the N-terminal tail, net 5661.35 Da, 8+ charge state) possessing two modifications. This particular proteoform (acH4K20me2) is one of the most commonly detected and represents an ideal benchmark histone.[44,48] **Figure 7.2** shows the dependence of sequence coverage on the N-terminal peptide on laser pulse number and power. Sequence coverage and P-score values were generated using Xtract to deconvolute the raw data and ProSight Lite to match the deconvoluted fragment ions to the theoretical modified sequence of histone H4 (residues 2-53). Using a single laser pulse, the sequence coverage increased with increasing laser power; however, with multiple pulses the increase in sequence coverage peaked or plateaued at 2.5 mJ. Optimal sequence coverage (69%) was obtained using two pulses at 2.5 mJ. Other combinations of laser conditions yielded similar performance, such as 3 pulses at 2 mJ (67%). The P-scores were used to discriminate between the best performing UVPD conditions (**Figure 7.3**). The P-score, based on the probability of observed spectra matching theoretical spectra by random chance, is a useful metric as it is often utilized by database searching algorithms such as MASCOT (used in this study) during LC-MS data analysis. Applying 2 pulses at 2.5 mJ gave the lowest P-score (2.6E-66), indicating the highest confidence in the fragment-to-theoretical spectral match. Secondary dissociation and generation of internal fragments occurs if the photon flux is too high or the ions are exposed to multiple pulses. These additional non-diagnostic ions can negatively influence spectral matching confidence,

which may be the case for the 3-pulse data, a factor that would explain the high sequence coverage (**Figure 7.2**) but non-optimal P-score (**Figure 7.3**).

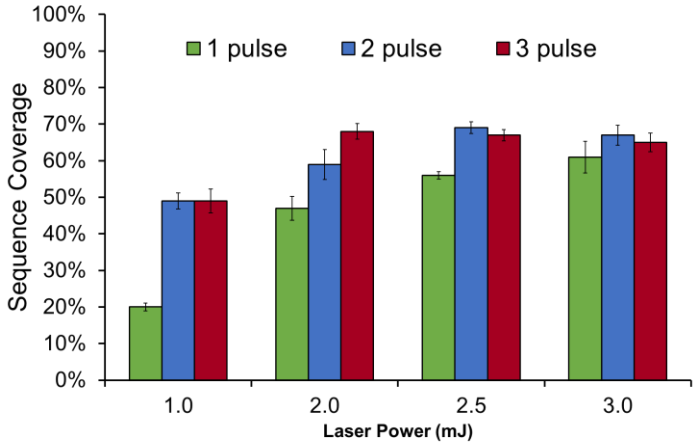


Figure 7.2 UVPD laser optimization: sequence coverage dependence on pulse number and power of (8+) acH4K20me2

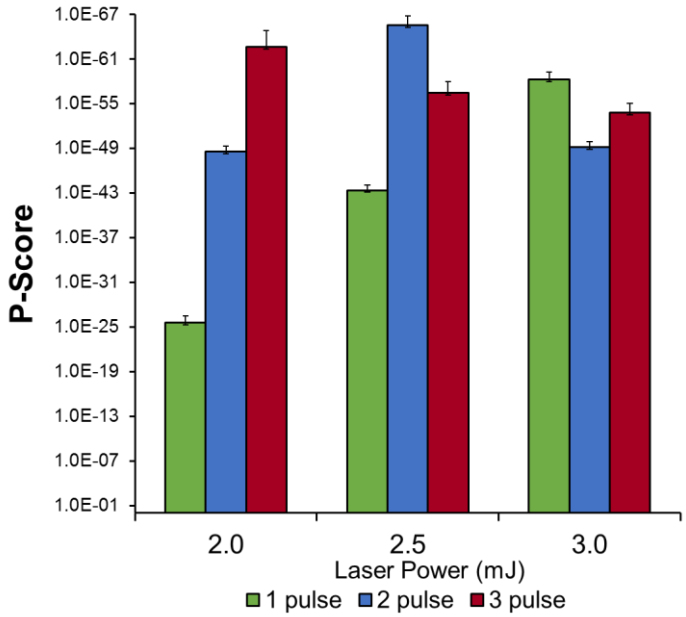


Figure 7.3 Effect of UVPD laser pulse number and power on P-scores of the UVPD mass spectra matched to (8+) acH4K20me2

7.4.2 Benchmarking UVPD against ETD

WCX-HILIC separations followed by high resolution ETD-MS has become the gold standard method for middle-down histone analysis, as originally implemented by Young *et al.*[17] Given the large sizes of the N-terminal peptides and their basic nature, under the acidified conditions utilized in the WCX-HILIC separation they are often multiply protonated and found in charge states ranging from 5+ to 12+. ETD proved to be an efficient means to characterize these multiply charged basic peptides while retaining their abundant modifications.[17] While UVPD is similar to ETD with respect to retention of PTMs and the ability to generate excellent sequence coverage, UVPD generates several additional ion types (UVPD: *a,a+1,b,c,x,x+1,y,y-1,z* compared to *c/z* for ETD).[60] These additional ion types have the potential to add confidence in localization of modification sites and improve sequence coverage, at the expense of

potentially reducing the S/N levels of the resulting MS/MS spectra owing to greater dispersion of the ion current.

In order to evaluate the viability of UVPD for LC-MS analysis of the many modified forms of histone H3, a mixture of H3 tails were subjected to WCX-HILIC separation and analyzed by ETD and UVPD. To maximize the sensitivity of the analysis, a narrow mass window bracketing the +8 charge state of the H3 tail and its modified forms was used, followed by data dependent selection of precursors for MS2 analysis.[17] The global performance of ETD and UVPD was evaluated based on the number of unique species identified, and detailed evaluation of the fragment ion spectra generated by both methods is discussed later. The number of unique species detected, after filtering out ambiguous matches and non-quantifiable species, was similar for histone H3 (175 proteoforms for ETD and 180 proteoforms for UVPD thus showing that UVPD is comparable to ETD with respect to number of identifications and is a competitive strategy for identification of heavily modified middle-down sized peptides.

The histone peptides identified by ETD and UVPD were heavily modified. In order to characterize the multitude of modifications, each modifiable site was considered, and the relative contribution of acetylation (ac:yellow), methylation (me1:green), dimethylation (me2:blue), and trimethylation (me3:red) are displayed in **Figure 7.4.** Overall, the relative distributions of modifications characterized by UVPD and ETD were similar; however, UVPD of the untreated set resulted in identification of a greater proportion of methylation sites on residues closer to the C-terminus. Several abundant proteoforms identified by UVPD contained K27me1 and K36me1 and contributed to this finding. Speculation why UVPD better for methylation close to C-term? I cant think of any mechanistic reason, my thought is that both of these sites are right next to another K or R and required a fragment at the K/R or K/K site in order to pass the isoScale filter. I presume that UVPD generated these specific (important) cleavages more frequently than ETD.

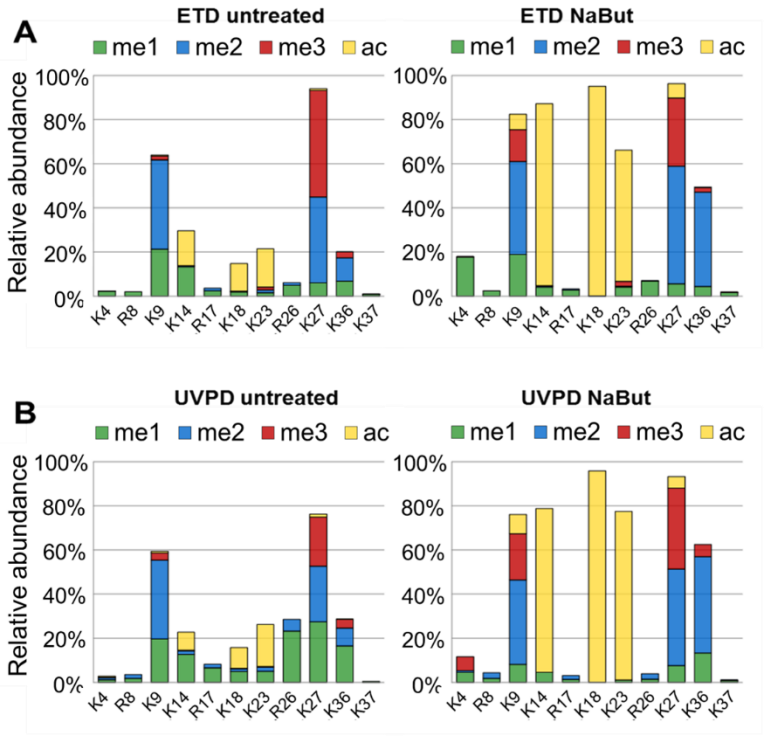


Figure 7.4 Distribution of specific modifications detected by ETD (A and C) and UVPD (B and D) based on analysis of the N-terminal peptides of histone H3 (residues 1-50) for untreated (A and B) and butyrate-treated (NaBut) (C and D) cells. The distributions represent the summation of modifications found on all forms of H3.

Among the proteoforms identified, K4, K9, K14, K18, K23 and K27 were found to be acetylated. After treatment with NaBut, acetylation of K14, K18 and K23 was detected at significantly increased levels by both UVPD and ETD, whereas acetylation of K9 and K27 increased slightly. NaBut has been shown to block histone deacetylase enzymes (HDACs) resulting in hyperacetylation,[17] so our results are consistent with this finding. Both UVPD and ETD yielded PTM distributions which were nearly identical and reflected a large increase in acetylation after NaBut treatment, confirming that UVPD should be applicable for relative quantitation of PTMs and is sufficiently sensitive to discriminate different modification distributions based on biological conditions (e.g. NaBut treatment vs. untreated).

Figure 7.5 shows the log fold change between NaBut-treated and control samples for the individual PTMs of histone H3 resulting from either the ETD or UVPD analysis. The abundance of the single PTMs was assessed by summing the relative abundance of all the quantified polypeptides carrying each individual PTM to obtain their total relative abundance. Change in acetylation is indicated by green data points in **Figure 7.5** . Significant increases in acetylation were found by both UVPD and ETD, an outcome consistent with inhibition of HDACs by sodium butyrate.

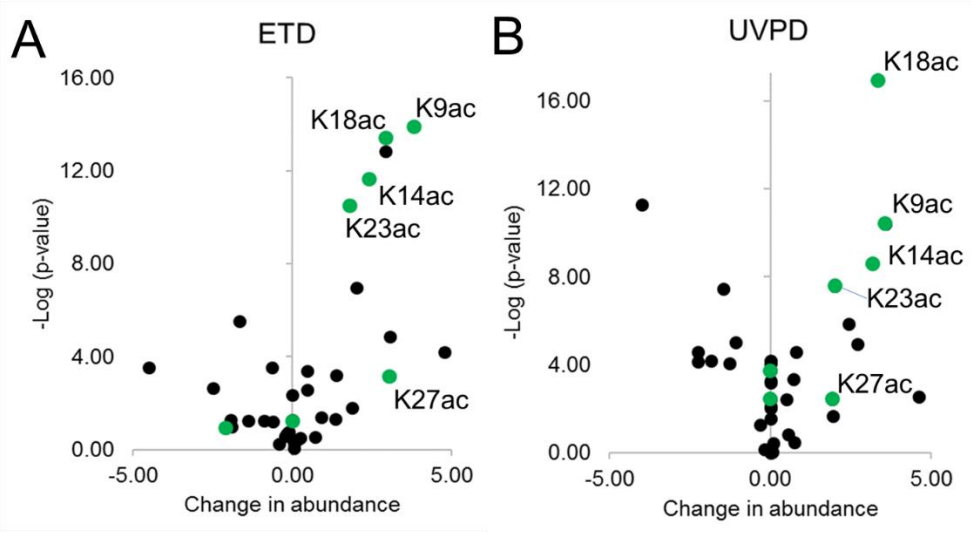


Figure 7.5 Fold change in modification relative abundance for Histone H3 identified by ETD (A) and UVPD (B). Green points indicate upregulated acetylation

7.4.3 Comparison of modified forms

The overlap of modified forms identified by both ETD and UVPD accounts for only 15% of the total H3 histoforms, as summarized in the Venn diagrams shown in **Figure 7.6**. In fact, the forms identified uniquely by either ETD or UVPD account for over 80% of the total forms identified, strongly suggesting the complementarity of these two methods. In many

cases, a histoform identified uniquely by UVPD diverges from a similar one found by ETD based on a difference in a single modification. For instance, histone H3R8me1K14acR17me1K18acR26me1K36me2 (containing six modifications) identified by UVPD differs in only one position, K23, from H3R8me1K14acR17me1K18acK23acR26me1K36me2 (containing 7 modifications) identified by ETD. For this histone, 6 out of 7 modifications were identified in common by both methods and confirmed by manual interpretation. However, the one identified by UVPD displayed acetylation of Lys18, whereas the one characterized by ETD exhibited acetylation of Lys 18 and Lys23. Despite the large difference in the specific proteoforms identified by each method, the relative distributions of modifications were similar (**Figure 7.4**). Inspection of the abundances of the modified forms from **Figure 7.6** indicates that the ~17% of modified histones found in common for UVPD and ETD account for approximately 30% of the total abundance of histoforms identified by UVPD and 40% of the total abundance found by ETD.

Because we identify and quantify intact histone tails, it is possible to assess similarities and differences with the estimated co-frequencies of PTMs (i.e. instances where two PTMs occur on the same peptide). **Figure 7.7** shows a web diagram illustrating the co-occurrence of modifications on untreated H4 as indicated by a weighted line connection. The abundance of these co-occurrences is denoted by the thickness of the line. Similar co-occurrences are observed by both ETD and UVPD, again confirming the reproducibility in PTM quantification despite some differences in the identified combinatorial codes

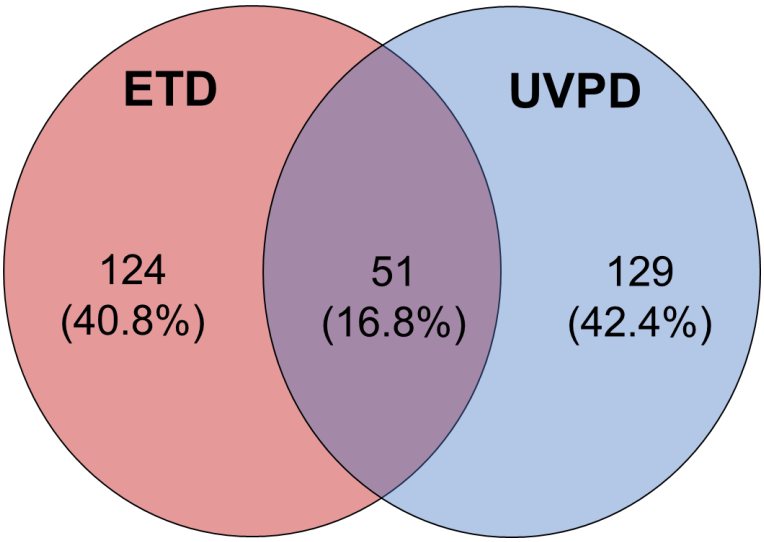


Figure 7.6 Venn diagram showing the number of modified forms of n-terminal histone H3 peptides (1-50) identified in common and uniquely by ETD and UVPD

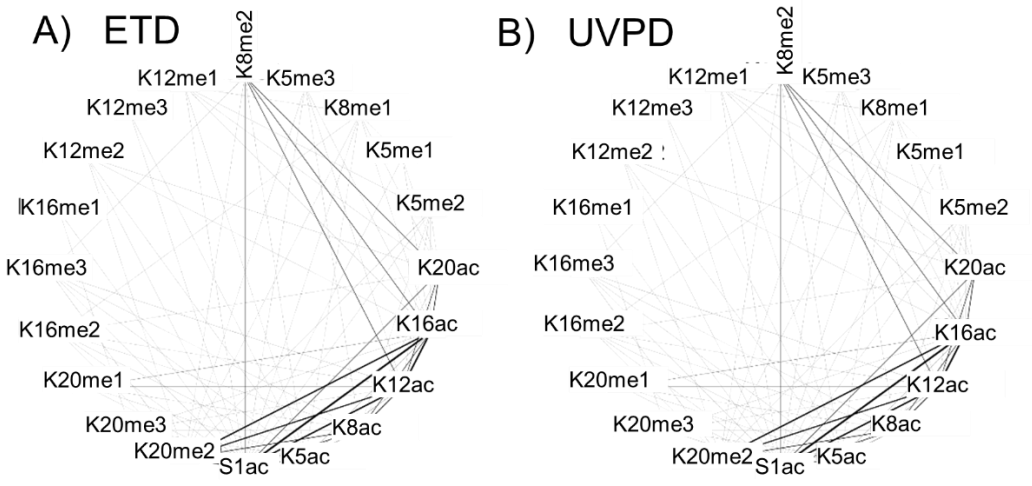


Figure 7.7 PTM co-existence web for histone H4 (1-50). The line connection indicates co-existence and the line thickness indicates frequency of co-existence. Thicker connection indicates those two modifications are more often found on the same peptides

7.4.4 ETD and UVPD fragmentation

One likely factor contributing to the differences in the distribution of PTMs, and both the number and overlap of identified species mentioned above is the significant number of ions generated by UVPD which are not utilized by MASCOT for scoring spectral matches. UVPD consistently generates many diverse ion types, including a, $a+1$, b, c, x, $x+1$, y, $y-1$, and z type ions.[61] MASCOT has been designed to utilize *a, b ,c, x, y, z*, and $z+1$ ions for scoring. **Figure 7.8** shows the distribution of ions generated by UVPD of one typical middle-down sized doubly-modified peptide, representing acH4K20me2, the same species used for the UVPD optimization. The fragment ions were matched at 10 ppm error using ProSight Lite. The results in **Figure 7.8** imply that MASCOT utilizes only 53% of the total number of UVPD

fragment ions possible (corresponding to only 35% of the total abundances of identified fragment ions). Moreover, the presence of the diagnostic *a*+1, *x*+1, and *y*-1 ions are not utilized and may be counted as noise, actually depressing the MASCOT scoring metrics for UVPD peptide spectral matches.

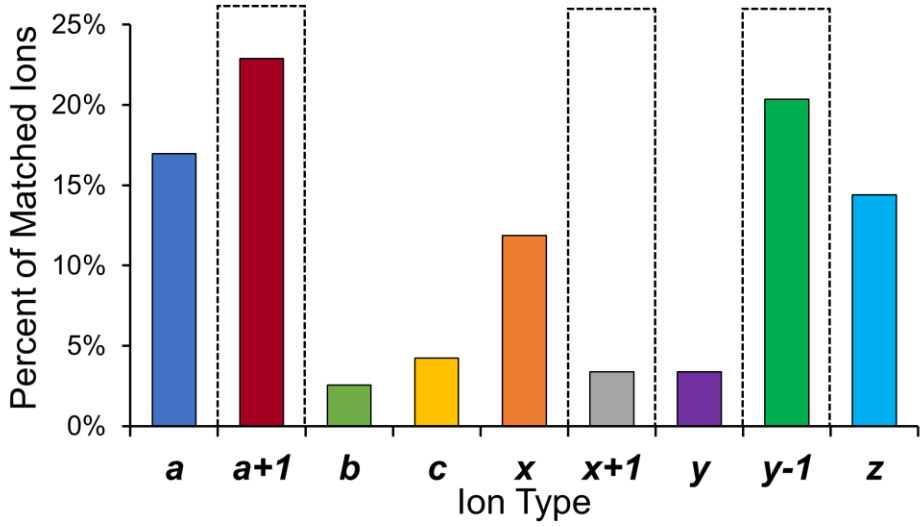


Figure 7.8 Ion type distribution (by number of matched ions) for UVPD spectra of acH4K20me2. Highlighted bars indicate ion types which are not used for scoring spectral matches.

Better utilization of the ion types characteristic of UVPD of middle-down size peptides could increase the confidence of UVPD PSMs, increase the number of overall matched forms, and reconcile some of the differences seen between ETD and UVPD results. Although training MASCOT (or another platform) for UVPD spectra would result in a more ideal performance, the use of isoScale (a program currently only compatible with

MASCOT output) justifies the workflow used in this study. IsoScale further processes the MASCOT output, specifically focusing on culling false positive modification assignments by virtue of localizing fragment ions. After isoScale processing, the final results are unambiguous (i.e. each localized modification is supported by bracketing fragment ions.) This feature is crucial for analyzing datasets containing heavily modified peptides with a high level of confidence.

7.4.5 Characterization of the most heavily modified species (> 5 PTMs)

In light of the shortcomings of the current automated workflow, manual annotation can be used to achieve the greatest sequence coverage and PTM site localization from UVPD spectra. Both ETD and UVPD are effective for characterization of lightly and moderately modified species. UVPD is especially useful for heavily modified forms (i.e. ones containing more than five modifications). In order to highlight the proficiency of UVPD for characterization of highly modified histones, ProSight Lite was used to manually annotate UVPD spectra acquired for the most heavily modified species.[62] **Figure 7.9** shows deconvoluted ETD and UVPD mass spectra of the hepta-modified peptide H3K4me1K9me2K14acK18acK23acK27acK36me3 (8+ charge state, a N-terminal tail containing 50 residues with mono-methylation of residues K4, dimethylation of residues K9, trimethylation of residues K36, and acetylation of residues K14,K18,K23 and K27). Both MS/MS methods adequately localized several of the modifications; however, UVPD was able to achieve higher confidence by virtue of production of multiple PTM-localizing fragment ions. For example, UVPD successfully characterized K14ac and K27ac,

generating the greatest number (three or more) of flanking fragment ions containing the modification, including both complementary C-terminal and N-terminal ions. K14ac was localized by $a_{14} + 1, z_{37}, y_{37} - 1, x_{37} + 1,$ and K27ac was localized by $a_{27} + 1, y_{24} - 1,$ and $x_{24} + 1$. By comparison, ETD best characterized K4me, K9me2 and K18, generating only one fragment ion containing the modification and one or more flanking ions facilitating localization

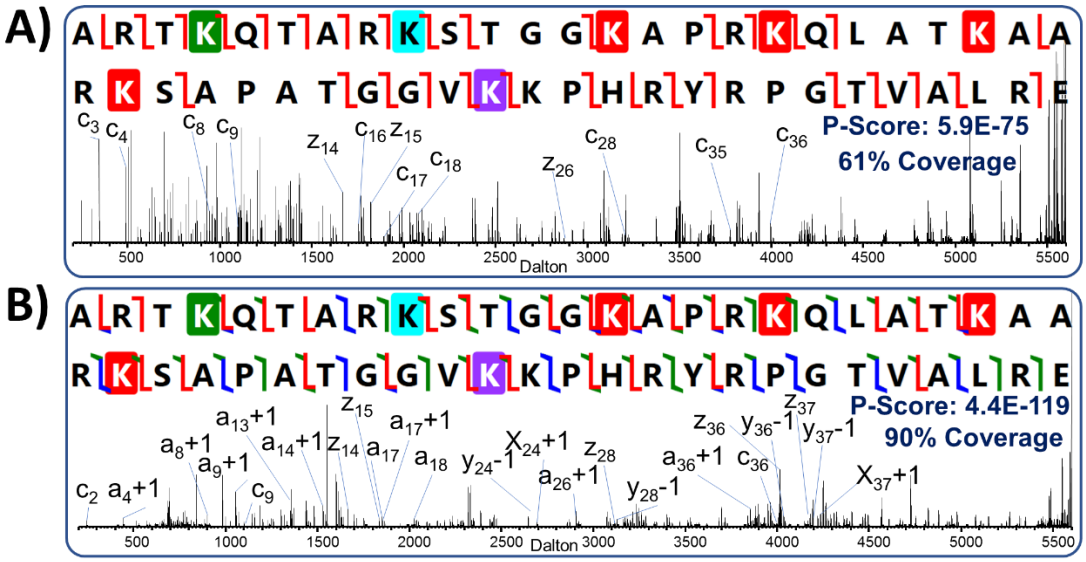


Figure 7.9 Comparison of deconvoluted (A) ETD and (B) UVPD spectra of H3K4me1K9me2K14acK18acK23acK27acK36me3 (8+) showing fragment ion maps, P-score, sequence coverage and labeled solidification site localizing ions.

7.4.6 Presence and use of neutral loss ions

Modified peptides can undergo informative neutral losses after activation, often exploited for characterization of phosphorylated and glycosylated peptides.[36,62] Traditional

bottom-up analysis of histone peptides has utilized neutral losses generated by HCD and ETD of methylated peptides, particularly 59.07 Da from trimethylated lysine and 45.06 Da from Arg residues of histones.[63] The presence and diagnostic nature of neutral losses upon UVPD of methylated species has not been reported previously. These neutral losses can be very useful for determining the specific nature of modified lysines. For instance, the mass difference between a trimethylated lysine and an acetylated lysine is 0.036 Da which for a 5-6 kDa peptide represents a 6-7 ppm mass difference, well within the accepted mass tolerance of 10 ppm. The heavily modified H3 peptide shown in **Figure 7.8** has several acetylated lysines and a trimethylated lysine residue. The presence of a 59.07 Da neutral loss upon UVPD can be used to discriminate between the K36ac and K36me3 forms. **Figure 7.10** shows the occurrence of the 59.07 Da loss, thus confirming the presence of a trimethylation of K36. Conversely, when the trimethylated K36 residue is replaced with acetylated K36 in the search, several scoring metrics degrade, including P-score, number of matched fragments and the ppm mass error. Although other metrics can be used to discriminate between acetylation and trimethylation, the presence of the 59.07 Da mass loss provides further evidence supporting the assignment of trimethylation of K36 for this peptide.

Figure 7.11 shows the sequences and deconvoluted UVPD mass spectra of two nearly isobaric N-terminal peptides of H3K4acR8me2K23acK27me2 (5478.15 Da) and H3K9me3K14acKme2K36me2 (5478.19 Da), differing only by the acetyl-trimethyl mass difference. For each of these proteoforms, the 8+ charge state was subjected to UVPD. The trimethylated species displays the expected 59.07 Da neutral loss which is absent from the

UVPD mass spectrum of the acetylated species, confirming the assignment of trimethylation. The presence of this diagnostic neutral loss ion upon UVPD offer notable utility for correctly interpreting ambiguous spectra.

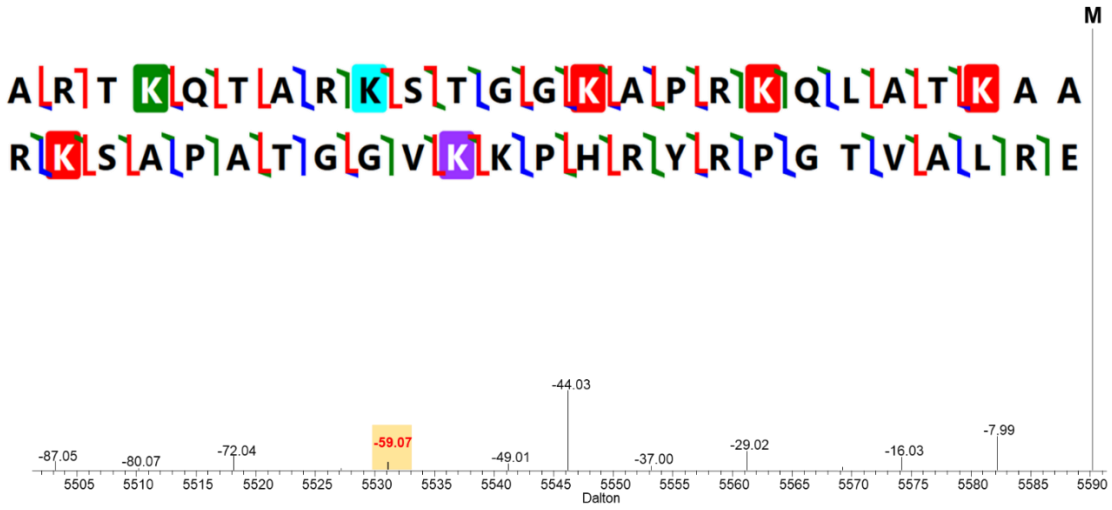


Figure 7.10 Deconvoluted fragment ion spectra showing a -59.07 Da neutral loss ion confirming the presence of a trimethylation on H3K4me1K9me2K14acK18acK23acK27acK36me3 (8+).

