Analysis of Post Translational Modifications by Tandem Mass Spectrometry

Stephanie Miller Lehman

Lancaster, Pennsylvania

Bachelor of Science in Chemistry, Eastern Mennonite University, 2006

A Dissertation Presented to the Graduate Faculty of the University of Virginia
in Candidacy for the Degree of Doctor of Philosophy

Department of Chemistry

University of Virginia
December, 2018

Abstract

This dissertation is composed of three projects, connected by the common goal of detection of post translational modifications on proteins by tandem mass spectrometry. From beginning to end, each of the projects is very different from the next, sampling the field of protein mass spectrometry.

The first project described in this dissertation is a proteomic based project at its core. Tryptic peptides, enriched for arginine methylation, are analyzed from cells treated under different conditions. A search algorithm is used to prioritize results; ultimately two spliceosome proteins are manually validated across all samples to better understand the constraints of the method. The second project is a detailed characterization of an acidic post translational modifications on larger peptide fragments from a histone chaperone protein, Nucleosome Assembly Protein 1-like. The final project applies a state-of-the-art technique for intact protein analysis and applies it to a complex biological sample, Nucleoplasmin purified from Xenopus eggs.

Acknowledgements

The work in this thesis is directly tied to the creative work of Dr. David Shechter and his graduate students and post doc, Dr. Christopher Warren, Benjamin Lorton, and Dr. Hongshan Chen. There have been countless hours of cell growth and protein purification that I am not even aware of. I am very appreciative of the opportunity to work with them.

The lab that Don and Jeff built has provided me with strong fundamentals of mass spectrometry, which I know that I will forever value. Don provides his students with space to creatively explore scientific questions, which they may or may not answer. Jeff's keen eye for detail and healthy dose of pessimism has taught me to become intimately knowledgeable about my data. I know I will incorporate both of these perspectives as I frame much of my professional career.

Dina offered optimistic enthusiasm to any project that needed a boost, including this dissertation. I am very grateful for the unjudging support and gentle encouragement while writing my dissertation. Thank you.

Thank you to Drs. Rebecca Pompano, Ken Hsu, Linda Columbus, and David Brautigan for serving on my defense committee. I am appreciative of your willingness to spend time and energy reviewing this thesis and participating at my defense.

Within the Hunt lab, I learned as much from my peers as from anyone else. Lissa Anderson provided me with a no-nonsense training my first summer in the Hunt lab. I have valued the continued conversations, encouragement, and friendship that has developed. Scott Ugrin, Ben Barnhill, and Paisley Myers helped me stay grounded in the lab and out of it. I appreciate that your cohort was willing to include Joel and me in things like Science Supper Club. Thank you. Josh Hinkle has been an excellent sounding board for discussing the lab or research after losing Ben and Scott to the professional world. I learned the simplest and most elegant explanations of mass spectrometry from Elizabeth Duselis. I am grateful to have had a yoga buddy and the chance to be your friend. I hope I can be more fun in the future. Ilsa Cooke, thank you for your dear friendship, it has meant the world.

Family has been a huge support for the last five years. My parents have provided enthusiastic encouragement. I appreciate the support you offered during this time. Glenn and Dorcas have been supportive during these years in Charlottesville. I really appreciated your continued curiosity and interest in my work. Sarah and Ryan and the kids were regular visitors to Charlottesville, provided entertaining breaks to the monotony of grad school. I am so grateful for the time and energy you spent traveling to visit us. I am excited to be able to spend more time with all of my family.

And Joel. These five years have been harder than each of us anticipated they would be. But we made it. Your patience, love, and understanding has been the best gift you have ever given me. Thank you. I am so grateful.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 3

Abbreviations

μ	micro ($1x10^{-6}$)
AGC	automated gain control
A, Ala	alanine
ADMA	asymmetric dimethylarginine
API	atmospheric pressure ionization
C, Cys	cysteine
^{13}C	carbon 13
C18	octadecylsilane
CAD	collision activated dissociation
CENP-A	centromeric protein A
D, Asp	aspartic acid
Da	Dalton
DC	direct current
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E, Glu	glutamic acid
ESI	electrospray ionization
ETD	electron transfer dissociation
F, Phe	phenylalanine
FWHM	full width at half maximum
FT	fourier transform, or high resolution
g	gravity
g	grams
G, Gly	glycine
H, His	histidine
HPLC	high performance liquid chromatography
I, Ile	isoleucine

i.d.	inner diameter
IIPT	ion/ion proton transfer
IT	ion trap or low resolution
k	kilo ($1x10^3$)
K, Lys	lysine
L	liter
L, Leu	leucine
LC	liquid chromatography
LIT	linear ion trap
m	milli ($1x10^{-3}$)
M	molar
M, Met	methionine
min	minute
MMA	monomethylarginine
mol	mole
mRNA	messenger ribonucleic acid
ms	milliseconds
MS	mass spectrometry
MS^1	full mass spectrum, scan of ions available without manipulation
MS^2, MS^n	tandem mass spectrum, scan of ions after isolation and reaction
Mw	molecular weight
m/z	mass-to-charge ratio
n	nano ($1x10^{-9}$)
N, Asn	asparagine
NA	Not applicable
NaCl	sodium chloride
NAP1L	nucleosome assembly protein 1-like
ND	not detected

NL	non-linear, way of normalizing ion counts
Npm	nucleoplasmin
NR	not reported
o.d.	outer diameter
OT	Orbitrap
p	pico ($1x10^{-12}$)
P, Pro	proline
PIP	parallel ion parking
ppm	parts per million
PRMT	protein arginine methyltransferase
PRTC	Pierce Retention Time Calibration
psm	peptide spectral match
PTM	post translational modification
Q, Gln	glutamine
R, Arg	arginine
RF	radio frequency
RMe2A	asymmetric dimethylarginine
RMe2S	symmetric dimethylarginine
RNA	ribonucleic acid
RT	retention time
S, Ser	serine
SDMA	symmetric dimethylarginine
S/N	signal-to-noise ratio
snRNA	small nuclear ribonucleic acid
T, Thr	threonine
TIC	total ion current
V	volt
V, Val	valine

w/w	weight-to-weight
W, Trp	tryptophan
Y, Tyr	tyrosine

1. Introduction to the Dissertation

1.1 Overview

This thesis presents three individual projects that are unified in their goal of detecting post translational modifications (PTMs) on proteins by mass spectrometry. Each of the chapters targets a specific PTM: Chapter 2 identifies arginine mono- and dimethylation, and Chapter 3 and 4 focus on the localization of glutamylation. However, the approach for the individual chapters is each unique and surveys the field of biological mass spectrometry as currently utilized to identify and site localize PTMs.

Chapter 2 is a proteomic project with a goal of developing a method to identify the arginine mono- and di-methylome in cells treated with methyltransferase inhibitors and to characterize the methylation symmetry. The project makes use of an antibody enrichment of modified peptides from treated cells. It requires chromatographically consistent methods with wide dynamic range. The peptides are highly charged, tryptic peptides; and the predominant fragmentation method is electron transfer dissociation. Furthermore, the method relies on the speed and decision tree advantages offered by the Thermo Orbitrap Fusion Tribrid instrument. Finally, because of the number of peptides and proteins, a database search is required to identify proteins for further manual validation.

In contrast, the goal of Chapter 3 is to characterize the modifications found on a single protein; specifically, the goal was to characterize glutamylation on the histone chaperone protein, nucleosome assembly protein 1-like. The analysis utilizes Endoproteinase Arg-C to preserve larger peptides for analysis. By evaluating larger peptides, identification of protein variants was possible. Rather than looking at peptides rich in basic amino acids, as in Chapter 2, the peptides have few basic residues and high mass; therefore, collisionally activated dissociation is used for analysis. Another goal that developed over the course of this project was to sufficiently sample peptides eluting over peaks to characterize the specific sites of modification rather than to sample a peak once, as in Chapter 2.

Finally, Chapter 4 represents a state-of-the-art application of recently developed methods to attempt to identify PTMs on intact proteins. Similar to Chapter 3, the work

focused on histone chaperone nucleoplasmin and recombinant truncations of the protein. The work for the chapter was performed on the Thermo Orbitrap Elite™ instrument with several custom modifications to the hardware and software. It represents the first biological application of parallel ion parking with ion/ion proton transfer (PIP-IIPT) towards identification of post translational modifications. The chapter represents an exploration into the opportunities and constraints of the method.

The focus of this first chapter will be to cover basic principles of mass spectrometry which will be the foundation for each of the following chapters.

1.2 Mass Spectrometry Instrumentation and Principles

1.2.1 Overview of MS experiment

In this chapter, peptides and proteins may be used interchangeably as the descriptor for multiply charged cations. A brief overview of an experiment utilizing mass spectrometry for biological samples is presented below.

Peptides are loaded on a reversed-phase liquid chromatography column for desalting and are gradient separated over time by hydrophobicity. Eluted peptides are ionized to the gas phase by electrospray and are introduced into the mass spectrometer. The peptide cations' mass-to-charge ratio (m/z) are analyzed with a high-resolution mass analyzer across a wide mass range, such as 300 to 1500 m/z, providing a snapshot of all the peptides that are eluting off the LC column at specific time; this is known as an MS1 spectrum. The peptides are individually selected by abundance, then isolated and fragmented in the mass spectrometer. The resulting fragment ions are acquired by either a low- or high-resolution mass analyzer; this is known as a tandem mass spectrum or MS^2. The precursors identified from the MS^1 continue to be selected by abundance for isolation, fragmentation, and analysis. After a designated number of MS^2 spectra, or a period of time, the MS^1 is repeated starting the process over again. Thousands of mass spectra are acquired over a given 60 min LC run. Events occurring in the instrument take place on the millisecond timescale. The MS^2 are then analyzed by searching a database of protein and peptide sequences and/or

manually evaluated. Concepts fundamental to these processes and utilized throughout the dissertation will be described in more detail within this chapter. (1)

1.2.2 HPLC Separation and Dynamic Exclusion

Prior to mass spectrometric analysis, the samples in each chapter of this dissertation are separated utilizing reverse-phase liquid chromatography, RP-HPLC, placed in-line with the mass spectrometer. While mass spectrometry separates ions by m/z ratio, RP-HPLC separates peptides by hydrophobicity over time as the mobile phase composition increases in organic solvent. The value in utilizing an orthogonal separation technique is that the dynamic range of the instrument is greatly improved; the instrument is able to detect peptides several orders of magnitude different in abundance. During a traditional data-dependent proteomic experiment, precursors eluting at a given point in time are identified by a high resolution MS^1 scan. The precursors are then selected for MS^2 analysis based on abundance in the MS^1, where the most abundant species are given preference. By eluting peptides over time, the instrument has more time to sample a broader range of peptides. Additionally, peptides with similar m/z will separate by hydrophobicity so that they can have separate fragmentation spectra (1).

After precursors are selected for fragmentation, they are placed on a “dynamic exclusion list” for a set period of time so that they are not re-selected for fragmentation. When the instrument takes another MS^1 scan, the process of selecting peptides for fragmentation repeats itself, but precursors on the dynamic exclusion list are excluded. This minimizes repeat analysis of identical precursors, enabling selection and fragmentation of lower abundance precursors; this more complete access to all peptides in the sample enables a more comprehensive analysis. Adjusting factors for dynamic exclusion, such as the number of times the precursor is selected before being placed on the dynamic exclusions list or the time it remains on the list, can dramatically impact the dynamic range of the method (1).

1.2.3 Electrospray Ionization

During electrospray ionization, peptides and proteins are converted from liquid phase to the gas phase, bridging the literal space from liquid chromatography to the gas phase for

detection by the mass spectrometer. HPLC columns used in the experiments presented in this dissertation are prepared from thin, fused silica that are only 360 µm in outer diameter and 75 µm inner diameter and have sharp laser pulled tips, a few µm in diameter, through which peptides elute in charged droplets (2). There is a 2kV potential between the inlet of the mass spectrometer and the column, causing the surface of the droplet to be positively charged. As the droplet traverses the length of the heated capillary, the solvent evaporates, shrinking the droplet and increasing the density of the charge on the surface. The droplet reaches its Raleigh limit, and either breaks into smaller droplets or expels a jet of droplets, much like the original column tip. The droplets continue under go this process until the protonated peptides are fully desolvated and are in the gas phase. (3)(4)

1.2.4 Charge State and m/z Ratios

Once protonated peptides are in the gas phase, they have specific mass-to-charge (m/z) ratios; the field of mass spectrometry relies on the ability to manipulate ions based on their m/z ratios. Ion motion in the gas phase is dictated by m/z ratio; this property allows for isolation, fragmentation, and analysis. Every molecule has a discrete mass, determined by its isotopic make-up. As molecules increase in size, a population of ions will then sample a greater variety of isotope distributions. For example, carbon 13 is a single neutron heavier than carbon 12 and is found in about 1.109% of the naturally occurring carbon. Thus, a peptide with a single carbon 13 will be 1.008 Da heavier than the peptide with no carbon 13. The larger the peptide, the more carbon atoms and therefore a greater number of carbon 13 atoms are likely to be exist within a molecule, creating a distribution of masses varying based on the isotope compositions. (1)

The charge state of a protein or peptide can be determined by measuring the m/z difference between the isotope peaks, see Figure 1 (1). It should be noted that high resolution mass analyzer such as an Orbitrap is required to be able to detect the individual isotopic distribution of peptides for charge states beyond a 2+ or 3+. A low resolution mass analyzer, such as a linear ion trap, is generally recognized as having 1 Da unit resolution.

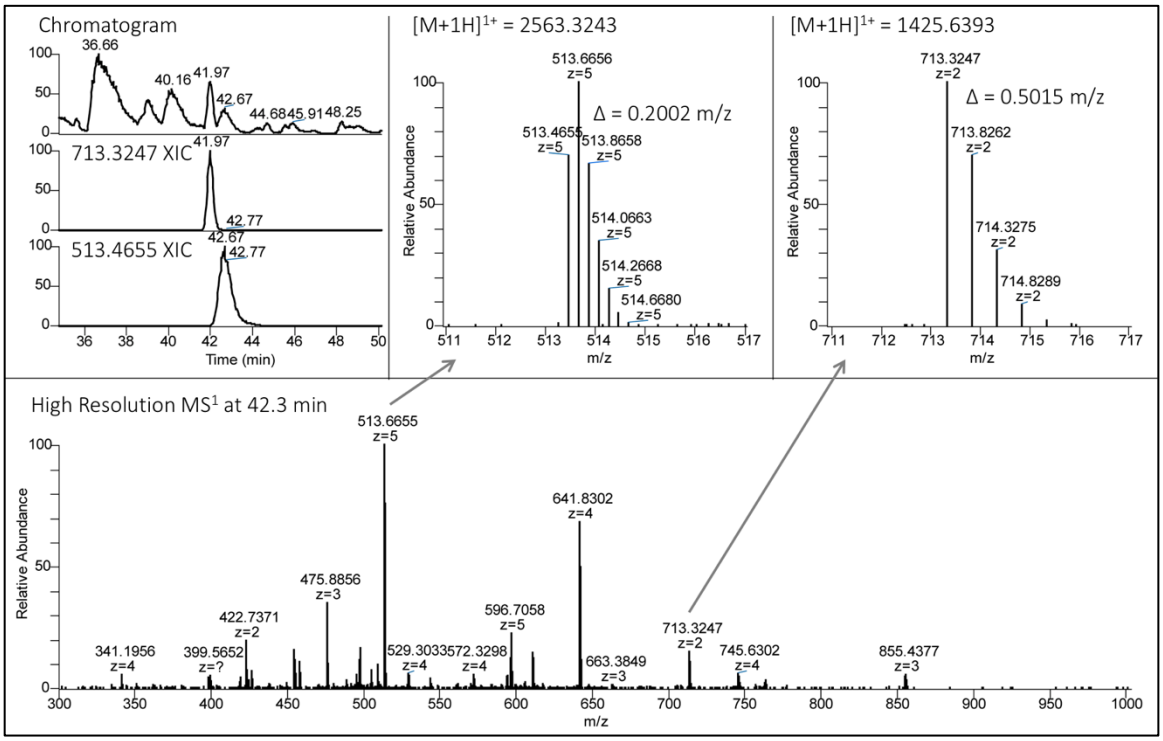

Figure 1. Peptide and Protein Isotope Distribution; the chromatogram in the A is a plot of the most abundant ions detected over time, with the extracted ion chromatogram (XIC) for two peptides plotted below. The MS^1 was taken at 42.3 min, the isotopic distributions for two distinct precursors are expanded above.

1.2.5 Quadrupole Mass Filter and Quadrupole Theory

One of the simplest mass analyzers is the quadrupole mass filter (QMF) composed of a set of four parallel rods, each 90° apart. Each pair of opposing rods have an equal DC offset and an AC potential referred to as the RF potential because it operates at radio frequencies.

Between the pairs of rods, the DC offset is oppositely charged and the RF potential is 180° out-of-phase with each other, see

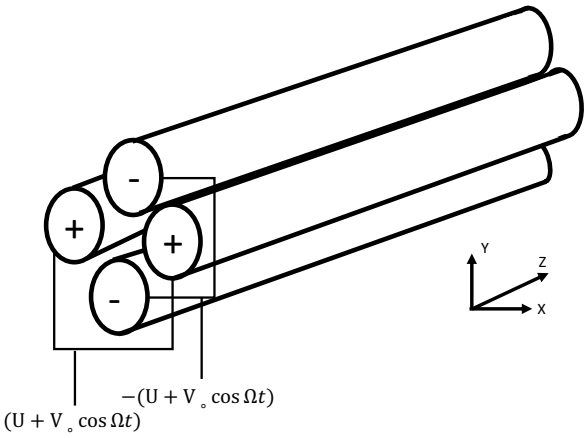

Figure 2. The oscillating RF potentials create a quadrupolar electric field between the electrodes, perpetually attracting and repelling ions. Ions within the electric field acquire characteristic, m/z-dependent frequencies of oscillation, known as secular frequencies (ω). (5)(6)

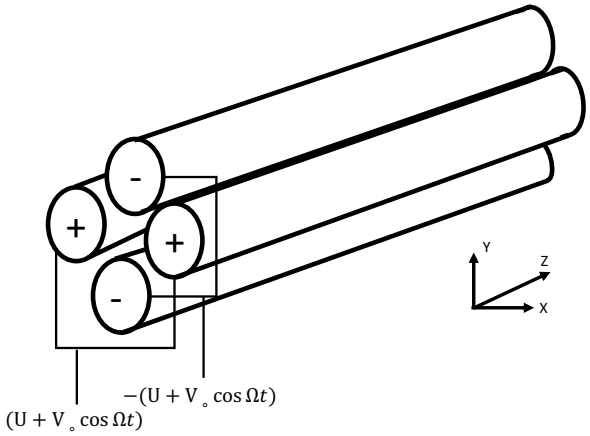

Figure 2. Quadrupole Mass Filter

Ion motion within the quadrupole electric field can be approximated by the Mathieu equation, a mathematical relationship created originally to model the vibrational patterns of stretch animal skins, much like a drum. When a drum is hit, the specific frequency to which the drum is tuned creates a unique pattern of peaks and nodes on the skin. Similarly, the amplitude of the RF potential can be compared to tuning a drum; adjusting the amplitude of the RF can define which ions are stable within the quadrupole field. The Mathieu equation, see Eq. 1, has two unitless parameters *q* and *a*, see Eq. 2 where V is the applied RF potential

and U is the DC offset. Plotting the Mathieu equation creates a stability diagram, depicted in Figure 3, where ions within the shaded region are defined as stable within the quadrupole field. A QMF can operate with either a wide range of stable frequencies or a narrow range of stable frequencies depending on the value of the AC to DC ratio. (5)(6)

(1) $\frac{d^2u}{d\xi^2} + (a_u - 2q_u \cos 2\xi)\, u = 0$

Where: *a* and *q* are unitless parameters related to stability of ions within the quadrupole field; u is ion displacement in the x, y, and z directions; ξ is related to an ion's secular frequency

(2) $a = \frac{-8eU}{m\omega^2 {r_0}^2} \qquad q = \frac{4eV}{m\omega^2 {r_0}^2}$

Where: *e* is charge of ion, *m* is molecular mass, *r* is internal radius of the quadrupole, ω is the frequency of the AC potential, U is the amplitude of the DC potential, V is the amplitude of the AC potential

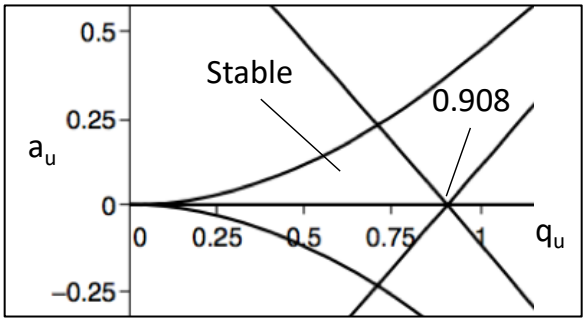

Figure 3. Mathieu Stability Diagram (Adapted from March, 2005)

1.2.6 Linear Ion Traps

Utilized in every chapter of this dissertation, the linear ion trap, LIT, was developed by Syka, Senko, and Schwartz and is a segmented, RF-only quadrupole (7). Its roles include ion isolation, fragmentation, reaction chamber, ion transfer, and mass analysis. There are four rods with hyperbolic surfaces which mimic the quadrupolar electric field created by the QMF. The rods are axially divided into three sections each, as shown in Figure 4, with the center section being longer than either of the end sections. Additionally, the center section contains small slits in the X direction for ejection during mass analysis. (7)

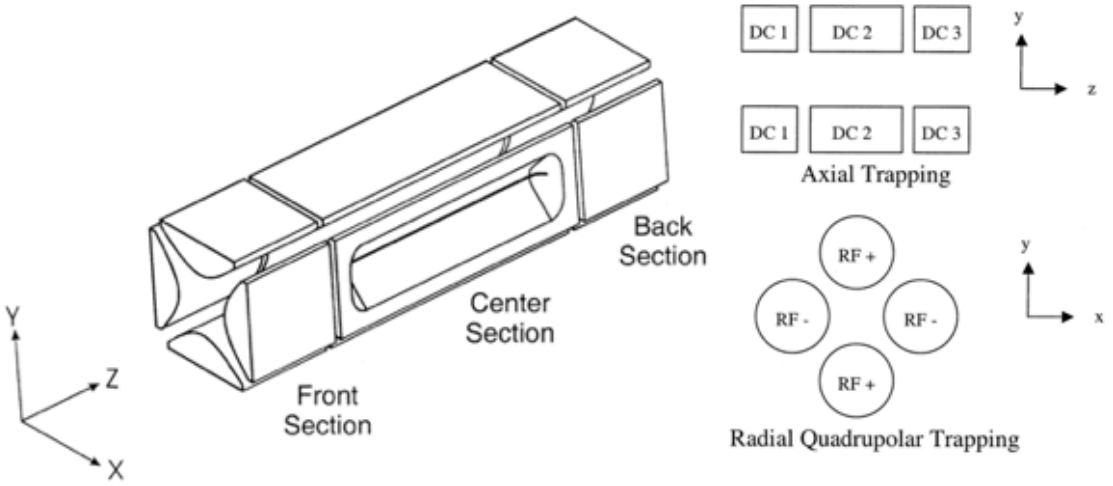

Figure 4. LIT Diagram (Adapted from Schwartz, 2002)

The LIT operates similarly to a QMF, but the sections are able to have different DC offsets (7). In addition to acting as a mass filter, the LIT has two distinct advantages over the QMF: ions can be trapped in the axial direction, and ions of opposite charge can be simultaneously trapped. Additionally, a supplemental electric AC potential may be applied to the LIT on the x-axis rods (7). The supplemental potential can be applied to a small range of frequencies, which acts to "activate" or resonantly kinetically excite specific ions, increasing the velocity of those ions. This supplemental potential can be used to eject ions in the x-direction for detection by electron multipliers and can also be used for collisionally activated dissociation (CAD) (*vide infra*) to fragment peptides. During much of ion residence time in the LIT, LIT are operated with a_u=0, which allows the q_u to be defined for a particular m/z value (7). For example, when an ion's q equals 0.908, the ion trajectory within the LIT is unstable; and the ion is ejected from the trap. Through employment of the supplemental potential on the x-rods, this phenomenom is exploited to resonantly kinetically excite ions such that that they become unstable and are ejected through the slits in the x-rods for detection by electron multipliers, creating a low resolution mass spectrum. As mass analyzers, LIT are traditionally reported as capable of 1 Da unit resolution. Modern LIT are capable of slightly better resolution, with the potential to differentiate a 3+ and 2+ charge states. (7)(6)

It is important to note that many of the ion manipulations for any given experiment occur in the LIT, such as accurate mass analysis, ion isolation, and fragmentation; and each of

these relies on the ability to accurately predict the frequencies at which an ion is moving. However, too many charges in a given space result in 'space charging,' meaning the cloud of charges from the ion population is contributing to an individual ion's perceived electric field from the quadrupolar electric field created by the 4 rods. This can affect accurate mass detection or appropriate isolation of the ion population. Therefore, the ion flux is measured constantly throughout the experiments, and ion injection events are dynamically altered to restrict space charging-induced phenomena. Note, the ion trap is filled based on the number of charges, not the number of ions, an important distinction in later concepts. (7)(6)

1.2.7 Orbitrap Mass Analyzer

High resolution mass analysis was performed utilizing Orbitrap mass analyzers. The Orbitrap is composed of a center electrode in the shape of a spindle with two outer electrodes, pictured in Figure 5. Ions are injected into the trap and orbit the center electron both radially and axially. The axial frequency for an ion is m/z dependent, see Eq. 3, and is measured by the two outer electrodes.

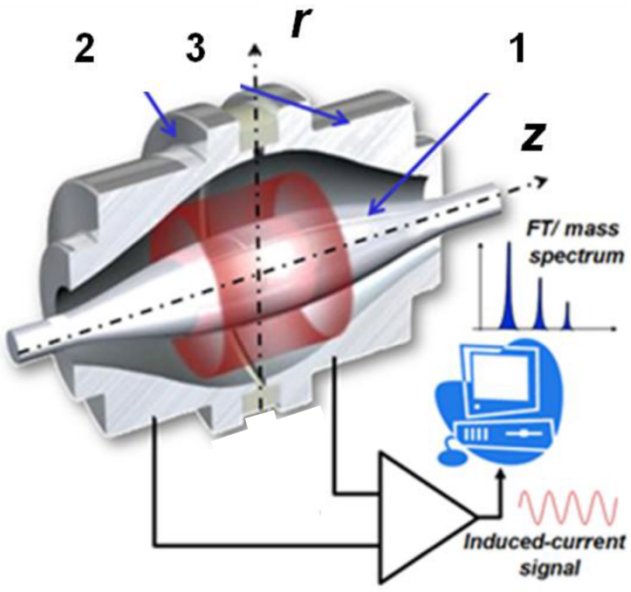

Figure 5. Thermo Orbitrap Mass Analyzer; ions are injected into the Orbitrap and orbit the (1) center electrode radially and axially, passing back and forth between the (2,3) outer electrodes. The ions induce a current and can be Fourier transformed into a mass spectrum. (Adapted from (8))

(3) $\omega = \sqrt{\frac{z}{m}k}$

Where: ω is angular frequency of the ions oscillating in the axial direction and k is the force constant of the quadratic restoring potential and related to the shape of the electrodes

As the ions pass back and forth between the detectors, the difference in detected current is measured as a complex waveform, referred to as a transient, which is then Fourier transformed into m/z measurements. Small m/z ions have high frequencies, and large m/z ions have low frequencies. Transients must be acquired long enough to accrue enough signal to detect low frequency ions; in that same time frame, many more oscillations from the high frequency ions can be recorded. Therefore, high frequency ions can be Fourier transformed with higher resolution than low frequency ions; the resolution of an Orbitrap varies across the mass range. (9)

As the Orbitrap operates at very low pressures, less than $1x10^{-10}$ Torr, ions can remain in motion for extended periods of time such as a second. The Orbitrap is capable of a range of resolutions, though it should be noted that the higher the resolution, the longer the acquisition of the transient. In the dissertation, the highest resolution utilized was 120,000 FWHM at 400 m/z, or 0.0033 m/z. (9, 10)

1.2.8 Instrumentation

The combination of ion optics and mass analyzers determine the utility and capabilities of each mass spectrometer instrument. Two specific instruments were utilized in this dissertation, a Thermo Orbitrap Fusion™ Tribrid™ and a Thermo Orbitrap Elite™, and should be discussed briefly to highlight specific differences and advantages.

Chapters 2 and 3 utilized the Orbitrap Fusion™ Tribrid™, an instrument with three different mass spectrometers: a QMF, a LIT, and an Orbitrap (see Figure 6) (11). The instrument also contains an ion-routing multipole (IRM) that is used as an ion accumulation and storage device. Ions are isolated using the QMF rather than the LIT and then accumulated in the IRM (11). With this configuration as well as several software changes, the

instrument performs multiple functions in parallel, such as acquisition of a high-resolution transient while simultaneously isolating the precursor in the IRM and acquiring a MS^2 in the LIT. The software controls for the method builder were also improved allowing for more complex decision tree type methods; this was utilized in Chapters 2 and 3 (11).

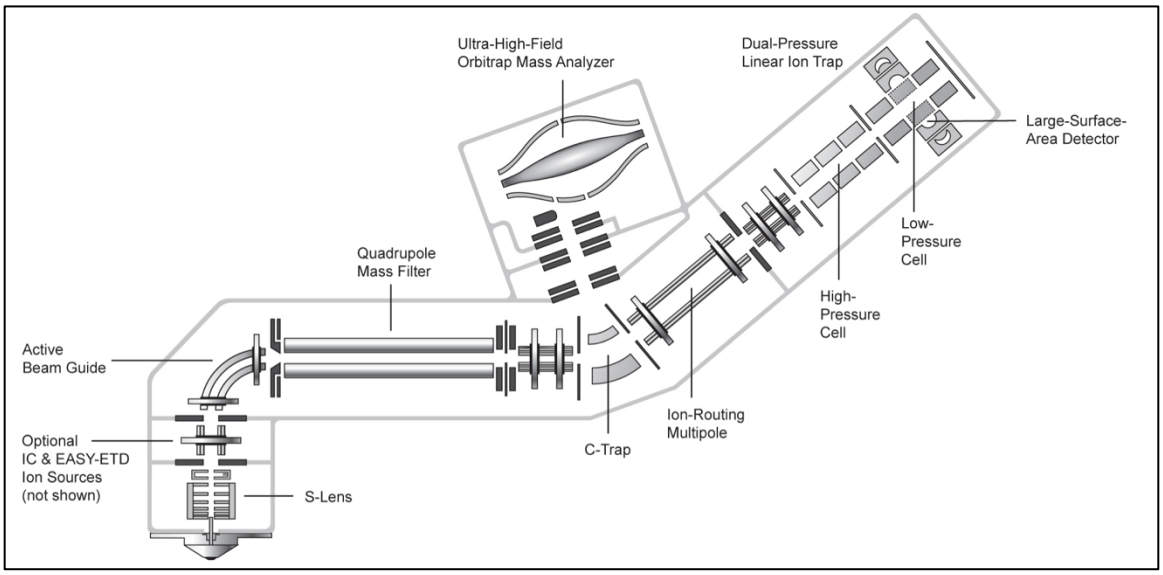

Figure 6. Thermo Orbitrap Fusion™ Tribrid™ (modified from 19)

Alternatively, the Thermo Orbitrap Elite™, see Figure 7, is a previous generation instrument (12). Peptide analysis has been the driving force for much of the field of biological mass spectrometry; Chapter 4 is focused on analyzing intact proteins. Intact protein analysis by mass spectrometry faces several challenges due to their larger size; this complicates the analysis of both the intact mass as well as fragments. The Hunt Lab has modified the instrument from the factory design with three specific changes to optimize the instrument towards intact protein analysis: the instrument has ion/ion proton transfer reagents in addition to electron transfer dissociation reagents, the ion source was customized to allow the reagents utilized in ion/ion reactions to be introduced from the front of the instrument, and the software code was modified to include multiple fills of ion/ion reaction products into the C-trap (13). Several of these concepts are described in more detail in Chapter 4.

The multiple C-trap fills (mf) is important due to the limited capacity of the LIT. During ion/ion reactions, the cations are trapped in the back section of the LIT while the center section is filled with reagent ions (14). Thus, the number of precursor ions allowed

into the LIT is limited by the number of charges that can be trapped in the small portion of the LIT before space charging occurs. Additionally, the total summed charge of the product ions is reduced due to the inherent neutralization of ion/ion reactions. The storage capacity of the C-trap is greater than small section of the LIT, therefore the product ions from multiple iterations of ion/ion reactions in the LIT can be stored in the C-trap prior to sending the ions to the Orbitrap for a single analysis (13). The ion population for each product ion is built up, directly improving the signal measured by in the Orbitrap.

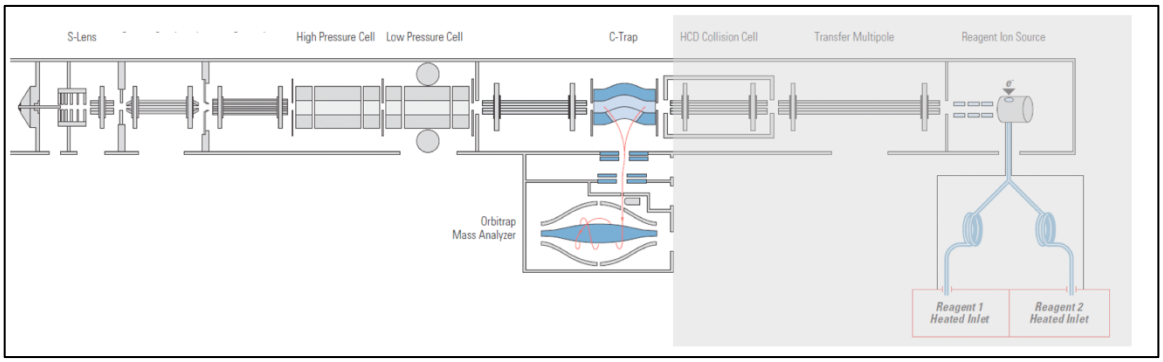

Figure 7. Thermo Orbitrap Elite™ Diagram (modified from 20)

1.2.9 Fragmentation and Peptide Sequencing

In order to accurately identify a given peptide or site localize PTMs, the peptide must be fragmented; and the MS^2 is acquired to provide sequence information. There are two fragmentation methods utilized in this dissertation: collisionally activated dissociation (CAD) and electron transfer dissociation (ETD). When both methods are performing optimally, each individual precursor ion will break into two fragments, also known as product ions, which contain either the peptide N-terminus or the C-terminus. The peptide should fragment across the peptide backbone, yielding a series of fragment ions increasing in m/z; the difference between each fragment ion will then correspond to the mass of an amino acid, illustrated in Figure 8. MS^2 spectra can be acquired in either the LIT or the Orbitrap. Detection in the LIT is faster and more sensitive than the Orbitrap but lacks the high resolution (11).

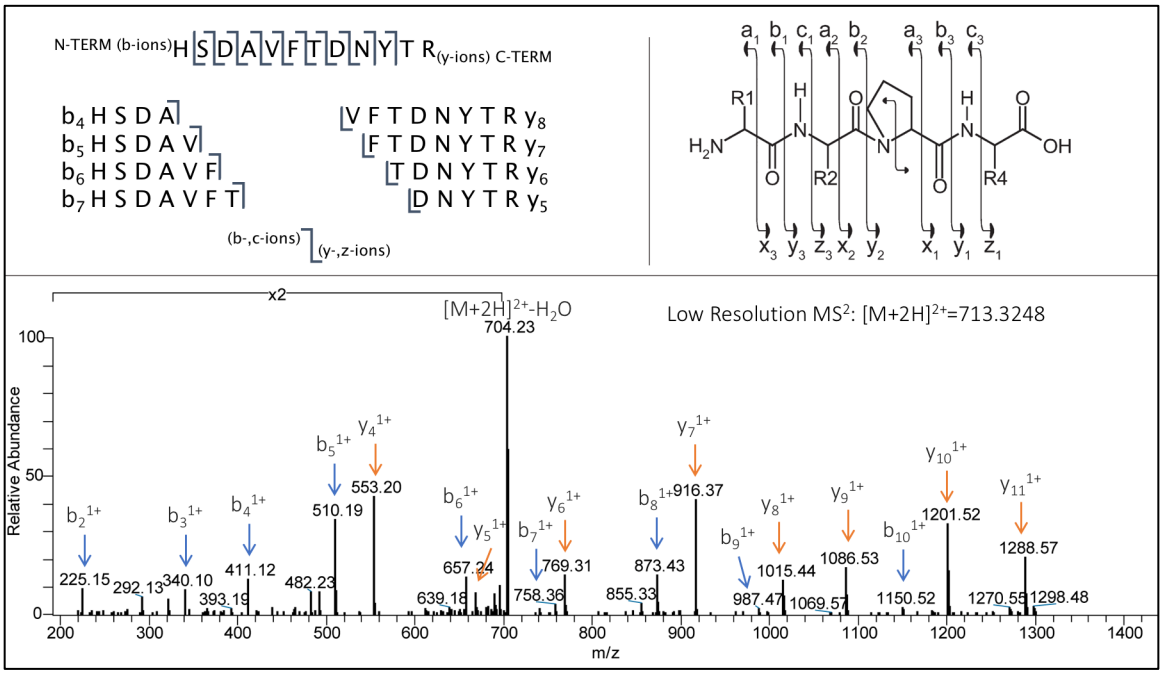

Figure 8. Peptide Sequencing: the mass difference between fragment ions corresponds to amino acid residues, ions detected in the low resolution spectrum are identified with hash marks in the sequence above. Image in upper right corner from (15).

1.2.9.1 Collisionally Activated Dissociation

Collisionally activated dissociation is a common fragmentation method that occurs in the LIT. A window of frequencies corresponding to a defined range of m/z values around a precursor for fragmentation are activated using supplemental activation (*vide supra*.) The ions with corresponding secular frequencies experience increased velocities. This causes higher energy collisions with the helium bath gas. The small amount of kinetic energy transferred in each of the collisions is converted to vibrational energy, eventually causing the peptide to fragment. The resulting fragment ions have characteristic frequencies outside the window of activation, preventing them from undergoing further fragmentation. Fragmentation by CAD results in b- and y-ion series, the characteristic ions resulting from the amide bond cleavage. (6, 16)

Two limitations with fragmentation by CAD occur with highly charged precursor ions and precursors with a PTM. Highly charged precursors result in highly charged product ions, difficult for sequencing with low resolution MS^2 spectra, the traditional approach. Precursors with PTMs frequently yield little sequence information, since the lowest energy pathway for

fragmentation is often labile PTMs. Theoretically, there should be a single fragmentation; however, the activation frequently results in detection of fragment ions with the loss of the PTM, complicating the ability to accurately place the modification. (6, 16)

1.2.9.2 Electron Transfer Dissociation

Electron transfer dissociation (ETD) is a complementary fragmentation technique to CAD (14). Also performed in the LIT, ETD is an ion/ion reaction between a multiply charged cation (peptide) and a radical anion, either fluoranthene or azulene. The precursor is trapped in the back section of the LIT, and the reagent ions are accumulated in the center section of the trap. The DC potentials sequestering the ion populations are removed, and the ions react following the mechanism outlined in Figure 9. A radical electron is transferred from the reagent to the peptide, hypothetically at the carbonyl carbon. The shifted electron density creates a basic site at this carbon, causing the carbonyl to abstract a proton from a nearby protonated residue, such as a lysine, arginine, or histidine. This causes a rearrangement of the peptide backbone, fragmenting frequently at the N-C_α bond, resulting in c- and z●-ions (14). Each ETD reaction results in the transfer of a radical electron that reduces the net charge of the peptide. Because the ETD mechanism reduces the charge of fragment ions, it performs better with more highly charged peptides. Additionally, ETD seems to produce better spectra when working with high charge density peptides or m/z ratios below 900 m/z. ETD is preferred when site localizing labile PTMs as the fragment ions are detected with the modification intact (14, 15).

In the mass spectrometer, gas phase ion/ion reactions occur without any sort of solvent or dielectric medium. Thus, the rate of ion/ion reactions is strongly governed by the coulombic attraction between the reacting species. McLuckey et al. showed that the rate of ion/ion reactions is dependent on the square of the charge of the ions; more highly charge peptides will react faster than the lower charged product ions (17). Historically, the number of second and third generation reactions can by controlled by restricting the time of the reaction (17, 18).

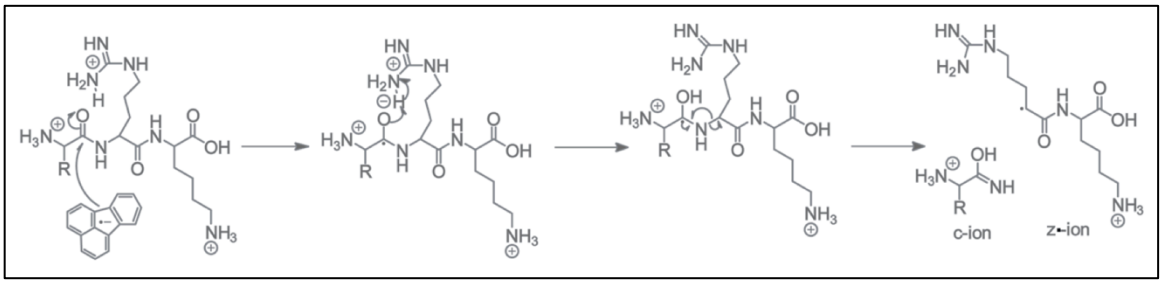

Figure 9. ETD Mechanism

1.3 References

1. Dass C (2001) *Principles and practice of biological mass spectrometry* (John Wiley).

2. Martin SE, Shabanowitz J, Hunt DF, Marto JA (2000) Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 72(18):4266–4274.

3. Fenn JB (2002) Electrospray ionization mass spectrometry: How it all began. *J Biomol Tech* 13(3):101–118.

4. Cech NB, Enke CG (2001) Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* 20(6):362–387.

5. March RE (1997) An introduction to quadrupole ion trap mass spectrometry. *J Mass Spectrom* 32(4):351–369.

6. March R, Todd J (2005) *Quadrupole Ion Trap Mass Spectrometry* (J. Wiley). 2nd Ed. Available at: https://search.lib.virginia.edu/catalog/u7299000.

7. Schwartz JC, Senko MW, Syka JEP (2002) A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* 13(6):659–669.

8. Grinfeld D, Monastyrskiy M, Makarov A (2015) Control of Aberration and Space-Charge Effects in the Orbitrap Mass Analyzer. *Microsc Microanal* 21(S4):176–181.

9. Hu Q, et al. (2005) The Orbitrap: A new mass spectrometer. *J Mass Spectrom* 40(4):430–443.

10. Makarov A, et al. (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* 78(7):2113–2120.

11. Senko MW, et al. (2013) Novel parallelized quadrupole/linear ion trap/orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Anal Chem* 85(24):11710–11714.

12. Michalski A, et al. (2012) Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Mol Cell Proteomics* 11(3):O111.013698.

13. Earley L, et al. (2013) Front-End Electron Transfer Dissociation: A New Ionization Source. *Anal Chem* 85(17):8385–8390.

14. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci* 101(26):9528–9533.

15. Mikesh LM, et al. (2006) The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta - Proteins Proteomics* 1764(12):1811–1822.

16. Kaiser RE, Cooks RG, Syka JEP, Stafford GC (1990) Collisionally activated dissociation of peptides using a quadrupole ion-trap mass spectrometer. *Rapid Commun Mass Spectrom* 4(1):30–33.

17. McLuckey SA, Stephenson JL, Asano KG (1998) Ion/Ion Proton-Transfer Kinetics: Implications for Analysis of Ions Derived from Electrospray of Protein Mixtures. *Anal Chem* 70(6):1198–1202.

18. Good DM, Wirtala M, McAlister GC, Coon JJ (2007) Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol Cell Proteomics* 6(11):1942–1951.

19. Planetorbitrap.com. (2018). *Thermo Fisher :: Orbitrap :: Orbitrap Fusion*. [online] Available at: http://planetorbitrap.com/orbitrap-fusion#tab:schematic [Accessed 18 Sept. 2018].

20. www.thermofisher.com. (2018). *Orbitrap Elite Hybrid Ion Trap-Orbitrap Mass Spectrometer*. [online] Available at: https://www.thermofisher.com/order/catalog/product/IQLAAEGAAPFA DBMAZQ?SID=srch-srp-IQLAAEGAAPFADBMAZQ [Accessed 18 Sept. 2018].

2. Development of an LC-MS/MS Method for Type I and Type II PRMT Treated Samples Enriched for Methylated Arginine Residues by PTMScan Technology

2.1 Abstract

Arginine methylation is dysregulated in cancer, frequently with overexpression of protein arginine methyltransferase (PRMTs) enzymes. PRMTs catalyze the post translational addition of methylation in to arginine residues as either monomethylation, or asymmetric or symmetric dimethylation. Previous work with histone proteins highlights the importance of the geometry of the arginine dimethylation. To identify substrates for Type I (asymmetric) and Type II (symmetric) PRMTs, cells were treated with either Type I or Type II inhibitors and enriched for methylated arginine containing peptides using antibodies from Cell Signaling Technologies. A method to analyze the samples using a Thermo Orbitrap Fusion Tribrid with electron transfer dissociation was developed for detection of highly charged tryptic peptides. To understand the scope of the developed method, two Sm proteins, known Type II substrates, were manually validated across all samples. SmD1 dimethylation is shown to be reduced in cells inhibited for Type II PRMT activity, while SmD3 dimethylation remained constant.

2.2 Introduction

Methylation, a conserved post translational modification (PTM), is the enzymatic transfer of a methyl group by a methyltransferase from the donor, S-adenosyl methionine, to a receiving group, see *Figure 10*. Methylation occurs naturally on lysines, arginines, protein amino and carboxyl termini, and has been found on acidic residues. Though less thoroughly studied than phosphorylation, methylation is thought to occur with similar abundance, with an estimated 7% of arginines thought to be methylated (1).

