Analysis of Post-translational Modifications of Histones and Myelin Basic Protein: Crosstalk and Beyond

by

Chunchao Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Bioinformatics)
in The University of Michigan
2013

Doctoral Committee:

Professor Philip C. Andrews, Chair
Associate Professor Yali Dou
Professor Kojo Seys John Elenitoba-Johnson
Associate Professor Kristaina I. Hakansson
Associate Professor Santiago David Schnell

DEDICATION

To my parents, my wife, and my kids

ACKNOLEDGEMENTS

I would give my deepest gratitude to my advisor, Prof. Philip C. Andrews, for his excellent guidance, continuous support, and immense knowledge during my ph.d studies. His key ideas and thoughtful remarks became the initial source of my thesis research, has led me into deep ground in the proteomics field, and also changed my research philosophy for ever. I also would like to extend my sincere gratitude to our collaborators: Dr. Yifan Liu (histone project) and Prof. Robert Zand (MBP project). In the histone project, Dr. Liu worked even like a co-advisor. His invaluable inputs and discussions in epigenetics and many other fields makes our work more attractive. With his strong backup, a lot of collaborations has been done and is still going on between two labs. Prof. Zand also added many critical comments and discussions during our collaborations in the myelin basic project.

I really appreciate all the members in Andrews's lab and the most significant ones in Liu's lab where the spirits and the wisdom are actively communicated and transferred. Finally, I would like to give my special thanks to Dr. Kristaina I. Hakansson, Prof. Kojo S. Elenitoba-Johnson, Dr. Santiago Schnell, and Dr. Yali Dou for serving on my thesis committee and for their valuable advice on my research.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Ac: Acetylation

ATXR: *Arabidopsis* Trithorax Related

AUC: Area under the curve

CID: Collision induced dissociation

CV: Coefficient of variation

ESI: Electrospray ionization

EZL2: Enhancer of Zeste -like 2

FA: Factor analysis

HAT: Histone acetyltransferase

HCD: Higher-energy collisional dissociation

HDAC: Histone deacetylase

HDACi: Histone deacetylase inhibitors

HDMT: Histone demethylase

HILIC: Hydrophilic interaction chromatography

HMT: Histone methyltransferase

HPLC: High-performance liquid chromatography

ICAT: Isotope-coded affinity tag

iTRAQ: Isobaric tags for relative and absolute quantification

LC: Liquid chromatography

MALDI: Matrix-assisted laser desorption/ionization

MBP:Myelin basic protein

Me1: Monomethylation

Me2: Dimethylation

Me3: Trimethylation

MRM: Multiple reaction monitoring

MS: Mass spectrometry

MS/MS (or MS2): Tandem mass spectrometry

O.D.: Optical density

PAGE: Polyacrylamide gel electrophoresis

PAM: Partitioning Around Medoids

PcG: Polycomb group

PCM: Polydimethylcyclosiloxane

PIP: PCNA-interacting protein

Pr: Propionylation

PTM: Post-translational modification

RP: Reverse phase

RT: Retention time

SDS: Sodium dodecyl sulfate

SET: Suppressor of variegation, Enhancer of Zeste, Trithorax

SILAC: Stable isotope labeling by amino acids in cell culture

SRM: Selected reaction monitoring

TMT: Tandem mass tags

TOF: Time of flight

TOF/TOF: Tandem time of flight

TXR1: *Tetrahymena* Trithorax Related 1

WT: Wild-type

XIC: Extracted ion chromatography

CHAPTER 1

Introduction

1.1 The language of protein post-translational modifications

Post-translational modifications (PTMs) are the chemical modifications deposited by different modifying enzymes on proteins after they are translated from mRNA by the ribosome. Most proteins undergo various site-specific covalent modifications and these PTMs are found for all classes of proteins and are especially common in eukaryotic cells, [1]. It is estimated that 5% of the the human proteome comprises enzymes dedicated to various protein modifications including 500 protein kinases, 150 protein phosphatases, 500 proteases and various methyl-transferases, acetyl-transferases, glycosyl-transferases, among many others[2,3]. Among the 20 primary amino acids encoded in eukaryotic DNA, 15 have been associated with PTMs[2,3]. More than 200 forms of covalent modifications have been found and characterized[4]. The most commonly encountered PTMs are: phosphorylation, acetylation, glycosylation, methylation, amidation, ubiquitination, and hydroxylation[5](Figure 1.1) although many others play key roles in cells.

Besides mRNA splicing at the post-transcriptional level, the functional diversity of the proteome is further greatly expanded due to myriad post-translational modifications (Figure 1.2) [2,6]. Although changes on protein primary structure triggered by PTMs are relatively small, they can have significant impacts on protein functions and biological

controls in a broad spectrum. The roles of PTMs include cell signalling, protein-protein or protein-DNA interactions, gene regulation, protein degradation, and many others[1]. The modified amino acids may serve as signals for recruiting effector molecules and subsequently initiate various downstream biological events, they may affect activity directly, or they may affect localization in the cell, etc.

Some key proteins undergo very high levels of complex modifications to control their functions. Histones are the most salient of these proteins due to their key roles in living cells and the extent of their modifications. Their functional diversity, the high complexity of their PTMs and the biological significance of their covalent modifications has led to the proposal of a protein barcode hypothesis for histone proteins (see further discussions in the following section)[7,8]. It is possible that the complex PTM code embedded in histone proteins may be a universal language in the cell as these observations can also be expanded onto other non-histone proteins such as p53[9] and potentially to myelin basic protein which contains many of the same PTMs as histones[10]. Within the protein code hypothesis lies the basic idea that PTMs are written or erased by modifying enzymes (code writers or erasers) and read by effector molecules or proteins (code readers). These different types of PTMs usually do not act independently but in concert[11]. In other words, there are inter-relationships or 'crosstalk' between them where one PTM serves as a signal for the addition or removal of a second PTM as part of a control or compensatory mechanism. Examples of PTM crosstalk has been found in both histone and non-histone proteins[11–14]. A few examples are provided as follows: a close relationship between phosphorylation and O-GlcNAcylation has been well-established[15]; systematic perturbations of protein phosphorylation and acetylation confirmed that these two

different types of PTMs reciprocally affect one another in *Mycoplasma pneumoniae*[16]; In histones, phosphorylation at H3S10 promotes H4K16 acetylation and mediates transcription elongation[17]; while many other studies also suggest that lysine acetylation may have the potential to 'crosstalk' with other different types of PTMs such as methylation, sumoylation, and ubiquitination[18,19].

Given their key roles in controlling cell physiology (e.g., cell cycle[20]) and protein functions, protein PTMs, must be finely regulated to maintain homeostasis or to respond to environmental challenges. Interruption or dysregulation of PTMs will affect cell fitness and has been associated with many diseases including cancer, diabetes, heart diseases, and infectious diseases[21,22].

1.2 Histones, histone PTMs, 'histone code' hypothesis, and histone crosstalk

In eukaryotic cells, genomic DNA wraps around a protein octamer formed by two of each copy of the core histones (H1, H2, H3, H4) as a basic unit of eukaryotic chromatin called a nucleosome as a unit structure of the higher order, extremely condensed structure of the chromosome in the nucleus[23,24]. Core histones generally consist of 100-150 residues with molecular weights ranging from 11 to 15 kd. Their high content of basic residues, lysine and arginine, makes histones strongly positively charged and contributes to their tight interactions with the negatively charged DNA backbone[23]. Both linker histones (H1) and core histones (H2A, H2B, H3, and H4) are heavily modified with numerous epigenetic marks including methylation, acetylation, phosphorylation, biotinylation, citrullination, ADP-ribosylation, and ubiquitination[25]. To date, more than 60 different residues on histones have been found to be covalently modified[23]. The recent identification of several new types of histone PTMs, such as propionylation,

butyrylation, crotonylation, and succinylation, further expanded the growing pool of known PTMs in the cell[26–28]. Most histone PTMs exhibit dynamic changes in cells in response to cell cycle, differentiation, environmental challenges, etc[29,30]. Modifications may affect the highly organized chromatin structure by changing the contacts between different histones in nucleosomes, with other proteins involved in transcription, or the interaction of positively charged histones with negatively charged DNA[23].

A combinatorial permutation of these chemical modifications on histone proteins constitutes the 'histone code' hypothesis which controls eukaryotic gene expression, replication, repair, and many other DNA metabolism processes[7,8] through their effects on chromatin structure and binding of accessory proteins. In other words, multiple regulatory layers can be achieved by changing the levels, types, and positions of PTMs in histones.

The numerous chemical modifications especially on the histone N-terminal tails makes positive or negative 'crosstalk' between PTMs possible via several mechanisms[11–14,23]. Crosstalk can exist at the level of a single histone tail[31], between two different histones[17,32,33], or beyond the histones. Although our current understanding of histone crosstalk and mechanisms involved in them are becoming increasingly clear, only limited examples of crosstalk have been identified in histones. Deciphering the interactions between histone PTMs and elucidating the details of their global regulation is an area of intense ongoing research.

1.3 Mass spectrometry-based quantitative proteomics methods and their applications in histone PTMs

Rapid developments in mass spectrometry(MS)-based proteomics over the past decade has eclipsed Edman sequencing as the predominant tool for protein identification. Its ability to identify and quantify thousands of components in a single step with great sensitivity when coupled with liquid chromatography gives it a number of advantages compared with other analytical techniques[34–36]. MS-based quantification can be either relative or absolute depending on whether internal standards with known quantities are used[37–39]. Most quantification methods are based on the intensities or peak areas of peptide precursors measured at the MS1 level[36,37,39], while protocols, such as Isobaric Tags for Relative and Absolute Quantification (iTRAQ), Selected Reaction Monitoring (SRM) and Multiple Reaction Monitoring (MRM), provide direct comparisons of different samples based on the relative intensities of fragment ions at the MS2 level[37,39–42]. Direct comparison of protein abundance by the MS signal or by the number of tandem spectra without using isobaric tags in the sampe preparation is usually referred to as label-free quantification[43,44]. While stable isotopic labeling methods, either by metabolic labeling *in vivo* or by chemical reactions *in vitro*, provide more accurate measurements on the changes of peptide/protein abundance between two or more physiological states[37].

Mass spectrometric analysis of histone proteins is technically challenging due to the excess of basic residues, the large number and extent of post-translational modifications, and the relatively low-level of some of these modifications in their natural states. This is particularly true for monitoring the dynamics of histone PTMs. A part of these effects is taken advantage of in some derivatization methods (e.g., propionylation, described below) to increase the size and hydrophobicity of histone fragments. These challenges emphasize

the need for development of robust quantitative proteomic strategies for histone PTMs using mass spectrometry. Here, I briefly review label-free and isotope enrichment methods and discuss their applications in the field of epigenetics.

1.3.1 Label-free quantification

Mass spectral peak intensities and spectral counting are two major widely used strategies for label-free relative quantification of proteins in complex biological samples[43,44]. The basic concept behind this technique arises from the observation that both MS signal intensity, measured as peak area under the curve (AUC) or extracted ion chromatography (XIC), and spectral counts for a selected analyte molecule (e.g., tryptic peptide) correlate well with the corresponding protein abundances present in a sample[43–45]. The spectral counting approach quantifies proteins by comparing the number of MS/MS spectra from the same protein identified across multiple LC-MS/MS experiments[43,44]. It is worth noting that quantification by spectral counting is applied for a protein not for a peptide. In contrast to the spectral counting approach, label-free quantification by peak intensities is relatively technical complicated. Mass to charge ratios, chromatographic peak areas and retention times for each ion across all experiments need to be precisely determined and aligned. For this reason, high reproduciblity in sample preparation and LC-MS/MS runs is a key factor for successful quantification[43–45]. Good linearity (r^2>0.95) and wide linear dynamic range (>100-fold) measured by AUC or spectral counting were obtained when standard proteins were spiked into different fractions, representing different chemical backgrounds[46]. A typical coefficient of variance (CV) approximating 30% for protein ratios was reported in two studies[44,46]. Although both spectral counting and label-free quantification exhibit

good reproducibility and sensitivity for protein quantification, spectral counting is not as useful for peptide quantification, while AUC was found to be more accurate[45]. For this reason the application of spectral counting to quantification of protein PTMs has been largely limited.

Several studies have used label-free methods to quantify protein PTMs. In a study of phosphorylation stoichiometry by the label-free approach, Steen *et al.* showed that precise quantification of protein phosphorylation was possible if an appropriate normalization procedure was applied[47]. For histone studies, Beck, *et al.* successfully demonstrated an ion intensity-based LC-MS/MS label-free method for quantitative analysis of PTMs from four human core histones[48]. Unmodified tryptic peptides with small statistical variations (CV<30%) were selected for data normalization in this study, while internal standards were used to correct dataset variations between LC-MS runs. In their study of the global impact of histone deacetylase inhibitors (HDACi) on histone acetylation in H3 and H4, a relative standard deviation between 17% and 24% was observed in three replicate LC-MS/MS runs[49]. Fold-change and statistical analysis showed that the majority of peptides (~90%) were unaffected in the HDACi-treated group. A label-free approach, which combined mass profiling analysis of intact histones with tryptic peptides generated from charge isoforms, was employed to study yeast histone PTM fold-changes in wild-type and histone acetyltransferase-deficient mutants[50]. The results showed that most PTMs were not affected in knockout cells but acetylation of H3K56 was found to be significantly down-regulated as investigators anticipated from previous studies. The label-free quantification method has also been

used to determine the percentage of intact individual H4 isoforms in human embryonic stem cells based on their peak areas[51].