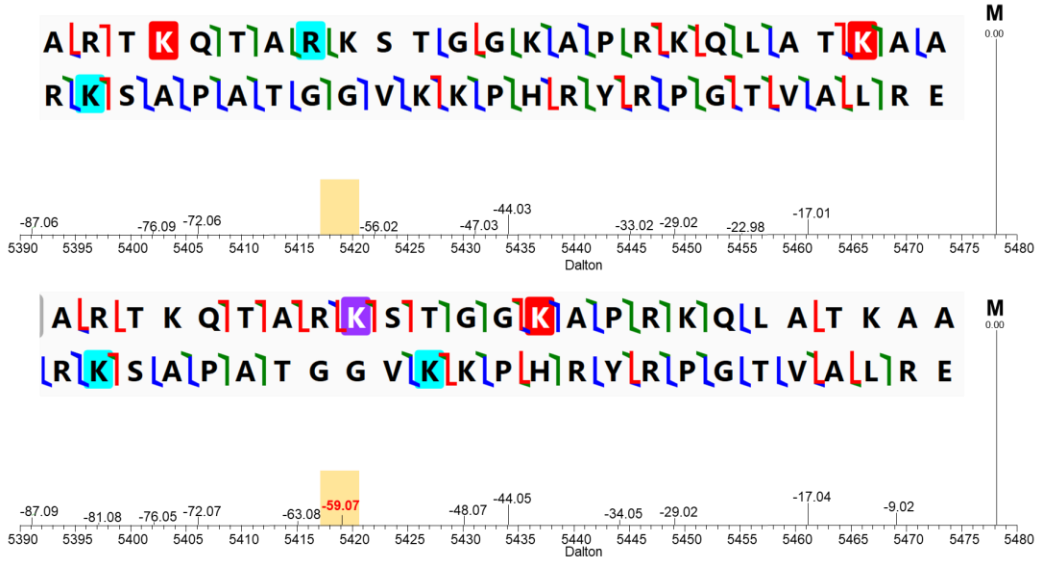


Figure 7.11 Deconvoluted UVPD spectra showing the utility of the 59.07 Da neutral loss in confirming the presence of a trimethylation on two nearly isobaric species: H3K4acR8me2K23acK27me2(5478.15Da)and H3K9me3K14acKme2K36me2 (5478.19 Da) both 8+.

7.5 CONCLUSION

Results from LC-MS analyses of GluC-generated middle-down sized N-terminal tails of histone H3 and H4 demonstrate that UVPD is broadly applicable for characterization of heavily modified histones. The two methods identified largely unique (only 15% overlap) combinatorial species. However, UVPD and ETD led to highly comparable results when assessing the abundance of single and co-existing modifications, implying that the high orthogonality in terms of which combinatorial codes are identified in every MS run does not affect significantly the ultimate conclusions.

UVPD was useful for deciphering changes in modifications related to specific cellular treatments, as shown by the 2-fold to 4-fold up-regulation of acetylation in the NaBut-treated Hela cells. Evaluation of the differences between ETD and UVPD revealed that the automated data processing workflow, which relies on MASCOT, utilizes only approximately 50% of the ions generated by UVPD. Approximately 35% of the total matched fragment ion population from UVPD originates from the *a*+1 ion series which is not considered by MASCOT. Moving forward, a fully trained search algorithm would extend the capabilities of UVPD for high throughput analysis of modified middle-down sized peptides. In order to evaluate the ability of UVPD to characterize the most heavily modified peptides, manual spectral interpretation facilitated by ProSight Lite was a key to success. The N-terminal peptide of hepta-modified histone H3K4me1K9me2K14acK18acK23acK27acK36me3 was well-characterized and yielded 90% sequence coverage, motivating future investigation of UVPD for interrogating other heavily modified middle-down sized peptides. UVPD also resulted in characteristic neutral loss pathways. For example, loss of 59.07 Da upon UVPD differentiated trimethylation from acetylation for histones H3K4acR8me2K23acK27me2 and H3K9me3K14acKme2K36me2.

7.6 REFERENCES

(1) Venne, A. S.; Kollipara, L.; Zahedi, R. P. The next Level of Complexity: Crosstalk of Posttranslational Modifications. *Proteomics* **2014**, *14* (4–5), 513–524.

(2) Jaenisch, R.; Bird, A. Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals. *Nat. Genet.* **2003**, *33*, 245–254.

(3) Appella, E.; Anderson, C. W. Post-Translational Modifications and Activation of P53 by Genotoxic Stresses. *Eur. J. Biochem.* **2001**, *268* (10), 2764–2772.

(4) Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **2007**, *128* (4), 693–705.

(5) Meek, D. W.; Anderson, C. W. Posttranslational Modification of P53: Cooperative Integrators of Function. *Cold Spring Harb. Perspect. Biol.* **2009**, *1* (6), 1–16.

(6) Dai, C.; Gu, W. P53 Post-Translational Modification: Deregulated in Tumorigenesis. *Trends Mol. Med.* **2010**, *16* (11), 528–536.

(7) Bode, A. M.; Dong, Z. Post-Translational Modification of P53 in Tumorigenesis. *Nat. Rev. Cancer* **2004**, *4* (10), 793–805.

(8) Jenuwein, T. Translating the Histone Code. *Science* **2001**, *293* (5532), 1074–1080.

(9) Bannister, A. J.; Kouzarides, T. Regulation of Chromatin by Histone Modifications. *Cell Res.* **2011**, *21* (3), 381–395.

(10) Luo, Y.; Yogesha, S. D.; Cannon, J. R.; Yan, W.; Ellington, A. D.; Brodbelt, J. S.; Zhang, Y. Novel Modifications on C-Terminal Domain of RNA Polymerase II Can Fine-Tune the Phosphatase Activity of Ssu72. *ACS Chem. Biol.* **2013**, *8* (9), 2042–2052.

(11) Sims, R. J.; Rojas, L. A.; Beck, D. B.; Bonasio, R.; Schüller, R.; Drury, W. J.; Eick, D.; Reinberg, D. The C-Terminal Domain of RNA Polymerase II Is Modified by Site-Specific Methylation. *Science* **2011**, *332* (6025), 99–103.

(12) Farriol-Mathis, N.; Garavelli, J. S.; Boeckmann, B.; Duvaud, S.; Gasteiger, E.; Gateau, A.; Veuthey, A.-L.; Bairoch, A. Annotation of Post-Translational Modifications in the Swiss-Prot Knowledge Base. *PROTEOMICS* **2004**, *4* (6), 1537–1550.

(13) Garavelli, J. S. The RESID Database of Protein Modifications as a Resource and Annotation Tool. *PROTEOMICS* **2004**, *4* (6), 1527–1533.

(14) Mann, M.; Kulak, N. A.; Nagaraj, N.; Cox, J. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. *Mol. Cell* **2013**, *49* (4), 583–590.

(15) Olsen, J. V.; Mann, M. Status of Large-Scale Analysis of Post-Translational Modifications by Mass Spectrometry. *Mol. Cell. Proteomics* **2013**, *12* (12), 3444–3452.

(16) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. Peptide Sequencing and Characterization of Post-Translational Modifications by Enhanced Ion-Charging and Liquid Chromatography Electron-Transfer Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2007**, *79* (24), 9243–9252.

(17) Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A. High Throughput Characterization of Combinatorial Histone Codes. *Mol. Cell. Proteomics* **2009**, *8* (10), 2266–2284.

(18) Sidoli, S.; Schwämmle, V.; Ruminowicz, C.; Hansen, T. A.; Wu, X.; Helin, K.; Jensen, O. N. Middle-down Hybrid Chromatography/Tandem Mass Spectrometry Workflow for Characterization of Combinatorial Post-Translational Modifications in Histones. *PROTEOMICS* **2014**, *14* (19), 2200–2211.

(19) Moradian, A.; Kalli, A.; Sweredoski, M. J.; Hess, S. The Top-down, Middle-down, and Bottom-up Mass Spectrometry Approaches for Characterization of Histone Variants and Their Post-Translational Modifications. *PROTEOMICS* **2014**, *14* (4–5), 489–497.

(20) Egelhofer, T. A.; Minoda, A.; Klugman, S.; Lee, K.; Kolasinska-Zwierz, P.; Alekseyenko, A. A.; Cheung, M.-S.; Day, D. S.; Gadel, S.; Gorchakov, A. A.; et al. An Assessment of Histone-Modification Antibody Quality. *Nat. Struct. Mol. Biol.* **2011**, *18* (1), 91–94.

(21) Pereira Morais, M. P.; Mackay, J. D.; Bhamra, S. K.; Buchanan, J. G.; James, T. D.; Fossey, J. S.; van den Elsen, J. M. H. Analysis of Protein Glycation Using Phenylboronate Acrylamide Gel Electrophoresis. *PROTEOMICS* **2010**, *10* (1), 48–58.

(22) Strack, R. Mass Spectrometry: Designer Proteases for Post-Translational Modifications. *Nat. Methods* **2017**, *14* (2), 106–107.

(23) Aebersold, R.; Mann, M. Mass-Spectrometric Exploration of Proteome Structure and Function. *Nature* **2016**, *537* (7620), 347–355.

(24) Larance, M.; Lamond, A. I. Multidimensional Proteomics for Cell Biology. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (5), 269–280.

(25) Villén, J.; Gygi, S. P. The SCX/IMAC Enrichment Approach for Global Phosphorylation Analysis by Mass Spectrometry. *Nat. Protoc.* **2008**, *3* (10), 1630–1638.

(26) Carlson, S. M.; Moore, K. E.; Green, E. M.; Martin, G. M.; Gozani, O. Proteome-Wide Enrichment of Proteins Modified by Lysine Methylation. *Nat. Protoc.* **2013**, *9* (1), 37–50.

(27) Li, Y.; Silva, J. C.; Skinner, M. E.; Lombard, D. B. Mass Spectrometry-Based Detection of Protein Acetylation. In *Sirtuins*; Hirschey, M. D., Ed.; Humana Press: Totowa, NJ, 2013; Vol. 1077, pp 81–104.

(28) Wang, K.; Dong, M.; Mao, J.; Wang, Y.; Jin, Y.; Ye, M.; Zou, H. Antibody-Free Approach for the Global Analysis of Protein Methylation. *Anal. Chem.* **2016**, *88* (23), 11319–11327.

(29) Zhang, L.; Liu, C.-W.; Zhang, Q. Online 2D-LC-MS/MS Platform for Analysis of Glycated Proteome. *Anal. Chem.* **2017**.

(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. Tyrosine Sulfation in a Gram-Negative Bacterium. *Nat. Commun.* **2012**, *3*, 1153.

(31) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. *Anal. Chem.* **2016**, *88* (4), 2303–2310.

(32) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. Analysis of Phosphorylation Sites on Proteins from Saccharomyces Cerevisiae by Electron Transfer Dissociation (ETD) Mass Spectrometry. *Proc. Natl. Acad. Sci.* **2007**, *104* (7), 2193–2198.

(33) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. *Proteomics* **2008**, *8* (21), 4466–4483.

(34) Liao, R.; Zheng, D.; Nie, A.; Zhou, S.; Deng, H.; Gao, Y.; Yang, P.; Yu, Y.; Tan, L.; Qi, W.; et al. Sensitive and Precise Characterization of Combinatorial Histone Modifications by Selective Derivatization Coupled with RPLC-EThcD-MS/MS. *J. Proteome Res.* **2017**, *16* (2), 780–787.

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

(36) Robinson, M. R.; Taliaferro, J. M.; Dalby, K. N.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation Mass Spectrometry for Phosphopeptide Characterization in the Positive and Negative Ion Modes. *J. Proteome Res.* **2016**, *15* (8), 2739–2748.

(37) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Anal. Chem.* **2000**, *72* (3), 563–573.

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

(39) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2012**, *84* (22), 9668–9673.

(40) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. *J. Proteome Res.* **2017**, *16* (7), 2653–2659.

(41) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-Nm Photodissociation of Singly and Multiply Charged Peptide Anions for Acidic Proteome Characterization. *PROTEOMICS* **2011**, *11* (7), 1329–1334.

(42) Sidoli, S.; Garcia, B. A. Middle-down Proteomics: A Still Unexploited Resource for Chromatin Biology. *Expert Rev. Proteomics* **2017**, *14* (7), 617–626.

(43) Peters, A. H. F. M.; Kubicek, S.; Mechtler, K.; O'Sullivan, R. J.; Derijck, A. A. H. A.; Perez-Burgos, L.; Kohlmaier, A.; Opravil, S.; Tachibana, M.; Shinkai, Y.; et al. Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. *Mol. Cell* **2003**, *12* (6), 1577–1589.

(44) Dang, X.; Scotcher, J.; Wu, S.; Chu, R. K.; Tolic, N.; Ntai, I.; Thomas, P. M.; Fellers, R. T.; Early, B. P.; Zheng, Y.; et al. The First Pilot Project of the Consortium for Top-down Proteomics: A Status Report. *PROTEOMICS* **2014**, *14* (10), 1130–1140.

(45) Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C. High-Throughput Middle-Down Analysis Using an Orbitrap. *J. Proteome Res.* **2010**, *9* (8), 3886–3890.

(46) Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W. P.; Mohammed, S.; Heck, A. J. R. Toward an Optimized Workflow for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89* (6), 3318–3325.

(47) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; et al. Informed-Proteomics: Open-Source Software Package for Top-down Proteomics. *Nat. Methods* **2017**, *14* (9), 909–914.

(48) Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. Quantitative Analysis of Modified Proteins and Their Positional Isomers by Tandem Mass Spectrometry: Human Histone H4. *Anal. Chem.* **2006**, *78* (13), 4271–4280.

(49) Zheng, Y.; Fornelli, L.; Compton, P. D.; Sharma, S.; Canterbury, J.; Mullen, C.; Zabrouskov, V.; Fellers, R. T.; Thomas, P. M.; Licht, J. D.; et al. Unabridged Analysis of Human Histone H3 by Differential Top-Down Mass Spectrometry Reveals Hypermethylated Proteoforms from MMSET/NSD2 Overexpression. *Mol. Cell. Proteomics* **2016**, *15* (3), 776–790.

(50) Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; et al. Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-Translational Modifications. *J. Proteome Res.* **2004**, *3* (3), 621–626.

(51) Plazas-Mayorca, M. D.; Zee, B. M.; Young, N. L.; Fingerman, I. M.; LeRoy, G.; Briggs, S. D.; Garcia, B. A. One-Pot Shotgun Quantitative Mass Spectrometry Characterization of Histones. *J. Proteome Res.* **2009**, *8* (11), 5367–5374.

(52) Schräder, C. U.; Ziemianowicz, D. S.; Merx, K.; Schriemer, D. C. Simultaneous Proteoform Analysis of Histones H3 and H4 with a Simplified Middle-Down Proteomics Method. *Anal. Chem.* **2018**, *90* (5), 3083–3090.

(53) Sidoli, S.; Lu, C.; Coradin, M.; Wang, X.; Karch, K. R.; Ruminowicz, C.; Garcia, B. A. Metabolic Labeling in Middle-down Proteomics Allows for Investigation of the Dynamics of the Histone Code. *Epigenetics Chromatin* **2017**, *10* (1), 34.

(54) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S.

Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. *Anal. Chem.* **2014**, *86* (4), 2185–2192.

(55) Greer, S. M.; Holden, D. D.; Fellers, R.; Kelleher, N. L.; Brodbelt, J. S. Modulation of Protein Fragmentation Through Carbamylation of Primary Amines. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (8), 1587–1599.

(56) Greer, S. M.; Brodbelt, J. S. Top-Down Characterization of Heavily Modified Histones Using 193 Nm Ultraviolet Photodissociation (UVPD) Mass Spectrometry. *J. Proteome Res.* **2018**.

(57) Lin, S.; Garcia, B. A. Chapter One - Examining Histone Posttranslational Modification Patterns by High-Resolution Mass Spectrometry. In *Methods in Enzymology*; Wu, C., Allis, C. D., Eds.; Nucleosomes, Histones & Chromatin Part A; Academic Press, 2012; Vol. 512, pp 3–28.

(58) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2016**, *88* (1), 1044–1051.

(59) Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem Soc Rev* **2014**, *43* (8), 2757–2783.

(60) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88* (1), 30–51.

(61) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. Integration of Ultraviolet Photodissociation with Proton Transfer Reactions and Ion Parking for Analysis of Intact Proteins. *Anal. Chem.* **2016**, *88* (1), 1008–1016.

(62) DeHart, C. J.; Fellers, R. T.; Fornelli, L.; Kelleher, N. L.; Thomas, P. M. Bioinformatics Analysis of Top-Down Mass Spectrometry Data with ProSight Lite. In *Protein Bioinformatics*; Methods in Molecular Biology; Humana Press, New York, NY, 2017; pp 381–394.

(63) Yang, Y.; Liu, F.; Franc, V.; Halim, L. A.; Schellekens, H.; Heck, A. J. R. Hybrid Mass Spectrometry Approaches in Glycoprotein Analysis and Their Usage in Scoring Biosimilarity. *Nat. Commun.* **2016**, *7*, 13397.

(64) Zhang, K.; Yau, P. M.; Chandrasekhar, B.; New, R.; Kondrat, R.; Imai, B. S.; Bradbury, M. E. Differentiation between Peptides Containing Acetylated or Tri-Methylated Lysines by Mass Spectrometry: An Application for Determining Lysine 9 Acetylation and Methylation of Histone H3. *PROTEOMICS* **2004**, *4* (1), 1–10.

Chapter 8

Top-Down Characterization of Heavily Modified Histones using 193 nm Ultraviolet Photodissociation (UVPD) Mass Spectrometry

8.1 Overview

The characterization of protein post-translational modifications (PTMs) remains a significant challenge for traditional bottom-up proteomics methods owing to the lability of PTMs and the difficulty of mapping combinatorial patterns of PTMs based on analysis of small peptides. These shortcomings have accelerated interest in top-down MS/MS methods which focus on analysis of intact proteins. Mapping all protein PTMs simultaneously requires the ability to obtain extensive sequence coverage to allow confident localization of the modifications. 193 nm Ultraviolet photodissociation (UVPD) has been shown to generate unparalleled sequence coverage for intact proteins compared to traditional MS/MS methods. This study focuses on identification and localization of PTMs of histones by UVPD, higher-energy collisional dissociation (HCD) and the hybrid method electron-transfer/higher-energy collision dissociation (EThcD) via a high throughput LC-MS strategy. In total over 500 proteoforms were characterized among these three activation methods with 46% of the identifications found in common by two or more activation methods. EThcD and UVPD afforded more extensive characterize of proteoforms than

HCD with average gains in sequence coverage of 15% and C-scores that doubled on average.

8.2 INTRODUCTION

Mass spectrometry-based strategies for proteomics have advanced to the point that hundreds to thousands of proteins can be identified in a single run.[1,2] Improvements in sample preparation methods and development of state-of-the-art LC-MS/MS technologies enable deep profiling of post-translational modifications (PTMs), including phosphorylation, glycosylation, acetylation, methylation among others.[3–8] Despite these significant advances, routine identification and characterization of co-existing modifications remains an outstanding challenge.[9,10] Characterization of multiply modified proteins is critical to facilitate the study of PTM crosstalk, the phenomena where multiple modifications act in concert to modulate a biological outcome. [11,12] Given that the typical size range of peptides generated from conventional bottom-up proteolytic methods is 1-2 kDa, multiple peptides must be tracked and re-assembled to indirectly piece together the PTM landscapes of the parent proteins.[13,14]

In order to directly observe and map combinatorial PTMs, a number of groups have developed mass spectrometry approaches to characterize larger peptides (middle-down) or intact proteins (top-down), thus facilitating simultaneous characterization of all PTMs in a key region of a protein or in the entire protein sequence.[15–18] Middle-down strategies use enzymatic or chemical procedures which limit the extent of protein digestion, thus producing peptides that are typically much larger than those generated in conventional

bottom-up workflows.[19] Top-down and middle-down mass spectrometry methods require the use of mass spectrometers that offer both high resolving power and high mass accuracy to allow the confident identification of fragment ions, typically that are large and multi-charged.[18,20,21] As the size of the peptides or proteins increases, assignment of the fragment ions becomes more challenging, especially as the signal-to-noise (S/N) of the spectra diminishes as the ion current is dispersed among more fragmentation channels. Another challenge is the significant demand of data processing and analysis, though recently new software packages have become widely available to handle this burden.[22–24] For cases in which the targeted proteins of greatest interest are relatively small or when the use of proteases might obscure the presence of endogenously truncated proteins that are indicative of certain biological scenarios or activities, then the top-down method offers substantial advantages [18,21,24]. Analysis of intact proteins offers the potential for characterization of the complete PTM landscape not confounded by possible truncations. The ability to create confident maps of the sites of PTMs arises from extensive sequence coverage and minimization of the loss of labile modifications, two critical factors of any ion activation method utilized for MS/MS characterization of proteins. Neutral losses can be advantageous for determining the presence or absence of certain PTMs and differentiating similar mass PTM's such as acetylation and trimethylation on peptides.[25] However, neutral losses from intact proteins is less feasible for identifying and localizing PTMs owing to the possibility of multiple modifications and multiple locations, thus obscuring the ability to pinpoint specific sites.

Histones are particularly well-suited for analysis as intact proteins as they are generally less than 25 kDa, an ideal size for top-down methods. Histones have been implicated in the regulation of transcription, cell cycle progression, and DNA damage repair through dynamic variations in their PTM repertoire, making characterization of the various combinations of great interest.[26–30] This dynamic modification landscape has been termed the "histone code" and is often very complex, involving multiple PTMs across multiple histone subunits. These codes enable unique recognition sites for complexes of chromatin readers, benchmarking DNA readout and other functions. In particular, histones are heavily modified on the first 50 amino acid residues of the N-terminus, and these information-rich tails are key features that are known to mediate the structure and functions of nucleosomes. The N-terminal stretches have unusually high frequencies of lysine and arginine residues, making them prone to excessive degradation by trypsin in conventional bottom-up mass spectrometric approaches. Although traditional antibody-based ELISA analyses are purportedly specific, targeting just one or a limited number of PTMs, often they result in false negatives and false positives and can be biased towards heavily modified proteoforms.[20,20,30,31] These factors have motivated the development of other mass spectrometry-based methods for histones, including middle-down and top-down approaches.[9,15,20,26,32–38]

Mass spectrometry of intact proteins has been pursued more widely in recent year owing to the increasing accessibility of high resolving power mass spectrometers.[21] Activation of multi-charged proteins can generate hundreds of fragment ions (i.e a,b,c,x,y,

and z type ions) in multiple charge states, thus reinforcing the need for high performance capabilities of the mass spectrometer.[21,35] In the context of top-down MS/MS analysis histones, both collisional and electron-based activation methods have been evaluated.[33,34] Traditional collisional activation methods have not proven to be ideal because the large number of positives charges on multi-protonated histones results in sparse fragmentation, and labile PTMs are commonly lost during collisional activation.[38] Electron-based dissociation methods such as electron transfer dissociation (ETD) and electron capture dissociation (ECD) are more well suited for histones owing to the ability to retain modifications and fragment even very basic proteins.[36,39] Electron-based activation methods are dependent on the charge state of the protein, a factor that is not readily controlled for histones.[40,41] The ideal activation technique would cleave at every inter-residue position, allow localization of PTMs, and yield high performance metrics regardless of charge state. Three new methods have been introduced to meet these demands: electron transfer higher-energy collisional dissociation (EThcD),[42,43] activated ion electron transfer dissociation (AI-ETD),[44,45] and 193 nm ultraviolet photodissociation (UVPD).[46,47] EThcD is a hybrid method that combines two activation techniques, HCD and ETD, to generate a greater array of fragment ion types (b,y and c,z type ions).[43] AI-ETD uses infrared laser irradiation during ETD as a form of supplemental activation to overcome charge state dependency and "ET-no-D" events which limit the extent of fragmentation and curtail sequence coverage.[45] UVPD utilizes absorption of energetic photons, typically 193 nm, to energize and dissociate proteins.[46–49] UVPD results in

production of an array of ion types, including a,b,c,x,y, and z types, and does not cause dislodgement of labile PTMs.[46–48] The performance of UVPD for high throughput analysis of intact proteins has been reported in two recent studies.[47,50] The present study focuses on the use of three activation techniques, EThcD, UVPD, and traditional HCD, to characterize a series of histones and evaluate the performance metrics of these activation methods.

8.3 EXPERIMENTAL

8.3.1 Materials and Methods

Lyophilized calf thymus histone extract was purchased from Sigma Aldrich (St. Louis, MO). LC-MS solvents were purchased from EMD-Millipore (Bilerica, MA).

8.3.2 LC-MS

The histone extract proteins were resuspended in 2% ACN and separated using a Dionex RSLC 3000 nano-LC system (Thermo Fisher, San Jose, CA, USA). Approximately 1 μg of extract was injected onto a 3 cm PLRP-S (5 μm particles, 1000 Å pore size, Agilent, Santa Clara CA) reverse phase trapping column (100 μm i.d.) packed with 5 μm particles (1000 Å pore size). Proteins were then transferred onto a 30 cm fritted (75 μm i.d.) pulled tip analytical column (New Objective, Woburn, MA) packed in-house with PLRP-S. Proteins were eluted at a flow rate of 300 nL/min using the following gradient: starting at 2% B, going to 10% B at 8 min, then to 23% B at 10 min, and finally to 40% B at 130 min.

Mobile phase A was water with 0.1% formic acid and mobile phase B was acetonitrile with 0.1% formic acid. Examples of chromatographic traces are shown in **Figure 8.1.**

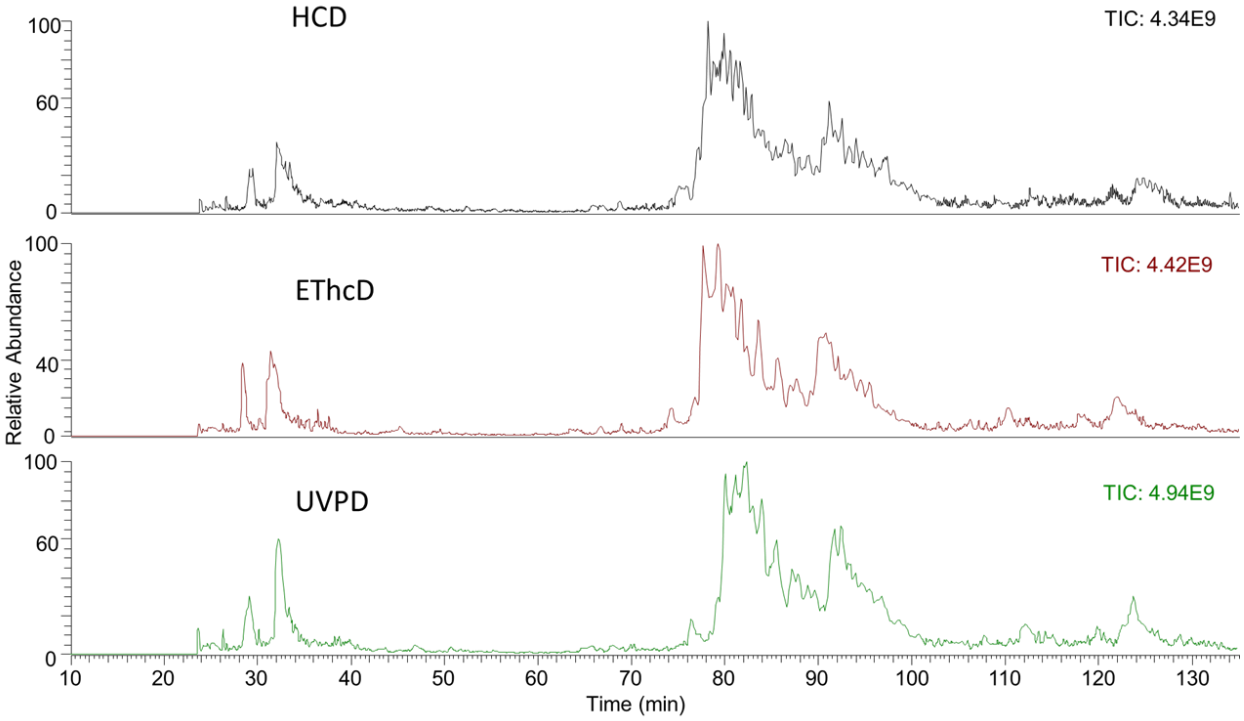


Figure 8.1 Examples of nano LC traces for HCD, EThcD and UVPD experiments illustrating the level of reproducibility.

