Analyses of Intact Proteins Using Novel Mass Spectrometric Techniques

Lissa Caitlin Anderson Woodbridge, Virginia

B.S. Chemistry, The College of William and Mary, 2009

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 August, 2014 Dedicated to my family, and to Jeff and Don.

i

Acknowledgements

First and foremost I have to thank my advisors, Dr. Donald Hunt, and Dr. Jeffrey Shabanowitz, for giving me the opportunity to work in the Hunt laboratory. I treasure my experiences and the knowledge that I have gained from working under their guidance. I am so grateful for the mentorship and friendship that they both bestowed upon me in my stay here at UVA.

I owe a large amount of my success to several former Hunt lab members, graduate students and post-docs. In particular, I'd like to thank Dr. Michelle English and Dr. John Syka for their many technical lessons, advice, and participation as members of the team that made the experiments presented in this dissertation possible. I must also thank the current members of the Hunt lab, especially Drs. Dina Bai and Weihan Wang, for their day-to-day contributions, enlightening discussions and support.

I cannot understate the importance of the contributions made by several collaborators in multiple disciplines during my time as a graduate student. It's been a pleasure working with Mr. Lee Early, Drs. Christopher Mullen, Benjamin Garcia, Ian Macara, David Shechter, Barney Bishop, and Monique van Hoek and the many post-docs, graduate students and lab technicians from their respective labs. I have had the good fortune of participating in several projects, some of which are not presented in this dissertation, and I owe a great deal to my collaborators for their willingness to participate and for the knowledge that they bring to the table.

I have been lucky to make several close friends, many of whom are fellow members of the chemistry department, in my time in Charlottesville. I thank them for all the wonderful experiences we've shared, and for their continued friendship and support. I appreciate the contributions from my committee members, who collectively comprised my candidacy and Ph.D. defense committees. I also thank them for their time spent reading this dissertation.

Finally, I must thank my family, who knew all along of what I was capable, even when I didn't believe it for myself.

Table of Contents

Table of Contents	iv
List of Figures and Tables	vii
Abbreviations	x
Abstract	xv
Chapter 1	1
1.1 Introduction	1
1.2 Mass Spectrometry—Principles and Instrumentation	3
1.2.1 Ionization Methods	3
1.2.2 Mass Analyzers	4
1.2.3 Tandem Mass Spectrometry (MS/MS)	11
1.2.4 Bottom-up, Middle-down and Top-down Proteomics	16
1.3 References	23
Chapter 2	23
2.1 Abstract	23
2.2 Introduction	24
2.2.1 A Brief History	24
2.2.2 Signal-to-noise Ratios in Top-down Mass Spectrometry	26
2.2.3 Spectral Complexity and Incomplete Fragmentation	34
2.3 Materials & Instrumentation	39
2.4 Methods	41
2.4.1 Instrument Modification and Operation	41
2.4.2 Gas-phase Ion/Ion Reactions	45
2.4.3 Sample Preparation	46
2.4.4 MS Analysis	47

2.5 Results & Discussion49
2.5.1 Multiple Fills Increase Ion Current of Fragment Ions
2.5.2 Gas-phase Ion/Ion Reactions Yield Near Complete Sequence Coverage of Apomyoglobin54
2.5.3 Protein Derivatization to Facilitate Sequence Analysis at the N- and C-termini of Proteins61
2.6 Conclusions
2.7 References
Chapter 371
3.1 Abstract71
3.2 Introduction
3.2.1 Histone Proteins, Chromatin and Epigenetics
3.2.2 Histone Post-translational Modifications
3.2.3 Histone Variants
3.2.4 Controlled Histone Proteolysis80
3.2.5 Why Top-Down Mass Spectrometry For Histones?
3.3 Materials and Instrumentation83
3.4 Methods
3.4.1 Histone Extraction From HeLa S3 Cells
3.4.2 Off-line Histone Fractionation
3.4.3 On-line RP-HPLC for Mass spectrometry85
<i>3.4.4 MS Analyses</i>
3.5 Results
3.5.1 Initial Experiments and MS Methods Determination with Fraction H2A2

3.5.2 Highlights from Other Histone Fractions	98
Fraction H1	
Fraction H2A1	102
Fraction H2B	
3.6 Conclusions and Discussion	113
3.7 References	117

List of Figures and Tables

Chapter 1

Figure 1.1 Stability diagram showing combined axial and radial stability withi quadrupole ion trap	n a . 6
Figure 1.2 Basic design of a two-dimentional linear ion trap (QLT)	.7
Figure 1.3 QLT voltages for the simultaneous storage of positive and negative i	ons . 8
Figure 1.4 Cutaway view of an Orbitrap mass analyzer	10
Figure 1.5 CAD produced b- and y-type fragment ions	14
Figure 1.6 ETD reaction mechanism	16

Chapter 2

Figure 2.1: Full MS1 sprectra showing the charge state distributions of VIP and apomyoglobin
Figure 2.2: Cartoon depiction of how charge distribution in precursors affects fragment ion charge following ETD
Figure 2.3 Fragment ion charge states are determined by the location of charge neutralization during ETD
Figure 2.4 Isotopic distributions of VIP and apomyoglobin
Figure 2.5 Ion trap charge capacity limits the number of molecules analyzed 32
Figure 2.6 ETD MS/MS spectra using 5 and 40 ms reaction times
Figure 2.7 Multiple electron transfers create internal fragment ions
Figure 2.8 Schematic of reagent gas delivery to the reagent ion source
Figure 2.9 SolidWorks rendering of an early reagent ion source for the SRIG interface color
Figure 2.10 Image of the current front-end reagent ion source
Figure 2.11 Mechanism for the reaction of sulfur hexafluoride with highly charged ETD-produced fragment ions
Figure 2.12 QLT voltages during ion/ion reactions46

Figure 2.13 Instrument diagram of the Orbitrap Velos Pro™ 50
Figure 2.14 Multiple Fills Increase Ion Current of Fragment Ions51
Figure 2.15 Fragment ion current as a function of the number of C-trap fills analyzed 53
Figure 2.16 Elapsed scan time comparisons for the analysis of 10 ion volumes 54
Figure 2.17 ETD/IIPT MS/MS spectra recorded on intact apomyoglobin56
Figure 2.18 Fragment ions observed with long and short IIPT reaction times57
Figure 2.19 Apomyoglobin sequence coverage determined by analysis of ETD and ETD/IIPT MS/MS spectra with ProSight PC [™] software 58
Figure 2.20 Heat map of the sequence coverage generated from intact apomyoglobin using sequential ETD/IIPT reactions60
Figure 2.21 Derivatization of human CLIP peptide with histamine enhances fragmentation observed by ETD
Figure 2.22 Protein modification strategies for observing extreme N- and C-terminal fragment ions and for enhancing fragmentation by ETD63
Figure 2.23 Sequence coverage of the N- and C-terminus of apomyoglobin following derivatization
Figure 2.24 Electrospray ionization mass spectra of unmodified and modified apomyoglobin

Chapter 3

Figure 3.1 X-ray crystal structures of the nucleosome	3
Figure 3.2 Some known N-terminal PTMs of nucleosomal histones74	4
Figure 3.3 Post translational modifications change the chemical properties of amir acid residues	10 6
Figure 3.4 Aligned sequences of some of the known isoforms of histone H2A79	9
Figure 3.5 Peptides resulting from proteolytic digest of H3 isoforms using GluC 82	2
Figure 3.6 Elution profile of histones purified from butyrate treated HeLa cells 89	9
Figure 3.7 Total ion current (TIC) and base peak chromatograms obtained by LC MS analysis of fraction H2A2 from butyrate treated HeLa cells	C- D
Figure 3.8 Averaged ESI full MS spectrum from "peak 3"92	2

Figure 3.9 ETD/IIPT MS/MS spectra of intact H2A.193
Figure 3.10 ESI full MS spectra of intact H4 and ETD/IIPT MS/MS spectra of the +8 and +15 charge states94
Figure 3.11 Narrowed data dependent selection scan ranges capture single charge states of different protein precursors
Figure 3.12 Full MS spectrum indicating the unique proteoforms of intact H2A identified by MS/MS
Figure 3.13 Sequence alignments of histone H2A isoforms identified in fraction H2A2
Figure 3.14 ADDT affords the opportunity to obtain optimized MS/MS spectra 100
Figure 3.15 ADDT allows for the selection of optimal isolation windows
Figure 3.16 Fraction H2A1 contains several peptidic species
Figure 3.17 Averaged ESI full MS spectrum depicting intact H2A species and their respective amino acid sequences
Figure 3.18 Averaged ESI full MS spectrum depicting intact H2B species and their respective amino acid sequences
Figure 3.19 Fraction H2B contains highly abundant truncated histones 108
Table 3.1 Unique proteoforms observed in histone fractions purified from butyrate treated HeLa cells and confirmed by MS/MS 110
Table 3.2 N-terminal truncated histones observed in H2A1 fraction from butyrate treated HeLa cells 111-112

Х

Abbreviations

αΝ	alpha amino group (N-terminus)
μ	micro (1x10 ⁻⁶)
AA	amino acid(s)
Ac	acetyl
ADDT	active data-dependent targeting
ADP	adenosine diphosphate
AEBSF	4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride
AGC	automated gain control
A, Ala	alanine
API	atmospheric pressure ionization
C, Cys	cysteine
°C	Celsius degrees
¹³ C	carbon-13
C18	octadecylsilane
CAD	collision-activated dissociation
CENP-A	variant of histone H3
CLIP	class II-associated invariant chain peptide
CSD	charge state distribution
D, Asp	aspartic acid
Da	Dalton(s)
dc	direct current
DNA	deoxyribonucleic acid
DTSSP	3,3'-dithiobis[sulfosuccinimidylpropionate]
DTT	dithiothreitol

E, Glu	glutamic acid
ECD	electron capture dissociation
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
ESI	electrospray ionization
ET	nondissociative electron transfer
ETD	electron transfer dissociation
f	femto (1x10 ⁻¹⁵)
F, Phe	phenylalanine
FT	Fourier transform or high resolution
FT-ICR	Fourier transform-ion cyclotron resonance
FTMS ⁿ	high resolution (orbitrap) analysis of MS/MS products
g	gravitational constant
g	gram(s)
G, Gly	glycine
G, Gly Gcn5	glycine gene coding for a histone acetyltransferase
G, Gly Gcn5 H, His	glycine gene coding for a histone acetyltransferase histidine
G, Gly Gcn5 H, His HAT	glycine gene coding for a histone acetyltransferase histidine histone acetyltransferases
G, Gly Gcn5 H, His HAT HDAC	glycine gene coding for a histone acetyltransferase histidine histone acetyltransferases histone deacetylases
G, Gly Gcn5 H, His HAT HDAC HCI	glycine gene coding for a histone acetyltransferase histidine histone acetyltransferases histone deacetylases hydrochloric acid
G, Gly Gcn5 H, His HAT HDAC HCI HP1	glycine gene coding for a histone acetyltransferase histidine histone acetyltransferases histone deacetylases hydrochloric acid heterochromatin protein 1
G, Gly Gcn5 H, His HAT HDAC HCI HP1 HPLC	glycine gene coding for a histone acetyltransferase histidine histone acetyltransferases histone deacetylases hydrochloric acid heterochromatin protein 1 high performance liquid chromatography
G, Gly Gcn5 H, His HAT HDAC HCI HP1 HPLC I, IIe	glycine gene coding for a histone acetyltransferase histidine histone acetyltransferases histone deacetylases hydrochloric acid heterochromatin protein 1 high performance liquid chromatography isoleucine
G, Gly Gcn5 H, His HAT HDAC HCI HP1 HPLC I, IIe i.d.	glycine gene coding for a histone acetyltransferase histidine histone acetyltransferases histone deacetylases hydrochloric acid heterochromatin protein 1 high performance liquid chromatography isoleucine inner diameter
G, Gly Gcn5 H, His HAT HDAC HCI HP1 HPLC I, IIe i.d. IgG1	glycine gene coding for a histone acetyltransferase histidine histone acetyltransferases histone deacetylases hydrochloric acid heterochromatin protein 1 high performance liquid chromatography isoleucine inner diameter
G, Gly Gcn5 H, His HAT HDAC HCI HP1 HPLC I, IIe i.d. IgG1 IIPT	glycine gene coding for a histone acetyltransferase histidine histone acetyltransferases histone deacetylases hydrochloric acid heterochromatin protein 1 high performance liquid chromatography isoleucine inner diameter Immunoglobulin G (subclass 1) ion/ion proton transfer

ITCL	ion trap control language or instrument control software
k	kilo (1x10 ³)
K, Lys	lysine
KDM	lysine demethylases
KMT	lysine methyltransferase
L	liter(s)
L, Leu	leucine
LC	liquid chromatography
LTQ	linear trap quadrupole (Thermo Scientific QLT)
m	milli (1x10 ⁻³)
m	meter(s)
М	molar
M, Met	methionine
MA	mass area
MALDI	matrix-assisted laser desorption/ionization
min	minute(s)
mol	mole(s)
mRNA	messenger ribonucleic acid
ms	millisecond(s)
MS	mass spectrometry
MS1	full MS spectrum
MS/MS, MS2, MS ⁿ	tandem mass spectrometry or tandem mass spectrum
MW	molecular weight
m/z	mass-to-charge ratio
n	nano (1x10 ⁻⁹)

Ν	normality
N, Asn	asparagines
NaCl	sodium chloride
o.d.	outer diameter
O-GlcNAc	O-linked N-acetylglucosamine
р	pico (1x10 ⁻¹²)
P, Pro	proline
Phos	phosphorylation
ppm	parts per million
PRMT	protein arginine methyltransferase
PTM	post translational modification
Q, GIn	glutamine
QIT	quadrupole ion trap
QLT	quadrupole linear trap
r	resolving power or resolution
R, Arg	arginine
RCC1	regulator of chromatin condensation 1
rf	radio frequency
RNA	ribonucleic acid
S, Ser	serine
sccm	standard cubic centimeters per minute
SF_6	sulfur hexafluoride
SNR	signal-to-noise ratio
SRIG	stacked ring ion guide
Su(var)3-9	suppressor of variegation 3-9, a lysine methyltransferase
T, Thr	threonine

- TIC total ion current
- TOF time of flight
- UV ultraviolet
- UVPD ultraviolet photodissociation
- V volt(s)
- V, Val valine
- VIP vasoactive intestinal peptide
- v/v volume-to-volume
- W, Trp tryptophan
- Y, Tyr tyrosine

Abstract

The regulation of nearly every cellular process is directly linked to the sequence and structure of the proteins involved. The study of proteins requires comprehensive knowledge of protein primary structure. This information cannot be derived from the genome alone, as the translation of mRNA into protein does not fully dictate the composition of the final protein complement. Alternative splicing, single nucleotide polymorphisms and the vast number of ways in which a protein can be post-translationally modified expand the number of possible gene products to over 1,000,000. Each has the potential to affect different cellular outcomes. Mass spectrometry (MS) has proven to be an invaluable tool in the study of protein structure and function. The wealth of information that MS is capable of generating, including amino acid sequence and post-translational modification (PTM) identification, makes it a premier method for analysis of complex biological samples.

The ultimate MS-based proteomics platform would be capable of unequivocally distinguishing closely related species from one another while concurrently characterizing any PTMs. To do this, the protein must be analyzed intact, as it exists in the system from which it is derived. However, this has proven an elusive endeavor as slow spectral acquisition rates, low signal-to-noise ratios and the complexity of the data generated in tandem mass spectrometry (MS/MS) experiments of whole proteins complicate the analysis. Herein, we describe new MS technologies and methodologies that facilitate the acquisition of high-quality MS/MS spectra capable of providing near-complete amino-acid sequence coverage of intact proteins on a chromatographic time-scale. The utility of our new approaches is demonstrated via analysis of intact histones derived from HeLa cells. We hope that the methods described will help guide the design of future instrumentation and data acquisition software for similar studies, as the knowledge

potentially gained from such studies could revolutionize our understanding of numerous biological processes.

1 Introduction to the Dissertation – A Description of Mass Spectrometry for the Analysis of Proteins

1.1 Introduction

Proteins are involved in nearly every aspect of cellular function and their characterization is key to understanding their roles in biological systems. The study of proteins is the foundation upon which advances in a number of biological disciplines rest and from which a relatively new discipline has emerged: proteomics. Proteomics deals with the identification and characterization of the final protein complement expressed by a genome (1). Since translation of mRNA into protein does not always yield the final protein product, this endeavor is complicated by the presence of multiple forms of the same protein in varying amounts at different times. This is the result of alternative splicing and variable and reversible post-translational chemical modifications of amino acid side chains and amino and carboxy termini of proteins. These post-translational modifications (PTMs), which include phosphorylation, acetylation, methylation, O-linked N-acetylglucosamine (O-GlcNAc), and over 200 others, are widespread and required for normal cellular function, as they have been shown to affect protein folding, alter protein activity and protein-protein interactions, as well as localize protein to various sites in the cell (2-5). The wealth of information that can be gleaned from the study of proteins and their PTMs has implications for the understanding of numerous diseases and may offer valuable therapeutic tools for the treatment of debilitating health conditions.

Despite major advances in the field of proteomics, studies probing the biological functions of proteins are subject to many analytical challenges owing to the low abundance of many of these proteins and to the complexity of the samples analyzed. Additionally, since the diversity of the proteome is rooted in the primary structural level, any tools that are used to study the proteome must be able to determine the protein's

1

primary sequence and identify any modifications that might be present. Such challenges necessitate the use of highly selective and sensitive analytical techniques. In this regard, mass spectrometry (MS)–based techniques have played a crucial role in proteomic studies, affording many advantages including fast analysis time, high sensitivity, large dynamic range, tremendous resolving power, and the ability to identify and quantify up to a few thousand proteins in a single analysis (6, 7). Specifically, tandem mass spectrometry has played a crucial role in the identification and sequencing of peptides and proteins, and in the characterization of their PTMs (8).

The expansion of the use of analytical mass spectrometry in proteomic analyses has coincided with several advances in MS technologies. The most important of these was the development of electrospray ionization (ESI) by Fenn et al. in 1985 (9). For the first time, peptides and proteins could be ionized efficiently at atmospheric pressure directly into the inlets of mass spectrometers (10). This allowed users to integrate high performance liquid chromatography (HPLC) for the separation of peptides and proteins directly with MS, which enabled the analysis of complex biological mixtures containing multiple proteins at varying abundances. With this technological achievement, coupled with the availability of gene and genome sequence databases, biological mass spectrometry grew into an indispensible tool for the study molecular and cellular biology. This introduction will provide context to the utility of MS in characterizing proteins and will explain the basics of several MS-proteomics platforms. We will also outline the advantages and disadvantages of each and discuss how they impact the analysis of biological samples.

1.2 Mass Spectrometry—Principles and Instrumentation

1.2.1 Ionization Methods

Mass spectrometric measurements determine the mass to charge ratios (m/z) of ionized analytes in the gas phase. Before these measurements can be made, these analytes must first be injected into the mass spectrometer. In proteomic analyses, physical introduction of peptide and protein cations is typically accomplished in one of two ways. These are matrix-assisted laser desorption/ionization (MALDI) and ESI. MALDI is performed using laser pulses to desorb and ionize samples from a solid, crystalline matrix (11). Since the sample must be crystallized, it cannot be coupled directly with chromatography, and thus it is usually used for samples of low complexity. Once ionized by MALDI, most analytes exhibit +1 charge states, and so the masses of the molecules that can be analyzed are limited to the scan range of the mass spectrometer (2,000-4,000 m/z for ion trap based analyzers). This is why MALDI is usually coupled with time-of-flight (TOF) mass analyzers, which, theoretically, have unlimited mass ranges.

ESI is accomplished by applying a high voltage (~2-3 kV) to an electrospray "tip", through which analytes in solution are passed at tens to hundreds of μ L/min. This tip is a sharp, capillary needle with a very small inner diameter. As the solution emerges from the tip, it is dispersed by Coulombic forces into a fine spray of polarized droplets. The droplets travel towards the inlet of the mass spectrometer, which is held at a low voltage potential relative to that of the electrospray tip. The droplets enter the mass spectrometer through a heated capillary, facilitating solvent evaporation. As solvent molecules evaporate, the droplets decrease in diameter and their surface charge densities increase until the Coulombic repulsive forces equal the surface tension and the droplet "explodes", producing smaller progeny droplets. This evaporation/fission cycle

3

repeats until droplets contain one or fewer analyte ions. The analytes enter the gas phase with the charges the droplet carried after the remaining solvent evaporates (12, 13). ESI has several major advantages over MALDI. First, since analytes are ionized directly from solution, ESI can be directly coupled to HPLC, allowing for the analysis of more complex mixtures. By limiting the number of species entering the mass spectrometer at any given moment via chromatographic separation, the dynamic range of the mass analyzer is improved, allowing the detection of low-level proteins and substoichiometric modifications. Second, ions produced by ESI exhibit multiple, high charge states, yielding signals of lower m/z values than those obtained by MALDI. In most cases, this allows ESI to be used with nearly any mass analyzer. However, this freedom is restricted for the analysis of intact proteins, which, for large proteins, is often hindered by limits in m/z scan range. Also, since ESI-produced analyte ions exist in multiple charge states and give signals that are split among several m/z values, spectral interpretation is more complicated compared to data generated by MALDI-TOF MS. This has led to the development of spectral deconvolution which is an algorithmic means of simplifying ESI-MS and tandem mass spectra by transforming the multiplicity of peaks into the single peaks that would be expected if all ions derived from the same molecule had a single charge (10, 14).

1.2.2 Mass Analyzers

Once inside the instrument, the ions are electrostatically guided to the mass analyzer, typically through a series of lenses and multipole ion guides. For the sake of brevity and simplicity, we will limit detailed discussion of mass analyzers to those that are used for the experiments described in this dissertation. Quadrupole ion traps are one of the preferred analyzers for the characterization of peptides and proteins because of their relatively low cost, the speed with which tandem MS experiments can be performed and their sensitivity (6). Data can be acquired on a chromatographic timescale at rates of over 10 scans/second. Their multifaceted utility and speed is derived from their ability to manipulate the motion of ions quickly. This ability is rooted in the basic theory of quadrupole interactions (15, 16). A quadrupole mass analyzer consists of a set of four parallel conductive metal rods arranged in a square array, and through the center of which ions drift lengthwise. To one pair of opposing rods, a positive direct current (dc) potential overlaid with a radio frequency (rf) potential is applied; potentials of the same amplitude are applied to the other pair of rods, but the dc potential is of negative polarity, and the rf potential is 180° out of phase with that applied to the first set of rods. This creates an oscillating field normal to the direction of ion drift, and through which ions of specific m/z ranges can stably pass to the detector. Ions of unstable trajectories collide with the rods and are not detected. Mass-to-charge ratios are determined by varying the amplitude of the potentials applied to the rods; ions detected under certain electrostatic field conditions possess unique m/z values.

In a quadrupole ion trap (QIT), is a three dimensional analog of the quadrupole mass analyzer (16). Instead of the square rod array, there are two end-cap electrodes, to which dc potentials can be applied, and a hyperbolic ring electrode, to which rf potential is applied, creating a quadrupolar field. Following ion injection and collisional cooling with helium atoms, an ion's motion as it oscillates a quadrupolar field can be defined by the Mathieu equation (**Equation 1.1**).

(1.1)
$$\frac{d^2u}{d\xi^2} + (a_u - 2q_u cos 2\xi)u = 0$$

The variable, *u* represents the ion's displacement in the x, y and z directions; ξ is equal to $\omega t/2$, where ω is the ion's radial frequency; *t* is time and *a* and *q* are dimensionless trapping parameters which describe the stability of ions within the quadrupolar field represented by the first stability diagram (**Figure 1.1**). If the differential is solved, the

values of a and q are given by **Equations 1.2** and **1.3**, respectively. These values define ion motions that are stable within the confines of the quadrupolar field.



(1.2)
$$a = \frac{-8eU}{m\omega^2 r_0^2}$$
 (1.3) $q = \frac{4eV}{m\omega^2 r_0^2}$

Figure 1.1 Stability diagram showing combined axial and radial stability within a quadrupole ion trap. Ions having values of *a* and *q* within the gridded region are stably trapped. Ions with *a* and *q* values outside this region exhibit unstable motion, and are ejected from the trap. For m/z determination the QITs is operated in rf-only mode, making a = 0. The value of *q* and therefore, stable motion within the trap, is determined by the amplitude of the rf and an ion's m/z. At $q \ge 0.908$, ions exhibit unstable trajectories and exit the trap. This image was taken from (16).

Here, U and V are the amplitudes of the applied dc and rf potentials, respectively; e is the charge of the ion (the z of m/z, defined as e to avoid confusion with the z direction); ω is the radial frequency of the applied potential; and r_0 is the internal radius of the ring electrode. Thus, values of a and q determine whether an ion will possess motion that is stable within the confines of the trap. Note that for most proteomic applications, following initial ion trapping, QITs are operated in rf-only mode, and since no dc potential is applied, a = 0 and ion stability is governed by the value of q. Mass spectra are recorded by increasing the amplitude of the rf (V) so that an ion's q value is "moved" along the q axis of the stability diagram (**Figure 1.1**) until it reaches a value of 0.908. At this point, an ion's motion is unstable, and the ion is ejected from the trap where it is detected; typically using an electron multiplier. If **Equation 1.3** is rearranged, the m/z values of ions can be determined by **Equation 1.4**.

(1.4)
$$m/z = \frac{4V}{0.908\omega^2 r_0^2}$$

Today, the most commonly used ion trap-based instrument is the two-dimentional quadrupole ion trap (quadrupole linear trap, or QLT). A QLT operates under the same principles as those which guide ion motion in a QIT. The main difference is that a QLT is stretched lengthwise to accommodate more ions, and thus, generate better sensitivity. Its design is shown in **Figure 1.2**. At the most basic level, it is a quadrupole. However, each rod is divided into three sections. The first and last sections are analogous to the end-cap electrodes of the QIT and the middle section is like the ring electrode. In addition to the ability to trap and store larger quantities of ions, the QLT has an added



Figure 1.2 Basic design of a two-dimentional linear ion trap (QLT). The rods are cut into three distinct sections, each with discrete electronic control. This image was taken from (17).

functionality that is unique to its design. Since the rods are arranged in three distinct sections, different dc potentials can be applied to each section to create potential wells within certain regions of the trap (17). This allows for the movement and storage of ions in discreet sections (**Figure 1.3**). This capability enables the storage of ions of opposite

polarity, which, in turn, enables the performance of ion/ion reactions within the trap; a development of great import to several proteomics applications.



Figure 1.3 QLT voltages for the simultaneous storage of positive and negative ions. A diagram of the QLT is shown at the top to indicate the voltage settings on each of the three sections of the trap as well as the front and back lenses. Panel (**A**) shows that by sequentially lowering the dc voltages applied to each section of the trap, positive ions are injected and move into the back section. In panel (**B**) positive potential is applied to the front and middle sections of the trap while the rear section is held at a negative potential. This allows for the injection of negative ions while the positive ions remain trapped in the rear section. Panel (**C**) shows the voltage of the front section being dropped to create a voltage gate so that negative ions don't leak out of the trap. Positive and negative ions are simultaneously stored in their discrete sections.