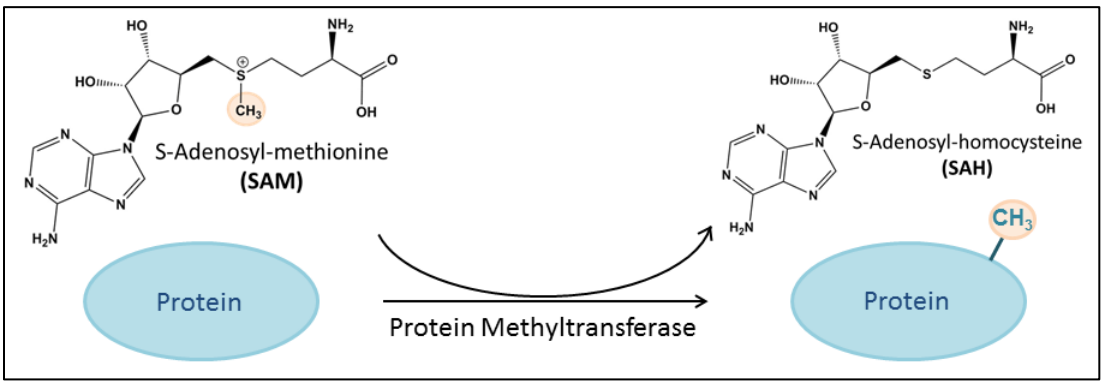

Figure 10. Transfer of methyl from S-adenosine methionine to protein

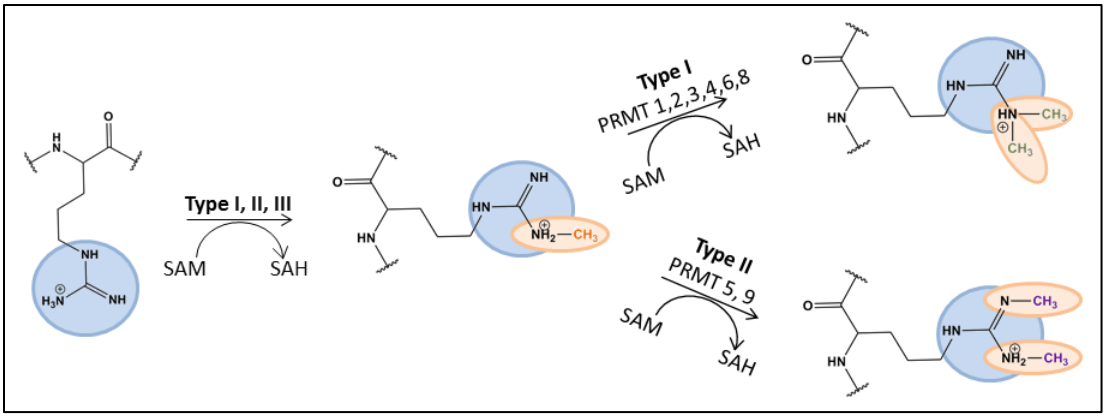

Figure 11. Arginine methylation patterns

Humans have nine protein arginine methyltransferases (PRMTs) that have been identified and divided into three categories depending on their final products, presented in

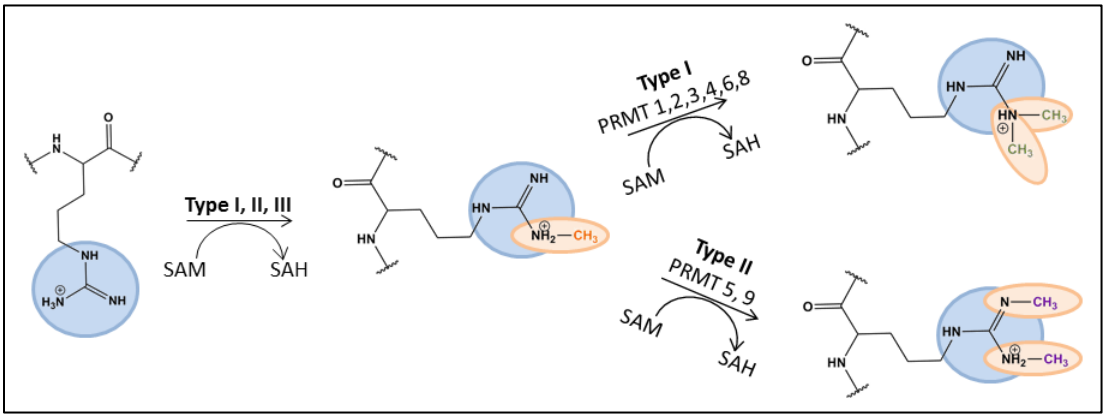

Figure 11 (2). All PRMTs are able to mono-methylate the terminal nitrogens of the guanidinium group of the arginine residues (ω-N^G-monomethylarginine) (MMA), though PRMT 7, the only identified Type III PRMT, is capable only of mono-methylating arginine

residues. Type I PRMTs (PRMT 1,2,3,4 (also known as CARM1), 6,8) are able to further modify the mono-methylated nitrogen, producing an asymmetrically di-methylated arginine residue (RMe2A or ADMA) (ω-N^G,N^G-asymmetric dimethylarginine). Alternatively, monomethylated arginine residues can be methylated on the unmodified nitrogen creating a symmetrically dimethylated arginine residue (RMe2S or SDMA) (ω-N^G,N^G-symmetric dimethylarginine). PRMT 5 and 9 are Type II PRMTs, though PRMT5 is thought to be the predominant Type II enzyme (2, 3). Both PRMT1 and PRMT5, the primary enzymes for RMe2A and RMe2S respectively, have been shown to be essential for cell viability; complete knockout of either enzyme affects embryonic viability (4, 5). While arginine methylation is thought to be a dynamic modification similar to lysine methylation, no arginine demethylase has been identified. However, a few lysine demethylases have been shown to demethylate modified arginine residues *in vitro* (3).

With only nine identified PRMTs, there are commonalities between them as well as significant differences. In general, PRMTs form head-to-tail homodimers with binding sites in a conserved catalytic core made out of a seven stranded beta barrel that binds both the substrate arginine and SAM (2). Within the catalytic core of all PRMTs, there is a "double-E loop" that correctly positions the guanidinium group of the arginine substrate through hydrogen binding with two glutamate residues. Type II enzymes have a conserved phenylalanine residue that catalyzes the formation of the symmetric dimethylation; Type I enzymes have two conserved methionine residues that appear to direct the asymmetric dimethylation (2, 6). Dimethylation by PRMT5 and its known binding partner MEP50 has been shown to be a distributive process, with RMe2S only being produced after MMA concentration was greater than unmodified substrate (7, 8). Alternatively, PRMT1 has been shown to be both semi-processive and semi-distributive, the re-binding process thought to be driven by the substrate peptide sequence (9). Another component to PRMT function is substrate motif. Glycine and arginine rich regions, or GAR motifs, are recognized as common targets for arginine methylation. Individual PRMTs show evidence to deviate from the GAR motif. CARM1 has a preference for PGM motifs (proline, glycine, and methionine rich

sequences). PRMT7 follows a 'RxR' motif. PRMT5 has the capability to modify both GAR and PGM motifs (2, 3).

Many of the roles for arginine methylation in the cell involve protein binding or DNA and RNA binding. Arginine residues use hydrogen binding and electrostatic interactions as the mechanism of binding. With its side chain guanidinium group, arginine is the most basic amino acid residue with a pKa value in solution of approximately 14.2 and is protonated at physiological pH (10). Methylation of arginine residues does not reduce the basicity of the residue as the pKa of a monomethylated arginine (MMA) has been measured at 13.4 (10). Rather, the mono and di-methylation increase the hydrophobicity and size of the residue while remaining positively charged. Dimethylation reduces the potential hydrogen binding donors of the guanidinium group from five to three, with the symmetry of the methylation pattern effecting the potential hydrogen binding pattern (11). Thus, the methylation pattern can direct binding preferences; not only is the number of methylations valuable but the symmetry is as relevant.

Arginine methylation of histone proteins has been studied and highlights the importance of the geometry of dimethylated arginine residues. Histone proteins H3 and H4 have arginine residues close to the amino termini of the proteins, H3R2 and H4R3, respectively. Both H3R2 and H4R3 are capable of being either symmetrically or asymmetrically modified (12). While the common assumption is that methylation is frequently viewed as a transcriptionally repressive mark, assisting in the histone/DNA binding, this is not consistently true. PRMT5 was shown to symmetrically dimethylate H3R2, a transcriptional promoter mark, by its recruitment of WDR5 at cell growth genes (13). PRMT5 can also symmetrically dimethylate H4R3, recognized as a transcriptional repressor mark, around genes connected with metastasis suppression. Alternatively, the asymmetric dimethylation of H3R2 has been correlated with transcriptional repression and H4R3 has been correlated with transcriptional promotion (12). Readers of dimethylated arginine residues are able to distinguish between RMe2S and RMe2A, pointing to a refined method of cell signaling.

Expression of PRMTs is upregulated in a variety of cancers (13, 14). This upregulation can impact RNA splicing or genetic transcription. PRMT5 is a known regulator of p53, a key regulator of DNA damage repair. Up-regulation of PRMT5 yields increased methylation of p53, which suppresses p53 activity (3). Tumors with over-expression of PRMT5 have increased risk of metastasis due to suppression of p53 (14).

To further understand the mechanisms by which dysregulated PRMTs effect cell function and why targeting PRMTs is an effective chemotherapy, a large-scale approach to a global understanding of Type I and Type II enzymes was designed by David Shechter at Albert Einstein School of Medicine. The project makes use of Type I and Type II specific inhibitors in multiple cell types to probe the mechanisms of cancer progression in a series of three aims: first, identify Type I and Type II specific substrates; second, evaluate the transcriptome and alternative splicing pathways affected by PRMTs; third, identify mechanisms for tumor development and identification of druggable targets. To identify substrates of Type I and Type II enzymes, an experiment was developed using methyl-targeting PTMScan kits developed by Cell Signaling Technology. The experiment uses multiple treated cell lines and three separate antibodies to target the methylarginine states in an effort to identify all Type I and Type II PRMT substrates. The scale and complexity of the experiment has not been previously tried for the arginine methylome. Thus, the goal of this chapter is the development of an LC-MS/MS analysis method for treated samples enriched for methylated arginine residues by PTMScan technology.

PTMScan is a protocol and immobilized-antibody based enrichment method sold by Cell Signaling Technologies. Individual kits were developed for each of the arginine methylation states, each with monoclonal antibodies developed in collaboration with the Bedford lab toward a specific arginine methylation state (15, 16). The basic premise for the

protocol is outlined in

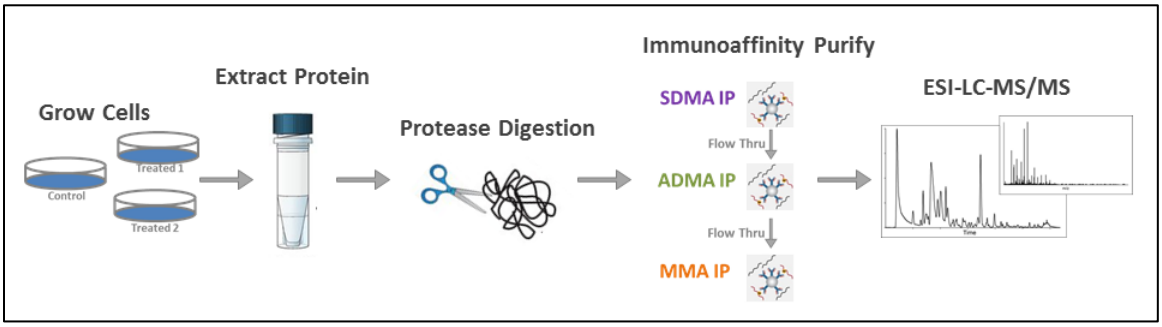

Figure 12. Cells are grown under the conditions of interest. In the experiments presented in this chapter, there were three sets of cells: control cells, Type I inhibited, and Type II inhibited. The cells were harvested, and the protein was extracted and digested by trypsin. The tryptic peptides were sequentially incubated with immobilized antibodies for anti-RMe2S, anti-RMe2A, and anti-MMA, respectively. The enriched, methylated peptides were then eluted from the antibodies for analysis.

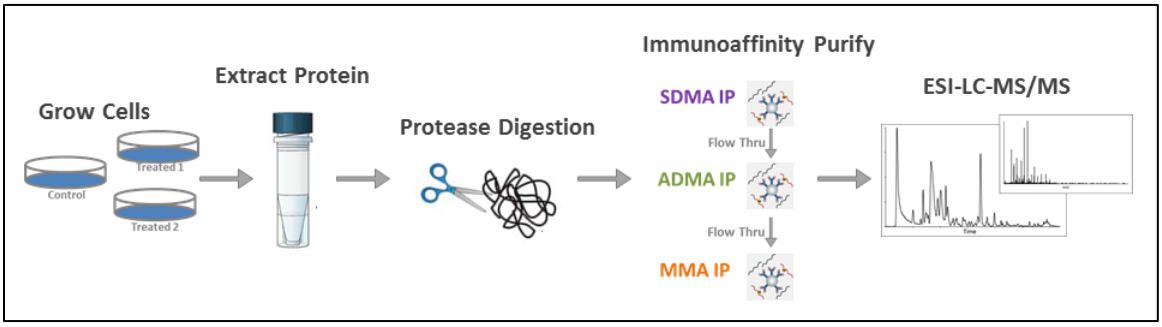

Figure 12. Experimental Design following the PTMScan Kit

The protocol outlined in

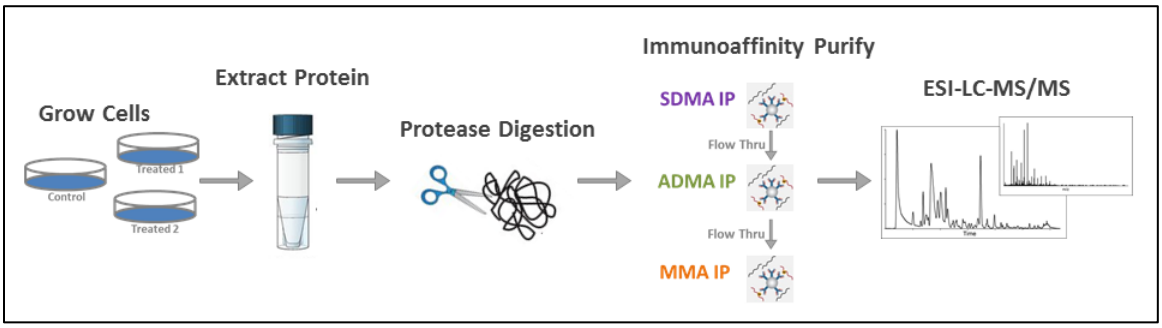

Figure 12 is similar to traditional, basic proteomic experiments such as those for phospho-proteomic experiments. Treated and untreated cells are then trypsin-digested and

immunoaffinity purified. The tryptic digest would generate peptides containing approximately two to four basic sites (amino terminus, lysine or arginine at the c-terminus, and perhaps another basic residue or two). These short peptides could be analyzed by LC-MS/MS using one of any number of traditional methods that have been published. There would be a modified peptide to account for (hopefully abundant and less abundant); and the results would be searched using traditional search methods: a handful of modifications and one or two missed cleavages.

However, there are specific constraints within the design of the methylarginine experiment and the protocol that create challenges compared to the basic proteomic structure. In order to understand the challenges, it is important to understand the impact of components within the analysis: the complex relationship between the motifs used to develop the antibodies, the trypsin digest, and the impact of the sample treatment.

The antibodies used for the immunoaffinity purification (IAP) were created by traditional methods. In all cases, rabbits were immunized with a mixture of degenerate peptides, containing a pattern of modified residues and randomized residues to develop antibodies that were specific toward the common motif, see

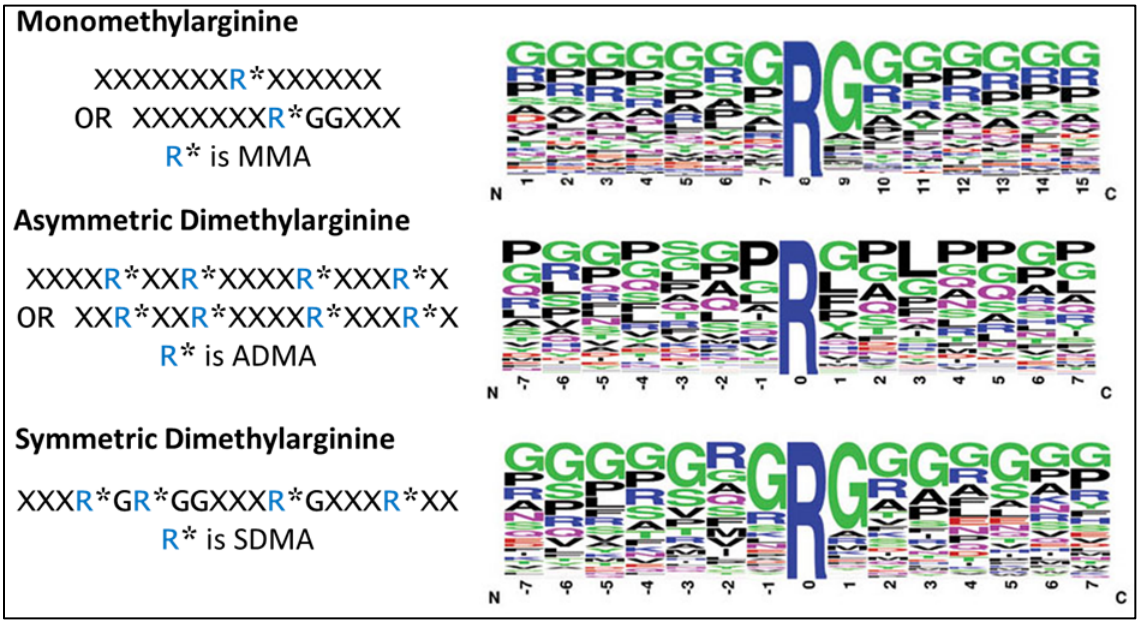

Figure 13. Cell Signaling Technologies does not published the epitopes used to develop the antibodies, though publications from Mark Bedford's lab are cited in the

provided protocols and have published motifs (15, 16).

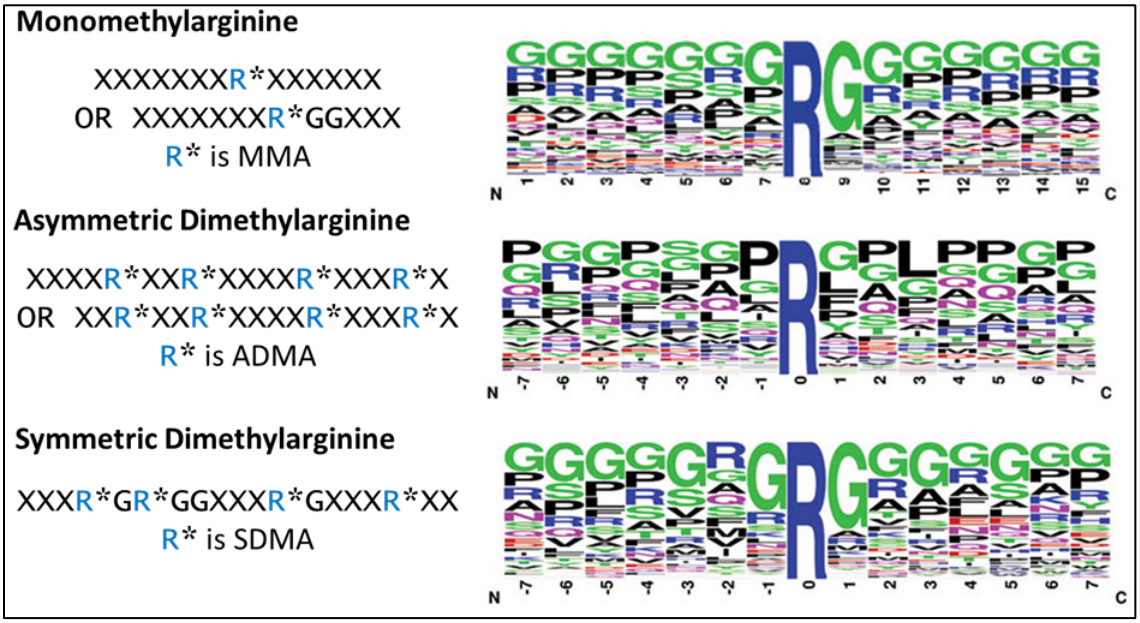

Figure 13 presents the motifs for each of the arginine methylation states as published from the Bedford research along with researchers from Cell Signaling Technologies (15, 16). The motifs used for both the anti-RMe2A and the anti-RMe2S, included four modified arginine residues, either all asymmetric or all symmetrically modified, respectively. As per the available references, the asymmetric and symmetric detecting antibodies were polyclonal antibodies; the monomethylated antibodies were potentially a panel of five monoclonal antibodies (16). In additional research, monoclonal antibodies against PRMT1 and PRMT4 substrates were created, thou Cell Signaling Technologies does not specify the antibody used for the PTMScan kit for ADMA (15). In the composite sequences, arginine residues are commonly enriched in peptides isolated from the samples they used, along with glycine and proline residues. The monomethylation motif included a single modified arginine residue. The peptides in the experiments in this chapter were sequentially enriched for RMe2S, RMe2A, and MMA, in that order.

Since the antibodies were developed towards multiply modified peptides, there should be a clear preference for the multiply modified peptides. Though the modified peptides for the antibody motif were all consistently modified as RMe2S or RMe2A, one cannot assume that a multiply modified peptide isolated in a particular IAP is consistently all

RMe2S or RMe2A. The peptide extract was depleted sequentially by the anti-RMe2S IAP, the anti-RMe2A IP and the anti-MMA IAP. The order of IAP should be considered in the experiment. Furthermore, if a peptide is no longer modified due to cell treatment options, it may not be detected from any of the IAP results.

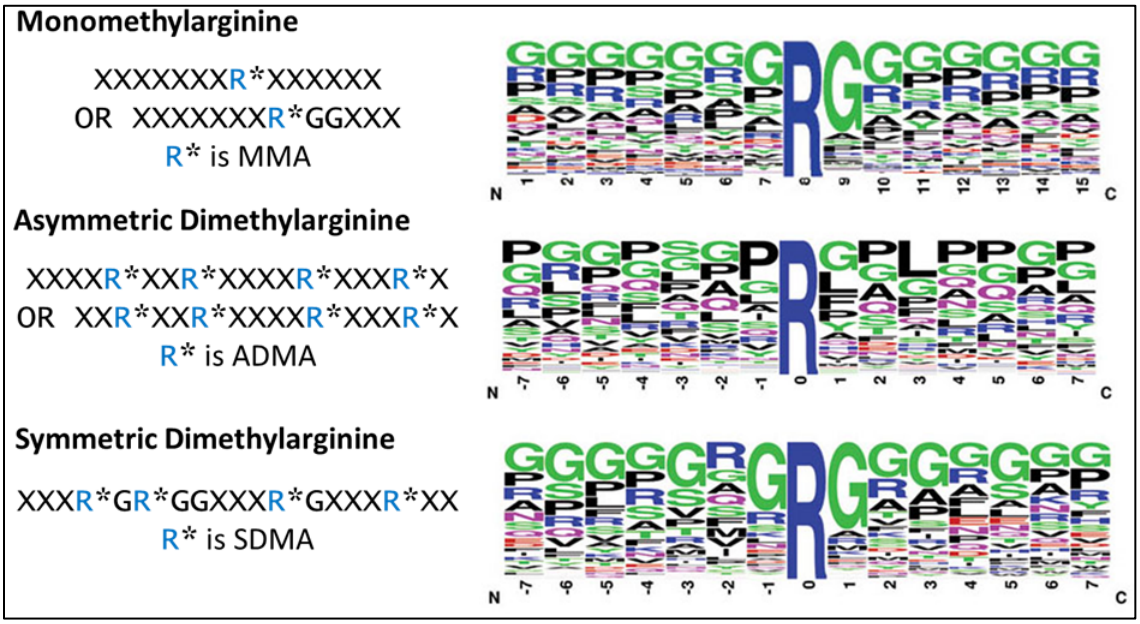

Figure 13. PTMScan Motifs; the motifs used are presented on the left with X representing any potential amino acid residue except tryptophan, cysteine, and tyrosine (15, 16). The sequences on the right are published by Cell Signaling Technologies, identified peptides are compared and the frequency of amino acids in positions relative to the modified arginine are represented by size. The peptide maps were generated using non-redundant human tryptic peptides immunoprecipitated by the MMA antibodies from Cell Signaling Technologies under their search conditions (37-39).

The antibodies were utilized to immunoaffinity purify peptides from the trypsin digest. Trypsin has a primary specificity to cleave C-terminal to lysine and arginine residues, with the exception of those N-terminal to a proline. However, trypsin is less reactive with modified arginine residues (1, 17). Thus, a peptide containing two dimethylated arginine residues will be expected to contain at least two missed cleavages and will have a minimum of four positive charges. Every modified arginine residue will lead to an increase in the number of missed cleavages and an increase to the number of positive charges. The experiment is complicated by both the PTM and the proteolytic cleavage occurring on the same amino acid residue. One would expect that changes to the number of modified arginine

residues due to a decrease in PRMT activity would impact the length and mass of the peptides created from the trypsin digest.

Finally, a central component to the experiment is that the cells are being treated for Type I and Type II enzyme inhibitors. There are three cell types: control cell (untreated), Type I inhibited, and Type II inhibited. Type I enzymes were selectively inhibited with MS023, a potent, small molecule inhibitor that binds in the substrate binding pocket of Type I enzymes with an IC_{50} = 4-119 nM (18). Type II enzymes were inhibited with GSK591 (EPZ015866), which competitively binds in the peptide substrate pocket of PRMT5, as observed with an analogue EPZ015666, and has an IC_{50} = 4 nM for PRMT5 (19, 20). The structures for MS023 and GSK591 are shown in Figure 14. The cell treatment is expected to change the arginine methylation pattern. This will have direct effects on the cleavage of the peptides by trypsin. A peptide from the control sample containing two dimethylated arginine residues could in a treated sample potentially now be cleaved at one or both of the arginine residues. The unmodified peptide would not be expected to be immunoaffinity purified; if one or both of the arginine residues was mono-methylated, it would be detected in the MMA enrichment.

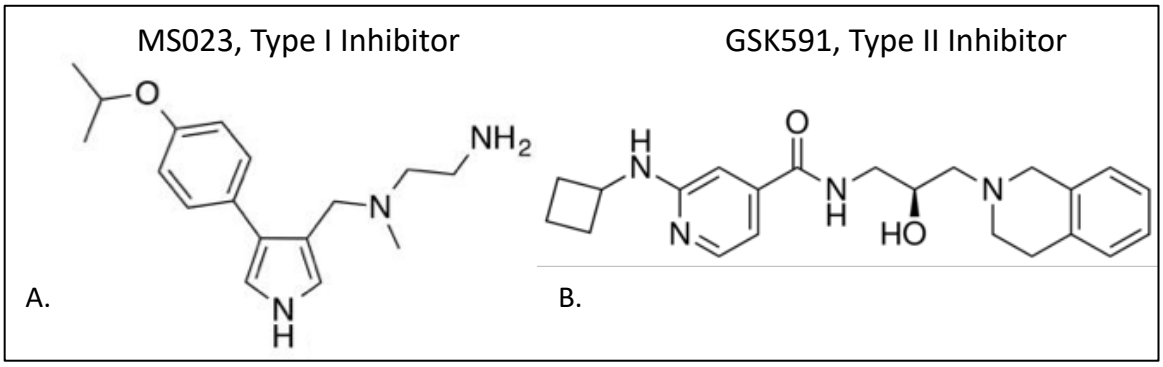

Figure 14. Type I inhibitor (A) and Type II Inhibitor (B)

Previous proteomic studies have been focused on arginine methylation. The studies focus on general arginine methylation identification on a proteome scale such as Uhlmann et al., Larsen et al., and Wang et al (1, 21, 22). Alternatively, knockdowns of a single PRMT are evaluated, probing samples to identify substrates specific to a single enzyme such as the investigation of CARM1 by Shishkova et al, or looking at methylation in a single cell type such as Geoghegan et al. Common to many of these tryptic peptide based studies are the MS^2

fragmentation type of HCD and the limitations within the database search to two or three missed cleavages (23, 24). Thus, the resulting data is limited by peptides that fit the instrument method parameters and the database search results.

The combination of the multiply modified peptides along with the trypsin digestion, sequential IPs, and cell treatment creates a complicated network of peptides. The goal of this chapter is to address these problems. A combination of the advancements from the Thermo Fusion™ Tribrid™ for sensitive and fast peptide analysis utilizing a customized instrument method makes methylated peptides a priority. A combination of Byonic by Protein Metrics and Proteome Discoverer 2.2 was utilized to search the spectra with adequate numbers of modifications and missed cleavages and provide a means of relative quantitation. Finally, manual inspection of side chain fragments for signature losses confirming dimethylated arginine residues is investigated. The developed method will be evaluated using detailed analysis of two Sm proteins.

2.3 Materials

Agilent Technologies (Palo Alto, CA)

1100 Series High Performance Liquid Chromatograph

1100 Series Vacuum Degasser

Anaspec, Inc. (Fremont, CA)

Vasoactive Intestinal Peptide; Human, porcine, rat; > 95% purity

ATCC (Manassas, VA)

A549 Cells (ATCC® CCL-185™)

Cayman Chemical (Ann Arbor, Mi)

MS023; Methyltransferase Type I Inhibitor, PN 18361

GSK591; Methyltransfersase Type II Inhibitor, PN 18354

Cell Signaling Technologies (Danvers, MA)

PTMScan® Mono-Methyl Arginine Motif [mme-RG] Kit, PN 12235

PTMScan® Asymmetric Di-Methyl Arginine Motif [adme-R] Kit, PN 13474

PTMScan® Symmetric Di-Methyl Arginine Motif [sdme-RG] Kit, PN 13563

Dr. Maisch GmbH

Reprosil-Pur 120 Å C18-AQ, 3 μm

Eppendorf (Hauppauge, NY)

5414R Benchtop centrifuge

Honeywell (Morristown, NJ)

Burdick and Jackson® Acetonitrile, LC-MS grade

Molex (Lisle, IL)

Polymicro Technologies™ polyimide coated fused silica capillary,

360 μm o.d. x 75 μm i.d.

New Objective (Woburn, MA)

PicoClear™ Union Assembly

Phoenix S&T (Chadds Ford, PA)

Pencil Column Heater

Column Heater Controller

PQ Corporation (Valley Forge, PA)

Kasil – Potassium silicate solution

Protein Metrics (Cupertino, CA)

Byonic™

Sigma Aldrich (St. Louis, MO)

Angiotensin I acetate salt hydrate, ≥99% purity (human)

Fluoranthene, >99% purity

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

Thermo Fisher Scientific (San Jose, CA; Bremen, Germany)

Formic Acid, LC-MS Grade

Orbitrap Fusion™ Tribrid™ mass spectrometer

Proteome Discoverer™ 2.2, Beta

Pierce Retention Time Calibration Mixture

Pierce Trypsin Protease

Pierce Water, LC-MS Grade

Waters (Milford, MA)

Sep-Pak® C18 Cartridge

2.4 Methods

Sections 2.4.1 and 2.4.2 were designed and performed by Drs. David Shechter and Hongshan Chen at Albert Einstein School of Medicine in Bronx, NY.

2.4.1 Experimental Design

Lung cancer cell line, A549, was grown in 4 confluent 10 cm cell culture dishes to $1x10^8$ cell equivalents. The culture was then divided into three groups of 15 cm plates and continued to grow for 7 days: 20 plates as a control, 30 plates as MS023 treated, and 25 plates as GSK591 treated. Treated cells were treated with 1 µM of MS023, a Type I methyltransferase inhibitor or 1 µM of GSK591, a Type II methyltransferase inhibitor. Cells were collected by plate scraping.

2.4.2 Immunoaffinity Purification of Methylation States

Cells were handled as per the PTMScan protocol developed by Cell Signaling Technologies. Briefly, harvested cells were diluted in a urea lysis buffer. Cells were lysed by sonication and centrifuged at 20,000 g for 15 min to extract proteins. Proteins were reduced with dithiothreitol and alkylated with iodoacetamide. The protein extract was then diluted 4x to a concentration of 2M urea, 20 mM HEPES at pH 8.0. Half of the protein extract was digested for 20 hours with 50 ug of trypsin. Complete digestion was confirmed by SDS-PAGE gel, as per Figure 15. The digest was acidified with TFA to halt the digestion and remove lipids. Following centrifugation for 15 min at 1780 g, peptides were separated from remaining cellular components and intact proteins by using a C18 Sep-Pak Column; 0.1% TFA and 40% acetonitrile was used to elute the peptides from the column. The peptides were lyophilized and diluted to 1.4 µL with immunoaffinity purification buffer containing 50 mM MOPS/NaOH, pH 7.2; 10 mM Na_2HPO_4, 50 mM NaCl. The peptides were then sequentially incubated with the antibody beads for 2 hours for each antibody: RMe2S, then RMe2A, then MMA. The beads were washed with immunoaffinity purification buffer, followed by HPLC

grade water. The peptides were eluted from each set of beads with 0.15% TFA and sent to University of Virginia for further analysis.

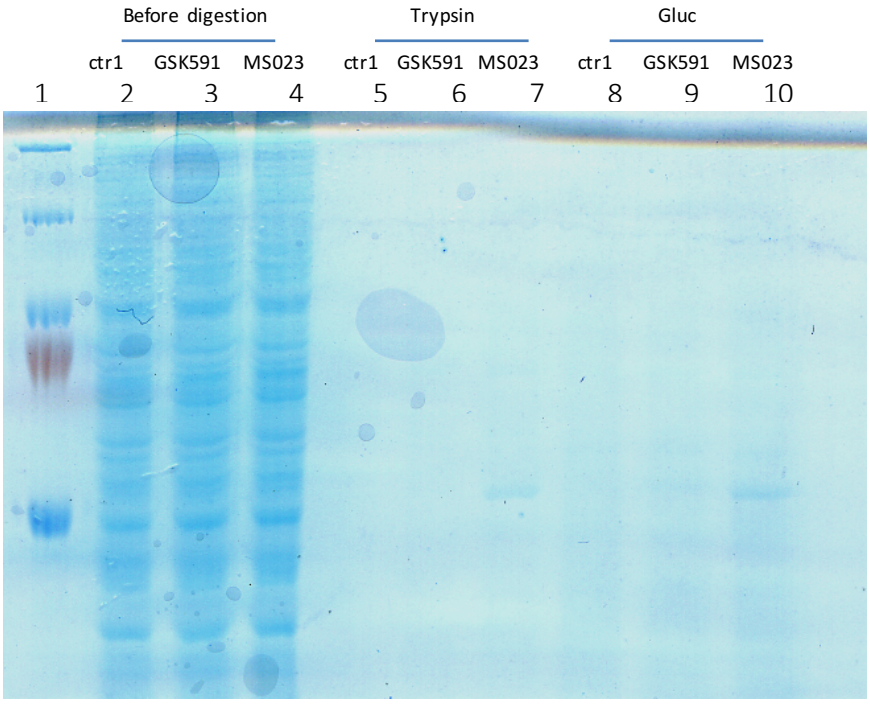

Figure 15. Gel shows protein extract and post-digestion remaining protein material; Coomassie stained gel with the following lanes 1) Molecular weight marker 2-4) Protein extract prior to digestion 5-7) Post trypsin digestion 8-10) Post GluC digestion

2.4.3 LC-MS/MS Method

To remove any remaining agarose beads, samples were centrifuged at 10,000 rpm for 30 seconds and an aliquot was transferred to a fresh tube. The column was pressure loaded with 10 fmol of Pierce Retention Time Calibration Mixture (PRTC); 100 fmol of angiotension and vasoactive; and 2.5 µL of sample, or approximately 5%, provided from the Shechter lab. The samples were analyzed using in-house prepared columns. Fused silica (360 um OD x 75 um ID) were fritted with a 1-2 mm kasil frit and packed to 22 cm with Dr. Maisch Reprosil, C18, 3 µm particles. The column was heated to 40C for separation. Reverse-phase separation was performed using 0.2% formic acid for solvent A and 0.2% formic acid in 80% acetonitrile for solvent B with a 60-minute gradient from 100% solvent A to 40% solvent B at approximately 100 µL/minute. Samples were introduced to the mass spectrometer by

electrospray ionization with an in-house laser pulled fused silica tip, approximately 1 µm diameter.

A mass spectrometric analysis was performed on a Thermo Orbitrap Fusion™ Tribrid™ in a data dependent manner. A high resolution MS^1 was acquired every 2 seconds from 300-1500 m/z with 60,000 resolution at 200 m/z, FWHM. Precursors were selected for fragmentation starting with the most abundant ions. Fragmentation and acquisition parameters were decided using a decision tree based on the precursors' charge state and m/z. The decision tree is presented in Figure 16. Low resolution MS^2 scans were acquired in the LIT with an AGC target of 10,000 for a maximum of 80 ms. High resolution MS^2 were acquired in the Orbitrap with 60,000 resolution and an AGC target of 75,000 for a maximum of 120 ms. A 2 m/z quadrupole isolation was used for each MS^2 scan.

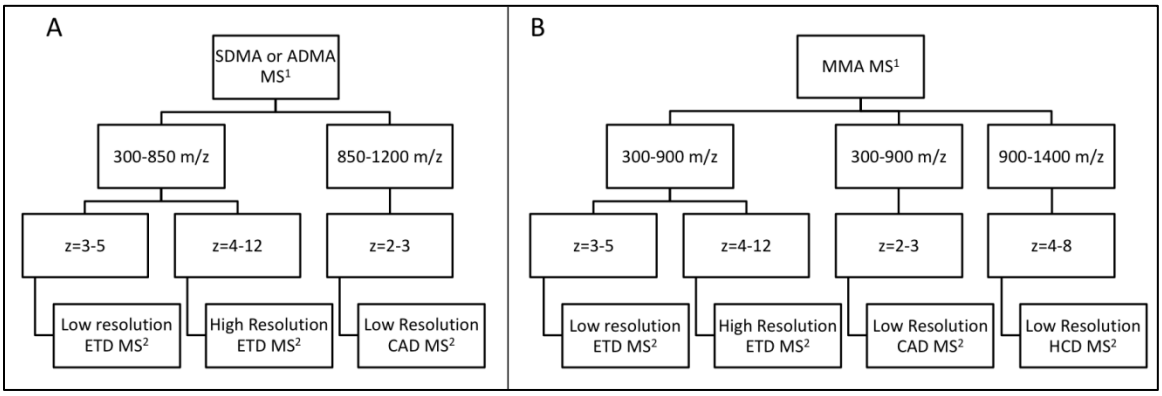

Figure 16. Fragmentation Decision Tree

2.4.4 Database Search Parameters

Byonic™ (Protein Metrics) was used to search the data against the human proteome. Proteome Discoverer™ 2.2 Beta (Thermo) was used to compile the results of the database searches. The results reported in section 2.5 have been manually confirmed. The human proteome database was downloaded from Uniprot on Dec. 15, 2015.

Raw data from dimethylated samples were search with Byonic in three searches: a PRTC Standard search, a low-resolution search, and a high-resolution search. The PRTC standard search utilized a database with decoys composed of the PRTC standards as well as Angiotension and Vasoactive. Low resolution scans were searched with 5 ppm precursor

mass tolerance and 0.4 Da fragment mass tolerance. The standard database was searched with no cleavages and 2 common modifications: dimethylarginine, and heavy labeled arginine and lysine. Low resolution ETD and CAD scans were also searched using the human database, including decoys and common contaminants. Results were searched with trypsin (c-term to K, R only) with 6 missed cleavages and maximum precursor mass of 4650 Da. Total common modifications were set at 8 and total rare modifications were set at 2. Table 1 presents the modifications searched for low resolution spectra. High resolution ETD scans were similarly searched using the human database, including decoys and common contaminants. Results were searched with trypsin (c-term to K, R only) with 12 missed cleavages, 5 ppm precursor tolerance, and 7000 Da maximum precursor mass. Total common modifications were set at 12 and total rare modifications were set at 2. Table 1 presents the modifications searched for high resolution spectra.

Raw data from monomethylated samples were search similarly to the dimethylated samples. In addition to the three searches described above, low resolution HCD spectra were searched using the human database, including decoys and common contaminants. Results were searched with trypsin (c-term to K, R only) with 6 missed cleavages, 5 ppm precursor tolerance, and 5500 Da maximum precursor mass. Total common modifications were set at 7, and total rare modifications were set at 2. Table 1 includes the modification settings.

Table 1. Static and Dynamic Modifications

Search	Modification	Mass	Residue	Fixed or Static	# Allowed
Low Resolution CAD/HCD/ETD	Carbamidomethyl	57.0215	C	Static	--
	Dimethyl	28.0313	R	Common	6
	Methyl	14.0157	R	Common	4
	Acetyl	42.0106	K; Protein N-term	Rare	1
	Trimethyl	42.04695	K	Rare	1
	Dimethyl	28.0313	K	Rare	2
	Methyl	14.0157	K	Common	2
Standards	Dimethyl	28.0313	R	Common	2
	$^{13}C_6{}^{15}N_4$	10.0083	R	Common	1
	$^{13}C_6{}^{15}N_2$	8.0142	K	Common	1
High Resolution ETD	Carbamidomethyl	57.0215	C	Static	--
	Dimethyl	28.0313	R	Common	10
	Methyl	14.0157	R	Common	5
	Acetyl	42.0106	K; Protein N-term	Rare	1
	Trimethyl	42.04695	K	Rare	1
	Dimethyl	28.0313	K	Rare	2
	Methyl	14.0157	K; Peptide N-term	Common	2

2.4.5 Arginine Side Chain Fragmentation

Electron transfer dissociation (ETD) is an optimal fragmentation method for highly charged peptides. As discussed in Chapter 1.2.9, ETD prefers to cleave the N-Cα bond in the peptide backbone, though a• and y- ions are also detected reaction products. Arginine residues are the most basic of the common amino acids with the guanidinium group being protonated at physiological pH (17). (Gehrig, 2004) This can result in cleavage of the arginine side chain during ETD reactions (25–27). Fragmentation of arginine side chains is also possible during CAD and HCD, both collision-based methods (17, 28).

While RMe2S and RMe2A are isobaric peptides, the guanidinium group is modified such that the side chain fragments from the two geometries have different masses (27). As discussed previously, RMe2S has two methylated nitrogens and RMe2A has a single doubly modified nitrogen. Both dimethylarginines will have a potential loss of 71 Da, corresponding to a dimethylcarbodiimidium (26, 27). However, the symmetric arginine will have a loss of 31 Da for monomethylamine and the asymmetric arginine will have a loss of 45 Da for the dimethylamine (26, 27).. Alternatively, MMA will have a loss of 31 Da for the

monomethylamine and a loss of 57 for the monomethylcarbodiimidium (26, 27). Figure 17 presents a summary of side chain fragments from modified arginines.

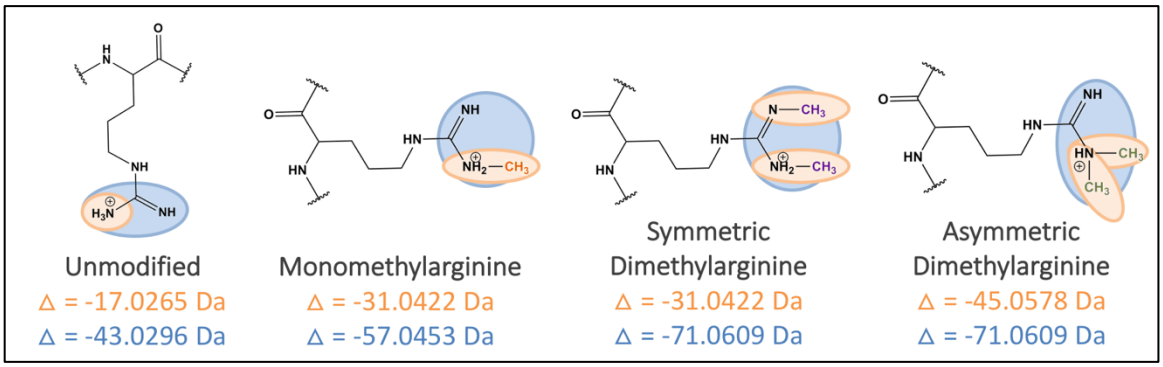

Figure 17. Arginine Side Chain Fragments, the orange masses correspond to the orange highlighted structure and the blue masses correspond to the loss of the blue highlighted structure.

The side chain fragments themselves were not detected in the MS^2 spectra for the peptides in question. Rather, the resulting peptides with the loss of the side chain were detected in the spectra. As a charge is neutralized during ETD fragmentation, the mass of the product ion is less the mass of the side chain fragment and a single charge. Example spectra are presented in Figure 18. Manual inspection was performed for all spectra reported below. The ratio of the relative abundances for the fragments of -45 Da and -31 Da was recorded when viewing the first charge reduced precursor to the precursor – 71 D. Figure 18B and D would be reported as (40/0) and (60/0), respectively.