The nanoLC system was coupled to a Fusion Lumos Orbitrap mass spectrometer (Thermo Scientific Instruments) modified for 193 nm UVPD, as previously described.[51] UVPD was performed in the high-pressure trap using a single pulse (1.7 mJ) from a 193 nm excimer laser (Coherent Excistar XS). HCD was performed in the ion routing multipole at 20% NCE. EThcD was performed in the ion trap (ECD) and ion routing multiple (HCD) with a 6 ms reaction time for ETD and 12% NCE for HCD, based on optimized conditions reported by Brunner et al.[42] The Orbitrap mass spectrometer was run using the following

parameters regardless of fragmentation type: MS1 at 120000 resolving power, 3 μscans averaged per spectrum, 1×10^6 AGC target, 20 V source fragmentation, intact protein monoisotopic precursor selection; and MS2 at 120000 resolving power with top speed mode enabled for 7 s, 6 μscans averaged per spectrum, and 1×10^6 AGC target. Precursor ions in charge-states 8+ to 24+ were selected for activation; additional precursors with undetermined charge state were also selected for activation. Top-speed mode allows the user to select the length of time between sequential MS1 scans. During this period the instrument isolates and activates as many precursors as possible for acquisition of MS/MS spectra. Therefore, the number of MS2 spectra per MS1 spectrum is variable unlike a traditional Top X method where X isolation/activation events (MS/MS spectra) are performed per MS1 spectrum.

8.3.3 Data Analysis

Raw data were uploaded to the National Resource for Translational and Developmental Proteomics (NRTDP, Northwestern University, Evanston, IL TDPortal13 high-performance computing environment for analysis of high-throughput top-down proteomics data (available for academic collaborators at: http://nrtdp.northwestern.edu/tdportal-request/). Details of the search strategy have been previously reported by Cleland et al.[50] Briefly, MS1 spectra were first averaged using the cRAWler algorithm, followed by deconvolution to monoisotopic masses by using Xtract (Thermo Fisher Scientific). Processed data were then searched against the Bos Taurus database. All searches entailed a three-pronged strategy, each mode of which was first

defined for ProSight PTM 2.0.42. For convenience each unique PTM-protein combination discussed in this study is referred to as a proteoform.[52]

8.4 RESULTS AND DISCUSSION

Despite the growing body of work showing the utility of 193 nm UVPD for the identification and characterization of intact proteins, there has been relatively little focus on characterization of PTMs.[47,50,53] In order to establish the feasibility of PTM identification and localization using UVPD and benchmark UVPD against other methods, a mixtures of histones was analyzed by high resolution LC-MS/MS using three different activation methods: UVPD, HCD, and EThcD. The performance of each of the MS/MS methods was evaluated based on proteoform level metrics including the number of proteoforms identified, sequence coverage, and C-score, the latter representing characterization of the protein backbone and modifications as defined by the number of matched fragment ions that localize PTMs. Examples of the MS/MS spectra obtained by HCD, EThcD, and UVPD for one representative histone, H3K9me2K27me, and the companion sequence maps are shown in **Figure 8.2**.

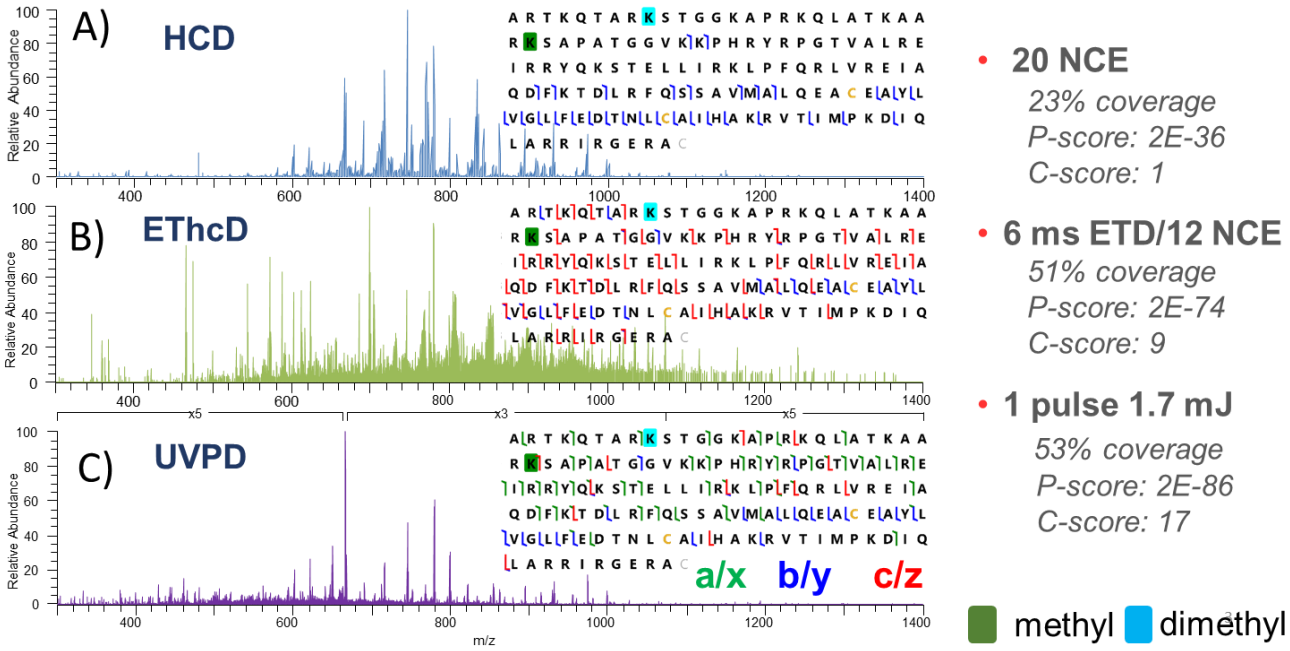


Figure 8.2 Examples of MS2 spectra for histone H3K9me2K27me (23+), sequence maps and score metrics for: A) HCD, B) EThcD and C) UVPD.

This proteoform was selected owing to its multiple modifications and its identification in the UVPD, EThcD and HCD datasets. A significant feature of the sequence maps, which are discussed in more detail later, is the striking difference in the types of fragment ions generated by UVPD (great diversity among a/x, b/y, and c/z) relative to HCD (all b/y) and EThcD (mostly c/z). While this study focuses exclusively on fragmentation metrics, the challenge posed by co-eluting isobaric proteoforms remains unsolved. In this report when multiple species are co-isolated and co-fragmented, the best PSM is reported. In the event of two equal scoring PSMs, both are reported.

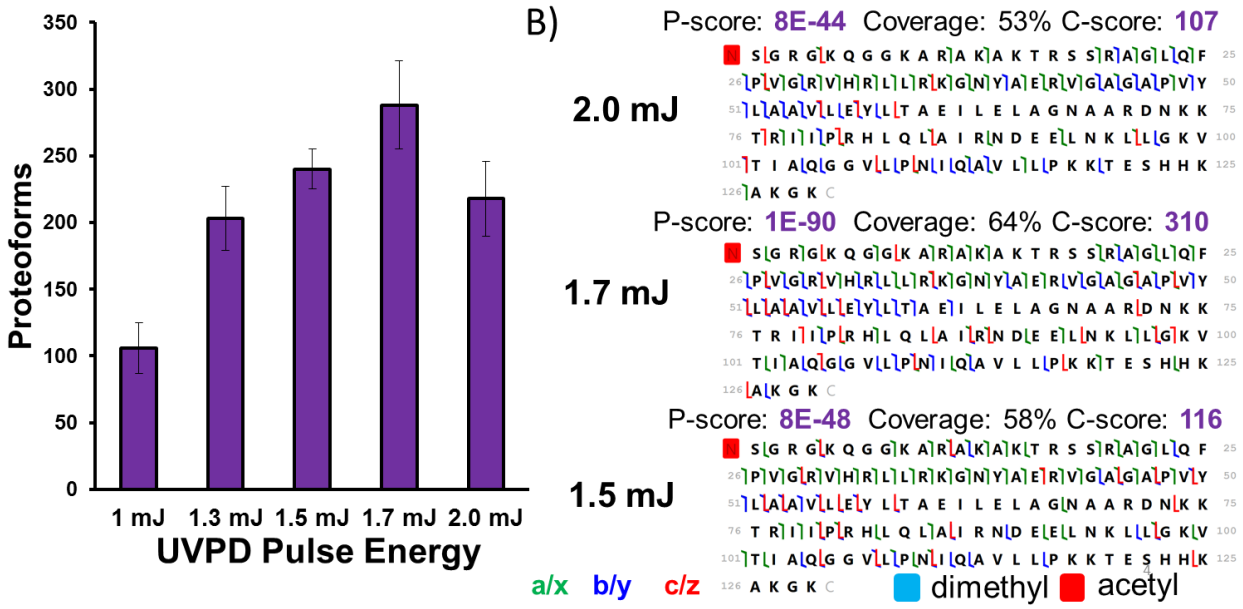
8.4.1 UVPD optimization

Absorption of even a single UV photon (193 nm in the present study) can activate ions into excited electronic states, resulting in access to many fragmentation channels. Both the laser flux (e.g., the power) and the number of consecutive pulses can be adjusted depending on the desired analytical outcome. For intact proteins generally a single lower energy laser pulse has been found to promote extensive backbone fragmentation and to yield the most informative spectra.[46,50,54] At the same time, using only a single pulse or a lower power pulse mitigates the prevalence of secondary fragmentation, a process which depletes the population of the larger fragments that are often most diagnostic and increases the production of smaller fragment ions and internal ions. Internal ions, which contain neither the C-terminus nor the N-terminus, are not assigned in conventional database searches owing to the exponentially larger search space. Because histones often have multiple modifications near the N-terminus, the impact of both the number of laser pulses and laser power on the total sequence coverage and localization of modifications near the N-terminus was evaluated in detail.

Figure 8.3A shows the number of proteoforms identified by UVPD using three laser powers ranging from 1.0 to 2.0 mJ and a single laser pulse. The optimal laser power was determined to be 1.7 mJ as it resulted in identification of 288 $\pm$ 33 (n=3) proteoforms, 48 more than obtained using 1.5 mJ. **Figure 8.3B** also displays examples of sequence maps and scoring metrics for one representative proteoform: N-terminal acetylated H2A obtained using a single UV laser pulse of 1.5 mJ, 1.7 mJ or 2.0 mJ. The P-score, C-score,

and sequence coverage were most favorable for the spectrum acquired using a single 1.7 mJ pulse, thus reflecting the importance of adjusting the laser power for optimal UVPD performance. The size range of the histones based on canonical sequence alone is relatively narrow (11.4 kDa to 20.8 kDa [based on the Uniprot entries for bovine histones]) which means that acquiring MS/MS spectra using a wide range of UVPD parameters is unlikely to return large gains in the number of identifications and level of coverage. **Figure 8.4A** illustrates the impact of using one or two pulses (1.7 mJ per pulse) on the number of proteoforms identified and average C-score obtained for the mixture of histones. For this comparison, using a single pulse resulted in identification of 306 proteoforms, approximately 50% more than obtained using two pulses. A similar level of enhancement was observed for the average C-scores obtained using a single pulse relative to two pulses. An example of the higher sequence coverage, P-score, and C-score obtained using a single pulse is shown for one mono-acetylated proteoform of H2A in **Figure 8.4B**, with a notable gain in sequence coverage near the N-terminal stretch that improved localization of the N-terminal acetylation.

Figure 8.3 A) Number of proteoforms identified using one laser pulse at different



energies. B) Examples of sequence coverage and proteoform metrics at various laser energies for H2A proteoform (19+).

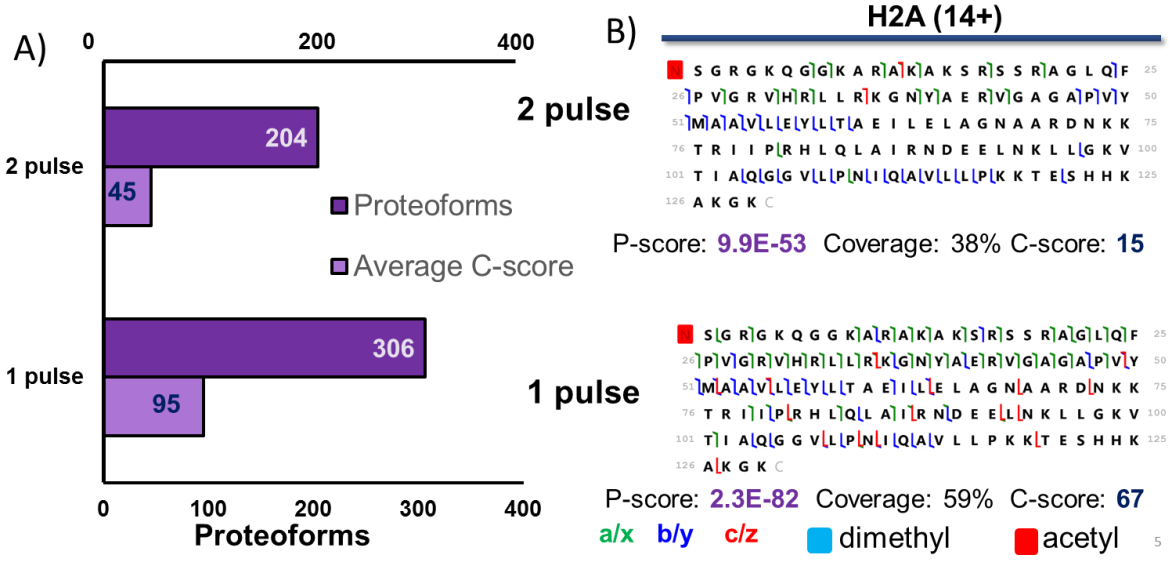


Figure 8.4 A) Results obtained for different numbers of laser pulses showing the number of identified proteoforms and average C-scores using one or two laser pulses at 1.7 mJ. B) Examples of sequence coverage and proteoform metrics obtained using one or two laser pulses (1.7 mJ).

8.4.2 Comparison of fragmentation of modified histone H3 by HCD, EThcD and UVPD

After optimization of the laser power and pulse number, UVPD was benchmarked against two other methods: the gold-standard collisional activation method HCD and another newly emerging hybrid activation technique EThcD. As mentioned earlier, **Figure 8.2** shows comparative MS2 spectra of histone H3K9me2K27me (23+) collected using the three activation methods. HCD yielded the lowest sequence coverage (23%) and generated few fragment ions around the two methylated lysines of the N-terminal region, leaving them ambiguously localized (**Figure 8.2A**). EThCD significantly extended the sequence coverage to 51%, however characterization of the two methylated lysines was only moderately improved relative to HCD. EThcD generated extensive fragmentation N-terminal to dimethylated K9 and C-terminal to methylated K27, but there was little fragmentation in the sequence stretch between these modified residues which impeded their localization (**Figure 8.2B**). UVPD of this H3 proteoform (**Figure 8.2C**) resulted in a net sequence coverage of 53%, similar to EThcD, yet with backbone cleavages occurring between the two methylated lysines which enhanced their localization as reflected by the higher C-score.

8.4.3 Global comparison of activation methods

A mixture of histones was separated and analyzed by LC-MS/MS (**Figure 8.5**) with each activation mode, HCD, EThcD or UVPD, run in triplicate to allow a more systematic comparison of the number of identified proteoforms, sequence coverage of proteoforms, and characterization of modified proteoforms for each activation method. **Figure 8.6a** shows the numbers and overlap of unique proteoforms identified by each method, and the numbers and overlap of unique proteoforms containing one or more modifications are compared in **Figure 8.6b**. HCD identified 321 proteoforms, slightly more than UVPD, an outcome attributed to its higher acquisition speed and efficiency at generating short "sequence-tag" stretches of proteoforms (see **Figure 8.7**). EThcD identified 278 unique proteoforms, whereas UVPD resulted in 313 identifications. Of the 530 total proteoforms identified, only 137 were found in common by all three methods, meaning that many unique proteoforms were identified by each method and recapitulating the complementary nature of the three MS/MS techniques. Considering only proteoforms containing modifications, **Figure 8.6b** shows that each MS/MS method yielded similar performance based solely on the number identified (ranging from 232 for EThcD to 262 for UVPD). For the 102 proteoforms identified uniquely by HCD, the average sequence coverage is 12%, average P-score is 3E-7, and average C-score is 10. For the 79 proteoforms uniquely identified by UVPD, the average sequence coverage is 23%, average P-score is 8E-10, and average C-score is 19. For the 65 proteoforms identified uniquely by EThcD, the average sequence coverage is 27%, average P-score is 3E-11, and C-score is 42.

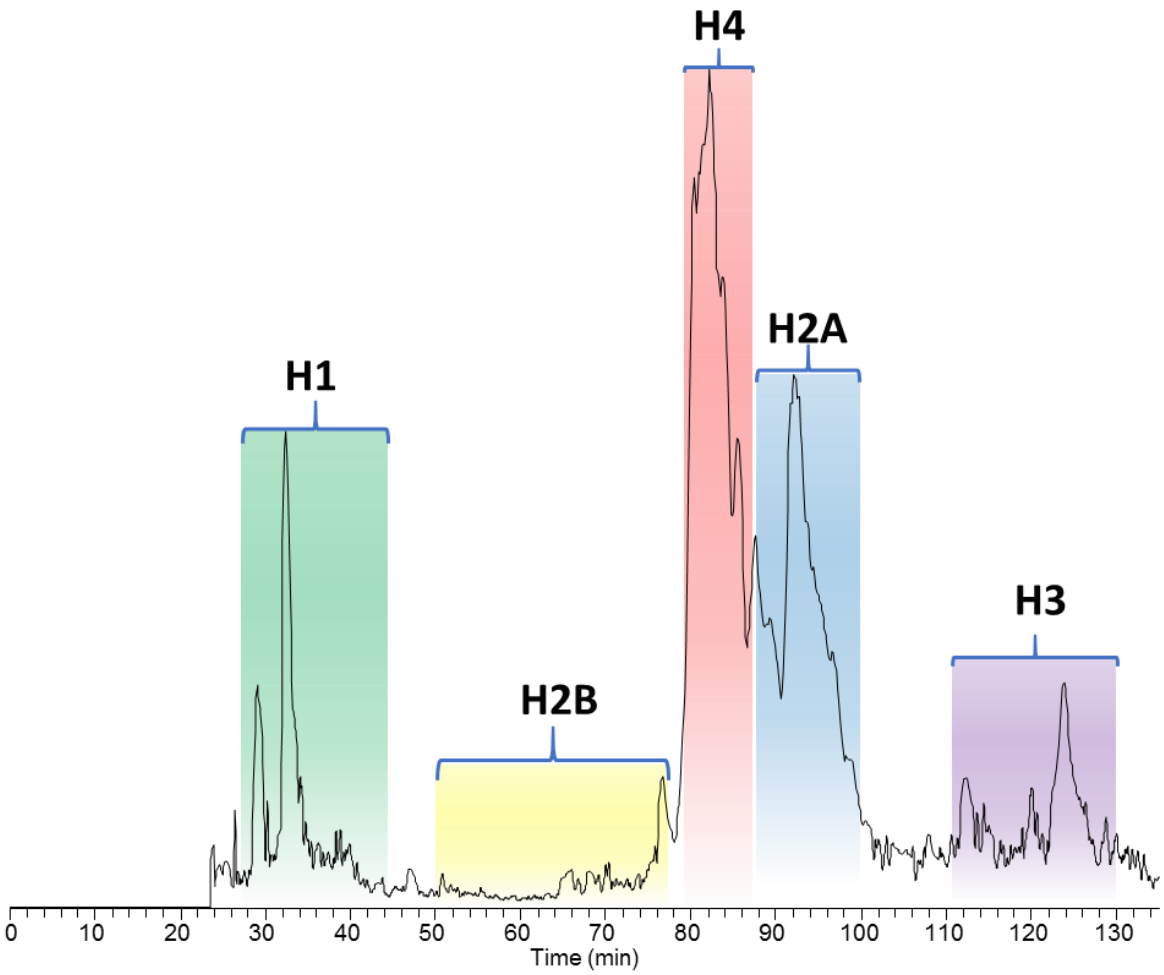


Figure 8.5 Typical liquid chromatogram of a mixture of intact calf histones separated on a 30 cm x 75 um column packed with PLRP-S stationary phase.

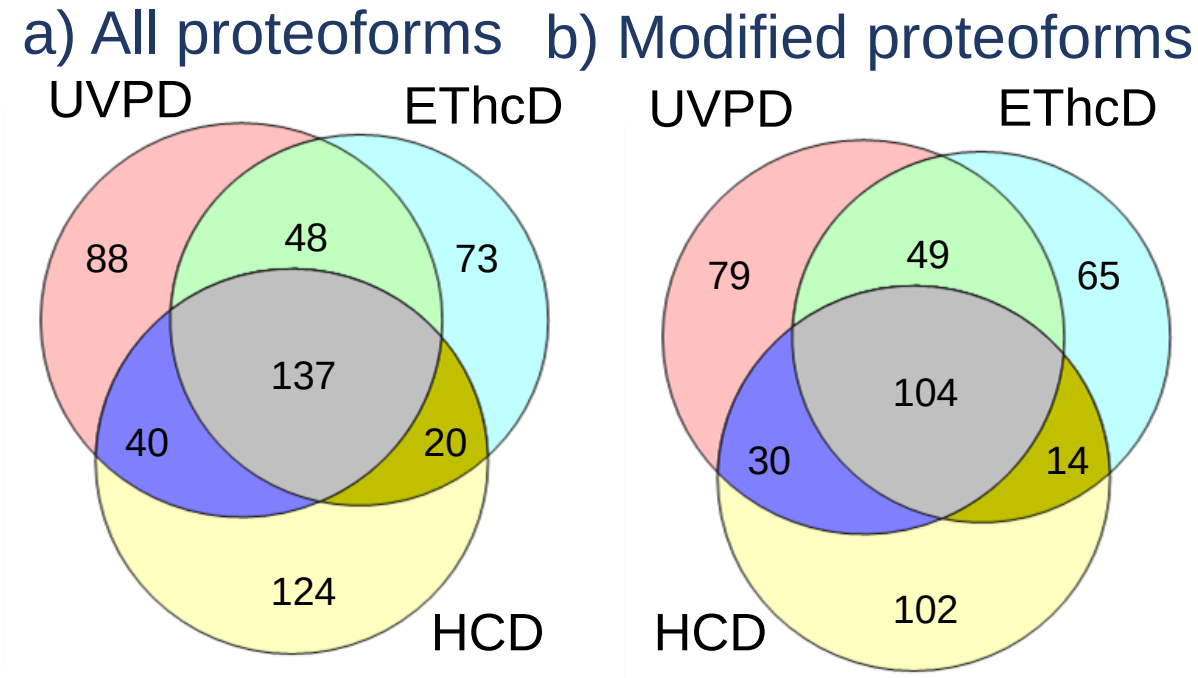


Figure 8.6 *Venn diagrams showing a) overlap of all proteoforms identified by UVPD (1 pulse 1.7 mJ), EThcD (6 ms reaction time for ETD and 12 NCE supplemental activation) and HCD (20 NCE), and b) overlap of modified proteoforms identified by UVPD, EThcD and HCD in combined triplicate runs filtered to 1% FDR.*

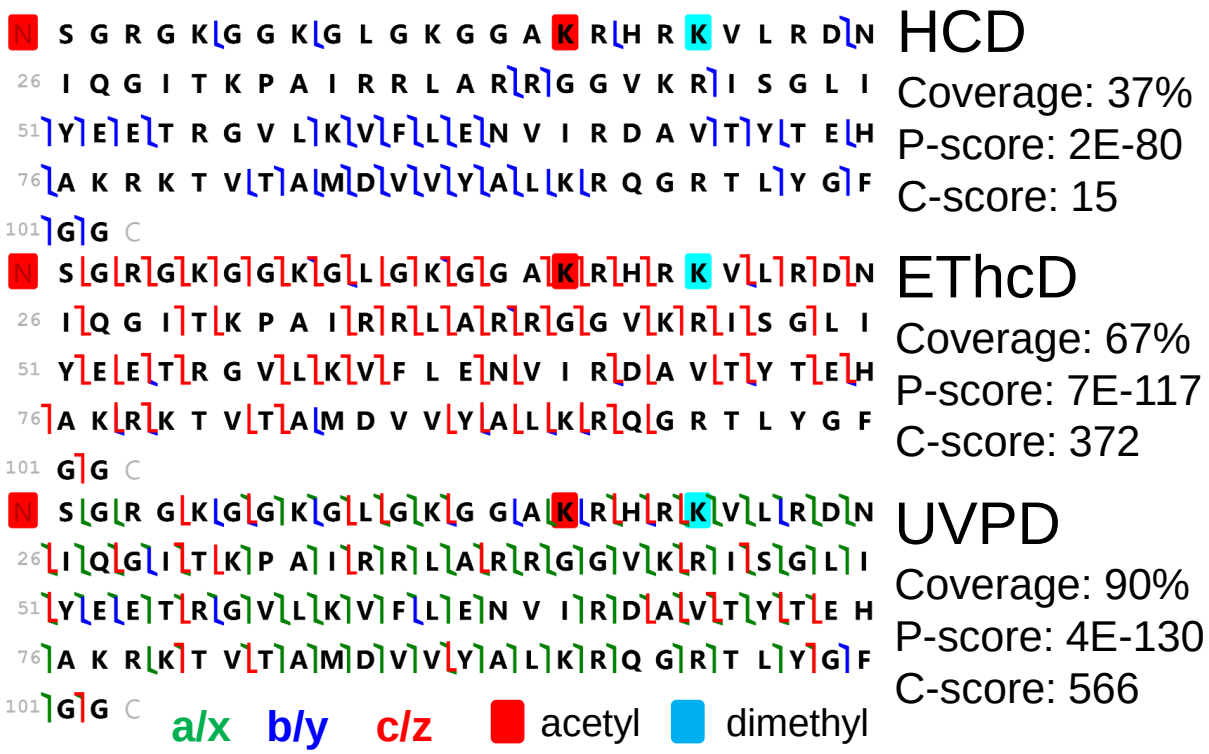


Figure 8.7 *Fragmentation maps, sequence coverage and scoring metrics of histone acH4K16acK20me2 (10+) obtained by HCD (20 NCE), EThcD (6 ms ETD reaction time and 12 NCE supplemental activation), and UVPD (1.7 mJ, 1 pulse).*

Despite the similar total number of identifications returned by each MS/MS method, the sequence maps and scoring metrics for one proteoform (acH4K16acK20me2, selected owing to its multiple modifications and positive identification by all three MS/MS methods) shown in **Figure 8.7** sheds light on the impact of the activation method in the context of PTM localization. The HCD sequence map displayed several short but well-sequenced stretches of the protein backbone which allowed unambiguous protein identification and resulted in a P-score of 2E-80 and sequence coverage of 37%. The

stretches of the adequately sequenced backbone occurred mainly along the C-terminal half of the protein, leaving the modifications of the N-terminus, K11 and K16 ambiguously localized (e.g., no fragment ions flanking the modified residues). This patchy coverage resulted in a low C-score of 15. EThcD showed improvement for several metrics, delivering 67% sequence coverage, and a significantly better P-score of 7E-117 and C-score of 372. Importantly, the localization of the modified residues was more confident upon EThcD owing to the fragment ions flanking acetylated K16, flanking ions within one residue of dimethylated K20, and the z_{101} and c_3 ions which effectively bracketed the acetylated N-terminus. UVPD exhibited a marked improvement for all metrics: 90% sequence coverage, with an impressive P-score of 4E-130 and C-score of 566. Although EThcD afforded good localization of modifications, UVPD generated key fragment ions that flanked all modified residues. In fact, owing to the unique ability to generate a/x, b/y, and c/z ions (**Figure 8.7**), UVPD produced five types of flanking ions that bracketed dimethylated K20 and afforded unparalleled confidence. As shown in **Figure 8.8** for histone acH2A (19+), one of the hallmarks of UVPD is the production of a greater array of fragment ion types compared to the more limited set for HCD (b/y) or EThcD (b/y, c/z). This greater array of fragment ions explains the higher sequence coverage and better characterization of PTMs (C-scores) often observed for UVPD, at the expense of S/N that may reduce the total number of PSMs. In this same vein, **Figure 8.9** shows the distributions of protein spectral matches (PSMs) relative to the number of identified fragment ions for the histone datasets generated for HCD, EThcD and UVPD. The majority of PSMs

obtained by HCD contained 70 or fewer fragment ions, whereas EThcD and UVPD generated PSMs based on more than 100 fragments per PSM which provided greater sequence coverage and enhanced localization of modifications.