In addition to their ability to serve as ion storage devices and mass analyzers, ion traps are capable of resonant excitation (15). Resonant excitation is the application of a small (on the order of millivolts) supplementary potential on the end-cap electrodes that oscillates at the frequency of a specific ionic species. This causes these ions to move away from the center of the trap. Their movement away from the center of the quadrupolar field causes them to experience a stronger trapping force and so they possess higher kinetic energies. In proteomics, resonant excitation is used in tandem MS experiments to remove unwanted ions from the trap and thereby isolate ions of interest, and then to increase the kinetic energy of those ions so that they undergo collisions with helium atoms and fragment (18-21). This will be discussed in more detail later in the chapter.

Despite their strengths, quadrupole devices are limited in their resolving power and mass accuracy. This makes it nearly impossible to unambiguously identify peptides and proteins derived from complex biological samples. As a result, QLTs are often coupled with high resolution mass analyzers creating hybrid instruments (22, 23). This way, sequence information from tandem MS experiments can be coupled with accurate mass and resolution measurements to identify analytes confidently. There are several high resolution mass analyzers that are can operate on a chromatographic timescale. The fastest of these is the recently developed Orbitrap mass analyzer (24).

Orbitraps, like ion traps, trap and store ions electrostatically. The device, pictured in **Figure 1.4**, is comprised of an outer barrel electrode and an inner spindle electrode. A 'quadrologarithmic' field is created when a dc potential is applied between the coaxially configured inner and outer electrodes (24). Ions are injected perpendicular to the long axis of the electrodes, accelerate towards the inner electrode and become trapped in elliptical orbits, such that the attractive force towards the inner electrode is balanced by the ions' centrifugal force. The ions quickly dephase radially, creating evenly dispersed

ion rings. The electrostatic field is normal to the surface of the electrode, which is curved, creating a field that is focused in the center of the trap. The point of injection is displaced from the center of the electrode, giving the ions potential energy along the axial direction of the trap (25). As a result, the rings of ions oscillate along this axis with an angular frequency, ω , given by **Equation 1.5** where *k* is the force constant of the quadratic restoring potential.

(1.5)
$$\omega = \sqrt{\left(\frac{z}{m}\right) * k}$$

The outer electrode, which is split along its central axis, detects the image current of the ions as they oscillate in the time domain for a certain period. The image current detected in this time is called a "transient". The transient is converted by Fourier transform (FT) to the frequency domain, from which m/z is determined via the relationship given in **Equation 1.5** with great accuracy and with high isotopic resolution.



Figure 1.4 Cutaway view of an Orbitrap mass analyzer. Ions are injected from the point indicated by the red arrow and follow the indicated path. After radial dephasing, ions orbit the electrode in rings which move back and forth along the axis of the electrode. Image modified from (26).

Accurate mass determination and resolution in the Orbitrap depend on the length of time ions of the same m/z maintain coherence. This coherence initially depends upon the ability to inject the ions into the Orbitrap all at once. This is accomplished using a C-

shaped quadrupole called a C-trap which is placed between two electrodes (23). The trap is filled with nitrogen gas at pressure of ~1 mTorr which is used for collisional damping of ion motion. Ions are injected into the C-trap where their motion is cooled, and they are stored in a long, thin line along the curved axis of the C-trap. The line is squeezed axially by applying a positive potential to the two electrodes. Then, the rf trapping voltage on the guadrupole rods is rapidly lowered and dc potentials are applied to the rods sequentially, creating a voltage distribution that forces the ions to move downward, orthogonally to the axis of curvature. The ions leave through a slot of decreasing diameter which is cut in one of the rods of the quadrupole. This causes the ions to be further compressed as they leave the C-trap and as a result, they are ejected into the Orbitrap as a tight packet of ions. The ions begin their axial and orbital motion almost simultaneously, allowing the detection of coherent image current before the rings of ions begin to spread due to collisions with gas molecules inside the Orbitrap. To extend this time, the Orbitrap is kept under an ultra-high vacuum (~1E-10 Torr). Under these conditions, the ions maintain coherence for over a second, allowing the acquisition of long transients, which equates to higher resolution (26). Today's Orbitrap mass analyzers can determine m/z to within less than 1 ppm and have resolving powers that approach 500,000 at 200 m/z. Transients can be acquired in as little as 256 ms (27).

1.2.3 Tandem Mass Spectrometry (MS/MS)

While accurate mass determination is necessary for the unambiguous identification of peptides and proteins, truly confident identifications can only be accomplished by determining/confirming the sequence of the analyte. Additionally, any PTMs or other modifications that are present should be site mapped. In this regard the rise of hybrid mass analyzers has expanded the use of mass spectrometry in proteomics applications by enabling the performance of high resolution tandem mass spectrometry (MS/MS). MS/MS analyses involve multiple steps which integrate the use of multiple mass analyzers. The goal is to accurately determine the mass of analytes as they are introduced into the instrument and then subsequently generate sequence information via analyte fragmentation. These processes should be completed on a timescale that is compatible with chromatography so that complex mixtures can be analyzed.

The process begins by acquiring an initial survey scan (full MS or MS1) in which the m/z of all ESI generated analyte ions are determined simultaneously by a high resolution mass analyzer. Each ion's m/z value is reported in a single spectrum along with their relative abundances, which are normalized to the most abundant ion. Then, certain analytes are selected for fragmentation in subsequent MS/MS experiments; these analytes are called "precursors". The selection process is generally accomplished by data-dependent selection. In data dependent selection, the data system of the instrument's on-board computer will use information collected in the MS1 scan to select and trigger MS/MS experiments on the fly. The selection is usually based on the observed relative abundance of the analytes in the full MS spectrum. Typically the top 1-10 most abundant ions in an MS1 spectrum are selected and fragmented. After all the MS/MS spectra have been acquired, the cycle repeats based upon the next full MS spectrum that is acquired. The number of MS/MS scans performed in the cycle is set by the user and should allow for a sufficient number of MS1 scans to be acquired as an analyte elutes to define a chromatographic peak. Since some analytes might be observed over several cycles, data-dependent selection is usually coupled with dynamic exclusion. Dynamic exclusion prevents the selection of the same precursor ions for an excessive number of scan cycles by setting a threshold for the number of times a precursor can be selected within a prescribed period of time. Once a precursor meets this threshold it cannot be reselected until an exclusion duration period has expired. In the mean time, if an ion has met the threshold for exclusion, the ion of the next highest intensity is selected in its stead. The coupling of data dependent selection with dynamic exclusion enables the generation of MS/MS spectra for precursors that possess abundances that are orders of magnitude lower than the most abundant species in the MS1 spectrum. This is critical for sequencing peptides and proteins that contain PTMs at substoichometric abundances. Specific precursors can also be selected by the user in "targeted" MS/MS experiments. When this is done, the m/z of a desired precursor ion is defined in the instrument method. MS/MS spectra of this precursor are acquired in every scan cycle, regardless of whether or not they are clearly observed in the MS1 spectrum. The two selection methods can be combined in a single method where the scan cycle includes a prescribed set of data dependent selections followed by a set of targeted selections.

Once an ion has been selected for MS/MS interrogation, it is isolated and fragmented. For decades, collision-activated dissociation (CAD) dominated the field of MS-based proteomics as the preferred method of fragmentation for peptides. In CAD, peptide precursors are isolated and subjected to multiple low energy collisions with inert gases such as helium and argon within an ion trap or triple quadrupole (18, 19, 28). The collisions are triggered by the application of a supplementary rf potential that is tuned to the resonant frequency of the precursor m/z. These collisions impart energy to the peptide ions, which is transferred and distributed throughout the peptides as vibrational excitation. Once the vibrational energy in a peptide bond exceeds the activation barrier required to dissociate the weakest bond, the bond is cleaved and the peptide ion fragments. CAD is generally observed to impart cleavage of the protonated amide bond of the peptide backbone, as this is the lowest energy pathway, resulting in the production of b- (N-terminus containing) and y-type (C-terminus containing) ions (**Figure 1.5**) (21). Fragment ions, which don't possess the same m/z as their precursor, are not excited by the supplemental rf and do not undergo further collisional activation.



Figure 1.5 CAD produced b- and y-type fragment ions. Cleavage of the protonated amide bond results in the production of complementary b- and y-ions. Consecutive ions within an ion series (b or y) differ in mass by a single amino acid residue. The masses of complementary b- and y-ions sum to the mass of the precursor ion.

Optimally, the peptide backbone is randomly protonated during ionization by ESI and all amide bonds of the peptide backbone are fragmented by CAD. The resulting spectrum (MS/MS or MS2 spectrum) should contain a series of b- and a series of y-type ions that correspond to cleavage at all amide bonds. The precursor peptide is sequenced by determining the mass differences between consecutive b- and y-ions; these differences correspond to the mass of a single amino acid residue (**Figure 1.5**). This process is referred to as *de novo* sequencing and is commonly used for the sequence elucidation of unknown peptides and proteins and PTM site mapping. Confident sequence determination depends upon the extent of amide bond dissociation and the subsequent detection of fragment ions.

Because efficient fragmentation by CAD is dependent upon the random protonation of amide bonds, it is generally limited to the analysis short peptides (~10-15 residues) containing low charge ($z \le +3$). When a peptide contains many basic residues, random protonation of the peptide backbone is inhibited and upon collisional activation, bond dissociation is directed to a few specific sites (20, 29). Thus, only a few fragment ions are observed giving little sequence information. Additionally, in the probing of posttranslationally modified peptides possessing labile moieties (phosphorylation, O-GlcNAcylation, etc), the bond which offers the lowest energy pathway for CAD fragmentation is often the bond containing the PTM (30, 31). Thus, spectra of PTMcontaining peptides are dominated by a peak corresponding to the loss of the PTM, and limited additional fragmentation is observed yielding little sequence information and leaving the site of the PTM undetermined if more than one possible modifiable residue is present on the peptide.

In order to improve the sequence information that is generated in MS/MS analyses, several researchers set out to develop alternative methods of fragmentation. The development of electron capture dissociation (ECD) in 1998 (32) provided a more efficient means of fragmenting highly charged precursor ions without the loss of labile PTMs. ECD involves the capture of thermal electrons by precursor cations, which triggers radical-driven dissociation of the peptide N-Cα bond. However, the use of ECD failed to take hold in the field as it can only be performed in the costly Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer and, early on, it was difficult to implement on a chromatographic time-scale. This prompted the development of electron transfer dissociation, the ion-ion analog of ECD, which is widely used today.

Electron transfer dissociation (ETD) is a relatively new fragmentation technique developed in the Hunt laboratory in 2004 (20). ETD exploits polycyclic aromatic radical anions for electron transfer to an amide bond along the peptide backbone. This generates a carbonyl radical anion which abstracts a proton from the nearby basic residue. This charge reduced species contains a radical site which triggers fragmentation of the N-C_{α} bond (20, 33). The result is the production of c- and z-type fragment ions which correspond to ions that contain the N- and C-termini, respectively (**Figure 1.6**). ETD easily fragments analytes containing multiple basic residues including large peptides and intact proteins. Additionally, ETD preserves labile PTMs making it useful for site mapping (34, 35). Since ETD reactions require that positive and negative ions be simultaneously trapped and stored, they are performed in QLT mass analyzers.

However, if high resolution mass analysis of the resulting fragment ions is desired, they can be transferred to the appropriate mass analyzer in hybrid instruments. Electron transfer reactions are performed rapidly (in milliseconds), and so they are easily implemented on a chromatographic timescale.



Figure 1.6 ETD reaction mechanism. Fragmentation scheme for the production of ions of types c and z by cleavage of the N-C α bond of a multiply charged precursor following the transfer of an electron from a radical anion of fluoranthene.

1.2.4 Bottom-up, Middle-down and Top-down Proteomics

Although ETD has been successful in overcoming the limitations of CAD in tandem mass spectrometry, it has not replaced CAD in proteomics analyses. In fact, both techniques are used, often side by side in sequential MS/MS experiments. The fragmentation technique that is selected for an MS analysis often depends upon how a sample is prepared and what information is desired. Early on, in the absence of alternatives to CAD, peptides and proteins were identified and sequenced "from the bottom up". In bottom-up approaches, proteins from biological samples are digested into peptides by trypsin, a serine protease which cleaves C-terminal to lysine and arginine

residues. The use of trypsin creates peptides that are short and optimal for fragmentation by CAD, containing a single basic residue (lysine or arginine) at the C-terminus and the charged amino N-terminus, which facilitates detection of both complementary b and y fragment ion pairs. The resulting mixture of peptides is separated using reverse-phase HPLC on a column integrated to an ESI source, which sprays the charged peptides into the inlet of the mass spectrometer as they elute. The ionized peptides are analyzed for their accurate m/z and analytes are selected and isolated for further structural interrogation by MS/MS using CAD. Proteins are identified usually by database search in which MS/MS spectra are compared with the predicted fragmentation patterns of tryptic peptides derived from proteins contained in the database (36, 37).

There are several advantages of using bottom-up approaches. The most important of these is the improved front-end separation of tryptic peptides compared to larger peptides and intact proteins, allowing for the successful analysis of even very complex mixtures, as the high degree of separation promotes higher levels of dynamic range. This coupled with the fact that small peptides exhibit significantly less ion current splitting when ionized by ESI results in a higher degree of sensitivity. Bottom-up approaches are ideal for the rapid identification of proteins in biological samples. However, no technique is without its drawbacks. The use of CAD complicates PTM site-mapping, since labile PTMs are preferentially cleaved. Also, since CAD is limited in its application to peptides containing few basic residues, dictating that they be short, the detection of multiple simultaneous PTMs on a single peptide is rarely reported. A protein's biological function can largely be determined by collective combinations of PTMs because the functional effects of a particular PTM may vary based on the presence of other PTMs on the protein (35). Therefore, the use of CAD not only limits the size of the peptides analyzed, but also the information that can be garnered regarding PTMs, the relationships between them, and their effect on the protein (38). Finally, it is common that many of the peptides generated in a tryptic digest are not recovered in the LC-MS analysis, making complete protein sequence determination impossible.

The invention of ETD sparked the development of the "middle-down" approach, which is very similar to bottom-up with the exception of the endoprotease used for digestion. Middle-down approaches use enzymes such as Asp-N or Glu-C to generate peptides that are much larger than their tryptic counterparts. These peptides are more suited for fragmentation by ETD in MS/MS analyses. As a result, labile PTMs are preserved and the chances of observing multiple PTMs on the same peptide are greatly increased, giving the data more value. While this is a step in the right direction, middle-down approaches are still not optimal for many of the same reasons that bottom-up approaches are not optimal. PTMs that are not in close enough proximity to coexist on the same peptide are not observed in the context of one another and the degree of sequence coverage hinges upon the number of peptides that are recovered and selected for MS/MS.

From a biological standpoint, the ultimate proteomics platform would be capable of identifying each unique "proteoform" within a biological sample. The word "proteoform" is used to designate all of the unique molecular forms in which a protein product of a single gene exists including changes due to genetic variations, splice variations and any biologically affected co- or post-translational modification (39). This can only be accomplished by analyzing the protein in its intact form, as it existed in the sample from which it was derived. To that end, top-down proteomics represents the "holy grail" of proteomics approaches and involves analyzing intact protein molecules.

There are several challenges associated with the analysis of whole proteins which keep top-down approaches from widespread use. These are outlined in detail in the **Chapter 2**. Briefly, they include limitations in front-end separation, incomplete

fragmentation, low signal-to-noise ratios, and difficulties in data analysis owing to the complexity of the MS/MS data generated. This dissertation outlines the implementation of new technologies and methodologies for the improved analysis of proteins by top-down mass spectrometry. The project also takes advantage of several existing techniques and mass analyzers that are reviewed in this introduction, but widens their functionality to enable the acquisition of high quality MS/MS spectra of whole proteins. Their utility is demonstrated in an application to intact histones derived from butyrate treated HeLa cells. We hope that the methods we introduce will guide the design of instrumentation and software that will bring top-down mass spectrometry further into the fold of tools that can be routinely applied to the study of biology.

1.3 References

1. Proteome and proteomics: New technologies, new concepts, and new words, Anderson N. L. & Anderson N. G., <u>Electrophoresis</u> **1998** (19): 1853-1861.

2. The language of covalent histone modifications, Strahl B. D. & Allis C. D., <u>Nature</u> **2000** (403): 41-45.

3. Phosphorylation of the proteasome activator PA28 is required for proteasome activation, Li N., Lerea K. M. & Etlinger J. D., <u>Biochem Biophys Res Commun</u> **1996** (225): 855-860.

4. O-GlcNAc modification is an endogenous inhibitor of the proteasome, Zhang F., Su K., Yang X., Bowe D. B., Paterson A. J. & Kudlow J. E., <u>Cell</u> **2003** (115): 715-725.

5. Proteomic analysis of post-translational modifications, Mann M. & Jensen O. N., <u>Nat</u> <u>Biotechnol</u> **2003** (21): 255-261.

6. Dass C (2001) *Principles and Practices of Biological Mass Spectrometry,* (John Wiley & Sons, Inc,

7. A stress test for mass spectrometry-based proteomics, Aebersold R., <u>Nat Methods</u> **2009** (6): 411-412.

8. Protein sequencing by tandem mass spectrometry, Hunt D. F., Yates J. R., 3rd, Shabanowitz J., Winston S. & Hauer C. R., <u>Proc Natl Acad Sci U S A</u> **1986** (83): 6233-6237.
9. Electrospray interface for liquid chromatographs and mass spectrometers, Whitehouse C. M., Dreyer R. N., Yamashita M. & Fenn J. B., <u>Anal Chem</u> **1985** (57): 675-679.

10. Electrospray ionization for mass spectrometry of large biomolecules, Fenn J. B., Mann M., Meng C. K., Wong S. F. & Whitehouse C. M., <u>Science</u> **1989** (246): 64-71.

11. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, Karas M. & Hillenkamp F., <u>Anal Chem</u> **1988** (60): 2299-2301.

12. Molecular beams of macroions. 3. zein and polyvinylpyrrolidone, Clegg G. A. & Dole M., <u>Biopolymers</u> **1971** (10): 821-826.

13. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism, Fernandez de la Mora J., <u>Anal Chim Acta</u> **2000** (406): 93-104.

14. A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra, Zhang Z. & Marshall A. G., <u>J Am Soc Mass</u> <u>Spectrom</u> **1998** (9): 225-233.

15. An introduction to quadrupole ion trap mass spectrometry, March R. E., <u>Journal of</u> <u>Mass Spectrometry</u> **1997** (32): 351-369.

16. Ion trap mass spectrometry, Wong P. S. & Graham Cooks R., <u>Current Separations</u> **1997** (16): 85-92.

17. A two-dimensional quadrupole ion trap mass spectrometer, Schwartz J. C., Senko M. W. & Syka J. E., <u>J Am Soc Mass Spectrom</u> **2002** (13): 659-669.

18. Metastable ion characteristics. VII. collision-induced metastables, Haddon W. & McLafferty F., <u>J Am Chem Soc</u> **1968** (90): 4745-4746.

19. Tandem mass spectrometry, McLafferty F. W., Science 1981 (214): 280-287.

20. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Syka J. E., Coon J. J., Schroeder M. J., Shabanowitz J. & Hunt D. F., <u>Proc</u> <u>Natl Acad Sci U S A</u> **2004** (101): 9528-9533.

21. Protein sequencing by tandem mass spectrometry, Hunt D. F., Yates J. R., 3rd, Shabanowitz J., Winston S. & Hauer C. R., <u>Proc Natl Acad Sci U S A</u> **1986** (83): 6233-6237.

22. Novel linear quadrupole ion trap/FT mass spectrometer: Performance characterization and use in the comparative analysis of histone H3 post-translational modifications, Syka J. E., Marto J. A., Bai D. L., Horning S., Senko M. W., Schwartz J. C., Ueberheide B., Garcia B., Busby S., Muratore T., Shabanowitz J. & Hunt D. F., J Proteome Res **2004** (3): 621-626.

23. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer, Makarov A., Denisov E., Kholomeev A., Balschun W., Lange O., Strupat K. & Horning S., <u>Anal Chem</u> **2006** (78): 2113-2120.

24. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis, Makarov A., <u>Anal Chem</u> **2000** (72): 1156-1162.

25. The orbitrap: A new mass spectrometer, Hu Q., Noll R. J., Li H., Makarov A., Hardman M. & Graham Cooks R., <u>J Mass Spectrom</u> **2005** (40): 430-443.

26. Orbitrap mass analyzer--overview and applications in proteomics, Scigelova M. & Makarov A., <u>Proteomics</u> **2006** (6 Suppl 2): 16-21.

27. Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates, Senko M. W., Remes P. M., Canterbury J. D., Mathur R., Song Q., Eliuk S. M., Mullen C., Earley L., Hardman M., Blethrow J. D., Bui H., Specht A., Lange O., Denisov E., Makarov A., Horning S. & Zabrouskov V., <u>Anal Chem</u> **2013** (85): 11710-11714.

28. Ion activation methods for tandem mass spectrometry, Sleno L. & Volmer D. A., <u>J</u> <u>Mass Spectrom</u> **2004** (39): 1091-1112.

29. Mobile and localized protons: A framework for understanding peptide dissociation, Wysocki V. H., Tsaprailis G., Smith L. L. & Breci L. A., <u>J Mass Spectrom</u> **2000** (35): 1399-1406.

30. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry, Schroeder M. J., Shabanowitz J., Schwartz J. C., Hunt D. F. & Coon J. J., <u>Anal Chem</u> **2004** (76): 3590-3598.

31. Analysis of phosphorylation sites on proteins from saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry, Chi A., Huttenhower C., Geer L. Y., Coon J. J., Syka J. E., Bai D. L., Shabanowitz J., Burke D. J., Troyanskaya O. G. & Hunt D. F., <u>Proc Natl Acad Sci U S A</u> **2007** (104): 2193-2198.

32. Electron capture dissociation of multiply charged protein cations. A nonergodic process, Zubarev R. A., Kelleher N. L. & McLafferty F. W., <u>J Am Chem Soc</u> **1998** (120): 3265-3266.

33. Probing the mechanism of electron capture and electron transfer dissociation using tags with variable electron affinity, Sohn C. H., Chung C. K., Yin S., Ramachandran P., Loo J. A. & Beauchamp J. L., <u>J Am Chem Soc</u> **2009** (131): 5444-5459.

34. Protein identification using sequential ion/ion reactions and tandem mass spectrometry, Coon J. J., Ueberheide B., Syka J. E., Dryhurst D. D., Ausio J., Shabanowitz J. & Hunt D. F., <u>Proc Natl Acad Sci U S A</u> **2005** (102): 9463-9468.

35. The utility of ETD mass spectrometry in proteomic analysis, Mikesh L. M., Ueberheide B., Chi A., Coon J. J., Syka J. E., Shabanowitz J. & Hunt D. F., <u>Biochim</u> <u>Biophys Acta</u> **2006** (1764): 1811-1822.

36. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, Eng J. K., McCormack A. L. & Yates J. R., <u>J Am Soc</u> <u>Mass Spectrom</u> **1994** (5): 976-989.

37. Probability-based protein identification by searching sequence databases using mass spectrometry data, Perkins D. N., Pappin D. J., Creasy D. M. & Cottrell J. S., <u>Electrophoresis</u> **1999** (20): 3551-3567.

38. Perspectives in spicing up proteomics with splicing, Godovac-Zimmermann J., Kleiner O., Brown L. R. & Drukier A. K., <u>Proteomics</u> **2005** (5): 699-709.

39. Proteoform: A single term describing protein complexity, Smith L. M., Kelleher N. L. & Consortium for Top Down Proteomics, <u>Nat Methods</u> **2013** (10): 186-187.

2 Novel Methods in Mass Spectrometry Facilitate Near-Complete Amino Acid Sequence Coverage of Intact Proteins

This work is a collaborative effort between members of the Hunt Laboratory, and Dr. John Syka, Lee Earley and Dr. Christopher Mullen of Thermo Fisher Scientific.

2.1 Abstract

Sequence determination of intact proteins by mass spectrometry allows for the study of unique protein isoforms that is not possible by other sequencing methods. The structure of each proteoform is known to play a significant role in determining the biological functions of the protein and thus, the ability to characterize all modifications that coexist on a particular protein is of utmost importance. Electron-transfer dissociation has significantly improved our ability to sequence whole proteins and large peptides. However, its successful implementation depends on several characteristics of the region of the chain to be fragmented. Additionally, these approaches suffer significant limitations as low signal-to noise ratios require the averaging of considerable numbers of spectra to produce a useful spectrum. These spectra are typically complex, with many overlapping fragments that are very difficult to interpret or deconvolute. These obstacles have proved a difficult hurdle to overcome in the burgeoning field of top-down proteomics. Here, we report novel MS technology and chemical derivatization methods that facilitate near-complete amino-acid sequence coverage of intact proteins on a chromatographic time-scale. We combine front-end enabled electron-transfer dissociation for fragmentation of intact proteins, multiple C-trap ion-fills for enhancing ion current observed for protein samples, ion-ion proton transfer reactions performed in the mass spectrometer that produce simplified fragmentation spectra to aid in sequence determination, and covalent labeling strategies that help direct the fragmentation to areas lacking sequence coverage. These techniques are combined to yield comprehensive coverage of the standard protein, apomyoglobin (17 kDa, 94%). This represents a significant improvement in coverage by similar techniques reported in the literature to date. The use of these novel mass spectrometry methods could transform the way we do research in areas such as drug development, immunology, epigenetics, microbiology and cell signaling.

2.2 Introduction

2.2.1 A Brief History

In the past decade, efforts to implement mass spectrometry-based methods for the analysis of intact protein molecules (top-down proteomics) have realized growing success (1, 2) Early top-down analyses were not routinely implemented as fragmentation by collision activated dissociation (CAD) failed to generate the sequence coverage necessary to characterize proteins completely. Additionally, the preferential cleavage some post-translational modifications (PTMs) made it difficult or impossible to perform analyses focused on PTM site localization. The development of electron capture dissociation (ECD) in 1998 (3) provided a more efficient means of fragmenting highly charged precursor ions without the loss of labile PTMs, but the use of ECD failed to take hold in the field as it can only be performed in the costly Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer and, early on, it was difficult to implement on a chromatographic time-scale.

In 2004, a technique analogous to ECD, electron transfer dissociation (ETD) (4) was developed and implemented on more conventional quadrapole linear ion trap (QLT) based instruments. ETD is easily implemented on a chromatographic time-scale, and preserves labile PTMs. Soon after its development, top-down applications using ETD were attempted on a low resolution, QLT platform. However, the resulting highly charged fragment ions were impossible to resolve on the QLT mass spectrometer,

making *de novo* sequencing very difficult. Thus, fragmentation by ETD was paired with sequential ion/ion proton transfer reactions (IIPT) (5-7) using benzoate anions to charge-reduce the mixtures of highly charged fragments down to +1 and +2 ions prior to m/z determination (8, 9). This made reading the sequence starting from the protein termini possible for up to 20 residues and greatly simplified spectral interpretation. The sequence information gathered, combined with accurate mass information is sufficient to identify proteins confidently using data base search algorithms. Note that this requires that the sequences of the proteins analyzed be present in the database being searched.