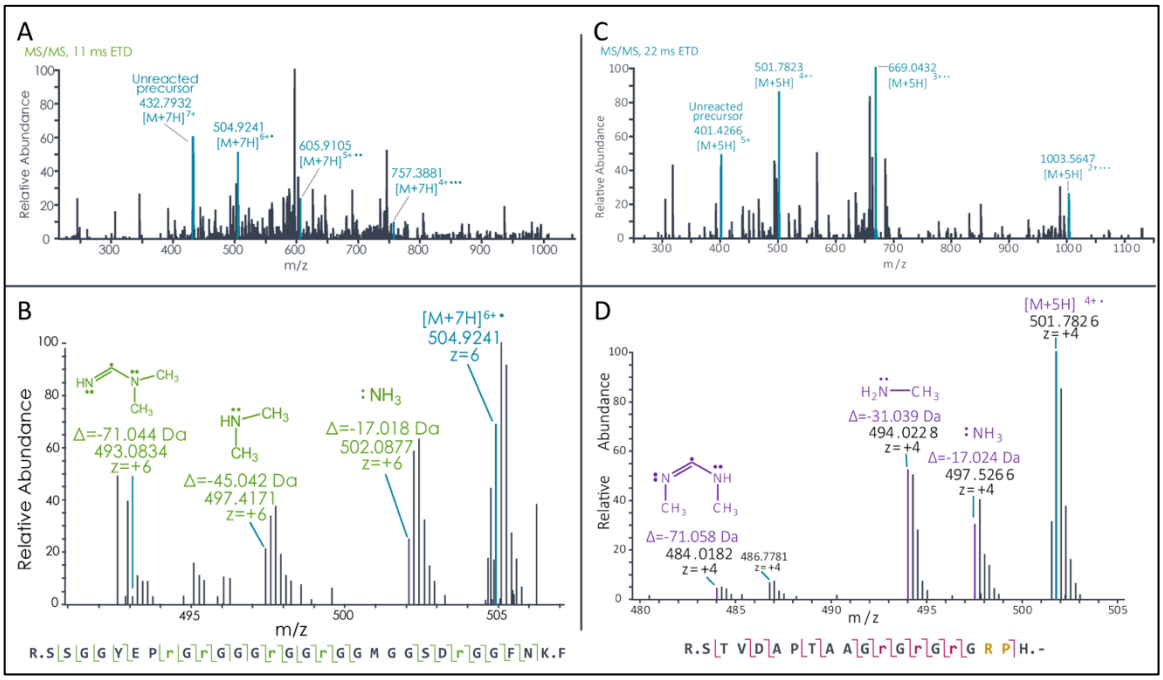

Figure 18. Example spectra with dimethyl side chain losses A) Full mass range 7+ MS^2 from RNA binding FUS protein B) Expanded region to show characteristic, asymmetric side chain losses C) Full mass range 5+ MS^2 from HIV-1 Nef-interacting protein D) Expanded region to show characteristic, symmetric side chain losses.

2.5 Results and Discussion

During the initial method development, ETD was prioritized over other fragmentation methods. The final methods rely solely on ETD fragmentation for precursors with charge states 4+ and less than 850 m/z. Initial results from methods acquiring both ETD and CAD for each precursor resulted in duplicate identifications at the protein level; only a few methylated proteins were uniquely identified by CAD as per Figure 19. The spectra of methylated peptides fragmented by ETD were rich with sequence informative ions compared to their CAD counterparts, consistent with results from Hart-Smith et al (29). Example spectra are presented in Figure 20. Finally, the charge state distribution of peptide spectral matches confirmed that methylated peptides had higher charge states on average than unmodified peptides, optimal for ETD fragmentation methods, see Figure 19.

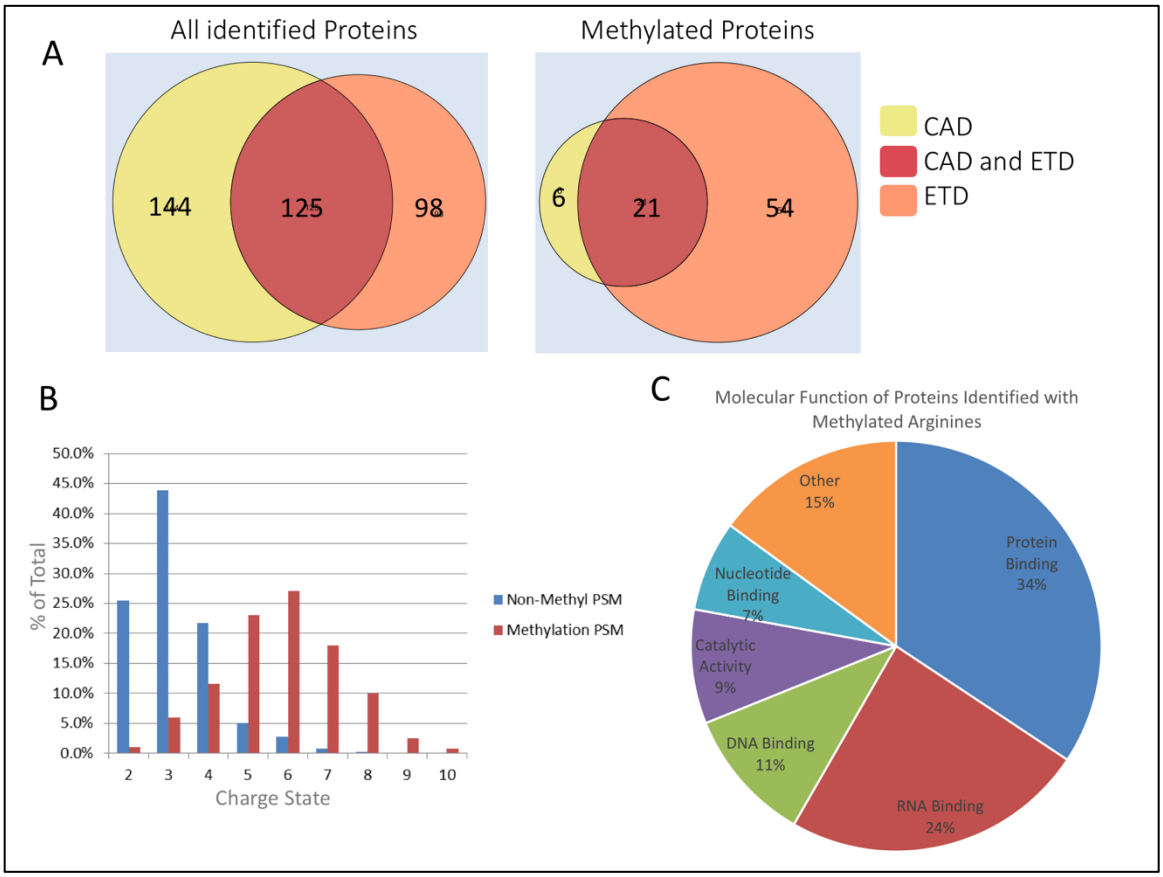

Figure 19. Trends identified during method development A) Human proteins identified from control sample show the trend that 50% more proteins are identified by CAD when including all identified proteins (left) but more proteins are exclusively identified by ETD when sorting for proteins identified from peptide spectral matches (PSM) containing methylarginine residues B) Methylated PSM have a broader charge state distribution C) Molecular function of proteins from all samples identified with methylarginine residues, (Figures were created in Proteome Discoverer 2.2 Beta based on the Human Uniprot database downloaded Dec. 15, 2015)

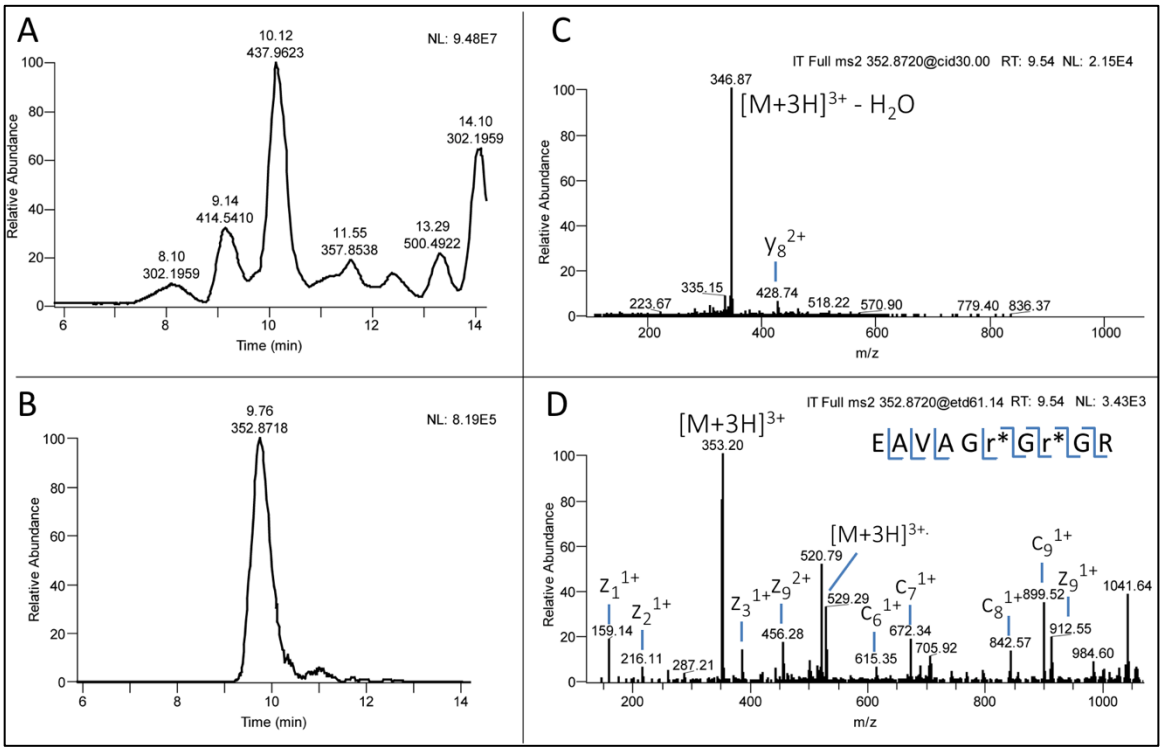

Figure 20. ETD and CAD comparison of 2x monomethylated peptide A) Base peak B) Extracted ion chromatogram of 3+ precursor, 352.8718 C) CAD spectrum D) ETD spectrum

The database search results were compiled; and normalized, methylated peptide abundances were summed for each protein in each condition. Ratios were calculated comparing treated protein abundance to control protein abundance for each pulldown, i.e. symmetric inhibited/control for symmetric pulldown and symmetric inhibited/control for asymmetric pulldown. Appendix 1, Appendix 2, and Appendix 3 present a summary of the top 25 potential results for Type I inhibited, Type II inhibited, and monomethylated samples, respectively. The proteins were ranked by abundance using for RMe2A Control for Appendix 1, RMe2S Control for Appendix 2, and summed methyl peptide abundance for Appendix 3. The ratios in the tables are coded in a gradient of red (lowest) to blue (highest) values. The experiments were performed using label free relative quantitation. For the purposes of this discussion, only results that are at least a 5-fold difference will be considered significant.

Due to several of the experimental conditions, interpretation of the protein ratios is not straight forward. The targeted PRMT activity could not be fully inhibited; PRMT1 and PRMT5 knockouts have been shown to be lethal for cells (4, 13). The digested peptides were

modified on the same residues that were targets for digestion. Therefore, decreasing the enzymatic activity with drug treatment is expected to change the modification patterns on arginine residues, which then changes the cleavage patterns of those peptides. Finally, the protein abundances are calculated on summed peptide abundances. It is possible that the treated sample has similar abundances to the untreated sample, but the peptides are modified in different patterns. Furthermore, methyltransferase activity has been shown to be less specific in enzymatically depleted cells as substrate scavenging has been reported in PRMT1 knockdown cell lines (16). Because of the constraints of the method, the tables summarizing the top results represent a way of organizing the data rather than final results.

The experimental design has the potential to create a complicated web of results. In addition to the challenges posed with the digestion, each of the three cell conditions was enriched for each of the three types of modified arginine residues. While one might think that the asymmetrically enriched samples can act as a control for the symmetrically treated cells, this might not always be the case. The understanding of why a particular arginine residue or series of residues are modified as symmetric or asymmetric is not clearly understood. There are reports of methylarginine scavenging by Type II PRMTs after a Type I knockdown, resulting Type II enzymes symmetrically modifying additional substrates (16). Thus, the opposite pulldown of the inhibited enzyme should not be expected to be similar to the control. Furthermore, while the epitopes to the antibodies sold by Cell Signaling Technologies can be speculated about, they have not provided the epitopes or evidence of the validation of the antibodies. For this reason, it is difficult to know if the antibodies are enriching for different motifs. The results returned from search algorithms include summed abundances at the protein level and at the peptide level. As the search algorithm uses intact mass to distinguish peptides, the same peptide may be returned with varying numbers of methylations as several different results. Thus, the same peptide may look depleted in a treated sample compared to the control only to see the same peptide with fewer methyl groups increased compared to the control. At the protein level, the search algorithm will return results summed for the protein. While a result may be reported as the same for a treated sample, it could be that a region of methylated peptides are missing compared to the

control but this would not be reflected on the summed protein level. Furthermore, because the results are compared by label free relative quantitation, a change in the order of magnitude would be considered to be appropriate to a real change. However, as the effectiveness of the antibodies is not known, it is difficult to know what is a significant change. Given the complexity of the experimental design, It is difficult to evaluate the effectiveness of the developed method and the results returned by automated search algorithms. For this reason, two proteins, SmD1 and SmD3, were chosen for detailed manual evaluation.

SmD1 and SmD3 have roles in pre-mRNA splicing as components of snRNPs, complexes of snRNA and RNA binding proteins that are major components of the spliceosome. Following transcription of small nuclear RNAs (snRNA), the survival of motor neuron (SMN) complex load a seven-membered protein ring (Sm core) onto the snRNA at a conserved nucleotide site (30, 31). The snRNP complex is then imported into the nucleus for use in pre-mRNA splicing. SmD1 and SmD3 are two of the seven Sm proteins (30, 31). Prior to association with the SMN complex, SmD1 and SmD3 are reported to be dimethylated in their RG-repeats at the C-termini by the PRMT5-MEP50 containing Methylosome 20S complex (32–34). The SMN proteins contain tudor domains that are known binding motifs of symmetrically modified arginine residues (35). Additionally, asymmetric dimethylarginines were reported by Miranda et al. for nuclear associating Sm proteins, though these results are inconsistent with known snRNP nuclear shuttling (36).

The results of the protein abundance ratios reported in the appendices for both SmD3 and SmD1 are summarized in Table 2. The results for SmD3 will be reported first, followed by the results for SmD1. The related tables and figures are presented at the end of this section.

Table 2. Summary of Sm Protein Results pulled from Appendix 1-3 (Protein abundance ratios are calculated using results from Proteome Discoverer, the sum of all methylated peptides from either Sm-D3 or Sm-D1 for detected for a treated sample are compared to the sum of the

abundances of methylated peptides for the same protein in the control sample of the same pulldown)

	Protein Abundance Ratio					
	Asymmetrically Inhibited Samples			Symmetrically Inhibited Samples		
Protein	**Symmetric Pulldown**	**Asymmetric Pulldown**	**Monomethyl Pulldown**	**Symmetric Pulldown**	**Asymmetric Pulldown**	**Monomethyl Pulldown**
Sm-D3	2	0.2	1	0.5	1	2
Sm-D1	0.6	1	3	0.001	3	0.9

The results for SmD3 in all enrichments are presented in Table 3, Table 4, and Table 5. The RMe2S pulldowns had similar results for all cell conditions, with results varying 2-fold between the control and each of the treated conditions. The total peptide abundances for the RMe2A and RMe1 were each less than 1% of the detected peptide abundances for the symmetric pulldowns. While cells treated for GSK591 had similar detected abundances of modified peptide, there was a greater variety in the detected patterns of modifications, see Table 5, suggesting that the dimethylation was less consistent in the Type II inhibited samples. Only a single peptide was detected that was cleaved prior to residue 124 in SmD3 (Table 5, line 4); all other detected peptides contained at least residues 115-124.

An example of the side chain losses for SmD3 is presented in Figure 22 for the RMe2A pulldowns and in Figure 23 for the RMe2S pulldowns. No side chain losses are presented for the RMe1 pulldowns; a loss of 31 Da would be expected for every mono-methylated peptide. The RMe2A pulldown results do not show much of a change in the pattern of side chain losses between the control and the GSK591 treated cells. As SmD3 is reported to be symmetrically modified, one would not expect to find SmD3 peptides in the pulldown for RMe2A. However, in the RMe2S pulldowns, there is a change in the relative abundance of the -45 Da side chain fragment in the GSK591 treated cells compared to the control and the MS023 treated cells. The relative abundance of the -45 Da fragment (suggesting asymmetric dimethylation) increases to approximately 50% relative abundance in the samples treated with a symmetric inhibitor.

Some similar patterns are observed with SmD1. Summaries of peptide results for each cell type in each pulldown are reported in Table 6, Table 7, and Table 8. Similar levels of SmD3 were detected in the control and MS023 treated cells for the RMe2S pulldowns, see Table 7. There were very few detected peptides for the GSK591 treated cells in the RMe2S

pulldown, less than 1% of the abundance detected in the control and MS023 treated cells. Additionally, there were a greater number of different peptides (cleavage/modification patterns) that were detected for the MS023 treated cells than the control in the symmetric pulldown. In the RMe2A pulldown, there were more peptides detected for the GSK591 treated cells than the other two conditions, see Table 6. Finally, the peptides from the control samples have a more consistent modification pattern than the treated samples as both MS023 and GSK591 have a greater number of clipped peptides in the RMe1 pulldown.

The side chain losses for a fully dimethylated peptide of SmD1 are presented in Figure 24 for the RMe2A pulldown and Figure 25 for the RMe2S pulldown. The asymmetric pulldown shows a loss of both 45 Da and 31 Da in the control. However, the loss of 45 is not detected in the MS023 treated cells, and the loss of 31 Da is not detected in the GSK591 treated cells. Additionally, in the RMe2S pulldown, the loss of 31 Da is detected in each of the cell lines, though only in the GSK591 cells is the relative abundance of the fragment for the loss of 45 greater than the loss of 31.

SmD1 and SmD3 are reported to be symmetrically dimethylated by PRMT5, a processive methyltransferase, (33). However, the Type II inhibited treatment appears to reduce the dimethylation of SmD1 but does not result in a reduction of SmD3 methylation. This finding suggests that they are not methylated through the same pathways. Yet, both SmD1 and SmD3 are still impacted by MS023 treatment with an increase in the variety of truncated peptides observed Furthermore, SmD3 treated with GSK591 has an increased variety of truncated peptides. Finally, SmD3 appears to have an increase in the abundance of asymmetrically modified arginine residues after treating to inhibit symmetric methylation. Taken along with observations from the Shechter lab, the results point to a better understanding of how PRMT5 regulates RNA splicing. qRT-PCR results reported in Figure 21, demonstrate that treatment with PRMT5 inhibitor GSK591 results in a large increase in retained introns, though treatment by MS023 does not appear to impact RNA splicing.

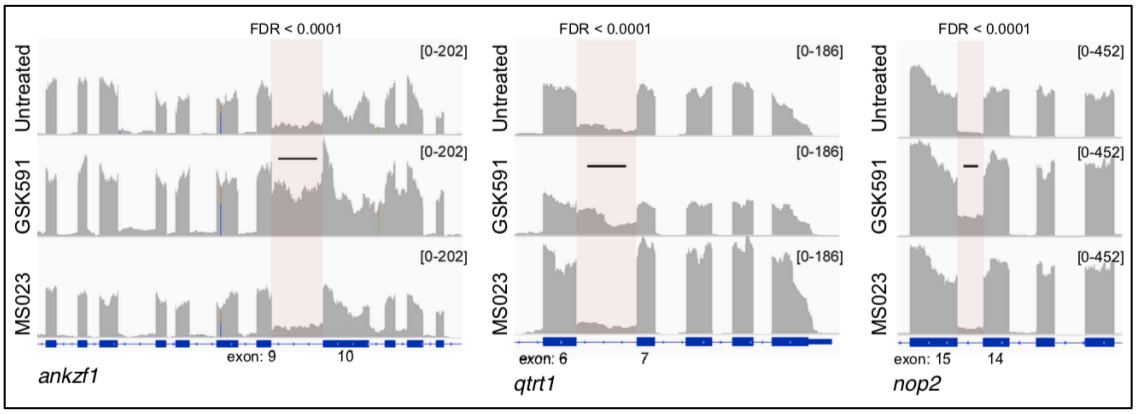

Figure 21. qRT-PCR results for three genes; the pink region denotes exon/intron junction measured by the primer in black (unpublished results)

Table 3. Summary of Peptides for SmD3 in Asymmetric Pulldowns

Asymmetric IP							Abundance			Symmetry		
	Residues			Methyl		[M+H]	Control	MS023 Treated	GSK591 Treated	Control	MS023 Treated	GSK591 Treated
1	105	124	K.AQVAARGRGRGMGRGNIFQK.R	8	AQVAAr28Gr28Gr28GMGr28GNIFQK	2242.2826	3.80E+08	PH: 6.0e5	1.40E+09	(30/90)	MS^1	(45/85)
2	111	124	R.GRGRGMGRGNIFQK.R	6	Gr28Gr28GMGr28GNIFQK	1617.9118	2.20E+08	ND	ND	(5/100)	--	--
3	111	124	R.GRGRGMGRGNIFQK.R	5	Gr14Gr28GMGr28GNIFQK	1603.8962	ND	1.20E+08	ND	--	(100/0)	--
4	111	124	R.GRGRGMGRGNIFQK.R	4	G[rGr]3x14GMGr14GNIFQK	1589.8806	1.60E+07	ND	ND	Undetermined	--	--
						Total	6.16E+08	1.20E+08	1.40E+09			
						Normalized	6.16E+08	1.07E+08	1.13E+09			

MS^1: no MS/MS available; ND: Not Detected

Table 4. Summary of Peptides for SmD3 in Symmetric Pulldowns

Symmetric IP							Abundance			Symmetry		
	Residues			# Methyl		[M+H]	Control	MS023 Treated	GSK591 Treated	Control	MS023 Treated	GSK591 Treated
1	105	124	K.AQVAARGRGRGMGRGNIFQK.R	8	AQVAAr28Gr28Gr28GMGr28GNIFQK	2242.2826	1.19E+11	9.84E+10	7.33E+10	(30/85)	(10/100)	(45/80)
2	105	124	K.AQVAARGRGRGMGRGNIFQK.R	8	AQVAAr28Gr28Gr28Gm16Gr28GNIFQK	2258.2775	4.14E+09	4.46E+09	5.03E+09	(5/100)	(5/100)	(15/100)
3	105	124	K.AQVAARGRGRGMGRGNIFQK.R	7	AQVAAr14Gr28Gr28GMGr28GNIFQK	2228.267	--	--	7.62E+09	--	--	(35/70)
					AQVAA[rGr]3x14Gr28GMGr28GNIFQK		1.13E+10	5.26E+10	--	(5/100)	(5/100)	--
4	105	124	K.AQVAARGRGRGMGRGNIFQK.R	6	AQVAAr14Gr14Gr28GMGr28GNIFQK	2214.2514	ND	7.70E+08	5.30E+08	--	(15/100)	(15/70)
5	111	124	R.GRGRGMGRGNIFQK.R	6	Gr28Gr28GMGr28GNIFQK	1617.9118	3.10E+10	2.40E+10	1.02E+09	(5/100)	(5/100)	(10/100)
6	111	124	R.GRGRGMGRGNIFQK.R	5	Gr14Gr28GMGr28GNIFQK	1603.8962	1.59E+09	3.62E+09	2.51E+08	(0/100)	(5/100)	(10/85)
7	111	124	R.GRGRGMGRGNIFQK.R	4	G[rGr]3x14GMGr14GNIFQK	1589.8806	4.13E+07	2.70E+08	5.48E+07	(10/90)	(15/100)	(5/80)
8	113	124	R.GRGMGRGNIFQK.R	4	Gr28GMGr28GNIFQK	1376.758	1.67E+08	ND	4.10E+08	(0/40)	--	(5/50)
9	113	124	R.GRGMGRGNIFQK.R	3	G[rGMGr]3x14GNIFQK	1362.7424	ND	ND	2.83E+07	--	--	(5/90)
10	115	124	R.GMGRGNIFQK.R	2	GMGr28GNIFQK	1135.6041	ND	ND	2.48E+07	--	--	(60/40)
						Total	1.67E+11	1.84E+11	8.83E+10			
						Normalized	1.67E+11	2.56E+11	1.30E+11			

Table 5. Summary of Peptides for SmD3 in Monomethylation Pulldowns

Monomethyl IP							Abundance			Symmetry		
	Residues			Methyl		[M+H]	Control	MS023 Treated	GSK591 Treated	Control	MS023 Treated	GSK591 Treated
1	105	124	K.AQVAARGRGRGMGRGNIFQK.R	7	AQVAAr14Gr28Gr28GMGr28GNIFQK	2228.267	ND	ND	9.42E+08	--	--	(75/30)
2	105	124	K.AQVAARGRGRGMGRGNIFQK.R	6	AQVAAr14Gr14Gr28GMGr28GNIFQK	2214.2514	ND	ND	4.60E+08	--	--	(65/50)
3	105	124	K.AQVAARGRGRGMGRGNIFQK.R	5	AQVAAr28Gr14Gr14GMGr14GNIFQK	2200.2358	ND	ND	2.60E+08	--	--	(15/45)
4	105	118	K.AQVAARGRGRGMGR.G	4	AQVAA[rGrGr]4x14GMGR	1498.8497	ND	ND	2.41E+07	--	--	(45/30)
5	111	124	R.GRGRGMGRGNIFQK.R	3	Gr14Gr14GMGr14GNIFQK	1575.865	6.84E+07	3.25E+08	3.00E+07	(0/55)	(0/60)	(0/80)
6	113	124	R.GRGMGRGNIFQK.R	3	G[rGMGr]3x14GNIFQK	1362.7424	ND	ND	4.54E+07	--	--	(15/25)
7	113	124	R.GRGMGRGNIFQK.R	3	Gr14GMGr14GNIFQK	1348.7268	4.77E+07	8.33E+07	1.85E+08	(0/100)	(0/45)	(0/35)
						Total	1.16E+08	4.08E+08	1.95E+09			
						Normalized	1.16E+08	3.35E+08	9.15E+08			

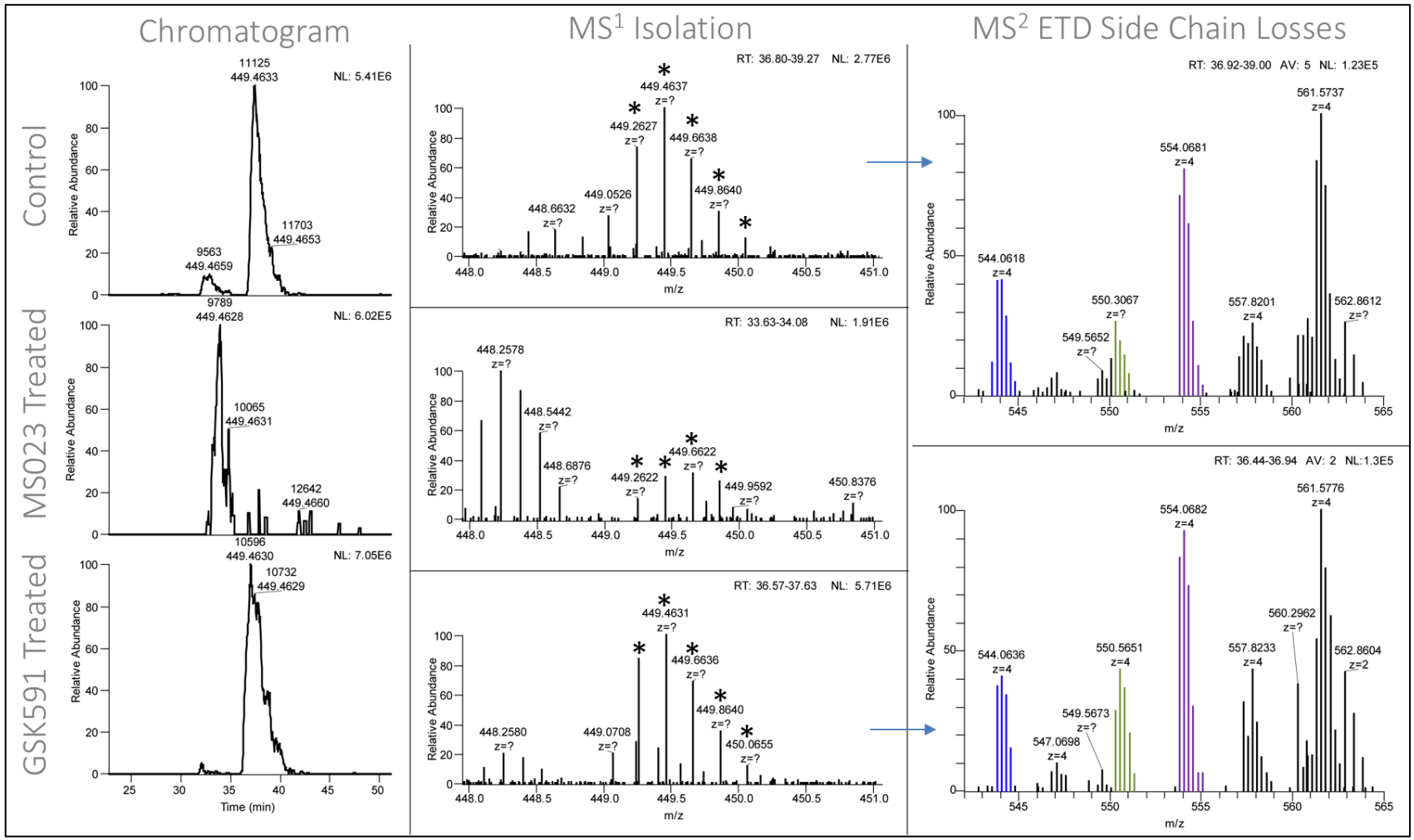

Figure 22. SmD3 Asymmetric Pulldown Side Chain Fragments for 5+ of peptide AQVAAr28Gr28r28GMGr28GNIFQK; ETD side chain losses marked by purple (-31 Da), green (-45 Da), blue (-71 Da)

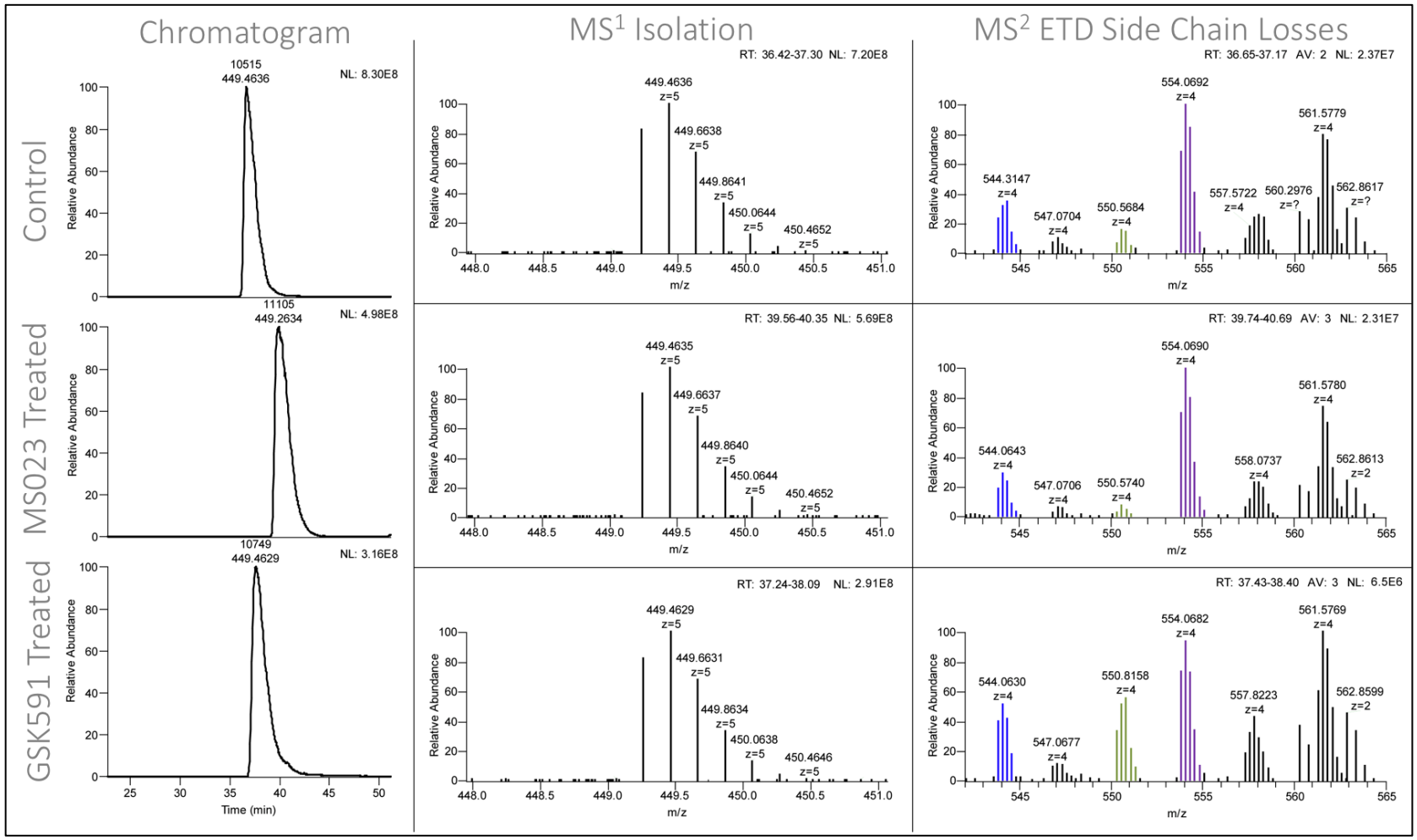

Figure 23. SmD3 Symmetric Pulldown Side Chain Fragments for 5+ of peptide AQVAAr28Gr28r28GMGr28GNIFQK; ETD side chain losses marked by purple (-31 Da), green (-45 Da), blue (-71 Da); Note the increase in relative abundance to the -45 Da of the GSK591 treated sample.

Table 6. Summary of Peptides for SmD1 in Asymmetric Pulldowns

Asymmetric IP							Abundance			Symmetry		
	Residues			Methyl		[M+H]	Control	MS023 Treated	GSK591 Treated	Control	MS023 Treated	GSK591 Treated
1	91	118	K.KREAVAGRGRGRGRGRGRGRGRGGPR.R	18	KREAVAGr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28GGPR	3210.9808	ND	ND	PH: 1.57e6	--	--	MS[1]
2	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	18	EAVAGr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28GGPR	2926.7848	PH: 4.3E6	PH: 2.6E6	2.10E+09	(20/70)	(0/45)	(100/5)
3	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	17	EAVAGr28G[rGrGrGr]7x14Gr28Gr28Gr28Gr28GGPR	2912.7692	PH: 2.3e6	ND	5.10E+08	--	--	(70/10)
4	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	16	EAVAGr28G[rGrGrGr]6x14Gr28Gr28Gr28Gr28GGPR	2898.7536	ND	ND	9.40E+07	--	--	(60/0)
5	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	15	EAVAGr28G[rGrGrGr]5x14Gr28Gr28Gr28Gr28GGPR	2884.7380	ND	ND	PH: 1.5e6	--	--	MS[1]
6	103	118	R.GRGRGRGRGRGGPR.R	12	Gr28Gr28Gr28Gr28Gr28Gr28GGPR	1833.1379	ND	ND	3.10E+08	--	--	(100/0)
7	105	118	R.GRGRGRGRGGPR.R	10	Gr28Gr28Gr28Gr28Gr28GGPR	1591.9840	ND	ND	8.20E+07	--	--	(100/0)
						Total	NR	NR	3.10E+09			
						Normalized	NA	NA	2.51E+09			

Table 7. Summary of Peptides for SmD1 in Symmetric Pulldowns

Symmetric IP							Abundance			Symmetry		
	Residues			Methyl		[M+H]	Control	MS023 Treated	GSK591 Treated	Control	MS023 Treated	GSK591 Treated
1	91	119	K.KREAVAGRGRGRGRGRGRGRGRGGPRR.cterm	18	KREAVAGr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28GGPRR	3367.0819	ND	PH: 2.1e6	ND	--	Undetermined	--
2	91	118	K.KREAVAGRGRGRGRGRGRGRGRGGPR.R	18	KREAVAGr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28GGPR	3210.9808	1.74E+08	3.39E+09	ND	(30/65)	(0/50)	--
3	93	119	R.EAVAGRGRGRGRGRGRGRGRGGPRR.cterm	19	EAVAGr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28GGPr14R	3096.9016	3.20E+08	4.19E+09	ND	(20/60)	(0/60)	--
4	92	118	K.REAVAGRGRGRGRGRGRGRGRGGPR.R	18	REAVAGr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28GGPR	3082.8859	2.11E+09	ND	ND	(0/70)	--	--
	93	119	R.EAVAGRGRGRGRGRGRGRGRGGPRR.cterm	18	EAVAGr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28GGPRR		ND	1.05E+10	ND	--	(0/80)	--
5	92	118	K.REAVAGRGRGRGRGRGRGRGRGGPR.R	15	EAVAGr28G[rGrGrGrGrGrGr]13x14GGPRR	3040.8391	ND	4.27E+08	ND	--	(10/25)	--
6	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	18	EAVAGr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28GGPR	2926.7848	1.87E+11	5.35E+10	PH: 3.8E6	(10/60)	(0/50)	(65/45)
7	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	17	EAVAGr28G[rGrGrGr]7x14Gr28Gr28Gr28Gr28GGPR	2912.7692	1.14E+10	6.27E+09	PH: 1.9e6	(15/20)	(5/35)	MS[1]
8	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	16	EAVAGr28G[rGrGrGr]6x14Gr28Gr28Gr28Gr28GGPR	2898.7536	4.70E+09	5.39E+09	ND	(15/25)	(<5/20)	--
9	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	15	EAVAGr28G[rGrGrGr]5x14Gr28Gr28Gr28Gr28GGPR	2884.7380	1.59E+09	4.80E+09	ND	(15/25)	(0/70)	--
10	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	14	EAVAG[rGrGrGrGrGrGr]10x14Gr28Gr28GGPR	2870.7224	PH: 5e6	3.20E+09	ND	(10/15)	(0/25)	--
11	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	13	EAVAG[rGrGrGrGrGrGr]9x14Gr28Gr28GGPR	2856.7068	PH: 1.6e7	1.05E+09	ND	MS[1]	(5/20)	--
12	93	106	R.EAVAGRGRGRGRGR.G	7	EAVAGr28Gr28Gr14Gr28GR	1552.9257	ND	2.40E+07	ND	--	Mixed	--
13	93	106	R.EAVAGRGRGRGRGR.G	6	EAVAGr28Gr14G[rGr]3x14GR	1538.9100	ND	7.70E+07	ND	--	(0/45)	--
14	93	106	R.EAVAGRGRGRGRGR.G	5	EAVAGr28Gr14Gr14Gr14GR	1524.8944	ND	2.40E+07	ND	--	(0/45)	--
15	93	104	R.EAVAGRGRGRGR.G	7	EAVAGr28Gr28Gr28Gr14	1339.8030	6.80E+06	ND	ND	(0/80)	--	--
16	93	104	R.EAVAGRGRGRGR.G	5	EAVAGr28Gr14Gr28GR	1311.7717	ND	2.90E+07	ND	--	(0/80)	--
17	93	104	R.EAVAGRGRGRGR.G	4	EAVAGr28Gr14Gr14GR	1297.7561	ND	PH: 7.1e5	ND	--	Undetermined	--
19	105	118	R.GRGRGRGRGRGGPR.R	10	Gr28Gr28Gr28Gr28Gr28GGPR	1591.9840	PH: 6.1e6	PH: 3.2e6	PH: 1.3e6	(35/70)	(20/75)	(60/75)
20	105	118	R.GRGRGRGRGRGGPR.R	9	G[rGrGrGrGr]9x14GGPR	1577.9684	ND	1.98E+08	ND	--	(25/75)	--
22	105	118	R.GRGRGRGRGRGGPR.R	7	G[rGr]2x14Gr28Gr28Gr28GGPR	1549.9372	ND	1.28E+08	ND	--	(0/30)	--
23	107	119	R.GRGRGRGRGGPRR.cterm	8	Gr28Gr28Gr28Gr28GGPRR	1506.9313	ND	3.00E+06	ND	--	(0/20)	--
24	107	118	R.GRGRGRGRGGPR.R	8	Gr28Gr28Gr28Gr28GGPR	1350.8301	PH: 4e6	1.95E+08	ND	MS[1]	(15/35)	--
25	107	118	R.GRGRGRGRGGPR.R	7	G[rGrGrGr]7x14GGPR	1336.8145	ND	2.29E+08	ND	--	(10/30)	--
26	107	118	R.GRGRGRGRGGPR.R	6	G[rGrGr]4x14Gr28GGPR	1322.7989	ND	1.03E+08	ND	--	(20/35)	--
						Total	2.07E+11	9.37E+10	NR			
						Normalized	2.07E+11	1.30E+11	NA			

Table 8. Summary of Peptides for SmD1 in Monomethylation Pulldowns

Monomethyl IP							Abundance			Symmetry		
	Residues			Methyl		[M+H]	Control	MS023 Treated	GSK591 Treated	Control	MS023 Treated	GSK591 Treated
1	91	118	K.KREAVAGRGRGRGRGR.G	5	KREAVAG[rGr]3x14Gr14Gr14GR	1809.0904	ND	3.57E+07	ND	--	(5/25)	--
2	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	16	EAVAGr28G[rGrGrGr]6x14Gr28Gr28Gr28Gr28GGPR	2898.7536	1.20E+08	ND	PH: 3.6e6	(15/80)	--	(65/5)
					EAVAG[rGrGr]4x14Gr28Gr28Gr28Gr28Gr28GGPR							
3	93	118	R.EAVAGRGRGRGRGRGRGRGRGGPR.R	15	EAVAGr28G[rGrGrGr]5x14Gr28Gr28Gr28Gr28GGPR	2884.7380	1.45E+08	ND	ND	(25/25)	--	--
					EAVAGr28Gr14Gr14Gr28Gr14Gr28Gr28Gr28Gr28GGPR							
4	93	106	R.EAVAGRGRGRGRGR.G	7	EAVAGr28Gr14Gr28Gr14GR	1766.0482	ND	1.90E+07	ND	--	(0/15)	--
5	93	106	R.EAVAGRGRGRGR.G	6	EAVAGr28Gr14G[rGr]3x14GR	1538.9100	PH: 5.76e5	2.87E+07	ND	MS[1]	(20/20)	--
6	93	106	R.EAVAGRGRGRGR.G	5	EAVAGr28Gr14Gr14Gr14GR	1524.8944	Review	3.30E+08	ND	(0/40)	(0/40)	--
7	93	104	R.EAVAGRGRGR.G	4	EAVAGr28Gr14Gr14GR	1297.7561	2.00E+05	7.60E+06	4.20E+07	(0/50)	(0/40)	(30/30)
					EAVAG[rGrGr]4x14GR							
8	93	104	R.EAVAGRGRGR.G	3	EAVAGr14Gr14Gr14GR	1283.7405	ND	5.30E+07	5.75E+07	--	(0/25)	(0/25)
9	93	102	R.EAVAGRGR.G	2	EAVAGr14Gr14GR	1056.6022	ND	ND	2.40E+07	--	--	(0/15)
10	101	118	R.GRGRGRGRGRGRGGPR.R	13	G[rGr]3x14Gr28Gr28Gr28Gr28GGPR	2060.2762	ND	ND	PH: 8.7e5	--	--	(75/0)
11	101	118	R.GRGRGRGRGRGRGGPR.R	10	Gr14Gr14Gr14Gr14Gr28Gr28Gr28GGPR.	2018.2294	ND	1.69E+07	ND	--	(10/60)	--
12	103	118	R.GRGRGRGRGRGRGGPR.R	11	G[rGrGrGr]7x14Gr28Gr28GGPR	1819.1223	ND	ND	8.88E+07	--	--	(60/15)
13	103	118	R.GRGRGRGRGRGRGGPR.R	10	G[rGrGrGGrGr]10x14GGPR	1805.1067	ND	ND	PH: 3.9e5	--	--	Undetermined
14	106	119	R.GRGRGRGRGRGGPRR.Cterm	8	Gr14Gr14Gr28Gr28Gr28GGPRR	1720.0539	ND	8.60E+06	ND	--	(0/60)	--
15	105	118	R.GRGRGRGRGRGGPR.R	9	G[rGrGrGrGr]9x14GGPR	1577.9684	ND	PH: 4.7e5	2.77E+07	--	Undetermined	(75/0)
					G[rGr]3x14Gr28Gr28Gr28GGPR.R							
16	105	118	R.GRGRGRGRGRGGPR.R	8	G[rGr]2x14Gr28Gr28Gr28GGPR	1563.9528	ND	4.27E+07	ND	--	(0/75)	--
17	105	118	R.GRGRGRGRGRGGPR.R	7	G[rGrGrGr]5x14Gr28GGPR	1549.9372	ND	6.54E+07	PH: 1.8e6	--	(0/60)	MS[1]
					G[rGrGrGr]5x14Gr28GGPR							
18	105	118	R.GRGRGRGRGRGGPR.R	6	G[rGr]2x14Gr14Gr14Gr28GGPR	1535.9214	ND	1.45E+07	ND	--	(20/60)	--
19	107	118	R.GRGRGRGGPR.R	7	G[rGrGr]5x14Gr28GGPR	1336.8145	ND	ND	4.51E+07	--	--	(35/0)
					G[rGrGrGr]7x14GGPR							
20	107	118	R.GRGRGRGRGGPR.R	6	G[rGrGr]4x14Gr28GGPR	1322.7989	ND	1.50E+07	PH: 2e5	--	(20/20)	--
						Totals	2.65E+08	6.37E+08	2.85E+08			
						Normalized	2.65E+08	5.22E+08	1.34E+08			

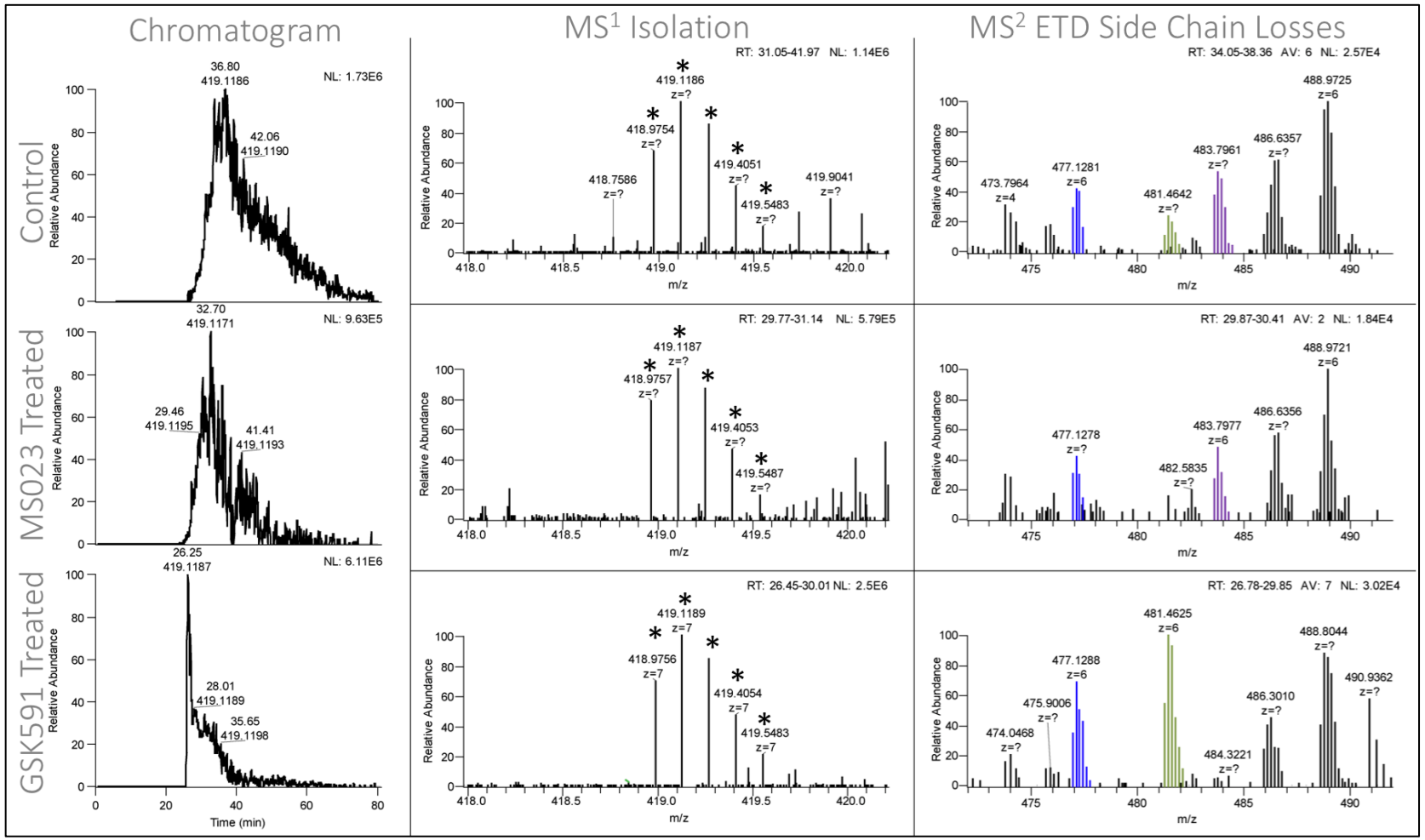

Figure 24. SmD1 Asymmetric Pulldown Side Chain Fragments for 7+ of peptide EAVAGr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28GGPR; ETD side chain losses marked by purple (-31 Da), green (-45 Da), blue (-71 Da); Note the depletion of the -31 Da fragment and increase in the relative abundance of the -45 Da fragment in the GSK591 treated sample.