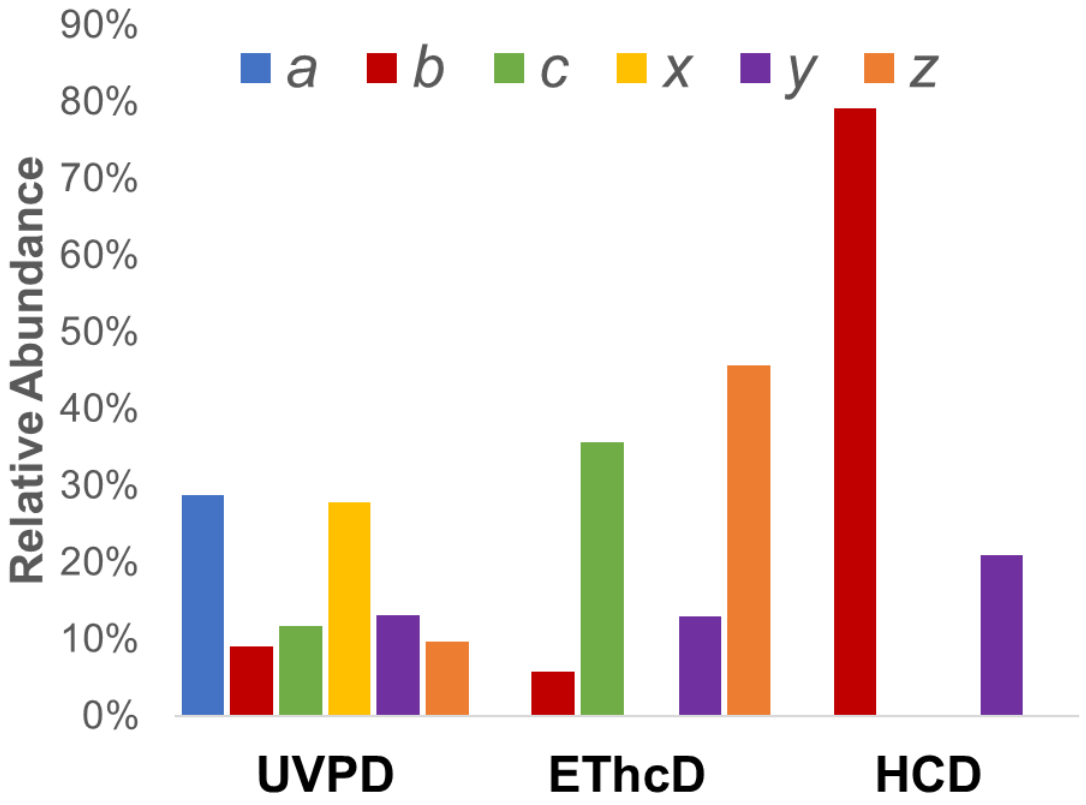


Figure 8.8 Typical distribution of fragment ion types generated by UVPD, EThcD and HCD of histone acH2A (19+).

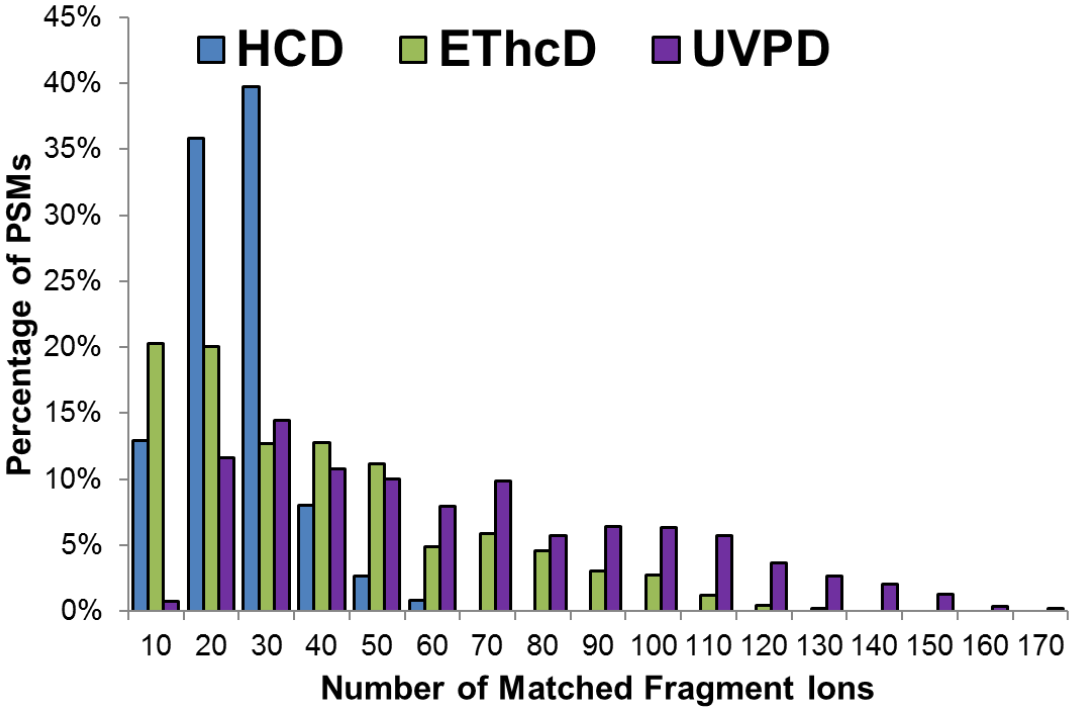


Figure 8.9 Histogram showing the percentage of all PSMs identified by HCD, EThcD or UVPD as a function of the number of matched diagnostic ions for all identified proteoforms.

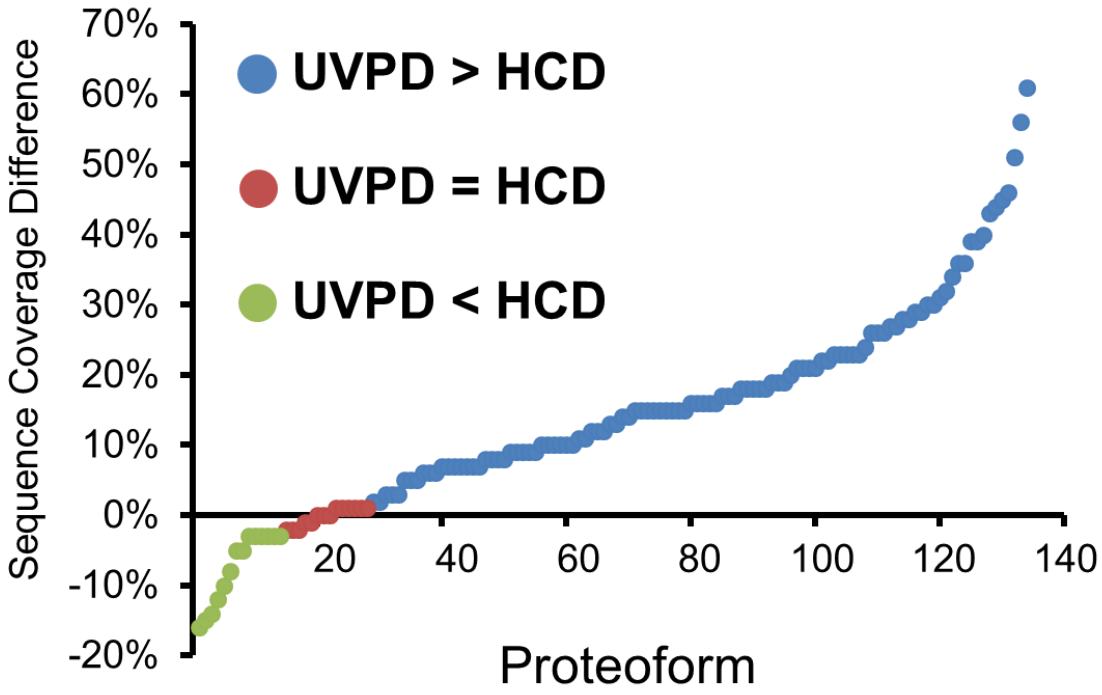


Figure 8.10 *Differences in sequence coverage of 134 modified proteoforms identified in common by HCD and UVPD from Figure 8.6b. Blue points indicate proteoforms for which UVPD generated greater sequence coverage than HCD (by more than 2%), green points indicate proteoforms for which HCD yielded greater sequence than UVPD (by more than 2%), and red points indicate proteoforms for which the sequence coverage differed by less than 2%. Proteoforms were ordered based on the difference in coverage. The activation conditions used were: HCD (20 NCE) and UVPD (1.7 mJ, 1 pulse).*

8.4.4 Sequence Coverage of Modified Proteoforms

Figure 8.10 compares sequence coverages generated by UVPD and HCD for the 134 modified proteoforms found in common by the two MS/MS methods. (**Figure 8.11** compares coverages for all 177 identified proteoforms). Each point represents the difference in sequence coverage between UVPD and HCD for a proteoform found in

common by each method (when a proteoform was identified multiple times, the highest sequence coverage was used). The proteoforms are ranked based on the difference in sequence coverage for HCD and UVPD and categorized for clarity. UVPD outperformed HCD for nearly 80% of the proteoforms found in common, in some cases afforded a gain of nearly 70% in coverage. A similar outcome was obtained when all 177 proteoforms (modified and unmodified) were compared (**Figure 8.11**).

C-scores of the same 134 proteoforms are displayed as a scatter plot in **Figure 8.12** in which the C-scores obtained by UVPD are plotted against the C-scores from HCD. A guiding line demarcates where the C-scores are equal for UVPD and HCD for the same proteoform. The majority of the points from the global histone dataset lie above the y = x guide-line, reflecting the enhanced ability of UVPD to characterize proteoforms relative to HCD, a result that echoes the single proteoform result shown in **Figure 8.7**.

A comparison of sequence coverages obtained by UVPD and EThcD for 153 modified proteoforms found in common is illustrated in **Figure 8.13**. The number of proteoforms is nearly equal above and below the 0% axis, indicating that both activation methods generate on average nearly identical sequence coverages for modified histone proteoforms. However, UVPD yields consistently higher sequence coverages by a wider margin for all of the cases in which UVPD outperforms EThcD. The average sequence coverage for all 153 proteoforms in **Figure 8.13** was 28±17% for UVPD versus 32±18% for EThcD (and 15±7% for HCD for comparison).

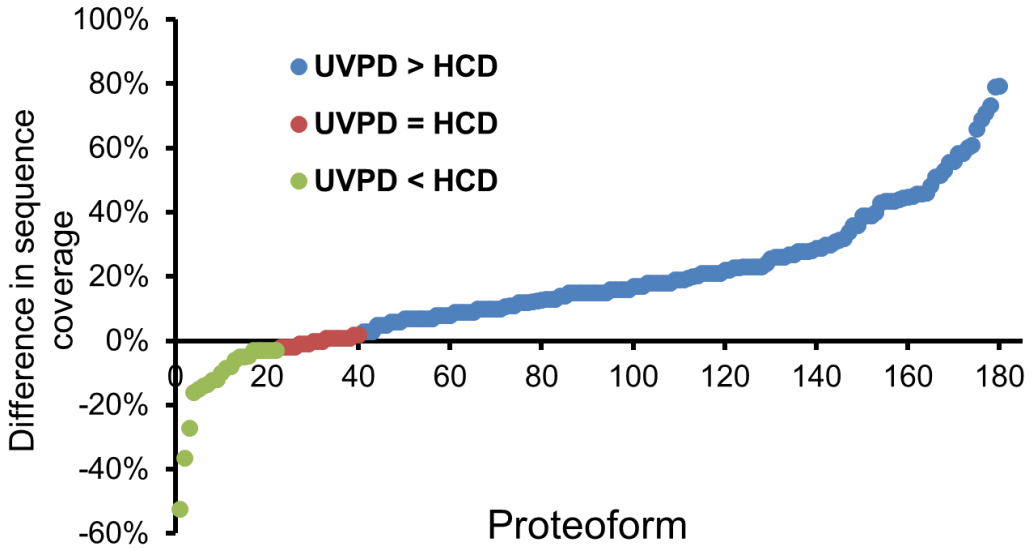


Figure 8.11 Difference in sequence coverage for the 177 proteoforms identified in common by HCD and UVPD from the Venn diagram in Figure 8.6. Each point represents the difference in sequence coverage between UVPD and HCD. Results are ordered based on the caliber of results for UVPD relative to HCD. All data points above the x-axis indicate better performance for UVPD.

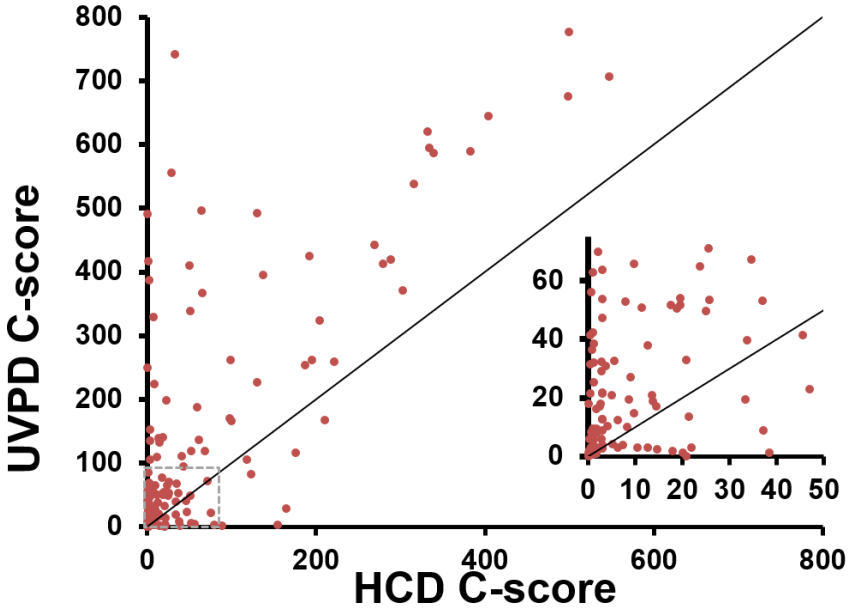


Figure 8.12 Scatter plot showing C-scores for 134 modified proteoforms identified in common by HCD and UVPD. A y = x line is plotted for reference. An expanded view of the densely populated region near the origin is shown in the inset.

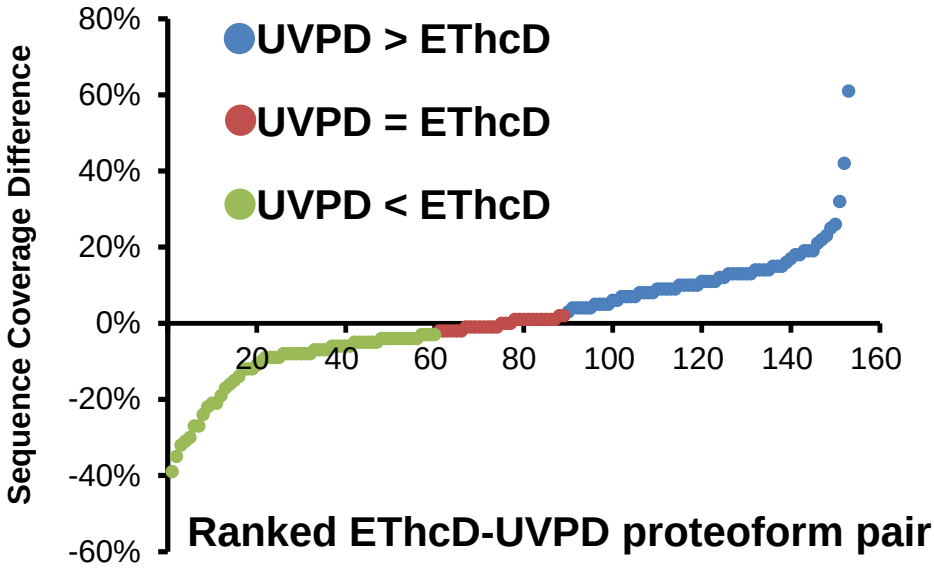


Figure 8.13. Differences in sequence coverage of 153 modified proteoforms identified in common by EThcD and UVPD from Figure 8.6b. Blue points indicate proteoforms for which UVPD generated greater sequence coverage than EThcD (by more than 2%), green points indicate proteoforms for which EThcD yielded greater sequence than UVPD (by more than 2%), and red points indicate proteoforms for which the sequence coverage differed by less than 2%. Proteoforms were ordered based on the difference in coverage. The activation conditions used were: EThcD (6 ms ETD reaction time and 12 NCE supplemental activation) and UVPD (1.7 mJ, 1 pulse).

To better understand the features of the proteoforms that differentiate the performance of EThcD and UVPD, a number of characteristics were considered, such as the size of the protein. The proteoforms were grouped into three bins based on molecular weight: 11-12 kDa, 12-14 kDa and 14-16 kDa. Core histones naturally fall into these rather narrow mass bins owing to their well-known molecular weights (H2A: 13.9 kDa, H2B: 13.7 kDa, H3: 15.3 kDa, H4: 11.2 kDa). **Figure 8.14** shows the box and whisker plots for sequence coverage of proteoforms obtained by EThcD and UVPD for the three mass bins.

For these plots, the height of the box conveys the range of values, the horizontal line within each box represents the median value, and the cross mark indicates the average value. For both activation methods, sequence coverage decreased as mass increased (**Figure 8.14**), an unsurprising trend considering the well-known mass dependence of top-down MS/MS proteomics methods. The performance of EThcD and UVPD for characterizing modifications as a function of protein size is summarized in **Figure 8.15**, again for the 153 modified proteoforms identified in common. For this plot positive values represent instances where UVPD generated a larger C-score (i.e. based on the difference in C-scores for UVPD and EThcD). **Figure 8.15** shows that as the proteoform mass and complexity increases (i.e. more modified forms), UVPD generally returns a higher C-score than EThcD, whereas the opposite outcome is true for proteoforms of lower mass. Proteoform mass is based on both the total number of residues in the protein plus additional modifications, thus the largest and most heavily modified proteoforms (generally histone H3 in the present study) fall into the largest mass bin, whereas the smaller and typically less heavily modified proteoforms (H4 and H2A) fall into the lower mass bins. In total, 77 proteoforms were identified by both EThcD and UVPD for H3, whereas only 38 were identified by both EThcD and UVPD for H2A and H4. This simple comparison of the number of identified proteoforms suggests that H3 has more combinatorial PTM variation which is consistent with prior reports [9,33,39]. The combined outcomes illustrated in **Figure 8.15** and **Figure 8.14** are interesting because it demonstrates that although sequence coverages are generally similar for UVPD and EThcD, the greater number of fragment ion

types generated by UVPD particularly increases confidence in localization of modifications for the more complex and larger histone proteoforms.

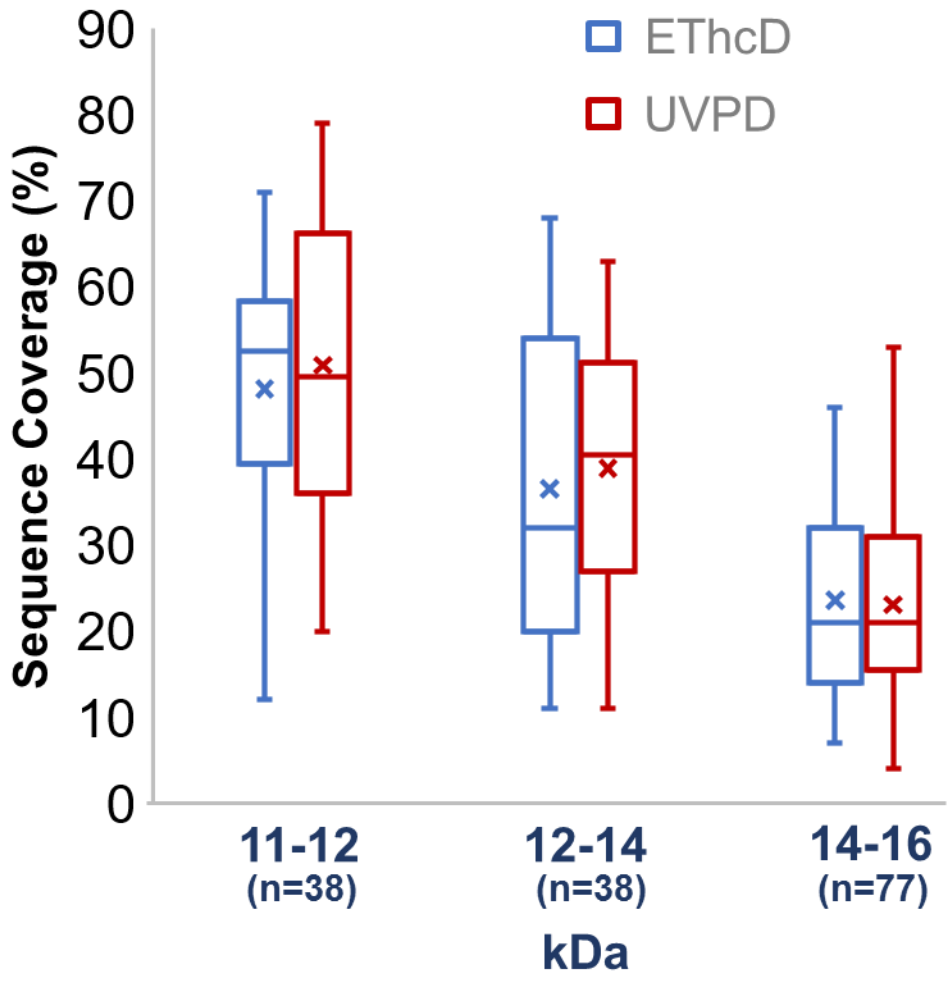


Figure 8.14 Box and whisker plot showing the sequence coverage of proteoforms common to UVPD and EThcD from Figure 8.6b. Results were sorted by mass to show impact of protein size.

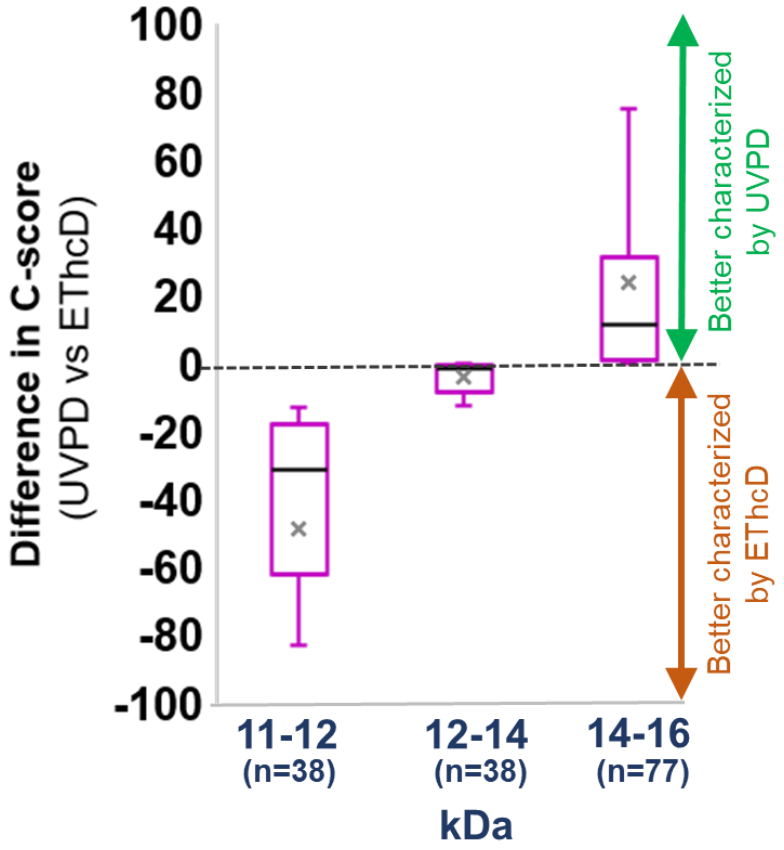


Figure 8.15 Box and whisker plot showing the difference in C-score of 153 proteoforms identified in common to UVPD and EThcD from Figure 8.6. Each bin contains the following number of proteoforms (11-12 kDa: 38, 12-14 kDa: 38, 14-16 kDa: 77)

The distribution of UVPD and EThcD C-scores for proteoforms modified 1 to 4 four times was generated for total identified by either method **Figure 8.16** and modified proteoforms found in common by both methods **Figure 8.17.** The C-scores were binned into three ranges (0-3, 3-40, >40) based on the principals laid out by Kelleher et al.[55] Briefly a proteoform with a C-score of 0-3 is poorly characterized, while a proteoform with a C-score of 3-40 is partially characterized, and a proteoform with C-score greater than 40

is considered well characterized. Both **Figure 8.16** and **Figure 8.17** show that as the number of modifications increase a greater percentage of proteoforms identified by UVPD garnered C-scores higher than 40 and UVPD identified more heavily modified proteoforms overall at higher numbers of modification (**Figure 8.16**).

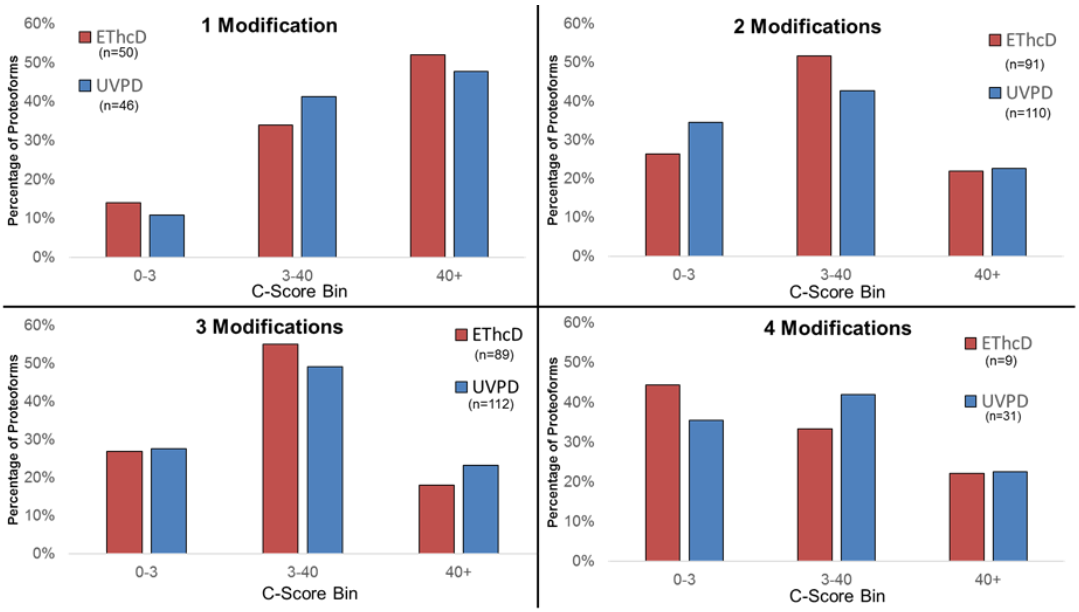


Figure 8.16 C-score distribution for all proteoforms identified by UVPD and EThcD with 1,2,3 or 4 modifications.