While these results represented a culmination of tremendous efforts by numerous notable scientists, the applications were still limited. Identifying and detecting multiple, unique protein isoforms (proteoforms) was possible on a chromatographic time-scale, but the sequence information gathered was localized to only the extreme N- and C-terminal regions of the proteins analyzed. The m/z scan range of a bench-top QLT mass spectrometer extends to just 2000. Since the average mass of an amino acid is about 110 Da, this range is generally exhausted after the first 18 residues if fragment ions carrying only a +1 charge are considered. If +2 fragment ions are considered, the range is extended to about 36 residues. Ions carrying charges of +3 or higher are difficult to consider as they cannot be resolved using QLT instruments. Thus, these top-down methods were only applicable to proteins the size of ubiquitin (76 residues, 8.6 kDa) or smaller for full sequence characterization or for use in the rapid identification of proteins without obtaining full sequence and PTM state information.

The commercial release of the Orbitrap mass analyzer in 2005 has sparked a tremendous effort to develop more routine top-down proteomics protocols on a high-resolution platform. The Orbitrap offers accurate mass and high resolution similar to those achievable with FT-ICR instrumentation (10, 11), but is much less costly and more user-friendly. Initially, ETD reagent ionization sources were not available for the Orbitrap

and so researchers first had to develop ways to introduce ETD reagent ions into the instrument. An approach using two electrospray ionization (ESI) emitters, one for samples and one for ETD reagents, failed to generate ideal reagent anions by electrospray (12). The reagent used, 9-anthracenecarboxylic acid, had to undergo CAD to produce the ETD anion and required excessively long injection times. Later, a negative chemical ionization source was placed at the rear of the instrument (13) similar to Thermo Fisher Scientific's commercial ETD-enabled LTQ series instruments (an LTQ is a Thermo brand QLT). This became the commercially available ETD platform for Thermo's Orbitrap Velos Pro[™] instruments.

2.2.2 Signal-to-noise Ratios in Top-down Mass Spectrometry

Aside from introducing reagent anions to enable recording high resolution ETD mass spectra, many technical challenges remain. Chiefly, top-down based methods suffer significantly lower signal-to-noise ratios (SNR) compared to their bottom-up based counterparts. These low SNR are inherent to the study of macromolecules using high resolution mass spectrometry and arise from numerous sources. First, we must consider the number of channels among which the precursor analyte signal can be split. The primary contributor to the splitting of precursor ion current is the number of charge states an analyte can exhibit using ESI. Figure 2.1 shows the charge state distributions (CSD) of a small peptide, vasoactive intestinal peptide (VIP) and an intact protein, apomyoglobin. Using ESI, VIP, a 12 residue peptide, exhibits charge states of +1 through +3, and so, the ion current is split among three channels. Intact apomyoglobin, a 153-residue protein, exhibits charge states of +9 to +27 through ESI, demonstrating its ion current is split among nineteen channels. In general, as the mass of the analyte increases, the number of charge states, and thus the number of channels among which signal is distributed increases (14, 15). Whereas for VIP the most abundant charge



Figure 2.1 Full MS1 sprectra showing the charge state distributions of VIP and apomyoglobin. (A)A full MS spectrum depicting the ESI charge state distribution of VIP. VIP is comprised of 12 amino acid residues (MW ~1426 Da) and can carry 1 to 3 positive charges. (B) A full MS spectrum depicting the ESI charge state distribution of intact apomyoglobin from equine skeletal muscle. Apomyoglobin is comprised of 153 amino acid residues (MW ~17 kDa) and carries 9 to 27 positive charges.

state comprises over 50% of the total ion current, the most abundant charge state in apomyoglobin comprises only about 10% of the total ion current. Typically, only a single charge state is selected and isolated for fragmentation by ETD. For small peptides, CSD are narrow enough that ion current splitting is not detrimental to MS/MS analysis. However, for whole proteins, the signal for any given charge state is significantly less than that observed for a small peptide. Additionally, inject times for intact protein precursors tend to be significantly longer due to the lower signal, and the chemical noise co-isolated with the precursor contributes to higher over-all noise levels.

After a precursor ion is selected, isolated and fragmented by ETD, the resulting fragment ions can also exhibit multiple charge states, further diluting the possible ion current observed for each sequence informative fragment. This effect is intensified when working with intact proteins due to the increased number of sites that are capable of holding positive charge. Fragment ion signal dilution is caused by multiple gas-phase phenomena. One of the primary contributors involves differences in the distribution of charge in the precursor cation that result from ESI. Just as ESI produces precursor ions with a distribution of overall charge, ESI can produce ions with variations in the location

of those charges. A cartoon representation of this is shown in **Figure 2.2**. The black bar represents a large, highly-charged precursor with eighteen positive charges distributed along the backbone. In panel **A**, the charge shown in red appears in a different location than in panel **B**. After ETD, identical fragment ions are produced. However, the charge states of the fragment ions differ due to the initial location of the charge.



Figure 2.2 Cartoon depiction of how charge distribution in precursors affects fragment ion charge following ETD. The top bars of panels (A) and (B) show two precursors of identical m/z undergoing fragmentation by ETD. The charge neutralized following electron transfer is shown in yellow. Note the difference in the placement of the charge shown in red. The bottom bars of panels (A) and (B) show the fragment ions, c_m and z_n , produced from each precursor. When the charge indicated in red is located elsewhere, the resulting fragment ion charge states shift by one positive charge.

The ETD process is also capable of producing fragment ions that exhibit different charge states. ETD exploits polycyclic, aromatic, radical anions for electron transfer to an amide bond along the peptide backbone. This generates a carbonyl, radical anion that subsequently abstracts a proton from a nearby positively-charged residue. The resulting charge reduced species contains a radical site which triggers fragmentation of the N-C_{α} bond (4, 16). Since ETD requires that there positively charged residue in close proximity for fragmentation to occur, differences in fragment ion charge states can be the result of the location from which the proton is abstracted (**Figure 2.3**).

A second major source of ion current splitting is the number of heavy isotopes that are present at high precursor and fragment ion mass. Isotope patterns of molecules follow a binomial expansion (17). If there are 100 carbon atoms in a molecule, each of which has approximately 1.07% chance of being a ¹³C isotope, then the whole molecule is not unlikely to contain at least one heavy carbon isotope. It follows that larger molecules contain more ¹³C atoms and have wider isotopic envelopes than smaller ones. This distribution is easily observed using high resolution mass spectrometry; an example is shown in Figure 2.4 A and B. Figure 2.4 C shows the theoretical m/z, the portion of the total ion current, and the relative abundance of each isotopic peak for VIP and apomyoglobin based on natural abundances of ¹³C as 1.070% and ¹⁵N as 0.368% (Pacific Northwest National Laboratories, http://omics.pnl.gov/). According to this isotope pattern calculator, the ion current of the +3 charge state of VIP can be distributed among 7 signals of unique m/z. The most abundant signal, the monoisotopic peak, accounts for 44.55% of the total ion current attributed to the +3 charge state. The total ion current of the +18 charge state of apomyoglobin can be distributed among significantly more isotopic peaks, with the most abundant, in theory, accounting for only 11.92% of the total ion current. While illustrated here using VIP and intact apomyoglobin, the issue of ion current splitting through isotopic distribution is intensified when dealing with large fragment ions, especially those of low abundance as the reduction of SNR due to the presence of heavy isotopes is inversely proportional to the square root of the mass (15).



Figure 2.3 Fragment ion charge states are determined by the location of charge neutralization during ETD. Two reaction pathways are shown. In the pathway on the left, a proton is abstracted from an arginine residue. The pathway on the right shows a proton abstracted from a lysine residue. Differences in the location of charge neutralization result in fragment ions, c_m and z_n , of different charge states (x and y).



Vasoactive Intestinal Peptide C ₆₁ H ₈₉ N ₁₈ O ₂₂			Apomyoglobin C ₇₂₉ H ₁₂₁₃ N ₂₁₀ O ₂₁₈ S ₂		
m/z	Portion of Total	Relative Intensity	m/z	Portion of Total	Relative Intensity
475.8846	45.55%	100.00%	942.1720	0.01%	0.05%
476.2189	33.93%	74.49%	942.2277	0.05%	0.43%
476.5531	14.53%	31.89%	942.2834	0.24%	2.01%
476.8873	4.55%	9.98%	942.3391	0.76%	6.38%
477.2215	1.15%	2.52%	942.3948	1.83%	15.37%
477.5556	0.25%	0.54%	942.4506	3.57%	29.94%
477.8898	0.05%	0.10%	942.5063	5.85%	49.09%
478.2239	0.01%	0.02%	942.5620	8.30%	69.63%
			942.6177	10.40%	87.24%
			942.6734	11.69%	98.02%
			942.7291	11.92%	100.00%
			942.7848	11.15%	93.53%
			942.8405	9.64%	80.85%
			942.8962	7.75%	65.04%
			942.9519	5.84%	48.96%
			943.0076	4.13%	34.65%
			943.0633	2.76%	23.17%
			943.1190	1.75%	14.68%
			943.1747	1.05%	8.84%
			943.2303	0.61%	5.08%
			943.2860	0.33%	2.78%
			943.3417	0.17%	1.46%
			943.3974	0.09%	0.73%
			943.4531	0.04%	0.35%
			943.5088	0.02%	0.16%
			943.5645	0.01%	0.07%
			943.6202	0.00%	0.03%

Figure 2.4 Isotopic distributions of VIP and apomyoglobin. (A) Segment of a full MS1 spectrum showing the isotopic distribution of the +3 charge state of VIP. (B) Segment of a full MS1 spectrum showing the isotopic distribution of the +18 charge state of apomyoglobin. The resolution setting for both scans is r = 60,000 at 400 m/z and the number of ¹³C atoms present is indicated for some signals. (C) The columns to the left show the theoretical isotopic distribution of the +3 charge state of VIP. The column on the right shows the same for the +18 charge state of intact apomyoglobin. Chemical formulas for each molecule are also shown. Theoretical isotopic distributions were calculated using Isotope Pattern Calculator software (Pacific Northwest National Laboratories, http://omics.pnl.gov/).

In addition to the impacts that charge state and isotopic distributions have on SNR, there is the problem of limited charge capacities of the instrumentation. A simple cartoon depicting this is shown in **Figure 2.5**. Commercial QLTs, like those used in the Orbitrap Velos Pro^{TM} , can hold a maximum of approximately 1E6 charges. Proteolytic peptides like those generated by trypsin are typically 10-15 amino acids in length and the charge states of these peptides exhibited using ESI range from +1 to +3. If we consider the +3 precursor, the linear ion trap can hold 3.33E5 peptide molecules. For an intact protein, such as apomyoglobin (153 amino acid residues), carrying 30 positive charges by ESI, the linear ion trap can only hold 3.33E4 precursor protein molecules. It is important to note that for every electron transfer that occurs during ETD a charge is neutralized, and so, after ETD is performed, the trap is no longer at capacity.



Figure 2.5 Ion trap charge capacity limits the number of molecules analyzed. (A) Tryptic peptides typically carry charges of +2 or +3. (B) Intact proteins are much larger and carry significantly more charge. If the charge of the protein precursor is +30, ten-times fewer ions can be stored in the trap compared to tryptic peptides carrying a charge of +3. Additionally, after fragmentation by ETD, charge is divided among many more fragment ion channels for intact proteins relative to peptides (each color represents a unique fragment ion).

Figure 2.5 illustrates another challenge associated with analyzing whole proteins. The initial precursor ion current is divided among many more fragment ion channels for whole proteins compared to peptides. If we again consider VIP, a 12 residue peptide, the maximum number of sequence informative c- and z'-fragment ions produced by ETD is 22. Assuming we begin with 3.33E5 precursor molecules and all N-Cα bonds are cleaved with equal frequency, the maximum number of each unique fragment ion that can be produced by ETD is about 15,136. However, for apomyoglobin (153 residues, 304 unique c- and z'-fragment ions), assuming we begin with 3.33E4 precursor molecules, the maximum number of each unique fragment ion produced is just 109. This effect is compounded when we consider that each fragment ion can exhibit multiple charge states and that large fragments possess wider isotopic distributions as previously discussed.

The simplest and most widely employed approach for improving the low SNR observed in high resolution MS analyses of intact proteins involves summing (or averaging) the time-domain transients produced from the image current generated by the ions analyzed. In theory, the electronic noise component of a signal is random, and so signal is increased at a higher rate than noise when it is summed, thereby increasing the SNR. This is preferably done prior to Fourier transform (FT), as it slightly reduces the acquisition time of each spectrum and because ion signals can be allowed to accumulate to a level that is above the noise level, making previously unobservable signals visible. Typically, the amount of signal averaging/summing that is performed is commensurate to the amount allowed by the chromatographic timescale necessary for the analysis. The instrument control software allows the user to sum multiple transients during acquisition. Alternatively, after the data is acquired, scans can be averaged to produce the spectrum that is analyzed. In an extreme application to intact monoclonal antibody, IgG1, thousands of transients (up to 10,000) from up to 40 LC-MS/MS analyses were summed prior to FT resulting in SNR increases proportional to the square root of the number of transients summed (18). Each run took about 15 min to complete and required 20 picomoles (pmol) of sample. Combining the results from essentially 4 experimental setups (varying isolation windows, ETD reaction times and the number of transients summed) the authors achieved 32.7% coverage of the intact antibody using

ETD on an Orbitrap Velos Pro[™]. Compared to similar studies performed on monoclonal antibodies, the degree of coverage is doubled. However, the amount of time required to acquire the data, the amount of sample consumption and the computing power necessary to compile and analyze the data does not lend itself to applications in which the sample may be limited or complex or for which high throughput is important.

2.2.3 Spectral Complexity and Incomplete Fragmentation

Once all of the tools are in place to allow for the acquisition of high-quality spectra of intact proteins, focus must shift to spectral interpretation. ETD MS/MS spectra of intact proteins are very complex. Using short ETD times, most fragment ion signals overlap and appear in a relatively small cluster surrounding the precursor mass (Figure 2.6 A). Spectra like this are impossible to interpret manually. Many researchers use longer ETD times to further fragment the larger, more highly charged fragments and spread out the fragment ion signals (Figure 2.6 B). However, this is detrimental to SNR and to protein sequence coverage as first generation fragment ions can undergo multiple rounds of ETD under extended reaction times (Figure 2.7), decaying or destroying the signal of many sequence informative fragment ions. For apomyoglobin, which has 304 possible sequence informative c- and z-- fragment ions, over 11,000 internal fragments are possible. While the abundance of any one unique internal fragment ion is so low that it does not give an appreciable signal, taken together, internal fragments comprise a significant portion of the chemical noise observed in MS/MS spectra. Their creation comes at the expense of large sequence informative fragments which are essential for extending coverage. If we consider a 40 kDa protein, assuming a mass spectrometer with a 4000 m/z range is used, fragment ions must exhibit charge states through +10 for complete coverage in both the c- and z-ion series. The only way to achieve this is to

use very short ETD reaction times so that some of the first generation fragment ions are preserved.



Figure 2.6 ETD MS/MS spectra using 5 and 40 ms reaction times. (**A**) ETD MS/MS spectrum recorded on the $[M+26H]^{+26}$ (m/z 653.02) ion of apomyoglobin using a 5 ms ETD reaction time. Noise levels are at approximately 7% relative to the most intense fragment ion signal. (**B**) ETD MS/MS spectrum recorded on the $[M+26H]^{+26}$ (m/z 653.02) ion of apomyoglobin using a 40 ms ETD reaction time. Noise levels are at approximately 22% relative to the most intense fragment ion signal.



Figure 2.7 Multiple electron transfers create internal fragment ions. As ETD reaction times are extended the occurrence of multiple electron transfers to the more highly charged fragment ions increases. This results in the destruction of sequence informative high m/z fragment ions and the generation of internal fragment ions which contribute to chemical noise levels.

Top-down-based methodologies for whole protein characterization have only just begun to gain traction in the field of proteomics. As a result, the few bioinformatics tools available for the analysis of intact protein MS/MS spectra are underdeveloped. They are also limited in their applicability as *a priori* knowledge of their sequences is necessary to utilize the software. Proteins whose sequences have not yet been determined, contain PTMs, or those that result from splice variants and genes containing single nucleotide polymorphisms require, to at least some degree, manual interpretation of the MS data. Additionally, in order to site map PTMs that occur on the protein, near complete sequence coverage of the intact form must be attained. This cannot be achieved unless the protein is fragmented at almost every possible position.

Although ETD has significantly improved our ability to sequence whole proteins and large peptides, the fragmentation efficiency depends on several regional characteristics of the protein being fragmented. One area where ETD falls short is in fragmenting proteins with regions containing low or disparate charge distribution. This is because there must be a positively charged residue near the N-C_{α} bond to be cleaved in order for fragmentation to occur. Nondissociative electron transfer (ET), where the N-C α bond is

cleaved, but the fragment ion pair does not come apart, also results in poor sequence coverage (19). It is believed that these non dissociated fragment ion pairs are held together by noncovalent intramolecular interactions like hydrogen bonding and salt bridge formations (20). Finally, because a single charge is neutralized when electron transfer occurs, sufficient remaining positive charge must be available to detect both fragment ions (one charge on the N-terminal-containing ion and one on the C-terminal-containing ion). This is usually not a problem when fragmenting intact proteins due to the high charge state the precursor exhibits except in the extreme C-terminal region of the protein, where there is no guarantee that a charge will exist (compared to the N-terminus, which contains the α -amino group). Several groups have employed chemical derivatization to increase the charge on peptides to enhance ETD fragmentation by converting thiols or carboxylic acids to groups that contain a fixed charge (e.g. timethylamine) (21-24). This approach has merit, as fixed-charge tagging discourages non-dissociative electron transfer and provides charge for detection. Unfortunately, it falls short of making mobile protons available to facilitate cleavage of more N-C α bonds.

As an alternative to modifying the protein chemically to promote extensive fragmentation by ETD, other dissociative methodologies are being explored. One alternative, ultraviolet photodissociation (UVPD) (25, 26), was recently implemented on an Orbitrap Elite[™] mass analyzer (27). This technique uses an argon fluoride laser to emit photons at 193 nm which are strongly absorbed by the protein backbone, leading to fragmentation. A major advantage of UVPD over more widely used techniques is that the fragmentation efficiency is independent of the charge characteristics of the region to be fragmented. MS/MS spectra contained fragment ions indicating the cleavage of all bonds of the peptide backbone (types a-, b-, c-, x-, y- and z-ions), and some side chain loses (d-, v-, and w-ions). The authors achieved

near complete sequence coverage of intact ubiquitin, apomyoglobin and carbonic anhydrase (99%, 93% and 87% respectively). This coverage represents the best coverage reported on these intact proteins to date and while the authors certainly overcame the hurdle of incomplete fragmentation, their methods were met with other issues. The spectral complexity associated with fragmenting every possible bond along the peptide backbone is mind boggling. Piecing the sequence together from such a vast number of fragment ions requires *a priori* knowledge of the sequence. Determining the theoretical fragment ion masses for all the fragment ions observed requires exhaustive calculations and advanced search algorithms and bioinformatics tools would have to be developed for routine implementation of UVPD proteins with known sequences. *De novo* sequencing proteins fragmented by UVPD would be nearly impossible. Finally, SNRs were still an issue as the spectra analyzed were the result of up to 500 averaged spectra; a number that typically cannot be obtained on a chromatographic time-scale. Thus, the analysis of complex mixtures of proteins would be very difficult to perform.

The difficulties of analyzing intact proteins on a timescale compatible with chromatography are many in number. However, the task of simultaneously overcoming them all is a worthy feat. The quality of the data generated from a successful top-down MS experiment is unquestionably more valuable than bottom-up approaches because we are able to see exactly how proteins exist inside the cell. Here, in collaboration with other members of the Hunt Lab (notably A. Michelle English) and Thermo Fisher Scientific, we attempt to address these problems using combinations of novel mass spectrometry technology, gas-phase ion-ion chemistry, and chemical derivatization methods.

2.3 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)

Alltech Associates, Inc. (Deerfield, IL)

Acetyl chloride, anhydrous

D₀-Methanol, anhydrous

Anaspec, Inc. (Fremont, Ca)

Vasoactive Intestinal Peptide; human, porcine, rat; ≥95% purity

Branson (Danbury, CT)

Branson 1200 ultrasonic bath

Eppendorf (Hauppauge, NY)

5414R Benchtop centrifuge

GTS-Welco (Allentown, PA)

10 ppm SF₆ in nitrogen gas

Honeywell (Morristown, NJ)

Burdick and Jackson® Acetonitrile, LC-MS grade

Labconco Corporation (Kansas City, MO)

Centrivap centrifugal vacuum concentrator

Millipore (Billerica, MA)

Amicon® Ultra-0.5 10 K centrifugal filter

Molex (Lisle, IL)

Polymicro Technologies[™] polyimide coated fused silica capillary

Sizes: (360 µm o.d. x 25, 50, 75 & 150 µm i.d.)

Omega Engineering (Stamford, CT)

Temperature controllers

Parker Hannifin Corporation (Cleveland, OH)

Porter Back pressure regulator

Porter Mass flow controller

SGE Analytical Science (Melbourne, Australia)

PEEKsil tubing 1/16" o.d., 0.025 mm i.d., 100 mm length

Sigma Aldrich (St. Louis, MO)

2-(2-Aminoethyl)benzimidazole dihydrochloride, 97%

N-(2-Aminoethyl)maleimide trifluoroacetate salt, ≥95% (HPLC), ≥98% (T)

Apomyoglobin from equine skeletal muscle,

protein sequencing standard, lyophilized powder

N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride, crystalline

DL-Dithiothreitol, ≥98% (TLC), ≥99% (T)

Glacial acetic acid, ≥99.99% trace metal basis

Fluoranthene, >99% purity

2-propanol, LC-MS grade

Pyridine, CHROMASOLV® Plus, for HPLC, ≥99.9%

Trichloroacitic acid, ≥99.0%

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

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

DTSSP (3,3'-Dithiobis[sulfosuccinimidylpropionate])

Formic Acid, LC-MS Grade

Hydrochloric Acid, Fisher, Certified ACS Plus

Orbitrap Velos Pro[™] mass spectrometer

Pierce® Water, LC-MS Grade Phosphate buffered saline Urea

2.4 Methods

The first front-end reagent ion source was designed by Dr. Jeffrey Shabanowitz, Dr. Philip Compton, et. al. under the guidance of Professor Donald Hunt and patented in collaboration with Thermo Fisher Scientific in 2009 (28). Thermo created multiple prototypes for commercial use, which were tested and optimized at Thermo's U.S. headquarters in San Jose, CA and here at UVA on the Hunt lab LTQ-Orbitrap Velos Pro[™] beginning in 2012. The source is described in (28, 29) and briefly below. In June of 2013, Thermo made the reagent ion source, called EASY-ETD[™], commercially available on its Orbitrap Fusion[™] Tribrid[™] Mass Spectrometer.

In order to implement the front-end reagent ion source and perform the experiments described in this chapter, it was necessary to make changes to the instrument control software. This was done by Dr. A. Michelle English, a member of the Hunt Lab, and the changes are discussed briefly below. Also, the analog electrical circuit board was modified by Dr. John Syka, Thermo Fisher Scientific, to enable instrument control of the high voltage power supply (mentioned below). All experiments were performed using a commercial Orbitrap Velos ProTM.

2.4.1 Instrument Modification and Operation

Reagent delivery—The reagent delivery system is diagramed in **Figure 2.8**. Nitrogen gas is delivered to a mass flow controller operating at 11 sccm that regulates the total gas flow to the reagent ion source. The flow is split at the back pressure regulator. This ensures that a constant vapor pressure of ETD reagent (fluoranthene) is delivered to the source and that the appropriate pressure for a stable discharge is maintained. After the split, the low flow (~0.1 sccm) nitrogen travels through the heated reagent vial and picks up gaseous ETD reagent molecules. Separately, the IIPT reagent, sulfur hexafluoride (SF₆), is contained in a 10 parts per million (ppm) mixture with nitrogen gas. This mixture is introduced to the makeup gas line at a very low flow rate after the backpressure regulator (the flow is so low that the change to the total flow is negligible). The ETD and IIPT reagent containing nitrogen gas mixtures join at the ETD source heater interface and the combined flow is delivered to the ion volume where reagent ionization takes place.



Figure 2.8 Schematic of reagent gas delivery to the reagent ion source. A low flow of nitrogen passes through a heated reagent vial containing fluoranthene and a second high flow of nitrogen containing gaseous IIPT reagent (sulfur hexafluoride, SF_6) joins reagent flow at the ETD source heater interface. The combined flow is delivered to the ion source.

Reagent source—The exit lens of the stacked ring ion guide (SRIG) was modified to a thickness of 7 mm to house an ionization source. An ion volume (4 mm i.d.), which doubles as a lens, defines the reagent ionization region and serves as the anode. Press fit against the ion volume is a 4 mm o.d. x 3 mm i.d. ceramic spacer in which sits a stainless steel, discharge cathode. Nitrogen carrying reagent is delivered through a reagent inlet capillary to the ion volume, which is held at a low pressure (1-5 Torr). A current-controlled high voltage supply applies a potential across the electrodes creating an electrical discharge (glow discharge) and reagent ions are formed via interaction with the discharge inside the ion volume. The high voltage supply is switched on only during reagent injection. Upon ionization, reagent ions travel through an outlet to an area of lower pressure (1E-5 Torr) and are allowed to pass through the SRIG exit lens onto successive compartments of the mass spectrometer in the same path ions from the atmospheric pressure ionization (API) source follow. An early revision of the source is shown in **Figure 2.9** and the current source is shown in **Figure 2.10**. The instrument's analog board was modified to allow enable instrument control of the discharge potential.



Figure 2.9 SolidWorks rendering of an early reagent ion source for the SRIG interface. Reprinted with permission from (29) © 2013 American Chemical Society. While this is not the source currently in use, the image contains all the vital components (indicated) and is shown in an orientation that is easily translated (*) to the current source configuration.



Figure 2.10 Image of the current front-end reagent ion source. Image kindly provided by Dr. Chistopher Mullen, Thermo Fisher Scientific. The pieces shown here press-fit against one another in the order shown.

ITCL code changes & ETD/IIPT scan types—The instrument control software (ITCL) was modified to control the reagent ion source, inject reagent anions, enable ion/ion reactions within the QLT using the front-end reagent ion source, and allow storage of multiple iterations of ion/ion reaction products in the C-trap.