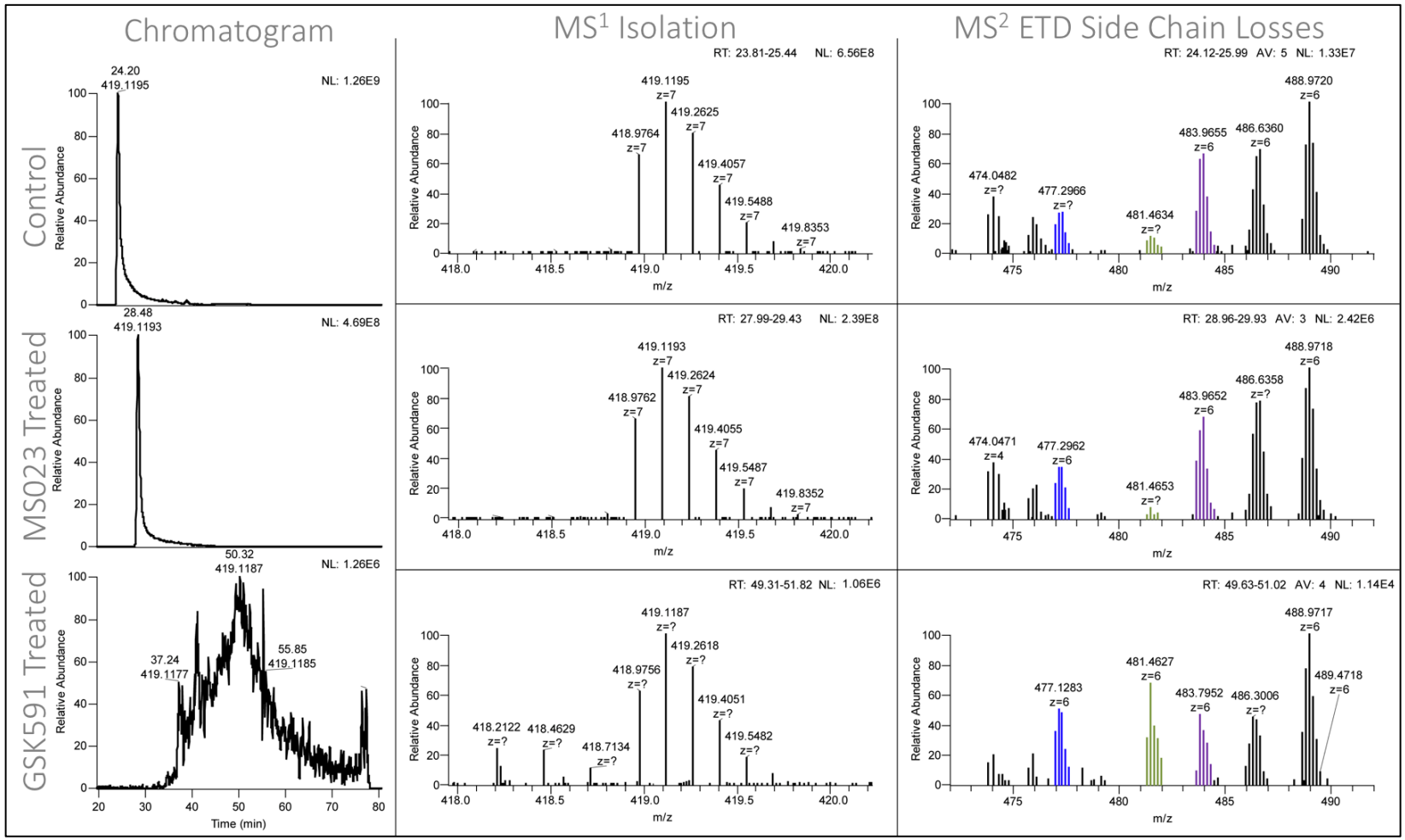

Figure 25. SmD1 Symmetric Pulldowne Side Chain Fragments for 7+ of peptide EAVAGr28Gr28Gr28Gr28Gr28Gr28Gr28Gr28GGPR; ETD side chain losses marked by purple (-31 Da), green (-45 Da), blue (-71 Da); Note the increase of relative abundance of the -45 Da fragment in the GSK591 treated sample.

2.6 Conclusion

A method was developed for the detection of methylarginine containing peptides enriched by immunoaffinity purification from control and treated cells by LC-MS/MS. The results were evaluated using two Sm proteins, SmD3 and SmD1, common components to the spliceosome. SmD3 highlights the advantages of ETD fragmentation and the Byonic search platform. The method was able to identify the full-length peptide for SmD3 that was detected from 8+ and 7+ charge states with 9 internal dimethylated arginine residues. GSK591 treatment of cells resulted in decreased detection of the methylated peptides for SmD3, including the truncated peptides. Results from qRT-PCR show retained introns for GSK591 treated cells, suggesting a disruption of the spliceosome. Interestingly, SmD1 is also reported to be dimethylated by PRMT5 but detected peptide levels in the GSK591 treated cells were consistent with the control sample suggesting an alternative dimethylation pathway.

2.7 References

1. Larsen SC, et al. (2016) Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. *Sci Signal* 9(443):rs9.
2. Morales Y, Cáceres T, May K, Hevel JM (2016) Biochemistry and regulation of the protein arginine methyltransferases (PRMTs). *Arch Biochem Biophys* 590:138–152.
3. Blanc RS, Richard S (2017) Arginine Methylation: The Coming of Age. *Mol Cell* 65(1):8–24.
4. Yu Z, Chen T, Hebert J, Li E, Richard S (2009) A Mouse PRMT1 Null Allele Defines an Essential Role for Arginine Methylation in Genome Maintenance and Cell Proliferation. *Mol Cell Biol* 29(11):2982–2996.
5. Cheng D, et al. (2004) Small Molecule Regulators of Protein Arginine Methyltransferases. *J Biol Chem* 279(23):23892–23899.
6. Stopa N, Krebs JE, Shechter D (2015) The PRMT5 arginine methyltransferase: Many roles in development, cancer and beyond. *Cell Mol Life Sci* 72(11):2041–2059.
7. Wilczek C, et al. (2011) Protein arginine methyltransferase Prmt5-Mep50 methylates histones H2A and H4 and the histone chaperone nucleoplasmin in Xenopus laevis eggs. *J Biol Chem* 286(49):42221–42231.
8. Wang M, Fuhrmann J, Thompson PR (2014) Protein arginine methyltransferase 5

catalyzes substrate dimethylation in a distributive fashion. *Biochemistry* 53(50):7884–7892.

9. Gui S, et al. (2013) Substrate-induced control of product formation by protein arginine methyltransferase 1. *Biochemistry* 52(1):199–209.
10. Evich M, Stroeva E, Zheng YG, Germann MW (2016) Effect of methylation on the side-chain pKa value of arginine. *Protein Sci* 25(2):479–486.
11. Fuhrmann J, Clancy KW, Thompson PR (2015) Chemical Biology of Protein Arginine Modifications in Epigenetic Regulation. *Chem Rev* 115(11):5413–5461.
12. Di Lorenzo A, Bedford MT (2011) Histone arginine methylation. *FEBS Lett* 585(13):2024–2031.
13. Chen H, Lorton B, Gupta V, Shechter D (2017) A TGFβ-PRMT5-MEP50 axis regulates cancer cell invasion through histone H3 and H4 arginine methylation coupled transcriptional activation and repression. *Oncogene* 36(3):373–386.
14. Yang Y, Bedford MT (2013) Protein arginine methyltransferases and cancer. *Nat Rev Cancer* 13(1):37–50.
15. Guo A, et al. (2014) Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Mol Cell Proteomics* 13(1):372–87.
16. Dhar S, et al. (2013) Loss of the major type i arginine methyltransferase PRMT1 causes substrate scavenging by other PRMTs. *Sci Rep* 3(1):1311.
17. Gehrig PM, Hunziker PE, Zahariev S, Pongor S (2004) Fragmentation pathways of NG-methylated and unmodified arginine residues in peptides studied by ESI-MS/MS and MALDI-MS. *J Am Soc Mass Spectrom* 15(2):142–149.
18. Eram MS, et al. (2016) A Potent, Selective, and Cell-Active Inhibitor of Human Type i Protein Arginine Methyltransferases. *ACS Chem Biol* 11(3):772–781.
19. Chan-Penebre E, et al. (2015) A selective inhibitor of PRMT5 with in vivo and in vitro potency in MCL models. *Nat Chem Biol* 11(6):432–437.
20. Duncan KW, et al. (2016) Structure and Property Guided Design in the Identification of PRMT5 Tool Compound EPZ015666. *ACS Med Chem Lett* 7(2):162–166.
21. Uhlmann T, et al. (2012) A Method for Large-scale Identification of Protein Arginine Methylation. *Mol Cell Proteomics* 11(11):1489–1499.
22. Wang K, et al. (2016) Antibody-Free Approach for the Global Analysis of Protein Methylation. *Anal Chem* 88(23):11319–11327.
23. Shishkova E, et al. (2017) Global mapping of CARM1 substrates defines enzyme specificity and substrate recognition. *Nat Commun* 8:15571.
24. Geoghegan V, Guo A, Trudgian D, Thomas B, Acuto O (2015) Comprehensive identification of arginine methylation in primary T cells reveals regulatory roles in cell signalling. *Nat Commun* 6(1):6758.
25. Cooper HJ, Hudgins RR, Håkansson K, Marshall AG (2002) Characterization of amino

acid side chain losses in electron capture dissociation. *J Am Soc Mass Spectrom* 13(3):241–249.

26. Xia Q, et al. (2011) Characterization and diagnostic value of amino acid side chain neutral losses following electron-transfer dissociation. *J Am Soc Mass Spectrom* 22(2):255–264.
27. Snijders APL, Hung ML, Wilson SA, Dickman MJ (2010) Analysis of Arginine and Lysine Methylation Utilizing Peptide Separations at Neutral pH and Electron Transfer Dissociation Mass Spectrometry. *J Am Soc Mass Spectrom* 21(1):88–96.
28. Brame CJ, Moran MF, McBroom-Cerajewski LDB (2004) A mass spectrometry based method for distinguishing between symmetrically and asymmetrically dimethylated arginine residues. *Rapid Commun Mass Spectrom* 18(8):877–881.
29. Hart-Smith G, Low JKK, Erce MA, Wilkins MR (2012) Enhanced methylarginine characterization by post-translational modification-specific targeted data acquisition and electron-transfer dissociation mass spectrometry. *J Am Soc Mass Spectrom* 23(8):1376–1389.
30. Battle DJ, et al. (2006) The SMN complex: An assembly machine for RNPs. *Cold Spring Harbor Symposia on Quantitative Biology* (Cold Spring Harbor Laboratory Press), pp 313–320.
31. Matlin AJ, Clark F, Smith CWJ (2005) Understanding alternative splicing: Towards a cellular code. *Nat Rev Mol Cell Biol* 6(5):386–398.
32. Brahms H, et al. (2000) The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. *J Biol Chem* 275(22):17122–17129.
33. Friesen WJ, et al. (2001) The Methylosome, a 20S Complex Containing JBP1 and pICln, Produces Dimethylarginine-Modified Sm Proteins. *Mol Cell Biol* 21(24):8289–8300.
34. Meister G, et al. (2001) Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Curr Biol* 11(24):1990–1994.
35. Friesen WJ, Massenet S, Paushkin S, Wyce A, Dreyfuss G (2001) SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Mol Cell* 7(5):1111–1117.
36. Miranda TB, et al. (2004) Spliceosome Sm proteins D1, D3, and B/B′ are asymmetrically dimethylated at arginine residues in the nucleus. *Biochem Biophys Res Commun* 323(2):382–387.

Appendix 1. Type I Inhibited (MS023 Treated) Dimethylated Pulldowns, Protein Abundance Ratios; red to blue gradient represent low to high ratios

		Summed Methylated Peptide Abundances per Protein				Protein Abundance Ratio	
		Control		Asymmetrically Inhibited		Asymmetrically Inhibited / Control	
Accession	Description	Symmetric Pulldown	Asymmetric Pulldown	Symmetric Pulldown	Asymmetric Pulldown	Symmetric Pulldown	Asymmetric Pulldown
47605556	RecName: Full=Bcl-2-associated transcription factor 1; Short=Btf	2.13E+07	2.20E+10	2.44E+08	7.86E+09	11.45	0.36
254763463	RecName: Full=Heterogeneous nuclear ribonucleoprotein U; Short=hnRNP U; AltName: Full=Scaffold attachment factor A; Short=SAF-A; AltName: Full=p120; AltName: Full=pp120	1.99E+08	1.59E+10	4.43E+08	1.61E+11	2.23	10.13
46403176	RecName: Full=Polyadenylate-binding protein 2; Short=PABP-2; Short=Poly(A)-binding protein 2; AltName: Full=Nuclear poly(A)-binding protein 1; AltName: Full=Poly(A)-binding protein II; Short=PABII; AltName: Full=Polyadenylate-binding nuclear protein 1	1.15E+07	1.28E+10	2.80E+08	1.89E+08	24.47	0.01
182676426	RecName: Full=Caprin-1; AltName: Full=Cell cycle-associated protein 1; AltName: Full=Cytoplasmic activation- and proliferation-associated protein 1; AltName: Full=GPI-anchored membrane protein 1; AltName: Full=GPI-anchored protein p137; Short=GPI-p137; Short=p137GPI; AltName: Full=Membrane component chromosome 11 surface marker 1; AltName: Full=RNA granule protein 105		8.88E+09		5.30E+10		5.96
97536050	RecName: Full=Y-box-binding protein 3; AltName: Full=Cold shock domain-containing protein A; AltName: Full=DNA-binding protein A; AltName: Full=Single-strand DNA-binding protein NF-GMB	1.78E+07	6.50E+09	6.25E+07	5.65E+07	3.52	0.01
8928305	RecName: Full=TATA-binding protein-associated factor 2N; AltName: Full=68 kDa TATA-binding protein-associated factor; Short=TAF(II)68; Short=TAFII68; AltName: Full=RNA-binding protein 56	2.98E+08	5.80E+09	1.29E+08	2.78E+10	0.43	4.80
97537467	RecName: Full=Thyroid hormone receptor-associated protein 3; AltName: Full=Thyroid hormone receptor-associated protein complex 150 kDa component; Short=Trap150	4.05E+08	5.22E+09	1.12E+08	8.82E+09	0.28	1.69
94707996	RecName: Full=Zinc finger CCCH domain-containing protein 4		4.67E+09	7.63E+07	6.39E+08		0.14
90101344	RecName: Full=Heterogeneous nuclear ribonucleoprotein U-like protein 1; AltName: Full=Adenovirus early region 1B-associated protein 5; AltName: Full=E1B-55 kDa-associated protein 5; Short=E1B-AP5	8.93E+06	3.99E+09	5.58E+08	9.50E+09	62.46	2.38
5921786	RecName: Full=Cold-inducible RNA-binding protein; AltName: Full=A18 hnRNP; AltName: Full=Glycine-rich RNA-binding protein CIRP	5.06E+06	2.68E+09	4.65E+07	8.20E+09	9.18	3.06
296452908	RecName: Full=Splicing factor 3B subunit 2; AltName: Full=Pre-mRNA-splicing factor SF3b 145 kDa subunit; Short=SF3b145; AltName: Full=SF3b150; AltName: Full=Spliceosome-associated protein 145; Short=SAP 145		2.64E+09		2.48E+09		0.94
71648673	RecName: Full=Protein LSM14 homolog A; AltName: Full=Protein FAM61A; AltName: Full=Protein SCD6 homolog; AltName: Full=Putative alpha-synuclein-binding protein; Short=AlphaSNBP; AltName: Full=RNA-associated protein 55A; Short=hRAP55; Short=hRAP55A		2.21E+09	2.76E+06	2.99E+09		1.36
1709851	RecName: Full=Splicing factor, proline- and glutamine-rich; AltName: Full=100 kDa DNA-pairing protein; Short=hPOMp100; AltName: Full=DNA-binding p52/p100 complex, 100 kDa subunit; AltName: Full=Polypyrimidine tract-binding protein-associated-splicing factor; Short=PSF; Short=PTB-associated-splicing factor	2.38E+08	1.83E+09	5.27E+08	2.59E+09	2.21	1.41
51338667	RecName: Full=Small nuclear ribonucleoprotein Sm D3; Short=Sm-D3; AltName: Full=snRNP core protein D3	1.19E+11	1.60E+09	2.25E+11	3.24E+08	1.88	0.20
74762712	RecName: Full=E3 ubiquitin-protein ligase RNF126; AltName: Full=RING finger protein 126		1.58E+09		2.84E+08		0.18
125083	RecName: Full=Keratin, type I cytoskeletal 18; AltName: Full=Cell proliferation-inducing gene 46 protein; AltName: Full=Cytokeratin-18; Short=CK-18; AltName: Full=Keratin-18; Short=K18		1.34E+09	1.03E+07	0.00E+00		0.00
51828015	RecName: Full=H/ACA ribonucleoprotein complex subunit 1; AltName: Full=Nucleolar protein family A member 1; AltName: Full=snoRNP protein GAR1		1.30E+09		3.61E+07		0.03
313104167	RecName: Full=RNA-binding protein 47; AltName: Full=RNA-binding motif protein 47	1.19E+07	1.26E+09	1.78E+07	6.30E+09	1.50	5.00
215274208	RecName: Full=Nucleoprotein TPR; AltName: Full=Megator; AltName: Full=NPC-associated intranuclear protein; AltName: Full=Translocated promoter region protein	1.39E+07	1.20E+09	6.50E+06	8.05E+07	0.47	0.07
51338665	RecName: Full=Small nuclear ribonucleoprotein Sm D1; Short=Sm-D1; AltName: Full=Sm-D autoantigen; AltName: Full=snRNP core protein D1	2.81E+11	1.19E+09	1.59E+11	1.38E+09	0.57	1.15
30173484	RecName: Full=5'-3' exoribonuclease 2; AltName: Full=DHM1-like protein; Short=DHP protein		1.05E+09		3.76E+08		0.36
54041574	RecName: Full=Chromatin target of PRMT1 protein; AltName: Full=Friend of PRMT1 protein; AltName: Full=Small arginine- and glycine-rich protein; Short=SRAG		1.01E+09	2.75E+08	2.77E+09		2.74
209572695	RecName: Full=pre-mRNA 3'end processing protein WDR33; AltName: Full=WD repeat-containing protein 33; AltName: Full=WD repeat-containing protein WDC146	2.47E+07	9.59E+08	9.78E+07	2.05E+09	3.96	2.14
57015374	RecName: Full=YLP motif-containing protein 1; AltName: Full=Nuclear protein ZAP3; AltName: Full=ZAP113		9.44E+08	3.17E+06	7.67E+09		8.13
544357	RecName: Full=RNA-binding protein FUS; AltName: Full=75 kDa DNA-pairing protein; AltName: Full=Oncogene FUS; AltName: Full=Oncogene TLS; AltName: Full=POMp75; AltName: Full=Translocated in liposarcoma protein	2.46E+08	8.94E+08	9.93E+09	1.04E+10	40.41	11.66

Appendix 2. Type II Inhibited (GSK591 Treated) Dimethylated Pulldowns, Protein Abundance Ratios; red to blue gradient represent low to high ratios

		Summed Methylated Peptide Abundances per Protein				Protein Abundance Ratio	
		Control		Symmetrically Inhibited		Symmetrically Inhibited / Control	
Accession	Description	Symmetric Pulldown	Asymmetric Pulldown	Symmetric Pulldown	Asymmetric Pulldown	Symmetric Pulldown	Asymmetric Pulldown
51338665	RecName: Full=Small nuclear ribonucleoprotein Sm D1; Short=Sm-D1; AltName: Full=Sm-D autoantigen; AltName: Full=snRNP core protein D1	2.72E+11	9.45E+08	2.70E+08	2.85E+09	0.00	3.01
51338667	RecName: Full=Small nuclear ribonucleoprotein Sm D3; Short=Sm-D3; AltName: Full=snRNP core protein D3	1.53E+11	1.65E+09	7.09E+10	2.14E+09	0.46	1.29
10720082	RecName: Full=U6 snRNA-associated Sm-like protein LSm4; AltName: Full=Glycine-rich protein; Short=GRP	4.94E+10	7.49E+08	3.81E+08	2.91E+08	0.01	0.39
1710632	RecName: Full=Heterogeneous nuclear ribonucleoprotein H; Short=hnRNP H; Contains: RecName: Full=Heterogeneous nuclear ribonucleoprotein H, N-terminally processed	1.49E+10	1.67E+08	2.47E+09	6.37E+08	0.17	3.81
3041738	RecName: Full=T-complex protein 1 subunit eta; Short=TCP-1-eta; AltName: Full=CCT-eta; AltName: Full=HIV-1 Nef-interacting protein; Contains: RecName: Full=T-complex protein 1 subunit eta, N-terminally processed	1.37E+10	9.08E+07	5.64E+08	1.09E+07	0.04	0.12
1173177	RecName: Full=40S ribosomal protein S10	1.26E+10	7.83E+07	1.99E+09	4.52E+07	0.16	0.58
17380137	RecName: Full=Double-strand break repair protein MRE11A; AltName: Full=Meiotic recombination 11 homolog 1; Short=MRE11 homolog 1; AltName: Full=Meiotic recombination 11 homolog A; Short=MRE11 homolog A	3.56E+09	2.74E+06	1.71E+07	6.46E+06	0.00	2.36
133257	RecName: Full=Heterogeneous nuclear ribonucleoproteins A2/B1; Short=hnRNP A2/B1	3.51E+09	0.00E+00	4.24E+09	8.96E+07	1.21	
313104306	RecName: Full=Far upstream element-binding protein 2; Short=FUSE-binding protein 2; AltName: Full=KH type-splicing regulatory protein; Short=KSRP; AltName: Full=p75	3.17E+09	1.42E+07	2.91E+06	6.05E+06	0.00	0.43
296439430	RecName: Full=Angiopoietin-related protein 5; AltName: Full=Angiopoietin-like protein 5; Flags: Precursor	1.25E+09		7.54E+07		0.06	
146334837	RecName: Full=Putative ribosome-binding factor A, mitochondrial; Flags: Precursor	9.65E+08	5.69E+06	5.85E+08		0.61	0.00
116242482	RecName: Full=Ras GTPase-activating protein-binding protein 2; Short=G3BP-2; AltName: Full=GAP SH3 domain-binding protein 2	8.32E+08	1.38E+07		2.68E+07	0.00	1.94
254763463	RecName: Full=Heterogeneous nuclear ribonucleoprotein U; Short=hnRNP U; AltName: Full=Scaffold attachment factor A; Short=SAF-A; AltName: Full=p120; AltName: Full=pp120	5.26E+08	1.27E+10	6.66E+08	7.72E+10	1.26	6.06
544357	RecName: Full=RNA-binding protein FUS; AltName: Full=75 kDa DNA-pairing protein; AltName: Full=Oncogene FUS; AltName: Full=Oncogene TLS; AltName: Full=POMp75; AltName: Full=Translocated in liposarcoma protein	4.68E+08	6.05E+09	6.01E+07	2.90E+10	0.13	4.79
585632	RecName: Full=Coilin; AltName: Full=p80-coilin	3.50E+08			9.46E+06	0.00	
46397025	RecName: Full=Polyadenylate-binding protein-interacting protein 1; Short=PABP-interacting protein 1; Short=PAIP-1; Short=Poly(A)-binding protein-interacting protein 1	3.38E+08		1.44E+07	1.79E+07	0.04	0.00
73622105	RecName: Full=DBIRD complex subunit ZNF326; AltName: Full=Zinc finger protein 326; AltName: Full=Zinc finger protein interacting with mRNPs and DBC1	1.71E+08				0.00	
74710467	RecName: Full=PERQ amino acid-rich with GYF domain-containing protein 2; AltName: Full=GRB10-interacting GYF protein 2; AltName: Full=Trinucleotide repeat-containing gene 15 protein	1.57E+08				0.00	
14916572	RecName: Full=Ras GTPase-activating protein-binding protein 1; Short=G3BP-1; AltName: Full=ATP-dependent DNA helicase VIII; Short=hDH VIII; AltName: Full=GAP SH3 domain-binding protein 1	1.43E+08	1.88E+08	7.23E+06	1.47E+08	0.05	0.78
52783206	RecName: Full=Plasminogen activator inhibitor 1 RNA-binding protein; AltName: Full=PAI1 RNA-binding protein 1; Short=PAI-RBP1; AltName: Full=SERPINE1 mRNA-binding protein 1	1.23E+08	5.56E+08	4.06E+05	2.34E+09	0.00	4.21
317373583	RecName: Full=Keratin, type II cytoskeletal 7; AltName: Full=Cytokeratin-7; Short=CK-7; AltName: Full=Keratin-7; Short=K7; AltName: Full=Sarcolectin; AltName: Full=Type-II keratin Kb7	1.21E+08	5.80E+07	1.86E+07	1.49E+07	0.15	0.26
90110781	RecName: Full=Nucleolin; AltName: Full=Protein C23	1.07E+08	1.29E+10	8.68E+07	3.91E+10	0.81	3.03
209572695	RecName: Full=pre-mRNA 3' end processing protein WDR33; AltName: Full=WD repeat-containing protein 33; AltName: Full=WD repeat-containing protein WDC146	1.07E+08	5.85E+09		6.75E+09	0.00	1.15
134037	RecName: Full=Small nuclear ribonucleoprotein-associated proteins B and B'; Short=snRNP-B; AltName: Full=Sm protein B/B'; Short=Sm-B/B'; Short=SmB/B'	9.60E+07		1.27E+09		13.27	
74735318	RecName: Full=Transcription elongation factor SPT5; Short=hSPT5; AltName: Full=DRB sensitivity-inducing factor 160 kDa subunit; Short=DSIF p160; AltName: Full=DRB sensitivity-inducing factor large subunit; Short=DSIF large subunit; AltName: Full=Tat-cotransactivator 1 protein; Short=Tat-CT1 protein	8.31E+07				0.00	

Appendix 3. Monomethylated Pulldown, Protein Abundance Ratios; red to blue gradient represent low to high ratios

			Summed Methylated Peptide Abundances per Protein			Protein Abundance Ratio	
			Monomethyl Arginine Pulldown			Monomethyl Arginine Pulldown	
	Accession	Description	Control	Asymmetric Inhibited	Symmetrically Inhibited	Asymmetric Inhibited / Control	Symmetric Inhibited / Control
1	133257	RecName: Full=Heterogeneous nuclear ribonucleoproteins A2/B1; Short=hnRNP A2/B1	1.46E+11	2.13E+09	6.54E+10	0.01	30.74
2	288558857	RecName: Full=Heterogeneous nuclear ribonucleoprotein A1; Short=hnRNP A1; AltName: Full=Helix-destabilizing protein; AltName: Full=Single-strand RNA-binding protein; AltName: Full=hnRNP core protein A1; Contains: RecName: Full=Heterogeneous nuclear ribonucleoprotein A1, N-terminally processed	7.67E+09	1.04E+11	1.49E+10	13.51	1.94
3	8928305	RecName: Full=TATA-binding protein-associated factor 2N; AltName: Full=68 kDa TATA-binding protein-associated factor; Short=TAF(II)68; Short=TAFII68; AltName: Full=RNA-binding protein 56	4.10E+10	6.65E+10	1.08E+10	1.62	0.26
4	20981688	RecName: Full=Cysteine-rich protein 1; Short=CRP-1; AltName: Full=Cysteine-rich heart protein; Short=CRHP; Short=hCRHP; AltName: Full=Cysteine-rich intestinal protein; Short=CRIP	6.93E+10	1.29E+10	5.81E+08	0.19	0.01
5	254763463	RecName: Full=Heterogeneous nuclear ribonucleoprotein U; Short=hnRNP U; AltName: Full=Scaffold attachment factor A; Short=SAF-A; AltName: Full=p120; AltName: Full=pp120	1.94E+09	5.37E+10	3.50E+09	27.63	1.80
6	544357	RecName: Full=RNA-binding protein FUS; AltName: Full=75 kDa DNA-pairing protein; AltName: Full=Oncogene FUS; AltName: Full=Oncogene TLS; AltName: Full=POMp75; AltName: Full=Translocated in liposarcoma protein	6.22E+09	3.74E+10	1.01E+10	6.01	1.63
7	51338779	RecName: Full=Heterogeneous nuclear ribonucleoprotein A3; Short=hnRNP A3	1.02E+09	3.69E+10	5.59E+08	36.33	0.55
8	48429103	RecName: Full=Heterogeneous nuclear ribonucleoprotein K; Short=hnRNP K; AltName: Full=Transformation up-regulated nuclear protein; Short=TUNP	1.13E+10	1.72E+10	6.87E+09	1.53	0.61
9	90101344	RecName: Full=Heterogeneous nuclear ribonucleoprotein U-like protein 1; AltName: Full=Adenovirus early region 1B-associated protein 5; AltName: Full=E1B-55 kDa-associated protein 5; Short=E1B-AP5	4.83E+09	2.51E+10	3.16E+09	5.20	0.66
10	54040031	RecName: Full=Nuclease-sensitive element-binding protein 1; AltName: Full=CCAAT-binding transcription factor I subunit A; Short=CBF-A; AltName: Full=DNA-binding protein B; Short=DBPB; AltName: Full=Enhancer factor I subunit A; Short=EFI-A; AltName: Full=Y-box transcription factor; AltName: Full=Y-box-binding protein 1; Short=YB-1	8.45E+09	1.89E+10	5.08E+09	2.24	0.60
11	125083	RecName: Full=Keratin, type I cytoskeletal 18; AltName: Full=Cell proliferation-inducing gene 46 protein; AltName: Full=Cytokeratin-18; Short=CK-18; AltName: Full=Keratin-18; Short=K18	1.37E+10	2.92E+08	1.71E+10	0.02	1.25
12	90110781	RecName: Full=Nucleolin; AltName: Full=Protein C23	3.99E+08	3.00E+10	6.20E+08	75.22	1.56
13	544261	RecName: Full=RNA-binding protein EWS; AltName: Full=EWS oncogene; AltName: Full=Ewing sarcoma breakpoint region 1 protein	4.12E+09	2.01E+10	2.55E+09	4.88	0.62
14	730773	RecName: Full=Serine/arginine-rich splicing factor 1; AltName: Full=Alternative-splicing factor 1; Short=ASF-1; AltName: Full=Splicing factor, arginine/serine-rich 1; AltName: Full=pre-mRNA-splicing factor SF2, P33 subunit	1.14E+10	3.34E+08	8.73E+09	0.03	0.77
15	23503093	RecName: Full=RNA-binding motif protein, X chromosome; AltName: Full=Glycoprotein p43; AltName: Full=Heterogeneous nuclear ribonucleoprotein G; Short=hnRNP G; Contains: RecName: Full=RNA-binding motif protein, X chromosome, N-terminally processed	6.41E+09	8.06E+09	4.82E+09	1.26	0.75
16	1709851	RecName: Full=Splicing factor, proline- and glutamine-rich; AltName: Full=100 kDa DNA-pairing protein; Short=hPOMp100; AltName: Full=DNA-binding p52/p100 complex, 100 kDa subunit; AltName: Full=Polypyrimidine tract-binding protein-associated-splicing factor; Short=PSF; Short=PTB-associated-splicing factor	1.26E+10	5.40E+09	8.92E+08	0.43	0.07
17	52783206	RecName: Full=Plasminogen activator inhibitor 1 RNA-binding protein; AltName: Full=PAI1 RNA-binding protein 1; Short=PAI-RBP1; AltName: Full=SERPINE1 mRNA-binding protein 1	4.59E+09	7.19E+09	3.19E+09	1.56	0.69
18	317373583	RecName: Full=Keratin, type II cytoskeletal 7; AltName: Full=Cytokeratin-7; Short=CK-7; AltName: Full=Keratin-7; Short=K7; AltName: Full=Sarcolectin; AltName: Full=Type-II keratin Kb7	8.91E+09	1.81E+09	1.56E+09	0.20	0.18
19	1173177	RecName: Full=40S ribosomal protein S10	3.57E+09	3.08E+09	4.62E+09	0.86	1.30
20	13124489	RecName: Full=Heterogeneous nuclear ribonucleoprotein D0; Short=hnRNP D0; AltName: Full=AU-rich element RNA-binding protein 1	2.58E+09	6.41E+09	2.02E+09	2.48	0.78
21	116241330	RecName: Full=ATP-dependent RNA helicase A; Short=RHA; AltName: Full=DEAH box protein 9; AltName: Full=Leukophysin; Short=LKP; AltName: Full=Nuclear DNA helicase II; Short=NDH II	3.81E+08	9.18E+09	5.55E+08	24.06	1.46
22	48429165	RecName: Full=THO complex subunit 4; Short=Tho4; AltName: Full=Ally of AML-1 and LEF-1; AltName: Full=Aly/REF export factor; AltName: Full=Transcriptional coactivator Aly/REF; AltName: Full=bZIP-enhancing factor BEF	2.00E+09	2.75E+09	5.23E+09	1.37	2.61
23	3929377	RecName: Full=Serine/arginine-rich splicing factor 9; AltName: Full=Pre-mRNA-splicing factor SRp30C; AltName: Full=Splicing factor, arginine/serine-rich 9	5.00E+09	1.94E+09	2.69E+09	0.39	0.54
24	62512150	RecName: Full=Interleukin enhancer-binding factor 3; AltName: Full=Double-stranded RNA-binding protein 76; Short=DRBP76; AltName: Full=M-phase phosphoprotein 4; Short=MPP4; AltName: Full=Nuclear factor associated with dsRNA; Short=NFAR; AltName: Full=Nuclear factor of activated T-cells 90 kDa; Short=NF-AT-90; AltName: Full=Translational control protein 80; Short=TCP80	5.01E+09	3.19E+09	4.78E+08	0.64	0.10
25	14916572	RecName: Full=Ras GTPase-activating protein-binding protein 1; Short=G3BP-1; AltName: Full=ATP-dependent DNA helicase VIII; Short=hDH VIII; AltName: Full=GAP SH3 domain-binding protein 1	5.84E+08	5.67E+09	1.78E+09	9.70	3.04

3. Nucleosome Assembly Protein 1 – Post-Translational Modification Characterization of a Glutamylated Histone Chaperone

3.1 Abstract

Nucleosome Assembly Protein 1-like, or Nap1L, is a histone chaperone protein known to bind H2A/H2B dimers and is capable undergoing glutamylation as a post-translational modification. Glutamylation is a rare modification that increases the negative charge of the protein, shown to be essential for histone binding in other histone chaperone proteins. In this chapter, I describe the characterization of post-translational modifications, specifically glutamylation and farnesylation, for Nap1L isolated from *Xenopus laevis* eggs. By utilizing traditional CAD methods and extensive manual validation of the data, the site localization of glutamylation of Nap1l1 is reported in greater detail than previous reports. Additionally, attention is paid to sequence variants and co-exisiting modifications. The glutamylation of Nap1L is determined to be an abundant modification that occurs at numerous sites within the acidic regions.

3.2 Introduction

Every cell in the human body contains the genetic information for all 20,000 protein coding genes; thus, each cell has the information necessary to transform into any other cell in the human body. However, what defines various differentiated cells and their properties is the limited genetic expression for a particular cell type, with transcription of necessary genes to RNA and ultimate translation to proteins that then allow the cell to differentiate into various roles. One of the driving goals for the last two decades has been to identify the ways in which each individual cell expresses the appropriate genes for its role. When this function is disrupted, cells become diseased. Genes that should be silenced are expressed and change the character of the cell. Understanding the process of genetic expression in normal cells will help to drive development of treatments to correct cellular misfunctions.

There are multiple layers of organization of the 6 billion base pairs of DNA to control which genes should be expressed or silenced. In its most basic unit, 147 base pairs of DNA wrap 1.7 times around a core of eight histone proteins, thus composing a single nucleosome

(1). The nucleosomes are then packed together, much like an old-fashioned telephone coil, allowing 6 billion base pairs of DNA to be neatly packed into the nucleus for each human cell. The histone core of the nucleosome contains two copies of the H2A/H2B dimer and two copies of both H3 and H4. Linker histone H1 caps off the nucleosome where DNA extends outwards toward the next nucleosome.

One of the mechanisms by which histone proteins organize such a large amount of genetic material is the use of specific, primary sequence variants (2, 3). Each of the canonical histones has numerous primary sequence variants, which have various roles. For example, during cell division, chromatin is tightly compacted into chromosomes, and microtubules will attach at the centromere to physically pull the chromosomes into the newly formed cells. The centromere is defined by the substitution of core histone H3 with CENP-A, an H3 variant, to which the kinetochore complex is then recruited (4). The removal of H3 and replacement with CENP-A as part of the nucleosome organization acts to assist the physical organization of jumbled strands of DNA to organize into individual chromosomes during cellular division. Another example of histone variant organization is H2A.Z, which is frequently found in the promoter region of the genetic code (5, 6). The canonical H2A/H2B complex is replaced at the promoter region with the H2A.Z/H2B complex, and is thought to be in a poised state, capable of recruiting transcriptional activators.

Another mechanism by which histone proteins organize genetic material is a complex series of post translational modifications (PTMs) commonly known as the histone code (3). Secondly, histones have a histone code that can help organize and signal information about the genetic code in its proximity. Core histone proteins are highly basic protein, approximately 10-15 kDa in size and are stacked together with a series of three alpha-helices in the C-terminal, globular domain (1). The amino tail, approximately 20-30 residues in length depending on the histone, is a highly basic and disordered region which can be modified with various combinations of PTMs, such as phosphorylation, acetylation, and methylation (3). The combination of PTMs on the histone tails, known as the histone code, can then be “read” by transcriptional regulators. For example, acetylation of lysine side chains occurs in numerous places on histone tails and is commonly associated with the transcriptionally active and less compact, euchromatin structure. The acetyl group neutralizes the basic site of the lysine side

chain, diminishing the electrostatic DNA/protein interactions, assisting in keeping a less compact chromatin structure for DNA transcription(3). Acetylated histone tails are also important in the process of silencing of the genetic code. For example, in yeast, after Origin Recognition Complex (ORC) is recruited to DNA loci known as HMR-E along with Rap1b and Abf1p, the Silent Information Regulator (Sir) proteins 1-4 are recruited to the complex (7). To extend the silencing region, Sir2p, a known NAD+ deacetylase, is then recruited along the chromosome to de-acetylate the modified histone tails and recruit the Sir2-3-4p complex.

While histones communicate a variety of information, they rely on proteins known as histone chaperone proteins for manipulation throughout the cell. Histone chaperone proteins are a family of proteins first identified by Laskey et al. in 1978, who identified a role of nucleosome assembly for an acidic protein, Nucleoplasmin (8). Histone chaperone proteins are defined by their roles and interactions with histone proteins, which include but are not limited to: nucleosome assembly and disassembly, histone substitution, preventing histone aggregation, histone storage, and cytoplasmic to nuclear transport (9).

While the DNA is inherently acidic, histone proteins are highly basic with an average isoelectric point of approximately 10 to 11. Without assistance, the highly charged nature of histone proteins leads to non-specific interactions with DNA or other molecules or to protein aggregation. To compete with DNA binding, histone chaperone proteins are frequently acidic in nature, often containing specific acidic regions. A recent review defined an acidic region as a region of 8 or more amino acids composed of at least 75% acidic residues(9). Histone chaperone proteins help to neutralize the charge of the histone proteins, preventing non-specific interactions (9, 10).

The focus of this chapter is on the histone chaperone protein, Nucleosome Assembly Protein 1 or Nap1, discovered in 1983 in HeLa cells (11). It was named for its identified function of assisting with nucleosome assembly (12). Nap1 is a conserved protein in eukaryotes, from yeast to humans; Figure 26 reports the sequence alignment across frogs, mice, humans, Arabidopsis, and yeast. Between frogs, humans, and mice, Nap1 has an 87.5% sequence similarity. *Xenopus laevis* contains two copies of Nap1 homologs, Nap1l1.L and Nap1l1.S (Nap1-like-1 long and short, previously identified as B and A, respectively). The Nap1l1 proteins, both similar in sequence, are approximately 45 kDa in size and contain three

separate acidic regions, highlighted in yellow in Figure 27. Additionally, Nap1 contains a nuclear localization signal, residues 174-180, highlighted in blue in Figure 27. Yeast Nap1 (yNap1) has also been reported to have a nuclear export signal (13). The structure foryNap1 was reported in 2005 by Park and Luger, using residues 70-370 (yNap1 contains 417 residues) (see Figure 28.) The N- and C- terminal regions were reported to be intrinsically disordered regions, lacking any specific tertiary structure (13). yNap1 was reported with series of alpha helices and beta folds and was shown to dimerize in an anti-parallel manner along the extended alpha-helix that ran the length of the folded protein (13).

Nap1L1 has been identified with numerous histone chaperone functions in a large variety of cell types for eukaryotes (14). Nap1L1 has been shown to bind both H2A/H2B dimers and histone linker H1 (15, 16). The H2A/H2B dimers have binding strong enough that they are used to immunoaffinity purify Nap1L in cell extracts. While found across cell development stages, Nap1L in human and *Drosophilia* have both been localized to the cytoplasm during cell development and identified in the nucleus during S-phase, which is associated with chromatin assembly (17). In cytoplasmic extracts, dimers of yNap1L have been shown to bind a single H2A/H2B dimer and are thought to potentially oligomerize in large clusters (18). The heterotetramer associates with Kap114p, a karyopherin, to form a nuclear shuttling complex and is transported across the nuclear membrane into the nucleus (16, 18).

Not only has Nap1L been associated with histone storage and nuclear membrane transport, but it has been shown to have roles in chromatin assembly as well as preventing non-nucleosomal DNA binding. The structure for the yNap1 dimer with the H2A/H2B dimer at 6.7 Å showed that yNap1 competes with DNA by creating an acidic concave binding pocket, electrostatically binding the same portion of H2A as that the acidic DNA backbone binds (18). By binding the same basic portion of the histones, yNap1 prevents non-specific binding by acidic cellular components such as RNA and non-nucleosome DNA (19, 20). Deletion of Nap1 from yeast was shown to result in smaller nucleosomes, consistent with hexomer sized particles, pointing to further evidence for a role in histone deposition (18). Nap1L is also implicated in histone dimer exchange and in octomer sliding (13).