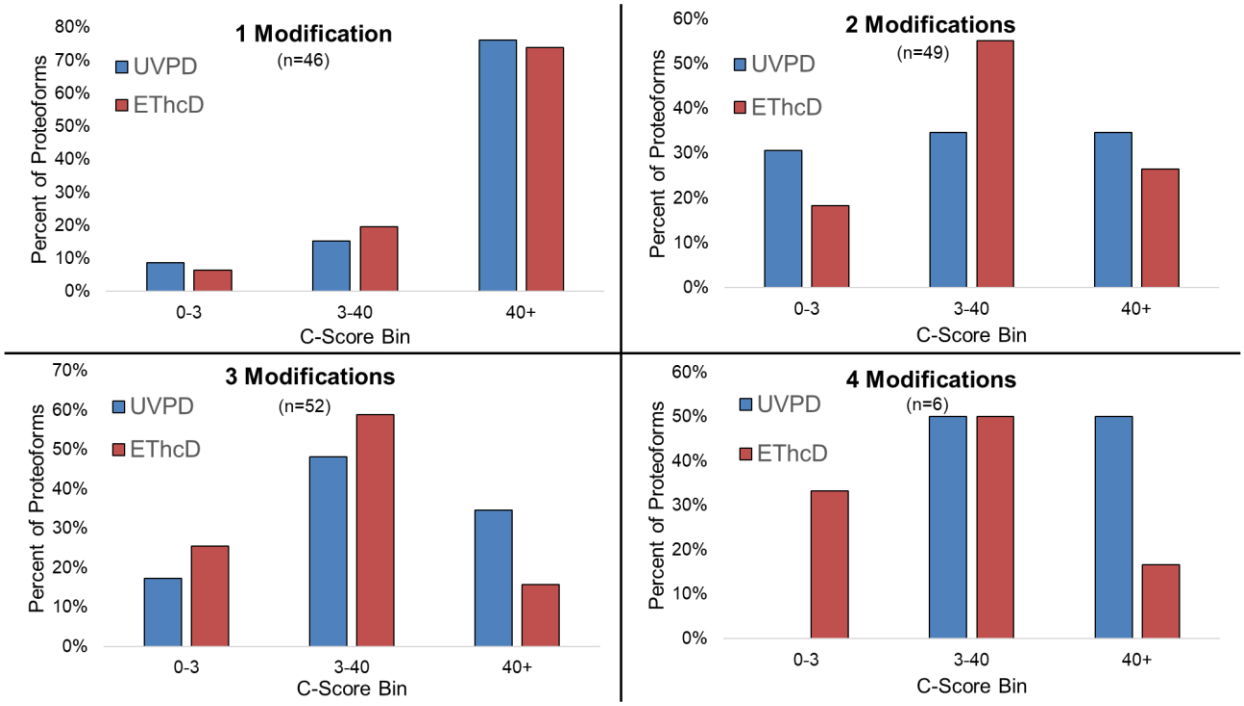


Figure 8.17 C-score distribution for proteoforms identified in common by UVPD and EThcD with 1,2,3 or 4 modifications.

Inspection of individual MS/MS spectra sheds light on the main reasons why HCD and EThcD outperform UVPD in some cases. In most instances in which HCD or EThcD significantly outperforms UVPD, the abundance of the precursor is low. The S/N of UVPD mass spectra is frequently lower than that of HCD or EThcD spectra for two reasons: (1) the production of a,b,c,x,y,z ions disperses the ion current among more channels, and (2) the laser power (photon flux) is kept relatively low to minimize secondary dissociation. Since the S/N of an MS/MS spectrum is related to the abundance of the precursor, it means that the overall performance of UVPD degrades for low abundance precursors owing to the inability to accurately deconvolve the resulting low abundance fragment ions. In cases in

which UVPD slightly underperforms HCD or UVPD, often HCD or EThcD generated a large series of fragment ions which entirely bracketed a particular modification, whereas UVPD only covered the modification with N-terminal or C-terminal ions but not both. In essence HCD or EThcD promoted more bidirectional fragmentation; some of the complementary fragment ions from UVPD might have been unassigned owing to low S/N and inadequate deconvolution of overly fragment-rich regions of the spectra.

8.5 CONCLUSIONS

Analysis of hundreds of histone proteoforms by HCD, EThcD and UVPD revealed the overall complementary nature of the three MS/MS methods based upon the large degree of shared identifications; however, the methods differ in their ability to characterize proteoforms. While HCD identified a greater number of proteoforms, EThcD and UVPD offered distinct advantages for histone analysis including greater sequence coverage (HCD: 15±7%, UVPD: 28±17%, EThcD: 32±18%) and proficiency for characterization, measured as average C-score (HCD: 34±78, UVPD: 76±149, EThcD 78±139) for modified histone proteoforms. EThcD and UVPD displayed a great degree of complementarity regarding the number of identified proteoforms and sequence coverage. The ability of these techniques to characterize proteoforms varied based on the mass of the proteoform: EThcD displayed enhanced characterization of smaller proteoforms, while UVPD resulted in better characterization of larger proteoforms.

8.6 REFERENCES

(1) Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A. M.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; et al. Mass-spectrometry-based draft of the human proteome. *Nature* **2014**, *509* (7502), 582–587.

(2) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The One Hour Yeast Proteome. *Mol. Cell. Proteomics* **2014**, *13* (1), 339–347.

(3) Sharma, K.; D'Souza, R. J.; Tyanova, S.; Schaab, C.; Wisniewski, J.; Cox, J.; Mann, M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Rep.* **2014**, *8* (5), 1583–1594.

(4) Villén, J.; Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* **2008**, *3* (10), 1630–1638.

(5) Li, Y.; Silva, J. C.; Skinner, M. E.; Lombard, D. B. Mass Spectrometry-Based Detection of Protein Acetylation. In *Sirtuins*; Hirschey, M. D., Ed.; Humana Press: Totowa, NJ, 2013; Vol. 1077, pp 81–104.

(6) Zhang, Y.; Song, L.; Liang, W.; Mu, P.; Wang, S.; Lin, Q. Comprehensive profiling of lysine acetylproteome analysis reveals diverse functions of lysine acetylation in common wheat. **2016**, *6*, 1–10.

(7) Carlson, S. M.; Moore, K. E.; Green, E. M.; Martin, G. M.; Gozani, O. Proteome-wide enrichment of proteins modified by lysine methylation. *Nat. Protoc.* **2013**, *9* (1), 37–50.

(8) Wang, K.; Dong, M.; Mao, J.; Wang, Y.; Jin, Y.; Ye, M.; Zou, H. Antibody-Free Approach for the Global Analysis of Protein Methylation. *Anal. Chem.* **2016**, *88* (23), 11319–11327.

(9) Zheng, Y.; Fornelli, L.; Compton, P. D.; Sharma, S.; Canterbury, J.; Mullen, C.; Zabrouskov, V.; Fellers, R. T.; Thomas, P. M.; Licht, J. D.; et al. Unabridged Analysis of Human Histone H3 by Differential Top-Down Mass Spectrometry Reveals Hypermethylated Proteoforms from MMSET/NSD2 Overexpression. *Mol. Cell. Proteomics* **2016**, *15* (3), 776–790.

(10) Yuan, Z.-F.; Arnaudo, A. M.; Garcia, B. A. Mass Spectrometric Analysis of Histone Proteoforms. *Annu. Rev. Anal. Chem.* **2014**, *7* (1), 113–128.

(11) Altelaar, A. F. M.; Munoz, J.; Heck, A. J. R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **2012**, *14* (1), 35–48.

(12) Peng, M.; Scholten, A.; Heck, A. J. R.; van Breukelen, B. Identification of Enriched PTM Crosstalk Motifs from Large-Scale Experimental Data Sets. *J. Proteome Res.* **2014**, *13* (1), 249–259.

(13) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. Six alternative proteases for mass spectrometry based proteomics beyond trypsin. *Nat. Protoc.* **2016**, *11* (5), 993–1006.

(14) Greer, S. M.; Parker, W. R.; Brodbelt, J. S. Impact of Protease on Ultraviolet Photodissociation Mass Spectrometry for Bottom-up Proteomics. *J. Proteome Res.* **2015**, *14* (6), 2626–2632.
(15) Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A. High Throughput Characterization of Combinatorial Histone Codes. *Mol. Cell. Proteomics* **2009**, *8* (10), 2266–2284.
(16) Phanstiel, D.; Brumbaugh, J.; Berggren, W. T.; Conard, K.; Feng, X.; Levenstein, M. E.; McAlister, G. C.; Thomson, J. A.; Coon, J. J. Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. *Proc. Natl. Acad. Sci.* **2008**, *105* (11), 4093–4098.
(17) Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. Quantitative Analysis of Modified Proteins and Their Positional Isomers by Tandem Mass Spectrometry: Human Histone H4. *Anal. Chem.* **2006**, *78* (13), 4271–4280.
(18) Moradian, A.; Kalli, A.; Sweredoski, M. J.; Hess, S. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *PROTEOMICS* **2014**, *14* (4–5), 489–497.
(19) Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W. P.; Mohammed, S.; Heck, A. J. R. Toward an Optimized Workflow for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89* (6), 3318–3325.
(20) Sidoli, S.; Garcia, B. A. Middle-down proteomics: a still unexploited resource for chromatin biology. *Expert Rev. Proteomics* **2017**, *14* (7), 617–626.
(21) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9* (1), 499–519.
(22) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; et al. Informed-Proteomics: open-source software package for top-down proteomics. *Nat. Methods* **2017**, *14* (9), 909–914.
(23) Kou, Q.; Xun, L.; Liu, X. TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **2016**.
(24) Fornelli, L.; Durbin, K. R.; Fellers, R. T.; Early, B. P.; Greer, J. B.; LeDuc, R. D.; Compton, P. D.; Kelleher, N. L. Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J. Proteome Res.* **2017**, *16* (2), 609–618.
(25) Zhang, K.; Yau, P. M.; Chandrasekhar, B.; New, R.; Kondrat, R.; Imai, B. S.; Bradbury, M. E. Differentiation between peptides containing acetylated or tri-methylated lysines by mass spectrometry: An application for determining lysine 9 acetylation and methylation of histone H3. *PROTEOMICS* **2004**, *4* (1), 1–10.
(26) Hauer, M. H.; Seeber, A.; Singh, V.; Thierry, R.; Sack, R.; Amitai, A.; Kryzhanovska, M.; Eglinger, J.; Holcman, D.; Owen-Hughes, T.; et al. Histone

degradation in response to DNA damage enhances chromatin dynamics and recombination rates. *Nat. Struct. Mol. Biol.* **2017**, *24* (2), 99–107.
(27) Kumar, R.; Horikoshi, N.; Singh, M.; Gupta, A.; Misra, H. S.; Albuquerque, K.; Hunt, C. R.; Pandita, T. K. Chromatin modifications and the DNA damage response to ionizing radiation. *Front. Oncol.* **2013**, *2*.
(28) Ewen, M. E. Where the cell cycle and histones meet. *Genes Dev.* **2000**, *14* (18), 2265–2270.
(29) Altheim, B. A.; Schultz, M. C. Histone modification governs the cell cycle regulation of a replication-independent chromatin assembly pathway in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (4), 1345–1350.
(30) Jenuwein, T. Translating the Histone Code. *Science* **2001**, *293* (5532), 1074–1080.
(31) Tanner, G. J.; Colgrave, M. L.; Blundell, M. J.; Goswami, H. P.; Howitt, C. A. Measuring Hordein (Gluten) in Beer – A Comparison of ELISA and Mass Spectrometry. *PLOS ONE* **2013**, *8* (2), e56452.
(32) Pesavento, J. J.; Kim, Y.-B.; Taylor, G. K.; Kelleher, N. L. Shotgun Annotation of Histone Modifications: A New Approach for Streamlined Characterization of Proteins by Top Down Mass Spectrometry. *J. Am. Chem. Soc.* **2004**, *126* (11), 3386–3387.
(33) Zhou, M.; Paša-Tolić, L.; Stenoien, D. L. Profiling of Histone Post-Translational Modifications in Mouse Brain with High-Resolution Top-Down Mass Spectrometry. *J. Proteome Res.* **2017**, *16* (2), 599–608.
(34) Tian, Z.; Tolić, N.; Zhao, R.; Moore, R. J.; Hengel, S. M.; Robinson, E. W.; Stenoien, D. L.; Wu, S.; Smith, R. D.; Paša-Tolić, L. Enhanced top-down characterization of histone post-translational modifications. *Genome Biol.* **2012**, *13* (10), R86.
(35) Dang, X.; Scotcher, J.; Wu, S.; Chu, R. K.; Tolic, N.; Ntai, I.; Thomas, P. M.; Fellers, R. T.; Early, B. P.; Zheng, Y.; et al. The first pilot project of the consortium for top-down proteomics: A status report. *PROTEOMICS* **2014**, *14* (10), 1130–1140.
(36) Siuti, N.; Roth, M. J.; Mizzen, C. A.; Kelleher, N. L.; Pesavento, J. J. Gene-Specific Characterization of Human Histone H2B by Electron Capture Dissociation. *J. Proteome Res.* **2006**, *5* (2), 233–239.
(37) Anderson, L. C.; Karch, K. R.; Ugrin, S. A.; Coradin, M.; English, A. M.; Sidoli, S.; Shabanowitz, J.; Garcia, B. A.; Hunt, D. F. Analyses of Histone Proteoforms Using Front-end Electron Transfer Dissociation-enabled Orbitrap Instruments. *Mol. Cell. Proteomics* **2016**, *15* (3), 975–988.
(38) Huang, H.; Lin, S.; Garcia, B. A.; Zhao, Y. Quantitative Proteomic Analysis of Histone Modifications. *Chem. Rev.* **2015**, *115* (6), 2376–2418.
(39) Jung, H. R.; Sidoli, S.; Haldbo, S.; Sprenger, R. R.; Schwimmle, V.; Pasini, D.; Helin, K.; Jensen, O. N. Precision Mapping of Coexisting Modifications in

Histone H3 Tails from Embryonic Stem Cells by ETD-MS/MS. *Anal. Chem.* **2013**, *85* (17), 8232–8239.
(40) Roman, M.; Gaskell, S. J. Charge state dependent top-down characterisation using electron transfer dissociation: Charge state dependent top-down characterisation using ETD. *Rapid Commun. Mass Spectrom.* **2012**, *26* (3), 282–286.
(41) Iavarone, A. T.; Paech, K.; Williams, E. R. Effects of Charge State and Cationizing Agent on the Electron Capture Dissociation of a Peptide. *Anal. Chem.* **2004**, *76* (8), 2231–2238.
(42) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. *Anal. Chem.* **2015**, *87* (8), 4152–4158.
(43) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2012**, *84* (22), 9668–9673.
(44) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal. Chem.* **2015**, *87* (14), 7109–7116.
(45) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. *J. Proteome Res.* **2017**, *16* (7), 2653–2659.
(46) 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. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135* (34), 12646–12651.
(47) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. *Anal. Chem.* **2014**, *86* (4), 2185–2192.
(48) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 nm and its Applicability to Proteomic Workflows. *J. Proteome Res.* **2010**, *9* (8), 4205–4214.
(49) Morrison, L. J.; Rosenberg, J. A.; Singleton, J. P.; Brodbelt, J. S. Statistical Examination of the a and a + 1 Fragment Ions from 193 nm Ultraviolet Photodissociation Reveals Local Hydrogen Bonding Interactions. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (9), 1443–1453.
(50) Cleland, T. P.; DeHart, C. J.; Fellers, R. T.; VanNispen, A. J.; Greer, J. B.; LeDuc, R. D.; Parker, W. R.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. *J. Proteome Res.* **2017**, *16* (5), 2072–2079.

(51) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2016**, *88* (1), 1044–1051.
(52) Smith, L. M.; Kelleher, N. L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla, L.; Kruppa, G.; Zubarev, R.; et al. Proteoform: a single term describing protein complexity. *Nat. Methods* **2013**, *10* (3), 186–187.
(53) Cotham, V. C.; Brodbelt, J. S. Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level using Middle-Down 193 nm Ultraviolet Photodissociation. *Anal. Chem.* **2016**, *88* (7), 4004–4013.
(54) Greer, S. M.; Holden, D. D.; Fellers, R.; Kelleher, N. L.; Brodbelt, J. S. Modulation of Protein Fragmentation Through Carbamylation of Primary Amines. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (8), 1587–1599.
(55) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. The C-Score: A Bayesian Framework to Sharply Improve Proteoform Scoring in High-Throughput Top Down Proteomics. *J. Proteome Res.* **2014**, *13* (7), 3231–3240.

Chapter 9

Conclusions

9.1 Conclusion

The growth of mass spectrometry based proteomics in the past decade is largely due to significant improvements in mass spectrometric instrumentation, tandem mass spectrometry methods, and database searching software. Along with these advancements, newer complex challenges such as intact protein analysis and analysis of PTMs have arisen, leaving much room for future improvements.[1,2] The ability of tandem MS to generate rich and informative fragmentation of difficult samples will continue to play a critical role in overcoming these challenges.[3] Therefore, continued development of novel fragmentation techniques such as UVPD which delivers an extensive and meaningful array of diagnostic product ions is crucial for the continued success of mass spectrometry based proteomics.[4]

In chapter 3, a simple, cheap, and highly efficient derivatization scheme (carbamylation) was utilized to block the basic charges found on lysine residues and the N-termini of peptides and enhance deprotonation. In turn these passivated peptides were analyzed in the negative polarity mode (deprotonating conditions) and showed enhanced charging and sensitivity. This method was used to enhance the traditionally underrepresented acidic proteome. After derivatization and negative mode UVPD analysis

a nearly 30% increase in peptide identifications was reported compared to an underivatized sample.

In chapter 4 peptide characteristics that directly influence the performance of two activation methods (HCD, UVPD) were compared on the basis of number of identified peptides, and sequence coverage. Overall HCD of basic tryptic peptides led to the greatest number of peptide identifications and greatest sequence coverage. However, when the peptides were longer (cleaved after only lysine residues) or had more acidic C-termini (cleaved after glutamic and aspartic acid) UVPD generated slightly better performance suggesting that UVPD is less dependent on mobile proton mediated dissociation. Additionally a very modest enhancement was noted for chromophore-bearing peptides (peptides having tryptophan, tyrosine and phenylalanine), reinforcing the role of photon absorption in the success of UVPD.[5]

Carbamylation was revisited in chapter 5 to investigate the role of protonation in HCD and UVPD dissociation of intact proteins and the effect on their chromatographic characteristics. Results from dissociation of six model proteins showed that UVPD was able to achieve excellent backbone sequencing of fully carbamylated proteins, comparable to the unmodified versions. HCD was unable to generate satisfactory dissociation and yielded poor sequence coverage for all modified proteins except for ubiquitin which was the smallest protein studied (8.5 kDa.) These results support the hypothesis that UVPD dissociation of intact proteins is not primarily driven by mobile protons and so is minimally influenced by precursor charge state and protein modifications. Additionally, upon LC

separation of unmodified and carbamylated *E. coli* ribosomal proteins two trends were noticed: longer retention due to greater hydrophobicity after modification and improved peak shape due to reduced electrostatic interaction with the stationary phase (via capping charged sites through carbamylation at lysine and the N-terminus).

Building on the established utility in chapter 1 of negative mode UVPD, in chapter 6 HCD, positive mode UVPD, and negative mode UVPD were collectively used to improve the depth and breadth of coverage for a cell lysate of hepatocyte cells. The combination of positive and negative mode UVPD increased the number of peptides and proteins identified by 25%. Additionally, for the first time a widely available database searching algorithm, Byonic, was trained to analyze both positive and negative mode UVPD data. Byonic expanded the number of identified peptides and proteins over currently available software by more than 15%

UVPD of large middle-down sized heavily modified histone peptides was investigated in chapter 7. Histone modifications are biologically important for gene regulation, generating a very complex epigenetic code. UVPD performed comparably to the currently adopted method of ETD for determination of the distribution of modifications along the backbone of histone H3. Importantly it was discovered that the current state of the art workflow, which was developed for ETD, only utilized approximately half of the information rich UVPD fragment ions (by abundance). Upon manual interpretation of the most heavily modified forms UVPD showed distinct advantages which were reflected by several performance metrics. Additionally, it was found that like other methods UVPD

generates informative neutral loss ions which can be used to discriminate between two nearly isobaric modifications (acetylation and trimethylation).

In chapter 8 shotgun UVPD analysis of intact histones was evaluated. When tested against HCD and EThcD, HCD identified the largest number of proteoforms. However, UVPD and EThcD were able to better sequence and characterize the identified proteoforms as measured by sequence coverage and C-score. Furthermore, UVPD better characterized the most heavily modified H3 proteoforms.

9.2 FUTURE WORK

Future work should focus on developing new UVPD methods for facilitating the analysis of middle-down sized and intact proteins which contain biologically relevant modifications. In particular effort should focus on training or developing software which takes full advantage of the rich array of ion types generated upon UVPD.[6]

One such class of proteins which would make an excellent candidate for UVPD analysis is glycoproteins. Glycoproteins serve many key biological functions such as protein-protein binding, receptor signaling and immune protection.[7,8] The glycosylation patterns associated with these functions often involve extensive modification of the host protein.[9] 193 nm UVPD could be used to enhance analysis of glycosylated proteins. The extensive fragmentation afforded by UVPD allows precise localization of glycosylation on the protein backbone as shown in **Figure 9.1** for avidin. Despite the impressive ability of UVPD to sequence the protein backbone and pinpoint glycosylations, enrichment

procedures and chromatographic methods also need to be developed to aid in the high-throughput analysis of glycopeptides and intact glycoproteins.[9,10]

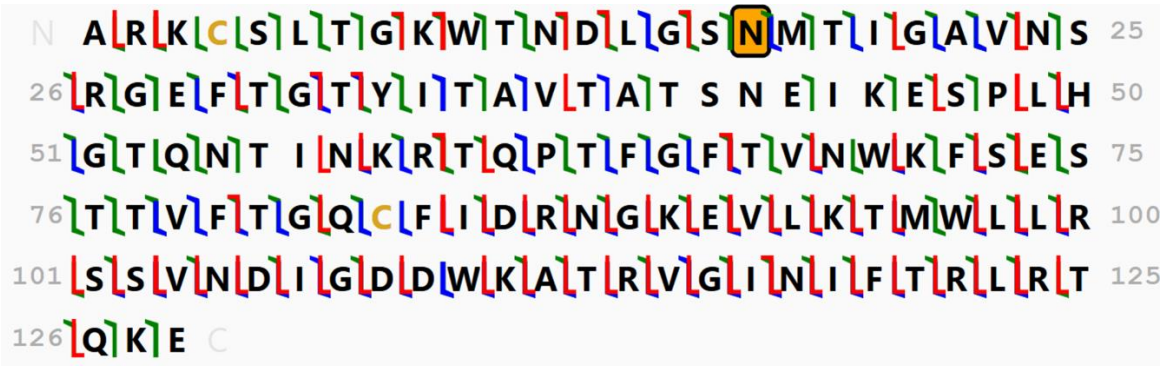


Figure 9.1 UVPD fragment ion map of avidin from egg white. UVPD was able to pinpoint the ($Man_5GlcNAc_4$) glycosylation to N41 with 4.72 ppm mass accuracy.

9.3 REFERENCES

(1) Peng, M.; Scholten, A.; Heck, A. J. R.; van Breukelen, B. Identification of Enriched PTM Crosstalk Motifs from Large-Scale Experimental Data Sets. *J. Proteome Res.* **2014**, *13* (1), 249–259.

(2) Venne, A. S.; Kollipara, L.; Zahedi, R. P. The next level of complexity: crosstalk of posttranslational modifications. *Proteomics* **2014**, *14* (4–5), 513–524.

(3) Dang, X.; Scotcher, J.; Wu, S.; Chu, R. K.; Tolic, N.; Ntai, I.; Thomas, P. M.; Fellers, R. T.; Early, B. P.; Zheng, Y.; et al. The first pilot project of the consortium for top-down proteomics: A status report. *PROTEOMICS* **2014**, *14* (10), 1130–1140.

(4) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88* (1), 30–51.

(5) Vasicek, L.; Brodbelt, J. S. Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores. *Anal. Chem.* **2010**, *82* (22), 9441–9446.

(6) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. Integration of Ultraviolet Photodissociation with Proton Transfer Reactions and Ion Parking for Analysis of Intact Proteins. *Anal. Chem.* **2016**, *88* (1), 1008–1016.

(7) Neue, K.; Mormann, M.; Peter-Katalinić, J.; Pohlentz, G. Elucidation of Glycoprotein Structures by Unspecific Proteolysis and Direct nanoESI Mass Spectrometric Analysis of ZIC-HILIC-Enriched Glycopeptides. *J. Proteome Res.* **2011**, *10* (5), 2248–2260.
(8) Yang, Y.; Liu, F.; Franc, V.; Halim, L. A.; Schellekens, H.; Heck, A. J. R. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* **2016**, *7*, 13397.
(9) Palmisano, G.; Lendal, S. E.; Engholm-Keller, K.; Leth-Larsen, R.; Parker, B. L.; Larsen, M. R. Selective enrichment of sialic acid–containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat. Protoc.* **2010**, *5* (12), 1974–1982.
(10) Mysling, S.; Palmisano, G.; Højrup, P.; Thaysen-Andersen, M. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. *Anal. Chem.* **2010**, *82* (13), 5598–5609.

References

Chapter 1

(1) Aebersold, R.; Mann, M. Mass-spectrometric exploration of proteome structure and function. Nature 2016, 537 (7620), 347–355.

(2) Bensimon, A.; Heck, A. J. R.; Aebersold, R. Mass Spectrometry–Based Proteomics and Network Biology. Annu. Rev. Biochem. 2012, 81 (1), 379–405.

(3) Yates, J. R. The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. J. Am. Chem. Soc. 2013, 135 (5), 1629–1640.

(4) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. Nature 2003, 422 (6928), 198–207.

(5) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization-principles and practice. Mass Spectrom. Rev. 1990, 9 (1), 37–70.

(6) Strupat, K.; Karas, M.; Hillenkamp, F. 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry. Int. J. Mass Spectrom. Ion Process. 1991, 111, 89–102.

(7) Ikonomou, M. G.; Blades, A. T.; Kebarle, P. Investigations of the electrospray interface for liquid chromatography/mass spectrometry. Anal. Chem. 1990, 62 (9), 957–967.

(8) Ling, V.; Guzzetta, A. W.; Canova-Davis, E.; Stults, J. T.; Hancock, W. S.; Covey, T. R.; Shushan, B. I. Characterization of the tryptic map of recombinant DNA derived tissue plasminogen activator by high-performance liquid chromatography-electrospray ionization mass spectrometry. Anal. Chem. 1991, 63 (24), 2909–2915.

(9) Tyers, M.; Mann, M. From genomics to proteomics. Nature 2003, 422 (6928), 193–197.

(10) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc. Natl. Acad. Sci. U. S. A. 1993, 90 (11), 5011–5015.

(11) Biemann, K. Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. Methods Enzymol. 1990, 193, 455–479.

(12) Olsen, J. V.; Mann, M. Status of Large-scale Analysis of Post-translational Modifications by Mass Spectrometry. Mol. Cell. Proteomics 2013, 12 (12), 3444–3452.