Multiple fragment ion fills—In order to increase fragment ion signals and enhance SNRs, multiple C-trap fills (termed 'multiple fills') were used. After each ion/ion (or series of ion/ion) reaction(s), product ions were transferred to and stored in the C-trap. This was repeated anywhere from 2-30 times, until the C-trap was filled to capacity. The accumulated fragment ions were then transferred to the Orbitrap for high resolution mass analysis.

2.4.2 Gas-Phase Ion/Ion Reactions

lon/lon proton transfer (IIPT) reactions using sulfur hexafluoride (SF₆) (30, 31) were used to charge reduce the fragment ions generated by ETD (which was performed using fluoranthene as the reagent). A proposed reaction mechanism is shown in **Figure 2.11**. SF₆ captures a thermal electron upon ionization within the discharge source to form SF₆⁻ which then acts as a base and abstracts protons from highly charged, ETD-produced fragments. The amount of charge reduction is tuned by varying the amount of time ion/ion reactions were allowed to take place (typically 20-160 ms). In order to perform ion/ion reactions, the ITCL was modified to incorporate the scan events diagrammed in **Figure 2.12** into the standard QLT MSⁿ scan function. This figure depicts the voltage changes occurring in the QLT during ETD and IIPT reactions when the reagent ion source described above is in use.



Figure 2.11 Mechanism for the reaction of sulfur hexafluoride with highly charged ETDproduced fragment ions. Sulfur hexafluoride radical anions are allowed to react with the cⁿ⁺ fragment ion, abstracting a proton, and reducing the charge of the fragment ion by one.



Figure 2.12 QLT voltages during ion/ion reactions. A diagram of the QLT is shown at the top of each panel to indicate the voltage settings on each of the three sections of the trap. Panels (**A**) and (**B**) show the voltage changes occurring in the QLT during ETD and IIPT reactions, respectively, after precursor ion selection and isolation: (1, 6) creation of a potential well for storage of precursor cations in the back section of the trap; (2, 7) reagent anion injection from the front-end source; (3, 8) reagent anion isolation via resonant ejection of all ions except the positively charged precursor and reagent anion; (4, 9) removal of the dc potential well and application of an rf voltage to the end lenses of the trap to allow positive and negative ion populations react; (5,10) termination of ion/ion reactions by removal of excess reagent anions via axial ejection of negatively charged reagent ions while retaining positive ions using a potential well created in the center section of the trap.

2.4.3 Sample Preparation

Amidation—One nanomole of lyophilized apomyoglobin was dissolved in a 20 µL solution containing 1 M each of pyridine, HCl and 2-(2-Aminoethyl)benzimidazole

dihydrochloride, pH \approx 5.5. A 5-µL aliquot of 0.1 M *N*-(3-Dimethylaminopropyl)-*N*ethylcarbodiimide hydrochloride in 1 M pyridine-HCI was added and the solution was sonicated for 2 hrs. The sample was then taken to dryness in a CentriVap® centrifugal vacuum concentrator, reconstituted in 200 µL 0.1% (v/v) acetic acid, and desalted using an Amicon® Ultra-0.5 10 kDa centrifugal filter. The sample was stored at -40 °C until analyzed.

Derivatization with DTSSP and aminoethyl maleimide—One nanomole of lyophilized apomyoglobin was dissolved in a 100 µL solution containing 6 M Urea, 100 mM PBS and 4 mM 3,3'-dithiobis[sulfosuccinimidylpropionate] (DTSSP), pH \approx 7.2 and allowed to cross-link for 2 hrs at 4 °C. The resulting disulfide bonds were reduced with 160 mM dithiothreitol (DTT) and alkylated by adding N-(2-aminoethyl) maleimide (32) to produce a 320 mM solution and allowing the reaction to proceed for 15 min at room temperature. The sample was desalted by precipitation with trichloro acetic acid, taken to dryness in a CentriVap® centrifugal vacuum concentrator and stored at -40 °C until analyzed.

Esterification—One nmol of lyophilized apomyoglobin was esterified using standard Fischer esterification techniques (33). The sample was then taken to dryness in a CentriVap® centrifugal vacuum concentrator and stored at -40 °C until analysis.

Sample preparation for direct infusion—Immediately prior to MS analysis, samples (unmodified and derivatized apomyoglobin) were dissolved in a 200-500 μ L solution containing 40% acetonitrile and 60%, 0.1% acetic acid (all percentages expressed as v/v). Samples were infused at concentrations of 2-5 pmol/ μ L.

2.4.4 MS Analysis

Unmodified apomyoglobin—Apomyoglobin (2 pmol/µL) was directly infused using a fused silica, microcapillary column with a laser-pulled electrospray emitter tip and ionized using microelectrospray ionization into the front-end ETD/IIPT-enabled Orbitrap Velos

Pro[™]. Mass analyses were performed by targeting a 6 m/z window centered at the $[M+26H]^{+26}$ charge state (m/z = 653.08) for high resolution (100,000 at 400 m/z) MS/MS interrogation (200-4000 m/z scan range, 5E4 ion counts FTMSⁿ automatic gain control (AGC) target, 3E5 ion counts reagent AGC target, 2-30 multiple fills, 5 ms ETD, 20-160 ms IIPT).

Aminoethyl benzimidazole derivatized apomyoglobin—Sample was directly infused at a concentration of 5 pmol/µL as described above. Mass analyses were performed by targeting a 15 m/z window centered at m/z 867 that contained a highly charged, derivatized form of the protein. MS/MS spectra were acquired at high resolution (60,000 at 400 m/z, 200-4000 m/z scan range, 2E5 ion counts FTMSⁿ and reagent AGC targets, 10 multiple fills, 7 ms ETD and 60-120 ms IIPT).

For LC-MS analysis, derivatized protein (1 pmol) was pressure loaded onto 360-µm o.d. x 150-µm i.d. fused-silica micro-capillary pre-column (packed 11-cm of POROSHELL 300SB-C18 resin) and washed with ~20 column volumes of 0.3% (v/v) formic acid in water. The pre-column was then connected to a 360-µm o.d. x 75-µm i.d. fused-silica micro-capillary pre-column packed similarly to the pre-column and equipped with a laser-pulled electrospray emitter tip (34). Samples were gradient eluted by nanoflow (60 nL/min) reverse-phase HPLC and ionized using micro electrospray ionization as previously described (35).The elution gradient utilized solvent A: 0.3% formic acid in water and solvent B: 0.3% formic acid, 72% acetonitrile, 18% isopropanol and 9.7% water (all v/v). The instrument was set to toggle FT (r = 60,000 at 400 m/z) and IT full MS scan types over a 300-2000 m/z scan range.

DTSSP-aminoethyl maleimide derivatized & esterified apomyoglobin—samples were directly infused at a concentration of 5 pmol/ μ L as described above. Mass analyses were performed by targeting a 10 m/z window centered around m/z 848. High resolution MS/MS scans (r = 60,000 at 400 m/z) were acquired (200-4000 m/z scan range, 1E5 ion

counts precursor FTMSⁿ AGC target, 2E5 ion counts reagent AGC target, 4 multiple Ctrap fills, 3 microscans each of data generated from 5-10 ms ETD and 80-100 ms IIPT).

Data analysis—Interpretation of all ETD/IIPT MS/MS data was performed manually on the unprocessed raw spectra. Theoretical fragment ion masses were calculated using software developed in-house (Fragment Mass Calculator). Percent sequence coverage was calculated by dividing the number of observed N-Cα bond cleavages by the total number of predicted N-Cα bond cleavages (note that cleavage of the N-Cα bond that is N-terminal to proline does not produce an observable fragment). ProSight PC 3.0 (Thermo Fisher Scientific) was used to evaluate the effect of IIPT reactions on the number of fragment ion assignments the software was able to make. Fragment mass tolerances were set to 10 ppm for all searches. Searches were performed against all equine proteins contained in the 04/23/2010 release of Refseq (36).

2.5 Results & Discussion

2.5.1 Multiple Fills Increase Ion Current of Fragment Ions

The challenges of sequencing intact proteins by mass spectrometry were outlined in the introduction of this chapter. Recall that chief among them was the low SNRs inherent to the analysis of whole proteins. In addition to the limited number of precursor ions that can be held in the QLT prior to fragmentation, ETD and IIPT reactions are destructive of charge and the remaining ion current is divided among many more channels, making each fragment signal weak relative to noise. To improve SNRs, total fragment ion current must be increased. This can be accomplished by performing multiple fragment ion fills of the C-trap prior to high resolution mass analysis. To do this, reagent anions must be introduced from the front of the instrument. The commercial instrument configuration is shown in **Figure 2.13**, with the reagent ion source located in

the rear of the instrument. In this setup, reagents must traverse the C-trap to reach the high pressure QLT, where the ion/ion reactions are performed. We modified a Thermo Scientific Orbitrap Velos Pro[™] to introduce reagent anions from the front by adding a glow discharge ionization source between the stacked ring ion guide and multiple 00. The discharge is used to ionize ETD and IIPT reagents (fluoranthene and sulfurhexafluoride, respectively). By placing the reagent ion source in the front of the instrument, the C-trap is made available to store fragment ions from multiple iterations of ion/ion reactions. Once a sufficient amount of fragment ions are collected, they are sent to the Orbitrap for high resolution mass analysis. In this way, we are able to increase the ion current we observe for fragment ions, drastically improving SNRs. Figure 2.14 shows ETD/IIPT MS/MS spectra of intact apomyoglobin using varying numbers of C-trap fills. Analyzing just a single C-trap fill, which represents a normal scan function, the noise levels are approximately 24% relative to the most intense signal. As the number of C-trap fills analyzed increases, the fragment ion current rises, and the relative noise level drops. Fragment ion signals that are not visible in **A** are clearly observed in **D**. The number of fills performed can be changed for each type of analysis, depending on the size of the protein being analyzed or the amount of charge reduction performed by



Figure 2.13 Instrument diagram of the Orbitrap Velos Pro™. The configuration of the commercial instrument is shown with the reagent ion source in the rear. Reagents must traverse the C-trap before reaching the high pressure QLT, where ETD and IIPT reactions are carried out. The new front-end reagent ion source was installed between the SRIG and MP00.



Figure 2.14 Multiple Fills Increase Ion Current of Fragment Ions. MS/MS spectra of intact apomyoglobin following ETD (5 ms) and IIPT (160 ms) reactions with various numbers of fragment ion fills of the C-trap. Arrows indicate the approximate noise level relative to the most intense fragment ion signal. (**A**) Spectrum produced from a single iteration of ion-ion reactions (1 fill, 1.34 sec elapsed scan time). (**B**) Spectrum produced from 5 iterations of ion-ion reactions (5 fills, 3.14 sec elapsed scan time). (**C**) Spectrum produced from 10 iterations of ion-ion reactions (10 fills, 5.39 sec elapsed scan time). (**D**) Spectrum produced from 20 iterations of ion-ion reactions (20 fills, 9.88 sec elapsed scan time).

IIPT reactions. This is important because a smaller protein requires fewer multiple fills than a larger protein since there will be more precursor ions prior to fragmentation and the fragment ion current is not split among as many channels. When extended IIPT reaction times are used, such that fragment ion charge is reduced to +1 and +2, more multiple fills should be used because a large amount of charge is destroyed, yielding a relatively empty C-trap, and by extension Orbitrap, following ion/ion reaction.

The multiple fill strategy for improving SNRs has a number of advantages over transient summing or spectral averaging. First, averaging spectra acquired on single fragment ion fills is not useful when the signal being averaged is below the limit of detection. Averaging single fills serves only to decrease noise relative to the detected signal. Performing multiple fills increases the signal above both the limit of detection and above the noise level, making previously unobservable signals visible. When spectra or transients are averaged, the observed SNR increases as the square root of the number of transients averaged (100 averaged transients affords an SNR enhancement of 10). When multiple C-trap fills are used, signal, and by extension SNRs, increases fairly linearly with the number of fills until the C-trap capacity is reached (**Figure 2.15**).

Additionally, the time that it takes to acquire spectra of multiple C-trap fills is significantly less than the time it takes to acquire spectra from the same number of single C-trap fills to be averaged or summed. Figure 2.16 shows the time required to acquire a high resolution MS/MS spectrum broken into 3 categories: 1) ion handling time (which includes precursor and reagent ion injection, isolation, activation, and scan overhead): 2) transient acquisition (which increases with higher resolution settings); and 3) transient Fourier transform. Depending on the resolution used, transient acquisition can take up to 1080 ms (r = 100,000 at m/z = 400) and comprises the majority of the total scan duty cycle. Compared to transient acquisition, ion handling comprises only a fraction of that time. Fourier transform of the transient signal takes even less time to perform. Panel A shows the acquisition time of 10 high resolution spectra. Panel B illustrates the slight reduction of the total spectral acquisition time multiple transients are acquired (microscans are performed). These transients are summed before they are Fourier transformed for a SNR improvement of $\sqrt{10}$ or 3.16. Panel **C** depicts the time it takes to acquire a single spectrum of 10 C-trap fragment ion fills. This scan cycle takes less than one third the time of **B** and results in an SNR improvement of 10 (assuming the C-trap is not filled to capacity in less than 10 C-trap fills). Thus, the time it takes to acquire spectra of multiple C-trap fills is significantly less than the time it takes to acquire spectra for the same number of single-fills to be summed or averaged. This effect is



Figure 2.15 Fragment ion current as a function of the number of C-trap fills analyzed. (A) Chromatogram showing the relative abundance of the total (fragment) ion current detected by the Orbitrap as the number of C-trap fragment ion fills increases. The precursor is the +26 charge state of apomyoglobin. ETD and IIPT reaction times are 5 and 40 ms, respectively. The apparent signal instability is caused by a loss of precursor ion inject time predictability as more C-trap fills are enabled, which exacerbates ESI spray instability. (B) A graph showing the average total fragment ion current as a function of the number of C-trap fills analyzed from (A).



Figure 2.16 Elapsed scan time comparisons for the analysis of 10 ion volumes. (A) Total elapsed scan time required to acquire 10 spectra, each of one C-trap fragment ion fill. (B) Total elapsed scan time to produce a single spectrum from 10 summed transients (10 microscans). (C) Total elapsed scan time to produce a single spectrum from 10 fragment ion fills of the C-trap.

augmented when considering the averaging/analysis of a larger number of C-trap fill equivalents. Additionally, the SNRs improve significantly faster (linearly vs. as the square root of the number scans averaged or summed), and so fewer spectra are needed for averaging to produce a useful spectrum for analysis. This gives us the ability to acquire high-quality spectra on a chromatographic time-scale.

2.5.2 Gas-phase Ion/Ion Reactions Yield Near Complete Sequence Coverage of Apomyoglobin

Shown in **Figure 2.17** is the MS/MS spectra of the $[M+26H]^{+26}$ (m/z 653) charge state of apomyoglobin (average MW 16,952.52 Da) following 5 ms ETD using fluoranthene radical anions as the reagent. For apomyoglobin, a protein of 153 amino acids, there are 152 possible cleavage sites and 304 possible fragments of type c and z. These fragments are largely concentrated in a 200-300 m/z range surrounding the precursor m/z, with many fragment ion signals overlapping one another, making their

isotope patterns indistinguishable. Fragment ion signals outside of the large cluster are sparse. Under these conditions, the majority of the 200-4000 m/z scan range is not utilized (>1500 m/z). Additionally, the fragment ions exhibit multiple charge states from +1 to over +20, compounding the complexity of the spectrum. To simplify the spectrum, we employed IIPT reactions using sulfur hexafluoride radical anions to charge reduce the complex mixture of fragments. The SF_6^{-1} functions as a Brønsted base to remove protons from multiply charged fragment ions produced by ETD. The degree of charge reduction can be tuned by varying the amount of time the reagent anions are allowed to react. Since the rate of an ion/ion reaction is proportional to the square of the substrate charge (+10 ions react 100 times faster than +1 ions), it is possible to preferentially deprotonate highly charged ions. In this way, the amount of charge reduction is "normalized". By extending IIPT reaction times such that fragment ions are significantly charge reduced, the fragment ion signals are increasingly distributed over the full scan range (**Figure 2.17 B-E**).



Figure 2.17 ETD/IIPT MS/MS spectra recorded on intact apomyoglobin. (A) ETD spectrum recorded on $(M+26H)^{26+}$ ions from apomyoglobin using a reaction time of 5 msec. (B-E) Spectra obtained by performing IIPT reactions on the ETD fragment ions in (A) for 20 ms, 40 ms, 80 ms, and 160 ms, respectively.
Figure 2.18 depicts the degree to which the spectrum is simplified using extended IIPT reaction times. In panel **A**, the 1000-1500 m/z scan range of an MS/MS acquired following 5 ms of ETD and 160 ms IIPT is shown. Recall that approximately half of this scan range goes unused when IIPT reactions are not utilized. This extended IIPT reaction time (160 ms) produces fragment ions that are mostly singly and doubly charged. In this region, sequence information derives from ~10-27 residues from both protein termini. Panel **B** shows the same m/z range from a spectrum acquired following 5 ms ETD and 20 ms IIPT. Fragment ions contain anywhere from 3 to 7 charges and the sequence information derives from residues ~37-90 in both fragment ion series. While the spectrum is rich in sequence information, it is quite complex. Piecing the sequence together from such a vast number of fragment ions requires *a priori* knowledge of the sequence.



Figure 2.18 Fragment ions observed with long and short IIPT reaction times. MS/MS spectra recorded on $[M+26H]^{26+}$ ions from apomyoglobin using an ETD reaction time of 5 ms and (A) an IIPT reaction time of 160 ms or (B) an IIPT reaction time of 20 ms. Only the 1000-1500 m/z region is shown.

A 5 ms ETD only

c1 - GtL - S - D - GTE - W - Q - Q - V - LTN - V - W - G - K - VTE - ATD - I - A - GTHTG - z129 c26 - Q - E - VTL - I - RTL - FTTTGTH - P - E - TTLTETK - FTDTKTFTKTH - L - KTz104 c51 - T - ETATETM - K - A - STETDTLTKTKTH - G - T - V - VTL - TTATL - GTGTI - z79 c76 - LtKTK - K - G - H - H - ETA - E - LTK - P - L - A - QTSTHtA - T - KtH - K - I - P - z54 c101 - I - KtY - L - E - F - I - S - DtA - I - I - HtV - LtH - StK - H - P - GtD - F - G - A - z29 c126 - DtAtQtGtA - MtT - KtA - LtE - LtFtRtNtDtI - AtAtK - Y - K - E - L - G - z4 c151 tF - Q - G - z1 B 5 ms ETD + 20 ms IIPT

 c1
 -G-L-S+D-G+E-W-Q-Q-V-L+N+V-W+G-K-V+E-A+D-I+A+G+H+G-z129

 c26
 -Q-E-V-L-I-R+L-F+T+G-H-P-E+T+L+E+K+F+D+K+F+K+H+L+K-z104

 c51
 -T-E-A+E+M+K-A-S+E+D-L+K+K+H-G+T-V-V-L+T+A+L+G+G+I-z79

 c76
 £L+K+K+K+G-H-H+E+A+E+L+K-P-L-A-Q-S+H+A-T+K+H-K-I-P-z54

 c101
 ·I+K+Y+L+E-F-I-S+D+A-I+L+I-H+V-L+H+S+K-H-P-G+D+F-G-A-z29

 c126
 ·D+A+Q+G+A-M+T+K+A+L+E-L+F+R+N+D-I-A-A+K+Y+K-E-L-G-z4

 c151
 ·F-Q-G
 ·Z1

C 5 ms ETD + 40 ms IIPT

 c1
 ·G
 -L-SID-GIE-W-Q-Q-V-L-N-V-WIG-K-VIEIAID-IIA-GIH-G-Z129

 c26
 ·Q-EIV-L-I-RIL-FITIGIH-P-EITILIEIKIFIDIKIFIKIHLIKIZ104

 c51
 ·TIEIAIEIMIK-A-SIEIDILIKIKIHIGIT-V-V-LITIAILIGIGIIIZ79

 c76
 LLKKKKKG-HIHIEIAIEILIK-P-L-AIQISIHIA-TIKKH-KtI-P-Z54

 c101
 ·IKIYILIEIF-IISIDIA-I-I-HIV-LIHISIK-H-P-GIDIF-G-A-Z29

 c126
 ·DIAIQIG-A-M-TIKIALLELLETERINIDII-AIAIKIYIK-E-L-G-Z4

 c151
 ·F-QIG

Figure 2.19 Apomyoglobin sequence coverage determined by analysis of ETD and ETD/IIPT MS/MS spectra with ProSight PC[™] software. (A) Sequence coverage from 5 ms of ETD only. (B) Sequence coverage from 5 ms of ETD & 20 ms of IIPT. (C) Sequence coverage from 5 ms of ETD & 40 ms of IIPT. Searches were performed using ProSightPC 3.0 (Thermo Fisher Scientific). Fragment mass tolerance was set to 10 ppm.

Use of IIPT reactions also improves the number of fragment ion assignments that can be made with search algorithms such as ProSightPC (37, 38). Since the fragment ions no longer overlap, they are more easily deconvoluted and more sequence coverage is obtained compared to the ETD only MS/MS spectra. As shown in **Figure 2.19** the total number c- and z-ions detected by ProSightPC for 5 ms ETD only, 5 ms ETD & 20 ms IIPT and 5 ms ETD & 40 ms IIPT is 65 to 95 to 122 respectively.

Proteins with undetermined sequences, those that contain PTMs, or those that result from splice variants and genes containing single nucleotide polymorphisms require, at least to some degree, manual interpretation of the MS/MS data. For this, the extended IIPT reaction times are optimal, as they produce significantly less dense spectra for *de novo* sequencing. Of course the larger the protein, the more initial charge it carries and the more charge its fragment ions must possess to achieve full coverage of the intact form (under the scan range limit of 4000 m/z). Thus, it follows that optimal IIPT reaction times must be scaled according to the size of the protein analyzed.

To *de novo* sequence an intact protein or site map PTMs that occur on a protein, complete sequence coverage must be obtained. Shown in **Figure 2.20** is a heat map of the manually determined sequence coverage obtained from intact apomyoglobin when IIPT reaction times are incrementally decreased from 160 ms. Nearly full sequence coverage is achieved from the 160 ms IIPT reaction time alone. Almost a continuous series of singly, doubly and triply charged ion in both the c and z• series are observed for ~83-86 residues from the N-terminus. When shorter IIPT reaction times are used, significant sequence overlap in the c- and z-ion series is observed. Contiguous fragment ion series cover the more central regions of the protein and sequence information from either protein terminus can be obtained for up to ~110-115 residues. If the sequence information from all IIPT reaction times is combined, 86% coverage is achieved. Noticeably absent from this set of data are sequence assignments for a few residues at the N- and C-termini of the protein. These regions are devoid of basic residues and thus fail to carry a charge. Additionally, low-mass cutoff limitations require that extreme N- and C-terminal fragment ions be at least 200 m/z for detection.



Figure 2.20 Heat map of the sequence coverage generated from intact apomyoglobin using sequential ETD/IIPT reactions. Fragment ions observed with 5 ms of ETD and 160 ms IIPT are lightly shaded and dominated by charge states of +1, +2 and +3. More highly charged fragment ions are observed for the first time at shorter IIPT reaction times (80, 40 and 20 ms) and are labeled in darker shades. Those shaded in black were only observed following chemical derivatization of the protein.

2.5.3 Protein Derivatization to Facilitate Sequence Analysis at the N- and C-termini of Proteins

To enhance the formation and detection of ETD fragment ions at the N- and Ctermini of the protein, we employed two chemical derivatization methods. The first involves using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) chemistry to convert carboxylic acids groups at the C-terminus and on the side chains of Asp and Glu residues to amides containing a basic imidazole group. In the gas-phase, these imidazole groups contain a mobile proton. As proof of principle, ETD spectra recorded on the highly-acidic, human CLIP peptide, before and after derivatization with histamine are shown in **Figure 2.21**. Panel **A** contains the ETD MS/MS spectra of the most abundant charge state of the unmodified peptide ($[M+3H]^{+3}$, m/z 822.91). The spectrum contains very few fragment ions and is dominated by the charge reduced $[M+3H]^{+++}$ precursor species. Panel **B** shows the ETD MS/MS spectra of the most abundant change state of the derivatized peptide ($[M+7H]^{+7}$, m/z 433.28), which contains six histamine labeled carboxylic acids. In contrast, this spectrum is now dominated by c and z+ type ions that provide full sequence coverage (panel **C**).

This same approach was employed to derivatize carboxyl groups of apomyoglobin with 2-(2-aminoethyl) benzimidazole. Aminoethylbenzimidazole is chemically similar to histamine, but has a mass that allows even the smallest amino acid, Gly, to be detected at the C-terminus of the protein. C-terminal carboxylic acid derivatization using the benzimidazole makes the lightest possible z_1^{\dagger} ion (59 m/*z*, unprotonated) protonatable with an m/*z* of 203 (**Figure 2.22 A**), thereby enabling its detection. We found that while labeling of the intact protein was incomplete (apomyoglobin contains 13 glutamic acid residues, 8 aspartic acid residues and the C-terminus) labeling of the first six acidic residues from the C-terminus was very efficient. As seen with the CLIP peptide, more extensive fragmentation is observed in this region after derivatization, with a continuous



Figure 2.21 Derivatization of human CLIP peptide with histamine enhances fragmentation observed by ETD. (A) ETD MS/MS spectrum of $[M+3H]^{3+}$ ions (m/z 822.91) from unmodified CLIP. (B) ETD MS/MS spectrum of $[M+7H]^{7+}$ ions (m/z 433.27) of CLIP derivatized with histamine. Following derivatization, carboxyl groups of five D and E residues, plus the C-terminus are converted to amides containing a basic imidazole ring. (C) Sequence coverage obtained from (B). Residues labeled with a histamine group are indicated by an asterisk. Spectra kindly provided by Dr. A. Michelle English.



Figure 2.22 Protein modification strategies for observing extreme N- and C-terminal fragment ions and for enhancing fragmentation by ETD. (A) Derivatization of the C-terminal carboxyl group with aminoethylbenzimidazole. (B) Derivatization of the amino terminus with a protected thiopropionic acid hydroxysuccinimide ester followed by reaction with aminoethylmaleimide.

series of z-ions observed through the first 39 residues C-terminal residues (**Figure 2.23 A**). Newly observed fragment ions are indicated in black in **Figure 2.20**. Additionally, the overall charge state distribution of the protein increases; the +19 being the most abundant charge state of the unmodified protein, and the +25 charge state being the most abundant in the derivatized protein (**Figure 2.24 A** and **B**). HPLC-MS experiments revealed that 15-22 carboxyl groups are successfully labeled. The most abundant form of the modified protein contains 18 aminoethylbenzimidazole units.

To observe fragment ion signals for the first two N-terminal amino acid residues of apomyoglobin, the α -N-amino and lysine side chain amino groups were cross-linked with a thiopropionic acid hydroxysuccinimide ester. The resulting disulfide bonds were reduced and alkylated with basic aminoethylmaleimide (**Figure 2.22 B**). Successful labeling of the N-terminus increases the m/*z* of the lightest possible c₁⁺ ion from 75 to 303 and the presence of the ethylamino group from the maleimide ensures that it remains charged for detection after derivatization. **Figure 2.23 B** shows the ETD/IIPT

spectrum recorded on the derivatized sample. Type c ions are now visible for the first 6 amino acid residues. With data from both derivatizations, total sequence coverage for apomyoglobin is 94%. Newly observed fragment ions are indicated in black in **Figure 2.20**.