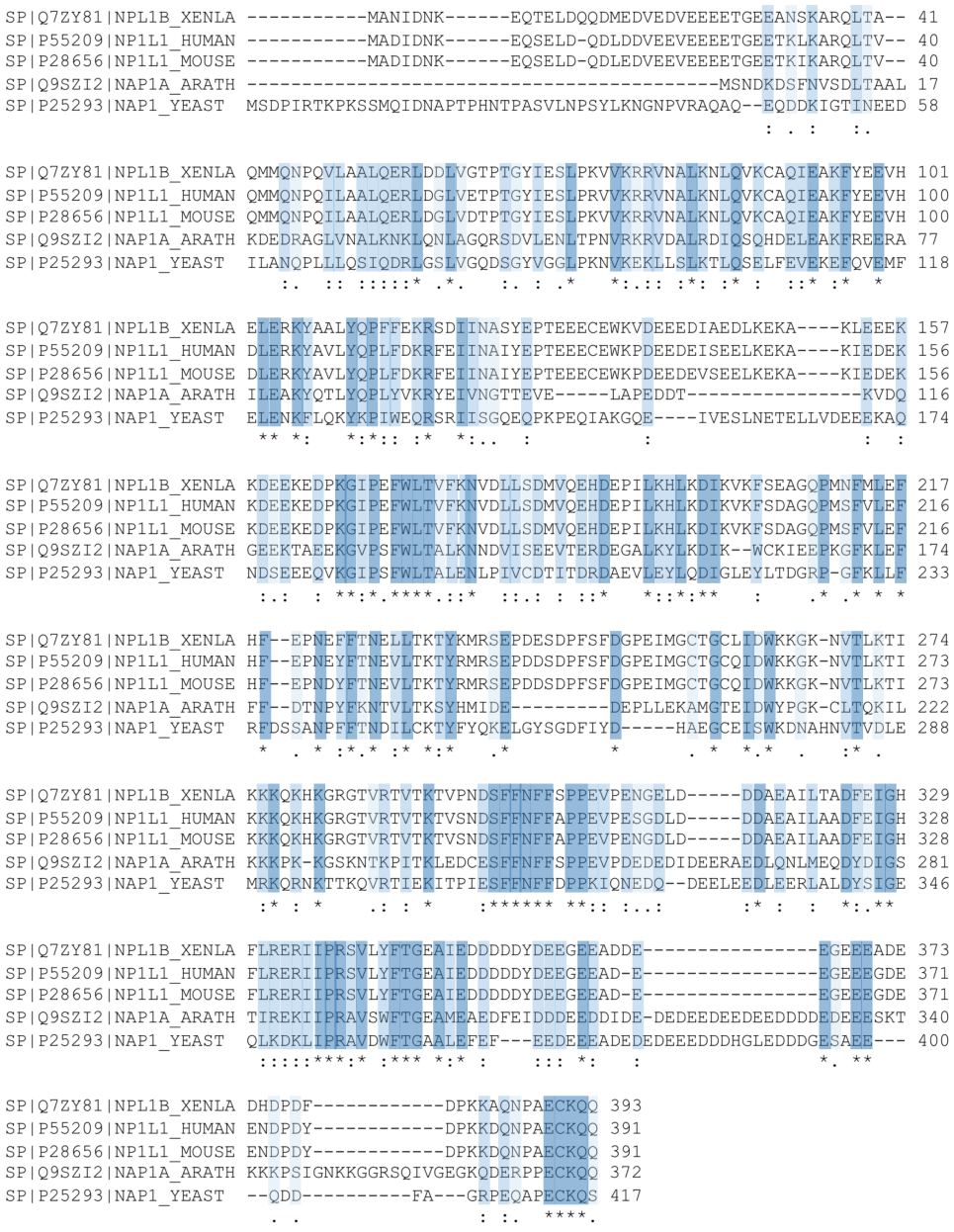

Figure 26. Sequence Alignment of Nap1 across cell types, 17% Identity across all; Xenla: *Xenopus laevis*; Arath: *Arabidopsis thaliana*; blue shading denotes sequence similarity.

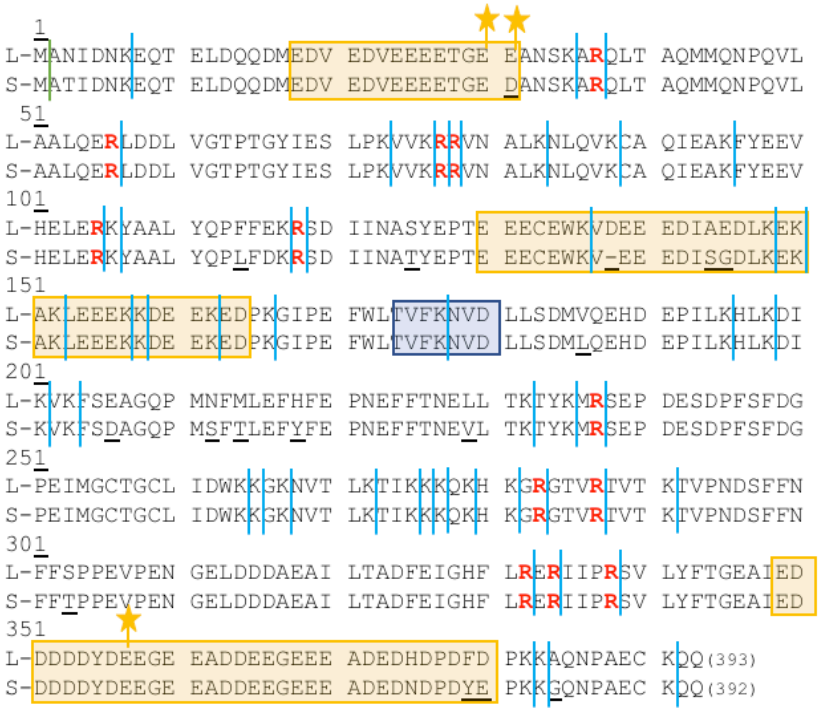

Figure 27. Sequence of Nap1L1 with highlights for the acidic regions in yellow, nuclear localization signal in blue, previously identified glutamylation sites (stars), tryptic cleavage sites (blue lines), and Arg-C cleavage sites (red R).

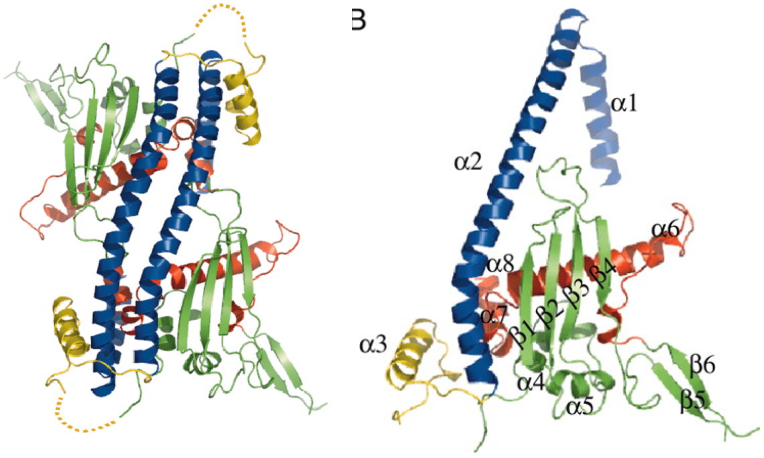

Figure 28. Crystal Structure for yNap1 (14); A) Dimerized Nap1 B) Single copy

Nap1L has been shown to not only bind H2A/H2B but is capable of binding H3/H4 tetramers and was identified for specific roles with H1 (21). In *Xenopus* eggs, it was shown that Nap1L1 controlled embryonic histone linker H1M deposition onto chromatin to provide condensed chromosome structure for mitosis (10, 22). Nap1 was also shown to assist in the removal of H1 from chromatin fibers, thereby creating a more open structure (15).

Numerous roles for Nap1L have been identified, yet in most of the studies, the role of post translational modifications for Nap1 have not been explored.

Two relatively unique modifications were previously identified and reported for Nap1: farnesylation and glutamylation (23–25). Farnesylation, a covalent lipid modification, has been reported for both hNap1L1 and A*rabidopsis thaliana* Nap1, AtNAP1;1 (24, 25). Nap1 was the first histone chaperone protein to be identified with glutamylation, the post translational addition of glutamic acid residues to the side chain of one or more primary sequence glutamic acid residues (23). Both farnesylation and glutamylation are discussed in more detail below.

Protein prenylation, the covalent attachment of either a farnesyl or a geranylgeranyl isoprenoid to a protein, occurs across eukaryotic cell types and is thought to have functions in in membrane associations, cell signaling, and protein-protein interactions (24, 26). The C-terminal motif of Nap1 is -CKQQ, known as a CaaX box, and is conserved in the Nap1 family across eukaryotes. The CaaX box recognition motif for prenylation is located at the extreme C-terminus of the protein and contains a cysteine, two typically aliphatic amino acids, and the final residue (X) is usually Cys, Met Ser, Ala, Gln (25). Because the second residue is a Lysine, Nap1 is a substrate for farnesylation, the covalent addition of a 15-carbon farnesyl diphosphate (FPP) to the c-terminal Cysteine residue by a protein farnesyl transferase (PFT). Farnesylation has been well studied in the Ras family of proteins as well as Lamin A (27). Prelamin A, a member of the nuclear lamin proteins, has a well curated pathway for processing the permanent modification of farnesylation. Following the covalent attachment of the farnesyl group to the cysteine, prelamin A is shuttled to the endoplasmic reticulum where the c-terminal tripeptide is clipped by Ste14. The C-terminal carboxyl group is then methylated by a prenyl protein specific methyltransferase, ICMT. As the protein then matures, the final 15 residues are clipped by a protease, ZMPSTE24, to remove the farnesyl group and signal a mature Lamin A. The farnesylation pathway of Lamin A has been identified as a factor in Hutchinson-Gillford progeria syndrome, characterized by accelerated aging. Additionally, farnesyltransferases have been explored as targets for cancer therapies, as farnesylation can transform proteins such as Ras (26). Galichet and Gruissem explored the role of farnesylation on AtNap1;1 in a 2006 study (24). They found that the farnesylated

AtNap1;1 was correlated with localization to the nucleus and associated with cell proliferation and division. In later stages of leaf development, AtNap1;1 was no longer found to be farnesylated and was accumulated in the cytoplasm (24). While hNap1L1 has been shown to be a substrate of PFT *in vitro*, the role of the modification for mammalian Nap1L1 has not been explored (24, 25).

Glutamylation is the post translational addition of a glutamate residue to the gamma carbon of a primary sequence glutamic acid in an ATP dependent manner (28). It was first identified on the acidic regions of tubulin proteins (29, 30). Similar to tubulin proteins, the modification was identified in the acidic region of Nap1 (22, 23). Acidic regions are defined as those with 75% acidic residues for at least 8 amino acid residues (9). The polyglutamylation serves to increase the negative charge to a protein by the number of additional glutamic acid residues plus one. In Nucleoplasmin, it was shown to enhance the segregation of the H2A/H2B histone dimer during xenopus egg development (31). Glutamylation is associated with microtubules, specifically alpha and beta tubulin proteins, where its chain length and frequency of polyglutamylation can distinguish different microtubule populations (28).

Several of the Tubulin Tyrosine Ligase Like (TTLL) proteins are identified as polyglutamylases. The modification is reversible, with several cytosolic carboxypeptidases identified for deglutamylation of tubulin chains, though no enzyme has been linked to histone chaperone proteins (32). Polyglutamylases are divided into two general categories, either chain initiating enzymes or elongating enzymes, though some enzymes may exhibit both activities (28). Initiating enzymes (TTLL 4, 5, and 7) catalyze the covalent attachment of the first glutamate residue to the gamma-carbon of primary chain glutamate (see Figure 29) creating an isopeptide bond. Elongating enzymes (TTLL 6, 11, and 13) catalyze the elongating the chain by covalent addition to either the gamma-carbon or to the alpha-carbon of the initial modification, similar to a peptide backbone, shown C and D in Figure 29. It is worth pointing out that it is assumed within the field that the extended chains are added as alpha additions (28). In tubulin proteins, it is recognized that glutamylation can occur at a single site or multiple sites with an extended poly-glu chain (upwards of 12 or 20 additions); or as identified with alpha-tubulin, the glutamylation occurs on multiple sites with shorter poly-glu chains (33, 34). Identification of the orientation of tubulin poly-glu chains has been limited to

comparison of retention times of synthetic, known peptides during chromatographic separation (30, 33, 34). Similar to tubulin tyrosine ligases, polyglutamylases use two Mg2+ to stabilize ATP in the binding pocket for the reaction between the carboxyl group and the amino group of the glutamic acid residue. It is a two-step reaction with an acylphosphate intermediate (35, 36). The TTLL enzymes have not been characterized beyond their elongating and initiating abilities. To a large extent, monoclonal antibodies have been used to characterize glutamylation chains (37, 38).

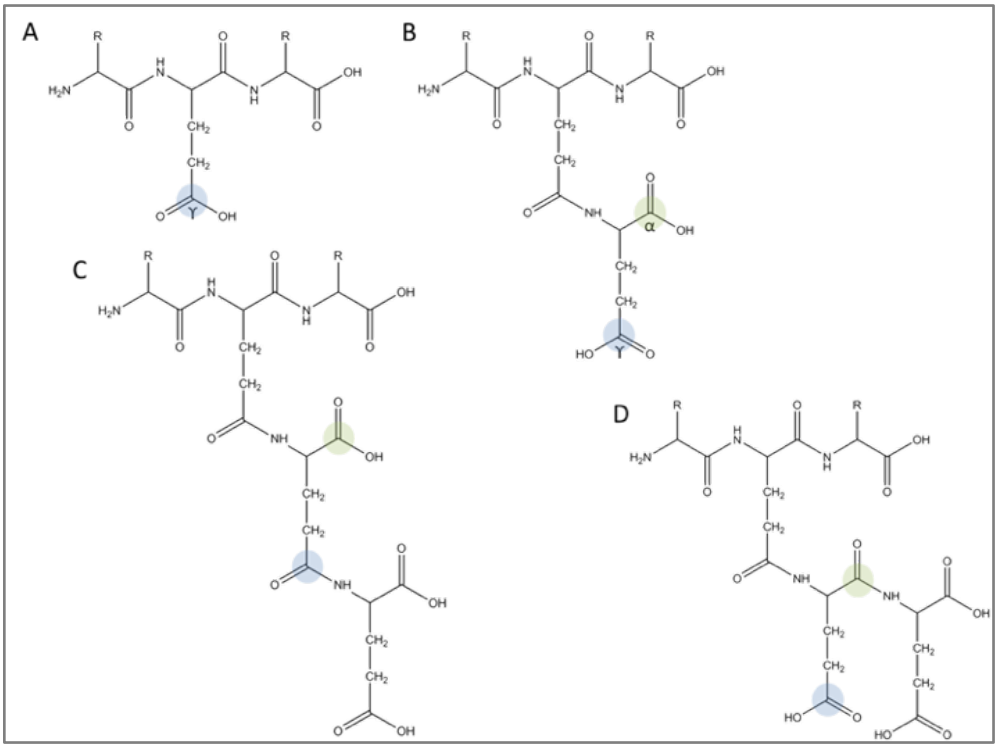

Figure 29. Glutamylation as a post translational modification. **A)** Unmodified, primary chain, glutamic acid residue **B)** Monoglutamylated peptide **C)** γ addition, diglutamylated peptide **D)** α addition, diglutamylated peptide. Blue circles represent γ-carbons and green circles represent α-carbons.

Nap1 was first identified as being polyglutamylated in 2000 by Regnard et al. in HeLa cells (23). The study utilized an antibody designed against polyglutamylated tubulin proteins to immuno-purify modified proteins followed by trypsin digestion and Edman sequencing. It showed a C-terminal peptide containing 3-9 glutamyl units on two potential sites, Glu-356 and Glu-357 for Nap1, as the Edman sequencing was inconclusive. The study also noted that tubulin proteins are frequently modified at a primary chain glutamic acid residue with a

polyglutamyl chain, and this was suggested as the modification for Nap1. In addition to the C-terminal peptide, an N-terminal peptide was identified but not characterized; noting that the peptide from residue 8-32 was modified with up to 5 glutamyl units. This work was confirmed by Miller and Heald in 2015, who published a study using liquid chromatography-tandem mass spectrometry to identify 12 potential residues that might be glutamylated, with E30, E31, and E357 being high confidence glutamylation sites (22). They prepared mutants, substituting multiple glutamic acid residues with aspartic residues to minimize glutamylation for studies. No further discussion was provided on the mass spectrometry sequencing. Of further note, farnesylation was not reported in the papers on glutamylation by either Miller or Regnard. The Miller and Heald study from 2015 of *Xenopus* Nap1l1 showed that glutamylation of Nap1l1 played a role in the ability of H1M (an embryotic isoform of histone linker H1) to be efficiently deposited on chromatin. Consistent with their results, previous work from David Shechter's lab shows that increasing levels of glutamylation on the histone chaperone, nucleoplasm, correlates to a change from histone sequestration to deposition (31). The specific function of glutamylation is still undefined.

In this chapter, I describe the post-translational modifications, specifically glutamylation and farnesylation, for Nap1l1 isolated from *Xenopus laevis* eggs. By utilizing traditional CAD methods and extensive manual validation of the data, the site localization of glutamylation of Nap1l1 is reported in greater detail than previous reports. Additionally, attention is paid to sequence variants and co-exisiting modifications. The goal of the chapter is to characterize the modifications utilizing high resolution mass spectrometry to build a more complete picture of the Nap1l1 protein.

3.3 Materials

Agilent Technologies (Palo Alto, CA)

1100 Series High Performance Liquid Chromatograph

1100 Series Vacuum Degasser

Poroshell 300SB-C18 resin, (5 μm, 300 Å)

Poroshell 300SB-C8 resin, (5 μm, 300 Å)

Dr. Maisch GmbH

Reprosil-Pur 120 Å C18-AQ, 3 μm

Reprosil-Pur 120 Å C18-AQ, 1.9 μm

Eppendorf (Hauppauge, NY)

5414R Benchtop centrifuge

Honeywell (Morristown, NJ)

Burdick and Jackson® Acetonitrile, LC-MS grade

Millipore (Billerica, MA)

Amicon Ultra-0.5 10K centrifugal filter

Molex (Lisle, IL)

Polymicro Technologies™ polyimide coated fused silica capillary, 360 μm x 75 μm i.d.

New Objective (Woburn, MA)

PicoClear™ Union Assembly

PQ Corporation (Valley Forge, PA)

Kasil – Potassium silicate solution

Promega Corporation, (Madison, WI)

Sequencing grade modified Endoproteinase Trypsin

Roche Diagnostics GmbH (Mannheim, Germany)

Endoproteinase Arg-C, Sequencing Grade

Sigma Aldrich (St. Louis, MO)

2-propanol, LC-MS grade

Ammonium Bicarbonate

Angiotensin I acetate salt hydrate, ≥99% purity (human)

Apomyoglobin from equine skeletal muscle, protein sequencing standard

Carbonic Anhydrase

DL-Dithiothreitol, ≥98% by TLC

Fluoranthene, >99% purity

Glacial Acetic Acid, ≥99.9% purity

Iodoacetamide (Bioultra), ≥99% purity

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

Thermo Fisher Scientific (San Jose, CA; Bremen, Germany)

Formic Acid, LC-MS Grade

Orbitrap Fusion™ Tribrid™ mass spectrometer

Pierce Water, LC-MS Grade

Urea

3.4 Methods

3.4.1 Nap1l1 Purification

Nap1 purification was designed and performed by David Shechter and Benjamin Lorton at Albert Einstein School of Medicine in Bronx, NY.

Eggs are harvested from *Xenopus Laevis* and incubated for several days before being lysed. The histone pair, H2A.S2/H2B, is immobilized on sepharaose resin by the streptactin antibodies. The resin is washed with egg lysis buffer (2.5 mM MgCl2, 50 mM KCl, 10 mM Hepes-KOH, pH 7.8). The egg extract was incubated with the immobilized histones for 1 hour. The resin was then packed into gravity chromatography columns and proteins were eluted with an increasing salt gradient from 0-2 M NaCl. Fractions containing Nap1l1 were then immunodepleted using an anti-Npm2 immobilized on Protein A sepharaose. Remaining soluble proteins were then separated by size exclusion chromatography with the purest fractions being combined for further analysis.

3.4.2 Digestions

Digestion buffer was prepared with 100 mM ammonium bicarbonate, occasionally including 2M urea buffer to mildly denature the protein, and the pH was adjusted to approximately 7.5 with 10% acetic acid. The samples of egg purified Nap1l1 or recombinant Nap1l1 were diluted 1:1 with digestion buffer. To reduce the disulfide bonds, dithiothreitol was added to the sample to a concentration of 2 mM and heated to 50C for 30 minutes. Iodoacetamide was prepared and added to a concentration of 6 mM and incubated in the dark for 30 minutes. Endoproteinase Arg-C (cleaves c-terminal to arginine) or Trypsin (cleaves

c-terminal to lysine/arginine) was added in a 1:100 (w/w) ratio with Nap1l1. Samples were incubated either at room temperature for 16 hours or at 37C for 6 hours. Samples were frozen until they were evaluated by LC-MS/MS.

3.4.3 LC-MS/MS Method

The samples were analyzed using in-house prepared columns. Both analytical and preparatory columns were made of fused silica columns (360 µm OD x 75 um ID for analytical and 100 um ID for preparatory) and had a 2 mm kasil frit. Analytical columns had a laser pulled tip. Two separate packing materials were used to evaluate samples: shorter peptides were evaluated with Dr. Maisch C18 packing material, both 3 and 5 µm particle size, packed to 10 cm; larger peptides were evaluated with Agilent Poroshell, SB-C18 5 µm particles packed to 10 cm. Reverse-phase separation was performed on an Agilent 1100 using 0.1% acetic acid for solvent A and 0.1% acetic acid in 60% acetonitrile for solvent B for short peptides or 0.2% formic acid for solvent A and 0.2% formic acid in 80% acetonitrile for solvent B for longer peptides. Gradients were run from 0-100% solvent B in 60 minutes at a flow rate of 100 µL/minute.

The preparatory columns were pressure loaded with 10 fmol equivalent of Pierce Retention Time Calibration Mixture and approximately 10-20 pmol of the digested material. The columns were rinsed with Solvent A prior to evaluation on a Thermo Fusion Tribrid ESI-QT-LIT-OT. A full mass spectrum was acquired in the Orbitrap with 60,000 resolution at 200 m/z from 300 – 2000 m/z. Precursors were selected for fragmentation utilizing a decision tree process based on charge state and charge density outlined below in Figure 5. Roughly, species with high charge density were fragmented by ETD; species with lower charge density were fragmented by CAD. Low-resolution MS/MS spectra were acquired for precursors with a charge state from +2 to +5, and high-resolution spectra were acquired for precursors with a charge state from +5 to +10. Precursors were excluded for 20 seconds after selection for fragmentation. The parameters outlined above changed slightly depending on the goal of the particular experiment.

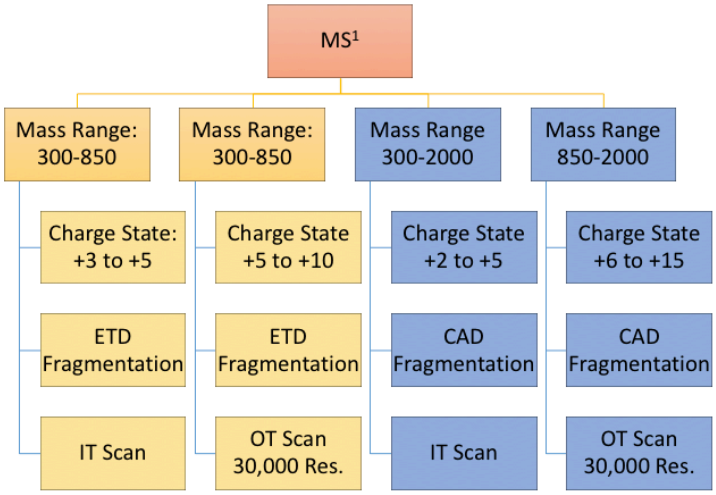

Figure 30. Fusion decision tree for Nap1l1 digests. The decision tree process optimizes the instrument time to the optimal fragmentation method and scan type.

3.4.4 Search Parameters

Data was searched using Byonic, by Protein Metrics. All results reported in the following sections were manually confirmed. Data was searched initially against the *Xenopus Laevis* database downloaded from Xenbase.org and then further searched a limited protein set of Nap1l1 variants along with PTMs. Variable modifications searched with Byonic included: N-terminal acetylation, lysine acetylation, arginine and lysine mono- and di-methylation, lysine trimethylation, serine/threonine/tyrosine phosphorylation, cysteine farnesylation, and up to 8 glutamylations.

3.4.5 Synthetic Glutamylated Peptides

Two synthetic peptides with glutamylated side chains were weighed and diluted with water to 1 mM stock concentration. The peptides, see Table 1, were combined at 2 µM in 50% 0.1% acetic acid and 50% Acetonitrile. The peptides were then infused on the Thermo Fusion Tribrid for experiments to confirm glutamylation fragmentation.

Full mass spectra were acquired with 100,000 resolution at 200 m/z. Precursors were fragmented by CAD with 35% collision energy, and MS/MS spectra at 60,000 resolution were acquired for the z=+2 charge state of each peptide, using a 2 m/z quadrupole isolation.

Table 9. Glutamylated Synthetic Peptides

Sequence	Glutamylation	[M+H]
MYeDDDEESEAQGPK E	Monoglutamylation	1871.71 Da
MYeDDDEESEAQGPK EE	Diglutamylation	2000.75 Da

3.5 Results and Discussion

The purification for Nap1l1 from *Xenopus laevis* eggs (eNap1l1) was successful at approximately 50% as shown in Figure 31, B. The purified Nap1l1 was identified as being glutamylated by staining with antibodies for glutamylation (see Figure 31, C) (38).

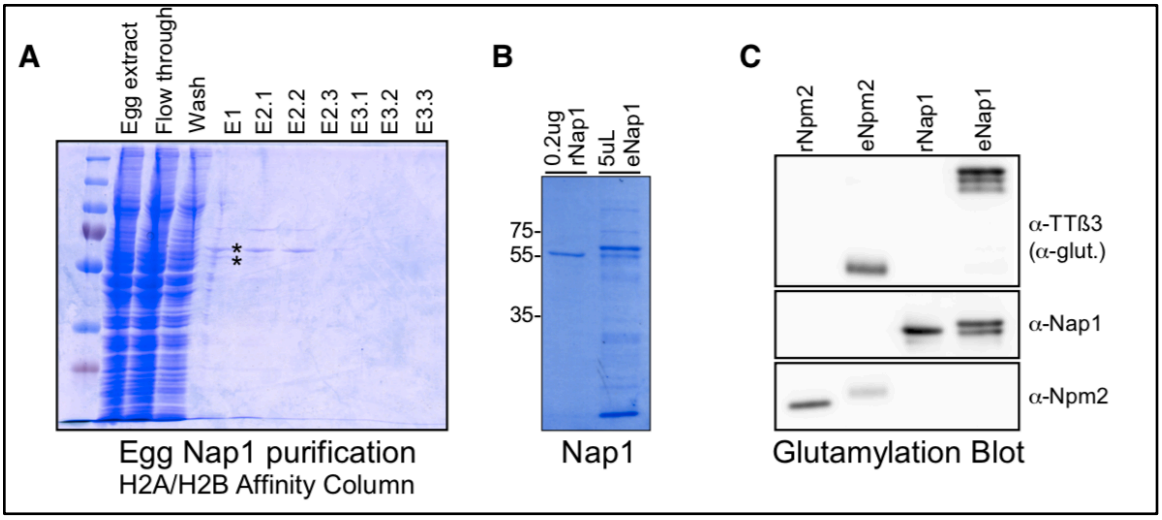

Figure 31. Nap1l1 purification results. A) The egg extract was run over streptactin bound H2A.S/H2B immobilized dimer pairs. The elution steps of the ion exchange were included and the bands corresponding to Nap1l1 are identified with *. (Coomassie stained) B) Recombinant Nap1l1 was run on a gel along with egg purified Nap1l1. Note there are two strong bands at 55 kDa for Nap1l1. (Coomassie stained) C) Western blots for the anti-TTB3 (glutamylation) (top), anti-Nap1 (middle), anti-nucleoplasm (bottom.) As a control, nucleoplasm is shown to be glutamylated and the recombinant Nucleoplasm is not glutamylated.

Sequence variants are one of the challenges posed by endogenously expressed. Published protein sequences are frequently the result of genomic sequencing, reporting the most abundant result from the genetic pool. When working with recombinant proteins, the sequence is known and controlled; each cell then produces copies of the exact same sequence of the over-expressed protein. Additionally, *Xenopus laevis* provides an additional

layer on to that challenge. Not only are the animals not homologous genetic replicates, but *Xenopus laevis* are tetraploids, and their resulting gamete cells are diploid. Thus, there are two gene loci for Nap1L1: Nap1l1.L is located on chromosome 3L, and Nap1l1.S is located on chromosome 3S. The sequences presented in Figure 27, Nap1l1.L (RefSeq NP_001080547) and Nap1l1.S (RefSeq XP_0018109639), were used as the base sequences due to their similarity to hNap1L1 and other reported species.

The tryptic digest provided nearly full sequence coverage of eNap1l1. Nap1l1 is an acidic protein, in which nearly 40% of its sequence are charged residues; Of those, 2/3 are acidic residues. However, of the 49 basic amino acid residues within the sequence, 3 out of 4 residues are lysine residues. With full tryptic digestion, 20% of the sequence is digested to pieces less than 600 Da in size, only observable at a 1+ charge state in the traditional 300-2000 m/z mass range and not selected for MS/MS analysis. Endoproteinase Arg-C was selected due to the significantly larger size of the digested fragments, with the optimal fragments ranging from 1,600 Da to 14,000 Da. With a full digest, only 4 peptides would be less than 600 Da in size, accounting for only 10 total residues. Figure 27 highlights the trypsin (blue lines) and the Arg-C (red R) peptides on Nap1l1.

With the Arg-C protease, both the first and third acidic regions were identified with glutamylation by both Regnard et al. and Miller and Heald (22, 23). As discussed above, two papers directly address the glutamylation of Nap1Ll1. Regnard et al. utilized trypsin to digest Nap1l1, mass spectrometry to analyze the intact mass of the tryptic peptides, and Edman degradation to sequence the peptides (23). The paper focused on the C-terminal peptides, identifying by mass up to 9 post translational glutamylations in HeLa cells and concluded that the glutamylation is most likely added in a single chain at a specific residue similar to tubulin proteins. Similar conclusions are drawn about the N-terminal peptides. The second paper by Miller and Heald also utilized a tryptic digest but analyzed the peptides using tandem mass spectrometry (22). It should be noted that they reference a different version of Nap1l1.S, Refseq NP 001080547, which is 303 residues long and does not include the N-terminal 90 residues. The paper reported that three residues are modified with “high confidence” but that other residues may also be modified.

The N-terminal peptide was identified for both Nap1l1.L and Nap1l1.S from the Arg-C digest in either 2-35 with a non-specific lysine cleavage, or residues 2-37 with the fully specific cleavage at the arginine residue. Predictably, the initial methionine is cleaved, and the protein is N-terminally acetylated. The sequences are reported below:

- Nap1l1.L 2-35: ANIDNKEQTELDQQDMEDVEDVEEEETGEEANSK.A
- Nap1l1.L 2-37: ANIDNKEQTELDQQDMEDVEDVEEEETGEEANSKAR.Q

- Nap1l1.S 2-35: ANIDNKEQTELDQQDMEDVEDVEEEETGEDANSK.A
- Nap1l1.S 2-37: ANIDNKEQTELDQQDMEDVEDVEEEETGEDANSKAR.Q

The first acidic region, was found to be polyglutamylated, consistant with the results from the Heald and Miller paper for Nap1l1.L. I identified Nap1l1.S (Figure 32) and Nap1l1.L (Figure 33) with a range of 0 to 4 and 6 glutamylations, respectively (22). It was observed, as one might predict, that increases in glutamylation caused an increase in the hydrophilic nature of the peptides and this was observable with a level of chromatographic separation detected as multiple peaks and shoulder peaks. While separation by m/z is a complementary to chromatographic separation, mass spectrometry does rely on chromatographic separation to separate isobaric peptides that may be differentially modified.

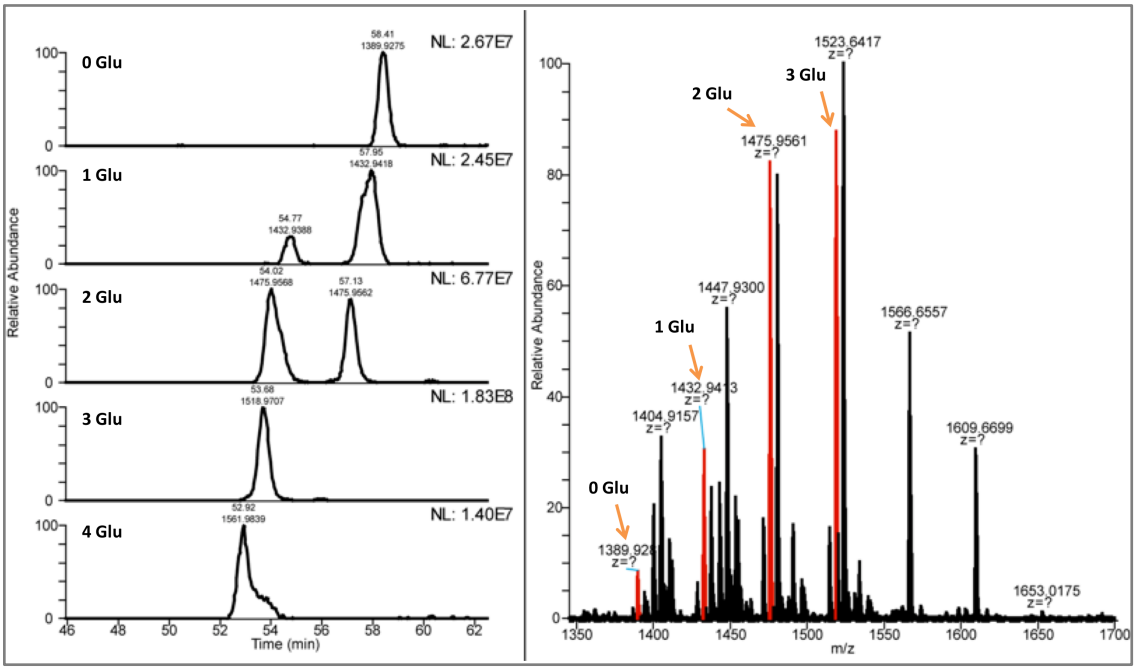

Figure 32. A) Extracted ion chromatograms for Nap1l1.S, peptide 2-37 with increasing numbers of glutamylations. B) Averaged MS1 from approximately 53-55 minutes; masses in red correspond to masses in A. Normalized abundances are presented as NL.

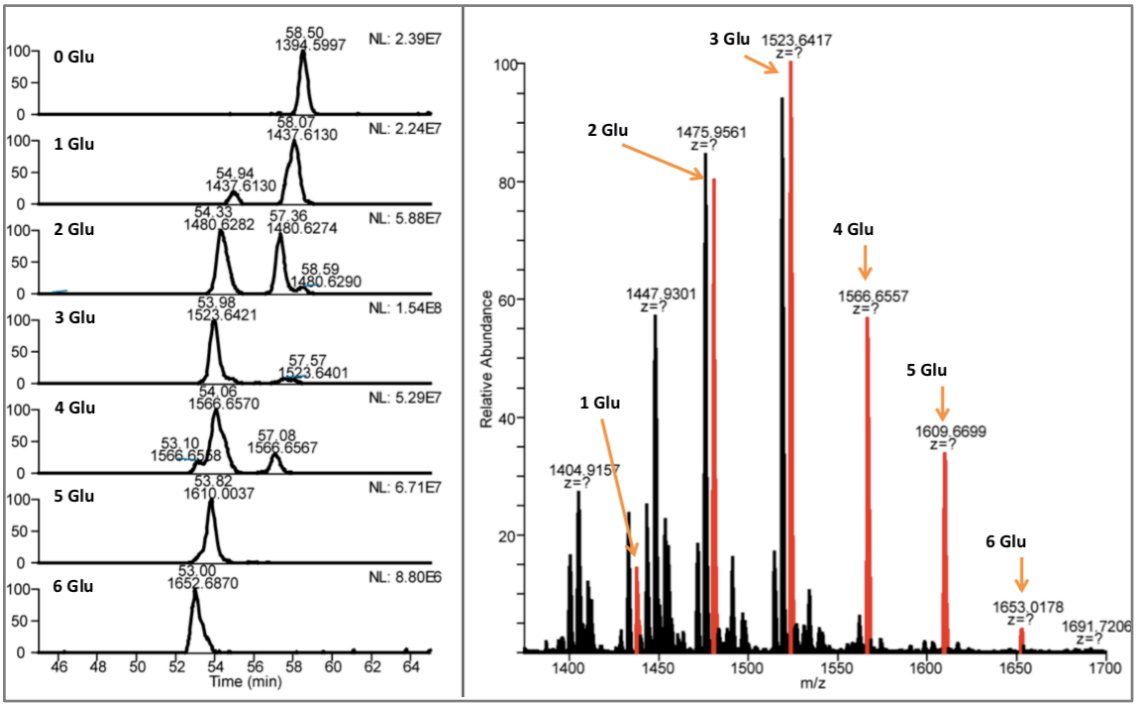

Figure 33. A) Extracted ion chromatograms for Nap1l1.L, peptide 2-37 with increasing numbers of glutamylations. B) Averaged MS1 from approximately 53-55 minutes; masses in red correspond to masses in A. Normalized abundances are presented as NL.

In order to site localize PTMs, fragmentation across the peptide back bone is required to confidently assign the location of the modification. One of the concerns for sequencing the glutamylated peptides is the inability to site localize the modification without ETD, which would ensure that the peptide would cleave at the N-C_α bond. However, the peptides were detected in the range of 1000-1600 m/z, representing a charge density of 1 proton for every 10-15 residues; optimal fragmentation by ETD requires a higher charge density as discussed in Chapter 1. The peptides were fragmented by CAD. One critique of CAD includes the potential loss of side chain fragments, creating ambiguous spectra that are not able to localize the modification. The side chain fragments complicate the mass spectrum and bring uncertainty to determining the site of the modification. Since the exact mass addition for a glutamylation (Δ = +129.04 Da) is identical with the delta mass for a glutamic acid residue, a complicated web of potential masses is created. Table 10 presents the first 19 y-ions for the unmodified Nap1l1.L 2-35 peptide and the y-ion fragments for monoglutamylated peptides at various residues. Additionally, the glutamylation is occurring in stretches with multiple glutamate residues; Nap1l1.L peptide contains 10 primary sequence glutamic acids, all potential locations for the post-translational modification. When comparing the fragments for the unmodified peptide with the monoglutamylated peptides in Table 10, it is important to note how many of the fragments are identical in predicted mass. For example, y_{12} at 1480.59 m/z is indicative of monoglutamylation occurring on or prior to that residue as compared to the unmodified peptide; but without a gap in sequence coverage at 964.45 m/z, there is no way to localize the modification to y_9. In theory, a peptide modified at y_9 would not have a mass detected at 964.45 m/z. Also, a side chain loss from a fragment ion during CAD yields a delta mass of -147.04 Da, similar to an uncharged y_1 Glu residue because of the C-terminal hydroxyl (see Figure 29). The side chain fragment losses are shown in Figure 34, which presents the fragmentation of a synthetic tubulin peptide containing mono or di-glutamylation at the same residue. The side chain loss from the y_{13} fragment is evident by the Δ = - 147 m/z, followed by the loss of the remaining residue (either Δ = -111 m/z for mono- or Δ = -129 and -111 m/z for di-glutamylation) to reach the y_{12} ion. Thus, if the peptide in

Table 10 is monoglutamylated but none of the "gaps" are detected between y_8 and y_{13}, the lack of "gaps" would indicate a population of different modification sites.

Table 10. y-ion fragment masses for the first 19 y-ions of the Nap1l1.L, 2-35 peptide predicted for CAD fragmentation. Each column indicates expected fragments for either an unmodified peptide or a peptide mono-glutamylated at specified residues.

Sequence	y ions	Unmodified	MonoGlu y_9 Mod	MonoGlu y_{10} Mod	MonoGlu y_{11} Mod	MonoGlu y_{12} Mod
M	19	2168.87	2297.91	2297.91	2297.91	2297.91
E	18	2037.83	2166.87	2166.87	2166.87	2166.87
D	17	1908.78	2037.83	2037.83	2037.83	2037.83
V	16	1793.76	1922.80	1922.80	1922.80	1922.80
E	15	1694.69	1823.73	1823.73	1823.73	1823.73
D	14	1565.64	1694.69	1694.69	1694.69	1694.69
V	13	1450.62	1579.66	1579.66	1579.66	1579.66
E	12	1351.55	1480.59	1480.59	1480.59	1480.59
E	11	1222.51	1351.55	1351.55	1351.55	1222.51
E	10	1093.46	1222.51	1222.51	1093.46	1093.46
E	9	964.42	1093.46	964.42	964.42	964.42
T	8	835.38	835.38	835.38	835.38	835.38
G	7	734.33	734.33	734.33	734.33	734.33
E	6	677.31	677.31	677.31	677.31	677.31
E	5	548.27	548.27	548.27	548.27	548.27
A	4	419.22	419.22	419.22	419.22	419.22
N	3	348.19	348.19	348.19	348.19	348.19
S	2	234.14	234.14	234.14	234.14	234.14
K	1	147.11	147.11	147.11	147.11	147.11
Missing Mass		**1480.5921**	**964.45**	**1093.46**	**1222.51**	**1351.55**

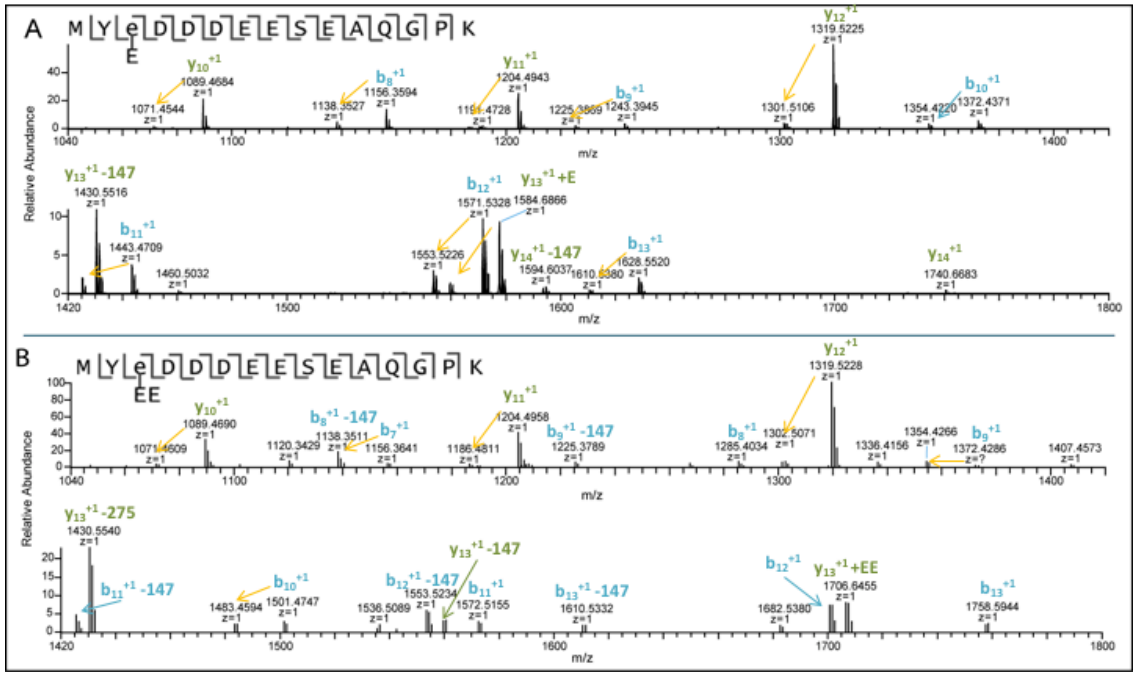

Figure 34. CAD spectra for a synthetic peptide created from an alpha-Tubulin peptide with A) monoglutamylation or B) diglutamylation on the third residue. The yellow arrows indicate a $\Delta = -18$, common for water loss. Note in B, b_8-147 is identical in mass to the b_7-18.

Despite the challenges posed by fragmenting the peptides by CAD, the goal of my work is to site localize the glutamylation. The non-glutamylated N-terminal peptides for both Nap1l1.L and Nap1l1.S confirmed that the protein was N-terminally acetylated and the primary chain glutamyl sequence matched the published sequences as noted above. For the modified peptides, the MS/MS spectra acquired across the peaks were evaluated to determine if the glutamylation could be site localized. By plotting the abundance of specific fragment ions (see Table 10) from the MS/MS spectra of the monoglutamylated precursor, as per Figure 35, changes in ion abundances were noted across the chromatogram. Peak A was determined to be modified at E24, observed by the 0% relative abundance of 1351.5 m/z. Peak B was determined to be modified at E25 and Peak B' was determined to be modified at E26 (see Figure 36). The monoglutamylated peptides were separated chromatographically by site specific modification.