1.3.2 Isotope enrichment methods: ^{15}N uniform labeling vs SILAC vs chemical tags

The use of non-radioactive isotopes introduced by metabolic incorporation *in vivo* or by isobaric tags *in vitro* for quantitative proteomics is referred to as stable isotope labeling[37]. The most common selected stable isotopes are ^{2}H, ^{13}C, ^{18}O, and ^{15}N[4,24,25]. Because the stable isotope-labeled peptides are chemically identical to their unlabeled counterparts, they behave almost identically in response to chromatography (retention time, losses due to adsorption, etc.,) and mass spectrometry (ionization)[53] leading to a more accurate determination compared with label-free quantification. In some cases ^{2}H-containing compounds exhibit significant changes in chromatographic retention times which makes integration more challenging but several of the major data analysis packages for LCMS correct for these differences and most methods avoid ^{2}H. The lower variance of isotope enrichment methods makes them more sensitive to changes and they are currently more generally used for quantifying PTMs than label-free approaches. ^{15}N uniform labeling reported for proteome analysis by Oda *et al*.[54] and stable isotope labeling by amino acids in cell culture (SILAC) introduced by Ong *et al*.[55] are two of the widely used metabolic labeling approaches for quantitative proteomics. The major difference between the two approaches is that all nitrogen atoms throughout the whole proteome are uniformly replaced by heavy isotopes in ^{15}N uniform labeling while in SILAC only specific amino-acid residues (e.g Lys or Arg) are selectively labeled. The mass offsets of paired peptides using SILAC are simple and well-defined[52]. Both strategies have been realized in a variety of species[52]. In

contrast to metabolic labeling, incorporation of isotope tags *in vitro* is achieved by labeling proteins or peptides chemically or enzymetically after cell lysis or protein preparation. Chemical reagents, such as ICAT, iTRAQ, tandem mass tags (TMT), and [^{18}O] water, have been developed for quantitative proteomics[53]. Both metabolic labeling and chemical labeling strategies have been implemented in epigenetic studies and a few representative examples are described here.

A SILAC-based quantitative proteomics study conducted by Jung *et al*.[56] investigated the effects of the polycomb protein SUZ12 on H3.2 and H3.3 modifications in wild type and mouse embryonic stem cells. A significant reduction in H3K27 di/tri-methylation and an increase in H3K27 acetylation were observed in *Suz12* knockout cells. Another SILAC study performed by Cuomo *et al*. revealed cancer-specific PTM patterns for histones H3 and H4 by comparing normal breast cells with breast cancer cell lines[57]. Some of these modifications have been implicated in human cancers. ^{15}N metabolic labeling demonstrated by Oda *et al*. also proved to be promising for accurate quantification of changes in protein phosphorylation[54]. Quantifying histone PTMs using chemical tags is an alternative strategy when *in vivo* labeling is not feasible. In one study, lysyl residues and peptide N-termini were labeled by differential chemical derivatization with d0/d10-propionic anhydride to map PTM changes in histone methyltransferase knockout cells[58,59]. A chemical method using differential labeling with deuterated methyl ester (D4-methanol) has also been developed to study histone PTMs[60]. Although the isobaric tag, iTRAQ, has not been previously used to quantify histone PTMs, we recently demonstrated feasibility by labeling both intact histones and tryptic peptides (Figure 1.3). The use of iTRAQ for intact histones was quite effective in

most respects but we chose to use the uniform labeling approach to reduce long-term costs and because iTRAQ does not resolve the low retention time issues for histone peptides as well as propionylation does. The use of hydrophilic interaction chromatography (HILIC) might prove to be effective for iTRAQ-tagged histone peptides. Other examples of the application of quantitative proteomics to chromatin biology have been discussed in recent reviews[25,61,62].

1.3.3 Compare and contrast

Label-free quantification is a simple, sensitive, and cost effective way to compare up to hundreds of biological samples, but it may require a significant number of biological replicates to reduce the p values for lower level PTMs[43,44]. While the most obvious weakness of this technique is that quantification accuracy is not comparable with labeling techniques and thus may yield poor results if LC-MS/MS runs are not reproducible[45]. Isotope enrichment methods provide more accurate quantification than label-free approaches as comparison of two or more samples is made within a single LC-MS/MS run. Thus variations derived from LC and sample handling are minimized. The use of stable isotopes by metabolic labeling has some advantages over chemical labeling: 1) Metabolic labeling minimizes experimental errors by introducing heavy isotopes into the entire cellular proteome before sample is mixed which reduces variance due to sample handling[62]. 2) Variations in labeling efficiency is typically low as incorporation of isotopes is generally highly efficient[55]. Unlike label-free approaches, metabolic labeling and chemical labeling only allow a limited number of samples to be compared in a single multiplex experiment. For example, only 2-8 states can be compared by SILAC or iTRAQ and larger sample numbers require use of a common sample (e.g., a pooled

control) across all analyses. Furthermore, the reagents required for SILAC and iTRAQ add to the expense over label-free approaches. In contrast, uniform labeling using an ^{15}N-labeled sample as the internal standard is a relatively cost effective alternative and allows us to compare many biological samples.

Although multiple methodologies are available for researchers in this field, there is no universal quantitative strategy that can deal with all analytical problems due to limitations inherent to each method. Therefore, the choice of a particular method for quantitative study of biological systems is highly flexible and case-dependent.

1.4 Thesis outline

My thesis focuses primarily on the PTMs of histones. The primary goals are: 1) to study how histone PTMs are affected under different epigenetic conditions; 2) address how these apparently independent chemical marks coordinate with each other to form functionally related clusters to control biological outcomes in the cell. In chapter 2, I present a new proteomic strategy to quantitatively study histone PTMs in the model organism, *Tetrahymena thermophila*, with an ^{15}N metabolic labelling technique. By virtue of this method, I was able to confirm a novel putative HMT, Txr1p, which preferentially affects monomethylation of H3K27. I also elucidated its relationship with Ezl2p, another HMT in the same organism. In chapter 3, I expand the success of this method to other core histones and identify several important PTMs that crosstalk with methylation of H3K27. In chapter 4, I describe quantitative analysis of histone PTMs across 15 epigenetic features to reveal global crosstalk among the four major histones. Multivariate analysis of histone PTM data revealed some functionally related subgroups

or clusters of these chemical marks which greatly expanded our understanding of the 'histone code' hypothesis.

The final chapter addresses the PTMs of myelin basic protein (MBP). Similar to histone proteins, MBP is a small basic protein which is also modified by a large number of PTMs, many of which shares with histones. In chapter 5, a detailed cross-species analysis of the MBP PTMs from bovine and rattlesnake revealed the differences in MBP modification patterns between mammals and reptiles. This study extends our knowledge about relationships between MBP PTMs and demyelinating diseases.

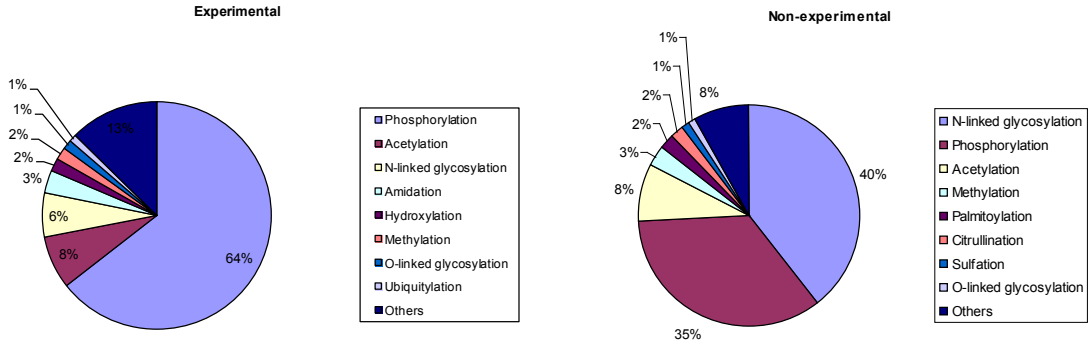


Figure 1.1 Summarized PTM statistics from experiments or computational prediction
From PTM Statistics Curator: http://selene.princeton.edu/PTMCuration/

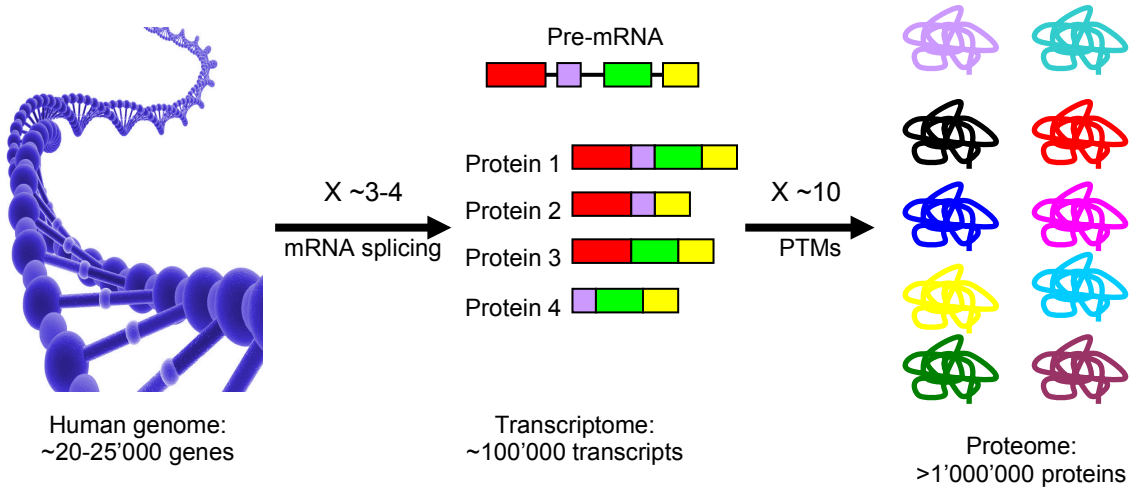


Figure 1.2 Post-translational modification is a key mechanism to increase proteomic diversity
Human genome consists of ~20-25,000 genes and its proteome comprises >1 million chemically distinct proteins. Alternative splicing and post-translational modifications significantly increase the complexity of the proteome. Statistical source: Creative Proteomics.

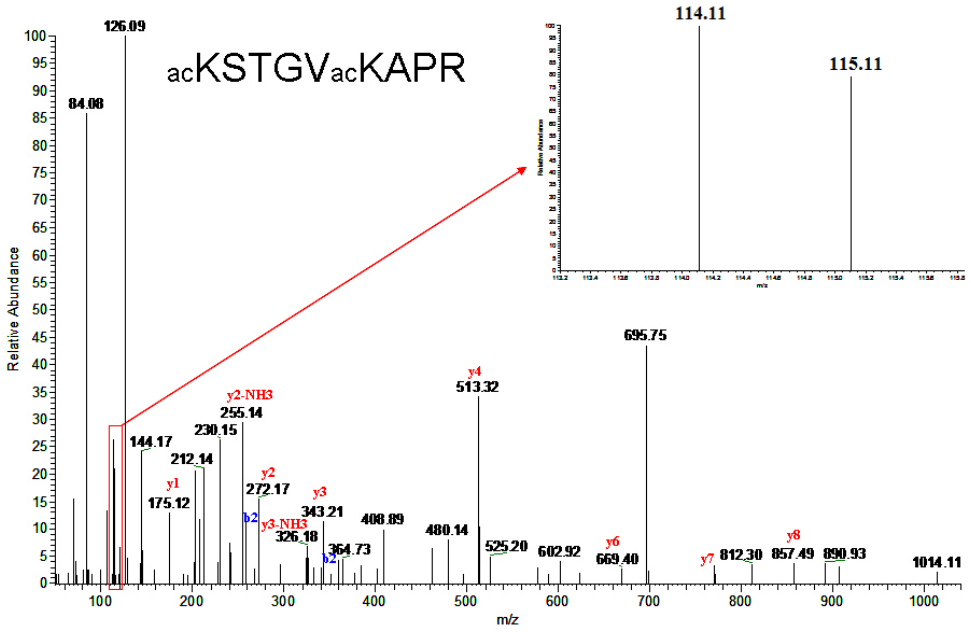


Figure 1.3 PTM quantification by iTRAQ labeling
Intact histone H3 from WT or KO cells was labeled with iTRAQ-114 or 115 and then equally mixed before trypsin digestion. The digests were analyzed with Orbitrap XL mass spectrometor in the data-dependent mode. Higher-energy C-trap Dissociation (HCD) technique was used to quantitate iTRAQ-labeled peptides to overcome the 1/3 rule in the CID. A peptide with two PTMs (acetylation) at Lys 1 and 6 was found moderately down-regulated in the HMT knocks.

1.5 References

[1] Jensen, O.N., Interpreting the protein language using proteomics. *Nature reviews. Molecular cell biology* 2006, 7, 391–403.

[2] Walsh, C.T., Garneau-Tsodikova, S., Gatto, G.J., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angewandte Chemie (International ed. in English)* 2005, 44, 7342–72.

[3] Wold, F., In vivo chemical modification of proteins (post-translational modification). *Annual review of biochemistry* 1981, 50, 783–814.

[4] Walsh, C., Postranslational Modification of Proteins: Expanding Nature's Inventory 2005, Colorado.

[5] Khoury, G.A., Baliban, R.C., Floudas, C.A., Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific reports* 2011, 1, pii:srep00090.

[6] Deribe, Y.L., Pawson, T., Dikic, I., Post-translational modifications in signal integration. *Nature structural & molecular biology* 2010, 17, 666–72.

[7] Strahl, B.D., Allis, C.D., The language of covalent histone modifications. *Nature* 2000, 403, 41–45.

[8] Jenuwein, T., Allis, C.D., Translating the Histone Code. *Science* 2001, 293, 1074–1080.

[9] Sims, R.J., Reinberg, D., Is there a code embedded in proteins that is based on post-translational modifications? *Nature Reviews Molecular Cell Biology* 2008, 9, 1–6.

[10] Zhang, C., Walker, A.K., Zand, R., Moscarello, M. a, et al., Myelin basic protein undergoes a broader range of modifications in mammals than in lower vertebrates. *Journal of proteome research* 2012, 11, 4791–802.

[11] Hunter, T., The Age of Crosstalk Phosphorylation, Ubiquitination, and Beyond. *Molecular Cell* 2007, 730–738.

[12] Latham, J. a, Dent, S.Y.R., Cross-regulation of histone modifications. *Nature structural & molecular biology* 2007, 14, 1017–24.