Figure 2.23 Sequence coverage of the N- and C-terminus of apomyoglobin following derivatization. (A) 200-1700 m/z subsection of an ETD/IIPT MS/MS spectrum of apomyoglobin derivatized with aminoethylbenzimidazole. Observed z-ions are indicated. (B) 250-1000 m/z subsection of an ETD/IIPT MS/MS spectrum of apomyoglobin derivatized with DTSSP-maleimide. Observed c- and z-ions are indicated. Labeled amino acid residues are given by an asterisk. ¥ denotes fragment ions not carrying a covalent label. Recall that c- and z-ions are not observed for cleavage N-terminal to proline residues.



Figure 2.24 Electrospray ionization mass spectra of unmodified and modified apomyoglobin. Distribution of charge states observed for (A) apomyoglobin, (B) apomyoglobin derivatized with aminoethylbenzimidazole and (C) esterified apomyoglobin.

We were also interested in mitigating the effects of potential salt-bridge formation on backbone fragmentation by ETD. To do this, the carboxylic acids of intact apomyoglobin were esterified with methanol. The resulting mixture was directly infused into the mass spectrometer. Esterification proved to be the most efficient derivatization technique. The most abundant modified form has all carboxylic acids converted to methyl esters. A few fragment ions near the N- and C-terminus that are not observed when the unmodified protein undergoes ETD are seen, but none in addition to what was gained from the previous derivatization reactions (data not shown). The fact that increased N- C α bond clevages were not observed suggests that salt bridge formation is a big contributor to non-dissociative electron transfer in apomyoglobin. The charge state distribution of esterified apomyoglobin was shifted to higher charge states, the most abundant being the +24 compared to the +18 for the unmodified protein (**Figure 2.24 C**).

2.6 Conclusions

Implementation of a front-end ETD source on an Orbitrap Velos Pro[™] instrument is described. This source enables multiple fills of the C-trap with product ions from ETD of intact proteins prior to high resolution mass analysis. The result is a dramatic enhancement of the observed fragment ion current and significant improvement in SNRs in MS/MS spectra. This eliminates the need for the acquisition of several time consuming transients. Sequential ETD and IIPT reactions can be used to simplify ETD MS/MS spectra and to disperse ETD fragment ions over the entire mass range in a controlled manner. We used these gas-phase reactions to obtain comprehensive (86%) sequence coverage of apomyoglobin. Covalent labeling strategies were employed with the goal of selectively enhancing the sequence information observed at the N- and C-termini of the protein and overcoming low-mass cutoff limitations. While these

derivatization reactions were less efficient in labeling targeted groups in the internal region of the protein, targeted groups near the N- and C-terminus are efficiently labeled based on the fragment ions observed in the MS/MS spectra. In the case of the aminoethylbenzimidazole derivatization, fragmentation of the C-terminal region of apomyoglobin was enhanced, resulting in the observation of several previously unseen fragment ions. The DTSSP-maleimide derivatization was successful in tagging the α -Namino group so that the first six C-terminal fragments were visible. Though not applicable here, the aminoethylmaleimide can be used to alkylate cysteine free thiols with basic sites. Esterification and derivatization with the aminoethylbenzimidazole were found to increase the charge on the protein significantly and thus facilitate more complete fragmentation by ETD. Using the additional information gleaned from these derivatization approaches, we achieved 94% sequence coverage of intact apomyoglobin compared to ~86% for the underivatized protein. These derivatization reactions are applicable to peptides and whole proteins alike. Future methods will focus on controlling ETD reaction rates and preventing product ions from undergoing secondary reactions. We hope that by successfully slowing the reaction, near complete sequence analysis of larger proteins (~50,000 Da) will soon be a reality. In the next chapter, we discuss the application of some of these methods to the analysis of intact histories, a class of proteins for which these techniques are particularly well suited.

2.7 References

1. What does the future hold for top down mass spectrometry? Garcia B. A., <u>J Am Soc</u> <u>Mass Spectrom</u> **2010** (21): 193-202.

2. Top-down mass spectrometry: Recent developments, applications and perspectives, Cui W., Rohrs H. W. & Gross M. L., <u>Analyst</u> **2011** (136): 3854-3864.

3. Electron capture dissociation of multiply charged protein cations. A nonergodic process, Zubarev R. A., Kelleher N. L. & McLafferty F. W., <u>J Am Chem Soc</u> **1998** (120): 3265-3266.

4. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry, Syka J. E., Coon J. J., Schroeder M. J., Shabanowitz J. & Hunt D. F., <u>Proc</u> <u>Natl Acad Sci U S A</u> **2004** (101): 9528-9533.

5. Simplification of product ion spectra derived from multiply charged parent ions via ion/ion chemistry, Stephenson J. L., Jr & McLuckey S. A., <u>Anal Chem</u> **1998** (70): 3533-3544.

6. Phosphorylation site identification via ion trap tandem mass spectrometry of whole protein and peptide ions: Bovine alpha-crystallin A chain, Hogan J. M., Pitteri S. J. & McLuckey S. A., <u>Anal Chem</u> **2003** (75): 6509-6516.

7. Whole protein dissociation in a quadrupole ion trap: Identification of an a priori unknown modified protein, Amunugama R., Hogan J. M., Newton K. A. & McLuckey S. A., <u>Anal Chem</u> **2004** (76): 720-727.

8. Protein identification using sequential ion/ion reactions and tandem mass spectrometry, Coon J. J., Ueberheide B., Syka J. E., Dryhurst D. D., Ausio J., Shabanowitz J. & Hunt D. F., <u>Proc Natl Acad Sci U S A</u> **2005** (102): 9463-9468.

9. Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry, Chi A., Bai D. L., Geer L. Y., Shabanowitz J. & Hunt D. F., <u>Int J Mass Spectrom</u> **2007** (259): 197-203.

10. The orbitrap: A new mass spectrometer, Hu Q., Noll R. J., Li H., Makarov A., Hardman M. & Graham Cooks R., <u>J Mass Spectrom</u> **2005** (40): 430-443.

11. Orbitrap mass analyzer--overview and applications in proteomics, Scigelova M. & Makarov A., <u>Proteomics</u> **2006** (6 Suppl 2): 16-21.

12. Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer, McAlister G. C., Phanstiel D., Good D. M., Berggren W. T. & Coon J. J., <u>Anal Chem</u> **2007** (79): 3525-3534.

13. A proteomics grade electron transfer dissociation-enabled hybrid linear ion traporbitrap mass spectrometer, McAlister G. C., Berggren W. T., Griep-Raming J., Horning S., Makarov A., Phanstiel D., Stafford G., Swaney D. L., Syka J. E., Zabrouskov V. & Coon J. J., <u>J Proteome Res</u> **2008** (7): 3127-3136.

14. Electrospray ionization for mass spectrometry of large biomolecules, Fenn J. B., Mann M., Meng C. K., Wong S. F. & Whitehouse C. M., <u>Science</u> **1989** (246): 64-71.

15. On the scalability and requirements of whole protein mass spectrometry, Compton P. D., Zamdborg L., Thomas P. M. & Kelleher N. L., <u>Anal Chem</u> **2011** (83): 6868-6874.

16. Probing the mechanism of electron capture and electron transfer dissociation using tags with variable electron affinity, Sohn C. H., Chung C. K., Yin S., Ramachandran P., Loo J. A. & Beauchamp J. L., <u>J Am Chem Soc</u> **2009** (131): 5444-5459.

17. Relative abundance calculations for isotopic molecular species, Margrave J. L. & Polansky R. B., <u>J Chem Educ</u> **1962** (39): 335.

18. Analysis of intact monoclonal antibody IgG1 by electron transfer dissociation orbitrap FTMS, Fornelli L., Damoc E., Thomas P. M., Kelleher N. L., Aizikov K., Denisov E., Makarov A. & Tsybin Y. O., <u>Mol Cell Proteomics</u> **2012** (11): 1758-1767.

19. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors, Swaney D. L., McAlister G. C., Wirtala M., Schwartz J. C., Syka J. E. & Coon J. J., <u>Anal Chem</u> **2007** (79): 477-485.

20. Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins, Horn D. M., Ge Y. & McLafferty F. W., <u>Anal Chem</u> **2000** (72): 4778-4784.

21. Enhanced electron transfer dissociation through fixed charge derivatization of cysteines, Vasicek L. & Brodbelt J. S., <u>Anal Chem</u> **2009** (81): 7876-7884.

22. Global amine and acid functional group modification of proteins, Krusemark C. J., Ferguson J. T., Wenger C. D., Kelleher N. L. & Belshaw P. J., <u>Anal Chem</u> **2008** (80): 713-720.

23. Modifying the charge state distribution of proteins in electrospray ionization mass spectrometry by chemical derivatization, Krusemark C. J., Frey B. L., Belshaw P. J. & Smith L. M., <u>J Am Soc Mass Spectrom</u> **2009** (20): 1617-1625.

24. Chemical derivatization of peptide carboxyl groups for highly efficient electron transfer dissociation, Frey B. L., Ladror D. T., Sondalle S. B., Krusemark C. J., Jue A. L., Coon J. J. & Smith L. M., <u>J Am Soc Mass Spectrom</u> **2013** (24): 1710-1721.

25. Peptide sequence analysis by laser photodissociation fourier transform mass spectrometry, Hunt D. F., Shabanowitz J. & Yates J. R., <u>Journal of the Chemical Society</u>, <u>Chemical Communications</u> **1987**: 548-550.

26. 193-nm photodissociation of singly and multiply charged peptide anions for acidic proteome characterization, Madsen J. A., Kaoud T. S., Dalby K. N. & Brodbelt J. S., <u>Proteomics</u> **2011** (11): 1329-1334.

27. Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation, Shaw J. B., Li W., Holden D. D., Zhang Y., Griep-Raming J., Fellers R. T., Early B. P., Thomas P. M., Kelleher N. L. & Brodbelt J. S., <u>J Am Chem Soc</u> **2013** (135): 12646-12651.

28. Shabanowitz, J., Compton, P.D., Early, L., Hunt, D.F., Hunt, D.F.&Mullen, C., Method and Apparatus for Generation of Reagent Ions in a Mass Spectrometer, US20090294649 A1, Dec 3, 2009.

29. Front-end electron transfer dissociation: A new ionization source, Earley L., Anderson L. C., Bai D. L., Mullen C., Syka J. E., English A. M., Dunyach J. J., Stafford G. C., Jr, Shabanowitz J., Hunt D. F. & Compton P. D., <u>Anal Chem</u> **2013** (85): 8385-8390.

30. Electron transfer versus proton transfer in gas-phase ion/ion reactions of polyprotonated peptides, Gunawardena H. P., He M., Chrisman P. A., Pitteri S. J., Hogan J. M., Hodges B. D. & McLuckey S. A., <u>J Am Chem Soc</u> **2005** (127): 12627-12639.

31. Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions, Coon J. J., Syka J. E. P., Schwartz J. C., Shabanowitz J. & Hunt D. F., International Journal of Mass Spectrometry **2004** (236): 33-42.

32. Hermanson G.T. (2008) in *Bioconjugate Techniques*, Academic Press: 160-161.

33. Phosphoproteome analysis by mass spectrometry and its application to saccharomyces cerevisiae, Ficarro S. B., McCleland M. L., Stukenberg P. T., Burke D. J., Ross M. M., Shabanowitz J., Hunt D. F. & White F. M., <u>Nat Biotechnol</u> **2002** (20): 301-305.

34. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry, Martin S. E., Shabanowitz J., Hunt D. F. & Marto J. A., <u>Anal Chem</u> **2000** (72): 4266-4274.

35. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry, Udeshi N. D., Compton P. D., Shabanowitz J., Hunt D. F. & Rose K. L., <u>Nat Protoc</u> **2008** (3): 1709-1717.

36. RefSeq: An update on mammalian reference sequences, Pruitt K. D., Brown G. R., Hiatt S. M., Thibaud-Nissen F., Astashyn A., Ermolaeva O., Farrell C. M., Hart J., Landrum M. J., McGarvey K. M., Murphy M. R., O'Leary N. A., Pujar S., Rajput B., Rangwala S. H., Riddick L. D., Shkeda A., Sun H., Tamez P., Tully R. E., Wallin C., Webb D., Weber J., Wu W., DiCuccio M., Kitts P., Maglott D. R., Murphy T. D. & Ostell J. M., <u>Nucleic Acids Res</u> **2014** (42): D756-63.

37. ProSight PTM: An integrated environment for protein identification and characterization by top-down mass spectrometry, LeDuc R. D., Taylor G. K., Kim Y. B., Januszyk T. E., Bynum L. H., Sola J. V., Garavelli J. S. & Kelleher N. L., <u>Nucleic Acids</u> <u>Res</u> **2004** (32): W340-5.

38. ProSight PTM 2.0: Improved protein identification and characterization for top down mass spectrometry, Zamdborg L., LeDuc R. D., Glowacz K. J., Kim Y. B., Viswanathan V., Spaulding I. T., Early B. P., Bluhm E. J., Babai S. & Kelleher N. L., <u>Nucleic Acids Res</u> **2007** (35): W701-6.

3 Analysis of Intact Histone Proteins Using New Mass Spectrometry Techniques

This work is a collaborative effort between members of the Hunt Laboratory and members of Dr. Benjamin Garcia's Lab at the University of Pennsylvania Perelman School of Medicine, Department of Biochemistry and Biophysics.

3.1 Abstract

In the previous chapter, we described new methods in mass spectrometry for the acquisition of high quality MS/MS spectra of whole proteins. Here we report a novel extension of these techniques. Sequential ion/ion reactions carried out in a modified Orbitrap Velos Pro[™] capable of multiple fragment ion fills of the C-trap, in combination with data dependent and targeted HPLC-MS experiments, are used to obtain high resolution MS/MS spectra of histones from butyrate treated HeLa cells. These spectra were used to identify over 35 unique intact histone isoforms with up to 80% sequence coverage. Several of these proteins contained multiple post-translational modifications (PTMs), the vast majority of which were successfully site mapped. Additionally, we've characterized over 100 forms of clipped H2A and H2B found within the histone fractions. These clipped histones are intriguing in that they appear to be cleaved at higher abundance at specific residues, suggesting this is a controlled histone proteolysis event. The discovery of such a vast array of N- and C-terminally truncated histones by mass spectrometry is the first of its kind and demonstrates the power of our top down methodologies as the chances of discovering clipped histones is greatly dimished when samples are digested for mass spectrometric analyses. We also demonstrate that traditional data-dependent acquisition methods, which were designed for bottom-up proteomics, are not easily translated to the analysis of intact proteins and suggest useful changes that could be made to the instrument control software to enable the acquisition of more meaningful MS/MS data of intact proteins.

3.2 Introduction

3.2.1 Histone Proteins, Chromatin and Epigenetics

Chromatin is the structural framework that packages DNA into chromosomes within the nucleus of a cell. It is composed of DNA, RNA and proteins. Histones comprise the primary protein component of chromatin. Approximately 147 base pairs of DNA are wound around an octamer of histone proteins; two each of H2A, H2B, H3 and H4 (1, 2). This bundle of DNA and protein is referred to as a nucleosome core particle (**Figure 3.1**) and repeats roughly every 200 base pairs creating a "beads on a string" structural motif (3). Linker histone, H1, is involved in the compaction of these beads into chromatin fibers (3, 4) which are further condensed into chromosomes by scaffolding proteins (5). Nucleosome formation and DNA compaction is inherently repressive to the action of DNA-associated proteins. The structure of the chromatin regulates the accessibility of DNA to transcription factors and gives rise to epigenetic phenomena.

Epigenetics is the study of how heritable changes to chromatin affect which genes are expressed without changing the DNA sequence (6). These alterations result in changes in phenotype and this is how, in humans, over 200 different cell types exist while all cells contain identical genetic information. Compact areas of chromatin, or heterochromatin, are inaccessible and are thus transcriptionally silenced (7). More accessible areas of chromatin are referred to as euchromatin and can undergo transcription (8). Differences in which genes are expressed arise from the structure of chromatin and result in dramatically different phenotypes among cells (6). Discussed below are three ways in which the chromatin template is modified.



Figure 3.1 X-ray crystal structures of the nucleosome (PDB ID: 1EQZ) (9). DNA is shown in red. Histones H2A, H2B, H3 and H4 (2 of each) are shown in green, yellow, blue and magenta respectively. Histone N-terminal tales can be seen extending beyond the core nucleosomal structure. Images were rendered using the UCSF Chimera package (10). Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

3.2.2 Histone Post-Translational Modifications

The structure of chromatin can be altered in several ways. First, the structure of the nucleosome can be changed via covalent modifications to histones. Nucleosomal histone proteins contain two distinct domains. The globular domain is highly ordered and is a component of the nucleosome core structure. The N-terminal domain, which contains roughly the first fifty amino acid residues, is unstructured, flexible and extends outside of the nucleosome (11, 12). These N-terminal tails are the targets of numerous post-translational modifications (PTMs) (**Figure 3.2**). These modifications are known to function in two ways: 1) by directly changing the structure of chromatin, thereby making surrounding regions of DNA more or less accessible for transcription (*cis*-effects) (13-15) and/or 2) by recruiting affector proteins that activate or inhibit gene transcription (*trans*-effects) (16-18) or affect other cellular outcomes such as DNA damage repair (19).

These PTMs can vary temporally depending on the stage of the cell cycle or upon stem cell differentiation and thus are often reversible depending on the type of modification (20, 21).



Figure 3.2 Some known N-terminal PTMs of nucleosomal histones. The first 39 residues of histones H2A, H2B, H3 and H4 are shown with some known PTMs including phosphorylations, acetylations and methylations. Adapted from (6).

An example of a *cis*-affecting modification, histone acetylation is found to correlate strongly with gene activation. Since acetylation of the lysine ε -amino group (NH₃⁺) eliminates positive charge (**Figure 3.3 A**), it is hypothesized that it serves to disrupt histone-DNA electrostatic interactions, thereby allowing DNA decompaction and increasing the accessibility of the DNA to RNA polymerase and other proteins involved in transcription (13, 22). Acetylation of histones is reversible; acetylation is catalyzed by histone acetyltransferases (HATs) and removed by histone deacetylases (HDACs) (23, 24). Acetylation is also known to bind several "reader" domains (25-27) and functions as a signaling (*trans*-affecting) marker, triggering several cellular processes (28).

On the other hand, methylation of histone lysine and arginine residues creates more varied affects, observed as both "activating" (29-31) and "repressive" (18, 32, 33) epigenetic marks. Methylation of lysine and arginine residues alters the character of the positive charge they carry, and, in the case of trimethylation of lysine residues, creates a permanent positive charge (Figure 3.3 B and C). The direct or *cis*-effects this has on chromatin structure have not been thoroughly studied. However, in RCC1, a DNA binding protein, trimethylation of the N-terminal α -amino group has been shown to facilitate DNA binding (34). One could hypothesize that this strengthens the interaction between the trimethylated amine and the π -electron clouds of the DNA's aromatic rings. Trimethylation in histones might act similarly in order to bind DNA more strongly. In fact, recently, a histone H3 variant, CENP-A, was found to be trimethylated on its N-terminus (35), but the effects have not yet been investigated. In deciphering trans-effects, inconsistencies among the functional outcomes of histone methylation, even in the same methylation sites, have lead to the idea that the function of methylation not set in stone, but rather is determined by the proteins "reading" these epigenetic marks (36, 37). Several readers recognize these methylation sites and all affect completely different outcomes. Like acetylation, methylation is also a reversible PTM. Methylation of lysines is catalyzed by lysine methyltransferases (KMTs) and is removed by lysine Arginines are methylated by protein arginine demethylases (KDMs) (38, 39). methyltransferases (PRMTs), which can subsequently dimethylate arginine residues in a symmetric (PRMT type I) or asymmetric (PRMT type II) fashion (Figure 3.3 C) (40). The existence of arginine demethylases is still yet to be determined (41, 42), but it is known that monomethylated arginines can be converted to citrullines by protein arginine deiminase type 4 (43).



Figure 3.3 Post translational modifications change the chemical properties of amino acid residues. (A) Acetylation of lysine abolishes the positive charge it usually carries. (B) Trimethylation of lysine residues creates a fixed positive charge. (C) Arginines can be dimethylated symmetrically or asymmetrically. Evidence supporting the existence of arginine demethylases exists, but further studies are needed to confirm the identity of the enzyme responsible. (D) Phosphorylation of serine (depicted), threonine and tyrosine residues converts them to negatively charged chemical moieties.

Histone phosphorylation, like methylation, does not strictly correlate with a given cellular outcome. The mechanism of the *cis*-acting effects of phosphorylation are not fully understood, but in *Tetrahymena*, phosphorylation of several clustered residues on H1 has been shown to disrupt its binding to DNA, making it more accessible to transcription machinery (44). This is due to the disruption of the electrostatic interactions between the numerous basic residues of H1 and the negatively charged DNA. The duality of the effects of H1 phosphorylation is perplexing as H1 phosphorylation has also been found to correlate with chromatin condensation (45, 46). Similarly, phosphorylation of H3 at serine 10 has been shown to trigger chromatin condensation and transcriptional activation in a *trans*-acting manner (47, 48), lending credence to the theory of affector protein context-dependent outcomes for many histone PTMs (42). Phosphorylation of histones is performed by kinases and is reversible by the action phosphatases.

Histones can be post-translationally modified in several other ways including (but not limited to) ubiquitylation, sumoylation, biotinylation, ADP-ribosylation, and O-GlcNAcylation (49-53). In addition to the sheer number of possible ways histones can be modified, another layer of complexity is added because many of these PTMs coexist on the same proteoform. The presence of more than one epigenetic mark on a given histone often has significant biological implications. For example, phosphorylation of serine 10 on histone H3 increases the binding affinity of HAT, Gcn5, by an order of magnitude, suggesting that phosphorylation is necessary for subsequent acetylations to occur (54). There is also evidence cross-talk exists among histone PTMs. As a classic example, H3 serine 10 phosphorylation inhibits subsequent methylation by the KMT, Su(var)3-9, and dimethylation of lysine 9 reduces the activity of Aurora B kinase, which is responsible for S10 phosphorylation (55). A later study revealed that trimethylation of lysine 9 is important for the association of heterchromatin protein 1 (HP1) with chromatin and that during mitosis, serine 10 is phosphorylated, resulting in the removal of HP1

(56). Following mitosis, removal of the phosphorylation allows HP1 to reassociate with the still trimethylated lysine. This example of PTM cross-talk is an important mechanism of transcriptional regulation and is called a *"methyl/phos switch"*. Methylation promotes the formation of heterochromatin as evidenced by the association of HP1. Phosphorylation prevents this binding, inhibiting the formation of heterochromatin. Additional examples of this type of regulation have led to the emergence of the *Histone Code Hypothesis* which postulates the linking of specific modifications or combinations of modifications to individual cellular events (57).

3.2.3 Histone Variants

Although histone PTMs have received the most attention, an alternative means of epigenetically marking chromatin is through the incorporation of variant histones. All histones, with the exception of histone H4, have variants (58). Each histone variant acts as a substitute for the core or canonical histone it replaces. Taken together, histone variants represent a minor proportion of the total histone content and as a result they are difficult to study. They can differ dramatically from their canonical counterparts by the addition of large domains or by just a few amino acid residues. By way of example, **Figure 3.4** highlights some of the differences observed among known isoforms of histone H2A. In all histones, these differences can have significant impacts in their delivery during chromatin assembly or in their function following nucleosomal assembly (59). Replacement of H3.1 with H3.3 marks regions of active chromatin (60, 61). H3.3 differs from H3.1 by only four amino acid residues, but this is enough to determine the pathway by which each proteoform is deposited (62) and therefore, ultimately, the region of chromatin in which it is placed. The opposite extreme, another H3 variant, CENP-A, shares the least sequence identity with canonical H3. CENP-A comprises the H3

component of discrete patches of centromeric nucleosomes (63, 64). Loss of CENP-A leads to defects in chromosome segregation and is embryonic lethal in mice (65, 66).

1	MSGRGKQGGKARAKAK <mark>T</mark> RSSRAGLQFPVGRVHRLLRKG-NYAERVGAGAPVY <mark>L</mark> AAVLE	57	H2A.1_Human
1	MSGRGKQGGKARAKAK <mark>S</mark> RSSRAGLQFPVGRVHRLLRKG-NYAERVGAGAPVY <mark>M</mark> AAVLE	57	H2A.2-A_Human
1	MSGRGKQGGKARAKAK <mark>S</mark> RSSRAGLQFPVGRVHRLLRKG-NYAERVGAGAPVY <mark>M</mark> AAVLE	57	H2A.2-C_Human
1	M <mark>SGRGK</mark> T <mark>GGKARAKAKSRSSRAGLQFPVGRVHRLLRKG</mark> -H <mark>YAERVGAGAPVYLAAVLE</mark>	57	H2A.X_Human
1	MAGGKA <mark>GK</mark> DS <mark>GKA</mark> KT <mark>KAVSRSQRAGLQFPVGRIHR</mark> HLKSRTTSHG <mark>RVGA</mark> TAAVYSAAILE	60	H2A.Z_Human
1	MSSR <mark>GGK</mark> KKSTKT <mark>SRS</mark> AKA <mark>G</mark> VI <mark>FPVGR</mark> ML <mark>R</mark> YIK <mark>KG</mark> -HPKY <mark>RIG</mark> V <mark>GAPVYMAAVLE</mark>	54	macroH2A.1_Human
58	YLTAEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLL	117	H2A.1_Human
58	YLTAEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLL	117	H2A.2-A_Human
58	YLTAEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLL	117	H2A.2-C_Human
58	YLTAEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLG <mark>G</mark> VTIAQGGVLPNIQAVLL	117	H2A.X_Human
61	<mark>YLTAEVLELAGNA</mark> SK <mark>DLK</mark> VK <mark>RITPRHLQLAIRG</mark> DEELDSLIK-ATIAGGGVIPHIHKSLI	119	H2A.Z_Human
55	<mark>YLTAEILELAGNAARDNKK</mark> G <mark>R</mark> VT <mark>PRH</mark> IL <mark>LA</mark> VA <mark>NDEELN</mark> Q <mark>LL</mark> KG <mark>VTIA</mark> S <mark>GGVLPNI</mark> HPE <mark>LL</mark>	114	macroH2A.1_Human
110	<u> מערדד פט<mark>טעאע</mark> מער מער מער מער מער מער מער מער מער מער</u>	130	<u>407 1 Шитар</u>
110		120	H2A.1_Hullall
110		120	H2A 2-C Human
110		142	HZA.2-C_Hullian
100	PKKISAIVGP <mark>K</mark> APSGGKKAIQASQEI	143	HZA.A_Human
T20	G <mark>KK</mark> GQQKTV	T78	HZA.Z_Human
115	AKKRG <mark>S</mark> KGKLEAIITPPPAKKAKSPSOKKPVSKKAGGKKGARKSKKKOGEVSKAA//DAN	372	macroH2A.1 Human

Figure 3.4 Aligned sequences of some of the known isoforms of histone H2A. Histones of type H2A.1 and H2A.2 are considered canonical histones (67). Residues of the histone variants, H2A.X, H2A.Z and macroH2A.1, which match any of the three shown canonical H2A isoforms are highlighted in yellow. Dissimilarities among the included canonical histones are indicated in cyan. Sequences were obtained and aligned using Uniprot (68).