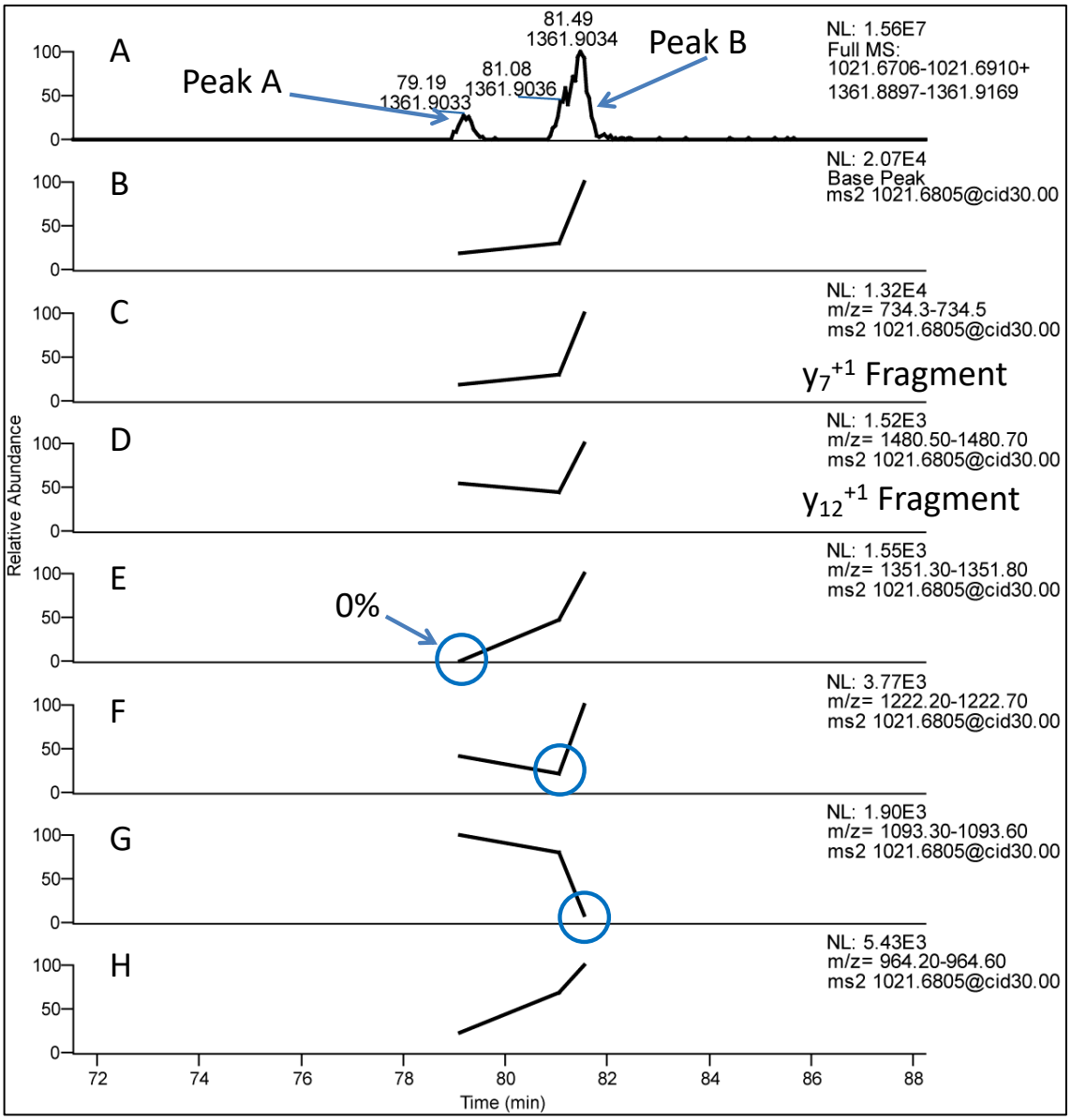

Figure 35. Extracted ion chromatograms for various y-ion fragments. A) Extracted ion chromatogram from MS1 for the +3 and +4 of monoglutamylated Nap1L 2-35 B) Base peak abundance of the CAD MS/MS of the +4 precursor (3 spectra acquired) C) Plot of the abundance of the y_7^{+1} fragment (GEEANSK) in the MS/MS (non-glutamylated) (734.4m/z) D) Plot of the abundance of the y_{12}^{+1} fragment (1480.6 m/z) E-H) Plots of abundance for the following ions, respectively: 1351.5, 1222.5, 1093.4, 964.4.

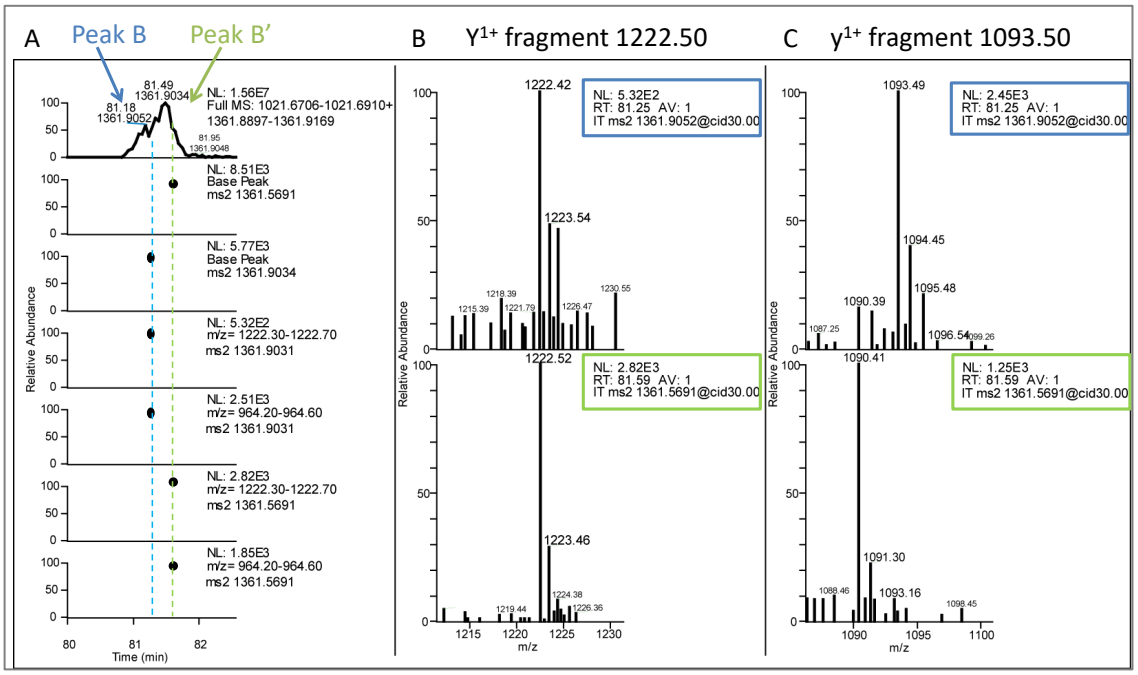

Figure 36. Expanded Chromatogram for Peak B from Figure 10. A) The same precursors, but different isotopes, were selected for MS/MS under Peak B (blue dotted line) and Peak B' (green dotted line). Specific fragment masses can be extracted from the ms/ms scans. B) Plots the abundance of 1222.5 m/z for each ms/ms. C) Plot of the abundance for 1093.5 m/z for each ms/ms. Peak B has a decreased abundance for the 1222 fragment and clear peak at 1093.5 indicating y_{11} is glutamylated. Peak B' has a detected abundance for 1222.5 and 1093.5 is not detected indicating that y_{10} is modified with glutamylation.

The diglutamylated peptides were analyzed in a similar manner, with multiple large chromatographic peaks. Table 11 presents the y-ion fragment masses for the differently modified peptides, and Figure 37 presents the chromatogram and extracted fragment ion abundances across the chromatographic peaks. Peak A was determined to be modified at E24 and E25, confirmed with depleted relative abundances at 1222.51 m/z and 1480.59 m/z and detected masses at 1093.46 m/z and 1351.55 m/z. By similarly evaluating the other peaks, Peak B was determined to be monoglutamylated at both E24 and E26, Peak C was monoglutamylated at E25 and E26, and Peak D was monoglutamylated at E30 and E31. An interesting note is that consistent with the monoglutamylated precursors, there is full chromatographic separation between the peaks containing a modified E24 and those without the modified E24.

Table 11. y-ion^{+1} fragments for diglutamylated Nap1l1.L, peptide 2-35. Differentially colored sections correspond to the peaks outlined in *Figure 37* and the sequence informative ions for determining locations of glutamylation.

Sequence	y ions	y_{11}, y_{12} Mod	y_{10}, y_{12} Mod	y_{10}, y_{11} Mod	y_5 and/or y_6
E	15	1952.77	1952.77	1952.77	1952.77
D	14	1823.73	1823.73	1823.73	1823.73
V	13	1708.70	1708.70	1708.70	1708.70
E**	12	1609.63	1609.63	1609.63	1609.63
E**	11	1351.55	1351.55	1480.59	1480.59
E**	10	1093.46	1222.51	1222.51	1351.55
E	9	964.42	964.42	964.42	1222.51
T	8	835.38	835.38	835.38	1093.46
G	7	734.33	734.33	734.33	992.42
E*	6	677.31	677.31	677.31	935.40
E*	5	548.27	548.27	548.27	677.31
A	4	419.22	419.22	419.22	419.22
N	3	348.19	348.19	348.19	348.19
S	2	234.14	234.14	234.14	234.14
K	1	147.11	147.11	147.11	147.11
Missing Masses		**1480.59**	**1480.59**	**1351.55**	**964.42**
		1222.51	**1093.46**	**1093.46**	**734.33**

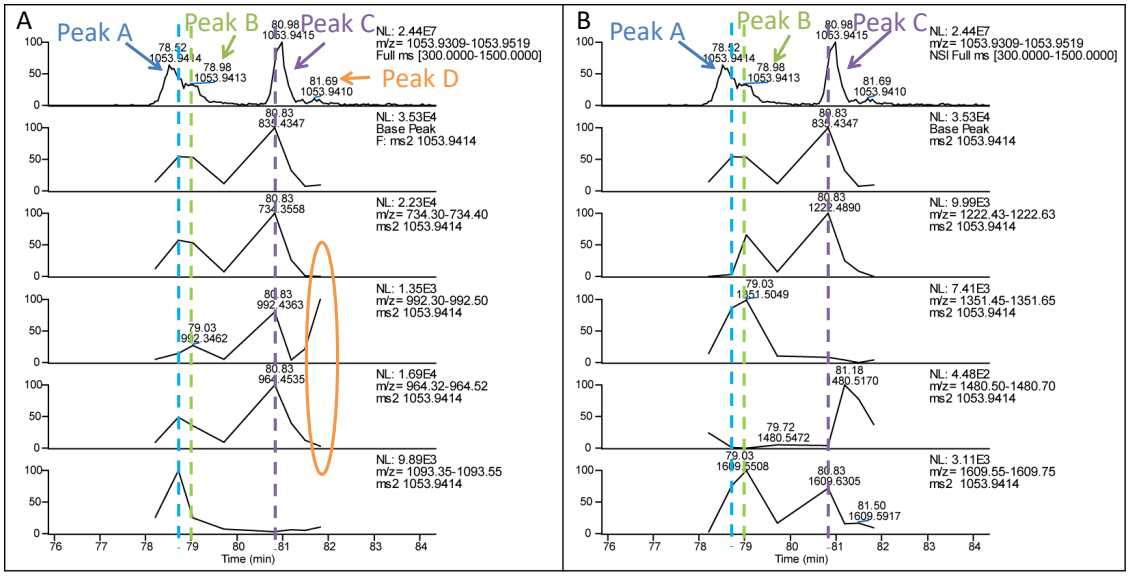

Figure 37. Extracted ion chromatograms for various y-ion fragments for the di-glutamylated peptides of Nap1l1.L, 2-35. A) Top to Bottom: Extracted ion chromatogram from MS1 for the +3 and +4 of diglutamylated Nap1l1.L 2-35; Base peak abundance of the CAD MS/MS of the +4 precursor (8 spectra acquired); Plot of the abundance of the y_7^{+1} fragment (GEEANSK) in the MS/MS (non-glutamylated) (734.4m/z); Plot of the abundance of the diglutamylated y_7^{+1} fragment (GEEANSK) in the MS/MS (992.4m/z); Plot of abundance 964.4 m/z; Plot of abundance for 1093.45 m/z. B) Top to Bottom: Extracted ion chromatogram from MS1 for the +3 and +4 of

diglutamylated Nap1l1.L 2-35; Base peak abundance of the CAD MS/MS of the +4 precursor (8 spectra acquired); Plot of abundance of 1222.53, 1351.55, 1480.6, 1609.65 m/z respectively.

Both the Nap1l1.S and Nap1l1.L were manually investigated to site map the major glutamylated species; the results of this work are presented in Table 12 and Table 13. Bracketed results represent sequences in which specific site localization was not possible; sequences labeled as major species represent confirmed modified sites. The predominant abundance of the glutamylation is monoglutamylation on 2 or 3 residues across the first acidic region of Nap1l1. The modification appears to initiate in the stretch of E24-26, a region with 4 consecutive glutamic acid residues. A maximum of 6 polyglutamylations were observed on Nap1l1.L and 4 polyglutamylations on Nap1l1.S; Nap1l1.L contains an additional primary chain E at E31, which appears to affect glutamylation at E30 and E31. Glutamylation after E30 on Nap1l1.S was not detected, and this appears to be the explanation for the reduced level of glutamylation detected for Nap1l1.S.

Table 12. Summary of peptides for N-terminus of Nap1S|1 with identified glutamylation patterns.

Glutamylation	Fragment	Peak Area	Sum/%	Modifications
0	2-35	1.1×10^8	7.1×10^8 (6%)	N-terminal Acetylation
0	2-37	6.0×10^8		
1	2-35	2.7×10^8	7.2×10^8 (6%)	Acetyl-ANIDNKEQTELDQQDMEDVEDVE*EEETGEDANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVEE*EETGEDANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVEEE*ETGEDANSKAR
1	2-37	4.5×10^8		
2	2-35	8.3×10^8	4.9×10^9 (39%)	Major Species: Acetyl-ANIDNKEQTELDQQDMEDVEDVE*E*EETGEDANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVEE*E*ETGEDANSKAR Undetermined Specificity: Acetyl-ANIDNKEQTELDQQDMEDVEDV[EEEE]**TGEDANSKAR
2	2-37	4.1×10^9		
3	2-35	8.7×10^8	5.7×10^9 (45%)	Major Species: Acetyl-ANIDNKEQTELDQQDMEDVEDVE*E*E*ETGEDANSKAR
3	2-37	4.8×10^9		
4	2-35	---	4.4×10^8 (4%)	Major Species: Acetyl-ANIDNKEQTELDQQDMEDVE*DVE*E*E*ETGEDANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVEE*E*E**TGEDANSKAR
4	2-37	4.4×10^8		
5	2-37	Faint	Not Reported	Intact Mass Detection Only

Table 13. Summary of peptides for N-terminus of Nap1L|1 with identified glutamylation patterns.

Glutamylation	Fragment	Peak Area	Sum/%	Modifications
0	2-35	4.0×10^8	9.2×10^8 (5%)	N-terminal Acetylation
0	2-37	5.2×10^8		
1	2-35	5.3×10^8	1.3×10^9 (6%)	Major Species: Acetyl-ANIDNKEQTELDQQDMEDVEDVE*EEETGEEANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVEE*EETGEEANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVEEE*ETGEEANSKAR
1	2-37	7.6×10^8		
2	2-35	1.9×10^9	5.5×10^9 (27%)	Major Species: Acetyl-ANIDNKEQTELDQQDMEDVEDVE*E*EETGEEANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVE*EE*ETGEEANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVEE*E*ETGEEANSKAR Undetermined Specificity: Acetyl-ANIDNKEQTELDQQDMEDVEDV[EEEETGEE]**ANSKAR
2	2-37	3.6×10^9		
3	2-35	2.0×10^9	7.0×10^9 (35%)	Major Species: Acetyl-ANIDNKEQTELDQQDMEDVEDVE*E*E*ETGEEANSKAR Acetyl-ANIDNKEQTELDQQDMEDVE*DVE*E*EETGEEANSKAR Undetermined Specificity: Acetyl-ANIDNKEQTELDQQDMEDV[EDVEEEETGEE]***ANSKAR
3	2-37	5.0×10^9		
4	2-35	5.7×10^8	3.3×10^9 (16%)	Major Species: Acetyl-ANIDNKEQTELDQQDMEDVE*DVE*E*E*ETGEEANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVE*E*EETGE*E*ANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVE*EE*ETGE*E*ANSKAR Acetyl-ANIDNKEQTELDQQDMEDVEDVEEE*E*TGE*E*ANSKAR Undetermined Specificity: Acetyl-ANIDNKEQTELDQQDMEDV[EDVEEEETGEE]****ANSKAR Evidence for: GE**EANSKAR or GE*E*ANSKAR
4	2-37	2.7×10^9		
5	2-35	2.0×10^8	2.0×10^9 (10%)	Major Species: Acetyl-ANIDNKEQTELDQQDMEDVEDV[EEEE]*****TGEEANSKAR Acetyl-ANIDNKEQTELDQQDMEDVE*DV[EEE]****ETGEEANSKAR Acetyl-ANIDNKEQTELDQQDMEDVE*DVEE*E*ETGE*E*ANSKAR
5	2-37	1.8×10^9		
6	2-37	Faint	Not Reported	Undetermined Specificity: Acetyl-ANIDNKEQTELDQQDMEDV[EDVEEEE]****TGE*E*ANSKAR

The maximum number of additional glutamylation levels for the first acidic region appear to be consistent with both Regnard et al. and Miller and Heald. However, this data presents a more detailed and nuanced look at the glutamylation. While evidence to suggest diglutamylated residues may have been observed, the predominant modification is a monoglutamylation across a few residues. The third acidic region was also identified with heavy glutamylation, consistent with the previous results. Furthermore, the C-terminal cysteine residue was found to be farnesylated concurrently with the glutamylation. While the farnesylation was confirmed with the tryptic digest, both modifications were detected on the same peptides in the Arg-C digest.

The CaaX motif for eNap1l1 is farnesylated, as confirmed by the tryptic peptide K.KAQNPAEcKQQ.Cterm containing several missed cleavages. The CAD spectrum had the characteristic Δ=-205 Da ion loss from the precursor as the major species. However, the sequence was unambiguously confirmed by ETD for a +3 peptide that contained the two

missed cleavages, presented in Figure 38. Farnesylation was confirmed for both Nap1l1.L and Nap1l1.S. Farnesylation is a permanent modification as discussed in the introduction. After farnesylation of the -CaaX motif, the modified protein has the potential to be further modified, the tripeptide -aaX can be clipped from the protein, followed by C-terminal carboxymethylation to minimize the negative charge (39). To remove the permanent farnesylation, such as occurs with Lamin A, the last 15 residues (18 in total) are clipped by an endoproteinase to produce a mature protein (40). The trypsin digest results were manually searched for the clipped tripeptide both with and without the carboxymethylation but none were identified. Within the tryptic result, (protease cleavage c-terminal to K and R) the predicted peptide would have an expected charge of +1, but no peptides were identified.

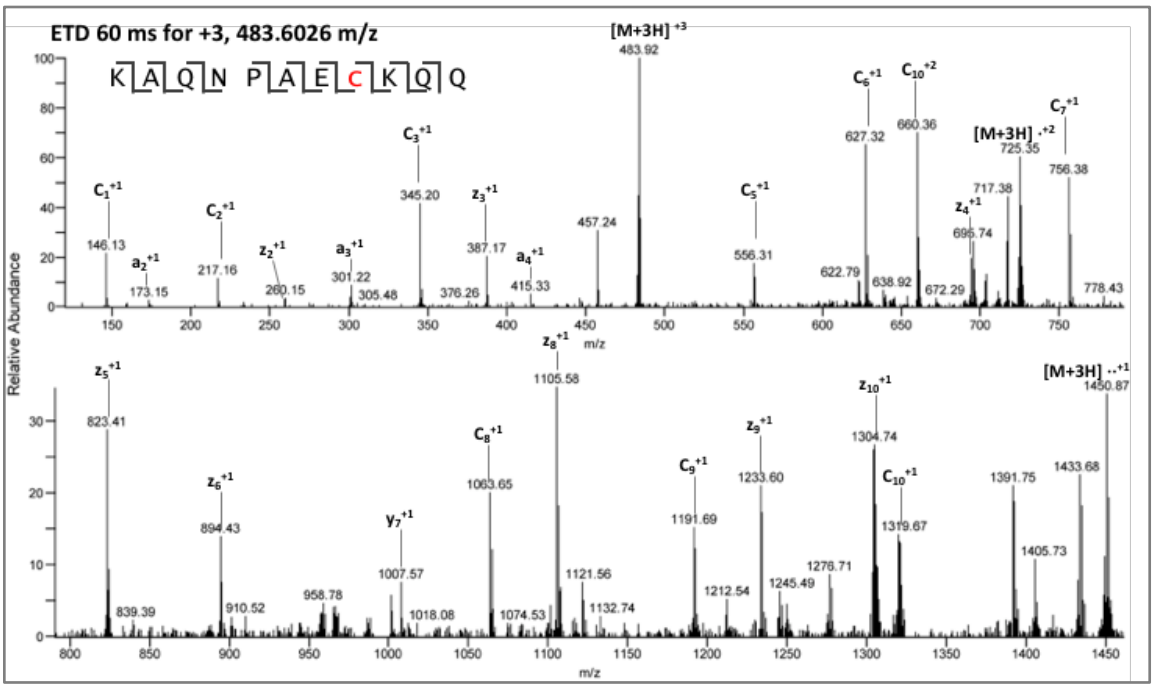

Figure 38. Farneslyated Cysteine sequence coverage was obtained for the +3 peptide; while the c/z fragments n-terminal to the proline are not observed with ETD, the a/y ions were detected. The peptide KGQNPAECKQQ for Nap1l1.S was confirmed in a similar manner with farnesylation.

The C-terminal peptides predicted for the Arg-C digest contain both the third acidic region as well as the -CaaX motif. The peptides were detected both with and without the farnesylated cysteine. Additionally, a third variant was identified, similar to the Nap1l1.L sequence and identified in Table 14 as Nap1l1.LX2; the peptide matches the final 45 residues for the previously reported variant in RefSeq, XP 018106203.1. The sequence has a lysine to

valine substitution at 383 and the sequence stops at V383, 10 residues shorter than the Nap1l1.L and Nap1l1.S sequences identified in the same sample. It does not contain the c-terminal cysteine. As a note, a similar variant for Nap1l1.S was not observed in the experiments. The mass difference for the Nap1l1.S would be +5 Da, within the isotopic distribution of Nap1l1.LX2 and difficult to observe at low levels. A summary of the most abundant C-terminal containing peptides are presented in Table 14 and the chromatographic separation is shown in Figure 39.

Table 14. List of identified C-terminal peptides from eNap1l1 in Arg-C digest

Variant	Fragment	Mass	Sequence	Approx. Ret. Time	Cys Modification	Detected Glutamylation
Nap1l1.L	339-393	6322.5025	R.SVLYFTGEAIEDDDDYDEEGEEADDEEGEEEADEDHDPDFDPKKAQNPAECKQQ.--	49.0	+57	4x to 9x
		6469.6693		64.4	+204	3x to 10x
	335-393	6801.8245	R.IIPRSVLYFTGEAIEDDDDYDEEGEEADDEEGEEEADEDHDPDFDPKKAQNPAECKQQ.--	57.1	+57	6x to 9x
		6948.9913		72.5	+204	5x to 9x
Nap1l1.LX2	339-383	5138.9625	R.SVLYFTGEAIEDDDDDYDEEGEEADDEEGEEEADEDHDPDFDPKV.--	51.3	No Cys	1x to 11x
	335-383	5618.2845	R.IIPRSVLYFTGEAIEDDDDDYDEEGEEADDEEGEEEADEDHDPDFDPKV.--	55.3	No Cys	2x to 11x
	333-383	5903.4282	R.ERIIPRSVLYFTGEAIEDDDDDYDEEGEEADDEEGEEEADEDHDPDFDPKV.--	60.2	No Cys	5x to 11x
Nap1l1.S	338-392	6315.4814	R.SVLYFTGEAIEDDDDYDEEGEEADDEEGEEEADEDNDPDYEPKKGQNPAECKQQ.--	48.0	+57	3x to 10x
		6462.6482		62.8	+204	2x to 10x
	334-392	6794.8034	R.IIPRSVLYFTGEAIEDDDDYDEEGEEADDEEGEEEADEDNDPDYEPKKGQNPAECKQQ.--	54.0	+57	4x to 9x
		6941.9702		64.4	+204	4x to 10x

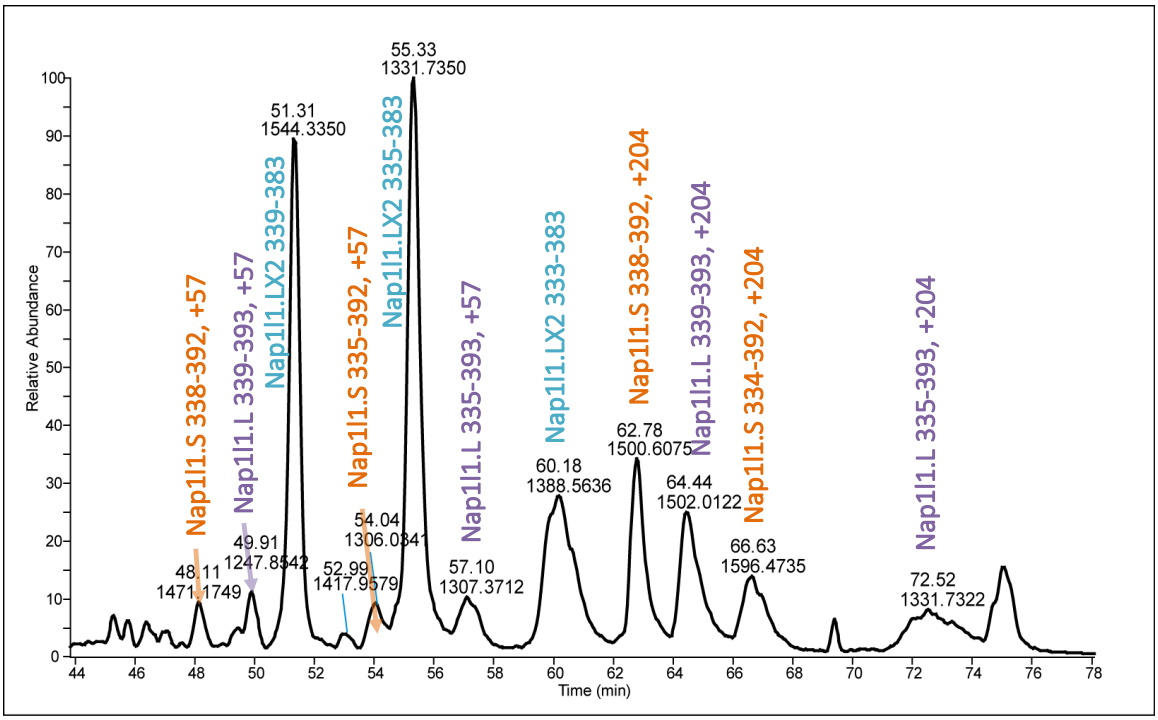

Figure 39. Extracted ion chromatogram (1200-2000 m/z) from the Arg-C digest highlighting the elution order of the C-terminal peptides; Nap1l1.L (purple), Nap1l1.S (orange), Nap1l1.LX2 (blue)

The third acidic region was identified with a wide range of glutamylation in Nap1l1.L and Nap1l1.S, both identified with 3 or 4 to 10 glutamylations; and Nap1l1.LX2 was identified with from 1 to 11 glutamylations (see Figure 40). There are several specific differences from the first acidic region which directly affect the depth to which the peptides can be characterized. One challenge was that the unmodified peptides were not detected. Since the primary sequence cannot be confirmed from an unmodified peptide, the results are assessed based on the theoretical sequences of the protein, which are reported in Table 14 . The third acidic region with the potential to be modified for all three variants is assumed to have the sequence DEEGEEADDEEGEEEA. Furthermore, the peptides did not significantly chromatographically separate by numbers of polyglutamylations. The more abundantly modified peptides eluted slightly earlier than the less modified peptides; but as observed in Figure 40, the most abundant species eluted at the same retention time.

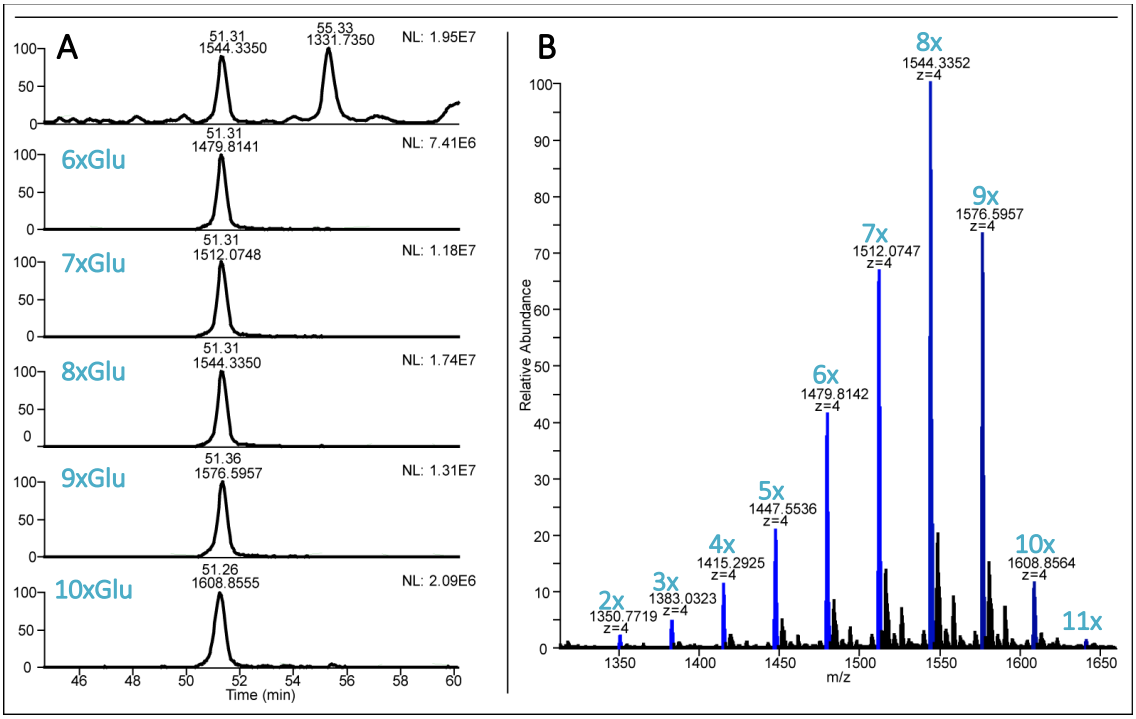

Figure 40. Extracted ion chromatograms of Nap1l1.LX2 339-383 A) Top to bottom: Extracted ion chromatogram base peak for 1200-2000 m/z; extracted ion chromatograms for Nap1l1.LX2, 339-383 with 6x to 10x glutamylations. NL represents normalized ion intensity. B) Full mass spectrum averaged across the peak at 51.13 minutes showing the +4 charge state for the Nap1l1.LX2 339-383 peptide with multiple glutamylations.

Finally, for confident ion assignments of the third acidic region, high resolution MS/MS spectra needed to be acquired. For the first acidic region, the low resolution MS/MS were evaluated because the peptides were shorter, contained fewer modifications, and contained fewer charge. Low resolution MS/MS are acquired in the ion trap, and the scan for the detected fragment masses takes a few milliseconds. By comparison, high resolution MS/MS are acquired in the Orbitrap. Not only does the acquisition of the transient itself takes tens of milliseconds to acquire depending on the desired resolution, but the ion target is higher for the Orbitrap requiring a longer ion injection times. Thus, fewer scans can be acquired in the same amount of time. The resulting combination of diminished chromatographic separation, increased numbers of modifications, increased numbers of available charge states, and fewer scans makes the characterization of the C-terminus difficult.

While the unmodified peptides were not detected, the peptides with 9 glutamylation sites were very abundant for each peptide. The extracted ion chromatograms for the Nap1l1.LX2 peptides for 339-383 and 335-383 are displayed in Figure 41 along with the ion chromatograms for the MS/MS events across the peaks. The peptide from 339-383 was sampled at either side of the apex of the peak at 51.36 minutes while the peptide from 335-383 was sampled at the apex of the peak at 55.29 minutes. Evaluating the extracted mass spectrum for the 335-383 peptide in Figure 42 gives a clear progression of nine monoglutamylations from E357 to E370. The spectrum has complementary pairs up to residue D364, confirming unambiguous monoglutamylation at E357-E361. The observation of the b-ion series to D364 is due to the missed cleavage at R338, allowing for multiply charged b-ions to be detected within the mass range.

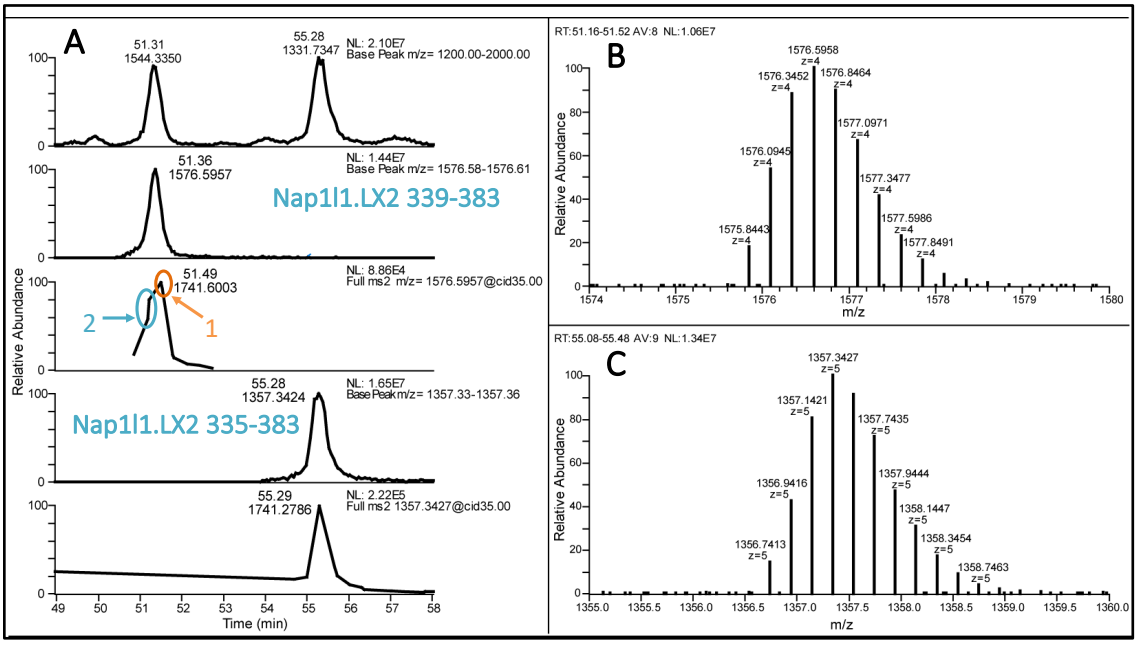

Figure 41. Chromatogram of the showing the MS/MS for the +4 and +5 for Nap1l1.LX2

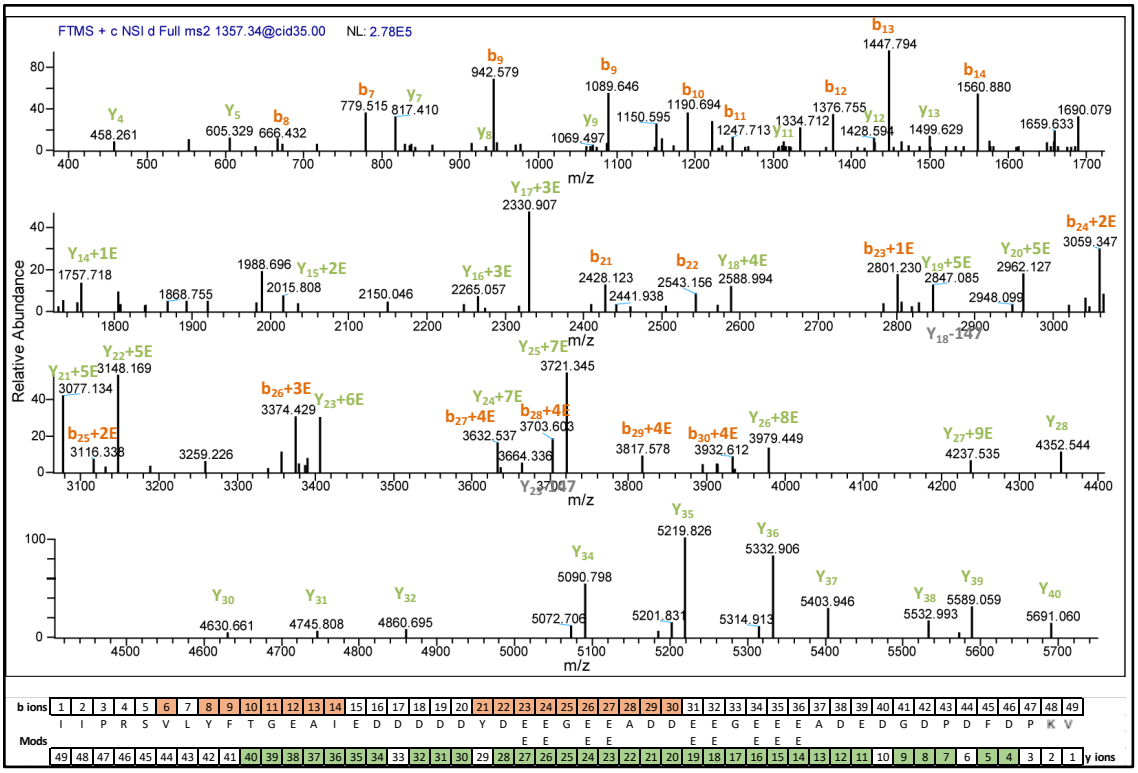

Figure 42. Deconvoluted ion spectrum [M+H] for the z=+5 of Nap1l1.LX2 335-383 with 9 glutamylations; ion fragment coverage is summarized at bottom. (Spectra deconvoluted with Xtract™ from Thermo)

Unlike the longer 335-383 peptide (Table 14), 339-383 does not have a basic residue near the N-terminus of the peptide, much of the b-ion series is singly charged and is not detected within the mass range. The most abundant MS/MS for 339-383 from Figure 41 at 51.49 minutes contains a strong y-ion series that matches the modification pattern observed in the longer peptide, see Figure 43. Also observed are the Δ = -147 Da side chain fragments from some of the glutamylated residues, where a Δ = -129 Da is not observed. In comparison, the same precursor sampled from earlier in the peak, 2 from Figure 41, does not give the same clear interpretation, presented in Figure 44. In the deconvoluted spectrum, two series can be observed for y_{19} to y_{22} (ADDE). For the blue series, y_{22} is detected, for the yellow series, y_{23}+E or Y_{24} is the next detected ion. Because the chromatographic separation between the two differently modified peptides series is not significant enough, the mixed populations of differently modified ions create ambiguous mass spectra that cannot be confidently sequenced. What is known is that a total of 9 glutamylations are on the peptide, and part of the population has 4 glutamylations by y_{19} and the other population has 5 glutamylations by y_{19}. For the population with 4 glutamylations at y_{19}, there is at least a diglutamylated residue by y_{27}, but the site cannot be confidently assigned, in part to the ambiguity of G367. There are too many potential combinations of modifications, a few of which are outlined in Figure 44. As the number of glutamylation events decreases across the peptide, there are increasing combinations of potential assignments.

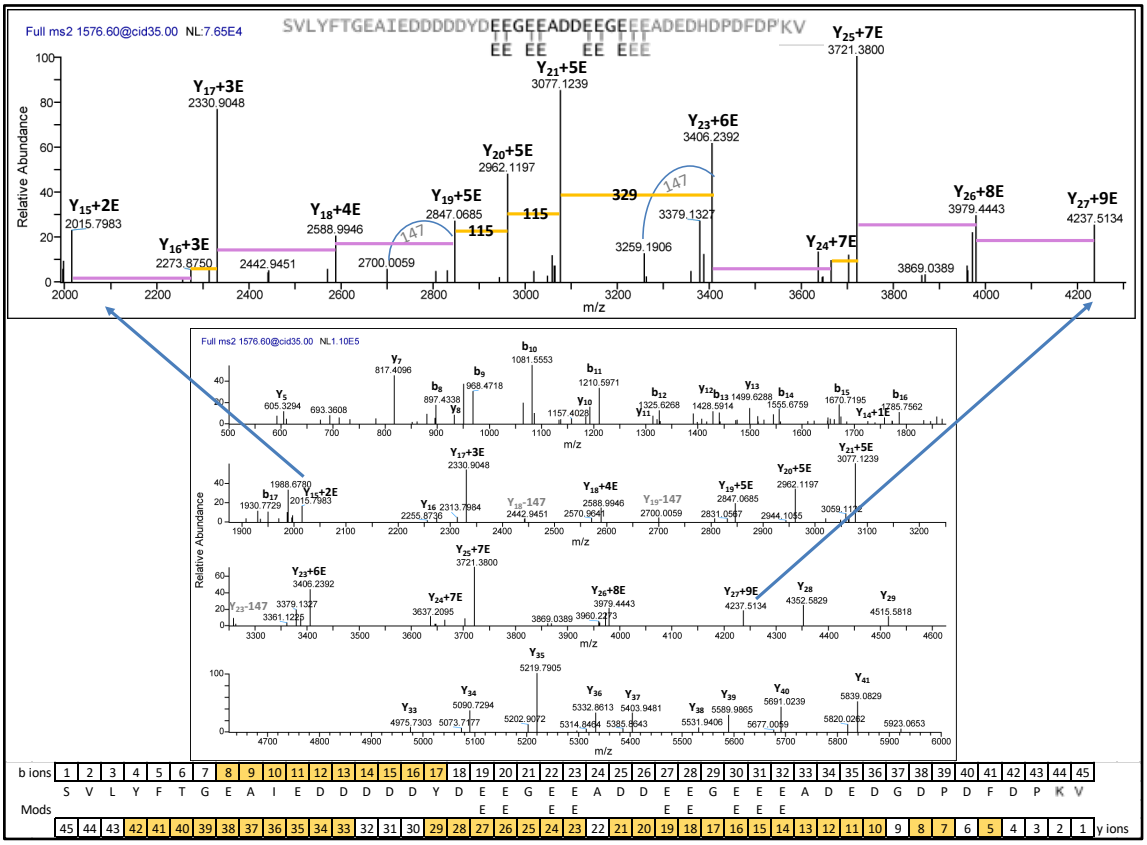

Figure 43. Deconvoluted ion spectrum [M+H] of the MS/MS for Nap1l1.LX2 339-383, 9xGlu at 51.49 minutes, labeled 1 in *Figure 41*. Top: Expanded region of the glutamylated region of the spectrum (Middle), pink lines represent Δ=258.1 Da; Bottom: Fragment ion coverage map of the presented spectrum

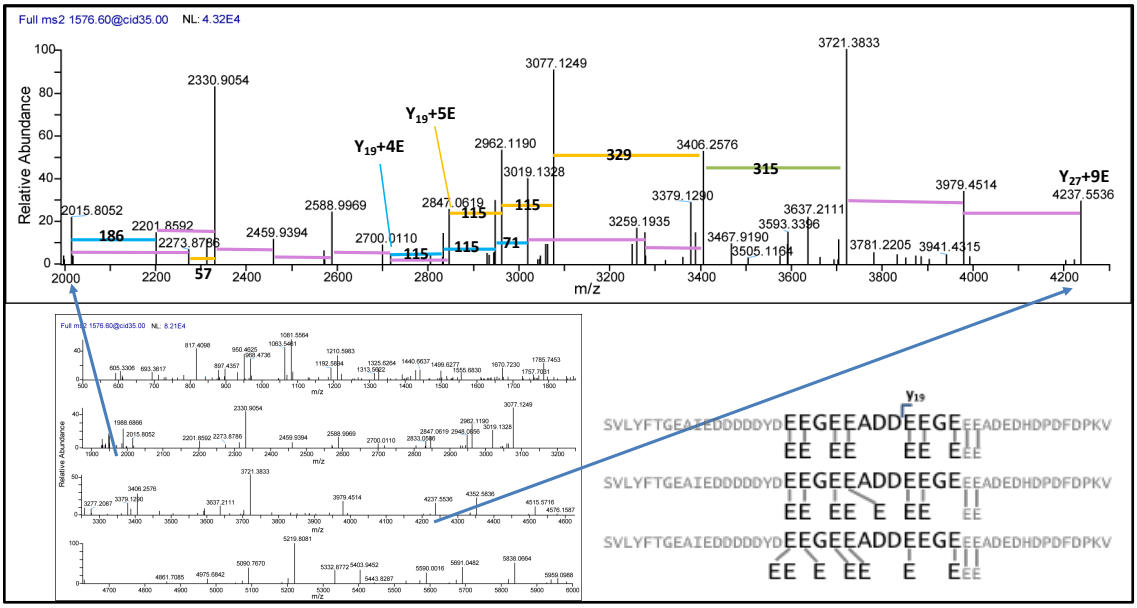

Figure 44. Deconvoluted ion spectrum [M+H] of the MS/MS for Nap1l1.LX2 339-383, 9xGlu at 51.25 minutes, labeled 2 in Figure 16. The expanded region from the full spectrum (bottom left) is shown (top). Pink lines represent Δ=+129.04 or +258.08 (E or EE). Several potential theoretical precursor sequences are shown (bottom right).

Table 15. Summary of glutamylation distribution of 3rd acidic region

Glutamylation	% of Total Nap1l1	% of Non-Farnesyl	% of Farnesyl
2x	<1%	<1%	<1%
3x	1%	1%	1%
4x	3%	2%	3%
5x	6%	6%	7%
6x	13%	12%	15%
7x	22%	21%	25%
8x	30%	31%	30%
9x	21%	23%	18%
10x	3%	4%	1%
11x	<1%	<1%	--

Variant	% Farnesyl	%Non-Farnesyl	% of Total Nap1l1
Nap1l1.S	74%	26%	22%
Nap1l1.L	74%	26%	19%
Nap1l1.LX2	--	100%	58%
Overall	31%	69%	--

The third acidic region was found to be glutamylated with a wide range of glutamylations, with the most abundant forms containing 6 to 9 additional glutamylations. It is thought that a majority of the glutamylations are mono-glutamylations with some poly-glutamylation. Only the X2 variant was detected by intact mass with 11 glutamylations; the X2 variant does not contain the last 10 residues of the protein. Finally, the populations of both Nap1l1.L and Nap1l1.S were approximately 75% farnesylated; however, the only 30% of whole population of Nap1l1 was farnesylated. The role of farnesylation is not known in Nap1l1. Furthermore, the specific mechanism by which glutamylation controls histone sequestration vs deposition is not clearly understood.

3.6 Conclusion

The eNap1L1 glutamylation for *Xenopus laevis* eggs has been characterized with more confidence than had ever previously been reported. Nap1l1.L was identified with up to 6 glutamylations across the first acidic region of the protein, most commonly with 2 or 3 additional monoglutamylations. Nap1L1.S was identified with up to 5 glutamylations in the first acidic region, but the most abundant forms contained 2 or 3 monoglutamylations. The third acidic region for Nap1l1.L and Nap1l1.S were identified with up to 10 glutamylations. Additionally, the peptides were detected both with and without C-terminal farnesylation. A third variant was identified as Nap1l1.LX2, which had the same sequence for Nap1l1.L up to a lysine to valine substitution at K383 and was not detected with the last 10 amino acid residues. The Nap1l1.LX2 variant was identified with a range of glutamylations from 1 to 11 PTMs in the third acidic region. It appears that many of these are monoglutamylations across the third acidic region.