[13] Suganuma, T., Workman, J.L., Crosstalk among Histone Modifications. *Cell* 2008, 135, 604–7.

[14] Fischle, W., Yanming Wang, Allis, C.D., Histone and chromatin cross-talk. *Current Opinion in Cell Biology* 2003, 172–183.

[15] Hart, G.W., Slawson, C., Ramirez-Correa, G., Lagerlof, O., Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annual review of biochemistry* 2011, 80, 825–58.

[16] Van Noort, V., Seebacher, J., Bader, S., Mohammed, S., et al., Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Molecular systems biology* 2012, 8, 571.

[17] Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., et al., Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* 2009, 138, 1122–36.

[18] Lu, Z., Cheng, Z., Zhao, Y., Volchenboum, S.L., Bioinformatic Analysis and Post-Translational Modification Crosstalk Prediction of Lysine Acetylation. *PLoS ONE* 2011, 6, e28228.

[19] Yang, X., Seto, E., Lysine acetylation: codified crosstalk with other posttranslational modifications. *Oncology* 2009, 31, 449–461.

[20] Vermeulen, K., Van Bockstaele, D.R., Berneman, Z.N., The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell proliferation* 2003, 36, 131–49.

[21] Chi, P., Allis, C.D., Wang, G.G., Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nature Reviews Cancer* 2010, 10, 457–69.

[22] Karve, T.M., Cheema, A.K., Small changes huge impact: the role of protein posttranslational modifications in cellular homeostasis and disease. *Journal of amino acids* 2011, 2011, 207691.

[23] Kouzarides, T., Chromatin modifications and their function. *Cell* 2007, 128, 693–705.

[24] Campos, E.I., Reinberg, D., Histones: annotating chromatin. *Annual review of genetics* 2009, 43, 559–99.

[25] Sidoli, S., Cheng, L., Jensen, O.N., Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *Journal of proteomics* 2012, Epub ahead.

[26] Zhang, K., Chen, Y., Zhang, Z., Zhao, Y., Identification and Verification of Lysine Propionylation and Butyrylation in Yeast Core Histones Using PTMap Software research articles. *Journal of Proteome Research* 2009, 8, 900–906.

[27] Tan, M., Luo, H., Lee, S., Jin, F., et al., Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* 2011, 146, 1016–1028.

[28] Zhang, Z., Tan, M., Xie, Z., Dai, L., et al., Identification of lysine succinylation as a new post-translational modification. *Nature Chemical Biology* 2010, 7, 58–63.

[29] Krebs, J.E., Moving marks: Dynamic histone modifications in yeast. *Molecular Biosystems* 2007, 3, 8–12.

[30] Zee, B.M., Levin, R.S., Xu, B., LeRoy, G., et al., In vivo residue-specific histone methylation dynamics. *The Journal of biological chemistry* 2010, 285, 3341–50.

[31] Wang, H., Cao, R., Xia, L., Erdjument-bromage, H., et al., Purification and Functional Characterization of a Histone H3-Lysine 4-Specific Methyltransferase. *Molecular cell* 2001, 8, 1207–1217.

[32] Sun, Z.-W., Allis, C.D., Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* 2002, 418, 104–8.

[33] Lee, J., Shukla, A., Schneider, J., Swanson, S.K., et al., Histone Crosstalk between H2B Monoubiquitination and H3 Methylation Mediated by COMPASS. *Cell* 2007, 1084–1096.

[34] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, 422, 198–207.

[35] Cox, J., Mann, M., Quantitative, high-resolution proteomics for data-driven systems biology. *Annual review of biochemistry* 2011, 80, 273–99.

[36] Schulze, W.X., Usadel, B., Quantitation in mass-spectrometry-based proteomics. *Annual review of plant biology* 2010, 61, 491–516.

[37] Ong, S.-E., Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nature chemical biology* 2005, 1, 252–62.

[38] Kirkpatrick, D.S., Gerber, S.A., Gygi, S.P., The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods (San Diego, Calif.)* 2005, 35, 265–73.

[39] Xie, F., Liu, T., Qian, W.-J., Petyuk, V. a, Smith, R.D., Liquid chromatography-mass spectrometry-based quantitative proteomics. *The Journal of biological chemistry* 2011, 286, 25443–9.

[40] Anderson, L., Hunter, C.L., Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Molecular & cellular proteomics* 2006, 5, 573–88.

[41] Carr, S. a, Anderson, L., Protein quantitation through targeted mass spectrometry: the way out of biomarker purgatory? *Clinical chemistry* 2008, 54, 1749–52.

[42] Pan, S., Aebersold, R., Chen, R., Rush, J., et al., Mass Spectrometry Based Targeted Protein Quantification: Methods and Applications reviews. *Journal of Proteome Research* 2009, 787–797.

[43] Zhu, W., Smith, J.W., Huang, C.-M., Mass spectrometry-based label-free quantitative proteomics. *Journal of biomedicine & biotechnology* 2010, 2010, 840518.

[44] Wang, M., You, J., Bemis, K.G., Tegeler, T.J., Brown, D.P.G., Label-free mass spectrometry-based protein quantification technologies in proteomic analysis. *Briefings in functional genomics & proteomics* 2008, 7, 329–39.

[45] Neilson, K. a, Ali, N. a, Muralidharan, S., Mirzaei, M., et al., Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* 2011, 11, 535–53.

[46] Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., et al., Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular & cellular proteomics* 2005, 4, 1487–502.

[47] Steen, H., Jebanathirajah, J.A., Springer, M., Kirschner, M.W., Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102, 3948–53.

[48] Beck, H.C., Nielsen, E.C., Matthiesen, R., Jensen, L.H., et al., Quantitative proteomic analysis of post-translational modifications of human histones. *Molecular & cellular proteomics* 2006, 5, 1314–25.

[49] Drogaris, P., Villeneuve, V., Pomiès, C., Lee, E.-H., et al., Histone deacetylase inhibitors globally enhance h3/h4 tail acetylation without affecting h3 lysine 56 acetylation. *Scientific reports* 2012, 2, 220.

[50] Drogaris, P., Wurtele, H., Masumoto, H., Verreault, A., Thibault, P., Comprehensive profiling of histone modifications using a label-free approach and

its applications in determining structure-function relationships. *Analytical chemistry* 2008, 80, 6698–707.

[51] Phanstiel, D., Brumbaugh, J., Berggren, W.T., Conard, K., et al., Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* 2008, 105, 4093–8.

[52] Beynon, R.J., Pratt, J.M., Metabolic labeling of proteins for proteomics. *Molecular & cellular proteomics* 2005, 4, 857–72.

[53] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry* 2007, 389, 1017–31.

[54] Oda, Y., Huang, K., Cross, F.R., Cowburn, D., Chait, B.T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* 1999, 96, 6591–6.

[55] Ong, S.-E., Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics* 2002, 1, 376–386.

[56] Jung, H.R., Pasini, D., Helin, K., Jensen, O.N., Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Molecular & cellular proteomics* 2010, 9, 838–50.

[57] Cuomo, A., Moretti, S., Minucci, S., Bonaldi, T., SILAC-based proteomic analysis to dissect the “histone modification signature” of human breast cancer cells. *Amino acids* 2011, 41, 387–99.

[58] Plazas-Mayorca, M.D., Zee, B.M., Young, N.L., Fingerman, I.M., et al., One-Pot Shotgun Quantitative Mass Spectrometry Characterization of Histones. *Journal of Proteome Research* 2009, 8, 5367–5374.

[59] Plazas-Mayorca, M.D., Bloom, J.S., Zeissler, U., Leroy, G., et al., Quantitative proteomics reveals direct and indirect alterations in the histone code following methyltransferase knockdown. *Molecular bioSystems* 2010, 6, 1719–29.

[60] Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S. a, et al., Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature protocols* 2007, 2, 933–8.

[61] Arnaudo, A.M., Molden, R.C., Garcia, B. a, Revealing histone variant induced changes via quantitative proteomics. *Critical reviews in biochemistry and molecular biology* 2011, 46, 284–94.

[62] Eberl, H.C., Mann, M., Vermeulen, M., Quantitative proteomics for epigenetics. *Chembiochem : a European journal of chemical biology* 2011, 12, 224–34.

CHAPTER 2

Quantification of histone modifications using ^{15}N metabolical labeling

2.1 Summary

Nuclear DNA in eukaryotic cells is assembled into the hierarchical chromatin structure, a process that is dynamically affected by the combinatorial set of post-translational modifications (PTMs) of histones in a dynamic manner responsive to physiological and environmental changes. Precise quantification of these complex modifications is challenging. Here we present a robust MS-based quantitative proteomics method for studying histone PTMs, using ^{15}N metabolically labeled histones as the internal reference. This approach identified Txr1p as a histone methyltransferase in *Tetrahymena thermophila* and characterized the relationships of the Txr1p and Ezl2p methyltransferases to histone H3 modification. We identified 32 PTMs in more than 60 tryptic peptides from histone H3 of the ciliate model organism *Tetrahymena thermophila*, and quantified them (average CVs: 13%). We examined perturbations to histone modification patterns in two knockout strains of SET domain-containing histone methyltransferases (HMT). Knockout of *TXR1* leads to progressively decreased mono-, di-, and tri-methylation of H3K27 and apparent reduced monomethylation of H3K36 *in vivo*. On the contrary, *EZL2* knockout resulted in dramatic reduction in both di- and tri-methylation of H3K27 *in vivo* while the levels of monomethylation of H3K27 increased significantly. This buildup in monomethyl-H3K27 is consistent with it being a substrate

for Ezl2p. These results were validated by immuno-blotting using modification site-specific antibodies. Taken together, our studies define Txr1p as a H3K27 monomethylation-specific HMT, which facilitates the build-up of H3K27 di- and tri-methylation by the canonical H3K27-specific HMT, Ezl2p. Our studies also delineate some of the inter-dependencies between various H3 modifications, as compensatory increases in monomethylation at H3K4, H3K23, and H3K56 were also observed for both *TXR1* and *ELZ2* mutants.

2.2 Introduction

Histones, especially their N-terminal tails, are subject to various covalent post-translational modifications (PTMs) including acetylation, methylation, phosphorylation, ubiquitination, and citrullination [1–3]. A combinatorial set of PTMs on one or more histones, deposited by histone modifying enzymes, effectively serve to modulate various DNA pathways, including gene expression and replication as postulated in the histone code hypothesis [4,5]. Prominent among the PTMs is the reversible epigenetic mark, lysine methylation, present in mono-, di-, and tri-methylation states. Methyl groups are added to the ε-amine of the lysyl residue by histone methyltransferases (HMT) and removed by histone demethylases (HDMT)[6–9]. Different lysine methylation states are often associated with different, sometimes even opposite, biological functions. For example, monomethylation of histone residues H3K9, H3K27, and H4K20 are linked to active transcription, while their trimethylation states are associated with transcriptional repression [6,10]. The functional distinction of the different methylation states is further underscored by the presence of divergent state-specific HMTs, such as SETDB1/SETDB2 (for H3K9Me1) and SUV39H1/SUV39H2 (for H3K9Me2 and

H3K9Me3) [11], SETD8 (for H4K20Me1) [12] and SUV4-20 (for H4K20Me2 and H4K20Me3) [13]. It is a major challenge in the field of epigenetics to unravel the mechanism regulating histone lysine methylation events, which are dynamically affected by many factors and implicated in various biological processes. The modifications of histone lysines are dynamic and mono-, di-, tri-methylated residues are generally considered to be progressively methylated *in vivo*[14].

Most HMTs for lysine methylation contain a conserved catalytic domain, the SET (Suppressor of variegation, Enhancer of Zeste, Trithorax) domain[15]. Histone methyltransferases (HMT), in particular lysine methyltransferases, have been implicated in human diseases, including cancers[16]. In the past decade, a large number of HMTs have been identified in a wide-range of eukaryotic organisms, and are classified according to their sequence homology into sub-families, whose members generally share the substrate specificity. ATXR5 (*Arabidopsis* Trithorax Related 5) and ATXR6, the founding members of a recently identified HMT sub-family, were first isolated as PCNAs (proliferating cell nuclear antigen) interacting proteins in *Arabidopsis thaliana*[17]. Both ATXR5 and ATXR6 feature a divergent SET domain[17], a PHD (plant homeo-domain) finger that binds the modified histones[18–21], and a PIP (PCNA-interacting protein) box that binds PCNA[22]. Homologues to ATXR5 and ATXR6 are found in plants but not in animals. The atxr5 atxr6 double mutant exhibits a reduction of H3K27Me1 levels[23], supporting that ATXR5 and ATXR6 function as the H3K27 monomethylation-specific HMTs.

Txr1p (*Tetrahymena* Trithorax Related 1) was recently identified as a putative HMT in the ciliate model organism, *Tetrahymena thermophila* (encoded by the *TXR1* gene) by

homology to *Arabidopsis* ATXR5 and ATXR6. Txr1p carries two PHD domains (PHD1 and PHD2), one PIP box (QKLIEDYF), and one C-terminal SET domain (Figure 2.1), all of which are consistent with it being a *bona-fide* member of the ATXR5/ATXR6 sub-family of HMTs. In *Tetrahymena*, there are also three homologues of the canonical H3K27-specific HMT Enhancer of Zeste, referred to as *EZL1*, *EZL2* and *EZL3*, respectively [24,25]. Only *EZL2* is expressed at significant levels [24,25] and required for H3K27 di- and tri-methylation in asexually dividing cells (see below).