Differences in the primary structure of histone variants sometimes lead to unique PTMs because their canonical counterparts do not contain those sites. For example, histone variant H2A.X possesses a C-terminal "S-Q-E motif" that is not present in canonical H2A (**Figure 3.4**). The serine is phosphorylated upon DNA double-strand breaks (69). Similarly, CENP-A contains serine residues at positions 16 and 18 that are not present in other H3 isoforms. These serines are phosphorylated and loss of this phosphorylation results in mitotic errors (70). On the other hand, there is evidence that some histone variants possess their own pattern of PTMs among conserved sites (71). H3.1 and H3.3 variants that are not yet incorporated into chromatin differ in the PTMs that occur on lysine 9 (monomethylation and acetylation, respectively) (72, 73). These PTMs, which occur prior to nucleosomal assembly, have influence in where these histones are incorporated in chromatin (73). It is important to note that some histone variants are incorporated into chromatin independently of DNA replication (61, 74).

Thus, variant histones can be exchanged for canonical histones at any point in the cell cycle and serve as additional scaffold for epigenetic regulation.

3.2.4 Controlled Histone Proteolysis

Considering the extensive remodeling the chromatin template undergoes through mitosis or upon stem cell differentiation (21), one might ask through what mechanisms such global "overhaul" is achieved? We have already discussed some of the enzymes that are responsible for adding and removing PTMs to histones as well as the removal of some canonical histones that are subsequently replaced by histone variants. A less understood mechanism involves controlled histone proteolysis. Histone N-terminal tails, as well as the C-terminus in the case of H2A, are flexible and extend away from the nucleosome core particle, making these regions accessible to proteases. Given the ease with which previously placed epigenetic markers can be "erased" by simply "clipping" them off the histone, it is postulated that the action of these proteases could be involved in histone turnover and epigenetic regulation (57). There are several reported examples of this, many of which correlate to specific cellular events (75-77). The first mammalian histone-specific protease was identified in 2008 as Cathepsin L, which cleaves H3 at its N-terminus at various sites between residues 21 and 28 during mouse embryonic stem cell differentiation (77). Intriguingly, histone variant H3.2 was preferentially cleaved over H3.1, suggesting that proteolysis might be regulated via isoform preference or variant specific PTMs.

Most of the literature regarding histone-specific proteolysis describes clipping of histone H3. However, there are a few reports of clipping in other histones. In *Tetrahymena*, H4 in the somatic macronucleus (which is transcriptionally silenced) is cleaved along the N-terminal tail domain (78). Histones purified from calf thymus have

been shown to contain a C-terminally clipped form of H2A (79). *In v*itro studies show that the loss of this C-terminal peptide results in dissociation of the histone octamer, suggesting that cleavage of H2A leads to the formation of a chromatin template that is more open to transcription (80). The action of the protease appears to be regulated through acetylation of the histone octamer (81), supporting the theory that this type of controlled proteolysis can be modulated by the presence of PTMs.

3.2.5 Why Top-Down Mass Spectrometry For Histones?

Comprehensive characterization of histone proteins by mass spectrometry has proven technically difficult for a number of reasons. Traditional bottom-up methods of sequence determination and PTM site localization are not practical. As shown in Figure 3.2, histone N-terminal regions are rich in lysine and arginine residues and thus proteolysis using trypsin generates peptides that are too small or are poorly retained on reverse-phase HPLC resins for subsequent MS detection (82). With the advent of ETD and more efficient ECD fragmentation methods, which are better suited for larger, more highly charged peptides, several studies utilizing middle-down-based approaches have emerged (82-85). While these methodologies do well to preserve the combinatorial PTM profiles of histone tails, in some cases it is still impossible to identify the protein isoforms from which these peptides originate. As depicted in Figure 3.5, depending on the sequence of the isoform, any form of proteolysis results in the loss of source information for some of the resulting peptides. This is why, in theory, top-down mass spectrometry is the best method for identifying unique histone proteoforms. Despite the advantages of top-down MS, this approach is not widely utilized because of the many challenges with which it is associated.

In the previous chapter, novel methods of improving signal-to-noise ratios (SNRs) in the analysis of intact proteins were discussed. These methods enable the acquisition of

ARTKQTARKSTGGKAPRKQLATKAARKSAPSTGGVKKPHRYRPGTVALREIRRYG TKAARKSAPSTGGVKKPHRYRPGTVALREIRRYG	KSTELLIRCLFFORLVI KSTELLIRCLFFORLVI	të LAQDEK TDLRFQSAAL GALQË të LAQDEK TDLRFQSAAL GALQË të LAQDEK TDLRFQSAAL GALQË GluC Digestion	ASEAYLVGI	.FEDINLCATHARNTIMPOLOLARRIRGERA H3.3_Huma .FEDINLCATHARNTIMPOLOLARRIRGERA ClippedH3.3	ر Human
artrotark stockaprediaticaark sapatogovecpherrpotvalre irrygeste	LLIRKLPFQRLVRE	IAQDFKTDLRFQSSAVMALQE	а <mark>с</mark> е ауг	NGLFE DINLCAIHACRUTIMPEDIQLARRIRGE RA H3.	I_Human
artrotark stograperolaticaark sapatogovicphritegtvaler i tryok ste artrotarkstograperolaticaark sapstogovicphritegtvaler i tryok ste	LLIRKLPFQRLVRE LLIRKLPFQRLVRE	IAODFKTDLRFOSSAVMALOE IAODFKTDLRFOSAAIGALOE	ASE AYL ASE AYL	VGLFE DTNLCAIHAKRVTIMPKDIQLARRIRGE RA H3. VGLFE DTNLCAIHAKRVTIMPKDIQLARRIRGE RA H3.	2_Human 3_Human
TKAARK SAPSTGGVKKPHRYRPGTVALRE IRRYQKSTE	LLIRUPFQRLVRE	IAQDFKTDLRFQSAAIGALQE	ASE AYL	VGLFE DINLCAIHACRVTIMPODIQLARRIRGE RA Clip	oedH3.3_Humar
1 2	ю	4	5	6 7 8	
Figure 3.5 Peptides resulting from proteolyt H3.3 and H3.3 following cleavage by Catheosir	tic digest of	f H3 isoforms us H3.3) are shown.	ing G I Resid	IuC. Sequences of histones H3.	1, H3.2, soforms
are indicated in red. Modified residues as report the afore mentioned criteria. Digestion using the	orted in the l	Uniproto database GluC. produces	e are ir eight p	ndicated in blue. Purple residues eptides (numbered). Peptides of	fit both tvpes 5
and 8 are very small and would be poorly retain types 2 3 6 7 and 8 are indistinguishable acr	ined on a rev	erse-phase HPL(C colun	nn and may elude detection. Per source isoforms are distinguisha	tides of

types 2, 3, 6, 7, and 8 are indistinguishable across all forms of H3. Type 1 peptide source isoforms are distinguishable only between histones H3.3 and clipped H3.3. For peptides of type 4, those derived from H3.1 and H3.2 cannot be distinguished from one another as is also the case for peptides derived from H3.3 and clipped H3.3. This highlights the wealth of potentially biologically relevant information that is lost upon digestion for MS analyses. **Figure** H3.3 a are ind and 8 the afc

high quality MS/MS spectra on a chromatographic timescale, which is essential to maximize the number of protein identifications that can be made. The use of sequential ETD and IIPT reactions yielding comprehensive coverage of apomyoglobin, a 17 kDa protein, is also described. As proof of principle that these new techniques can be used for the identification and characterization of more complex samples of intact proteins coupled with chromatography, we applied them to histones purified from butyrate treated HeLa cells. Histones represent a class of proteins ideally suited for our top-down techniques due to their size, the high ETD-compatible charge states they exhibit, and the potential to gain valuable information concerning combinatorial PTMs and variants.

3.3 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)

Anaspec, Inc. (Fremont, Ca)

Vasoactive Intestinal Peptide; human, porcine, rat; ≥95% purity

GTS-Welco (Allentown, PA)

10 ppm SF₆ in nitrogen gas

Honeywell (Morristown, NJ)

Burdick and Jackson® Acetonitrile, LC-MS grade

Molex (Lisle, IL)

Polymicro Technologies[™] polyimide coated fused silica capillary

Sizes: (360 µm o.d. x 75 µm i.d.)

Sigma Aldrich (St. Louis, MO)

Fluoranthene, >99% purity

2-propanol, LC-MS grade

Sutter Instrument Co. (Navato, CA)

P-2000 microcapillary laser puller

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

Angiotensin I human acetate hydrate, ≥ 90% purity

Formic Acid, LC-MS Grade

Orbitrap Velos Pro[™] mass spectrometer

Pierce® Water, LC-MS Grade

3.4 Methods

This work was done in collaboration with Dr. Benjamin Garcia, a former Hunt lab member and currently Presidential Associate Professor of Biochemistry and Biophysics in the epigenetics program at the University of Pennsylvania Perelman School of Medicine. The samples analyzed in this study were prepared in his laboratory and the methods are described briefly below.

3.4.1 Histone Extraction From HeLa S3 Cells

Cultured HeLa S3 cells were treated with 10 mM sodium butyrate overnight, harvested and the nuclei isolated as described in (90). The nuclei pellet was resuspended in 0.4 M NaCl containing 1 mM dithiothreitol (DTT), 0.3 mM 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), 10 mM sodium butyrate and 5 µM Halt protease inhibitor cocktail in a ratio of 10:1 (v/v), salt buffer:pellet. This was incubated with constant rotation for 2 hours at 4 °C and then spun at 3000 x g for 5 minutes. Supernatant was discarded and the pellet was resuspended in a 2.5 M NaCl

solution containing DTT, AEBSF, sodium butyrate and protease inhibitors as above in a ratio of 5:1 (v/v), salt buffer:pellet. The histones were then acid extracted and TCA precipitated as described in (90).

3.4.2 Off-line Histone Fractionation

Histones purified from HeLa S3 cells (~100 µg) were dissolved in water and chromatographically separated on a Vydac C18 column (4.6 mm i.d., 250 mm length, 5 µm particle size) at a flow rate of 0.8 mL/min using a gradient of 30-60% B in 100 minutes, 60-100% B in 20 minutes, then back down to 30% B in 10 minutes (solvent A = 5% (v/v) acetonitrile, 0.2% (v/v) trifluoroacetic acid in HPLC grade water; solvent B = 95% (v/v) acetonitrile, 0.188% (v/v) trifluoroacetic acid in HPLC grade water). Fractions were collected using an automatic sample collector in 1 minute intervals between 15 and 80 minutes from the start of the gradient. The UV detector was set to monitor absorbance at 214 nm and fractions containing histones H1, H2A, H2B and H4 were identified and appropriately combined based on their known chromatographic retention times (90). These were transferred to microcentrifuge tubes, taken to dryness using SpeedVac concentrator and stored at -35 °C.

3.4.3 On-Line RP-HPLC for Mass spectrometry

Dried samples were reconstituted in water at concentrations of 5-10 pmol/µL. Aliquots of 1-2.5 pmol of protein along with 100 fmol standards were pressure-loaded onto a 360-µm o.d. x 75-µm i.d. fused-silica micro-capillary analytical column (packed 11-cm of POROSHELL 300SB-C18 resin) equipped with a laser-pulled electrospray emitter tip (91). The column was washed with ~20 column volumes of 0.3% (v/v) formic acid in water. Samples were eluted at 60 nL/min using a gradient of 0-33% B in 10 minutes, 33-48% B in 60 minutes, 48-100%B in 10 minutes, then back down to 0% B in

10 minutes. The elution gradient utilized solvent A: 0.3% formic acid in water and solvent B: 0.3% formic acid, 72% acetonitrile, 18% isopropanol and 9.7% water (all v/v). Following reverse-phase HPLC separation, samples were directly ionized using micro electrospray ionization as described (92). For some samples, initial screens revealed that better chromatography might be achieved by enacting simple changes in the elution gradient. These changes included doubling the 0-33% B portion of the gradient to a time of 20 minutes for samples containing clipped histones, and altering the gradient used for H1 to 0-20% B in 10 minutes, 20-40% B in 60 minutes, then 40-100% B in 10 minutes. All other aspects of the RP-HPLC separation and ionization for mass spectrometry remained the same across all analyses.

3.4.4 MS Analyses

Samples were ionized by microelectrospray into the front-end of an in-house modified ETD/IIPT-enabled Orbitrap Velos Pro^{TM} (see **Chapter 2** for a description of the instrument). Initially, the instrument was operated in a data-dependent mode in which the MS1 scan was taken from m/z 300-2,000 in the Orbitrap mass analyzer (r = 60,000 at m/z 400) followed by a second MS1 scan acquired in the ion trap (IT, a QLT). A single ETD/IIPT MS/MS scan was also taken in the Orbitrap. MS/MS scan parameters were set as follows: r = 30,000 at m/z 400; 200-2000 m/z scan range; 1E5 ion counts FTMSⁿ automatic gain control (AGC) target; 3E5 ion counts reagent AGC target; 15 fragment ion fills of the C-trap; 3 microscans; 3 m/z precursor isolation window; 4 ms ETD; and 25 ms IIPT. The instrument was also operated using charge state exclusion (+1, +2 and +3) and dynamic exclusion with a repeat count of 2, repeat duration of 60 seconds, and an exclusion duration of 90 seconds. Each MS/MS scan took approximately 11.5 seconds to acquire and the full scan cycle was repeated roughly every 13 seconds.

After inspecting the data from the first LC-MS experiment, it became clear that operating the instrument in a traditional data dependent fashion would not work for the analysis of intact proteins. The reasons for this are discussed in the results section. To overcome this, we altered the instrument methods to include a narrowed IT Full MS scan window upon which data dependent selection was based. All samples were initially screened using a 600-900 m/z data dependent selection window and then subsequently reanalyzed using more optimal selection windows tuned to the analytes contained in the sample. For all analyses, we lowered the dynamic exclusion repeat count from 2 to 1 and precursor isolation windows from 3 m/z to 2 m/z. Additionally, because analytes of different sizes and charge densities require different ETD/IIPT reaction times, C-trap fill counts and resolving power; these were varied depending on the fraction being analyzed. For example, when analyzing H1, a 21.9 kDa protein, we used shorter ETD and IIPT reaction times, more multiple fills and higher resolution settings than we used for H4, an 11.4 kDa protein.

Some fractions contained larger proteins that required MS/MS scan ranges over 2000 m/z for sequence determination and PTM site mapping. However, Thermo's Xcalibur[™] Instrument Setup Software does not allow the acquisition of spectra with scan ranges exceeding 2000 m/z for ETD/IIPT MS/MS scan types in data dependent mode even though the instrument is capable of scanning over a higher range. This prompted us to perform "on-the-fly" targeted analyses based upon full MS data acquired in the initial sample screenings. We call this type of data acquisition, Active Data Dependent Targeting (ADDT). Instrument method parameters were varied at our discretion based on our evaluation of the MS/MS spectra generated as data were acquired. This enabled us to obtain higher quality spectra at scan ranges above 2000 m/z. These methods are discussed in further detail in the results section of this chapter.

High resolution full MS and MS/MS data were manually inspected using Qual Browser software (Thermo Fisher Scientific). Interpretation of all ETD/IIPT MS/MS data was performed manually on the unprocessed raw spectra. Theoretical fragment ion masses were calculated using an in-house developed fragment mass calculator. Theoretical isotopic distributions were modeled using Isotope Pattern Calculator software (Pacific Northwest National Laboratories, http://omics.pnl.gov/). Percent sequence coverage was calculated by dividing the number of observed N-C α bond cleavages by the total number of predicted N-C α bond cleavages (note that cleavage of the N-C α bond that is N-terminal to proline does not produce an observable fragment). Masses for all intact proteoforms identified agreed to within 5-ppm and were confirmed by MS/MS. Relative abundance information for each clipped histone species was determined by taking the area under the curve of the extracted ion chromatogram for the most abundant isotope for every charge state present utilizing a 10-ppm mass window (93).

3.5 Results

3.5.1 Initial Experiments and MS Methods Determination with Fraction H2A2

HeLa cells were treated with sodium butyrate prior to histone isolation and separation. Butyrate treatment of cells results in histone hyperacetylation; butanoate acts as an inhibitor of HDAC activity (94). The histones were acid extracted from the nuclei and purified by off-line reverse-phase HPLC fractionation. **Figure 3.6** shows the chromatogram from the second round of purification. The histones were fractionated and identified based on their well known chromatographic retention times (95). Note that H2A elutes in two distinct peaks, which were collected as separate fractions.



Figure 3.6 Elution profile of histones purified from butyrate treated HeLa cells. Histones are separated into their respective fractions (indicated above) using a C18, reverse-phase HPLC column. Absorbance was monitored at 214 nm.



Figure 3.7 Total ion current (TIC) and base peak chromatograms obtained by LC-MS analysis of fraction H2A2 from butyrate treated HeLa cells. The top trace (A) shows the TIC chromatogram for all high resolution, MS1 scan types (FT) and the bottom trace (B) shows the base peak chromatogram for the same scan type. Differences in relative abundance can be attributed to the degree of ion current splitting whole proteins exhibit when ionized by electrospray ionization, which is not captured by the base peak chromatogram.

Initial MS experiments were performed using the H2A2 fraction. Intact protein was directly loaded onto a reverse-phase C18 poros-shell column and gradient eluted into the mass spectrometer. Analysis was performed on the Orbitrap Velos Pro[™] modified for gas-phase ion/ion chemistry (ETD and IIPT) and multiple-fill capabilities described in the previous chapter. The instrument was operated in data dependent mode to select the most abundant ion from the previous full MS scan for MS/MS interrogation. The full scan cycle was repeated roughly every 13 seconds. For MS/MS scans, 3 transients were acquired, each using 15 multiple fills. ETD and IIPT reaction times were set for 4 ms and 25 ms respectively. Shown in **Figure 3.7** are the total ion current (TIC) and base-peak chromatograms (FT) from the LC-MS experiment. Note the differences in the

counts and relative abundance among the chromatograms. The two larger peaks eluting near the beginning of the chromatogram correspond to 100 fmol each of the peptidic standards, vaso active intestinal peptide (VIP) and angio tensin. Due to the degree of ion current splitting that is observed in mass spectrometry of intact proteins, base-peak chromatograms, which are only a trace of the most abundant ion at a given point in time, are not accurate reflections of the total amount of any given species. The TIC is a better scaled reflection of the total signal detected from any source because it takes into account all detected ionic species. Examination of the chromatogram reveals three distinct regions of intact protein elution (labeled 1, 2 and 3 in **Figure 3.7**).

Peaks 1 and 2 correspond to H4 in various modified forms containing oxidized and unoxidized methionines, respectively. As shown in Figure 3.6, the H4 and H2A2 peaks are not baseline resolved and so the presence of H4 in the H2A2 fraction is expected. Displayed in **Figure 3.8** is the averaged, high-resolution ESI spectrum recorded on peak 3 in the base peak chromatogram. Signals in the observed charge state distribution carry 8-23 positive charges. The region of this spectrum containing species carrying +13 charges (grey bracket) is enlarged in the inset. The most abundant species in this region is observed near m/z 1078, with the most abundant signal in this isotopic envelope occurring at m/z 1078.0766. The ETD/IIPT MS/MS spectrum recorded on the corresponding [M+19H]¹⁹⁺ ion (m/z 738.06) is shown in Figure 3.9 A. Residues that are identified by ions of types c and z are indicated in Figure 3.9 B and define roughly the first and last 50 residues of the histone protein H2A.1 (UniParc P0C0S8); approximately 80% sequence coverage was obtained. The isotopic distribution of this protein was modeled using Isotope Pattern Calculator software (Pacific Northwest National Laboratories, http://omics.pnl.gov/). The previously referenced signal at m/z 1078.0766 corresponds to an isoform containing eight ¹³C atoms, which is predicted to be the most


Figure 3.8 Averaged ESI full MS spectrum from "peak 3". MS1 spectrum depicting the ESI charge state distribution and basic relative abundance information for the proteins eluting in peak 3. All proteins that elute in this time possess similar charge state distributions, which range from +8 to +23. The inset shows an enlarged region of the spectrum (indicated by the grey bracket) containing species carrying 13 positive charges.

abundant isotopomer. However, the calculated and the observed m/z differ by a mass of 42 Da. All observed ions of type c are also 42 Da heavier than predicted. This, combined with accurate fragment ion masses enables us to conclude that this proteoform of H2A.1 is α -N-terminally acetylated. Once the acetylation is considered, the calculated (m/z 1078.0797) and observed masses for the most abundant isotopomer differ by only -2.8 ppms. The other signals within this isotopic envelope observe roughly equal mass accuracy and abundances that match the modeled isotopic distribution.



Figure 3.9 ETD/IIPT MS/MS spectra of intact H2A.1. (A) ETD/IIPT spectrum recorded on $[M+19H]^{19+}$ ions from H2A.1 using 4 ms ETD and 25 ms IIPT. (B) Sequence coverage of intact H2A.1 obtained from MS/MS spectrum shown in (A).

As can be seen in **Figure 3.8**, several other H2A (and H4) proteoforms were detected in the analysis of this sample. However, most of these species eluded MS/MS interrogation under the instrument method used in the initial analysis. Further examination of the data file elucidated the reasons behind the lack of more meaningful MS/MS spectra. Firstly, several of the species that were selected to undergo fragmentation possessed relatively low charge states. Shown in **Figure 3.10 A** is a full

MS spectrum showing the charge state distribution of intact H4. As H4 elutes, the +8 charge state is the most abundant charge state for all H4 proteoforms in more than half of the full MS spectra, upon which data dependent selection is based. Panel **B** displays the MS/MS spectra resulting from fragmentation of the +8 species. The spectrum is devoid of all but a few fragment ions and is dominated by charge reduced precursor, suggesting that it possesses a strong tendency towards non-dissocative electron transfer. The +15 species (panel **C**) exhibits significantly more efficient fragmentation by ETD. Thus, it is in our best interest to limit the selection window to one that possesses higher charge states, thereby promoting more extensive fragmentation and providing more meaningful spectra. Another related issue is that because the charge state distribu



Figure 3.10 ESI full MS spectra of intact H4 and ETD/IIPT MS/MS spectra of the +8 and +15 charge states. (A) Full MS spectra depicting the charge state distribution of intact H4 proteoforms. (B and C) ETD/IIPT MS/MS spectra of H4 precursor carrying 8 and 15 positive charges, respectively. Spectra were acquired under identical conditions in the same LC-MS experiment.

-tion of intact proteins is so wide and the modified forms of these proteins are present at significantly lower abundance, they are rarely selected for MS/MS because there are other charge states of the unmodified precursor species that are always present at higher levels, and are thus, always selected. In order to overcome these issues, the method was altered to include a low resolution full MS scan of a window that, at any point in the analysis, would contain a single, high (ETD-amenable) charge state of each unique protein species. The scan is taken in the ion trap (IT, a QLT) and is completed in about 100 ms, so it does not contribute much to the overall time it takes to complete a full scan cycle (FT MS1, IT MS1, MS/MS).

The sample was reanalyzed using the modified data acquisition method, which incorporated an IT full MS scan over an 850-910 m/z range. This range encompasses the +13 charge state of H4 related species and the +16 charge state of H2A related species (**Figure 3.11**). These charge states were selected with an aim at acquiring quality MS/MS spectra while also maximizing ion intensity for the selection of lower level proteoforms. This narrowed selection range window was combined with altered dynamic exclusion parameters to mimic the data dependent selection of the five most abundant ions from the IT data dependent selection scan. Under the original data acquisition method, only the two most abundant proteoforms of H2A.1 were selected for MS/MS. With the modified method, twelve masses corresponding to unique H2A proteoforms were selected.

Depicted in **Figure 3.12** is a full MS spectrum indicating signals for H2A isoforms that were identified and confirmed by ETD/IIPT MS/MS. In addition to H2A.1, also detected were the isoforms H2A.1-C and H2A.1-H (UniParc Q93077 and Q96KK5 respectively). Their sequences are shown in **Figure 3.13**. H2A.1-C differs from H2A.1 by the substitution of T16 with a serine residue, and by substitution of K99 with an arginine residue. These substitutions give rise to a delta mass of +13.9904 Da from the



Figure 3.11 Narrowed data dependent selection scan ranges capture single charge states of different protein precursors. The scan range in which data dependent selection takes place is indicated by grey brackets and is identical in both spectra. (A) Full MS spectrum depicting that all unique H4 proteoforms carrying 13 positive charges are present in the selection window. (B) Full MS spectrum depicting that all unique H2B proetoforms carrying 16 positive charges are present in the selection window.





1	MSGRGKQGGKA	RAKAK	TRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYLAAVLEYLT	60	H2A.1_Human
1	MSGRGKQGGKA	RAKAK	SRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYLAAVLEYLT	60	H2A.1-C_Human
1	MSGRGKQGGKA	RAKAK	TRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYLAAVLEYLT	60	H2A.1-H_Human
61	AEILELAGNAA	RDNKK	TRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKK	120	H2A.1_Human
61	AEILELAGNAA	RDNKK	TRIIPRHLQLAIRNDEELNKLLGRVTIAQGGVLPNIQAVLLPKK	120	H2A.1-C_Human
61	AEILELAGNAA	RDNKK	TRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKK	120	H2A.1-H_Human
121 121 121	TESHHKAKGK TESHHKAKGK TESHHKAK	130 130 130	H2A.1_Human H2A.1-C_Human H2A.1-H Human		

Figure 3.13 Sequence alignments of histone H2A isoforms identified in fraction H2A2. All isoforms were confirmed by MS/MS. Histones of type H2A.1 and H2A.2 are considered canonical histones (67). Sequence variations are highlighted in grey. Sequences were obtained and aligned using Uniprot (68).