There are many open questions still remaining. The work presented here gives a clear understanding of the extent of the glutamylation occurring in the first and third acidic regions of Nap1L. Understanding the functions of the modifications in each region and in conjunction with the unmodified second acidic region would present a logical next step. Further understanding of gamete development would additionally be a way to expand on the work presented here as Nap1l1 was isolated from eggs. Characterization of oocytes as an earlier time point might determine whether the modifications are increasing or decreasing. Further

questions remain regarding the role of farnesylation in Nap1l1. This work is the first reporting of glutamylation occurring simultaneously with farnesylation. Previous work evaluated either farnesylation or glutamylation independently but not concurrently. As Nap1l1 has been reported in other models to localize in the cytoplasm until chromatin assembly, exploring the differently modified Nap1l1 from cytoplasm versus nuclear extracts may point to roles for modifications.

3.7 References

1. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389(6648):251–260.
2. Talbert PB, Henikoff S (2017) Histone variants on the move: Substrates for chromatin dynamics. *Nat Rev Mol Cell Biol* 18(2):115–126.
3. Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 403(6765):41–45.
4. Foltz DR, et al. (2006) The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol* 8(5):458–469.
5. Guillemette B, et al. (2005) Variant histone H2A.z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol* 3(12):1–11.
6. Billon P, Côté J (2012) Precise deposition of histone H2A.Z in chromatin for genome expression and maintenance. *Biochim Biophys Acta - Gene Regul Mech* 1819(3–4):290–302.
7. Young TJ, Kirchmaier AL (2012) Cell cycle regulation of silent chromatin formation. *Biochim Biophys Acta - Gene Regul Mech* 1819(3–4):303–312.
8. Laskey RA, Honda BM, Mills AD, Finch JT (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* 275(5679):416–420.
9. Warren C, Shechter D (2017) Fly Fishing for Histones: Catch and Release by Histone Chaperone Intrinsically Disordered Regions and Acidic Stretches. *J Mol Biol* 429(16):2401–2426.
10. Shintomi K, et al. (2005) Nucleosome assembly protein-1 is a linker histone chaperone in Xenopus eggs. *Proc Natl Acad Sci U S A* 102(23):8210–5.
11. Ishimi Y, Yasuda H, Hirosumi J, Hanaoka F, Yamada M atsu (1983) A protein which facilitates assembly of nucleosome-iike structures In Vitro in mammalian cells. *J*

Biochem 94(3):735–744.

12. Ishimi Y, Kikuchi A (1991) Identification and molecular cloning of yeast homolog of nucleosome assembly protein I which facilitates nucleosome assembly in vitro. *J Biol Chem* 266(11):7025–7029.

13. Park YJ, Chodaparambil J V, Bao Y, McBryant SJ, Luger K (2005) Nucleosome assembly protein 1 exchanges histone H2A-H2B dimers and assists nucleosome sliding. *J Biol Chem* 280(3):1817–1825.

14. Park Y-J, Luger K (2006) Structure and function of nucleosome assembly proteins. *Biochem Cell Biol* 84(4):549–549.

15. Kepert JF, Mazurkiewicz J, Heuvelman GL, Tóth KF, Rippe K (2005) NAP1 modulates binding of linker histone H1 to chromatin and induces an extended chromatin fiber conformation. *J Biol Chem* 280(40):34063–34072.

16. Mosammaparast N, Ewart CS, Pemberton LF (2002) A role for nucleosome assembly protein 1 in the nuclear transport of histones H2A and H2B. *EMBO J* 21(23):6527–6538.

17. Steer WM, et al. (2003) Xenopus nucleosome assembly protein becomes tissue-restricted during development and can alter the expression of specific genes. *Mech Dev* 120(9):1045–1057.

18. Aguilar-Gurrieri C, et al. (2016) Structural evidence for Nap1-dependent H2A–H2B deposition and nucleosome assembly. *EMBO J* 35(13):1465–1482.

19. Andrews AJ, Chen X, Zevin A, Stargell LA, Luger K (2010) The Histone Chaperone Nap1 Promotes Nucleosome Assembly by Eliminating Nonnucleosomal Histone DNA Interactions. *Mol Cell* 37(6):834–842.

20. D'Arcy S, et al. (2013) Chaperone Nap1 Shields Histone Surfaces Used in a Nucleosome and Can Put H2A-H2B in an Unconventional Tetrameric Form. *Mol Cell* 51(5):662–677.

21. Ohtomo H, et al. (2016) C-terminal acidic domain of histone chaperone human NAP1 is an efficient binding assistant for histone H2A-H2B, but not H3-H4. *Genes to Cells* 21(3):252–263.

22. Miller KE, Heald R (2015) Glutamylation of Nap1 modulates histone H1 dynamics and chromosome condensation in Xenopus. *J Cell Biol* 209(2):211–220.

23. Regnard C, et al. (2000) Polyglutamylation of nucleosome assembly proteins. *J Biol Chem* 275(21):15969–15976.

24. Galichet A, Gruissem W (2006) Developmentally Controlled Farnesylation Modulates AtNAP1;1 Function in Cell Proliferation and Cell Expansion during Arabidopsis Leaf Development. *PLANT Physiol* 142(4):1412–1426.

25. Kho Y, et al. (2004) A tagging-via-substrate technology for detection and

proteomics of farnesylated proteins. *Proc Natl Acad Sci* 101(34):12479–12484.

26. Galichet A, Gruissem W (2003) Protein farnesylation in plants - Conserved mechanisms but different targets. *Curr Opin Plant Biol* 6(6):530–535.

27. Davies BSJ, Fong LG, Yang SH, Coffinier C, Young SG (2009) The Posttranslational Processing of Prelamin A and Disease. *Annu Rev Genomics Hum Genet* 10(1):153–174.

28. van Dijk J, et al. (2008) Polyglutamylation is a post-translational modification with a broad range of substrates. *J Biol Chem* 283(7):3915–22.

29. Eddé B, et al. (1990) Posttranslational glutamylation of α-tubulin. *Science (80-)* 247(4938):83–85.

30. Alexander JE, et al. (1991) Characterization of posttranslational modifications in neuron-specific class III beta-tubulin by mass spectrometry. *Proc Natl Acad Sci U S A* 88(11):4685–4689.

31. Onikubo T, et al. (2015) Developmentally regulated post-translational modification of nucleoplasmin controls histone sequestration and deposition. *Cell Rep* 10(10):1735–1748.

32. Rogowski K, et al. (2010) A family of protein-deglutamylating enzymes associated with neurodegeneration. *Cell* 143(4):564–578.

33. Redeker V (2010) Mass spectrometry analysis of C-terminal posttranslational modifications of tubulins. *Methods in Cell Biology* (Academic Press), pp 77–103.

34. Redeker V, Le Caer JP, Rossier J, Prome JC (1991) Structure of the polyglutamyl side chain posttranslationally added to α-tubulin. *J Biol Chem* 266(34):23461–23466.

35. Galperin MY, Koonin E V. (2008) A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Sci* 6(12):2639–2643.

36. Janke C, et al. (2005) Tubulin polyglutamylase enzymes are members of the TTL domain protein family. *Science (80-)* 308(5729):1758–1762.

37. Wolff A, Houdayer M, Chillet D, de Néchaud B, Denoulet P (1994) Structure of the polyglutamyl chain of tubulin: Occurrence of alpha and gamma linkages between glutamyl units revealed by monoreactive polyclonal antibodies. *Biol Cell* 81(1):11–16.

38. Spano AJ, Frankfurter A (2010) Characterization of Anti-β-tubulin antibodies. *Methods in Cell Biology* (Academic Press), pp 33–46.

39. Nambara E, McCourt P (1999) Protein farnesylation in plants: A greasy tale. *Curr Opin Plant Biol* 2(5):388–392.

40. Weber K, Plessmann U, Traub P (1989) Maturation of nuclear lamin A involves a

specific carboxy-terminal trimming, which removes the polyisoprenylation site from the precursor; implications for the structure of the nuclear lamina. *FEBS Lett* 257(2):411–414.

4. Detection of Low-Level Modifications on Truncated Nucleoplasmin by Parallel Ion Parking with Ion/Ion Proton Transfer Reactions (PIP-IIPT)

4.1 Abstract

Intact protein analysis is an emerging field, underdeveloped in comparison to proteomic analysis. The process of electrospray ionization, commonly used in LC-MS/MS, results in a wide distribution of charge states with low signal to noise ratios for intact proteins. This makes accurate determination of intact mass challenging and limits the detection of low abundance species, key when identifying post translational modifications. Nucleoplasmin (Npm) is a histone chaperone protein with a unique post translational modification of glutamylation. The constraints of intact protein analysis are highlighted with truncations of Npm, created to investigate the role of glutamylation. The Hunt lab implemented parallel ion parking with ion/ion proton transfer (PIP-IIPT) on a Thermo Orbitrap Elite™ instrument, enabling a wide range of charge states to be reduced and coalesce into a single, lower charge state. This technique was applied to the truncations of Npm and egg purified Npm and represents the first biological application of PIP-IIPT. Truncated species were identified by intact mass with multiple glutamylations at less than 10% relative abundance. Individual species of endogenously purified Npm were differentiated due to PIP-IIPT.

4.2 Introduction

Mass spectrometric analysis of an intact protein can yield unambiguous information about sequence and post translational modifications (PTMs). Analysis of the intact molecule allows for the simultaneous analysis of sequence variants, splice variants, or cross-talk of PTMs without the need for labeling or correlations drawn across multiple peptide fragments. Various research groups have been developing analytical methods for improving intact protein sequence analysis by mass spectrometry, including improving available ion current and novel fragmentation methods (1–4). However, sequence analysis of an intact protein can be challenging when evaluating complex mixtures of species. Without accurate intact mass,

the problem becomes even more challenging. Accurate determination of the intact mass for a protein is challenging because it requires high resolution and accurate detection of the isotopic envelope to predict the monoisotopic mass. Modifications, such as deamidation or single amino acid substitutions, differ by only a few daltons; accurate detection of intact mass plays a key role in sequence analysis.

Many of the recently analytical methods have been developed for Orbitrap instruments and therefore use electrospray ionization (ESI, discuss previously in 1.2.3) (1, 3, 5) ESI (which has "made elephants fly") enables analysis of intact proteins under denaturing conditions and is largely the source of one problem plaguing the field, low signal-to-noise (S/N) ratios (6, 7). As proteins are introduced into the mass spectrometer by ESI, a distribution of charge states can be observed, with larger proteins sampling a greater number of charge states (7). Individual charge states are made up of a collection of isotope peaks; each peak is the result of a specific isotopic distribution of atoms. Accurate measurement of the mass difference between isotope peaks allows for determination of that particular charge state of a protein. The measured isotopic distribution for a charge state of a protein can be compared to a theoretical distribution of isotopes, which is estimated from the chemical composition of the protein and relies on models to predict the distribution of isotope peaks for a particular charge state (7). Therefore, clear detection of the isotopic envelope is necessary to accurately determine the intact mass but can be difficult to obtain due to reasons explained below.

Low S/N ratios for proteins are the direct result of the division of available ion current among the greater number of charge states and an increase in the number of isotopes for larger molecular weight molecules (7). The ion current available for a high-resolution scan in an Orbitrap is limited by the capabilities of the trapping quadrupole prior to the mass analyzer (8). As an RF trap is filled to a predetermined number of charges (described in 1.2.6), the number of molecules captured in a RF trap is dictated by the net charge of the cation precursor. Therefore, the higher the number of charges on a protein, the fewer protein molecules are available for a particular scan. This is depicted in Figure 45. Thus, there are fewer available precursors for any given charge state and each isotope peak, which both increase in number as the protein increases in size. While detection of a single protein may

be fairly straightforward, co-eluting species yield complex mass spectra, with decreased sensitivity for low abundance species.

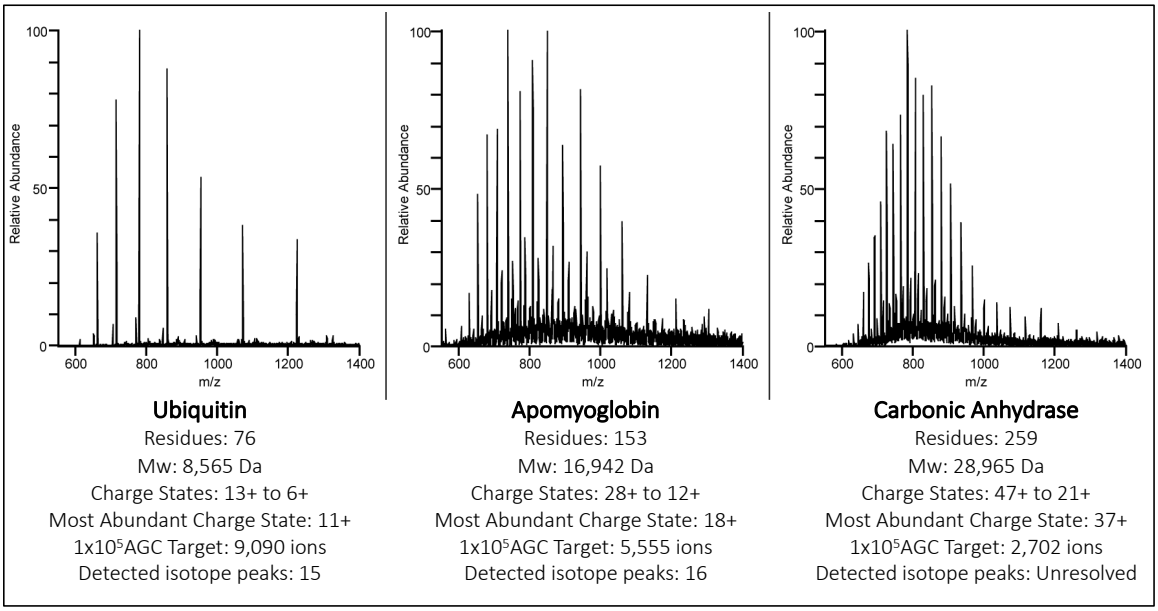

Figure 45. Comparison of ubiquitin, apo-myoglobin, and carbonic anhydrase

Co-isolation of precursors is another challenge with intact protein analysis. Protein chromatography has improved significantly with ultra-high-pressure liquid chromatography, though the field continues to chase more complex mixtures (9). However, because of the decreased S/N ratios described above, it is not always apparent that there are low-level species co-isolated with a given precursor. Furthermore, as charge states increase, the mass difference between precursors decreases. For example, a peptide at a 4+ charge state with an oxidized methionine (+16 Da) will have a mass difference of 4 m/z between the unmodified and modified precursors; alternatively, this will be 0.8 m/z for a protein at 20+ charge state. While an oxidized protein may chromatographically separate, any overlapping peaks will be co-isolated.

Finally, accurate charge state determination is a challenge for large, highly charged proteins in an Orbitrap mass analyzer (8). Increasingly higher charge states have increasing requirements for resolution. In an Orbitrap, the acquisition of the transient requires that the precursors pass between the two electrodes as distinct packets; higher resolution scans require longer acquisitions of the transient (8). Though the pressure in the Orbitrap is kept

low, $0.3x10^{-10}$ Torr, proteins will begin to de-phase at longer scan times as the molecules collide with the low level of gas in the mass analyzer (8). An example of this is presented in Figure 46 A and B, comparing single spectra between low and high resolution, respectively. The charge state distribution is not easily observed in the high resolution spectrum. This is problematic because fragmentation methods such as HCD and ETD can rely on charge state detection for improved fragmentation conditions (10, 11). Instrument methods frequently utilize charge state identification to identify appropriate precursors and/or determine whether to exclude or sample related charge states. In a mixed spectrum, the problem of charge state identification can be amplified for low abundance species.

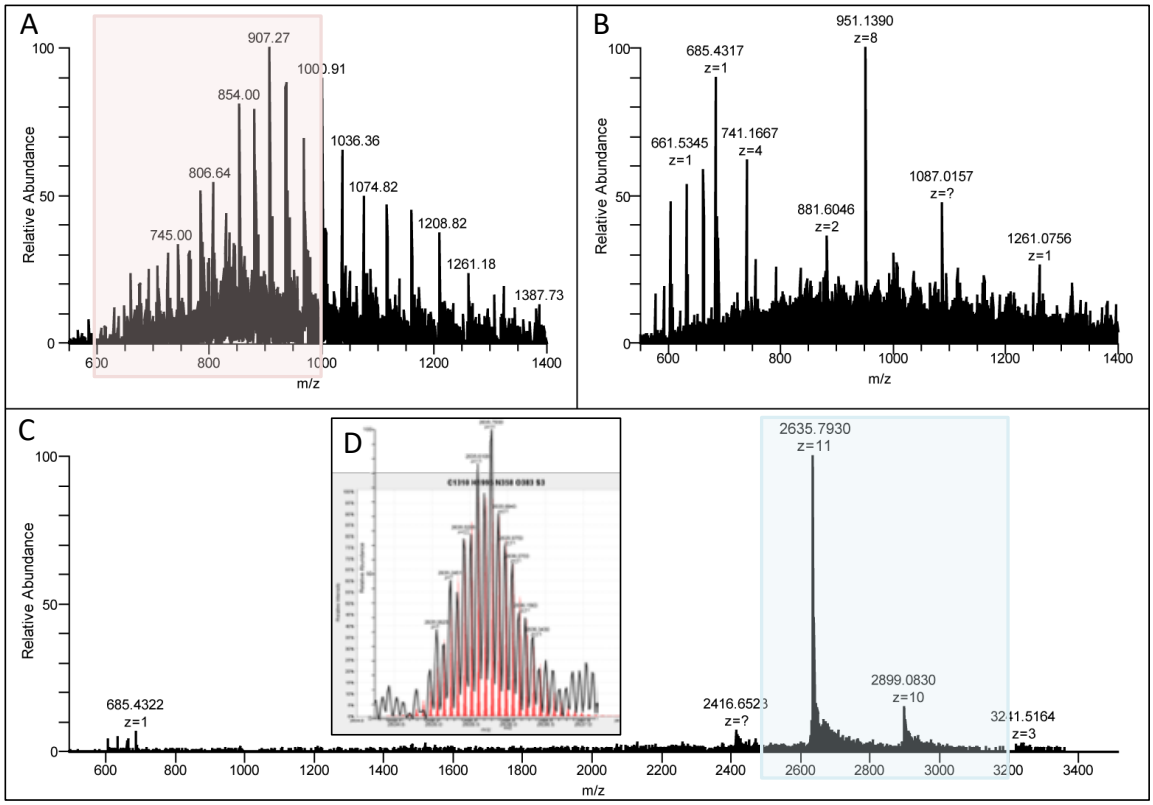

Figure 46. Carbonic Anhydrase A) a single low-resolution scan B) a single high-resolution scan at 120,000 resolution C) a single scan of the PIP-IIPT parked spectrum using a parking window at 2500 m/z with 5 mf and 200 ms IIPT D) Isotope distribution from C for the 11+ is expanded and overlaid on the theoretical isotope distribution. Note the isolation window is highlighted in pink in A and the mass range corresponding to the parking frequencies is highlighted in blue in C.

A solution to the complexity of determining the accurate intact mass of proteins is to simplify the spectra for the intact mass by reducing the charge of the protein and minimizing

the number of detected charge states. Toward this end, a combination of ion/ion proton transfer (IIPT) reactions with parallel ion parking (PIP) were developed in the Hunt laboratory for use on a Thermo Dual LIT-Orbitrap Elite™ tandem mass spectrometer; the implementation of the technology was the focus of Scott Ugrin's dissertation (12, 13). The goal of this novel method is to coalesce multiple charge states into a single, reduced charge state. Stevenson and Scott McLuckey identified that proton transfer reactions could be performed under RF trapping conditions between radical anion reagents and multiply charged protein cations (14, 15). For every ion/ion reaction, the net charge of the protein was reduced by a single charge. Today, the reaction is performed in the linear ion trap (LIT) by simultaneous trapping of multiply charged protein cations with a radical anion reagent, similar to ETD as described in Chapter 1. Decreasing the charge of the protein reduces the instrumental requirements for resolution and increases the mass to charge ratio of the product ion. When a multiply charged precursor undergoes an ion/ion reaction, either ETD or IIPT, the result is a distribution of final charge states. Therefore, IIPT alone is not sufficient to simplify the resulting spectra.

Parallel ion parking was a technique introduced by Scott McLuckey at Purdue as a way to control the kinetics of ion/ion reactions (16, 17). The goal of parallel ion parking during IIPT is to control the kinetics of ion/ion reactions such that there is only a single, final charge state. McLuckey and his group highlighted that the collisional cross section of a reaction is inversely proportional to the cube of the velocity of the ions (Eq. 1); by controlling the velocity, they could impact the rate of reaction (16). Equation 1 approximates the collisional cross section of the rate determining step, which is the formation of the ion/ion orbiting complex:

$$\boldsymbol{\sigma}_{ion/ion} = \boldsymbol{\pi}\left[\frac{\boldsymbol{z_1 z_2 e^2}}{\boldsymbol{4\pi\varepsilon_0 \mu v^2}}\right]^{\boldsymbol{2}} \quad (1)$$

where υ is the relative velocity, z_1 and z_2 are the charges of the ions, e is the charge of the electron, ε_0 is the vacuum permittivity constant, and μ is the reduced mass of the ion pair (17, 18). The velocity of ions in the LIT is commonly manipulated for events such as isolation, ejection, and CAD (see 1.2.9.1). Ions in a LIT at different m/z have unique frequencies of motion, or secular frequencies (see 1.2.5 and 1.2.6) (19). During events such as CAD, a

supplemental electric potential is applied in addition to the RF trapping potential (19, 20). The additional potential will resonantly, kinetically excite ions with a similar secular frequency, thereby selectively increasing that particular ion's velocity.

During PIP, a low level of supplemental activation is applied during an ion/ion reaction causing the ion to increase its velocity without fragmentation. The collisional cross section of the reaction decreases as per Eq. 1, and the rate of the reaction is slowed. During pip-IIPT, supplemental activation is applied to a range of frequencies corresponding to the m/z of the final product ions rather than a single frequency associated with the precursor (13, 17). While the precursor is not excited, the reagent is resonantly excited with a low level of supplemental activation, thereby slowing the kinetics of all reactions (13). The precursor undergoes IIPT reactions with the reagent until the product ions fall within the window of lower frequencies for the supplemental activation. The product ions are excited and increase their velocity, preventing further reactions and effectively "parking" the product ions into a single charge state. An example of a parked spectrum is presented in Figure 46: ions were isolated from the region in 2A from 600-1000 m/z and reacted with the IIPT reagent for 200 ms while a supplemental waveform was applied, depicted as the blue box in C. Ions reacted until their charge was reduced, and their m/z dependent secular frequencies matched the frequencies in the parking waveform. To simplify the above description, the goal of PIP-IIPT is to reduce the charge and number of charge states of a multiply charged precursor in order to improve precursor ion detection of intact proteins.

Consolidating the number of charge states simplifies a mass spectrum and enables the available ion current to be concentrated, thereby increasing the S/N ratio. This is exemplified in Figure 46 where there are half the number of protein molecules in C with $1.5x10^5$ charges than in B with $1.0x10^6$ charges. Furthermore, increasing the m/z ratio of intact proteins potentially allows for improved identification of modified species. To date, there has been a limited application of this technology (13). The goal of this chapter is to highlight the capabilities of PIP-IIPT towards the identification of low level, post-translationally modified truncations of Nucleoplasmin. At the end of the chapter, the opportunities of the method and some of the constraints of the technique are discussed.

Nucleoplasmin (Npm) was the first identified histone chaperone protein and is responsible for H2A/H2B histone dimer storage during *Xenopus Laevis* gamete development (21, 22). Npm is a heat stable protein that forms a homopentamer in biological conditions, with each copy of Npm binding a histone dimer (23). A hydrophobic core region of Npm folds together to form the pentamer; however, Npm is an acidic protein, containing three acidic regions that are important for binding the predominantly basic histone proteins (24). Previous work with Npm utilized peptide digestion to characterize the PTMs for Npm purified from both *Xenopus Laevis* oocytes and eggs (24). The peptide-based methods required tens of pmol of material per experiment for low level detection (24, 25). The pentamer is stable, and denaturation and digestion of the core is difficult. Furthermore, ideally mass spectrometry samples for analysis are desalted and acidified prior to analysis; Npm has an isoelectric point of approximately 4, potentially causing samples to precipitate out of solution (24).

Npm is abundantly modified with numerous PTMs by the stage of egg development, as portrayed in Figure 47 (23–25). Hyperphosphorylation was identified in both the amino- and carboxyl- termini regions of the protein. An arginine residue in the carboxyl-terminal tail region, R187, was identified to be dimethylated and was shown to be a substrate of the PRMT5-MEP50 complex. Additionally, the protein is glutamylated in the third acidic region during both the oocyte phase and the egg phase of development. Glutamylation is described in detail in the introduction to Chapter 3. Previous work from David Shechter's lab showed that phosphorylation occurs prior to glutamylation with electrostatic interactions playing a key role in histone binding (24, 26, 27). The PTMs for Npm are an example of the inter-connected relationships that modifications play in signaling and protein development and provide an excellent system to explore the capabilities in intact protein analysis by using PIP-IIPT.

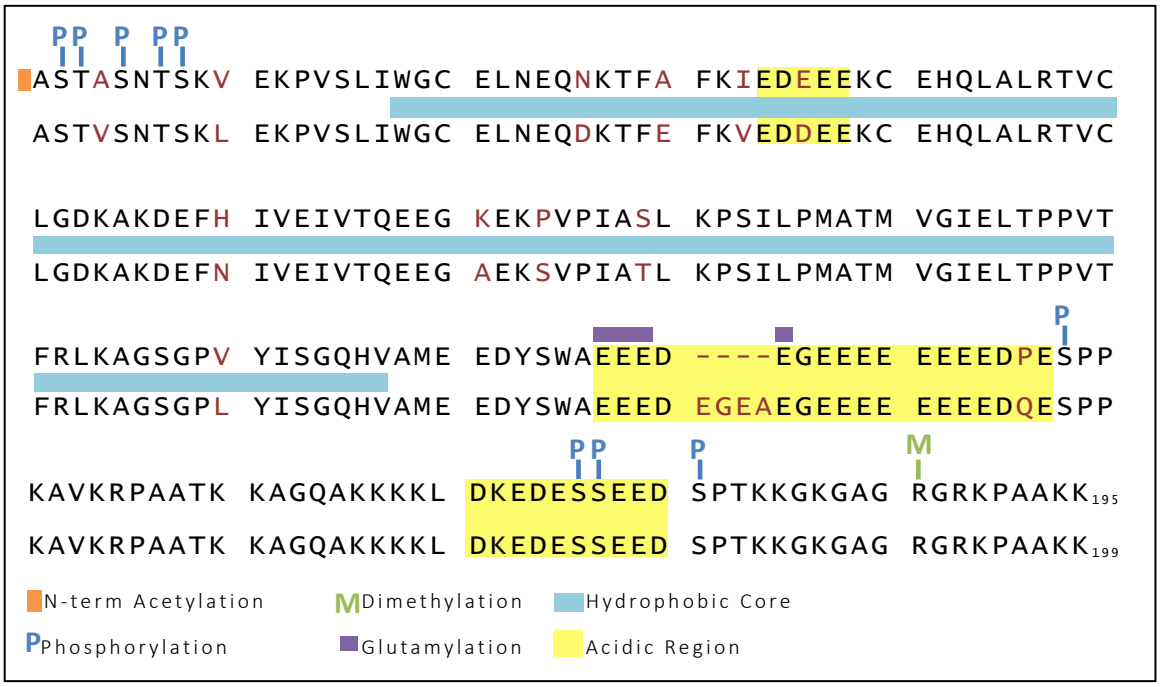

Figure 47. Nucleoplasmin sequence and modification, Npm2b and Npm2a presented, respectively

4.3 Materials and Instrumentation

Agilent Technologies (Palo Alto, CA)

1100 Series High Performance Liquid Chromatograph

1100 Series Vacuum Degasser

Poroshell 300SB-C18 resin, (5 μm, 300 Å)

Poroshell 300SB-C8 resin, (5 μm, 300 Å)

Honeywell (Morristown, NJ)

Burdick and Jackson Acetonitrile, LC-MS grade

Millipore (Billerica, MA)

Amicon Ultra-0.5 10K centrifugal filter

Molex (Lisle, IL)

Polymicro Technologies polyimide coated fused silica capillary, 360 μm x 75 μm i.d.

New Objective (Woburn, MA)

PicoClear™ Union Assembly

Oakwood Chemical (Estill, SC)

Perfluoromethyldecalin

Omega Engineering (Stamford, CT)

Temperature controllers

Parker Hannifin Corportation (Cleveland, OH)

Porter back pressure regulator

Porter mass flow controller

PQ Corporation (Valley Forge, PA)

Kasil – Potassium silicate solution

Sigma Aldrich (St. Louis, MO)

Apomyoglobin from equine skeletal muscle, protein sequencing standard

Carbonic Anhydrase

DL-Dithiothreitol, ≥98% by TLC

Fluoranthene, >99% Purity

Ubiquitin from bovine erythrocytes

2-propanol, LC-MS grade

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

Thermo Fisher Scientific (San Jose, CA; Bremen, Germany)

Formic Acid, LC-MS Grade

Orbitrap Elite™ mass spectrometer, custom modified with front-end ETD

Pierce Water, LC-MS Grade

4.4 Methods

Sample Preparation

All Npm (recombinant, egg purified, truncated recombinant) purification was designed and performed by David Shechter, Christopher Warren, and Benjamin Lorton at Albert Einstein School of Medicine in Bronx, NY as described previously (24, 26). Briefly, recombinant proteins were expressed in *E. coli* with His-tags and purified using Ni affinity and cation exchange chromatography. Five separate recombinant proteins are referenced in the results: rNpm (full length recombinant Npm); rNpmΔ10 (1-185 of Npm); rNpmΔ16 (1-179 of Npm); Core+A2 (1-145 of Npm); and SDII (phosphomimetic mutant containing Ser or Thr to

Asp mutations at 7 sites). *In vitro* glutamylation assays were performed by combining 2 ug Npm with 5 ug GST tagged TTLL4Δ526, 0.1 mM L-glutamate, and ATP and reacting overnight. Western blots were stained with rabbit TBIII anti-glutamylation antibody. To purify eNpm from *Xenopus laevis* eggs, laid eggs were harvested, lysed, and treated with protease and phosphatase inhibitors. The extract was heated and centrifuged; and the supernatant was purified with weak anion chromatography followed by hydrophobic interaction chromatography.

Chromatography

Samples provided by the Shechter lab were analyzed using RP-HPLC with in-house made columns. Fused silica columns (75 μm ID, 360 μm OD) containing a kasil frit, were packed with Poroshell 300SB packing material, either C8 or C18, 5 μm particles to 10-12 cm, and had a laser pulled emitter tip. If a pre-column was used, the fused silica column had a 100 μm ID, 360 μm OD with a kasil frit and was packed to 15 cm with similar packing material. Samples were diluted with water or cleaned up using a 10 kDa spin filter to remove salts. Columns were pressure loaded with approximately 500 fmol to 1 pmol protein and rinsed using 0.1% formic acid in water for approximately 20 column volumes. Proteins were separated by RP-HPLC using 0.1% formic acid in water for Solvent A and 0.1% formic acid in 60% Acetonitrile, 20% Isopropanol for Solvent B at an approximate flow rate of 100 nL/min. Proteins were ionized by electrospray and introduced to the mass spectrometer.

Mass Spectrometry

Samples were analyzed on an in-house modified Thermo Orbitrap Elite™ mass spectrometer. Modifications to the instrument and software code have been previously described (1, 13). The hardware modifications include changes to enable front-end reagent introduction as well as a delivery system for the introduction of perfluoromethyldecalin (PFMD) to the reagent ionization source. Software modifications were made previously to allow for IIPT, multiple c-trap fills, and parallel ion parking (1, 13). Briefly, multiple c-trap fills are performed to increase the ion current for detection after ion/ion reactions. Multiple, iterative sets of product ions from reactions in the LIT are collected in the C-trap prior to analysis in the Orbitrap, increasing the population of the product ions to directly increase the S/N ratio.

IIPT Parking

Samples were screened for chromatographic separation of intact species as well as for charge state distribution of precursors by taking high resolution MS^1 at 120,000 resolution FWHM at 400 m/z with an AGC target of $1x10^6$ charges as well as a low resolution, IT MS^1 scan with an AGC target of $1x10^5$ charges. PIP-IIPT spectra were acquired by isolating multiply charged precursor cations from either a narrow or wide isolation window in the LIT. IIPT reagent was separately isolated in the LIT and the ion populations were allowed to mix for a predefined amount of time. A supplemental parking waveform was applied during the ion/ion reaction to a range of frequencies corresponding to the reagent as well as the anticipated frequencies of the product ions. Ions reacted, undergoing multiple ion/ion proton transfer reactions, until the secular frequency of the charge-reduced product ions falls within the range of frequencies of the parking waveform. The product ions were then resonantly excited, thus slowing the rate of the ion/ion reaction. As PIP-IIPT was not completely efficient, the product ions were effectively "parked" into a few reduced charge states, see Figure 48. The parked MS^2 spectra were acquired with 120,000 resolution FWHM at 400 m/z, with an AGC target of $1x10^5$ charges for the precursor and $2x10^5$ charges for the reagent, and typically 5 multiple fills (mf).

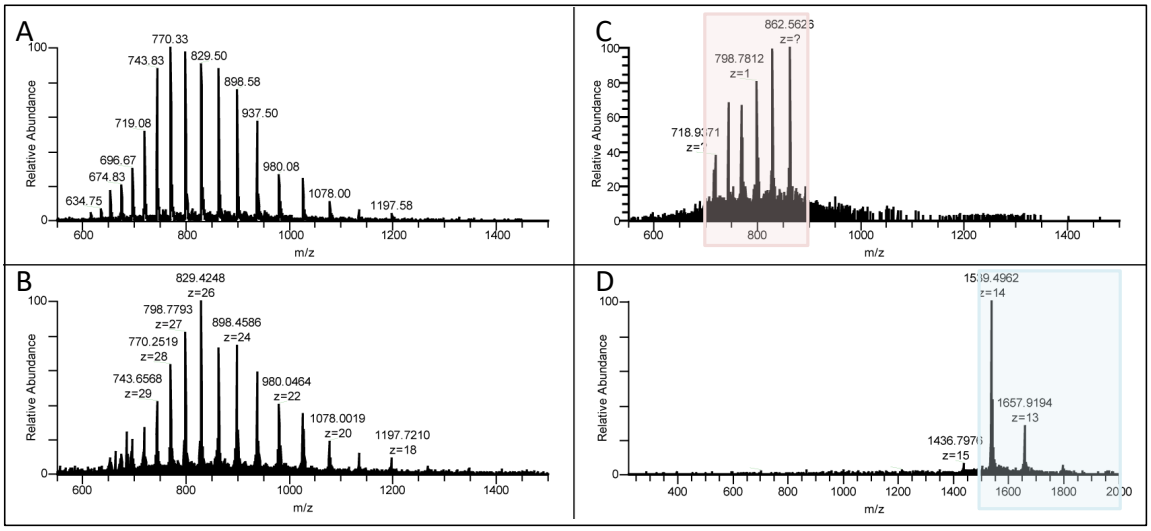

Figure 48. *rNpm depicted through a parking experiment A) Low resolution MS^1 scan B) Low resolution MS^1, 5 average scans C) High resolution scan of the isolation window in pink D) Single high resolution scan of PIP-IIPT spectrum, parking window shown in blue, 50 ms IIPT and 1 mf*

Analysis

Spectrum were manually analyzed or deconvoluted using Xcalibur™ software (Thermo Fisher Scientific version 4.0) and the embedded Xtract™ algorithm. Isotope Pattern Calculator™ from Pacific Northwest Laboratories was used to model theoretical isotope distributions based on chemical composition and charge state.

Appropriate precursors were identified from the MS^1 (vide supra) and targeted for ETD followed by IIPT using a 4 m/z isolation window and acquisition in the Orbitrap at 60,000 resolution, FWHM. Reaction times and numbers of multiple fills are noted below with each fragmentation spectrum. Spectra were manually validated utilizing in house developed software for predicted fragment masses.

4.5 Results

Nucleoplasmin is an ideal protein to highlight novel applications of parallel ion parking with ion/ion proton transfer (PIP-IIPT) in support of intact protein analysis. The protein is approximately 195 residues, or 22 kDa, of which 34 are basic residues, creating a high charge density protein, shown in Figure 49. It should be noted that a charge state of 32+ can be detected, though at 10% relative abundance (unlabeled at approximately 670 m/z in Figure 49.) The high charge density of the protein makes it an optimal candidate for fragmentation by ETD. The protein has 3 acidic regions and a hydrophobic core region with very few polar residues; half of the basic residues are in the last 55 residues of the protein. In the MS^2 spectrum presented in Figure 50 where ETD fragmentation was performed followed by IIPT, 50% sequence coverage (excluding proline residues) is obtained on a chromatographic timescale, coverage sufficient for a confident identification.

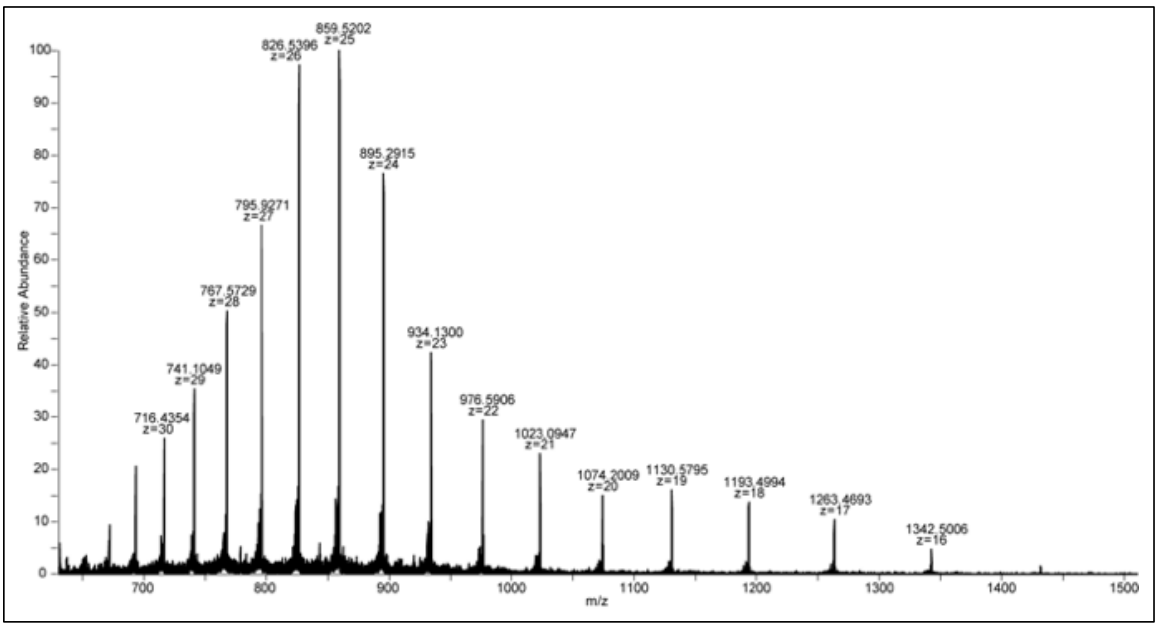

Figure 49. Full Charge State Distribution of rNpm

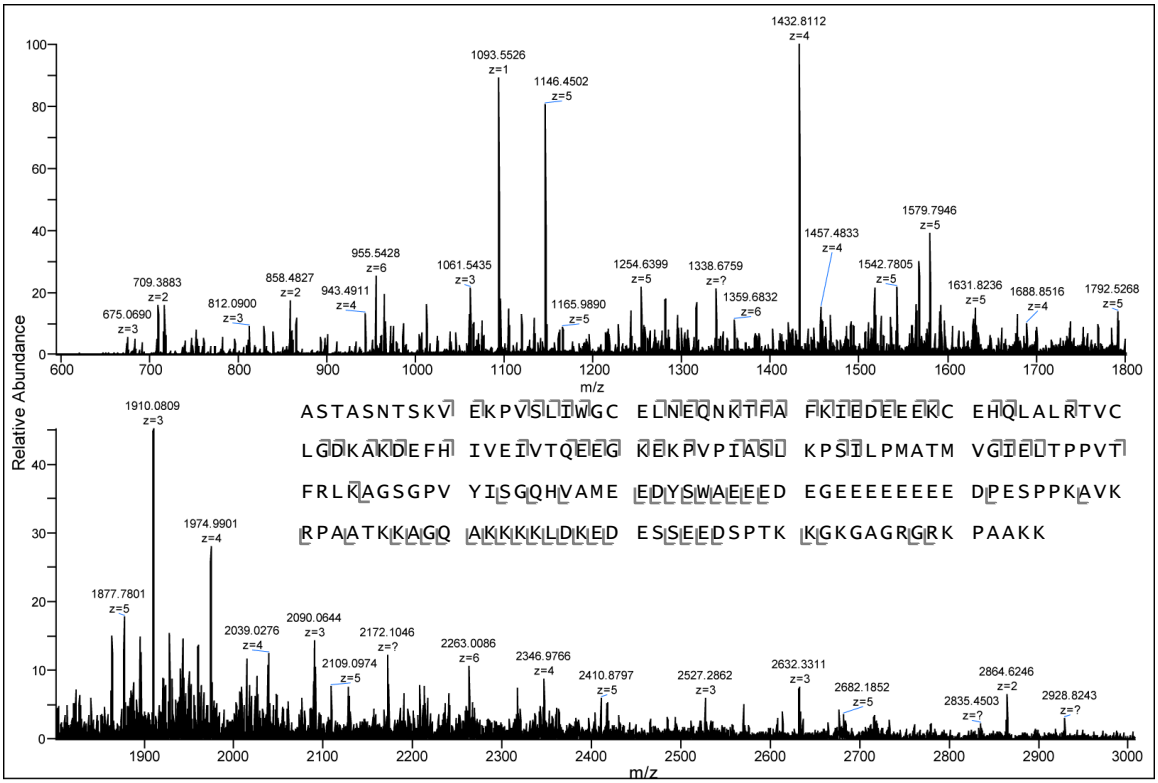

Figure 50. MS/MS Spectra of rNpm (716.50 m/z precursor (z=30), 5 ms ETD/20 ms IIPT and 5 multiple fills)

Previous work from David Shechter's group, presented in Figure 51a, showed that the second acidic region (see Figure 47 for sequence) was necessary for glutamylation by TTLL4; neither the full length recombinant nucleoplasmin (rNpm) nor the hydrophobic core were modified (24). This suggested that TTLL4 was modifying the second acidic region of the protein, but it was unknown if TTLL4 was modifying the recombinant protein on the same acidic residues as were found to be modified on egg purified Npm. Furthermore, the abundance of glutamylation or numbers of modifications could not be determined from the western blot.

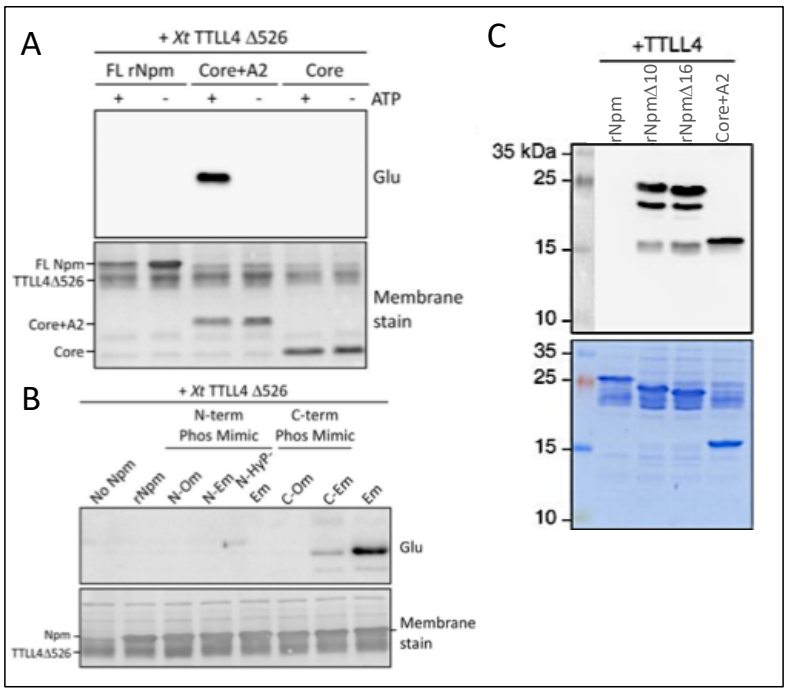

Figure 51. Western Blot of Glutamylated Truncations of rNpm A) *In vitro* glutamylation reaction of rNpm, Core+A2, and Core with TTLL4 B) C-Em and Em are phosphomimetic mutants with S to D substitutions at identified phosphorylation sites on the C-terminus and all identified phosphorylation sites, respectively C) Removal of the last 10 and 16 C-terminal residues of rNpm enable the protein to be glutamylated (24, 26)

The Core+A2 truncation contained the hydrophobic core as well as the second acidic region, residues 1-145, and was shown in Figure 51a to be glutamylated by TTLL4 *in vitro*. The glutamylated Core+A2 protein was analyzed intact, and the results are presented in Figure 52. It should be noted that the truncation was detected with a side product from β-mercaptoethanol (a component of the storage buffer) at C_{40}, an additional mass of 75.998Da

(H_4C_2OS). While approximately half of the basic residues were removed due to the truncation, the truncated protein still contained sufficient electron density for ETD with up to 19 charges. The Core+A2 is abundantly modified and was detected with up to four post translational glutamylations. It is estimated that approximately 60% of the Core+A2 was glutamylated by TTLL4. The monoglutamylated precursor was targeted for fragmentation using ETD and IIPT and is presented in Figure 53. The sequence coverage was 37% for the monoglutamylated species, however the glutamylation could not be site localized within the second acidic region. While the sequence contains 16 basic residues, R102 and K104 are the only two basic residues in the region from 85 – 145, leaving the C-terminus void of charge for any z● ions to be detected.