Mass spectrometry (MS) has played an important role in the study of histone PTMs because: 1) MS is capable of simultaneously monitoring multiple PTMs; 2) it can identify and quantify known and unknown PTMs in histones that cannot be easily determined by other analytical approaches such as micro-sequencing by Edman degradation or immuno-blotting with modification site-specific antibodies[26–30]. Quantification of histone PTMs can be achieved by a label-free strategy based on the relative intensities of extracted ion chromatograms (XICs) of precursors[31], or more accurately by stable isotope labeling techniques such as SILAC or iTRAQ[32–35]. However, analysis of histone PTMs by LC-MS is particularly challenging due to the enormous number of isoforms generated by the combination of various densely deposited PTMs[29]. The problem is further exacerbated by the basicity of histones, which, after trypsin digestion, generate peptides too small or hydrophilic to be effectively retained on reversed-phase HPLC columns and analyzed by MS. Chemical derivatization by propionic anhydride was introduced to alleviate some of these challenges [36,37]. Briefly, the propionylated histones will only be cleaved after arginyl residues when digested with trypsin and thus generate nicely sized, more hydrophobic peptides that can be more readily analyzed by

LC-MS. When chemically labeled with light and heavy isotopes (d0/d10-propionic anhydride), the relative levels of individual modifications from the two samples can be quantified by MS[32]. The deuterium isotope effect on chromatography and the variations of differential labeling from one peptide to another, however, place some limitations on precise quantification of histone PTMs. If metabolic labeling is used for quantification, variations in propionylation can be minimized because different physiological samples are combined prior to reaction with propionic anhydride.

The major aims of this study were to determine the roles of the Txr1p and Ezl2p methyltransferases in shaping histone modification patterns, with the focus on H3. This was achieved through quantifying the levels of various histone modifications in *TXR1* and *EZL2* knockout cells, as well as the wild-type cells. For this purpose, we developed a robust MS-based quantitative proteomics method for the study of histone PTMs using ^{15}N metabolically labeled histones as internal standards spiked into histone preparations as references. The general strategy and experimental design for this uniform labeling technique are illustrated in Figure 2.2. Similar studies using isotope labeled tissue or cells as global internal standards have been applied in mammals[38,39]. Overall, more than 60 unique H3 tryptic peptides were successfully quantified using this technique with small statistical variation (average CVs: ~13%). Similar to Super-SILAC which combines a mixture of several stable-isotope labeled cell lines to serve as internal standards for mass spectrometry-based analysis[40], our method provides a cost-effective alternative to study protein PTMs, especially in systems for which the SILAC medium is not available or cannot be easily formulated, and iTRAQ-based quantitative proteomics techniques are not easily applicable.

2.3 Experimental procedures

2.3.1 Construction of TXR1 and EZL2 knockout strain

To generate the *TXR1* and *ELZ2* knockout constructs, the genomic regions flanking *TXR1* or *ELZ2* were PCR amplified from the wild-type *Tetrahymena* cells and fused with the neo4 cassette[41], which confers paromomycin resistance to *Tetrahymena* cells. The constructs were introduced into *Tetrahymena* cells by standard biolistic transformations[42]. Transformants were selected for paromomycin resistance and complete replacement was finally confirmed by quantitative PCR.

2.3.2 Cell culture and ^{15}N metabolic labeling of *Tetrahymena* cellular proteome

Tetrahymena thermophila wild-type strains CU428 (*Tetrahymena* Stock Center), *TXR1* and *ELZ2* knockout strains were grown in 1× SPP medium (2% Protease Peptone, 0.2% Dextrose, 0.1% Yeast Extract, 0.003% Sequestrine) at 30°C with gentle shaking. Logarithmic-phase cells (2×10^5/ml) were collected for subsequent experiments.

For ^{15}N labeling of wild-type *Tetrahymena* cells, ^{15}N-labeled *Escherichia coli* BL21 cells were grown in the [^{15}N] M9 minimal medium (30mM Na_2HPO_4, 2g/L KH_2PO_4, 0.5g/L NaCl, 300μM, Na_2SO_4, 1mM $MgSO_4$, 0.3mM $CaCl_2$, 1μg/ml Biotin, 1μg/ml, Thiamine, 10g/L Glucose, and 1g/L $(^{15}NH_4)_2SO_4$ (Cambridge Isotope Laboratories), supplemented with ^{15}N-substituted Bioexpress (Cambridge Isotope Laboratories). Briefly, *E. coli* BL21 cells were inoculated into a small LB starter culture, and incubated at 37°C with vigorous shaking until it reaches the logarithmic-phase ($O.D._{600}$: 0.5 to 1). 1.0 ml of the starter culture was inoculated into 500 ml of the [^{15}N] M9 media, and incubated overnight at 37°C with vigorous shaking. Stationary-phase *E. coli* BL21 cells were collected by centrifugation and the cell pellet was resuspended in 500 ml of 1× phosphate

buffer (0.2g/L KCl, 1.15g/L Na_2HPO_4, 0.2g/L KH_2PO_4). Into this labeling medium were inoculated *Tetrahymena* cells from a logarithmic-phase small starter culture grown in 1× SPP medium. The culture was incubated at 30°C with gentle shaking for 48 hours, with ^{15}N-labeled *Escherichia coli* BL21 cells as the only nitrogen source. The labeled *Tetrahymena* cells were collected by centrifugation for subsequent nuclear preparations.

2.3.3 Nuclear preparation, histone acid extraction, and HPLC purification

The procedure for isolating the macronuclei from *Tetrahymena* cells was adapted from a protocol as reported[43]. Briefly, *Tetrahymena* cells were resuspended in 200 ml of medium A (0.1 M sucrose, 2 mM $MgCl_2$, 4% gum arabic, 10 mM Tris, 5mM EDTA, 10 mM butyric acid, 1mM iodoacetamide, 1mM PMSF, adjusted to pH 6.5). Cells were disrupted by vigorous blending in the presence of 1-octanol (0.7 ml). *Tetrahymena* macronuclei were pelleted by differential centrifugation. They were then acid-extracted with 1 ml of 0.4 N sulfuric acid, as reported [44]. The acid-extracted histones were precipitated by TCA (20% w/v). After washing once with acidified actone (0.2% HCl) and once with acetone, the histone samples were air-dried and resuspended in 500 µl of water.

The histone samples were further purified on a C8 reversed-phase HPLC column (Vydac Part No. 208TP54, 250mm× x 4.6mm) on a Rainin Rabbit HPLC with 5 ml/min pump heads, with the HPLC run conditions as reported[45]. Briefly, the HPLC column was equilibrated with 100% solvent A (5% acetonitrile in 0.1% TFA) for 5 min, then increased to 35% solvent B (90% acetonitrile in 0.1% TFA) for another 5 min. A 60 min gradient to 65% solvent B was applied to elute core histones from the column. Finally, the column was washed with 100% Solvent B for 10 min followed by re-equilibration

with 100% Solvent A. HPLC fractions were vacuum-dried, resuspended in deionized water, and evaluated by 15% SDS-PAGE, and those containing individual histones were combined (Figure S2.1). Concentrations of the purified histones were determined by the Bradford method (Bio-Rad).

2.3.4 Chemical derivatization, protein digestion, and quantitative mass spectrometry analysis of histone PTMs

For each biological replicate (n=3), 5 μg of histone H3 from wild-type, Δ*TXR1* or Δ*EZL2* cells grown in 1× SPP medium was mixed with an equal amount of ^{15}N-labeled H3 separated and purified from wild-type cells. The two-step chemical derivatization of histone H3 with propionic anhydride, adopted from Garcia *et al.*[37], was performed before and after trypsin digestion to increase the hydrophobicity of tryptic peptides as histones are very basic proteins with relatively short retention times on a reversed-phase column. Briefly, dried histone H3 samples were resuspended in 5μl, 100 mM ammonium bicarbonate. Samples were then treated with 20 μl of propionylation reagent made with 3:1 (v/v) anhydrous methanol (Alfa Aesar) : propionic anhydride (Sigma Aldrich), immediately followed by addition of ~15 μl ammonium hydroxide (Sigma Aldrich) to raise the pH to 8.0. The reaction mixtures were incubated for 15 min at 50ºC and then concentrated to ~5 μl in a SpeedVac concentrator. The propionylation reaction was performed twice to ensure maximum conversion of primary amines to propionyl amides. Generally, more than 95% propionylation efficiency was achieved after 2 rounds of chemical derivatization. No significant evidence of Asn or Gln deamidation (a potential side reaction of propionylation) through database searches was found for H3 peptides. Propionylated samples were again brought up in 100mM ammonium bicarbonate buffer

and in-solution digested with sequencing-grade trypsin (Promega, Madison, WI) at a ratio of 1:20 (enzyme:substrate) at 37ºC, 6h. The reactions were quenched by TFA (10% w/v). A second round of propionylation as described above was performed to convert the newly generated N-termini to propionyl amides. Finally, the reaction mixture was vacuum-dried and reconstituted in 0.1% formic acid. After filtering (Millipore Ultracel YM-10), the sample was stored in a -20 degree freezer until the mass spectrometry analysis.

The histone-derived peptides were resolved with a C18 capillary column (3μm, 300Å, 150mm x 100μm; CVC Technologies) on an Eksigent HPLC with a non-ferrous solvent path. A linear gradient for peptide separation was used as follows: Run 100% solvent A (HPLC-grade water with 0.1% formic acid) in 5 min; Run a 0% to 40% gradient against solvent B (90% acetonitrile in 0.1% formic acid) in 90 min; and finally run 10 min 100% solvent B followed by 15 min re-equilibrium with 100% solvent A. The resolved peptides were then introduced into a Thermo Fisher Orbitrap XL mass spectrometer at 200 nl/min flow rate connected with a nano-ESI source operated in positive ion mode. The mass spectrometer was operated in data dependent mode at a resolution of 30,000 on MS1 followed by 8 CID tandem MS with 30% normalized collision energy. All precursor ions were placed on a dynamic exclusion list for 120 seconds. Two polydimethylcyclosiloxane (PCM) ions (m/z=429.088735 and 445.12002) were selected for internal mass calibration.

Raw data were processed in Mascot distiller (Matrix Sciences, Version 2.4) and spectra were searched against the NCBI Tetrahymena database using the Mascot search engine (Matrix Sciences, version 2.2.07). A false discovery rate was estimated from the protein decoy database. Mass error tolerance was 10ppm for the precursor ion and 0.8 Da for the

fragment ions. N-terminal propionylation was set as a fixed modification and variable modifications were as follows: Acetylation(K), Methylation (KR), Propionylation (K), Monomethylation & Propionylation (K, +70 Da), and Phosphorylation (STY). We allowed up to five missed cleavages for trypsin digestion to compensate for the expected missed cleavages due to propionylation and PTMs on lysyl residues. Peptides were analyzed and quantified using the ^{15}N metabolic labeling method in the Mascot Distiller software. Peptide ratios were normalized against total histone H3 as the light and heavy forms were equally loaded based on protein quantification (Bradford method) of histone samples and standards. In detail, all replicates were normalized by the weighted average of the ratios of all H3 peptides according to their intensities. That is, normalization forces the average peptide ratio of L/H to be 1. Statistical analysis was done in Microsoft Excel or R. All spectra assigned with PTMs were manually validated based on the following criteria: 1) Most abundant ions should be assigned as b or y ions; 3) They should have more than 3 spectra observed; 4) Generally, there were at least 3 consecutive peaks identified bracketing an assigned PTM-residue; 5) Precursor has a mass error less than 6 ppm and fragment ion mass accuracy is less than 0.8 Da. 6) Rules such as proline effect, loss of water on ST, and loss of ammonia on K,R,N,Q were also taken into consideration.

2.3.5 Immuno-blot analysis of H3K27 methylation and acetylation

Wild-type, Δ*TXR1* or Δ*EZL2* histone H3 were resolved by 15% SDS-PAGE and transferred onto the Immobilon® P membrane (Millipore). Blots were then washed, blocked with non-fat dry milk, and incubated overnight with rabbit polyclonal antibodies to H3K27Me1 (1:5000, Millipore, Cat. No.: 07-448), H3K27Me2 (1:5000, Millipore, Cat. No.: 07-452), H3K27Me3 (1:5000, Millipore, Cat. No.: 07-449), and mouse monoclonal

antibody to H3K27Ac (1:50000, Wako Chemicals, Cat. No.: 306-3484). Blots were washed and then incubated with appropriate peroxidase-labeled secondary antibodies. Finally, the signal was visualized and processed with the Bio-Rad imaging system, after developing using the ECL chemiluminescent reagent (GE Healthcare).

2.4 Results and discussion

We developed a MS-based quantitative proteomics method based on spiking uniformly ^{15}N-labeled histones into normally ^{14}N-labeled histone preparations to allow quantitative analysis of histone PTMs in *Tetrahymena thermophila*. In this method, wild-type *Tetrahymena* cells were metabolically labeled with ^{15}N stable isotopes as described in Materials and Methods and the ^{15}N-labeled histones purified as described. In parallel, wild-type, Δ*TXR1* and Δ*EZL2 Tetrahymena* cells were grown in the standard medium without ^{15}N and the histones were similarly purified. Purified histone H3 from wild-type or knockout cells was spiked with ^{15}N-labeled histone H3, which served as an internal reference in the experiment. This allows a single preparation of ^{15}N-labeled histones to be used across many biological experiments with normally-labeled histones, providing a common standard for quantification and minimizing one potential source of variation. The peptide ratios can be precisely determined from the isotope distributions of the light and heavy forms of the tryptic peptides. As both forms were treated under the same experimental conditions and ^{15}N-labeled H3 was equally spiked across all biological replicates prior to propionylation and trypsinization, variations of chemical reactions are minimized leading to more accurate determination of the ratios. Finally, the peptide ratio of wild-type vs Δ*TXR1* or Δ*EZL2* cell can be calculated from each ratio of light vs heavy peptide. The overall experimental procedure is outlined in Figure 2.2.

There are several reasons why we used this uniform metabolic labeling strategy instead of the general metabolic labeling technique in which cellular proteomes from two different physiological states are directly compared with each other as they were grown in chemically identical media. SILAC medium is not commercially available and cannot be easily prepared for growth of *Tetrahymena* cells. In addition, *Tetrahymena* cells grow poorly in chemically defined media. As *Tetrahymena* can feed on bacteria, we achieved metabolic incorporation of ^{15}N stable isotopes from *E.coli* grown with ^{15}N ammonium sulfate as the sole nitrogen source. However, preparation of ^{15}N-labeled *E.coli* and using those cells routinely as the food source for binary experiments on *Tetrahymena* histones would be expensive and would contribute to variability in labeling. In addition, growing *Tetrahymena* cells on *E. coli* greatly slows their growth rates and can potentially change the histone modification patterns. In our method, both control (wild-type) and knockout cells are grown in a standard medium without bacteria and peptides in each group are compared with the ^{15}N labeled histones which are equally spiked into the samples across all experiments. This provides an internal reference for histones from different groups to be compared to. It also removes the practical limitation on the number of different types of cells to be compared in commercial SILAC, where only a triplex experiment can be performed.