(13) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc. Natl. Acad. Sci. 2004, 101 (33), 12130–12135.

(14) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. Anal. Chem. 2015, 87 (8), 4152–4158.

(15) Humphrey, S. J.; Azimifar, S. B.; Mann, M. High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. Nat. Biotechnol. 2015, 33 (9), 990–995.

(16) Li, Y.; Silva, J. C.; Skinner, M. E.; Lombard, D. B. Mass Spectrometry-Based Detection of Protein Acetylation. In Sirtuins; Hirschey, M. D., Ed.; Humana Press: Totowa, NJ, 2013; Vol. 1077, pp 81–104.

(17) Zhang, Y.; Song, L.; Liang, W.; Mu, P.; Wang, S.; Lin, Q. Comprehensive profiling of lysine acetylproteome analysis reveals diverse functions of lysine acetylation in common wheat. 2016, 6, 1–10.

(18) Carlson, S. M.; Moore, K. E.; Green, E. M.; Martin, G. M.; Gozani, O. Proteome-wide enrichment of proteins modified by lysine methylation. Nat. Protoc. 2013, 9 (1), 37–50.

(19) Guo, A.; Gu, H.; Zhou, J.; Mulhern, D.; Wang, Y.; Lee, K. A.; Yang, V.; Aguiar, M.; Kornhauser, J.; Jia, X.; et al. Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. Mol. Cell. Proteomics 2014, 13 (1), 372–387.

(20) Wang, K.; Dong, M.; Mao, J.; Wang, Y.; Jin, Y.; Ye, M.; Zou, H. Antibody-Free Approach for the Global Analysis of Protein Methylation. Anal. Chem. 2016, 88 (23), 11319–11327.

(21) Altelaar, A. F. M.; Munoz, J.; Heck, A. J. R. Next-generation proteomics: towards an integrative view of proteome dynamics. Nat. Rev. Genet. 2012, 14 (1), 35–48.

(22) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The One Hour Yeast Proteome. Mol. Cell. Proteomics 2014, 13 (1), 339–347.

(23) Bern, M.; Kil, Y. J.; Becker, C. Byonic: advanced peptide and protein identification software. Curr. Protoc. Bioinforma. 2012, Chapter 13, Unit13.20.

(24) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. J. Proteome Res. 2011, 10 (4), 1794–1805.

(25) Frese, C. K.; Altelaar, A. F. M.; Hennrich, M. L.; Nolting, D.; Zeller, M.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. J. Proteome Res. 2011, 10 (5), 2377–2388.

(26) Shen, Y.; Tolić, N.; Xie, F.; Zhao, R.; Purvine, S. O.; Schepmoes, A. A.; Moore, R., J.; Anderson, G. A.; Smith, R. D. Effectiveness of CID, HCD, and ETD with FT MS/MS for Degradomic-Peptidomic Analysis: Comparison of Peptide Identification Methods. J. Proteome Res. 2011, 10 (9), 3929–3943.

(27) Thakur, S. S.; Geiger, T.; Chatterjee, B.; Bandilla, P.; Fröhlich, F.; Cox, J.; Mann, M. Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation. Mol. Cell. Proteomics 2011, 10 (8), M110.003699.

(28) Cristobal, A.; Hennrich, M. L.; Giansanti, P.; Goerdayal, S. S.; Heck, A. J. R.; Mohammed, S. In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. The Analyst 2012, 137 (15), 3541.

(29) Kay, R. G.; Gregory, B.; Grace, P. B.; Pleasance, S. The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum. Rapid Commun. Mass Spectrom. 2007, 21 (16), 2585–2593.

(30) Jorgenson, J. W. Capillary Liquid Chromatography at Ultrahigh Pressures. Annu. Rev. Anal. Chem. 2010, 3 (1), 129–150.

(31) Michalski, A.; Cox, J.; Mann, M. More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC−MS/MS. J. Proteome Res. 2011, 10 (4), 1785–1793.

(32) Xu, P.; Duong, D. M.; Peng, J. Systematical Optimization of Reverse-Phase Chromatography for Shotgun Proteomics. J. Proteome Res. 2009, 8 (8), 3944–3950.

(33) Ducret, A.; Oostveen, I. V.; Eng, J. K.; Yates, J. R.; Aebersold, R. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. Protein Sci. 1998, 7 (3), 706–719.

(34) Mahoney, W. C.; Hermodson, M. A. Separation of large denatured peptides by reverse phase high performance liquid chromatography. Trifluoroacetic acid as a peptide solvent. J. Biol. Chem. 1980, 255 (23), 11199–11203.

(35) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal. Bioanal. Chem. 2008, 391 (1), 151–159.

(36) McNulty, D. E.; Annan, R. S. Hydrophilic Interaction Chromatography Reduces the Complexity of the Phosphoproteome and Improves Global Phosphopeptide Isolation and Detection. Mol. Cell. Proteomics 2008, 7 (5), 971–980.

(37) Palmisano, G.; Lendal, S. E.; Engholm-Keller, K.; Leth-Larsen, R.; Parker, B. L.; Larsen, M. R. Selective enrichment of sialic acid–containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. Nat. Protoc. 2010, 5 (12), 1974–1982.

(38) Mysling, S.; Palmisano, G.; Højrup, P.; Thaysen-Andersen, M. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. Anal. Chem. 2010, 82 (13), 5598–5609.

(39) Neue, K.; Mormann, M.; Peter-Katalinić, J.; Pohlentz, G. Elucidation of Glycoprotein Structures by Unspecific Proteolysis and Direct nanoESI Mass Spectrometric Analysis of ZIC-HILIC-Enriched Glycopeptides. J. Proteome Res. 2011, 10 (5), 2248–2260.

(40) Alpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J. Chromatogr. A 1990, 499, 177–196.

(41) Washburn, M. P.; Wolters, D.; Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 2001, 19 (3), 242–247.

(42) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 1999, 17 (7), 676–682.

(43) Zhu, M.-Z.; Li, N.; Wang, Y.-T.; Liu, N.; Guo, M.-Q.; Sun, B.; Zhou, H.; Liu, L.; Wu, J.-L. Acid/Salt/pH Gradient Improved Resolution and Sensitivity in Proteomics Study Using 2D SCX-RP LC–MS. J. Proteome Res. 2017, 16 (9), 3470–3475.

(44) Zhang, L.; Liu, C.-W.; Zhang, Q. Online 2D-LC-MS/MS Platform for Analysis of Glycated Proteome. Anal. Chem. 2017.

(45) Wang, S.; Shi, X.; Xu, G. Online Three Dimensional Liquid Chromatography/Mass Spectrometry Method for the Separation of Complex Samples. Anal. Chem. 2017, 89 (3), 1433–1438.

(46) Navarro-Reig, M.; Jaumot, J.; Baglai, A.; Vivó-Truyols, G.; Schoenmakers, P. J.; Tauler, R. Untargeted Comprehensive Two-Dimensional Liquid Chromatography Coupled with High-Resolution Mass Spectrometry Analysis of Rice Metabolome Using Multivariate Curve Resolution. Anal. Chem. 2017, 89 (14), 7675–7683.

(47) Li, D.; Jakob, C.; Schmitz, O. Practical considerations in comprehensive two-dimensional liquid chromatography systems (LCxLC) with reversed-phases in both dimensions. Anal. Bioanal. Chem. 2015, 407 (1), 153–167.

(48) Mann, M.; Kelleher, N. L. Precision proteomics: The case for high resolution and high mass accuracy. Proc. Natl. Acad. Sci. 2008, 105 (47), 18132–18138.

(49) Capriotti, A. L.; Cavaliere, C.; Foglia, P.; Samperi, R.; Laganà, A. Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics. J. Chromatogr. A 2011, 1218 (49), 8760–8776.

(50) Shen, Y.; Tolić, N.; Piehowski, P. D.; Shukla, A. K.; Kim, S.; Zhao, R.; Qu, Y.; Robinson, E.; Smith, R. D.; Paša-Tolić, L. High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics. J. Chromatogr. A 2017, 1498, 99–110.

(51) Aguilar, M. I.; Hearn, M. T. High-resolution reversed-phase high-performance liquid chromatography of peptides and proteins. Methods Enzymol. 1996, 270, 3–26.

(52) Riley, N. M.; Mullen, C.; Weisbrod, C. R.; Sharma, S.; Senko, M. W.; Zabrouskov, V.; Westphall, M. S.; Syka, J. E. P.; Coon, J. J. Enhanced Dissociation of Intact Proteins with High Capacity Electron Transfer Dissociation. J. Am. Soc. Mass Spectrom. 2016, 27 (3), 520–531.

(53) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. Annu. Rev. Anal. Chem. 2016, 9 (1), 499–519.

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

(55) Smith, L. M.; Kelleher, N. L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla, L.; Kruppa, G.; Zubarev, R.; et al. Proteoform: a single term describing protein complexity. Nat. Methods 2013, 10 (3), 186–187.

(56) Huang, T.; Wang, J.; Yu, W.; He, Z. Protein inference: a review. Brief. Bioinform. 2012, 13 (5), 586–614.

(57) Armirotti, A.; Damonte, G. Achievements and perspectives of top-down proteomics. PROTEOMICS 2010, 10 (20), 3566–3576.

(58) Tian, Z.; Tolić, N.; Zhao, R.; Moore, R. J.; Hengel, S. M.; Robinson, E. W.; Stenoien, D. L.; Wu, S.; Smith, R. D.; Paša-Tolić, L. Enhanced top-down characterization of histone post-translational modifications. Genome Biol. 2012, 13 (10), R86.

(59) Moradian, A.; Kalli, A.; Sweredoski, M. J.; Hess, S. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. PROTEOMICS 2014, 14 (4–5), 489–497.

(60) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. The C-Score: A Bayesian Framework to Sharply Improve Proteoform Scoring in High-Throughput Top Down Proteomics. J. Proteome Res. 2014, 13 (7), 3231–3240.

(61) Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W. P.; Mohammed, S.; Heck, A. J. R. Toward an Optimized Workflow for Middle-Down Proteomics. Anal. Chem. 2017, 89 (6), 3318–3325.

(62) Sidoli, S.; Garcia, B. A. Middle-down proteomics: a still unexploited resource for chromatin biology. Expert Rev. Proteomics 2017, 14 (7), 617–626.

(63) Schräder, C. U.; Lee, L.; Rey, M.; Sarpe, V.; Man, P.; Sharma, S.; Zabrouskov, V.; Larsen, B.; Schriemer, D. C. Neprosin, a Selective Prolyl Endoprotease for Bottom-up Proteomics and Histone Mapping. Mol. Cell. Proteomics 2017, 16 (6), 1162–1171.

(64) Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C. High-Throughput Middle-Down Analysis Using an Orbitrap. J. Proteome Res. 2010, 9 (8), 3886–3890.

(65) Sidoli, S.; Schwämmle, V.; Ruminowicz, C.; Hansen, T. A.; Wu, X.; Helin, K.; Jensen, O. N. Middle-down hybrid chromatography/tandem mass spectrometry workflow for characterization of combinatorial post-translational modifications in histones. PROTEOMICS 2014, 14 (19), 2200–2211.

(66) Busman, M.; Schey, K. L.; Oatis, J. E.; Knapp, D. R. Identification of phosphorylation sites in phosphopeptides by positive and negative mode electrospray ionization-tandem mass spectrometry. J. Am. Soc. Mass Spectrom. 1996, 7 (3), 243–249.

(67) Robinson, M. R.; Moore, K. L.; Brodbelt, J. S. Direct Identification of Tyrosine Sulfation by using Ultraviolet Photodissociation Mass Spectrometry. J. Am. Soc. Mass Spectrom. 2014, 25 (8), 1461–1471.

(68) Doerr, A. Proteomics: Navigating the negative-mode proteome. Nat. Methods 2015, 12 (9), 808–808.

(69) Riley, N. M.; Rush, M. J. P.; Rose, C. M.; Richards, A. L.; Kwiecien, N. W.; Bailey, D. J.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. The Negative Mode Proteome with Activated Ion Negative Electron Transfer Dissociation (AI-NETD). Mol. Cell. Proteomics MCP 2015, 14 (10), 2644–2660.

(70) 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. High-throughput Database Search and Large-scale Negative Polarity Liquid Chromatography-Tandem Mass Spectrometry with Ultraviolet Photodissociation for Complex Proteomic Samples. Mol. Cell. Proteomics 2013, 12 (9), 2604–2614.

(71) Yamashita, M.; Fenn, J. B. Negative ion production with the electrospray ion source. J. Phys. Chem. 1984, 88 (20), 4671–4675.

(72) McClory, P. J.; Håkansson, K. Corona Discharge Suppression in Negative Ion Mode Nanoelectrospray Ionization via Trifluoroethanol Addition. Anal. Chem. 2017, 89 (19), 10188–10193.

(73) Madsen, J. A.; Ko, B. J.; Xu, H.; Iwashkiw, J. A.; Robotham, S. A.; Shaw, J. B.; Feldman, M. F.; Brodbelt, J. S. Concurrent Automated Sequencing of the Glycan and Peptide Portions of O -Linked Glycopeptide Anions by Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2013, 85 (19), 9253–9261.

(74) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. Anal. Chem. 2016, 88 (4), 2303–2310.

(75) 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. Analysis of the Acidic Proteome with Negative Electron-Transfer Dissociation Mass Spectrometry. Anal. Chem. 2012, 84 (6), 2875–2882.

(76) Ewing, N. P.; Cassady, C. J. Dissociation of multiply charged negative ions for hirudin (54–65), fibrinopeptide B, and insulin A (oxidized). J. Am. Soc. Mass Spectrom. 2001, 12 (1), 105–116.

(77) Griffiths, W. J.; Wang, Y. Mass spectrometry: from proteomics to metabolomics and lipidomics. Chem. Soc. Rev. 2009, 38 (7), 1882.

(78) Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. Electron transfer dissociation of peptide anions. J. Am. Soc. Mass Spectrom. 2005, 16 (6), 880–882.

(79) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. Anal. Chem. 2016, 88 (1), 30–51.

(80) Shukla, A. K.; Futrell, J. H. Tandem mass spectrometry: dissociation of ions by collisional activation. J. Mass Spectrom. 2000, 35 (9), 1069–1090.

(81) 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. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. J. Am. Chem. Soc. 2013, 135 (34), 12646–12651.

(82) Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L. Analyzing Internal Fragmentation of Electrosprayed Ubiquitin Ions During Beam-Type Collisional Dissociation. J. Am. Soc. Mass Spectrom. 2015, 26 (5), 782–787.

(83) Zhang, Z.; Wu, S.; Stenoien, D. L.; Paša-Tolić, L. High-Throughput Proteomics. Annu. Rev. Anal. Chem. 2014, 7 (1), 427–454.

(84) Brodbelt, J. S. Photodissociation mass spectrometry: new tools for characterization of biological molecules. Chem Soc Rev 2014, 43 (8), 2757–2783.

(85) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. Mol. Cell. Proteomics 2007, 6 (11), 1942–1951.

(86) Diedrich, J. K.; Pinto, A. F. M.; Yates, J. R. Energy Dependence of HCD on Peptide Fragmentation: Stepped Collisional Energy Finds the Sweet Spot. J. Am. Soc. Mass Spectrom. 2013, 24 (11), 1690–1699.

(87) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. J. Mass Spectrom. 2000, 35 (12), 1399–1406.

(88) Bythell, B. J.; Suhai, S.; Somogyi, Á.; Paizs, B. Proton-Driven Amide Bond-Cleavage Pathways of Gas-Phase Peptide Ions Lacking Mobile Protons. J. Am. Chem. Soc. 2009, 131 (39), 14057–14065.

(89) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. J. Am. Chem. Soc. 1996, 118 (35), 8365–8374.

(90) Laskin, J.; Kong, R. P. W.; Song, T.; Chu, I. K. Effect of the basic residue on the energetics and dynamics of dissociation of phosphopeptides. Int. J. Mass Spectrom. 2012, 330–332, 295–301.

(91) Douglas, D. J.; Frank, A. J.; Mao, D. Linear ion traps in mass spectrometry. Mass Spectrom. Rev. 2005, 24 (1), 1–29.

(92) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; et al. A Dual

Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. Mol. Cell. Proteomics 2009, 8 (12), 2759–2769.

(93) March, R. E.; Todd, J. F. Quadrupole Ion Trap Mass Spectrometry; John Wiley & Sons, 2005.

(94) Bowie, J. H.; Brinkworth, C. S.; Dua, S. Collision-induced fragmentations of the (M-H)- parent anions of underivatized peptides: An aid to structure determination and some unusual negative ion cleavages. Mass Spectrom. Rev. 2002, 21 (2), 87–107.

(95) Bokatzian-Johnson, S. S.; Stover, M. L.; Dixon, D. A.; Cassady, C. J. A Comparison of the Effects of Amide and Acid Groups at the C-Terminus on the Collision-Induced Dissociation of Deprotonated Peptides. J. Am. Soc. Mass Spectrom. 2012, 23 (9), 1544–1557.

(96) Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. Influence of Secondary Structure on the Fragmentation of Protonated Peptides. J. Am. Chem. Soc. 1999, 121 (22), 5142–5154.

(97) Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R. Influence of Basic Residue Content on Fragment Ion Peak Intensities in Low-Energy Collision-Induced Dissociation Spectra of Peptides. Anal. Chem. 2004, 76 (5), 1243–1248.

(98) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc. Natl. Acad. Sci. 2007, 104 (7), 2193–2198.

(99) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, Ü. A.; Tsybin, Y. O. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. Chem. Soc. Rev. 2013, 42 (12), 5014.

(100) Riley, N. M.; Coon, J. J. The Role of Electron Transfer Dissociation in Modern Proteomics. Anal. Chem. 2017.

(101) Shaw, J. B.; Kaplan, D. A.; Brodbelt, J. S. Activated Ion Negative Electron Transfer Dissociation of Multiply Charged Peptide Anions. Anal. Chem. 2013, 85 (9), 4721–4728.

(102) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors. Anal. Chem. 2007, 79 (2), 477–485.

(103) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. Anal. Chem. 2015, 87 (14), 7109–7116.

(104) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. J. Proteome Res. 2017, 16 (7), 2653–2659.

(105) Ledvina, A. R.; Rose, C. M.; McAlister, G. C.; Syka, J. E. P.; Westphall, M. S.; Griep-Raming, J.; Schwartz, J. C.; Coon, J. J. Activated Ion ETD Performed in a Modified Collision Cell on a Hybrid QLT-Oribtrap Mass Spectrometer. J. Am. Soc. Mass Spectrom. 2013, 24 (11), 1623–1633.

(106) Reilly, J. P. Ultraviolet photofragmentation of biomolecular ions. Mass Spectrom. Rev. 2009, 28 (3), 425–447.

(107) R. Julian, R. The Mechanism Behind Top-Down UVPD Experiments: Making Sense of Apparent Contradictions. J. Am. Soc. Mass Spectrom. 2017, 28 (9), 1823–1826.

(108) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 nm and its Applicability to Proteomic Workflows. J. Proteome Res. 2010, 9 (8), 4205–4214.

(109) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-nm photodissociation of singly and multiply charged peptide anions for acidic proteome characterization. PROTEOMICS 2011, 11 (7), 1329–1334.

(110) Greer, S. M.; Holden, D. D.; Fellers, R.; Kelleher, N. L.; Brodbelt, J. S. Modulation of Protein Fragmentation Through Carbamylation of Primary Amines. J. Am. Soc. Mass Spectrom. 2017, 28 (8), 1587–1599.

(111) Lesur, A.; Domon, B. Advances in high-resolution accurate mass spectrometry application to targeted proteomics. PROTEOMICS 2015, 15 (5–6), 880–890.

(112) Hendrickson, C. L.; Quinn, J. P.; Kaiser, N. K.; Smith, D. F.; Blakney, G. T.; Chen, T.; Marshall, A. G.; Weisbrod, C. R.; Beu, S. C. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. J. Am. Soc. Mass Spectrom. 2015, 26 (9), 1626–1632.

(113) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. Annu. Rev. Anal. Chem. 2015, 8 (1), 61–80.

(114) Makarov, A.; Denisov, E.; Lange, O. Performance evaluation of a high-field orbitrap mass analyzer. J. Am. Soc. Mass Spectrom. 2009, 20 (8), 1391–1396.

(115) Denisov, E.; Damoc, E.; Lange, O.; Makarov, A. Orbitrap mass spectrometry with resolving powers above 1,000,000. Int. J. Mass Spectrom. 2012, 325–327, 80–85.

(116) Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra. J. Proteome Res. 2014, 13 (8), 3679–3684.

(117) Brosch, M.; Swamy, S.; Hubbard, T.; Choudhary, J. Comparison of Mascot and X!Tandem Performance for Low and High Accuracy Mass Spectrometry and the Development of an Adjusted Mascot Threshold. Mol. Cell. Proteomics MCP 2008, 7 (5), 962–970.

(118) Branca, R. M. M.; Orre, L. M.; Johansson, H. J.; Granholm, V.; Huss, M.; Pérez-Bercoff, Å.; Forshed, J.; Käll, L.; Lehtiö, J. HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. Nat. Methods 2014, 11 (1), 59.

(119) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; et al. Informed-Proteomics: open-source software package for top-down proteomics. Nat. Methods 2017, 14 (9), 909–914.

(120) Granholm, V.; Kim, S.; Navarro, J. C. F.; Sjölund, E.; Smith, R. D.; Käll, L. Fast and Accurate Database Searches with MS-GF+Percolator. J. Proteome Res. 2014, 13 (2), 890–897.

(121) Wenger, C. D.; Coon, J. J. A Proteomics Search Algorithm Specifically Designed for High-Resolution Tandem Mass Spectra. J. Proteome Res. 2013, 12 (3), 1377–1386.

(122) Chapman, J. D.; Goodlett, D. R.; Masselon, C. D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. Mass Spectrom. Rev. 2014, 33 (6), 452–470.

(123) Egertson, J. D.; Kuehn, A.; Merrihew, G. E.; Bateman, N. W.; MacLean, B. X.; Ting, Y. S.; Canterbury, J. D.; Marsh, D. M.; Kellmann, M.; Zabrouskov, V.; et al. Multiplexed MS/MS for improved data-independent acquisition. Nat. Methods 2013, 10 (8), 744–746.

(124) Hu, A.; Noble, W. S.; Wolf-Yadlin, A. Technical advances in proteomics: new developments in data-independent acquisition. F1000Research 2016, 5, 419.

(125) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. Mol. Cell. Proteomics 2012, 11 (6), O111.016717.

(126) Lambert, J.-P.; Ivosev, G.; Couzens, A. L.; Larsen, B.; Taipale, M.; Lin, Z.-Y.; Zhong, Q.; Lindquist, S.; Vidal, M.; Aebersold, R.; et al. Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. Nat. Methods 2013, 10 (12), 1239–1245.

(127) Elias, J. E.; Gygi, S. P. Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. Methods Mol. Biol. Clifton NJ 2010, 604, 55–71.

(128) Medzihradszky, K. F.; Chalkley, R. J. Lessons in de novo peptide sequencing by tandem mass spectrometry. Mass Spectrom. Rev. 2015, 34 (1), 43–63.

(129) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 1994, 5 (11), 976–989.

(130) Klammer, A. A.; Park, C. Y.; Noble, W. S. Statistical Calibration of the SEQUEST XCorr Function. J. Proteome Res. 2009, 8 (4), 2106–2113.

(131) Carvalho, P. C.; Fischer, J. S. G.; Xu, T.; Cociorva, D.; Balbuena, T. S.; Valente, R. H.; Perales, J.; Yates, J. R.; Barbosa, V. C. Search Engine Processor: filtering and organizing PSMs. Proteomics 2012, 12 (7), 944–949.

(132) Cociorva, D.; L. Tabb, D.; Yates, J. R. Validation of Tandem Mass Spectrometry Database Search Results Using DTASelect. In Current Protocols in Bioinformatics; John Wiley & Sons, Inc., 2002.

(133) Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S. Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases. J. Proteome Res. 2008, 7 (1), 29–34.

(134) Choi, H.; Nesvizhskii, A. I. False Discovery Rates and Related Statistical Concepts in Mass Spectrometry-Based Proteomics. J. Proteome Res. 2008, 7 (1), 47–50.

(135) Reiter, L.; Claassen, M.; Schrimpf, S. P.; Jovanovic, M.; Schmidt, A.; Buhmann, J. M.; Hengartner, M. O.; Aebersold, R. Protein Identification False Discovery Rates for Very Large Proteomics Data Sets Generated by Tandem Mass Spectrometry. Mol. Cell. Proteomics 2009, 8 (11), 2405–2417.

(136) Aggarwal, S.; Yadav, A. K. False Discovery Rate Estimation in Proteomics. Methods Mol. Biol. Clifton NJ 2016, 1362, 119–128.

(137) Matthiesen, R. Mass Spectrometry Data Analysis in Proteomics; Springer Science & Business Media, 2007.

Chapter 2

(1) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. Annu. Rev. Anal. Chem. 2015, 8 (1), 61–80.

(2) 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.; et al. Novel Parallelized

Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. Anal. Chem. 2013, 85 (24), 11710–11714.

(3) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. Anal. Chem. 2000, 72 (6), 1156–1162.

(4) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization-principles and practice. Mass Spectrom. Rev. 1990, 9 (1), 37–70.

(5) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. Anal. Chem. 2013, 85 (1), 2–9.

(6) Karpievitch, Y. V.; Polpitiya, A. D.; Anderson, G. A.; Smith, R. D.; Dabney, A. R. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. Ann. Appl. Stat. 2010, 4 (4), 1797–1823.

(7) Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A. High Throughput Characterization of Combinatorial Histone Codes. Mol. Cell. Proteomics 2009, 8 (10), 2266–2284.

(8) Lin, S.; Garcia, B. A. Chapter One - Examining Histone Posttranslational Modification Patterns by High-Resolution Mass Spectrometry. In Methods in Enzymology; Wu, C., Allis, C. D., Eds.; Nucleosomes, Histones & Chromatin Part A; Academic Press, 2012; Vol. 512, pp 3–28.

(9) Angel, P. M.; Orlando, R. Quantitative carbamylation as a stable isotopic labeling method for comparative proteomics. Rapid Commun. Mass Spectrom. RCM 2007, 21 (10), 1623–1634.

(10) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 1994, 5 (11), 976–989.

(11) Xu, H.; Freitas, M. A. A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data. BMC Bioinformatics 2007, 8 (1), 133–143.

(12) Bern, M.; Kil, Y. J.; Becker, C. Byonic: advanced peptide and protein identification software. Curr. Protoc. Bioinforma. 2012, Chapter 13, Unit13.20.

(13) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical software to analyze top-down mass spectrometry data. PROTEOMICS 2015, 15 (7), 1235–1238.

Chapter 3

(1) Bensimon, A.; Heck, A. J. R.; Aebersold, R. Annu. Rev. Biochem. 2012, 81, 379−405.

(2) Gunaratne, J.; Schmidt, A.; Quandt, A.; Neo, S. P.; Sarac, O. S.; Gracia, T.; Loguercio, S.; Ahrne, E.; Xia, R. L. H.; Tan, K. H.; Lossner, C.; Bahler, J.; Beyer, A.; Blackstock, W.; Aebersold, R. Mol. Cell. Proteomics 2013, 12, 1741−1751.

(3) Gygi, S.; Villen, J.́ Nat. Protoc. 2008, 3, 1630−1638.

(4) Fonslow, B. R.; Stein, B. D.; Webb, K. J.; Xu, T.; Choi, J.; Park, S. L.; Yates, J. R., Jr. Nat. Methods 2013, 10, 54−56.

(5) Kuhn, E.; Whiteaker, J.; Mani, D. R.; Jackson, A.; Lei, Z.; Pope, M.; Smith, D.; Rivera, K.; Anderson, N. L.; Skates, S. J.; Pearson, T. W.; Paulovich, A. G.; Carr, S. A. Mol. Cell. Proteomics 2012, 9, 184− 196.

(6) 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, 2875−2882.

(7) 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, 2604−2614.