H2A.1 isoform, which is very close to the delta mass that would be exhibited in the event of a monomethylation (+14.0157 Da). The calculated and observed mass differences for the H2A.1-C and the monomethylated form of H2A.1 are 3.5 and 5.3 ppms respectively, making H2A.1-C the better fit by mass accuracy. However, based on this information alone, neither could be conclusively excluded. The MS/MS spectra contained c-ions that exhibited a delta mass of -14 Da following T16 and z-ions that exhibited a mass shift of +28 Da following K99. The +28 Da mass shift is also indicative of a dimethylated lysine, but this possibility was ruled out based on the mass accuracy of the fragment ions observed. Thus, we were able to conclude that this H2A isoform is H2A.1-C. This demonstrates the importance of obtaining high quality MS/MS spectra for the confirmation of proteoform identity. H2A.1-H differs from H2A.1 by the deletion of the last two C-terminal residues, emphasizing the importance of analyzing the intact forms of these proteins as digestions would produce identical peptides with the exception of those containing the C-termini. This difference was easily discerned upon examination of the z fragment ion series in the MS/MS spectra. H2A.1-C and H2A.1-H are present at levels roughly 20-25% and 15-20% respectively relative to H2A.1. Eluting later in the gradient and also confirmed by MS/MS are the respective mono-K-acetylated forms of these proteins. The acetylation is localized to K5 in all H2A types.

3.5.2 Highlights from Other Histone Fractions

For the remaining histone fractions, we began by screening each fraction so that an optimal data dependent selection window could be determined. In order to improve the spectra garnered from the initial screen, the data dependent selection window was narrowed to a range of 600-900 m/z. Based on this data, an optimized data dependent selection window was determined and the sample was then reanalyzed using the modified instrument method. MS/MS scan ranges for analyses of this type extended to 2000 m/z, which is high enough for most proteins to be identified confidently. Higher scan ranges require more acquisition time, and thus, reduce the total number of MS/MS scans that can be acquired, which is detrimental to on-line analysis. All results are summarized in **Table 3.1** (which can be found near the end of this section). Only those protoforms that could be identified confidently by MS/MS are included. Columns 1 and 2 list each protein's common name and accession number. Column 3 lists each proteoform that was observed. The fraction in which each proteofrom was detected is indicated in column 4. Note that the initiating methionine was removed from all proteoforms observed in each fraction. Cleavage of the initiating methionine is usually observed for proteins that contain G, A, S, P, V, T, or C in position 2 and is catalyzed by methionine amino peptidases (96). Consistent with conventional histone nomenclature, when referring to positions of amino acid residues, position 1 is defined as the first residue following the initiating methionine.

Fraction H1— Compared to other histones, H1 presents a more difficult situation for on-line LC-MS analysis of the intact form. The most common human isoforms of H1 range from 194-226 amino acids in length and, due to their larger size, require higher resolution, shorter gas-phase ion/ion reaction times, more multiple fills of the C-trap, and an extended mass range up to at least 4000 m/z. However, Thermo's Xcalibur[™]

Instrument Setup Software does not allow the acquisition of spectra with scan ranges exceeding 2000 m/z for ETD/IIPT MS/MS scan types in data dependent mode even though the instrument is capable of scanning over a higher range. Initial sample runs revealed the presence of at least two isoforms of H1. However, confident identification by MS/MS was difficult to obtain due to the limited scan range. This led us to perform a rather unconventional style of data acquisition we've termed, Active Data Dependent Targeting (ADDT). In ADDT, the sample is first screened using high resolution (r = 100,000 at 400 m/z) to determine the m/z of unique proteoforms that we wish to interrogate by MS/MS as well as the times at which they elute. In a second run, the sample is loaded onto the same HPLC column as used in the previous analysis, eluted using the same gradient, and the user actively observes the proteins eluting from the column and determines the optimal instrument parameters for targeted for MS/MS. This allows the user more flexibility to acquire MS/MS spectra that are optimized for proteins of different types. Since a method is not being used, the number of C-trap fills is not fixed as it is in a data dependent run, and so the user can adjust according to the size of the protein that is being eluted from the column. We can similarly vary the ETD/IPT reaction times according to protein size. It also gives us the opportunity to evaluate the quality of the MS/MS spectra as they are acquired and make meaningful adjustments. For example, the m/z chemical interferents are impossible to predict, and so, in some data dependent analysis, these interferents are co-isolated with species of interest, producing weak MS/MS spectra for the desired species (Figure 3.14). Since we are able to quickly assess the quality of the spectra that are acquired, we can choose to target different charge states, away from chemical interferents, that provide higher quality spectra. We can also optimize the isolation window by adjusting it to a slightly higher or slightly lower m/z so that two unique protein species are not co-isolated, which produces mixed spectra (Figure 3.15).



Figure 3.14 ADDT affords the opportunity to obtain optimized MS/MS spectra. (A) High resolution full MS spectrum showing the +23 and +24 charge states of intact H1. Note the species present at ~30% relative abundance. (B) If the +24 charge state of this species is selected, a high quality MS/MS spectrum is obtained. (C) If the +23 charge state is selected, the resulting MS/MS spectrum is of lower quality due to the presence of a chemical interferent that is not observed in the full MS spectrum.



Figure 3.15 ADDT allows for the selection of optimal isolation windows. If the instrument were run in a traditional data-dependent fashion, the isolation window indicated in red would be used and the MS/MS spectra produced would contain 2 precursor species. If the user is performing ADDT, isolation windows can be centered at m/z values that exclude one of the two species (shown in green) so that MS/MS spectra are not mixed and therefore easier to interpret.

Using ADDT, more informative MS/MS spectra were acquired and used to identify several unique proteoforms contained in the fraction (**Table 3.1**). These proteins include two known isoforms of H1, H1.2 and H1.4 (UniParc P16403 and P10412, respectively). Both are α -N-terminally acetylated. A form of H1.2 containing a previously uncharacterized single amino acid substitution, A144 \rightarrow T, was discovered along with mono- and di-phosphorylated H1.4. In the mono-phosphorylated form, the phosphorylation site occurs as a mixture between S171 and T145. In the di-phosphorylated form, the two phosphorylations are shared among three sites, T145, S171 and S186. Serine 171 is a novel phosphorylation site. Intriguingly, also identified

in this fraction were two non-histone proteins, 60S ribosomal proteins L27A and L22 (UniParc P46776 and P35268, respectively).

Fraction H2A1—Inspection of full MS spectra acquired in the initial screen of fraction H2A1 revealed the presence of several peptides that elute prior to the intact histones contained in the sample (**Figure 3.16**). These could not be identified because the MS/MS parameters were not appropriate for the characterization of species this size. ETD reaction times were too short for their efficient fragmentation and the number of C-trap fragment ion fills (n=15) was unnecessarily high and created a scan cycle that was too long to acquire MS/MS spectra for the majority of the peptides. The sample was reanalyzed using parameters that, while inappropriate for intact protein characterization, were more optimal for large peptides. MS/MS spectra revealed that these species possessed c- or z- fragment ion series, but not both, that were identical to those observed for the intact H2A isoforms found in the sample. This combined with accurate mass information led us to conclude that the peptides we were observing were both N-and C-terminal truncated, or "clipped", versions of these histones.



Figure 3.16 Fraction H2A1 contains several peptidic species. Inspection of the TIC chromatogram and full MS spectra revealed the presence of several peptidic species in a wide range of sizes. (A) TIC chromatogram from an LC-MS analysis of H2A1. (B) Averaged full MS spectrum from spectra acquired over minutes 12.5-14.5 of the analysis showing several highly charged peptides. (C) Averaged full MS spectrum from spectra acquired over minutes 16-20 of the analysis showing several, larger peptides.

To define in greater detail the levels at which these truncated histone peptides are observed, we performed semi-quantitative analysis of the data. In the analysis, we were met with some rather significant challenges/limitations and thus, there are a few caveats that bear mentioning. Firstly, the intact forms of H2A identified in the fraction were H2A.2-A and H2A.2-C, which differ only by their last 5-6 C-terminal residues (**Figure 3.17**). Because of this, there is no way to determine if the N-terminal peptides we observed are derived from H2A.2-A or H2A.2-C (UniParc Q6FI13 and Q16777, respectively). Similarly, for C-terminal peptides, it is impossible to determine whether these were derived from the mono- (α -N-acetyl) or di- (α -N-acetyl & K5-acetyl) acetylated form of the intact protein. Another challenge involved accounting for the



Figure 3.17 Averaged ESI full MS spectrum depicting intact H2A species and their respective amino acid sequences. (A) MS1 spectrum depicting the +18 charge state of the intact proteins found in the H2A1 fraction. (B) Aligned amino acid sequences of the two intact H2A isoforms identified in the fraction. Sequence dissimilarities are highlighted in grey. Sequences were obtained and aligned using Uniprot (68).

amount of variation in the degree of ion current splitting that each individual peptide exhibits. The peptides we identified range in size from 9 to 85 amino acids in length. As a result, the number of charge states that these species exhibit and the width of their isotopic distributions differ significantly.

Sequences of some of the observed peptides are listed in Table 3.2 (which can be

found at the end of this section). The C-terminal peptides included in the table are

derived only from H2A.2-A. Also listed in the table are the calculated relative abundances of each peptide expressed as "mass area" (MA). Mass areas were calculated with great care to account for variation in charge state distributions by including all charge states each peptide exhibited. Note that the mass areas listed for N-terminal peptides are given as the sum of the mono- and di-acetylated forms and might also contain area derived from H2A.2-C peptides. To estimate the abundance of the clipped histones relative to the intact protein, we compared mass areas of the C-terminal peptides with the areas derived from the mono- and di-acetylated intact forms of H2A.2-A. Based on these calculations, we estimate that the abundance of the clipped histones is roughly 0.5-1% relative to the intact protein.

Table 3.2 is not a comprehensive list of all the peptides observed in fraction H2A1. Based on observed retention times, some of these peptides may have eluded detection because they were not retained on the HPLC column and eluted during column washing. For example, the 1-9 N-terminal peptide can be observed in the first full MS scan of the analysis, indicating that it had begun eluting prior to the start of data acquisition. It is fair to assume that other peptides near this size had similar chromatographic difficulties. The retention times of the largest peptides we identified suggests that several of the larger clipped species may have eluted with the intact histones and were not detected. These peptides are present at levels less than 1%, and so their signals might be buried under those of the intact protein. Several very large clipped peptides elute after the intact protein, suggesting that several must have co-eluted with the intact proteins.

Peptides in **Table 3.2** shown in red are those that exhibit at least a five-fold increase in abundance compared to the N-terminal peptide that precedes it. When we view the data this way, patterns emerge. The data suggests that the preferred cleavage sites are C-terminal to some glycine residues and to aspartic and glutamic acids. Peptides that are observed at abundances of 1E7 or higher are highlighted in yellow, indicating regions of increased clipping. Based on the N-terminal peptides we observed, there does not appear to be any bias towards clipping of the di-acetylated over the mono-acetylated protein. Both forms are observed in abundances that mirror their intact sources. Similarly, C-terminal peptides derived from H2A.2-C, when observed, were found at the levels we'd expect, considering the lower abundance of this intact isoform. We believe that some H2A.2-C peptides were not observed simply because they were present below our limit of detection.

Fraction H2B—Intact H2B proved the most difficult to characterize by mass spectrometry. This was due to the amount of heterogeneity within the sample. There are over a dozen known isoforms of H2B. Many of these differ by only one amino acid residue. Such small differences are not significant enough to cause dramatic shifts in chromatographic retention times. If the multiple PTMs that have been found to occur on H2B are considered, the sample becomes even more complex. Shown in **Figure 3.18** is the averaged full MS spectra acquired over the peak in the chromatogram containing intact forms of H2B. For charge states that give adequate MS/MS spectra, the degree of separation between unique proteoforms is less than 1 m/z. Since we use 2 m/z isolation windows in our analysis so that precursor ions are trapped efficiently without running the risk of losing labile PTMs, any MS/MS spectra we acquire may contain up to three unique protein species. However, the quality of our spectra was high enough to identify several unique H2B proteoforms. The first three species in the cluster are unmodified forms of H2B.1-K, H2B.1-C/E/F/G/I (referred to here as H2B.1-C) and H2B.2-E (UniParc O60814, P62807 and Q16778, respectively). The remaining species are most likely various acetylated forms of these three isoforms. Mono-, di- and triacetylated H2B.1-C were confirmed and site mapped by MS/MS.



Figure 3.18 Averaged ESI full MS spectrum depicting intact H2B species and their respective amino acid sequences. (A) MS1 spectrum depicting the +18 charge state of the intact proteins identified by MS/MS in the H2AB fraction. The dashed red lines indicate the width of the isolation window centered around the m/z of the most abundant protein. Precursor ions derived from two other proteins are also contained in this window. (B) Aligned amino acid sequences of the three intact H2B isoforms identified in the fraction. Sequence dissimilarities are highlighted in grey. Sequences were obtained and aligned using Uniprot (68).

Eluting after the intact H2B proteoforms were several fairly abundant forms of clipped

H2B. The base peak chromatogram and the sequence of the most abundant isoform of

intact H2B are shown in Figure 3.19. Recall that the base peak chromatogram is not an

accurate reflection of relative abundance; we chose to use it because it better defines



Figure 3.19 Fraction H2B contains highly abundant truncated histones. Inspection of the base peak chromatogram and full MS spectra revealed the presence of several, highly abundant truncated histone species cleaved at residues 25, 51 and 50. (**A**) Base peak chromatogram from an LC-MS analysis of H2B. (**B**) The sequence of the most abundant H2B isoform contained in the fraction. The bold solid line indicates the "primary" cleavage sites mapped by MS. The grey lines indicate additional, less abundant cleavage sites. Lysines to which acetylations have been site mapped are marked with blue hexagons.

where each species elutes. The solid black lines indicate "primary" cleavage sites mapped by MS. Less abundant cleavage sites are marked with grey lines. This is not a comprehensive account of all the cleavages that are observed, but rather only those that are most abundant. Similar to the clipped H2A, we observe several other cleavage sites at lower levels. However, the cleavages seem far more specific and more closely clustered C-terminal to aspartic acid residues. Complementary N-terminal peptides, 125 and 1-51 are present and elute near the front of the gradient. They are less abundant, and are more difficult to characterize because their signals are split among multiple acetylated forms (+1-3 K-acetyl confirmed by MS/MS). The acetylations are shared among multiple sites including K11, K12, K15, K16, and K20, which are all sites that we identified in the intact form and that have previously been identified by MS in an earlier study (97). Truncations of the other H2B isoforms (H2B.2-E and H2B.1-K) are also present, though their analysis meets challenges similar to those experienced with the clipped forms of H2A. Specifically, the sequences of the intact forms found in the sample differ by only 1 residue (V39 \rightarrow I for H2B.2-E or S124 \rightarrow A for H2B.1-K) and so distinction between truncated peptides can only be made if they contain both residues. If only one of these residues is present, only one of the two additional isoforms can be excluded. There is also no way to distinguish the modified (acetylated) form of the intact protein from which the C-terminal peptides are derived. Taken together, in contrast with the truncated H2A, we estimate the clipped H2B is present at levels approaching 5% relative to the intact histones.

Table 3.1 Unique proteoforms observed in histone fractions purified from butyrate treatedHeLa cells and confirmed by MS/MS.*Indicates PTM that was not conclusively site mapped.

Protein	Accession Numbers	Unique Proteoforms	Histone Fractions
Н1 2	P16403	α-N-Acetyl	
111.2	110403	α-N-Acetyl + A144-→T	_
		a-N-Acetyl	_
	_	α-N-Acetyl + 1Phos	
H1.4	P10412	(mixture between S171 and T145)	H1
		α-N-Acetyl + 2Phos	
	D (0770	(shared among 1145, S171 and S186)	
60S RP L27A	P46776	NA	
60S RP L22	P35268	NA TRN Asstal	110.4.0
			H2A2
	Deces	C-IN-ACETYI + K5-ACETYI	H4
HZA.1	P0C058	d-IN-Acetyl + Ubiqutin (C-terminal tall)"	
		a-IN-Acetyl + K5-Acetyl	H4
		+ Obiqutin (C-terminal tail)	
H2A.1-C	Q93077		-
		a N Apotul	H2A2
H2A.1-H	Q96KK5	u-N-Acelyi	-
H2A.1-B/E	P04908		-
		a N Apotul	H4
H2A.1-D	P20671	u-N-Acelyi	-
		a N Apotul	
		u-N-Acelyi	-
H2A.2-A	Q6FI13		-
		$\frac{1}{27}$	
		a N Apotul	
	016777	a N Apotul - KE Apotul	-
пzА.2-С	QIOTT		
		(shared among K12, K15 and K16)	
H2B.1-	P62807	+ 2 K-Acetyl	-
C/E/F/G/I	1 02007	(shared among K11 K12 K15 and K16)	H2B
		+ 3 K-Acetyl	
		(shared among K11, K12, K15, K16 and K20)	
H2B.2-E	Q16778	NA	-
H2B.1-K	O60814	NA	-
		α-N-Acetvl + K20-diMethvl	
		a-N-Acetyl + K20-diMethyl	-
		+ K16-Acetyl	
		α-N-Acetyl + K20-diMethyl	
		+ K16-Acetyl	H2A1 H2A2
	P62805	+ K12-Acetyl	
		α-N-Acetyl + K20-diMethyl	
H4		+ K16-Acetyl	
	1 02000	+ K12-Acetyl	H2B
		+ K-Ac	H4
		(shared between K8 and K5)	4
		α-N-Acetyl + K20-diMethyl	
		+ K16-Acetyl	
		+ K12-Acetyl	
		+ NO-ACEIVI	
1	1		

Truncated Sequence	Residues	MA
SGRGKQGGKA	1-10	4.612E+05
SGRGKQGGKAR	1-11	2.687E+06
SGRGKQGGKARA	1-12	1.654E+06
SGRGKQGGKARAK	1-13	4.871E+05
SGRGKQGGKARAKA	1-14	6.503E+05
SGRGKQGGKARAKAK	1-15	1.548E+06
SCRCAQGGAARARAAS	1-16	5.164E+05
SCHORVOCHARANANSK	1 1 9	3.010E+00
SCREWCICKARAKARAR	1-10	2 347E+06
SGREKOGGKARAKAKSRSSR	1-20	7 186E+06
SGREKOGGKARAKAKSRSSRA	1-21	1727E+07
SGRGKOGGKARAKAKSRSSRAG	1-22	2.666E+07
SGRGKQGGKARAKAKSRSSRAGL	<mark>1-23</mark>	1.189E+07
SGRGKQGGKARAKAKSRSSRAGLQ	1-24	5.689E+06
SGRGKQGGKARAKAKSRSSRAGLQF	<mark>1-25</mark>	1.255E+07
SGRGKQGGKARAKAKSRSSRAGLQFP	1-26	5.117E+05
SGRGKQGGKARAKAKSRSSRAGLQFPV	1-27	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVG	<mark>1-28</mark>	3.040E+07
SGRGKQGGKARAKAKSRSSRAGLQFPVGR	1-29	4.898E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRV	1-30	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVH	1-31	4.647E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHR	1-32	4.078E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRL	1-33	2.758E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLL	1-34	5.283E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLR	1-35	5.355E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRK	1-36	1.812E+07
SGRGKQGGKARAKAKSKSSKAGLQFPVGRVHRLLIRKG	1-37 1.20	6.719E+07
SGRGKQGGKARAKAKSKSSKAGLQF PVGKVHKLLKKGN	1-30	0.900E+00
	1-39	9.333E+00
SCRCKQCCRX RAXARSKSSKASLQF FYORVIRKLIKRCNY A	1-40 1_/1	8 502E±06
SCRCK/CCCKARAKAKS/CS/CARDOLD/FP/CR/HRLLRKCN/AFP	1-41	8.173E+06
SGREKOGGKARAKAKSRSSRAGIOFPUGRVHRLIJRKGNYAERV	1-42	4 048E+06
SGRGKOGGKARAKAKSRSSRAGLOFPVGRVHRLLRKGNYAERVG	1-44	1.122E+08
SGREKOGGKARAKAKSRSSRAGLOFPVGRVHRLLRKGNYAERVGA	1-45	5.379E+07
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAG	1-46	3.876E+07
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGA	<mark>1-47</mark>	2.000E+07
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAP	1-48	8.763E+05
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPV	1-49	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVY	1-50	2.073E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYM	1-51	1.075E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMA	1-52	1.108E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAA	1-53	1.158E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAV	1-54	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVL	1-55	8.453E+05
SGRGKQGGKARAKAKSRSSRAGLOFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLE	1-56	1.345E+07
SGRGKQGGKARAKAKSRSSRAGLOFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEY	1-57	1.454E+06
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYL	1-58	9.524E+06
SGRGKQGGKARAKAKSKSSKAGLQFPVGRVHRLLRKGNYAEVGAGAPVYMAAVLEYLT	1-59	
	1-60	0.292E3
	1-62	4.049L0
SGREKOGGKARAKAKSRSSRAGIOFPUGRVHRLI.RKGNYAERVGAGAPVYMAVUETITAETI	1-63	ND
SGREKOGEKARAKAKESSSRAGLOFPUGRUHRLIRKGNYAERUGAGAPUYMAAULEYLTAEILE	1-64	ND
SGRGKOGGKARAKAKSRSSRAGLOFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILEL	1-65	ND
SGRGKOGGKARAKAKSRSSRAGLOFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELA	1-66	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELAG	1-67	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELAGN	1-68	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELAGNA	1-69	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELAGNAA	1-70	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELAGNAAR	1-71	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELAGNAARD	1-72	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELAGNAARDN	1-73	ND
${\tt SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELAGNAARDNK}$	1-74	ND
SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYMAAVLEYLTAEILELAGNAARDNKK	1-75	ND

Table 3.2 N-terminal truncated histones observed in H2A1 fraction from butyrate treated HeLa cells. Sequences shown in red exhibit at least a 5-fold increase compared to the N-terminal peptide listed before it. Highlighted peptides possess mass areas of 1E7 or higher.

Table 3.2 (continued) N-terminal truncated histones observed in H2A1 fraction from butyrate treated HeLa cells. Sequences shown in red exhibit at least a 5-fold increase compared to the N- terminal peptide listed before it. Highlighted peptides possess mass areas of 1E7 or higher.

Truncated Sequence	Residues	MA
VLEYLTAEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	54-129	ND
$\label{eq:legender} LEYLTAEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK$	55-129	ND
EYLTAEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	56-129	8.152E+05
YLTAEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	57-129 58 120	1.305E+07
TAETLELAGNAARDNKKTRITPRHLQLAIRNDEELNKLLGKVTTAOGGVLDNTQAVLLDKKTESHHKAKGK	59-129	8.828E+06
AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	60-129	3.675E+05
EILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	61-129	9.815E+05
${\tt ILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK$	62-129	7.764E+06
LELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	63-129	ND
ELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	64-129	2.203E+06
AGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVIIAOGGVLPNIQAVLLPKKTESHHKAKGK	66-129	9.383E+06
GNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	67-129	1.874E+07
NAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	68-129	6.762E+07
AARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	<mark>69-129</mark>	1.024E+07
ARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	70-129	8.689E+06
	71-129	9.611E+06
NKKTRIFRHIGHATKUBELINKLIGKVITAGGGVLPNIGAVLLPKKTESHHKAKGK	73-129	1.201E+07
KKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	74-129	7.947E+06
KTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	<mark>75-129</mark>	1.218E+07
TRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	<mark>76-129</mark>	1.689E+07
RIIPRHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	77-129	4.346E+06
I I PRHLQLA I RNDEELNKLLGK VT I AQGGVLPN I QAVLLPKK TESHHKAKGK	78-129	4.805E+06
PRHLOLATRNDEELNKLLGKVTTAOGGVLPNTOAVLLPKKTESHHKAKGK	80-129	3.001L+03
RHLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	81-129	5.944E+05
HLQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	82-129	2.300E+06
LQLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	83-129	2.131E+06
QLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	84-129	1.343E+06
LAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	85-129	1.405E+06
TRNDEELNKLLGKVTTAOGGVLENTQAVLLEKKTESHHKAKGK	87-129	4.034E+06
RNDEELNKLLGKVTIAOGGVLPNIOAVLLPKKTESHHKAKGK	88-129	ND
NDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	89-129	3.574E+06
DEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	90-129	6.639E+06
EELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	91-129	9.759E+06
ELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	92-129	1.495E+06
NKLLGKVTI AOGGVLPNI QAVLLPKKTESHHKAKGK	93-129	1.201E+00
KLLGKVTIAOGGVLPNIOAVLLPKKTESHHKAKGK	95-129	2.519E+06
LLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	96-129	1.500E+06
LGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	97-129	2.254E+06
GKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	98-129	2.310E+07
KVTIAQGGVLPNIQAVLLPKKTESHHKAKGK	99-129	3.303E+07
VIIAQGGVLENIQAVLLEKKTESHHKAKGK	100-129	6.904E+06
IAQGGUPNIQAVLEEKKTESHIKAKGK	102-129	2.448E+06
AQGGVLPNIQAVLLPKKTESHHKAKGK	103-129	7.020E+05
QGGVLPNIQAVLLPKKTESHHKAKGK	104-129	9.318E+06
GGVLPNIQAVLLPKKTESHHKAKGK	<mark>105-129</mark>	2.558E+07
GVLPNIQAVLLPKKTESHHKAKGK	106-129	6.375E+07
V DPNIQAV DDPKK I ESHIKAKGK	107-129	2.406E+07
PNIOAVLLPKKTESHHKAKGK	109-129	6.237E+07
NIQAVLLPKKTESHHKAKGK	110-129	4.050E+06
IQAVLLPKKTESHHKAKGK	111-129	7.644E+06
QAVLLPKKTESHHKAKGK	<mark>112-129</mark>	1.283E+07
AVLLPKKTESHHKAKGK	113-129	2.142E+07
	114-129 115-120	1.395E+07
I DRKABCHHRYRCK TTLVKI FOHHRYRCK	116-129	5 184E+06
PKKTESHHKAKKK	117-129	1.076E+07
KKTESHHKAKGK	118-129	ND
KTESHHKAKGK	119-129	ND
TESHHKAKGK	120-129	ND
ESHHKAKGK	121-129	ND

3.6 Conclusions and Discussion

The power of top town mass spectrometry lies in its ability to provide an unambiguous view of protein expression dynamics by distinguishing closely related isoforms from one another while also observing any combinations of PTMs. Bottom up MS strategies are not capable of this, which is why there has been a significant push in the field of proteomics to develop instruments and methodologies that enable highthroughput, top down mass spectrometry of intact proteins. Here, we have demonstrated the power of some of the novel top down MS technologies discussed in **Chapter 2** through their coupling with on-line HPLC chromatography and application to the analysis of intact histones. Sequential ion/ion reactions carried out in a modified Orbitrap Velos Pro[™] capable of multiple fragment ion fills of the C-trap, in combination with data dependent and targeted LC-MS experiments, have been employed to record high resolution MS/MS spectra of histones from butyrate treated HeLa cells. Using these spectra, we unambiguously identified over 35 unique forms of intact proteins. Several of these contain multiple modifications including single amino acid substitutions, methylations, acetylations, phosphorylations and ubiquitination. All modifications, with the exception of the ubiquitination of H2A.1, were site localized. Additionally, well over 100 forms of clipped H2A and H2B have been characterized.