2.4.1 False discovery rate, sequence coverage, and number of PTMs identified and quantified in Histone H3

Overall, the False Discovery Rate (FDR) of peptide identifications evaluated from the decoy database is below 1% at the Mascot identity threshold and less than 3% at the homology threshold, respectively (see Table S2.1). The sequence coverage generated

from trypsin digestion is 76% for histone H3 (Figure 2.3), which includes the major H3 and two minor variants, H3.3 and H3.4[46,47]. A total of 64 chemically unique peptides have been successfully identified and statistically quantified in histone H3, 22 of which are unique to the minor variant H3.3. A total of 32 PTMs were successfully identified and localized in the major H3, H3.3, H3.4 forms and 18 of these were statistically quantified. The PTMs identified in this study were: K4Me1, K9Ac, K14Ac, K14Me1, K18Ac, K23Ac, K23Me1, K23Me2, K23Me3, K27Me1, K27Me2, K27Me3, K27Ac, K36Me1, K36Me2, R40Me2, K56Me1, R83Me1, K79Me1 in the major H3 histone; K9Ac, K14Ac, K18Ac, K23Ac, K23Me2, K27Me1, K27Me2, K36Me1, K36Me2, K37Me3, K56Ac in H3.3; and K36Ac, K36Me1 in H3.4 (Figure 2.3). The peptides containing H3K4 were short and hydrophilic, thus poorly retained in the RP-HPLC column and scarcely recovered by the MS. This accounts for the discrepancy between the well-documented presence of high levels of H3K4 methylation[48] and the low representation in our MS analysis. A list of all peptides and PTMs statistically quantified in the major H3 and H3.3 are shown in Tables 2.1 and S2.2 respectively. All PTMs identified from wild-type and Δ*TXR1* cells and their modification site information in all three H3 variants are listed in Table S2.3.

2.4.2 ^{15}N-labeled versus ^{14}N-labeled histone H3

$(^{15}NH_4)_2SO_4$ was used as the sole nitrogen source for metabolically labeling *E. coli* and ultimately *Tetrahymena*. Each protein or peptide will incorporate a variable number of heavy nitrogen atoms depending on the length and composition of amino acids present in the protein or peptide[49]. We then separated *Tetrahymena* cells from the bacteria, collected nuclei, extracted bulk histones from the nuclei, and isolated and purified

individual histones using reversed-phase HPLC as we did in wild-type and knockout cells. The ratios of heavy vs light peptides calculated using Mascot Distiller software were mostly over 99% (Figure S2.2, Table S2.4), indicating that metabolic labeling was highly efficient. This was also confirmed by the Mascot search results, which indicated that very few peptides could be identified as unlabeled.

In the next step, wild-type histone H3 from *Tetrahymena* grown on light medium was mixed with equal amount of ^{15}N-labeled histone H3. The ratios of light/heavy histone H3 were evaluated in three biological replicates. Only peptides having p-values less than 0.05 with a 20% threshold change from unity are considered to significantly change in the mutants. Mass spectrometric analysis of intact histone H3 also suggests that gross changes in histone modifications do not occur between heavy and light isotope forms of H3 (see Figure S2.3). More importantly, most of the peptides are quantified with relatively small statistical variations (average CV=~12.7% was observed for the wild-type cells). Note that two final ratios are provided in the tables to aid in validation, raw ratios and normalized ratios which compensate for small differences in histone levels between experiments. The normalized ratios are used in this discussion. Both the p-values and the CVs are reported for each experiment. The former are useful for identifying ratios which have changed significantly but less useful for those which have not and for which the CVs are more useful. A few peptides have relatively large CVs which may be ascribed to low intensity spectra, or trypsin cleavage variability (e.g., at –RR- sites where the products may be -R, -RR, R-, and -).

2.4.3 Relative quantification of histone H3 PTMs in wild-type and Δ*TXR1* cells

As illustrated in Figure 2.2, the light forms of histone H3 purified from wild-type and Δ*TXR1* cells were each equally mixed with the heavy form of histone H3. Both wild-type and Δ*TXR1* cells were evaluated in three biological replicates. The statistical differences in the ratios of light/heavy peptides from each group can be inferred from the biological replicates, while the final ratios of peptides in each group can be directly converted from their ratios of light vs heavy forms. Most of the peptides in Δ*TXR1* cells have ratios similar to their counterparts in wild-type cells (Tables 2.1 and S2.2). Most of peptides also exhibited small statistical variations within each group with average CVs of 12.7% and 14.2% achieved respectively for wild-type and Δ*TXR1*. Furthermore, the peptide ratios were quite consistent for different charge states or different degrees of propionylation such as KQLASKAAR, [Ac]**K**QLAS[Ac]**K**AAR, KQLAS[Ac]**K**AAR, KSAPATGGIKKPHR, [Me1]**K**SAPATGGIKKPHR, [Me2]**K**SAPATGGIKKPHR, [Me3]**K**SAPATGGIKKPHR, etc. Taken together, the use of uniformly labeled internal histone standards provides a robust, reproducible, MS-based quantitative proteomics method for studying histone PTMs in this species.

2.4.4 *TXR1* knockout leads to dramatic reduction of H3K27Me1

The *TXR1* gene encodes a putative HMT as suggested from sequence homology and domain structure (Figure 2.1), but its substrate specificity was unproven. Quantitative analyses of the post-translational modification patterns of histone H3 were performed in wild-type and Δ*TXR1* cells in order to ascertain the relative effects of the *TXR1* knockout on H3 modification patterns. We found that monomethylation of H3K27 ($_{pr/me1}K^{27}SAPATGGI_{pr}K^{36}{}_{pr}K^{37}PHR$) in the major H3 was significantly decreased (~70%,

p=8.29E-05, see Figure 2.4a). In addition, dimethylation of H3K27 ($_{pr/me2}K^{27}$SAPATGGI$_{pr}K^{36}{}_{pr}K^{37}$PHR) was decreased by ~40% (Figure 2.4b) and trimethylation of H3K27 ($_{pr/Me3}K^{27}$SAPATGGI$_{pr}K^{36}{}_{pr}K^{37}$PHR) was slightly depressed by ~10% (Figure 2.4c). Note that the significant amount of trimethylation at K27 remaining in the *TXR1* mutant indicates that while monomethylation of K27 by Txr1p contributes to the pool it is not the sole source of substrate for dimethyl- and trimethyl-H3K27. The corresponding unmodified peptides ($_{pr}K^{27}$SAPATGGI$_{pr}K^{36}{}_{pr}K^{37}$PHR, $_{2pr}K^{27}$SAPATGGI$_{pr}K^{36}{}_{pr}K^{37}$PHR) increased in ratio in Δ*TXR1* cells, consistent with under-methylation at K27. These results are summarized in Table 2.1. Monomethylation of H3K36 ($_{2pr}K^{27}$SAPATGGI$_{pr/me1}K^{36}{}_{pr}K^{37}$PHR) also appeared to be reduced in Δ*TXR1* cells (Figure 2.4d). However, as this peptide had the same m/z ratio and partially co-eluted with the peptide mono-methylated at K27 ($_{pr/me1}K^{27}$SAPATGGI$_{pr}K^{36}{}_{pr}K^{37}$PHR) on the C18 analytical column, we were not able to unambiguously conclude from this peptide that H3K36Me1 was also affected by *TXR1* knockout. Nonetheless, we also observed that a peptide mono-methylated at both K27 and K36 ($_{2pr/me1}K^{27}$SAPATGGI$_{pr/me1}K^{36}{}_{pr}K^{37}$PHR) had the largest decrease in Δ*TXR1* cells (with a ratio of 0.11, see Figure 2.4e), which is consistent with H3K36Me1 also being a target of Txr1p. Monomethylation of H3K27 in the H3.3 variant was also significantly reduced (by 60%), as suggested from the diagnostic peptide $_{pr/me1}K^{27}$SAPVSGGV$_{pr}K^{36}{}_{pr}K^{37}$PH$_{pr}K^{40}$FRPGTVALR unique to H3.3 (see Table S2.2). The corresponding unmodified peptide ($_{pr}K^{27}$SAPVSGGV$_{pr}K^{36}{}_{pr}K^{37}$PH$_{pr}K^{40}$FRPGTVALR) ratio was increased, which is also consistent with under-methylation of H3.3 K27 in Δ*TXR1* cells.

Additional changes in PTMs were also observed and are summarized in Table 2.1. Monomethylation of H3K4, H3K23, and H3K56 were increased in Δ*TXR1* cells (1.8, 2.5, 2.5, respectively). Increased off-site methylation was also observed for H3K4 and H3K23 in the Δ*EZL2* cells (see below). All this may represent either a compensatory mechanism or an indirect response. Acetylation levels of H3 appear to be essentially unchanged in both Δ*TXR1* and Δ*EZL2* cells.

2.4.5 Ezl2p is a histone methyltransferase specific for H3K27Me2/Me3 and its activity is not tightly coupled to Txr1p

Based on previous studies [24,25], Ezl2p is one of the homologues of the *Drosophila* Enhancer of Zeste, which is specific for H3K27 methylation. We applied the same strategy used for Δ*TXR1* to Δ*EZL2* for analysis of the global PTM profile changes in histone H3. In Δ*EZL2* cells, a significant decrease in H3K27 di- and tri-methylation (by >80%, $p<0.01$) and a significant increase in H3K27 monomethylation (by >70%, $p<0.01$) was observed (Table S2.5), presumably because the monomethylation state is no longer being converted to higher methylation states efficiently. Note that *EZL2* knockout did not completely eliminate H3K27 di- and tri-methylation, suggesting the presence of an alternative pathway. *EZL2* knockout also led to increased monomethylation at K4 and K23, as we observed in Δ*TXR1* cells (see above).

Principal component analysis (PCA) and a PCA biplot analysis of histone PTM data generated from wild-type and knockout cells revealed that Δ*TXR1* and Δ*EZL2* were not tightly correlated in their changes in H3 modification patterns (Figure 2.5. Note that cosines of angles between vectors reflect relationships between variables). This analysis is consistent with the observation that while Ezl2p may utilize monomethyl-H3K27

generated by Txr1p, this is not the sole feedstock for synthesis of dimethyl- and trimethyl-H3K27. Ezl2p may be the major source of trimethyl H3K27 but these observations are consistent with multiple pathways for formation of trimethyl-H3K27. PCA analysis also revealed that di- and tri-methylation of K27 were closely correlated, but they were less correlated with monomethylation of K27, suggesting that the three methylation states of K27 may be differentially regulated in *Tetrahymena*.

2.4.6 Validation of the changes in H3K27 methylation states by immuno-blotting

Following LC-MS analysis, we performed immuno-blotting analysis to validate the changes in H3K27 methylation states in wild-type, Δ*TXR1* and Δ*EZL2* cells with site-specific antibodies recognizing mono-, di-, and tri-methylated H3K27 (Figure 2.6). As controls, we also included antibodies against H3K27Ac and general H3 (no bias for modifications). The *TXR1* knockout exhibited attenuation of mono-, di-, and tri-methylation at H3K27, with the strongest effect being on monomethyl- and dimethyl-K27, respectively, and weak attenuation of trimethyl-K27. In contrast, the *ELZ2* knockout showed greatly reduced H3K27Me2 and H3K27Me3 levels, but significantly increased H3K27Me1 levels. Little effect was observed for H3K27Ac. The results are consistent with the MS-based quantification. Commercial antibodies against H3K36Me1 failed to detect the modification in *Tetrahymena* (data not shown), probably due to the significant sequence divergence around K36 in histone H3 of *Tetrahymena* and higher eukaryotes.

It is noteworthy that all unmodified peptides in the variant H3.3 were found as up-regulated in Δ*TXR1* cells and the increase in the H3.3 protein ratio in Δ*TXR1* vs wild-type is around 2.5 (see Table S2.2), suggesting that knockout of *TXR1* induces over-expression of H3.3, possibly as an effort to compensate for the loss of H3K27Me1.

However, no significant change in H3.3 was observed in Δ*EZL2* cells. It is also worth noting with regards to compensatory mechanisms that Δ*TXR1* cells also exhibit increases in H3 monomethylation at K4, K23, and K56.

2.5 Conclusions and implications

In this study, we have successfully characterized the global PTM changes in histone H3 of wild-type, Δ*TXR1*, and Δ*EZL2 Tetrahymena* cells. Application of our modified ^{15}N uniform labeling technique allows identification and quantification of a total of 64 H3-derived peptides, which covers 18 PTMs in the major H3 and its minor variant H3.3, with acceptable coefficients of variation. Our quantitative proteomics data along with the immuno-blotting validation constitute a body of evidence strongly supporting that Txr1p is a HMT mainly responsible for H3K27Me1 in *Tetrahymena*. The result is in line with Txr1p being a member of the ATXR5/ATXR6 sub-family of HMTs [23]. Indeed, our work provides the first evidence for the ancient origin in evolution and functinal conservatin of the ATXR5/ATXR6 sub-family. Interestingly, *TXR1* knockout may also lead to a significant reduction in H3K36Me1, potentially making H3K36 an alternative modification site for Txr1p and this sub-family of HMTs.