(8) Schwartz, R.; Ting, C. S.; King, J. Genome Res. 2001, 11, 703− 709.

(9) Wells, J. M.; McLuckey, S. A. In Methods in Enzymology; Elsevier: Amsterdam, The Netherlands, 2005; Vol. 402, pp 148−185.

(10) Cooper, H. J.; Hakansson, K.; Marshall, A. G.˚ Mass Spectrom. Rev. 2005, 24, 201−222.

(11) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Proc. Natl. Acad. Sci. U.S.A. 2004, 101, 9528−9533.

(12) Gardner, M. A.; Ledvina, A. R.; Smith, S.; Madsen, J.; Schwartz, G. C.; Stafford, G. C.; Coon, J. J.; Brodbelt, J. S. Anal. Chem. 2009, 81, 8109−8118.

(13) Kjeldsen, F.; Hørning, O. B.; Jensen, S. S.; Giessing, A. M. B.; Jensen, O. N. J. Am. Soc. Mass Spectrom. 2008, 19, 1156−1162.

(14) Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. J. Am. Soc. Mass. Spectrom. 2005, 16, 880−882.

(15) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. Proteomics 2011, 11, 1329−1334.

(16) Shaw, J. B.; Madsen, J. A.; Xu, H.; Brodbelt, J. S. J. Am. Soc. Mass Spectrom. 2012, 23, 1707−1715.

(17) Henderson, J. C.; Fage, C. D.; Cannon, J. R.; Brodbelt, J. S.; Keatinge-Clay, A. T.; Trent, M. S. ACS Chem. Biol. 2014, 9, 2382− 2392.

(18) Luo, Y.; Yogesha, S. D.; Cannon, J. R.; Yan, W.; Brodbelt, J. S.; Zhang, Y. ACS Chem. Biol. 2013, 8, 2042−2052.

(19) Han, S. W.; Lee, S. W.; Bahar, O.; Schwessinger, B.; Robinson, M. R.; Shaw, J. B.; Madsen, J. A.; Brodbelt, J. S. Nat. Commun. 2012, 3, 1153.

(20) Robinson, M.; Moore, K.; Brodbelt, J. S. J. Am. Soc. Mass Spectrom. 2014, 25, 1461−71.

(21) Angel, P. M.; Orlando, R. Rapid Commun. Mass Spectrom. 2007, 10, 1623−1634.

(22) Allen, S. J.; Schwartz, A. M.; Bush, M. F. Anal. Chem. 2013, 85, 12055−12061.

(23) Douglass, K. A.; Venter, A. R. Anal. Chem. 2013, 85, 8212− 8218.

(24) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Anal. Chem. 2013, 85, 2−9.

(25) Vasicek, L. A.; Ledvina, A. R.; Shaw, J. B.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J.; Brodbelt, J. S. J. Am. Soc. Mass Spectrom. 2011, 22, 1105−1108.

(26) Liu, H.; Sadygov, R. G.; Yates, J. R. Anal. Chem. 2004, 76, 4193−4201.

(27) Michalski, A.; Cox, J.; Mann, M. J. Proteome Res. 2011, 10, 1785−1793. Analytical Chemistry Article 12290 dx.doi.org/10.1021/ac5035314 | A

Chapter 4

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

(2) Cox, J.; Mann, M. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. Annu. Rev. Biochem. 2011, 80, 273–299.

(3) Zhang, Z.; Wu, S.; Stenoien, D. L.; Paša-Tolić, L. High-Throughput Proteomics. Annu. Rev. Anal. Chem. 2014, 7, 427–454.

(4) Mann, M.; Kulak, N. A.; Nagaraj, N.; Cox, J. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. Mol. Cell 2013, 49, 583–590.

(5) Yates, J. R. The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. J. Am. Chem. Soc. 2013, 135, 1629–1640.

(6) Vandermarliere, E.; Mueller, M.; Martens, L. Getting intimate with trypsin, the leading protease in proteomics: TRYPSIN IN PROTEOMICS. Mass Spectrom. Rev. 2013, 000–000.

(7) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. Anal. Chem. 2014, 86, 3005–3012.

(8) Swaney, D. L.; Wenger, C. D.; Coon, J. J. Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. J. Proteome Res. 2010, 9, 1323–1329.

(9) Wiśniewski, J. R.; Mann, M. Consecutive Proteolytic Digestion in an Enzyme Reactor Increases Depth of Proteomic and Phosphoproteomic Analysis. Anal. Chem. 2012, 84, 2631–2637.

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

(11) McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R. Comparison of three directly coupled HPLC MS/MS strategies for identification of

proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. Int. J. Mass Spectrom. 2002, 219, 245–251.

(12) Cheung, W. C.; Beausoleil, S. A.; Zhang, X.; Sato, S.; Schieferl, S. M.; Wieler, J. S.; Beaudet, J. G.; Ramenani, R. K.; Popova, L.; Comb, M. J.; et al. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. Nat. Biotechnol. 2012, 30, 447–452.

(13) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc. Natl. Acad. Sci. 2007, 104, 2193–2198.

(14) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. Anal. Chem. 2012, 84, 9668–9673.

(15) 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. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. J. Am. Chem. Soc. 2013, 135, 12646–12651.

(16) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-nm photodissociation of singly and multiply charged peptide anions for acidic proteome characterization. PROTEOMICS 2011, 11, 1329–1334.

(17) Yeh, G. K.; Sun, Q.; Meneses, C.; Julian, R. R. Rapid peptide fragmentation without electrons, collisions, infrared radiation, or native chromophores. J. Am. Soc. Mass Spectrom. 2009, 20, 385–393.

(18) Ly, T.; Julian, R. R. Ultraviolet Photodissociation: Developments towards Applications for Mass-Spectrometry-Based Proteomics. Angew. Chem. Int. Ed. 2009, 48, 7130–7137.

(19) Hendricks, N. G.; Lareau, N. M.; Stow, S. M.; McLean, J. A.; Julian, R. R. Bond-Specific Dissociation Following Excitation Energy Transfer for Distance Constraint Determination in the Gas Phase. J. Am. Chem. Soc. 2014, 136, 13363–13370.

(20) Feketeová, L.; Khairallah, G. N.; Brunet, C.; Lemoine, J.; Antoine, R.; Dugourd, P.; O'Hair, R. A. J. Fragmentation of the tryptophan cluster [Trp9-2H]2- induced by different activation methods. Rapid Commun. Mass Spectrom. RCM 2010, 24, 3255–3260.

(21) Aravind, G.; Klærke, B.; Rajput, J.; Toker, Y.; Andersen, L. H.; Bochenkova, A. V.; Antoine, R.; Lemoine, J.; Racaud, A.; Dugourd, P. Photodissociation pathways and lifetimes of protonated peptides and their dimers. J. Chem. Phys. 2012, 136, 014307.

(22) Girod, M.; Sanader, Z.; Vojkovic, M.; Antoine, R.; MacAleese, L.; Lemoine, J.; Bonacic-Koutecky, V.; Dugourd, P. UV Photodissociation of Proline-containing Peptide Ions: Insights from Molecular Dynamics. J. Am. Soc. Mass Spectrom. 2014.

(23) Moon, J. H.; Yoon, S. H.; Bae, Y. J.; Kim, M. S. Dissociation kinetics of singly protonated leucine enkephalin investigated by time-resolved photodissociation tandem mass spectrometry. J. Am. Soc. Mass Spectrom. 2010, 21, 1151–1158.

(24) Shin, Y. S.; Moon, J. H.; Kim, M. S. Observation of phosphorylation site-specific dissociation of singly protonated phosphopeptides. J. Am. Soc. Mass Spectrom. 2010, 21, 53–59.

(25) Yoon, S. H.; Moon, J. H.; Kim, M. S. Dissociation mechanisms and implication for the presence of multiple conformations for peptide ions with arginine at the C-terminus: time-resolved photodissociation study. J. Mass Spectrom. JMS 2010, 45, 806–814.

(26) Yoon, S. H.; Moon, J. H.; Chung, Y. J.; Kim, M. S. Influence of basic residues on dissociation kinetics and dynamics of singly protonated peptides: time-resolved photodissociation study. J. Mass Spectrom. JMS 2009, 44, 1532–1537.

(27) 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. Tyrosine sulfation in a Gram-negative bacterium. Nat. Commun. 2012, 3, 1153.

(28) Guan, Z.; Kelleher, N. L.; O'Connor, P. B.; Aaserud, D. J.; Little, D. P.; McLafferty, F. W. 193 nm photodissociation of larger multiply-charged biomolecules. Int. J. Mass Spectrom. Ion Process. 1996, 157-158, 357–364.

(29) Vasicek, L.; Brodbelt, J. S. Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores. Anal. Chem. 2010, 82, 9441–9446.

(30) Vasicek, L. A.; Ledvina, A. R.; Shaw, J.; Griep-Raming, J.; Westphall, M. S.; Coon, J. J.; Brodbelt, J. S. Implementing Photodissociation in an Orbitrap Mass Spectrometer. J. Am. Soc. Mass Spectrom. 2011, 22, 1105–1108.

(31) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. J. Mass Spectrom. 2000, 35, 1399–1406.

(32) Papayannopoulos, I. A. The interpretation of collision-induced dissociation tandem mass spectra of peptides. Mass Spectrom. Rev. 1995, 14, 49–73.

Chapter 5

(1) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. Annu. Rev. Anal. Chem. 2016, 9 (1), 499–519.

(2) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. Middle-Down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. Anal. Chem. 2014, 86 (6), 3005–3012.

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

(4) Smith, L. M.; Kelleher, N. L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla, L.; Kruppa, G.; Zubarev, R.; et al. Proteoform: a single term describing protein complexity. Nat. Methods 2013, 10 (3), 186–187.

(5) Senko, M. W.; Speir, J. P.; McLafferty, F. W. Collisional Activation of Large Multiply Charged Ions Using Fourier Transform Mass Spectrometry. Anal. Chem. 1994, 66 (18), 2801–2808.

(6) Zhou, M.; Paša-Tolić, L.; Stenoien, D. L. Profiling of Histone Post-Translational Modifications in Mouse Brain with High-Resolution Top-Down Mass Spectrometry. J. Proteome Res. 2017, 16 (2), 599–608.

(7) Fornelli, L.; Durbin, K. R.; Fellers, R. T.; Early, B. P.; Greer, J. B.; LeDuc, R. D.; Compton, P. D.; Kelleher, N. L. Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. J. Proteome Res. 2017, 16 (2), 609–618.

(8) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc. Natl. Acad. Sci. 2004, 101 (26), 9528–9533.

(9) Anderson, L. C.; Karch, K. R.; Ugrin, S. A.; Coradin, M.; English, A. M.; Sidoli, S.; Shabanowitz, J.; Garcia, B. A.; Hunt, D. F. Analyses of Histone Proteoforms Using Front-end Electron Transfer Dissociation-enabled Orbitrap Instruments. Mol. Cell. Proteomics 2016, 15 (3), 975–988.

(10) Cammarata, M. B.; Brodbelt, J. S. Characterization of Intra- and Intermolecular Protein Crosslinking by Top Down Ultraviolet Photodissociation Mass Spectrometry. ChemistrySelect 2016, 1 (3), 590–593.

(11) Cannon, J. R.; Kluwe, C.; Ellington, A.; Brodbelt, J. S. Characterization of green fluorescent proteins by 193 nm ultraviolet photodissociation mass spectrometry. PROTEOMICS 2014, 14 (10), 1165–1173.

(12) Cannon, J. R.; Martinez-Fonts, K.; Robotham, S. A.; Matouschek, A.; Brodbelt, J. S. Top-Down 193-nm Ultraviolet Photodissociation Mass Spectrometry for Simultaneous Determination of Polyubiquitin Chain Length and Topology. Anal. Chem. 2015, 87 (3), 1812–1820.

(13) Cammarata, M.; Lin, K.-Y.; Pruet, J.; Liu, H.; Brodbelt, J. Probing the Unfolding of Myoglobin and Domain C of PARP-1 with Covalent Labeling and Top-Down Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2014, 86 (5), 2534–2542.

(14) 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. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. J. Am. Chem. Soc. 2013, 135 (34), 12646–12651.

(15) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. Integration of Ultraviolet Photodissociation with Proton Transfer Reactions and Ion Parking for Analysis of Intact Proteins. Anal. Chem. 2016, 88 (1), 1008–1016.

(16) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. Anal. Chem. 2014, 86 (4), 2185–2192.

(17) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. The C-Score: A Bayesian Framework to Sharply Improve Proteoform Scoring in High-Throughput Top Down Proteomics. J. Proteome Res. 2014, 13 (7), 3231–3240.

(18) Armirotti, A.; Damonte, G. Achievements and perspectives of top-down proteomics. PROTEOMICS 2010, 10 (20), 3566–3576.

(19) Chanthamontri, C.; Liu, J.; McLuckey, S. A. Charge state dependent fragmentation of gaseous α-synuclein cations via ion trap and beam-type collisional activation. Int. J. Mass Spectrom. 2009, 283 (1–3), 9–16.

(20) Reid, G. E.; Wu, J.; Chrisman, P. A.; Wells, J. M.; McLuckey, S. A. Charge-State-Dependent Sequence Analysis of Protonated Ubiquitin Ions via Ion Trap Tandem Mass Spectrometry. Anal. Chem. 2001, 73 (14), 3274–3281.

(21) Amunugama, R.; Hogan, J. M.; Newton, K. A.; McLuckey, S. A. Whole Protein Dissociation in a Quadrupole Ion Trap: Identification of an a Priori Unknown Modified Protein. Anal. Chem. 2004, 76 (3), 720–727.

(22) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In Methods in Enzymology; Elsevier, 2005; Vol. 402, pp 148–185.

(23) Cobb, J. S.; Easterling, M. L.; Agar, J. N. Structural characterization of intact proteins is enhanced by prevalent fragmentation pathways rarely observed for peptides. J. Am. Soc. Mass Spectrom. 2010, 21 (6), 949–959.

(24) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. J. Mass Spectrom. 2000, 35 (12), 1399–1406.

(25) Bythell, B. J.; Suhai, S.; Somogyi, Á.; Paizs, B. Proton-Driven Amide Bond-Cleavage Pathways of Gas-Phase Peptide Ions Lacking Mobile Protons. J. Am. Chem. Soc. 2009, 131 (39), 14057–14065.

(26) Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R. Influence of Basic Residue Content on Fragment Ion Peak Intensities in Low-Energy Collision-Induced Dissociation Spectra of Peptides. Anal. Chem. 2004, 76 (5), 1243–1248.

(27) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation

Efficiency: Evidence for the Mobile Proton Model. J. Am. Chem. Soc. 1996, 118 (35), 8365–8374.

(28) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. Anal. Chem. 2015, 87 (14), 7109–7116.

(29) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, Ü. A.; Tsybin, Y. O. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. Chem. Soc. Rev. 2013, 42 (12), 5014.

(30) Pesavento, J. J.; Kim, Y.-B.; Taylor, G. K.; Kelleher, N. L. Shotgun Annotation of Histone Modifications: A New Approach for Streamlined Characterization of Proteins by Top Down Mass Spectrometry. J. Am. Chem. Soc. 2004, 126 (11), 3386–3387.

(31) Karch, K. R.; DeNizio, J. E.; Black, B. E.; Garcia, B. A. Identification and interrogation of combinatorial histone modifications. Front. Genet. 2013, 4.

(32) Morrison, L. J.; Rosenberg, J. A.; Singleton, J. P.; Brodbelt, J. S. Statistical Examination of the a and a + 1 Fragment Ions from 193 nm Ultraviolet Photodissociation Reveals Local Hydrogen Bonding Interactions. J. Am. Soc. Mass Spectrom. 2016, 27 (9), 1443–1453.

(33) Greer, S. M.; Cannon, J. R.; Brodbelt, J. S. Improvement of Shotgun Proteomics in the Negative Mode by Carbamylation of Peptides and Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2014, 86 (24), 12285–12290.

(34) Iavarone, A. T.; Jurchen, J. C.; Williams, E. R. Supercharged Protein and Peptide Ions Formed by Electrospray Ionization. Anal. Chem. 2001, 73 (7), 1455–1460.

(35) Iavarone, A. T.; Williams, E. R. Collisionally Activated Dissociation of Supercharged Proteins Formed by Electrospray Ionization. Anal. Chem. 2003, 75 (17), 4525–4533.

(36) Krusemark, C. J.; Frey, B. L.; Belshaw, P. J.; Smith, L. M. Modifying the charge state distribution of proteins in electrospray ionization mass spectrometry by chemical derivatization. J. Am. Soc. Mass Spectrom. 2009, 20 (9), 1617–1625.

(37) Pitteri, S. J.; Reid, G. E.; McLuckey, S. A. Affecting Proton Mobility in Activated Peptide and Whole Protein Ions via Lysine Guanidination. J. Proteome Res. 2004, 3 (1), 46–54.

(38) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2016, 88 (1), 1044–1051.

(39) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical software to analyze top-down mass spectrometry data. PROTEOMICS 2015, 15 (7), 1235–1238.

(40) Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. Influence of Secondary Structure on the Fragmentation of Protonated Peptides. J. Am. Chem. Soc. 1999, 121 (22), 5142–5154.

(41) Qin, J.; Chait, B. T. Preferential Fragmentation of Protonated Gas-Phase Peptide Ions Adjacent to Acidic Amino Acid Residues. J. Am. Chem. Soc. 1995, 117 (19), 5411–5412.

(42) Jockusch, R. A.; Schnier, P. D.; Price, W. D.; Strittmatter, E. F.; Demirev, P. A.; Williams, E. R. Effects of Charge State on Fragmentation Pathways, Dynamics,

and Activation Energies of Ubiquitin Ions Measured by Blackbody Infrared Radiative Dissociation. Anal. Chem. 1997, 69 (6), 1119–1126.

(43) Breci, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H. Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra. Anal. Chem. 2003, 75 (9), 1963–1971.

(44) Newton, K. A.; Pitteri, S. J.; Laskowski, M.; McLuckey, S. A. Effects of Single Amino Acid Substitution on the Collision-Induced Dissociation of Intact Protein Ions: Turkey Ovomucoid Third Domain. J. Proteome Res. 2004, 3 (5), 1033–1041.

(45) Paizs, B.; Suhai, S. Fragmentation pathways of protonated peptides. Mass Spectrom. Rev. 2005, 24 (4), 508–548.

Chapter 6

(1) Bensimon, A.; Heck, A. J. R.; Aebersold, R. Mass Spectrometry–Based Proteomics and Network Biology. Annu. Rev. Biochem. 2012, 81 (1), 379–405.

(2) Cox, J.; Mann, M. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. Annu. Rev. Biochem. 2011, 80 (1), 273–299.

(3) Mann, M.; Kulak, N. A.; Nagaraj, N.; Cox, J. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. Mol. Cell 2013, 49 (4), 583–590.

(4) Yates, J. R. The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. J. Am. Chem. Soc. 2013, 135 (5), 1629–1640.

(5) Sharma, K.; D'Souza, R. J.; Tyanova, S.; Schaab, C.; Wisniewski, J.; Cox, J.; Mann, M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. Cell Rep. 2014, 8 (5), 1583–1594.

(6) Villén, J.; Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat. Protoc. 2008, 3 (10), 1630–1638.

(7) Carlson, S. M.; Moore, K. E.; Green, E. M.; Martin, G. M.; Gozani, O. Proteome-wide enrichment of proteins modified by lysine methylation. Nat. Protoc. 2013, 9 (1), 37–50.

(8) Li, Y.; Silva, J. C.; Skinner, M. E.; Lombard, D. B. Mass Spectrometry-Based Detection of Protein Acetylation. In Sirtuins; Hirschey, M. D., Ed.; Humana Press: Totowa, NJ, 2013; Vol. 1077, pp 81–104.

(9) Zhang, Y.; Song, L.; Liang, W.; Mu, P.; Wang, S.; Lin, Q. Comprehensive profiling of lysine acetylproteome analysis reveals diverse functions of lysine acetylation in common wheat. 2016, 6, 1–10.

(10) 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.; et al. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. Anal. Chem. 2013, 85 (24), 11710–11714.

(11) Bern, M. W.; Kil, Y. J. Two-dimensional target decoy strategy for shotgun proteomics. J. Proteome Res. 2011, 10 (12), 5296–5301.

(12) Riley, N. M.; Bern, M.; Westphall, M. S.; Coon, J. J. Full-Featured Search Algorithm for Negative Electron-Transfer Dissociation. J. Proteome Res. 2016, 15 (8), 2768–2776.

(13) 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. High-throughput Database

Search and Large-scale Negative Polarity Liquid Chromatography-Tandem Mass Spectrometry with Ultraviolet Photodissociation for Complex Proteomic Samples. Mol. Cell. Proteomics 2013, 12 (9), 2604–2614.

(14) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: An open-source MS/MS sequence database search tool. PROTEOMICS 2013, 13 (1), 22–24.

(15) Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra. J. Proteome Res. 2014, 13 (8), 3679–3684.

(16) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. J. Proteome Res. 2011, 10 (4), 1794–1805.

(17) Chi, H.; He, K.; Yang, B.; Chen, Z.; Sun, R.-X.; Fan, S.-B.; Zhang, K.; Liu, C.; Yuan, Z.-F.; Wang, Q.-H.; et al. pFind–Alioth: A novel unrestricted database search algorithm to improve the interpretation of high-resolution MS/MS data. J. Proteomics 2015, 125, 89–97.

(18) Yamashita, M.; Fenn, J. B. Negative ion production with the electrospray ion source. J. Phys. Chem. 1984, 88 (20), 4671–4675.

(19) Straub, R. F.; Voyksner, R. D. Negative ion formation in electrospray mass spectrometry. J. Am. Soc. Mass Spectrom. 1993, 4 (7), 578–587.

(20) Garcia, M. The effect of the mobile phase additives on sensitivity in the analysis of peptides and proteins by high-performance liquid chromatography–electrospray mass spectrometry. J. Chromatogr. B 2005, 825 (2), 111–123.

(21) McCalley, D. V. Effect of buffer on peak shape of peptides in reversed-phase high performance liquid chromatography. J. Chromatogr. A 2004, 1038 (1–2), 77–84.

(22) Busman, M.; Schey, K. L.; Oatis, J. E.; Knapp, D. R. Identification of phosphorylation sites in phosphopeptides by positive and negative mode electrospray ionization-tandem mass spectrometry. J. Am. Soc. Mass Spectrom. 1996, 7 (3), 243–249.

(23) Bokatzian-Johnson, S. S.; Stover, M. L.; Dixon, D. A.; Cassady, C. J. A Comparison of the Effects of Amide and Acid Groups at the C-Terminus on the Collision-Induced Dissociation of Deprotonated Peptides. J. Am. Soc. Mass Spectrom. 2012, 23 (9), 1544–1557.

(24) Bowie, J. H.; Brinkworth, C. S.; Dua, S. Collision-induced fragmentations of the (M-H)- parent anions of underivatized peptides: An aid to structure determination and some unusual negative ion cleavages. Mass Spectrom. Rev. 2002, 21 (2), 87–107.

(25) Tran, T. T. N.; Brinkworth, C. S.; Bowie, J. H. The identification of disulfides in ricin D using proteolytic cleavage followed by negative-ion nano-electrospray ionization mass spectrometry of the peptide fragments: Identification of disulfides in ricin by negative-ion MS. Rapid Commun. Mass Spectrom. 2015, 29 (2), 182–190.

(26) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-nm photodissociation of singly and multiply charged peptide anions for acidic proteome characterization. PROTEOMICS 2011, 11 (7), 1329–1334.

(27) Shaw, J. B.; Madsen, J. A.; Xu, H.; Brodbelt, J. S. Systematic Comparison of Ultraviolet Photodissociation and Electron Transfer Dissociation for Peptide Anion Characterization. J. Am. Soc. Mass Spectrom. 2012, 23 (10), 1707–1715.

(28) Coon, J. J.; Shabanowitz, J.; Hunt, D. F.; Syka, J. E. P. Electron transfer dissociation of peptide anions. J. Am. Soc. Mass Spectrom. 2005, 16 (6), 880–882.

(29) Huzarska, M.; Ugalde, I.; Kaplan, D. A.; Hartmer, R.; Easterling, M. L.; Polfer, N. C. Negative Electron Transfer Dissociation of Deprotonated Phosphopeptide Anions: Choice of Radical Cation Reagent and Competition between Electron and Proton Transfer. Anal. Chem. 2010, 82 (7), 2873–2878.

(30) 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. Analysis of the Acidic Proteome with Negative Electron-Transfer Dissociation Mass Spectrometry. Anal. Chem. 2012, 84 (6), 2875–2882.

(31) Shaw, J. B.; Kaplan, D. A.; Brodbelt, J. S. Activated Ion Negative Electron Transfer Dissociation of Multiply Charged Peptide Anions. Anal. Chem. 2013, 85 (9), 4721–4728.

(32) Riley, N. M.; Rush, M. J. P.; Rose, C. M.; Richards, A. L.; Kwiecien, N. W.; Bailey, D. J.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. The Negative Mode Proteome with Activated Ion Negative Electron Transfer Dissociation (AI-NETD). Mol. Cell. Proteomics MCP 2015, 14 (10), 2644–2660.

(33) Robinson, M. R.; Taliaferro, J. M.; Dalby, K. N.; Brodbelt, J. S. 193 nm Ultraviolet Photodissociation Mass Spectrometry for Phosphopeptide Characterization in the Positive and Negative Ion Modes. J. Proteome Res. 2016, 15 (8), 2739–2748.

(34) Greer, S. M.; Cannon, J. R.; Brodbelt, J. S. Improvement of Shotgun Proteomics in the Negative Mode by Carbamylation of Peptides and Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2014, 86 (24), 12285–12290.

(35) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. Anal. Chem. 2015, 87 (14), 7109–7116.

(36) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2016, 88 (1), 1044–1051.

(37) Chalkley, R. J.; Bandeira, N.; Chambers, M. C.; Clauser, K. R.; Cottrell, J. S.; Deutsch, E. W.; Kapp, E. A.; Lam, H. H. N.; McDonald, W. H.; Neubert, T. A.; et al. Proteome Informatics Research Group (iPRG)_2012: A Study on Detecting Modified Peptides in a Complex Mixture. Mol. Cell. Proteomics MCP 2014, 13 (1), 360–371.

(38) Bern, M.; Cai, Y.; Goldberg, D. Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. Anal. Chem. 2007, 79 (4), 1393–1400.

(39) Madsen, J. A.; Ko, B. J.; Xu, H.; Iwashkiw, J. A.; Robotham, S. A.; Shaw, J. B.; Feldman, M. F.; Brodbelt, J. S. Concurrent Automated Sequencing of the Glycan and Peptide Portions of O -Linked Glycopeptide Anions by Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2013, 85 (19), 9253–9261.

(40) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. Anal. Chem. 2015, 87 (8), 4152–4158.

(41) Robinson, M. R.; Moore, K. L.; Brodbelt, J. S. Direct Identification of Tyrosine Sulfation by using Ultraviolet Photodissociation Mass Spectrometry. J. Am. Soc. Mass Spectrom. 2014, 25 (8), 1461–1471.

(42) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. Anal. Chem. 2016, 88 (4), 2303–2310.

Chapter 7

(1) Venne, A. S.; Kollipara, L.; Zahedi, R. P. The next Level of Complexity: Crosstalk of Posttranslational Modifications. Proteomics 2014, 14 (4–5), 513–524.

(2) Jaenisch, R.; Bird, A. Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals. Nat. Genet. 2003, 33, 245–254.

(3) Appella, E.; Anderson, C. W. Post-Translational Modifications and Activation of P53 by Genotoxic Stresses. Eur. J. Biochem. 2001, 268 (10), 2764–2772.

(4) Kouzarides, T. Chromatin Modifications and Their Function. Cell 2007, 128 (4), 693–705.

(5) Meek, D. W.; Anderson, C. W. Posttranslational Modification of P53: Cooperative Integrators of Function. Cold Spring Harb. Perspect. Biol. 2009, 1 (6), 1–16.