We also demonstrate that traditional data-dependent acquisition methods, which were designed for bottom-up proteomics, are not easily translated to the analysis of intact proteins. The wide range of charge states which whole proteins possess renders data dependent selection, as it is most commonly employed, practically useless. The limitations we encountered necessitated the development of more practical data acquisition strategies. By first screening the sample and determining an optimal datadependent selection window, we were able greatly narrow the pool of selectable ions to those that only possessed a single ETD-compatible charge state, thereby enabling the generation of MS/MS spectra of several lower-level species that previously eluded tandem MS interrogation. Additionally, using ADDT we were able to tailor ion/ion reaction parameters and optimize isolation windows to each unique protein so that higher quality MS/MS spectra were obtained. This has taught us that several of the data acquisition tools we have in place could benefit from redevelopment if they are to be used in the analysis of intact proteins. New developments should include charge state distribution recognition software to prevent same-species reselection and charge state dependent ETD and IIPT reaction time determination.

The discovery of such a vast array of N- and C-terminally truncated histones by topdown mass spectrometry is, to our knowledge, the first of its kind. Typically, biologists use immunoblotting to probe cell extracts with histone antibodies and notice faster migrating bands in their samples. These bands are then excised and characterized often by Edman degradation or by bottom up mass spectrometry. That we are able to observe these cleavage events in simple top-down LC-MS experiments further demonstrates the utility of our methodologies as it is very likely that bottom up or middle down based proteomics approaches would have missed cleavage events like those we describe in our analyses. This could happen for a number of reasons. First, given the low levels at which these truncations were observed (which, for any single truncated form, is less than 0.1% relative to the intact histone), along with the added complexity that is generated when samples are digested for MS proteomic studies, peptides of this type may have not been detected or may have simply gone unnoticed by investigators. If the samples had been digested, several of the generated peptides derived from originally truncated species might have been too small to be retained by the HPLC column. If sample digestion was employed and peptides terminating in unexpected amino acid residues were found, there would be no way to uncouple the odd proteolytic

activity from the endoprotease used for digestion. The question of whether the odd cleavage sites were the result of the protease used for sample preparation as opposed to endogenous protelytic activity could not be answered. Quantitative analysis of the relative abundance of unique clipped forms, like that given here, would be far more difficult to perform as evidence for some unique truncated species could be lost, or could be below the limit of detection. Also, the ability to correlate the endogenous histone proteolysis with any specific histone variant or PTM state would not be possible unless those PTMs or sequence variations and the cleavage sites were contained within the same peptide produced in the sample preparative digestion. Only by determining the primary structures of these species, as they existed in the cell, can one draw these potentially important biological parallels.

Discerning whether any histone cleaving event is the result of continuous histone turnover and degradation versus an alternative biological process with epigenetic implications is very difficult. At this time, we do not know how these clipped histones were created; whether there are present *in vivo* or generated during sample preparation cannot be determined with the data we have. When relative abundances of the clipped histones in this study are compared, patterns emerge indicating sites of primary cleavage as well as several secondary sites. This is more apparent in the H2B clipping patterns, where cleavage directly C-terminal to aspartic acid residues is the clearly the primary site, and cleavage at a handful of residues C-terminal to this site are also observed at lower levels. In H2A, we observe increased cleavage C-terminal to some glycine and acidic residues. Such patterns are reminiscent of what was observed by Duncan et. al in 2008 in H3 from mouse embryonic stem cells. CathepsinL was found to be the enzyme responsible for clipping of the H3 primarily at A21, but also at 5 nearby C-terminal residues. In a much earlier study, several Cathepsins were found to hydrolyze histones (98). Perhaps the cleavages we are observing are the result of

similar proteases. Also in the CathepsinL study, as we saw, several complementary cleavage products were contained in the samples analyzed by mass spectrometry. This proves that both the N- and C-terminal cleavage products, which are conclusively made *in vivo*, can be retained through HPLC fractionation.

Recall that the cells from which our samples were derived were treated with sodium butyrate, which inhibits the activity of histone deacetylases, resulting in hyperacetylated histone states. Our analysis revealed that the intact H2B possessed up to 4 acetyllysines, which we mapped by MS/MS to mixtures of K11, K12, K15, K16 and K20. The complementary N-terminal cleavage products that we identified also possessed several N-terminal acetyl-lysines. The primary cleavage site was mapped to D25, and the Cterminal cleavage product (H2B 26-125) was found at surprisingly high levels relative to intact H2B. While we cannot conclude that there is a relationship between histone acetylation and controlled proteolysis, we speculate that, if these cleavages are catalyzed by proteases, there may be a link between the two. Cleavage of the Nterminal tail at D25 presents a "convenient" way to eliminate these acetylation marks. In fact, histone acetylation has been linked to proteosomal degradation of core histones during DNA repair and spermatogenesis (99). To explore this possible link, our collaborators are in the process of preparing histories purified from HeLa cells without butyrate treatment. We will perform identical LC-MS analyses and compare abundances of clipped histones, which may reveal a correlation. Our collaborators have also provided us with histone fractions from several other cell lines which have not undergone butyrate treatment. While the H2B fractions have not yet been analyzed, the H2A fractions exhibit significantly lower levels of truncated H2A.

Histone proteolysis has been implicated in several important regulatory processes, yet it remains poorly understood. Only 8 histone-specific proteases are known and their biological significance has not been extensively defined (100). This area of study is

clearly poised for significant expansion, as evidenced by the identification of two H2A specific proteases and several novel histone clipping events in just the past few years (100-103). We hope that the tools we developed and describe here will aid in deciphering this multi-faceted, biologically regulated phenomenon.

3.7 References

1. Chromatin structure: A repeating unit of histones and DNA, Kornberg R. D., <u>Science</u> **1974** (184): 868-871.

2. Chromatin structure; oligomers of the histones, Kornberg R. D. & Thomas J. O., <u>Science</u> **1974** (184): 865-868.

3. Involvement of histone H1 in the organization of the nucleosome and of the saltdependent superstructures of chromatin, Thoma F., Koller T. & Klug A., <u>J Cell Biol</u> **1979** (83): 403-427.

4. Solenoidal model for superstructure in chromatin, Finch J. T. & Klug A., Proc Natl Acad Sci U S A **1976** (73): 1897-1901.

5. Controlling the double helix, Felsenfeld G. & Groudine M., <u>Nature</u> 2003 (421): 448-453.

6. Epigenetics, Allis CD, Jenuwein T & Reinberg D, 2007: 28.

7. Heterochromatin and gene regulation in drosophila, Elgin S. C., <u>Curr Opin Genet Dev</u> **1996** (6): 193-202.

8. Heterochromatin and epigenetic control of gene expression, Grewal S. I. & Moazed D., <u>Science</u> **2003** (301): 798-802.

9. Asymmetries in the nucleosome core particle at 2.5 A resolution, Harp J. M., Hanson B. L., Timm D. E. & Bunick G. J., <u>Acta Crystallogr D Biol Crystallogr</u> **2000** (56): 1513-1534.

10. UCSF chimera--a visualization system for exploratory research and analysis, Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C. & Ferrin T. E., <u>J Comput Chem</u> **2004** (25): 1605-1612.

11. Crystal structure of the nucleosome core particle at 2.8 A resolution, Luger K., Mader A. W., Richmond R. K., Sargent D. F. & Richmond T. J., <u>Nature</u> **1997** (389): 251-260.

12. The histone tails of the nucleosome, Luger K. & Richmond T. J., <u>Curr Opin Genet</u> <u>Dev</u> **1998** (8): 140-146.

13. Acetylation and methylation of histones and their possible role in the regulation of rna synthesis, ALLFREY V. G., FAULKNER R. & MIRSKY A. E., <u>Proc Natl Acad Sci U S</u> <u>A</u> **1964** (51): 786-794.

14. On the biological role of histone acetylation, Csordas A. , <u>Biochem J</u> **1990** (265): 23-38.

15. Histone H4-K16 acetylation controls chromatin structure and protein interactions, Shogren-Knaak M., Ishii H., Sun J. M., Pazin M. J., Davie J. R. & Peterson C. L., <u>Science</u> **2006** (311): 844-847.

16. Signaling to chromatin through histone modifications, Cheung P., Allis C. D. & Sassone-Corsi P., <u>Cell</u> **2000** (103): 263-271.

17. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation, Cheung P., Tanner K. G., Cheung W. L., Sassone-Corsi P., Denu J. M. & Allis C. D., <u>Mol Cell</u> **2000** (5): 905-915.

18. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain, Bannister A. J., Zegerman P., Partridge J. F., Miska E. A., Thomas J. O., Allshire R. C. & Kouzarides T., <u>Nature</u> **2001** (410): 120-124.

19. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139, Rogakou E. P., Pilch D. R., Orr A. H., Ivanova V. S. & Bonner W. M., <u>J Biol Chem</u> **1998** (273): 5858-5868.

20. Origins and formation of histone methylation across the human cell cycle, Zee B. M., Britton L. M., Wolle D., Haberman D. M. & Garcia B. A., <u>Mol Cell Biol</u> **2012** (32): 2503-2514.

21. Chromatin organization and differentiation in embryonic stem cell models, Giadrossi S., Dvorkina M. & Fisher A. G., <u>Curr Opin Genet Dev</u> **2007** (17): 132-138.

22. Chromatin higher-order structures and gene regulation, Li G. & Reinberg D., <u>Curr</u> <u>Opin Genet Dev</u> **2011** (21): 175-186.

23. Histone acetyltransferases, Roth S. Y., Denu J. M. & Allis C. D., <u>Annu Rev Biochem</u> **2001** (70): 81-120.

24. Deacetylase enzymes: Biological functions and the use of small-molecule inhibitors, Grozinger C. M. & Schreiber S. L., <u>Chem Biol</u> **2002** (9): 3-16.

25. Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14, Kasten M., Szerlong H., Erdjument-Bromage H., Tempst P., Werner M. & Cairns B. R., <u>EMBO J</u> **2004** (23): 1348-1359.

26. Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex, Lange M., Kaynak B., Forster U. B., Tonjes M., Fischer J. J., Grimm C., Schlesinger J., Just S., Dunkel I., Krueger T., Mebus S., Lehrach H., Lurz R., Gobom J., Rottbauer W., Abdelilah-Seyfried S. & Sperling S., <u>Genes Dev</u> **2008** (22): 2370-2384.

27. Structural basis for recognition of H3K56-acetylated histone H3-H4 by the chaperone Rtt106, Su D., Hu Q., Li Q., Thompson J. R., Cui G., Fazly A., Davies B. A., Botuyan M. V., Zhang Z. & Mer G., <u>Nature</u> **2012** (483): 104-107.

28. Histone acetyltransferases, Roth S. Y., Denu J. M. & Allis C. D., <u>Annu Rev Biochem</u> **2001** (70): 81-120.

29. Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor, Wang H., Huang Z. Q., Xia L., Feng Q., Erdjument-Bromage H., Strahl B. D., Briggs S. D., Allis C. D., Wong J., Tempst P. & Zhang Y., <u>Science</u> **2001** (293): 853-857.

30. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling, Wysocka J., Swigut T., Xiao H., Milne T. A., Kwon S. Y., Landry J., Kauer M., Tackett A. J., Chait B. T., Badenhorst P., Wu C. & Allis C. D., <u>Nature</u> **2006** (442): 86-90.

31. High-resolution profiling of histone methylations in the human genome, Barski A., Cuddapah S., Cui K., Roh T. Y., Schones D. E., Wang Z., Wei G., Chepelev I. & Zhao K., <u>Cell</u> **2007** (129): 823-837.

32. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins, Lachner M., O'Carroll D., Rea S., Mechtler K. & Jenuwein T., <u>Nature</u> **2001** (410): 116-120.

33. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin, Schotta G., Lachner M., Sarma K., Ebert A., Sengupta R., Reuter G., Reinberg D. & Jenuwein T., <u>Genes Dev</u> **2004** (18): 1251-1262.

34. N-terminal alpha-methylation of RCC1 is necessary for stable chromatin association and normal mitosis, Chen T., Muratore T. L., Schaner-Tooley C. E., Shabanowitz J., Hunt D. F. & Macara I. G., <u>Nat Cell Biol</u> **2007** (9): 596-603.

35. Posttranslational modification of CENP-A influences the conformation of centromeric chromatin, Bailey A. O., Panchenko T., Sathyan K. M., Petkowski J. J., Pai P. J., Bai D. L., Russell D. H., Macara I. G., Shabanowitz J., Hunt D. F., Black B. E. & Foltz D. R., <u>Proc Natl Acad Sci U S A</u> **2013** (110): 11827-11832.

36. How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers, Taverna S. D., Li H., Ruthenburg A. J., Allis C. D. & Patel D. J., <u>Nat Struct Mol Biol</u> **2007** (14): 1025-1040.

37. Chromatin modifications and their function, Kouzarides T., Cell 2007 (128): 693-705.

38. The SET-domain protein superfamily: Protein lysine methyltransferases, Dillon S. C., Zhang X., Trievel R. C. & Cheng X., <u>Genome Biol</u> **2005** (6): 227.

39. Molecular mechanisms and potential functions of histone demethylases, Kooistra S. M. & Helin K., <u>Nat Rev Mol Cell Biol</u> **2012** (13): 297-311.

40. Histone arginine methylation, Di Lorenzo A. & Bedford M. T., <u>FEBS Lett</u> **2011** (585): 2024-2031.

41. Control of seed germination by light-induced histone arginine demethylation activity, Cho J. N., Ryu J. Y., Jeong Y. M., Park J., Song J. J., Amasino R. M., Noh B. & Noh Y. S., <u>Dev Cell</u> **2012** (22): 736-748.

42. Fundamentals of chromatin, Workman J. L. & Abmayr S. M.,

43. Histone deimination antagonizes arginine methylation, Cuthbert G. L., Daujat S., Snowden A. W., Erdjument-Bromage H., Hagiwara T., Yamada M., Schneider R., Gregory P. D., Tempst P., Bannister A. J. & Kouzarides T., <u>Cell</u> **2004** (118): 545-553.

44. Regulation of transcription by H1 phosphorylation in tetrahymena is position independent and requires clustered sites, Dou Y. & Gorovsky M. A., <u>Proc Natl Acad Sci</u> <u>U S A</u> **2002** (99): 6142-6146.

45. Inhibition of histone phosphorylation by staurosporine leads to chromosome decondensation, Th'ng J. P., Guo X. W., Swank R. A., Crissman H. A. & Bradbury E. M., <u>J Biol Chem</u> **1994** (269): 9568-9573.

46. Characterization of the mitotic specific phosphorylation site of histone H1. absence of a consensus sequence for the p34cdc2/cyclin B kinase, Gurley L. R., Valdez J. G. & Buchanan J. S., <u>J Biol Chem</u> **1995** (270): 27653-27660.

47. Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation, Clayton A. L., Rose S., Barratt M. J. & Mahadevan L. C., <u>EMBO J</u> **2000** (19): 3714-3726.

48. Making the auroras glow: Regulation of aurora A and B kinase function by interacting proteins, Carmena M., Ruchaud S. & Earnshaw W. C., <u>Curr Opin Cell Biol</u> **2009** (21): 796-805.

49. Maintenance of low histone ubiquitylation by Ubp10 correlates with telomereproximal Sir2 association and gene silencing, Emre N. C., Ingvarsdottir K., Wyce A., Wood A., Krogan N. J., Henry K. W., Li K., Marmorstein R., Greenblatt J. F., Shilatifard A. & Berger S. L., <u>Mol Cell</u> **2005** (17): 585-594.

50. Histone sumoylation is associated with transcriptional repression, Shiio Y. & Eisenman R. N., <u>Proc Natl Acad Sci U S A</u> **2003** (100): 13225-13230.

51. Biological functions of biotinylated histones, Kothapalli N., Camporeale G., Kueh A., Chew Y. C., Oommen A. M., Griffin J. B. & Zempleni J., <u>J Nutr Biochem</u> **2005** (16): 446-448.

52. Histone shuttling by poly(ADP-ribosylation), Realini C. A. & Althaus F. R., <u>J Biol</u> Chem **1992** (267): 18858-18865.

53. Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code, Sakabe K., Wang Z. & Hart G. W., <u>Proc Natl Acad Sci U S A</u> **2010** (107): 19915-19920.

54. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation, Cheung P., Tanner K. G., Cheung W. L., Sassone-Corsi P., Denu J. M. & Allis C. D., <u>Mol Cell</u> **2000** (5): 905-915.

55. Regulation of chromatin structure by site-specific histone H3 methyltransferases, Rea S., Eisenhaber F., O'Carroll D., Strahl B. D., Sun Z. W., Schmid M., Opravil S., Mechtler K., Ponting C. P., Allis C. D. & Jenuwein T., <u>Nature</u> **2000** (406): 593-599.

56. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation, Fischle W., Tseng B. S., Dormann H. L., Ueberheide B. M., Garcia B. A., Shabanowitz J., Hunt D. F., Funabiki H. & Allis C. D., <u>Nature</u> **2005** (438): 1116-1122.

57. Translating the histone code, Jenuwein T. & Allis C. D., <u>Science</u> **2001** (293): 1074-1080.

58. Non-allelic variants of histones 2a, 2b and 3 in mammals, Franklin S. G. & Zweidler A., <u>Nature</u> **1977** (266): 273-275.

59. The nucleosome: A little variation goes a long way, Bernstein E. & Hake S. B., Biochem Cell Biol **2006** (84): 505-517.

60. Nucleosomal histones of transcriptionally active/competent chromatin preferentially exchange with newly synthesized histones in quiescent chicken erythrocytes, Hendzel M. J. & Davie J. R., <u>Biochem J</u> **1990** (271): 67-73.

61. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly, Ahmad K. & Henikoff S., <u>Mol Cell</u> **2002** (9): 1191-1200.

62. Assembly of variant histones into chromatin, Henikoff S. & Ahmad K., <u>Annu Rev Cell</u> <u>Dev Biol</u> **2005** (21): 133-153.

63. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma, Earnshaw W. C. & Rothfield N., <u>Chromosoma</u> **1985** (91): 313-321.

64. Conserved organization of centromeric chromatin in flies and humans, Blower M. D., Sullivan B. A. & Karpen G. H., <u>Dev Cell</u> **2002** (2): 319-330.

65. A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis, Stoler S., Keith K. C., Curnick K. E. & Fitzgerald-Hayes M., <u>Genes Dev</u> **1995** (9): 573-586.

66. Early disruption of centromeric chromatin organization in centromere protein A (cenpa) null mice, Howman E. V., Fowler K. J., Newson A. J., Redward S., MacDonald A. C., Kalitsis P. & Choo K. H., <u>Proc Natl Acad Sci U S A</u> **2000** (97): 1148-1153.

67. Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry, Bonenfant D., Coulot M., Towbin H., Schindler P. & van Oostrum J., <u>Mol Cell Proteomics</u> **2006** (5): 541-552.

68. Activities at the universal protein resource (UniProt), UniProt Consortium, <u>Nucleic</u> <u>Acids Res</u> **2014** (42): D191-8.

69. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139, Rogakou E. P., Pilch D. R., Orr A. H., Ivanova V. S. & Bonner W. M., <u>J Biol Chem</u> **1998** (273): 5858-5868.

70. Posttranslational modification of CENP-A influences the conformation of centromeric chromatin, Bailey A. O., Panchenko T., Sathyan K. M., Petkowski J. J., Pai P. J., Bai D. L., Russell D. H., Macara I. G., Shabanowitz J., Hunt D. F., Black B. E. & Foltz D. R., <u>Proc Natl Acad Sci U S A</u> **2013** (110): 11827-11832.

71. Assembly of variant histones into chromatin, Henikoff S. & Ahmad K., <u>Annu Rev Cell</u> <u>Dev Biol</u> **2005** (21): 133-153.

72. The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39mediated K9me3 in pericentric heterochromatin, Loyola A., Tagami H., Bonaldi T., Roche D., Quivy J. P., Imhof A., Nakatani Y., Dent S. Y. & Almouzni G., <u>EMBO Rep</u> **2009** (10): 769-775.

73. PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state, Loyola A., Bonaldi T., Roche D., Imhof A. & Almouzni G., <u>Mol Cell</u> **2006** (24): 309-316.

74. Separation of basal histone synthesis from S-phase histone synthesis in dividing cells, Wu R. S. & Bonner W. M., <u>Cell</u> **1981** (27): 321-330.

75. Proteolytic processing of histone H3 in chromatin: A physiologically regulated event in tetrahymena micronuclei, Allis C. D., Bowen J. K., Abraham G. N., Glover C. V. & Gorovsky M. A., <u>Cell</u> **1980** (20): 55-64.

76. Foot-and-mouth disease virus protease 3C induces specific proteolytic cleavage of host cell histone H3, Falk M. M., Grigera P. R., Bergmann I. E., Zibert A., Multhaup G. & Beck E., <u>J Virol</u> **1990** (64): 748-756.

77. Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation, Duncan E. M., Muratore-Schroeder T. L., Cook R. G., Garcia B. A., Shabanowitz J., Hunt D. F. & Allis C. D., <u>Cell</u> **2008** (135): 284-294.

78. Proteolytic removal of core histone amino termini and dephosphorylation of histone H1 correlate with the formation of condensed chromatin and transcriptional silencing during tetrahymena macronuclear development, Lin R., Cook R. G. & Allis C. D., <u>Genes</u> <u>Dev</u> **1991** (5): 1601-1610.

79. A chromatin-bound proteolytic activity with unique specificity for histone H2A, Eickbush T. H., Watson D. K. & Moudrianakis E. N., <u>Cell</u> **1976** (9): 785-792.

80. H2a-specific proteolysis as a unique probe in the analysis of the histone octamer, Eickbush T. H., Godfrey J. E., Elia M. C. & Moudrianakis E. N., <u>J Biol Chem</u> **1988** (263): 18972-18978.

81. Regulation of H2a-specific proteolysis by the histone H3:H4 tetramer, Elia M. C. & Moudrianakis E. N., <u>J Biol Chem</u> **1988** (263): 9958-9964.

82. Chemical derivatization of histones for facilitated analysis by mass spectrometry, Garcia B. A., Mollah S., Ueberheide B. M., Busby S. A., Muratore T. L., Shabanowitz J. & Hunt D. F., <u>Nat Protoc</u> **2007** (2): 933-938.

83. Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry, Bonenfant D., Coulot M., Towbin H., Schindler P. & van Oostrum J., <u>Mol Cell Proteomics</u> **2006** (5): 541-552.

84. Analysis of histones in xenopus laevis. II. mass spectrometry reveals an index of cell type-specific modifications on H3 and H4, Nicklay J. J., Shechter D., Chitta R. K., Garcia B. A., Shabanowitz J., Allis C. D. & Hunt D. F., <u>J Biol Chem</u> **2009** (284): 1075-1085.

85. Revealing histone variant induced changes via quantitative proteomics, Arnaudo A. M., Molden R. C. & Garcia B. A., <u>Crit Rev Biochem Mol Biol</u> **2011** (46): 284-294.

86. Pervasive combinatorial modification of histone H3 in human cells, Garcia B. A., Pesavento J. J., Mizzen C. A. & Kelleher N. L., <u>Nat Methods</u> **2007** (4): 487-489.

87. High throughput characterization of combinatorial histone codes, Young N. L., DiMaggio P. A., Plazas-Mayorca M. D., Baliban R. C., Floudas C. A. & Garcia B. A., <u>Mol</u> <u>Cell Proteomics</u> **2009** (8): 2266-2284.

88. Combined bottom-up and top-down mass spectrometry analyses of the pattern of post-translational modifications of drosophila melanogaster linker histone H1, Bonet-Costa C., Vilaseca M., Diema C., Vujatovic O., Vaquero A., Omenaca N., Castejon L., Bernues J., Giralt E. & Azorin F., <u>J Proteomics</u> **2012** (75): 4124-4138.

89. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications, Moradian A., Kalli A., Sweredoski M. J. & Hess S., <u>Proteomics</u> **2014** (14): 489-497.

90. Examining histone posttranslational modification patterns by high-resolution mass spectrometry, Lin S. & Garcia B. A., <u>Methods Enzymol</u> **2012** (512): 3-28.

91. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry, Martin S. E., Shabanowitz J., Hunt D. F. & Marto J. A., <u>Anal Chem</u> **2000** (72): 4266-4274.

92. Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry, Udeshi N. D., Compton P. D., Shabanowitz J., Hunt D. F. & Rose K. L., <u>Nat Protoc</u> **2008** (3): 1709-1717.

93. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC), Ong S. E. & Mann M., <u>Nat Protoc</u> **2006** (1): 2650-2660.

94. Sodium butyrate inhibits histone deacetylation in cultured cells, Candido E. P., Reeves R. & Davie J. R., <u>Cell</u> **1978** (14): 105-113.

95. Extraction, purification and analysis of histones, Shechter D., Dormann H. L., Allis C. D. & Hake S. B., <u>Nat Protoc</u> **2007** (2): 1445-1457.

96. Methionine or not methionine at the beginning of a protein, Sherman F., Stewart J. W. & Tsunasawa S., <u>Bioessays</u> **1985** (3): 27-31.

97. Quantitative proteomic analysis of post-translational modifications of human histones, Beck H. C., Nielsen E. C., Matthiesen R., Jensen L. H., Sehested M., Finn P., Grauslund M., Hansen A. M. & Jensen O. N., <u>Mol Cell Proteomics</u> **2006** (5): 1314-1325.

98. Hydrolysis of histones by proteinases, Harvima R. J., Yabe K., Fraki J. E., Fukuyama K. & Epstein W. L., <u>Biochem J</u> **1988** (250): 859-864.

99. Acetylation-mediated proteasomal degradation of core histones during DNA repair and spermatogenesis, Qian M. X., Pang Y., Liu C. H., Haratake K., Du B. Y., Ji D. Y., Wang G. F., Zhu Q. Q., Song W., Yu Y., Zhang X. X., Huang H. T., Miao S., Chen L. B., Zhang Z. H., Liang Y. N., Liu S., Cha H., Yang D., Zhai Y., Komatsu T., Tsuruta F., Li H., Cao C., Li W., Li G. H., Cheng Y., Chiba T., Wang L., Goldberg A. L., Shen Y. & Qiu X. B., <u>Cell</u> **2013** (153): 1012-1024.

100. Proteolytic clipping of histone tails: The emerging role of histone proteases in regulation of various biological processes, Azad G. K. & Tomar R. S., <u>Mol Biol Rep</u> **2014** (41): 2717-2730.

101. Purification and characterization of a novel histone H2A specific protease (H2Asp) from chicken liver nuclear extract, Panda P., Chaturvedi M. M., Panda A. K., Suar M. & Purohit J. S., <u>Gene</u> **2013** (512): 47-54.

102. Neutrophil elastase in the capacity of the "H2A-specific protease", Dhaenens M., Glibert P., Lambrecht S., Vossaert L., Van Steendam K., Elewaut D. & Deforce D., <u>Int J Biochem Cell Biol</u> **2014** (51C): 39-44.

103. Identification of histone H3 clipping activity in human embryonic stem cells, Vossaert L., Meert P., Scheerlinck E., Glibert P., Van Roy N., Heindryckx B., De Sutter P., Dhaenens M. & Deforce D., <u>Stem Cell Research</u>