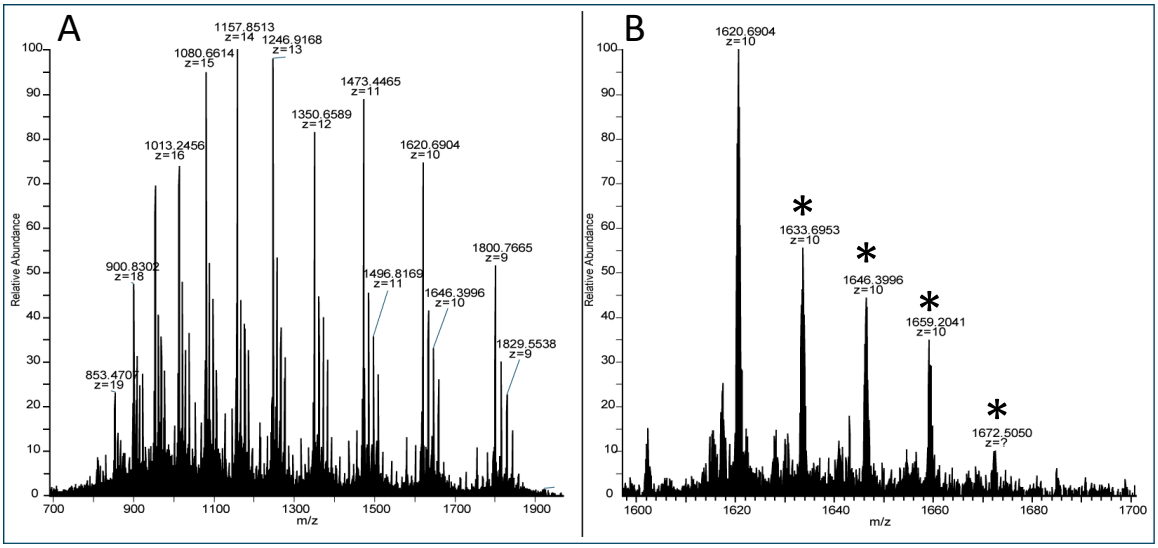

Figure 52. Glutamylated Core+A2 Truncation MS1; A) Full charge state distribution of Core+A2 B) Expanded view of +10 charge state showing by mass difference the non-glutamylated species to four glutamylations.

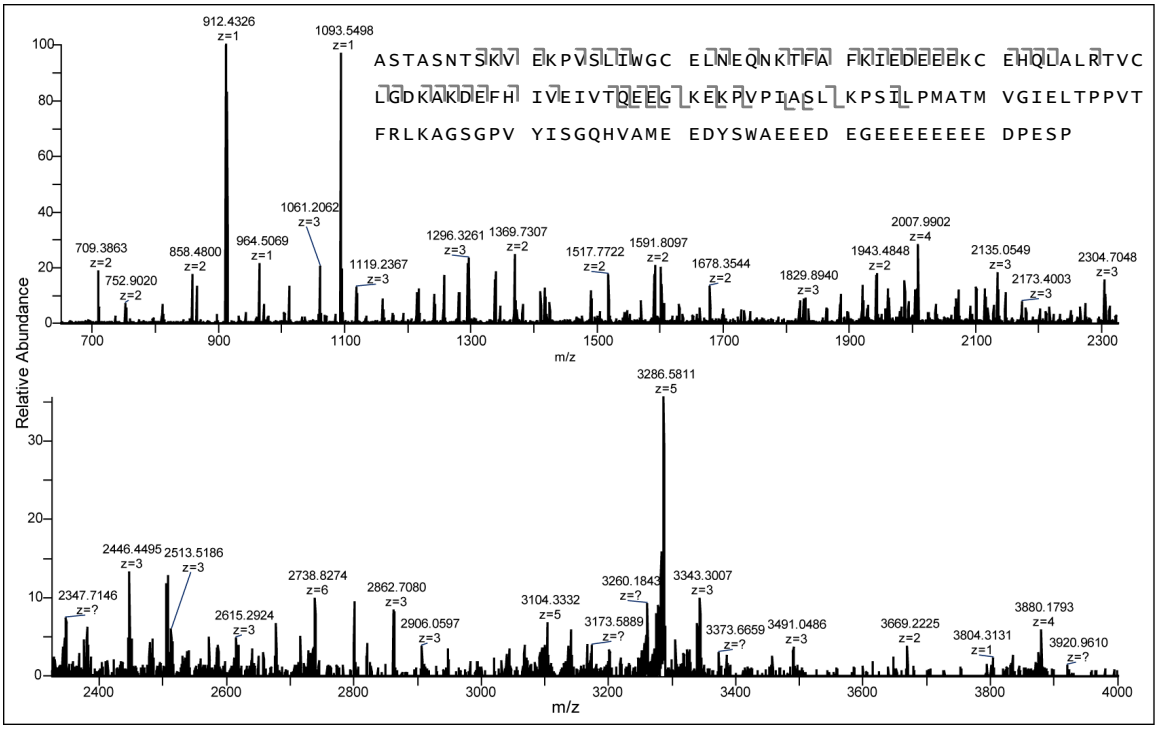

Figure 53. MS/MS Spectrum of 19+ mono-glutamylated Core+A2 (5 ms ETD/20 ms IIPT and 5 multiple fills)

Two additional truncations were created by the Shechter lab to further investigate the interaction between C-terminal basic residues and the glutamylation of the second acidic region. The first was a 10-residue deletion from the C-terminus (rNpmΔ10*)* and the second was a 16-residue deletion from the C-terminus (rNpmΔ16*)*. These extended past the second acidic region to include a series of basic residues followed by the third acidic region and were shown to be glutamylated by TTLL4Δ526, see Figure 51c. rNpmΔ10 contains three C-terminal basic residues (K180, K181, K183); rNpmΔ16 excludes these residues.

Figure 54 and Figure 55 present the results for rNpmΔ10, and Figure 56 presents the results for rNpmΔ16. The non-glutamylated species of rNpmΔ10 is the major species in Figure 54a. In the high resolution MS^1, no glutamylated species are detected as confirmed in the expanded spectrum presented in Figure 54b. A 200 m/z window of masses from Figure 54a, highlighted in blue, was isolated and contained seven charge states ranging from 30+ to 23+. The precursors were allowed to react with the IIPT reagent and were parked in the 7+ charge state, presented in Figure 54c. Figure 55 presents an expanded view with up to six

glutamylations detected on rNpmΔ10 after IIPT-parking. Charge states from 12+ to 8+ are shown in Figure 55 at low abundance and the 6+ charge state is detected at approximately 25% of the 7+ charge state. The ability to coalesce charge states into a single charge state within the parking window is not completely efficient. However, the multiply glutamylated species were detected with a S/N ratio of approximately 2, see inset in Figure 55, and were previously undetected in the MS^1 spectrum. The experiment was repeated with rNpmΔ16 and the results are presented comparing single scans for the MS^1 and the PIP-IIPT spectrum in Figure 56. In a single spectrum, the mono- and di-glutamylated species are unobservable in Figure 56b. Additionally, the intact average mass is detected at 902.35 m/z, slightly off from the theoretical average mass of 903.02 m/z for z=22. The parked unmodified products, shown in Figure 56c, have a S/N ratio of 70 with up to six glutamylated species clearly abundant; the diglutamylated species has a S/N ratio of 5. The parked spectrum simplifies the analysis of the intact mass detected at $[M+H]^{1+}$=19818.8190 Da; the theoretical monoisotopic is $[M+H]^{1+}$=19832.7696 Da. Sequencing of the unmodified species indicated that the mass shift was in the middle of the protein (data not shown); several amino acid substitutions vary by 14 Da. Multiply glutamylated species were detected for both truncations and have the potential to be selected for fragmentation and sequencing from the IIPT-parked spectra as an MS^3. However, the relative abundance of the mono-glutamylated species for either of the precursors was less than 15% compared to the unmodified species in the parked spectra. The available ion current is dictated by the capacity of the LIT from the initial precursor isolation; thus, the resulting fragmentation spectra would be significantly diminished in quality.

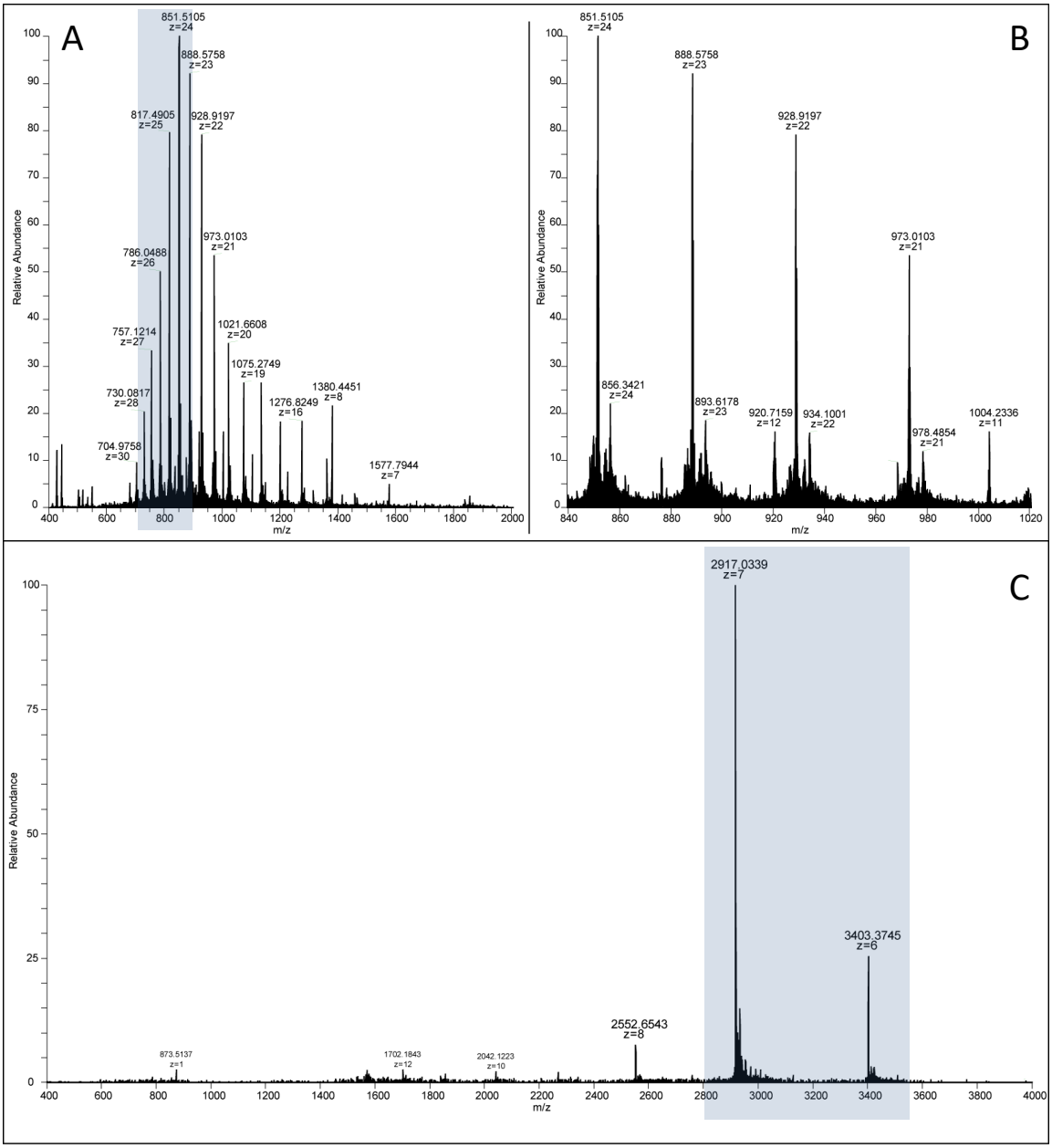

Figure 54. Glutamylated rNpmΔ10; A: Full MS^1 charge state distribution, isolation window highlighted with shaded window B: Expanded view of 840 to 1020 m/z; C: IIPT parked MS^2 (Isolated 800 m/z ± 100 m/z, 150 ms IIPT, shaded region shows parking window at 2800 m/z)

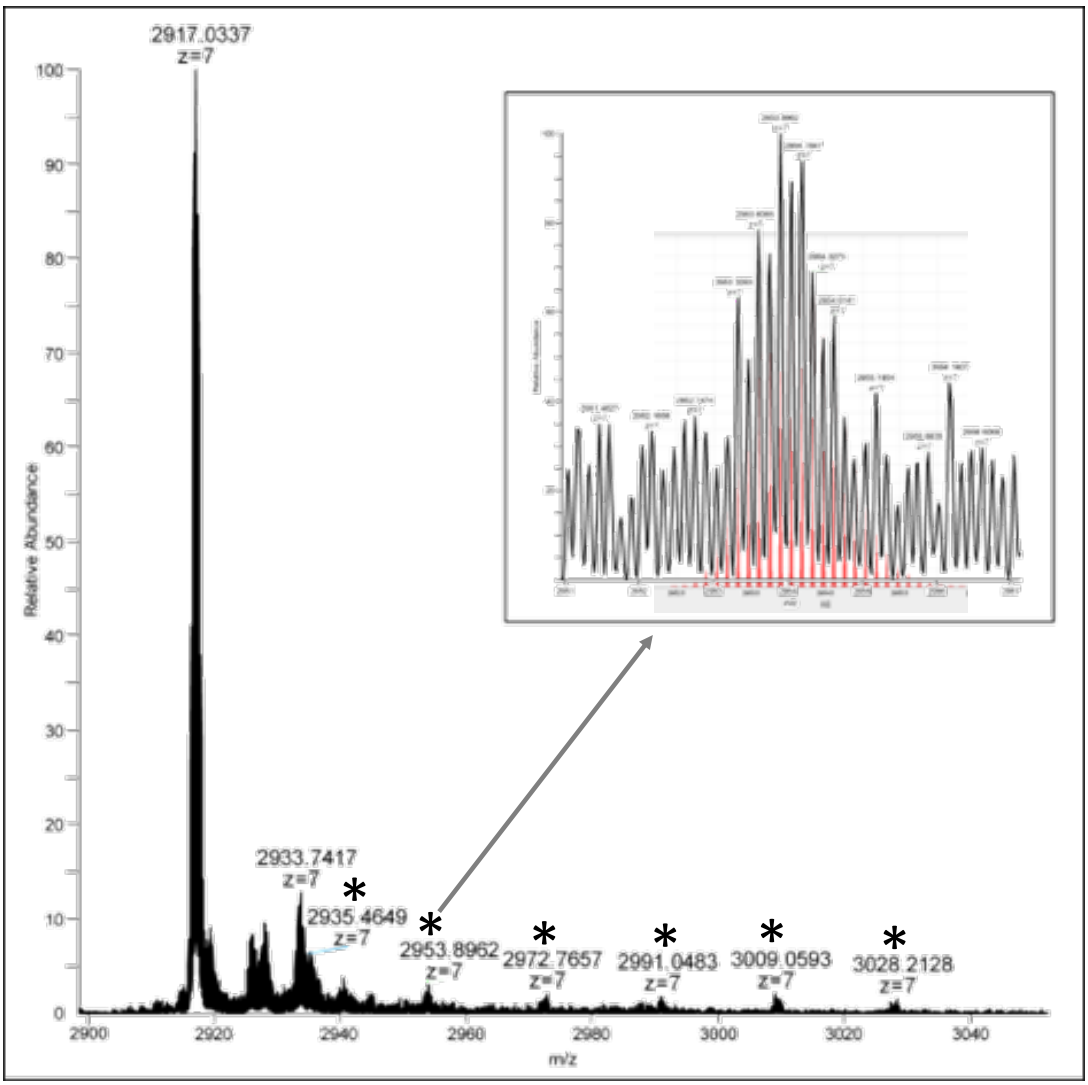

Figure 55. Expanded view of parked rNpmΔ10 at the 7+ charge state; Inset: isotope distribution of di-glutamylated rNpmΔ10 with the theoretical distribution in red.

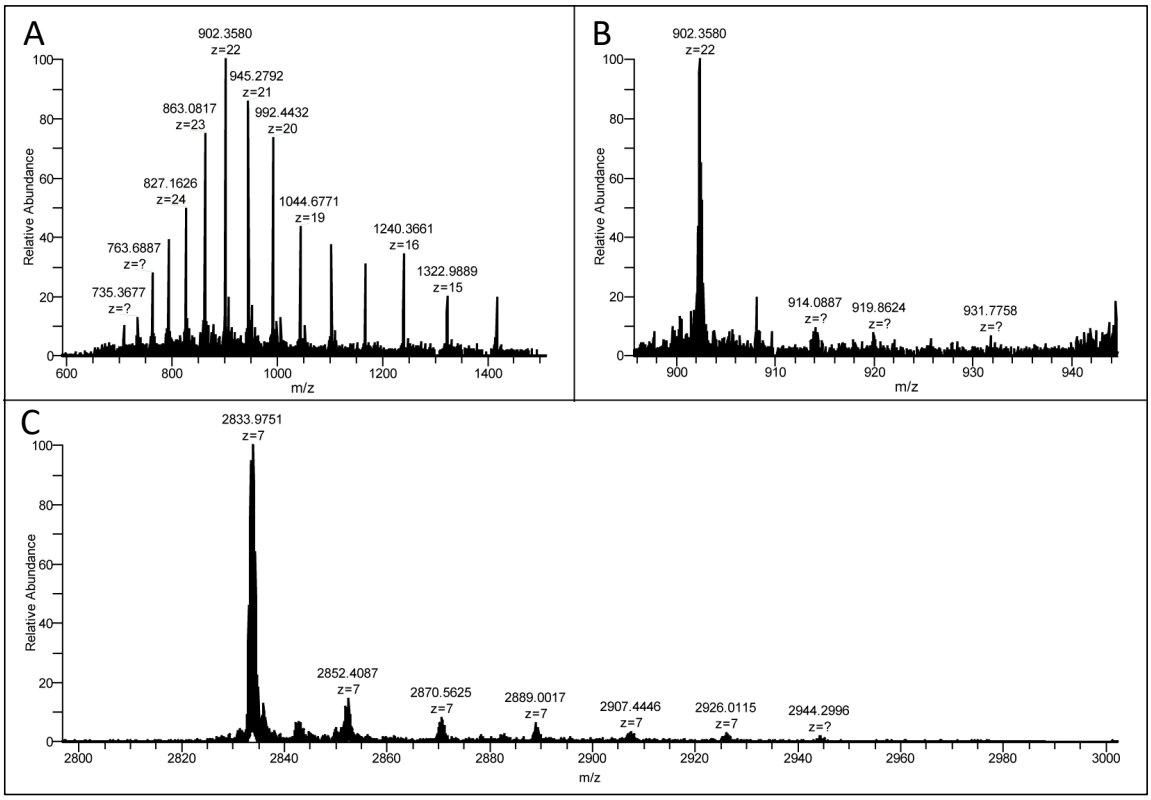

Figure 56. Glutamylated rNpmΔ16 A) MS^1 B) Expanded MS^1 charge state of 22+; C) IIPT Parked MS^2 (Isolated 800 m/z ± 100 m/z, 150 ms IIPT, parking window at 2800 m/z); theoretical average masses are 903.2 m/z for z=22 and 2835.89 m/z for z=7

Previous examples have discussed the detection of low level species. Highly complex spectra can be challenging to parse individual species because of the exponential mass difference between charge states from different molecular weight species. Complex spectra can be simplified by coalescing charge states into a single reduced charge state that makes use of a greater mass range. Figure 51C presents two distinct glutamylated bands in each lane for rNpmΔ16 and rNpmΔ10. The apparent secondary truncation is difficult to detect in a traditional MS^1 scan in Figure 57A, but is detected at approximately 10% relative abundance after deconvolution using the Xtract algorithm, shown in Figure 57C. By utilizing the PIP-IIPT technique, the truncation as well as two glutamylated species can be clearly detected in a single scan of the MS^2 as well as in the deconvoluted spectrum. The mass difference of the two truncations is 1460.73 Da; peptides from either extreme of the protein do not match this mass, suggesting it comes from an internal deletion. However, this highlights the challenges posed with intact mass analysis: any isolation for fragmentation in the MS^1 spectrum could

be a mixed isolation as the overlapping distributions will be spread across the mass range. Analysts traditionally rely on the MS^1 to evaluate potential unwanted species in the isolation window. With intact protein analysis, S/N ratios are already low and hampering the identification of clean isolation windows. PIP-IIPT directly improves the limit of detection.

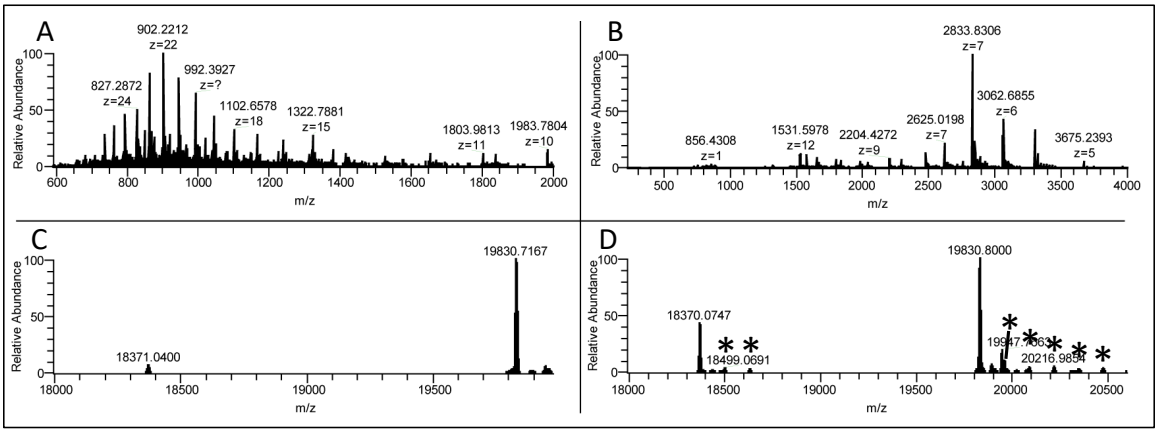

Figure 57. rNpmΔ16 mix spectra A) High resolution scan of mixed spectrum B) High resolution single scan of PIP-IIPT (Isolation 700-900 m/z, 150 ms IIPT, 5 mf) C) Deconvolution of spectrum in A, note truncation at 18371.04 Da D) Deconvolution of spectrum in B, note truncation at 18370.07 Da and two glutamylated species marked with *

A phosphomimetic mutant (SDII) was created to explore the role of phosphorylation in the N and C-terminal tails. The mutant contained seven aspartic acid mutations at known phosphorylation sites (S5, S6, T7, S144, S172, S173, S77). While rNpm was not glutamylated *in vitro* with TTLL4Δ526, see Figure 51b, SDII was glutamylated. Figure 51b includes two mutants labeled N-Em and C-Em, in which only the phosphorylated residues in the N-terminal tail and C-terminal tail were mutated, respectively. C-Em is able to be *in vitro* glutamylated. Though the N-Em mutant could not be modified, there was an increase in the glutamylation detected in the western blot from the C-Em mutant to the Em mutant, or SDII.

The SDII sample was confirmed to be glutamylated with up to three glutamylations by intact mass using PIP-IIPT as reported in Figure 58. It should be noted that the SDII species were detected with a side product from β-mercaptoethanol (a component of the storage buffer) at C_{40}, an additional mass of 75.998Da (H_4C_2OS). The mono-glutamylated species was targeted for ETD fragmentation, as reported in Figure 59. The modification could only be confirmed between residues 127 to 139, within the residues of the 2nd acidic region.

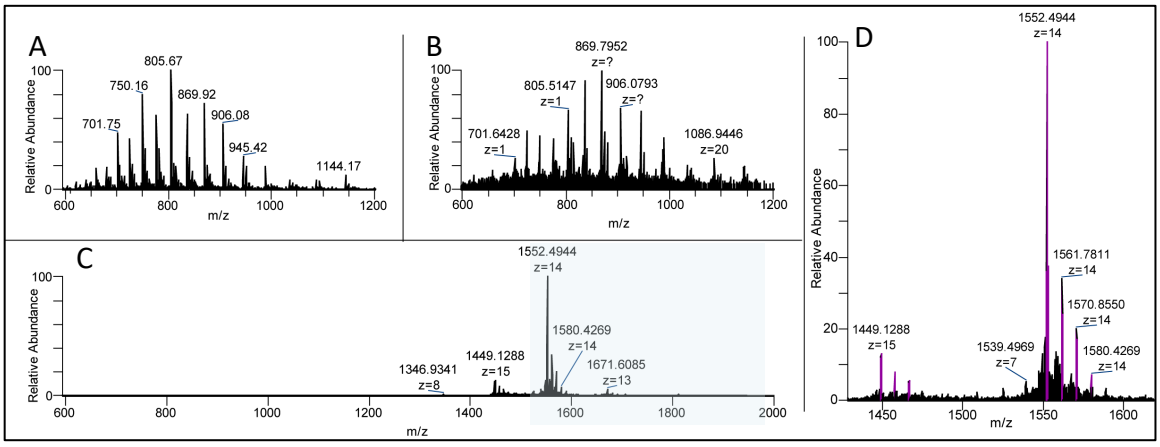

Figure 58. MS[1] of Glutamylated SDII A) Low resolution MS[1] B) High Resolution MS[1] C) PIP-IIPT MS[2] with parking window at 1510 m/z, 50 ms IIPT, and 10 mf D) Expanded PIP-IIPT MS[2] spectrum, purple denotes SDII species

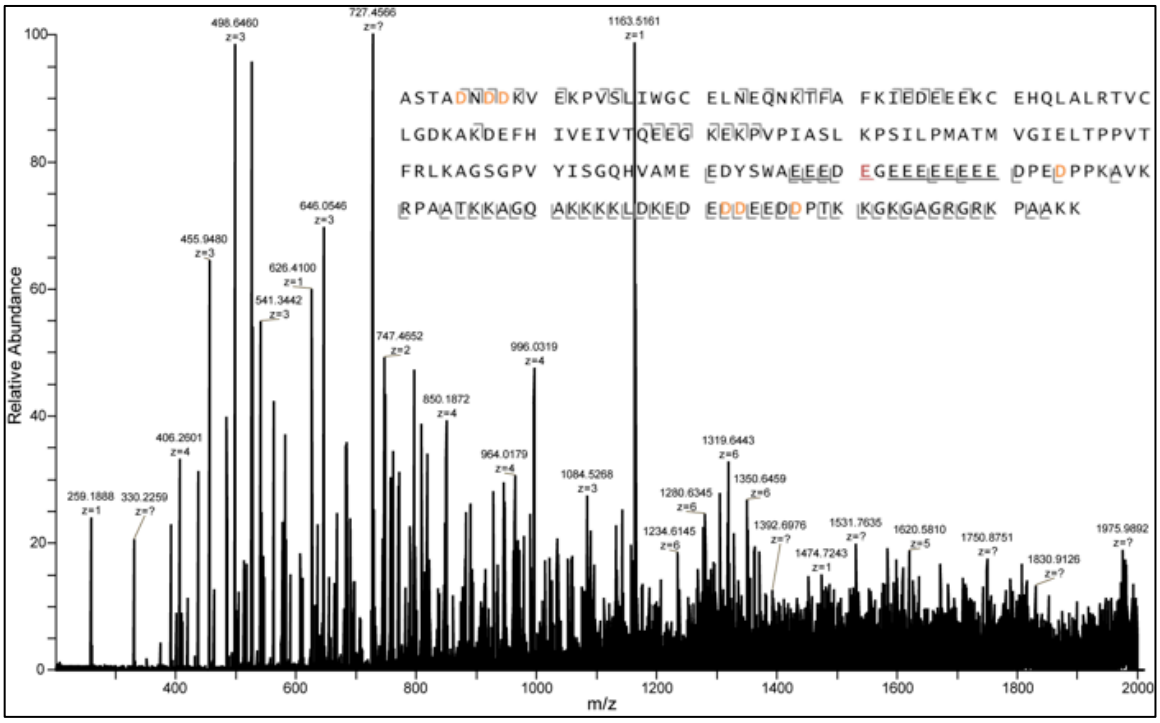

Figure 59. ETD MS/MS of Glutamylated SDII (727.25 m/z, 2 ms ETD, 11 ms IIPT, 5 mf); orange letters are aspartic acid substitutions at known phosphorylation sites

Endogenous Npm2 purified from *Xenopus Laevis* eggs (eNpm) is a richly modified protein. The characterization of PTMs for eNpm was previously reported in Onkinubo et al. 2015; eNpm was reported to have N-terminal acetylation, a dimethylated arginine, up to 8 relatively abundant phosphorylations, and up to 3 constitutive glutamylations (oocyte Npm

was reported with up to five glutamylations.) Furthermore, there are two known variants for Npm: Npm2a and Npm2b, which vary by 11 point-mutations and a four-residue insertion at 131 for Npm2a (see Figure 47 for sequences.) A representative full, unmodified MS^1 spectrum of eNpm is presented in Figure 60. The ion current is a representation of the complexity of various modification patterns, variants, and charge states all overlapping in an 800m/z wide region of the spectrum. While charge states are labeled and there is a periodic pattern to the ion current, an individual scan does not present this and individual species cannot be visually detected (data not shown).

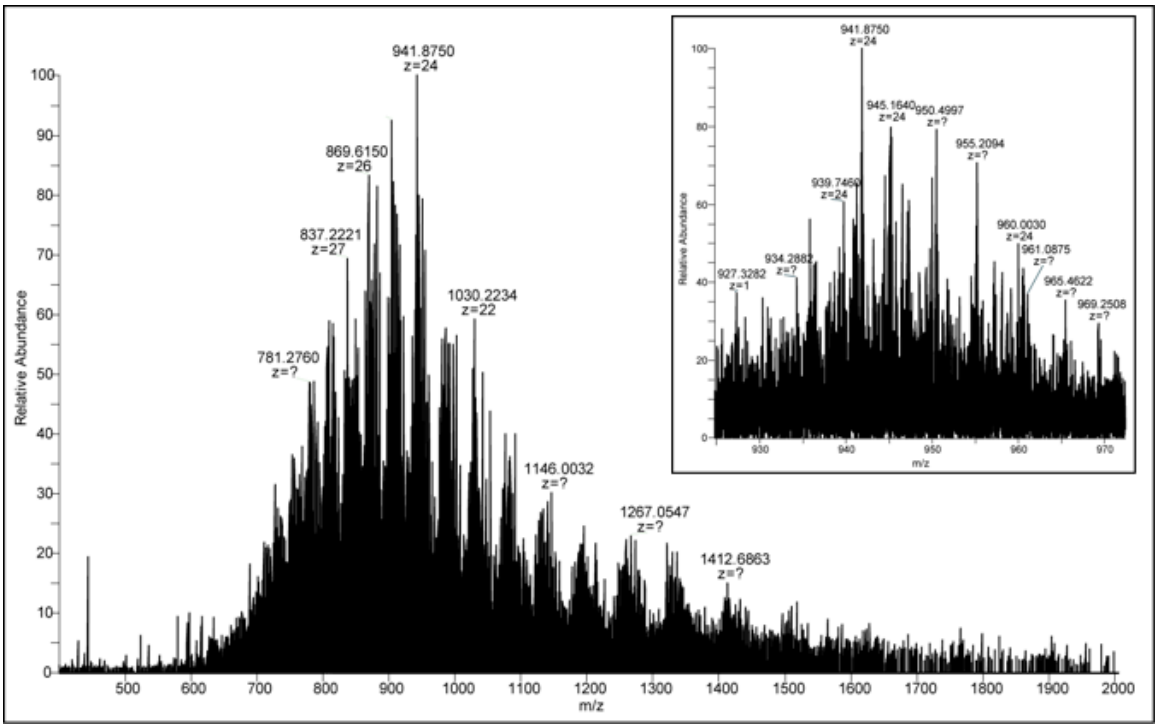

Figure 60. eNpm MS1 (5 averaged scans); Inset: expanded view of 920 – 975 m/z

The PIP-IIPT experiment was performed with the purified eNpm. The precursors were isolated from 800 to 1000 m/z and PIP-IIPT was performed with the parking window starting at 3100 m/z. The precursors were charge depleted by ion/ion proton transfer reactions until the precursors' m/z was within the parking window at 3100 m/z; the 7+ charge state was the first charge state within the window. The product ions were parked by resonant excitation and generally inhibited from further ion/ion reactions. A representative charge depleted spectrum is presented in Figure 61 with an expanded view of the 7+ charge state. In the

expanded view, four marked product ions have mass differences of 18.43 m/z or 129.04 Da, the mass of glutamylation. The intact mass is 21,915 Da for Npm2a or 21,464 Da for Npm2b, in their unmodified states (excluding the initial Met); peptide-based characterization methods identified 1355 Da of post translational modifications for Npm during gamete development. In Figure 61b, the average mass of the most abundant species can be explained by Npm2a with an acetylated N-terminus, a dimethylated arginine, seven phosphorylated residues, and four glutamylated modifications ($[M+7H]^{7+}$: 3293.4193 m/z; average mass: 3295.43 m/z), see isotope comparison in Figure 62.

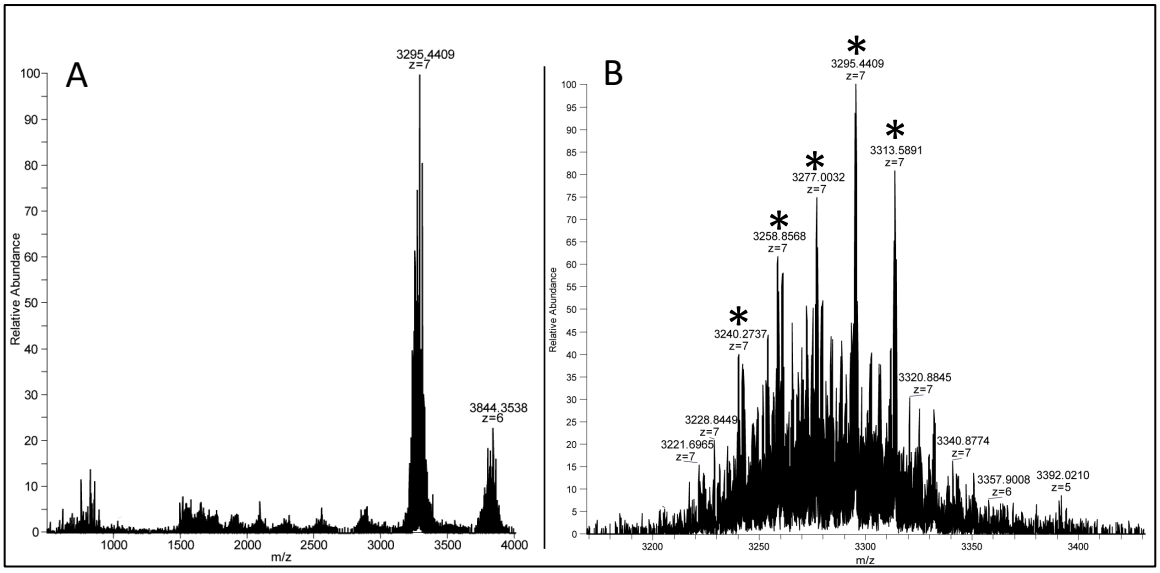

Figure 61. Parked Spectrum of eNpm; A: Full spectrum showing 7+ and 6+ charge states; B: Expanded view of 3150-3450 m/z from A, species marked by * are separated by 129.04 Da or 18.43 m/z for z=7+

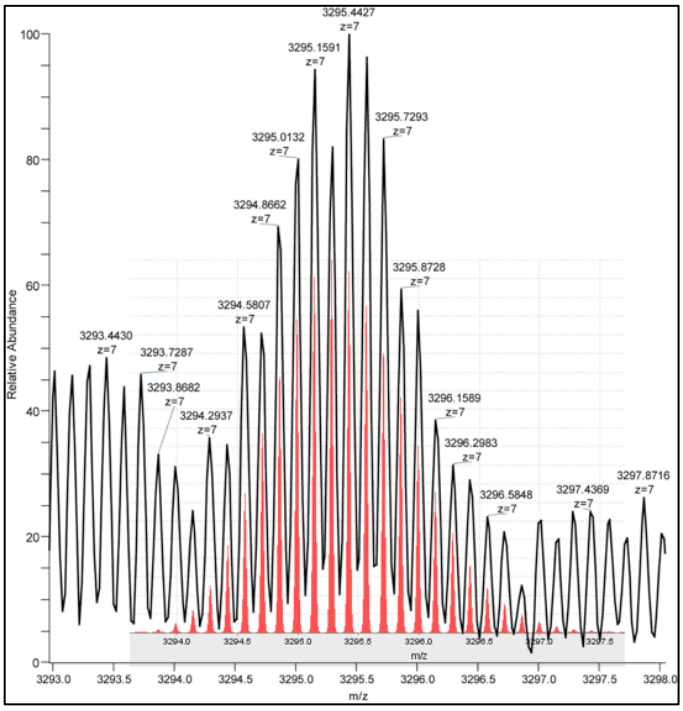

Figure 62. Isotopic comparison of theoretical precursor (red) with most abundant species from Figure 18b

4.6 Conclusions and Discussion

The glutamylation of Npm truncations by the glutamylase, TTLL4, was successfully evaluated by PIP-IIPT. Initial *in vitro* experiments to modify rNpm were unsuccessful. However, the Core+A2 trunctation, which is missing half of the basic residues and the short third acidic region, was abundantly glutamylated with at least four modifications. Unfortunately, it was not possible to site localize the modifications within the last 60 residues due to the lack of charge. The phosphomimetic mutant, SDII, was also modified by TTLL4. These two mutants point to potential functions of phosphorylated serine and threonine residues as well potential electrostatic interactions between the basic residues in the C-terminal tail and the second acidic region. The structure of the C-terminal tail is an intrinsically disordered region and is key to histone dimer binding. Shechter et al. hypothesized that the negative charge and bulk of the phosphorylated serine and threonine residues electrostatically repelled the acidic region, thereby allowing TTLL4 to modify the second acidic region. To explore the role of the basic residues, the rNpmΔ10 and rNpmΔ16 truncations were created. Both mutants were found to be glutamylated with up to 6

additional residues, though the ratio of modified to unmodified species was drastically reduced compared to the Core+A2. As Core+A2 was not investigated with PIP-IIPT, it is not known if the glutamylation extends to 6x modified species. Along with additional experiments, Warren et al. concluded that the last 10 residues protected the second acidic region by electrostatically binding the area (26). One interesting point to note is that while a majority of rNpmΔ10 and rNpmΔ16 species were unmodified, there were not drastic differences in the abundances between glutamylated species. Similar levels of a 6x modified protein to a 1x modified protein could suggest the potential of processive mechanism of modification by TTLL4.

The opportunities of PIP-IIPT for intact protein analysis were highlighted by the analysis of modified truncations of Npm. First, the spectra are simplified by parking species into only a few charge states. The complexity of multiple charge states overlaid from multiple species is reduced to only a few lower charge states. Reducing the charge state increases the delta mass between related species, improving the identification of precursors. Additionally, condensing the ion current into fewer charge states per precursor enables the analyst to more easily identify related species and improves the isotope distribution for accurate mass determination. This is highlighted with the PIP-IIPT experiment for the purified eNpm. In the unparked spectra, species could not be teased apart, isotopic distributions were unclear, and accurate intact mass could not be determined. With the variety of combinations of PTMs on variants and distribution across charge states and isotopes, it is possible that the isotope population for a given charge state was not significant enough for a detection of distributions; rather individual ions were being quantized (7). After the PIP-IIPT experiment, the species of eNpm at a 7+ charge state were easily differentiated, with isotopic distributions, a quintessential component of intact mass determination. Some of the complexity detected in the parked spectra could be due to limitations of the ion transmission efficiency of the instrument. Ion transmission has been improved with recent instrument design such as the Thermo Orbitrap Fusion™ Tribrid™ (28).

As important as simplifying the spectra is an improvement to the sensitivity of the method. In rNpmΔ10, none of the modified species could be detected in the unparked, MS^1 spectra; the available ion current was divided across a range of charge states for each

species. After parking, the ion current was condensed into predominantly a single, lower charge state; the modified species were dispersed over a wider mass range; and the isotopic distribution was detected for the modified species with a S/N ratio of 3. While the ion current for the parked spectra is less than the ion current for the unparked spectra, the available current per charge state is higher in the parked spectra.

The true opportunity of the PIP-IIPT method is the ability to use it on a chromatographic timescale. Yet constraints of the current method pose issues when applying it to a biologically relevant sample. There are several key components that hinder the method from being more broadly applied, which are: extended scan cycle times, a lack of 'on-the-fly' deconvolution programs, and the need to experimentally determined instrument parameters. Currently, a single ion/ion reaction on the Thermo Orbitrap Elite™ takes approximately 100 ms in addition to the ion/ion reaction time or the time needed to fill the LIT with the precursors. A single PIP-IIPT scan with 5 mf is acquired in 1 second. Implementation of PIP-IIPT on the Fusion™ instrument would provide a faster cycle time due to the use of quadrupole isolation for the precursor and reagent and the ability to parallelize the operation of the instrument (28). The ability to deconvolute the parked spectrum as the instrument is operating would open up method development opportunities. The instrument could modify isolation windows for MS^3 experiments, focusing its isolation to an abundant precursor at a higher charge state. Finally, all instrument parameters are experimentally determined with a protein of similar mass and charge prior to working with a running sample. More thorough characterization of the PIP-IIPT kinetics would aid in better understanding and a more robust implementation of the method.

4.7 References

1. Earley L, et al. (2013) Front-End Electron Transfer Dissociation: A New Ionization Source. *Anal Chem* 85(17):8385–8390.
2. Riley NM, Westphall MS, Hebert AS, Coon JJ (2017) Implementation of Activated Ion Electron Transfer Dissociation on a Quadrupole-Orbitrap-Linear Ion Trap Hybrid Mass Spectrometer. *Anal Chem*:acs.analchem.7b00213.
3. Shaw JB, et al. (2013) Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J Am Chem Soc* 135(34):12646–12651.

4. Weisbrod CR, et al. (2017) Front-End Electron Transfer Dissociation Coupled to a 21 Tesla FT-ICR Mass Spectrometer for Intact Protein Sequence Analysis. *J Am Soc Mass Spectrom* 28(9):1787–1795.

5. Riley NM, Westphall MS, Coon JJ (2017) Activated Ion-Electron Transfer Dissociation Enables Comprehensive Top-Down Protein Fragmentation. *J Proteome Res* 16(7):2653–2659.

6. Fenn JB (2002) Electrospray ionization mass spectrometry: How it all began. *J Biomol Tech* 13(3):101–118.

7. Compton PD, Zamdborg L, Thomas PM, Kelleher NL (2011) On the scalability and requirements of whole protein mass spectrometry. *Anal Chem* 83(17):6868–6874.

8. Makarov A, Denisov E, Lange O (2009) Performance Evaluation of a High-field Orbitrap Mass Analyzer. *J Am Soc Mass Spectrom* 20(8):1391–1396.

9. Riley NM, et al. (2018) The Value of Activated Ion Electron Transfer Dissociation for High-Throughput Top-Down Characterization of Intact Proteins. *Anal Chem* 90(14):8553–8560.

10. Olsen J V, et al. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* 4(9):709–712.

11. Rose CM, et al. (2015) A Calibration Routine for Efficient ETD in Large-Scale Proteomics. *J Am Soc Mass Spectrom* 26(11):1848–1857.

12. Michalski A, et al. (2012) Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes. *Mol Cell Proteomics* 11(3):O111.013698.

13. Ugrin S (2017) Ion-Ion Chemistry and Parallel Ion Parking for Advances in Mass Spectrometric Analysis of Intact Proteins. Dissertation (University of Virginia, Charlottesville, VA). doi:10.18130/V3Z62T.

14. Stephenson JL, McLuckey SA (1996) Ion/ion reactions in the gas phase: Proton transfer reactions involving multiply-charged proteins. *J Am Chem Soc* 118(31):7390–7397.

15. Stephenson JL, McLuckey SA (1998) Simplification of Product Ion Spectra Derived from Multiply Charged Parent Ions via Ion/Ion Chemistry. *Anal Chem* 70(17):3533–3544.

16. McLuckey SA, Reid GE, Wells JM (2002) Ion parking during ion/ion reactions in electrodynamic ion traps. *Anal Chem* 74(2):336–346.

17. Chrisman PA, Pitteri SJ, McLuckey SA (2006) Parallel ion parking of protein mixtures. *Anal Chem* 78(1):310–316.

18. Chrisman PA, Pitteri SJ, McLuckey SA (2005) Parallel ion parking: Improving conversion of parents to first-generation products in electron transfer dissociation. *Anal Chem* 77(10):3411–3414.

19. Schwartz JC, Senko MW, Syka JEP (2002) A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* 13(6):659–669.

20. March RE (1997) An introduction to quadrupole ion trap mass spectrometry. *J Mass Spectrom* 32(4):351–369.

21. Laskey RA, Honda BM, Mills AD, Finch JT (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* 275(5679):416–420.

22. Laskey RA, Earnshaw WC (1980) Nucleosome assembly. *Nature* 286(5775):763–767.

23. Taneva SG, et al. (2009) A Mechanism for Histone Chaperoning Activity of Nucleoplasmin: Thermodynamic and Structural Models. *J Mol Biol* 393(2):448–463.

24. Onikubo T, et al. (2015) Developmentally regulated post-translational modification of nucleoplasmin controls histone sequestration and deposition. *Cell Rep* 10(10):1735–1748.

25. Bañuelos S, et al. (2007) Phosphorylation of both nucleoplasmin domains is required for activation of its chromatin decondensation activity. *J Biol Chem* 282(29):21213–21221.

26. Warren C, et al. (2017) Dynamic intramolecular regulation of the histone chaperone nucleoplasmin controls histone binding and release. *Nat Commun* 8(1):2215.

27. Warren C, Shechter D (2017) Fly Fishing for Histones: Catch and Release by Histone Chaperone Intrinsically Disordered Regions and Acidic Stretches. *J Mol Biol* 429(16):2401–2426.

28. Senko MW, et al. (2013) Novel parallelized quadrupole/linear ion trap/orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Anal Chem* 85(24):11710–11714.