In addition, we have also defined the distinct roles of Txr1p and Ezl2p in regulating H3K27 methylation states. *TXR1* knockout exhibits a significant impact on H3K27Me1 levels, which we assume is direct. The diminishing effects on H3K27Me2 and H3K27Me3, as well the increase in unmodified H3K27 levels, are most likely the indirect consequences of the inhibition of monomethylation of H3K27. In contrast to *TXR1* knockout, *EZL2* knockout greatly reduces levels of H3K27Me2 and H3K27Me3, while significantly increases levels of H3K27Me1. Apparently, Ezl2p can build up H3K27Me2

and H3K27Me3 levels utilizing H3K27Me1 deposited by Txr1p. Most likely, Ezl2p can also direct monomethylate H3K27, thus responsible for the remaining H3K27Me1 in Δ*TXR1* cells. This also explains why the higher methylation states of H3K27 are not dramatically affected in the absence of Txr1p. On the other hand, H3K27Me2 and H3K27Me3 are also not completely abolished in Δ*EZL2* cells. This leaves open the possibility that Txr1p may be responsible for at least part of the higher methylation states. All these results are consistent with a model in which the H3K27 methylation states are jointly regulated by Txr1p and Ezl2p: Txr1p is mainly responsible for H3K27Me1 and plays a minor role in H3K27Me2/Me3, while Ezl2p is mainly responsible for H3K27Me2/Me3 and plays a minor role in H3K27Me1. As a theoretical alternative, there may be some additional HMTs that can affect H3K27 methylation states in *Tetrahymena*.

The differential impact of Txr1p and Ezl2p on the methylation states of H3K27 strongly suggests that they function in different pathways, and H3K27Me1 may have biological functions distinct from H3K27Me2 and H3K27Me3. This is supported by the observation that ATXR5 and ATXR6 are involved in replication control in *Arabidopsis* [50]. Furthermore, studies of histone modification turnover reveals significant accumulation of H3K27Me1 on newly-deposited H3 in the S-phase of the cell cycle [51]. All this potentially points to a novel role for H3K27Me1 in DNA replication, which is apparently separated from the well-established roles for H3K27Me2 and H3K27Me3 in transcriptional repression and heterochromatin formation.

Establishment of this uniform labeling method to identify the histone substrates and target sites for HMTs was the primary goal of our reseach and our initial focus on the Txr1p and Ezl2p methyl transferases on H3 was to fill a gap in our functional knowledge

and form a foundation for future studies. Methylation may occur via multiple routes and the relationships between histone modifications are complex. Development of an accurate quantitative method is a necessary part of clarifying these relationships for H3 and other histones. Our approach will now allow us to explore functional crossover between HMTs and their relative relationships as well as the *in vivo* relationships between methylation and other PTMs which is a continuing goal of our research. In continuing studies, we will explore the global changes of PTMs across other core histones.

Table 2.1 Quantification of all valid peptides and PTMs in histone H3 major form

					WT/^{15}N WT Mean (SD)	***TXR1*/^{15}N WT Mean (SD)**	**p-value***	**Final ratio *TXR1*/WT**	**Normalized *TXR1*/WT**
						Protein-level Ratio			
H3 major					1.26 (0.03)	1.01 (0.19)	0.13464	0.80	1.07
start	end	Sequence	Modifications**	z		Peptide Ratio-level Ratios			
3	8	TKQTAR	Me1(K4); Pr(N-term)	2	0.74 (0.06)	0.93 (0.04)	0.01503	1.26	1.78
9	17	KSTGAKAPR	Pr(N-term, K14)	2	0.94 (0.04)	0.68 (0.09)	0.022982	0.73	1.02
9	17	KSTGAKAPR	Pr(N-term, K9, K14)	1	0.82 (0.04)	0.53 (0.05)	0.002279	0.65	0.91
9	17	KSTGAKAPR	Ac(K14); Pr(N-term)	2	1.40 (0.35)	1.25 (0.11)	0.538428	0.89	1.24
18	26	KQLASKAAR	Ac(K18, K23); Pr(N-term)	1	1.19 (0.08)	1.17 (0.13)	0.816231	0.98	1.36
				2	1.15 (0.04)	1.16 (0.12)	0.872236	1.01	1.41
18	26	KQLASKAAR	Ac(K23); Pr(N-term, K18)	1	0.98 (0.04)	0.90 (0.11)	0.322315	0.91	1.27
				2	0.99 (0.03)	0.96 (0.11)	0.685553	0.97	1.35
18	26	KQLASKAAR	Ac(K18); Pr(N-term,K23)	2	0.99 (0.03)	0.96 (0.10)	0.669132	0.97	1.35
18	26	KQLASKAAR	Me1(K23); Pr(N-term, K18, K23)	2	0.63 (0.17)	1.13 (0.09)	0.019933	1.79	2.52
18	26	KQLASKAAR	Pr(N-term, K18, K23)	1	1.18 (0.05)	0.87 (0.10)	0.016895	0.74	1.02
				2	1.20 (0.05)	0.96 (0.17)	0.126337	0.80	1.12
27	40	KSAPATGGIKKPHR	Pr(N-term, K36, K37)	2	2.16 (0.40)	5.25 (1.03)	0.023132	2.43	3.34
				3	2.18 (0.48)	5.20 (0.98)	0.018246	2.39	3.29
27	40	KSAPATGGIKKPHR	Me1(K27); Pr(N-term, K36, K37)	2	1.36 (0.05)	0.25 (0.03)	8.29E-05	0.19	0.26
				3	1.30 (0.04)	0.27 (0.03)	1.47E-05	0.21	0.29
27	40	KSAPATGGIKKPHR	Me2(K27); Pr(N-term, K36, K37)	2	0.98 (0.08)	0.40 (0.05)	0.000824	0.41	0.56
				3	0.87 (0.03)	0.34 (0.04)	0.000115	0.40	0.55
27	40	KSAPATGGIKKPHR	Me3(K27); Pr(N-term, K36, K37)	2	0.63 (0.03)	0.44 (0.04)	0.040226	0.70	0.92
				3	0.61 (0.04)	0.38 (0.05)	0.005474	0.63	0.91
27	40	KSAPATGGIKKPHR	Me1(K27, K36); Pr(N-term, K36, K37)	2	0.59 (0.22)	0.05 (0.02)	0.049941	0.09	0.13
				3	0.74 (0.08)	0.06 (0.00)	0.004257	0.08	0.11
27	40	KSAPATGGIKKPHR	Ac(K27); Pr(N-term, K36, K37)	2	0.99 (0.10)	1.06 (0.09)	0.37916	1.08	1.50
27	40	KSAPATGGIKKPHR	Me1(K27); Pr(N-term, K27, K36, K37)	2	1.35 (0.04)	0.18 (0.02)	4.17E-05	0.13	0.18
27	40	KSAPATGGIKKPHR	Pr(N-term, K27, K36, K37)	2	2.19 (0.43)	4.98 (0.92)	0.020405	2.27	3.12
27	40	KSAPATGGIKKPHR	Me1(K36); Pr(N-term, K27, K36, K37)	2	1.30 (0.09)	0.15 (0.01)	0.001829	0.12	0.16

27	40	KSAPATGGIKKPHR	Me1(K27, K36); Pr(N-term, K27, K36, K37)	2	0.79 (0.10)	0.06 (0.00)	0.005714	0.08	0.11
27	40	KSAPATGGIKKPHR	Me1(K36); Me2(K27); Pr(N-term, K36, K37)	3	0.58 (0.08)	0.37 (0.02)	0.160328	0.64	1.02
28	40	SAPATGGIKKPHR	Pr(N-term, K36, K37)	2	4.92 (0.82)	4.98 (2.56)	0.978241	1.01	1.24
41	49	FRPGTVALR	Pr(N-term)	2	1.31 (0.15)	1.01 (0.17)	0.087001	0.77	1.07
53	63	KYQKSTDLLIR	Pr(N-term, K56)	2	1.24 (0.12)	1.00 (0.15)	0.093877	0.80	1.11
53	63	KYQKSTDLLIR	Me1(K56); Pr(N-term, K56)	2	0.66 (0.14)	1.17 (0.04)	0.019107	1.77	2.49
54	63	YQKSTDLLIR	Pr(N-term)	1	1.13 (0.01)	0.87 (0.01)	0.394434	0.77	0.90
				2	1.38 (0.21)	1.01 (0.40)	0.001585	0.73	1.23
70	83	LVRDIAHEFKAELR	Pr(N-term, K79)	2	1.06 (0.19)	1.02 (0.16)	0.825064	0.96	1.34
73	83	DIAHEFKAELR	Pr(N-term, K79)	1	1.26 (0.24)	1.04 (0.24)	0.569269	0.83	1.26
				2	1.13 (0.24)	1.03 (0.15)	0.461435	0.91	1.06
73	83	DIAHEFKAELR	Me1(R83); Pr(N-term, K79)	2	1.16 (0.03)	1.06 (0.25)	0.675039	0.92	1.17
116	128	RVTIMTKDMQLAR	Pr(N-term, K122)	2	0.99 (0.30)	0.99 (0.12)	0.746262	1.00	1.30
				3	1.00 (0.35)	0.92 (0.12)	0.982936	0.92	1.39
117	128	VTIMTKDMQLAR	Pr(N-term, K122)	1	1.05 (0.25)	1.10 (0.19)	0.808235	1.05	1.45
				2	1.07 (0.24)	1.06 (0.17)	0.959774	0.99	1.37

* Species with significant p-values ($p < 0.05$) are boxed. The p-value is calculated from peptides which are present at least twice in replicate WT or knockout cells. The average CV for peptides from histone H3 in WT and *TXR1* knockout cells is 12.7% and 14.2% respectively.

** PTM abbreviations: Ac: Acetylation; Me: Monomethylation; Me2: Dimethylation; Me3: Trimethylation; Pr: Propionylation.

Table S2.1 False discovery rate, number of PTMs identified and sequence coverage in WT and Δ*TRR1* cells

FDR (identity, %)	0.64(0.51)		
FDR (homology, %)	2.36(0.38)		
H3 isoforms	H3	H3.3	H3.4
Sequence Coverage	76%	76%	76%
# Peptides Statistically Quantified*	42	22	N/A
# PTMs Identified	20	10	2
# PTMs Statistically Quantified*	12	6	0

*: number of peptides/PTMs statistically quantified means that those peptides are presented at least twice in replicate wild-type or knockout cells.

Table S2.2 Summary of valid diagnostic peptides and PTMs quantified in variant H3.3 for *TRR1* mutant.

					WT/^{15}N WT Mean (SD)	*TRR1*/^{15}N WT Mean (SD)	p-value*	Final ratio *TRR1*/WT	Normalized *TRR1*/WT
						Protein-level Ratio**			
H3.3					0.78 (0.09)	1.41(0.15)	0.007	1.81	2.53
start	end	Sequence	Modifications	z		Peptide-level Ratios			
9	17	KSTGVKAPR	Pr(N-term,K9,K14)	1	0.57 (0.14)	0.72 (0.06)	0.199274	1.26	2.01
18	26	KQLATKAAR	Pr(N-term, K23)	2	0.64 (0.15)	1.06 (0.07)	0.023676	1.66	2.38
18	26	KQLATKAAR	Ac(K18,K23);Pr(N-term)	2	0.99 (0.03)	0.96 (0.11)	0.677255	0.97	1.35
18	26	KQLATKAAR	Ac(K23); Pr(N-term,K18)	1	2.29 (1.49)	0.8 (0.03)	0.39201	0.35	0.56
				2	1.20 (0.05)	0.96 (0.17)	0.118136	0.8	1.11
18	26	KQLATKAAR	Pr(N-term,K18,K23)	1	0.47 0.13)	0.82 (0.04)	0.033675	1.76	2.5
				2	0.59 (0.14)	1.05 (0.06)	0.020046	1.78	2.53
27	49	KSAPVSGGVKKPHKFRPGTVALR	Pr(N-term,K36,K37,K40)	3	0.83 (0.02)	3.81 (0.1)	0.000174	4.62	6.5
				4	0.84 (0.01)	3.89 (0.03)	1.80E-06	4.66	6.59
27	49	KSAPVSGGVKKPHKFRPGTVALR	Me1(K27);Pr(N-term,K36,K37,K40)	3	0.73 (0.15)	0.2 (0.02)	0.025915	0.27	0.39
27	49	KSAPVSGGVKKPHKFRPGTVALR	Me1(K27);Pr(N-term,K27, K36,K37,K40)	3	0.73 (0.15)	0.22 (0.01)	0.028321	0.3	0.43
27	49	KSAPVSGGVKKPHKFRPGTVALR	Me1(K27);Me1(K36);Pr(N-term, K36,K37,K40)	4	0.55 (0.12)	0.17 (0.09)	0.029351	0.3	0.43
27	49	KSAPVSGGVKKPHKFRPGTVALR	Ac(K37);Me1(K36);Pr(N-term,K36,K40)	3	0.84 (0)	3.86 (0.06)	0.009465	4.6	6.04
				4	0.85 (0.01)	3.89 (0.04)	0.005618	4.58	6.04
53	63	KYQKTTDLLIR	Ac(K56);Pr(N-term)	2	1.25 (0.12)	1 (0.15)	0.093929	0.8	1.11
53	63	KYQKTTDLLIR	Pr(N-term,K56)	2	0.66 (0.14)	1.17 (0.04)	0.018789	1.76	2.49
53	63	KYQKTTDLLIR	Pr(N-term,K1,K56)	2	0.68 (0.13)	1.18 (0.03)	0.01697	1.74	2.77
73	83	DIAMEMKSDIR	Pr(N-term,K79)	2	0.61 (0.19)	1.22 (0.04)	0.027935	2	2.83
116	128	RVTIMTKDLHLAR	Pr(N-term,K122)	2	0.60 (0.25)	1.17 (0.08)	0.047206	1.94	2.77
				3	0.58 (0.23)	1.16 (0.05)	0.042797	2.01	2.87
117	128	VTIMTKDLHLAR	Pr(N-term,K122)	2	0.64 (0.21)	1.27(0.06)	0.027222	1.98	2.8
117	129	VTIMTKDLHLARR	Pr(N-term,K122)	2	0.48 (0.08)	1.11 (0.07)	0.013342	2.33	3.68

* Species with significant p-values ($p < 0.05$) are boxed. The p-value is calculated from peptides which are present at least twice in replicate WT or KO cells. The average CV for peptides from histone H3.3 in WT and *TRR1* knockout cells is 17.9% and 6.3% respectively.