(6) Dai, C.; Gu, W. P53 Post-Translational Modification: Deregulated in Tumorigenesis. Trends Mol. Med. 2010, 16 (11), 528–536.

(7) Bode, A. M.; Dong, Z. Post-Translational Modification of P53 in Tumorigenesis. Nat. Rev. Cancer 2004, 4 (10), 793–805.

(8) Jenuwein, T. Translating the Histone Code. Science 2001, 293 (5532), 1074–1080.

(9) Bannister, A. J.; Kouzarides, T. Regulation of Chromatin by Histone Modifications. Cell Res. 2011, 21 (3), 381–395.

(10) Luo, Y.; Yogesha, S. D.; Cannon, J. R.; Yan, W.; Ellington, A. D.; Brodbelt, J. S.; Zhang, Y. Novel Modifications on C-Terminal Domain of RNA Polymerase II Can Fine-Tune the Phosphatase Activity of Ssu72. ACS Chem. Biol. 2013, 8 (9), 2042–2052.

(11) Sims, R. J.; Rojas, L. A.; Beck, D. B.; Bonasio, R.; Schüller, R.; Drury, W. J.; Eick, D.; Reinberg, D. The C-Terminal Domain of RNA Polymerase II Is Modified by Site-Specific Methylation. Science 2011, 332 (6025), 99–103.

(12) Farriol-Mathis, N.; Garavelli, J. S.; Boeckmann, B.; Duvaud, S.; Gasteiger, E.; Gateau, A.; Veuthey, A.-L.; Bairoch, A. Annotation of Post-Translational Modifications in the Swiss-Prot Knowledge Base. PROTEOMICS 2004, 4 (6), 1537–1550.

(13) Garavelli, J. S. The RESID Database of Protein Modifications as a Resource and Annotation Tool. PROTEOMICS 2004, 4 (6), 1527–1533.

(14) Mann, M.; Kulak, N. A.; Nagaraj, N.; Cox, J. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. Mol. Cell 2013, 49 (4), 583–590.

(15) Olsen, J. V.; Mann, M. Status of Large-Scale Analysis of Post-Translational Modifications by Mass Spectrometry. Mol. Cell. Proteomics 2013, 12 (12), 3444–3452.

(16) Kjeldsen, F.; Giessing, A. M. B.; Ingrell, C. R.; Jensen, O. N. Peptide Sequencing and Characterization of Post-Translational Modifications by Enhanced Ion-Charging and Liquid Chromatography Electron-Transfer Dissociation Tandem Mass Spectrometry. Anal. Chem. 2007, 79 (24), 9243–9252.

(17) Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A. High Throughput Characterization of Combinatorial Histone Codes. Mol. Cell. Proteomics 2009, 8 (10), 2266–2284.

(18) Sidoli, S.; Schwämmle, V.; Ruminowicz, C.; Hansen, T. A.; Wu, X.; Helin, K.; Jensen, O. N. Middle-down Hybrid Chromatography/Tandem Mass Spectrometry Workflow for Characterization of Combinatorial Post-Translational Modifications in Histones. PROTEOMICS 2014, 14 (19), 2200–2211.

(19) Moradian, A.; Kalli, A.; Sweredoski, M. J.; Hess, S. The Top-down, Middle-down, and Bottom-up Mass Spectrometry Approaches for Characterization of Histone Variants and Their Post-Translational Modifications. PROTEOMICS 2014, 14 (4–5), 489–497.

(20) Egelhofer, T. A.; Minoda, A.; Klugman, S.; Lee, K.; Kolasinska-Zwierz, P.; Alekseyenko, A. A.; Cheung, M.-S.; Day, D. S.; Gadel, S.; Gorchakov, A. A.; et al. An Assessment of Histone-Modification Antibody Quality. Nat. Struct. Mol. Biol. 2011, 18 (1), 91–94.

(21) Pereira Morais, M. P.; Mackay, J. D.; Bhamra, S. K.; Buchanan, J. G.; James, T. D.; Fossey, J. S.; van den Elsen, J. M. H. Analysis of Protein Glycation Using Phenylboronate Acrylamide Gel Electrophoresis. PROTEOMICS 2010, 10 (1), 48–58.

(22) Strack, R. Mass Spectrometry: Designer Proteases for Post-Translational Modifications. Nat. Methods 2017, 14 (2), 106–107.

(23) Aebersold, R.; Mann, M. Mass-Spectrometric Exploration of Proteome Structure and Function. Nature 2016, 537 (7620), 347–355.

(24) Larance, M.; Lamond, A. I. Multidimensional Proteomics for Cell Biology. Nat. Rev. Mol. Cell Biol. 2015, 16 (5), 269–280.

(25) Villén, J.; Gygi, S. P. The SCX/IMAC Enrichment Approach for Global Phosphorylation Analysis by Mass Spectrometry. Nat. Protoc. 2008, 3 (10), 1630–1638.

(26) Carlson, S. M.; Moore, K. E.; Green, E. M.; Martin, G. M.; Gozani, O. Proteome-Wide Enrichment of Proteins Modified by Lysine Methylation. Nat. Protoc. 2013, 9 (1), 37–50.

(27) Li, Y.; Silva, J. C.; Skinner, M. E.; Lombard, D. B. Mass Spectrometry-Based Detection of Protein Acetylation. In Sirtuins; Hirschey, M. D., Ed.; Humana Press: Totowa, NJ, 2013; Vol. 1077, pp 81–104.

(28) Wang, K.; Dong, M.; Mao, J.; Wang, Y.; Jin, Y.; Ye, M.; Zou, H. Antibody-Free Approach for the Global Analysis of Protein Methylation. Anal. Chem. 2016, 88 (23), 11319–11327.

(29) Zhang, L.; Liu, C.-W.; Zhang, Q. Online 2D-LC-MS/MS Platform for Analysis of Glycated Proteome. Anal. Chem. 2017.

(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. Tyrosine Sulfation in a Gram-Negative Bacterium. Nat. Commun. 2012, 3, 1153.

(31) Fort, K. L.; Dyachenko, A.; Potel, C. M.; Corradini, E.; Marino, F.; Barendregt, A.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. R. Implementation of Ultraviolet Photodissociation on a Benchtop Q Exactive Mass Spectrometer and Its Application to Phosphoproteomics. Anal. Chem. 2016, 88 (4), 2303–2310.

(32) Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F. Analysis of Phosphorylation Sites on Proteins from Saccharomyces Cerevisiae by Electron Transfer Dissociation (ETD) Mass Spectrometry. Proc. Natl. Acad. Sci. 2007, 104 (7), 2193–2198.

(33) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. Proteomics 2008, 8 (21), 4466–4483.

(34) Liao, R.; Zheng, D.; Nie, A.; Zhou, S.; Deng, H.; Gao, Y.; Yang, P.; Yu, Y.; Tan, L.; Qi, W.; et al. Sensitive and Precise Characterization of Combinatorial Histone Modifications by Selective Derivatization Coupled with RPLC-EThcD-MS/MS. J. Proteome Res. 2017, 16 (2), 780–787.

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

(36) Robinson, M. R.; Taliaferro, J. M.; Dalby, K. N.; Brodbelt, J. S. 193 Nm Ultraviolet Photodissociation Mass Spectrometry for Phosphopeptide Characterization in the Positive and Negative Ion Modes. J. Proteome Res. 2016, 15 (8), 2739–2748.

(37) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. Anal. Chem. 2000, 72 (3), 563–573.

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

(39) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. Anal. Chem. 2012, 84 (22), 9668–9673.

(40) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. J. Proteome Res. 2017, 16 (7), 2653–2659.

(41) Madsen, J. A.; Kaoud, T. S.; Dalby, K. N.; Brodbelt, J. S. 193-Nm Photodissociation of Singly and Multiply Charged Peptide Anions for Acidic Proteome Characterization. PROTEOMICS 2011, 11 (7), 1329–1334.

(42) Sidoli, S.; Garcia, B. A. Middle-down Proteomics: A Still Unexploited Resource for Chromatin Biology. Expert Rev. Proteomics 2017, 14 (7), 617–626.

(43) Peters, A. H. F. M.; Kubicek, S.; Mechtler, K.; O'Sullivan, R. J.; Derijck, A. A. H. A.; Perez-Burgos, L.; Kohlmaier, A.; Opravil, S.; Tachibana, M.; Shinkai, Y.; et al. Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. Mol. Cell 2003, 12 (6), 1577–1589.

(44) Dang, X.; Scotcher, J.; Wu, S.; Chu, R. K.; Tolic, N.; Ntai, I.; Thomas, P. M.; Fellers, R. T.; Early, B. P.; Zheng, Y.; et al. The First Pilot Project of the Consortium for Top-down Proteomics: A Status Report. PROTEOMICS 2014, 14 (10), 1130–1140.

(45) Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C. High-Throughput Middle-Down Analysis Using an Orbitrap. J. Proteome Res. 2010, 9 (8), 3886–3890.

(46) Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W. P.; Mohammed, S.; Heck, A. J. R. Toward an Optimized Workflow for Middle-Down Proteomics. Anal. Chem. 2017, 89 (6), 3318–3325.

(47) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; et al. Informed-Proteomics: Open-Source Software Package for Top-down Proteomics. Nat. Methods 2017, 14 (9), 909–914.

(48) Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. Quantitative Analysis of Modified Proteins and Their Positional Isomers by Tandem Mass Spectrometry: Human Histone H4. Anal. Chem. 2006, 78 (13), 4271–4280.

(49) Zheng, Y.; Fornelli, L.; Compton, P. D.; Sharma, S.; Canterbury, J.; Mullen, C.; Zabrouskov, V.; Fellers, R. T.; Thomas, P. M.; Licht, J. D.; et al. Unabridged Analysis of Human Histone H3 by Differential Top-Down Mass Spectrometry Reveals Hypermethylated Proteoforms from MMSET/NSD2 Overexpression. Mol. Cell. Proteomics 2016, 15 (3), 776–790.

(50) Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; et al. Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-Translational Modifications. J. Proteome Res. 2004, 3 (3), 621–626.

(51) Plazas-Mayorca, M. D.; Zee, B. M.; Young, N. L.; Fingerman, I. M.; LeRoy, G.; Briggs, S. D.; Garcia, B. A. One-Pot Shotgun Quantitative Mass Spectrometry Characterization of Histones. J. Proteome Res. 2009, 8 (11), 5367–5374.

(52) Schräder, C. U.; Ziemianowicz, D. S.; Merx, K.; Schriemer, D. C. Simultaneous Proteoform Analysis of Histones H3 and H4 with a Simplified Middle-Down Proteomics Method. Anal. Chem. 2018, 90 (5), 3083–3090.

(53) Sidoli, S.; Lu, C.; Coradin, M.; Wang, X.; Karch, K. R.; Ruminowicz, C.; Garcia, B. A. Metabolic Labeling in Middle-down Proteomics Allows for Investigation of the Dynamics of the Histone Code. Epigenetics Chromatin 2017, 10 (1), 34.

(54) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. Anal. Chem. 2014, 86 (4), 2185–2192.

(55) Greer, S. M.; Holden, D. D.; Fellers, R.; Kelleher, N. L.; Brodbelt, J. S. Modulation of Protein Fragmentation Through Carbamylation of Primary Amines. J. Am. Soc. Mass Spectrom. 2017, 28 (8), 1587–1599.

(56) Greer, S. M.; Brodbelt, J. S. Top-Down Characterization of Heavily Modified Histones Using 193 Nm Ultraviolet Photodissociation (UVPD) Mass Spectrometry. J. Proteome Res. 2018.

(57) Lin, S.; Garcia, B. A. Chapter One - Examining Histone Posttranslational Modification Patterns by High-Resolution Mass Spectrometry. In Methods in Enzymology; Wu, C., Allis, C. D., Eds.; Nucleosomes, Histones & Chromatin Part A; Academic Press, 2012; Vol. 512, pp 3–28.

(58) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2016, 88 (1), 1044–1051.

(59) Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. Chem Soc Rev 2014, 43 (8), 2757–2783.

(60) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. Anal. Chem. 2016, 88 (1), 30–51.

(61) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. Integration of Ultraviolet Photodissociation with Proton Transfer Reactions and Ion Parking for Analysis of Intact Proteins. Anal. Chem. 2016, 88 (1), 1008–1016.

(62) DeHart, C. J.; Fellers, R. T.; Fornelli, L.; Kelleher, N. L.; Thomas, P. M. Bioinformatics Analysis of Top-Down Mass Spectrometry Data with ProSight Lite. In Protein Bioinformatics; Methods in Molecular Biology; Humana Press, New York, NY, 2017; pp 381–394.

(63) Yang, Y.; Liu, F.; Franc, V.; Halim, L. A.; Schellekens, H.; Heck, A. J. R. Hybrid Mass Spectrometry Approaches in Glycoprotein Analysis and Their Usage in Scoring Biosimilarity. Nat. Commun. 2016, 7, 13397.

(64) Zhang, K.; Yau, P. M.; Chandrasekhar, B.; New, R.; Kondrat, R.; Imai, B. S.; Bradbury, M. E. Differentiation between Peptides Containing Acetylated or Tri-Methylated Lysines by Mass Spectrometry: An Application for Determining Lysine 9 Acetylation and Methylation of Histone H3. PROTEOMICS 2004, 4 (1), 1–10.

Chapter 8

(1) Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A. M.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; et al. Mass-spectrometry-based draft of the human proteome. Nature 2014, 509 (7502), 582–587.

(2) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The One Hour Yeast Proteome. Mol. Cell. Proteomics 2014, 13 (1), 339–347.

(3) Sharma, K.; D'Souza, R. J.; Tyanova, S.; Schaab, C.; Wisniewski, J.; Cox, J.; Mann, M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. Cell Rep. 2014, 8 (5), 1583–1594.

(4) Villén, J.; Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat. Protoc. 2008, 3 (10), 1630–1638.

(5) Li, Y.; Silva, J. C.; Skinner, M. E.; Lombard, D. B. Mass Spectrometry-Based Detection of Protein Acetylation. In Sirtuins; Hirschey, M. D., Ed.; Humana Press: Totowa, NJ, 2013; Vol. 1077, pp 81–104.

(6) Zhang, Y.; Song, L.; Liang, W.; Mu, P.; Wang, S.; Lin, Q. Comprehensive profiling of lysine acetylproteome analysis reveals diverse functions of lysine acetylation in common wheat. 2016, 6, 1–10.

(7) Carlson, S. M.; Moore, K. E.; Green, E. M.; Martin, G. M.; Gozani, O. Proteome-wide enrichment of proteins modified by lysine methylation. Nat. Protoc. 2013, 9 (1), 37–50.

(8) Wang, K.; Dong, M.; Mao, J.; Wang, Y.; Jin, Y.; Ye, M.; Zou, H. Antibody-Free Approach for the Global Analysis of Protein Methylation. Anal. Chem. 2016, 88 (23), 11319–11327.

(9) Zheng, Y.; Fornelli, L.; Compton, P. D.; Sharma, S.; Canterbury, J.; Mullen, C.; Zabrouskov, V.; Fellers, R. T.; Thomas, P. M.; Licht, J. D.; et al. Unabridged Analysis of Human Histone H3 by Differential Top-Down Mass Spectrometry

Reveals Hypermethylated Proteoforms from MMSET/NSD2 Overexpression. Mol. Cell. Proteomics 2016, 15 (3), 776–790.

(10) Yuan, Z.-F.; Arnaudo, A. M.; Garcia, B. A. Mass Spectrometric Analysis of Histone Proteoforms. Annu. Rev. Anal. Chem. 2014, 7 (1), 113–128.

(11) Altelaar, A. F. M.; Munoz, J.; Heck, A. J. R. Next-generation proteomics: towards an integrative view of proteome dynamics. Nat. Rev. Genet. 2012, 14 (1), 35–48.

(12) Peng, M.; Scholten, A.; Heck, A. J. R.; van Breukelen, B. Identification of Enriched PTM Crosstalk Motifs from Large-Scale Experimental Data Sets. J. Proteome Res. 2014, 13 (1), 249–259.

(13) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. Six alternative proteases for mass spectrometry based proteomics beyond trypsin. Nat. Protoc. 2016, 11 (5), 993–1006.

(14) Greer, S. M.; Parker, W. R.; Brodbelt, J. S. Impact of Protease on Ultraviolet Photodissociation Mass Spectrometry for Bottom-up Proteomics. J. Proteome Res. 2015, 14 (6), 2626–2632.

(15) Young, N. L.; DiMaggio, P. A.; Plazas-Mayorca, M. D.; Baliban, R. C.; Floudas, C. A.; Garcia, B. A. High Throughput Characterization of Combinatorial Histone Codes. Mol. Cell. Proteomics 2009, 8 (10), 2266–2284.

(16) Phanstiel, D.; Brumbaugh, J.; Berggren, W. T.; Conard, K.; Feng, X.; Levenstein, M. E.; McAlister, G. C.; Thomson, J. A.; Coon, J. J. Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. Proc. Natl. Acad. Sci. 2008, 105 (11), 4093–4098.

(17) Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. Quantitative Analysis of Modified Proteins and Their Positional Isomers by Tandem Mass Spectrometry: Human Histone H4. Anal. Chem. 2006, 78 (13), 4271–4280.

(18) Moradian, A.; Kalli, A.; Sweredoski, M. J.; Hess, S. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. PROTEOMICS 2014, 14 (4–5), 489–497.

(19) Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W. P.; Mohammed, S.; Heck, A. J. R. Toward an Optimized Workflow for Middle-Down Proteomics. Anal. Chem. 2017, 89 (6), 3318–3325.

(20) Sidoli, S.; Garcia, B. A. Middle-down proteomics: a still unexploited resource for chromatin biology. Expert Rev. Proteomics 2017, 14 (7), 617–626.

(21) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. Annu. Rev. Anal. Chem. 2016, 9 (1), 499–519.

(22) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; et al. Informed-Proteomics: open-source software package for top-down proteomics. Nat. Methods 2017, 14 (9), 909–914.

(23) Kou, Q.; Xun, L.; Liu, X. TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. Bioinformatics 2016.

(24) Fornelli, L.; Durbin, K. R.; Fellers, R. T.; Early, B. P.; Greer, J. B.; LeDuc, R. D.; Compton, P. D.; Kelleher, N. L. Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. J. Proteome Res. 2017, 16 (2), 609–618.

(25) Zhang, K.; Yau, P. M.; Chandrasekhar, B.; New, R.; Kondrat, R.; Imai, B. S.; Bradbury, M. E. Differentiation between peptides containing acetylated or tri-methylated lysines by mass spectrometry: An application for determining lysine 9 acetylation and methylation of histone H3. PROTEOMICS 2004, 4 (1), 1–10.

(26) Hauer, M. H.; Seeber, A.; Singh, V.; Thierry, R.; Sack, R.; Amitai, A.; Kryzhanovska, M.; Eglinger, J.; Holcman, D.; Owen-Hughes, T.; et al. Histone degradation in response to DNA damage enhances chromatin dynamics and recombination rates. Nat. Struct. Mol. Biol. 2017, 24 (2), 99–107.

(27) Kumar, R.; Horikoshi, N.; Singh, M.; Gupta, A.; Misra, H. S.; Albuquerque, K.; Hunt, C. R.; Pandita, T. K. Chromatin modifications and the DNA damage response to ionizing radiation. Front. Oncol. 2013, 2.

(28) Ewen, M. E. Where the cell cycle and histones meet. Genes Dev. 2000, 14 (18), 2265–2270.

(29) Altheim, B. A.; Schultz, M. C. Histone modification governs the cell cycle regulation of a replication-independent chromatin assembly pathway in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. 1999, 96 (4), 1345–1350.

(30) Jenuwein, T. Translating the Histone Code. Science 2001, 293 (5532), 1074–1080.

(31) Tanner, G. J.; Colgrave, M. L.; Blundell, M. J.; Goswami, H. P.; Howitt, C. A. Measuring Hordein (Gluten) in Beer – A Comparison of ELISA and Mass Spectrometry. PLOS ONE 2013, 8 (2), e56452.

(32) Pesavento, J. J.; Kim, Y.-B.; Taylor, G. K.; Kelleher, N. L. Shotgun Annotation of Histone Modifications: A New Approach for Streamlined Characterization of Proteins by Top Down Mass Spectrometry. J. Am. Chem. Soc. 2004, 126 (11), 3386–3387.

(33) Zhou, M.; Paša-Tolić, L.; Stenoien, D. L. Profiling of Histone Post-Translational Modifications in Mouse Brain with High-Resolution Top-Down Mass Spectrometry. J. Proteome Res. 2017, 16 (2), 599–608.

(34) Tian, Z.; Tolić, N.; Zhao, R.; Moore, R. J.; Hengel, S. M.; Robinson, E. W.; Stenoien, D. L.; Wu, S.; Smith, R. D.; Paša-Tolić, L. Enhanced top-down characterization of histone post-translational modifications. Genome Biol. 2012, 13 (10), R86.

(35) Dang, X.; Scotcher, J.; Wu, S.; Chu, R. K.; Tolic, N.; Ntai, I.; Thomas, P. M.; Fellers, R. T.; Early, B. P.; Zheng, Y.; et al. The first pilot project of the consortium for top-down proteomics: A status report. PROTEOMICS 2014, 14 (10), 1130–1140.

(36) Siuti, N.; Roth, M. J.; Mizzen, C. A.; Kelleher, N. L.; Pesavento, J. J. Gene-Specific Characterization of Human Histone H2B by Electron Capture Dissociation. J. Proteome Res. 2006, 5 (2), 233–239.

(37) Anderson, L. C.; Karch, K. R.; Ugrin, S. A.; Coradin, M.; English, A. M.; Sidoli, S.; Shabanowitz, J.; Garcia, B. A.; Hunt, D. F. Analyses of Histone Proteoforms Using Front-end Electron Transfer Dissociation-enabled Orbitrap Instruments. Mol. Cell. Proteomics 2016, 15 (3), 975–988.

(38) Huang, H.; Lin, S.; Garcia, B. A.; Zhao, Y. Quantitative Proteomic Analysis of Histone Modifications. Chem. Rev. 2015, 115 (6), 2376–2418.

(39) Jung, H. R.; Sidoli, S.; Haldbo, S.; Sprenger, R. R.; Schwimmle, V.; Pasini, D.; Helin, K.; Jensen, O. N. Precision Mapping of Coexisting Modifications in Histone H3 Tails from Embryonic Stem Cells by ETD-MS/MS. Anal. Chem. 2013, 85 (17), 8232–8239.

(40) Roman, M.; Gaskell, S. J. Charge state dependent top-down characterisation using electron transfer dissociation: Charge state dependent top-down characterisation using ETD. Rapid Commun. Mass Spectrom. 2012, 26 (3), 282–286.

(41) Iavarone, A. T.; Paech, K.; Williams, E. R. Effects of Charge State and Cationizing Agent on the Electron Capture Dissociation of a Peptide. Anal. Chem. 2004, 76 (8), 2231–2238.

(42) Brunner, A. M.; Lössl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F. M.; Heck, A. J. R. Benchmarking Multiple Fragmentation Methods on an Orbitrap Fusion for Top-down Phospho-Proteoform Characterization. Anal. Chem. 2015, 87 (8), 4152–4158.

(43) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. Anal. Chem. 2012, 84 (22), 9668–9673.

(44) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. Anal. Chem. 2015, 87 (14), 7109–7116.

(45) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. J. Proteome Res. 2017, 16 (7), 2653–2659.

(46) 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. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. J. Am. Chem. Soc. 2013, 135 (34), 12646–12651.

(47) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. Anal. Chem. 2014, 86 (4), 2185–2192.

(48) Madsen, J. A.; Boutz, D. R.; Brodbelt, J. S. Ultrafast Ultraviolet Photodissociation at 193 nm and its Applicability to Proteomic Workflows. J. Proteome Res. 2010, 9 (8), 4205–4214.

(49) Morrison, L. J.; Rosenberg, J. A.; Singleton, J. P.; Brodbelt, J. S. Statistical Examination of the a and a + 1 Fragment Ions from 193 nm Ultraviolet Photodissociation Reveals Local Hydrogen Bonding Interactions. J. Am. Soc. Mass Spectrom. 2016, 27 (9), 1443–1453.

(50) Cleland, T. P.; DeHart, C. J.; Fellers, R. T.; VanNispen, A. J.; Greer, J. B.; LeDuc, R. D.; Parker, W. R.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. High-Throughput Analysis of Intact Human Proteins Using UVPD and HCD on an Orbitrap Mass Spectrometer. J. Proteome Res. 2017, 16 (5), 2072–2079.

(51) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 2016, 88 (1), 1044–1051.

(52) Smith, L. M.; Kelleher, N. L.; Linial, M.; Goodlett, D.; Langridge-Smith, P.; Ah Goo, Y.; Safford, G.; Bonilla, L.; Kruppa, G.; Zubarev, R.; et al. Proteoform: a single term describing protein complexity. Nat. Methods 2013, 10 (3), 186–187.

(53) Cotham, V. C.; Brodbelt, J. S. Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level using Middle-Down 193 nm Ultraviolet Photodissociation. Anal. Chem. 2016, 88 (7), 4004–4013.

(54) Greer, S. M.; Holden, D. D.; Fellers, R.; Kelleher, N. L.; Brodbelt, J. S. Modulation of Protein Fragmentation Through Carbamylation of Primary Amines. J. Am. Soc. Mass Spectrom. 2017, 28 (8), 1587–1599.

(55) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Thomas, P. M.; Kelleher, N. L. The C-Score: A Bayesian Framework to Sharply Improve Proteoform Scoring in High-Throughput Top Down Proteomics. J. Proteome Res. 2014, 13 (7), 3231–3240.

Chapter 9

(1) Peng, M.; Scholten, A.; Heck, A. J. R.; van Breukelen, B. Identification of Enriched PTM Crosstalk Motifs from Large-Scale Experimental Data Sets. J. Proteome Res. 2014, 13 (1), 249–259.

(2) Venne, A. S.; Kollipara, L.; Zahedi, R. P. The next level of complexity: crosstalk of posttranslational modifications. Proteomics 2014, 14 (4–5), 513–524.

(3) Dang, X.; Scotcher, J.; Wu, S.; Chu, R. K.; Tolic, N.; Ntai, I.; Thomas, P. M.; Fellers, R. T.; Early, B. P.; Zheng, Y.; et al. The first pilot project of the consortium for top-down proteomics: A status report. PROTEOMICS 2014, 14 (10), 1130–1140.

(4) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. Anal. Chem. 2016, 88 (1), 30–51.

(5) Vasicek, L.; Brodbelt, J. S. Enhancement of Ultraviolet Photodissociation Efficiencies through Attachment of Aromatic Chromophores. Anal. Chem. 2010, 82 (22), 9441–9446.

(6) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. Integration of Ultraviolet Photodissociation with Proton Transfer Reactions and Ion Parking for Analysis of Intact Proteins. Anal. Chem. 2016, 88 (1), 1008–1016.

(7) Neue, K.; Mormann, M.; Peter-Katalinić, J.; Pohlentz, G. Elucidation of Glycoprotein Structures by Unspecific Proteolysis and Direct nanoESI Mass Spectrometric Analysis of ZIC-HILIC-Enriched Glycopeptides. J. Proteome Res. 2011, 10 (5), 2248–2260.

(8) Yang, Y.; Liu, F.; Franc, V.; Halim, L. A.; Schellekens, H.; Heck, A. J. R. Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. Nat. Commun. 2016, 7, 13397.

(9) Palmisano, G.; Lendal, S. E.; Engholm-Keller, K.; Leth-Larsen, R.; Parker, B. L.; Larsen, M. R. Selective enrichment of sialic acid–containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. Nat. Protoc. 2010, 5 (12), 1974–1982.

(10) Mysling, S.; Palmisano, G.; Højrup, P.; Thaysen-Andersen, M. Utilizing Ion-Pairing Hydrophilic Interaction Chromatography Solid Phase Extraction for Efficient Glycopeptide Enrichment in Glycoproteomics. Anal. Chem. 2010, 82 (13), 5598–5609.