** Protein Ratios are calculated from diagnostic peptides to H3.3.

Table S2.3 Summary of all PTMs identified in WT and knockout cells

H3 major									
Start	End	Sequence	Modifications	z	Observed	ΔM*	Ion Score	Intensity	RT(min)
3	8	TKQTAR	K4Me1	2	387.7275	1	37	9.46E03	41.1
9	17	KSTGAKAPR	K9Ac	2	542.2835	1	29	5.49E04	53.6
9	17	KSTGAKAPR	K14Ac	2	514.2699	0	63	3.51E04	48.4
9	17	KSTGAKAPR	K14Me1	2	556.2968	-3	42	5.96E05	64.6
18	26	KQLASKAAR	K18Ac	2	571.3108	1	47	2.10E05	78.1
18	26	KQLASKAAR	K18Ac, K23Ac	2	556.8246	0	56	3.10E05	74.5
18	26	KQLASKAAR	K23Ac	2	563.8319	-1	57	2.31E05	78.1
18	26	KQLASKAAR	K23Me1	2	557.3122	-1	53	3.60E05	61.9
18	26	KQLASKAAR	K23Me2	2	556.8409	-4	28	8.62E05	71.4
18	26	KQLASKAAR	K23Me3	2	571.3286	0	34	2.37E05	69.9
27	40	KSAPATGGIKKPHR	K27Me1	2	815.476	1	83	8.13E05	70.9
27	40	KSAPATGGIKKPHR	K27Me2	2	822.4858	3	96	2.59E04	63.1
27	40	KSAPATGGIKKPHR	K27Me3	2	829.4924	2	62	2.58E04	64.1
27	40	KSAPATGGIKKPHR	K27Ac	2	829.474	1	74	6.57E05	69.3
27	40	KSAPATGGIKKPHR	K27Me1, K36Me1	2	850.4947	-2	86	3.06E04	77.5
27	40	KSAPATGGIKKPHR	K27Me2, K36Me1	2	840.4605	2	72	5.64E04	67.5
27	40	KSAPATGGIKKPHR	K27Me1, K36Me2	2	829.4922	1	67	1.02E05	72.4
27	40	KSAPATGGIKKPHR	K27Ac, K36Me2	2	815.4771	2	62	1.56E04	62.9
27	40	KSAPATGGIKKPHR	K27Me1, K36Me1, K37Me3	2	843.5031	-4	60	9.49E04	80.8
27	40	KSAPATGGIKKPHR	K36Me1	2	843.4901	2	83	1.46E05	74.3
27	40	KSAPATGGIKKPHR	K36Me2	2	794.4721	3	55	1.23E05	63.3
27	49	KSAPATGGIKKPHRFRPGTVALR	K27Me1, R40Me2	3	885.8521	-3	59	3.03E05	82.8
27	49	KSAPATGGIKKPHRFRPGTVALR	K27Me1, K36Me1, R40Me2	3	890.5257	-4	64	1.33E05	83.1
53	63	KYQKSTDLLIR	K56Me1	2	745.9297	-4	67	1.23E06	84.9
73	83	DIAHEFKAELR	K79Me1	2	727.8858	0	57	1.62E05	105.1
73	83	DIAHEFKAELR	R83Me1	2	727.8863	1	54	1.16E05	107.1
H3.3									
9	17	KSTGVKAPR	K9Ac, K14Ac	2	549.2912	1	45	2.86E05	56.4

9	17	KSTGVKAPR	K14Ac	2	556.2992	1	47	8.09E05	65.6
18	26	KQLATKAAR	K18Ac, K23Ac	2	563.8322	-1	56	6.56E05	78.7
18	26	KQLATKAAR	K23Ac	2	570.8404	0	62	1.39E05	93.2
18	26	KQLATKAAR	K23Me2	2	571.329	1	48	5.44E05	73.6
27	49	KSAPVSGGVKKPHKFRPGTVALR	K27Me1	3	885.8548	2	121	3.09E05	84.9
27	49	KSAPVSGGVKKPHKFRPGTVALR	K27Me2	4	668.1464	1	63	6.54E05	79.2
27	49	KSAPVSGGVKKPHKFRPGTVALR	K36Me1	3	904.5299	1	77	1.89E05	88.9
27	49	KSAPVSGGVKKPHKFRPGTVALR	K36Me2	3	871.8516	2	58	3.46E04	79.5
27	49	KSAPVSGGVKKPHKFRPGTVALR	K27Me1, K36Me1	3	890.5257	0	118	1.33E05	83.1
27	52	KSAPVSGGVKKPHKFRPGTVALREIR	K36Me1, K37Ac	3	1013.9262	2	52	1.37E05	85.0
28	49	SAPVSGGVKKPHKFRPGTVALR	K36Ac, K37Me1	3	838.4809	-3	59	2.65E04	84.8
53	63	KYQKTTDLLIR	K56Ac	2	738.9255	1	74	6.51E04	93.1
H3.4									
27	49	KSAPISGGIKKPHKFRPGTVALR	K36Ac	3	904.5299	1	55	1.89E05	89.0
27	49	KSAPISGGIKKPHKFRPGTVALR	K36Me1	3	895.1992	2	41	2.28E04	87.1

*: Mass error is measured as parts per million (ppm).

Table S2.4 Full list of all ^{15}N H3 peptide ratios (Heavy vs Light)

H3

Start	End	Sequence	Modifications	z	H/L
9	17	KSTGAKAPR	Ac(K14)	2	0.94451
10	17	STGAKAPR	Ac(K14)	2	0.999426
18	26	KQLASKAAR	Ac(K18,K23)	1	0.992764
18	26	KQLASKAAR	Ac(K23)	2	0.994031
18	26	KQLASKAAR	Me2(K23)	2	0.999974
18	26	KQLASKAAR	Me3(K23)	2	0.998306
19	26	QLASKAAR	Ac(K23)	2	0.97774
27	36	KSAPATGGIK	Ac(K27)	1	0.999981
27	36	KSAPATGGIK	Me2(K27)	2	0.98534
27	36	KSAPATGGIK	Me1(K27)	1	0.999974
27	36	KSAPATGGIK	Me1(K27)	2	0.998732
27	36	KSAPATGGIK	Me3(K27)	2	0.991533
28	36	SAPATGGIK		1	0.98837
27	40	KSAPATGGIKKPHR	Ac(K27),Me2(K36)	2	0.992491
27	40	KSAPATGGIKKPHR	Ac(K27),Me1(K36)	2	0.997828
27	40	KSAPATGGIKKPHR	Ac(K27),Me1(K36)	3	0.98475
27	40	KSAPATGGIKKPHR	Me2(K27),Me1(K36)	2	0.996233
27	40	KSAPATGGIKKPHR	Me1(K27),Me2(K36)	2	0.996233
27	40	KSAPATGGIKKPHR	Me1(K27),Me1(K36)	2	0.995498
27	40	KSAPATGGIKKPHR	Me1(K27),Me1(K36)	3	0.98896
27	40	KSAPATGGIKKPHR	Me1(K36)	2	0.999973
28	40	SAPATGGIKKPHR	Me2(K36)	2	0.991499
28	40	SAPATGGIKKPHR	Me1(K36)	2	0.998617
28	40	SAPATGGIKKPHR	Me1(K36)	3	0.98299
43	49	PGTVALR		1	0.997312
70	83	LVRDIAHEFKAELR		2	0.9886
70	83	LVRDIAHEFKAELR		3	0.99698
73	79	DIAHEFK		1	0.998278
73	79	DIAHEFK		2	0.992109
73	83	DIAHEFKAELR		1	0.999972
73	83	DIAHEFKAELR		2	0.993934
123	128	DMQLAR		1	0.95991
130	135	IRGERF		1	0.997647

H3.3

Start	End	Sequence	Modifications	z	H/L
18	26	KQLATKAAR	Ac(K18,K23)	1	0.99997
18	26	KQLATKAAR	Me2(K23)	2	0.998306
27	36	KSAPVSGGVK	Me2(K27)	1	0.999977
27	36	KSAPVSGGVK	Me2(K27)	2	0.99104
27	36	KSAPVSGGVK	Me1(K27)	1	0.999979
27	36	KSAPVSGGVK	Me1(K27)	2	0.998358
27	36	KSAPVSGGVK		1	0.999959
27	36	KSAPVSGGVK		2	0.999973
27	40	KSAPVSGGVKKPHK	Me1(K27),Me1(K36)	2	0.999706
28	40	SAPVSGGVKKPHK	Me1(K36)	2	0.997124
57	63	TTDLLIR		1	0.997749

57	63	TTDLLIR	2	0.997586
73	79	DIAMEMK	1	0.98676

Table S2.5 Peptides used in PCA analysis of 16 PTM sites in H3 and H3.3

Peptide ID	Modifications	WT	*TRR1*	p-value	*EZL2*	p-value	*TRR1*/WT	*EZL2*/WT
1	H3K4Me1	0.61(0.06)	1.09(0.24)	0.01503	1.36(N/A)	N/A	1.78	2.21
2	H3K14Un	0.68(0.04)	0.61(0.12)	0.022982	0.83(0.02)	0.88193	0.91	1.23
3	H3K14Ac	1.17(0.31)	1.45(0.22)	0.538428	1.22(N/A)	N/A	1.24	1.05
4	H3K18Ac	0.82(0.04)	1.11(0.16)	0.669132	0.88(N/A)	N/A	1.35	1.07
5	H3K18AcK23Ac	0.95(0.05)	1.34(0.19)	0.872236	1.12(0.22)	0.967862	1.41	1.18
6	H3K23Ac	0.83(0.04)	1.11(0.15)	0.685553	0.95(0.13)	0.394129	1.35	1.15
7	H3K23Me1	0.52(0.13)	1.32(0.25)	0.019933	1.55(0.47)	0.006643	2.52	2.97
8	H3K18UnK23Un	1(0.06)	1.11(0.21)	0.126337	1.15(0.08)	0.967352	1.12	1.16
9	H3K27Ac	0.82(0.08)	1.23(0.21)	0.37916	0.94(0.29)	0.759725	1.5	1.15
10	H3K27Me1*	1.12(0.05)	0.21(0.03)	4.17E-05	1.94(0.27)	0.000358	0.18	1.73
11	H3K27Me2	0.81(0.06)	0.46(0.07)	0.000824	0.13(0.01)	0.001446	0.56	0.16
12	H3K27Me3	0.52(0.02)	0.47(0.1)	0.040226	0.1(0.024)	0.000263	0.92	0.2
13	H3K27Me1K36Me1	0.65(0.07)	0.07(0.02)	0.004257	0.78(0.19)	0.864776	0.11	1.2
14	H3K27Me2K36Me1	0.48(0.06)	0.49(0.01)	0.160328	0.50(0.15)	0.606316	0.86	0.94
15	H3K27UnK36Un	1.83(0.38)	5.7(0.91)	0.020405	2.37(0.61)	0.61335	3.12	1.3
16	H3K56Un	1.03(0.12)	1.15(0.16)	0.093877	1(N/A)	N/A	1.11	0.97
17	H3K56Me1	0.55(0.11)	1.37(0.29)	0.019107	0.85(N/A)	N/A	2.49	1.54
18	H3R83Un	0.94(0.2)	1.18(0.17)	0.569269	1.1(0.01)	0.97823	1.26	1.17
19	H3R83Me1	0.96(0.04)	1.12(0.06)	0.675039	1.02(N/A)	N/A	1.17	1.07
20	H3.3K18AcK23Ac	0.82(0.04)	1.11(0.16)	0.677255	0.95(0.14)	0.008598	1.35	1.16
21	H3.3K23Ac	1(0.06)	1.1(0.2)	0.118136	1.16(0.08)	0.25722	1.11	1.16
22	H3.3K18UnK23Un	0.49(0.11)	1.23(0.32)	0.020046	1.59(0.48)	0.006213	2.53	3.26
23	H3.3K56Ac	1.04(0.12)	1.15(0.16)	0.093929	1.21(N/A)	N/A	1.11	1.17

*: The peptide, $_{(2pr/me1)}K^{27}SAPATGGI_{pr}K^{36}{}_{pr}K^{37}PHR$, was used to calculate the ratio of K27Me1 as it was present in both Δ*TRR1* and Δ*EZL2* cells.
Note: Ratios in ths table were normalized as described in the Materials and Methods. Species with significant p-values ($p < 0.05$) are boxed.

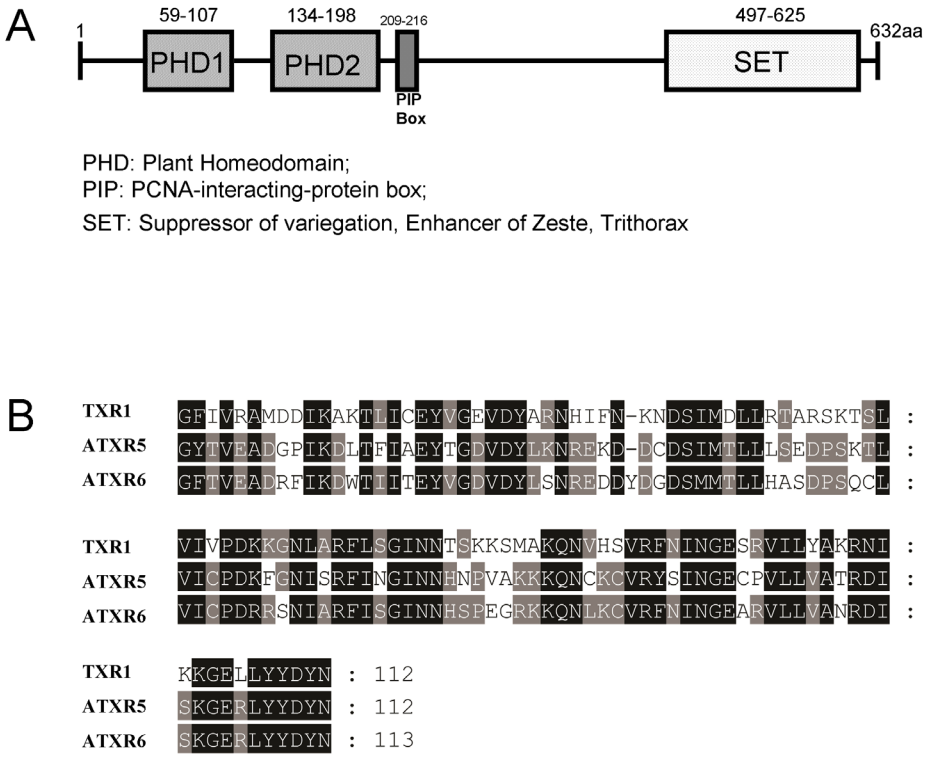


Figure 2.1 Domain structure analysis of protein Txr1p
a). The PIP box, C-terminal SET domain and two N-terminal PHD domains (PHD1 and PHD2) suggest Txr1p is a putative HMT homologous to the plant ATXR5/ATXR6; b). Amino acid sequence alignment of SET domains in Txr1p and *Arabidopsis* homologues ATXR5 and ATXR6.
Note: Identical and similar residues are darkly and lightly shaded, respectively.

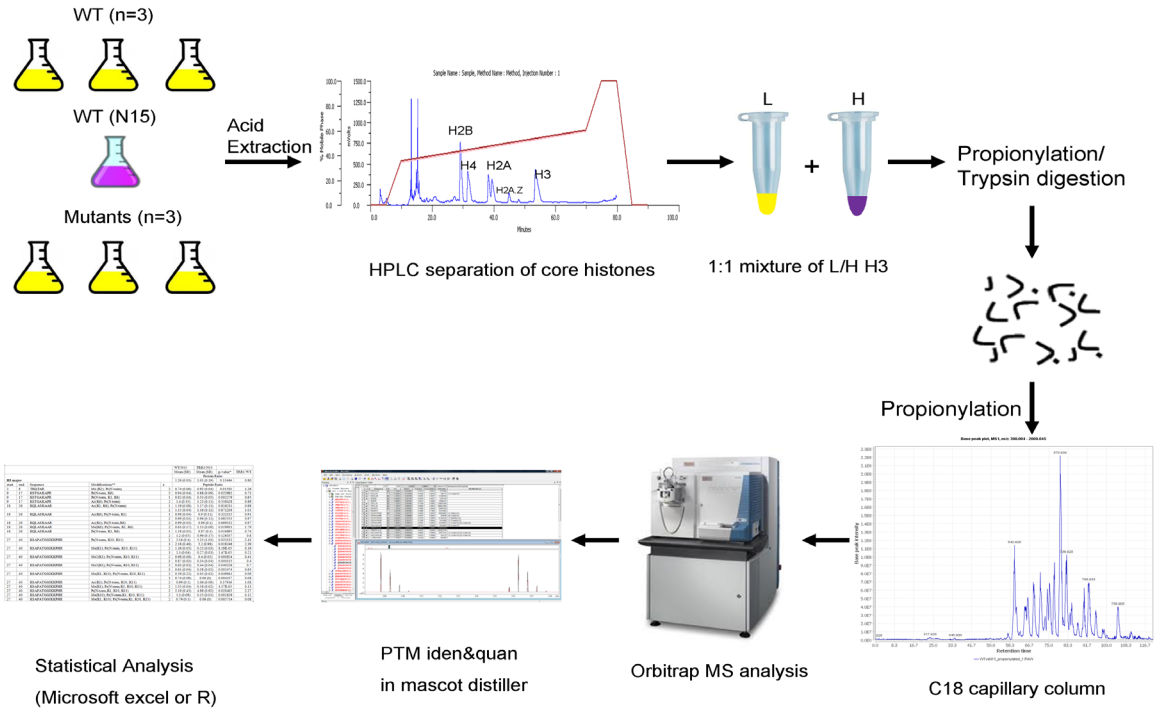


Figure 2.2 Experimental design

Three wild-type *Tetrahymena* cells and HMT mutants were normally grown in the SPP medium while one wild-type *Tetrahymena* cells were metabolically labeled with ^{15}N stable isotopes as a global reference. Once bulk histones were acid extracted, individual core histones were subsequently separated by reversed-phase HPLC. The light form and heavy form of H3 were equally mixed and chemically propionylated before and after trypsin digestion. The digests were resolved by a C18 capillary column before they were analyzed by mass spectrometer. A commercial software, Mascot distiller, was used to quantitate PTM changes in wild-type and knockout cells. Finally, normalization and statistical analysis of histone PTMs were performed by Microsoft excel or R.

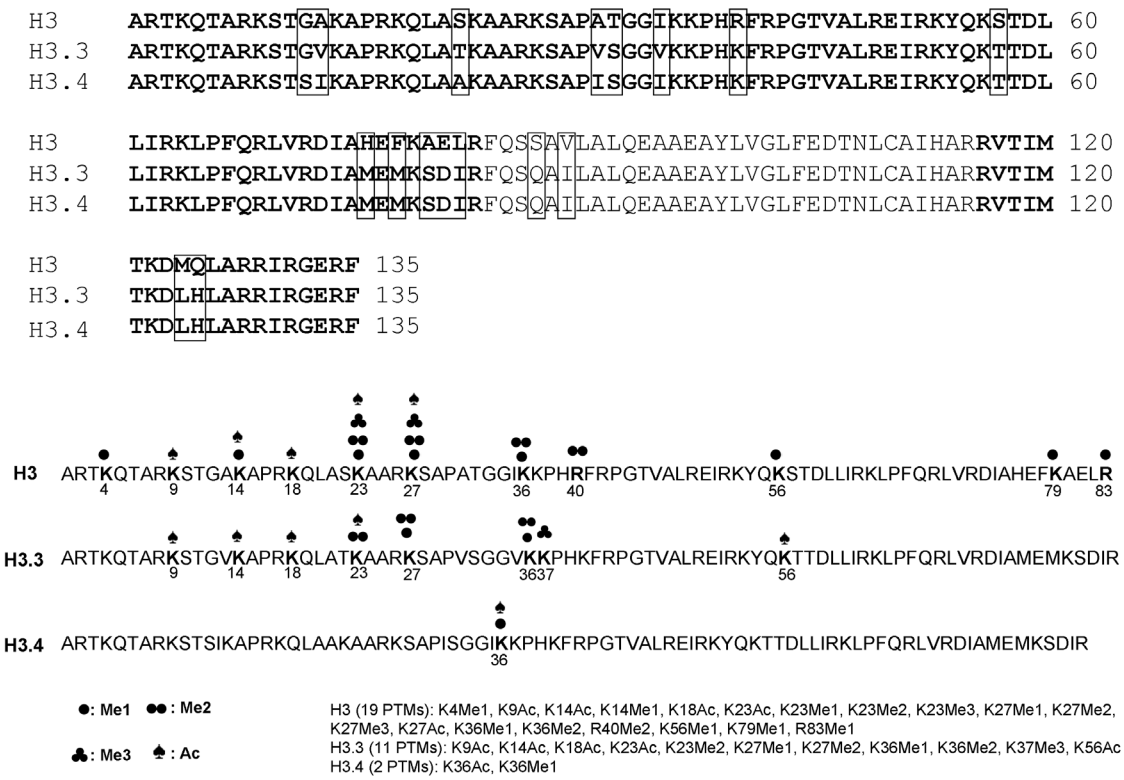


Figure 2.3 Histone H3 variants in *Tetrahymena thermophila*
About 76% sequence coverage (Bold letters) was achieved in all 3 histone H3 variants. Sequence variations in these variants are shown in rectangular frames. A total of 32 PTMs identified in 3 variants are labeled by different types of marks.

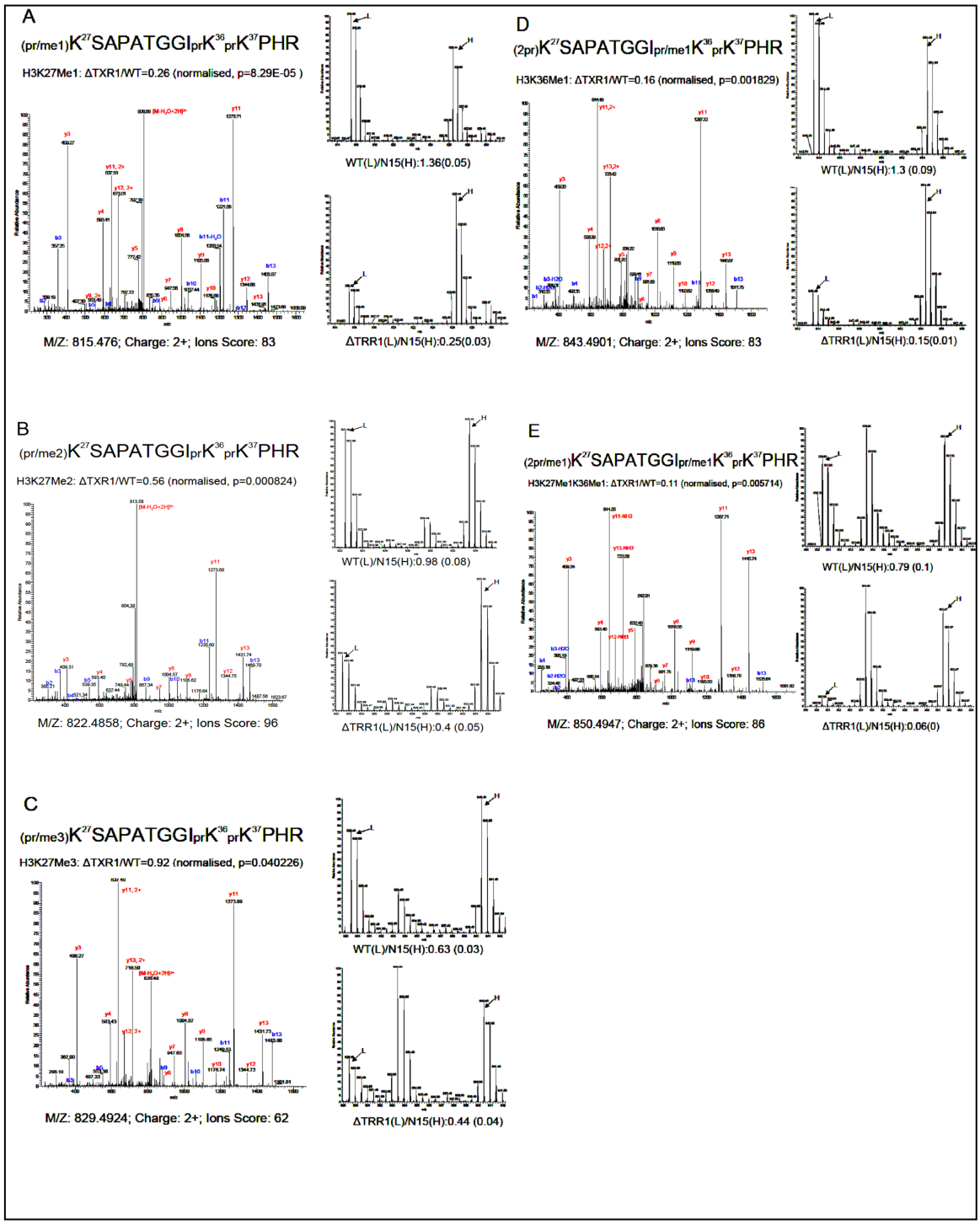


Figure 2.4 Analysis of Lys27 and Lys36 methylation by quantitative mass spectrometry
Monomethylation of H3K27 is signficantly down-regulated by ~70% in Δ*TXR1* cells a); Dimethylation of H3K27 is decreased by ~40% b), and trimethylation by ~10% c). Monomethylation of H3K36 may also be significantly affected as suggested from d) and e).

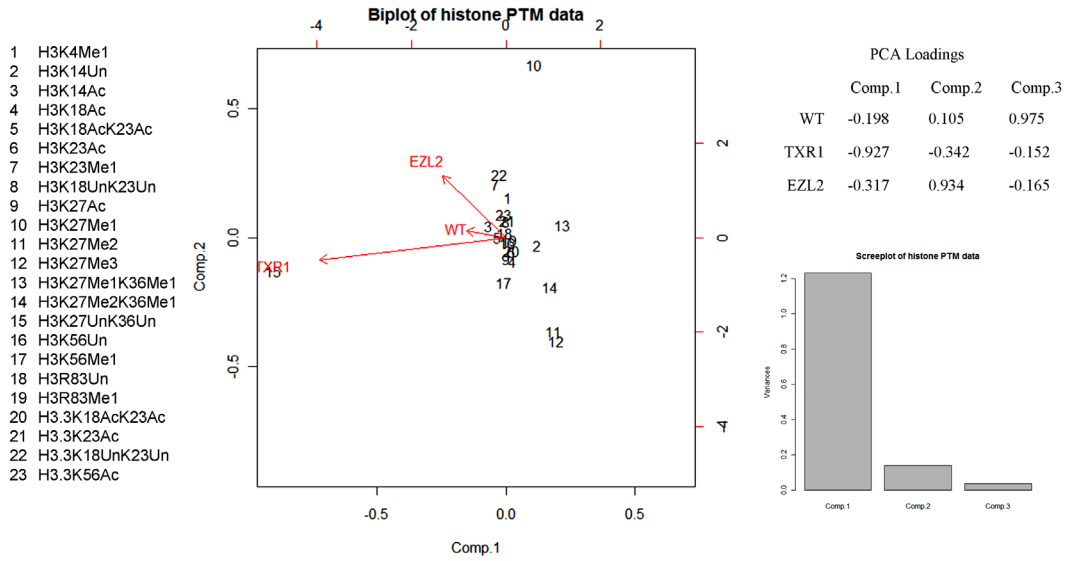


Figure 2.5 PCA biplot of histone PTM data

PCA and PCA biplot analysis of histone PTM data suggest that ΔTXR1 and ΔEZL2 act independently in regulation of H3K27 mono-, di- and trimethylation states. Di- and trimethylation of K27 are closely co-regulated and they have a far relationship with monomethylation of K27.

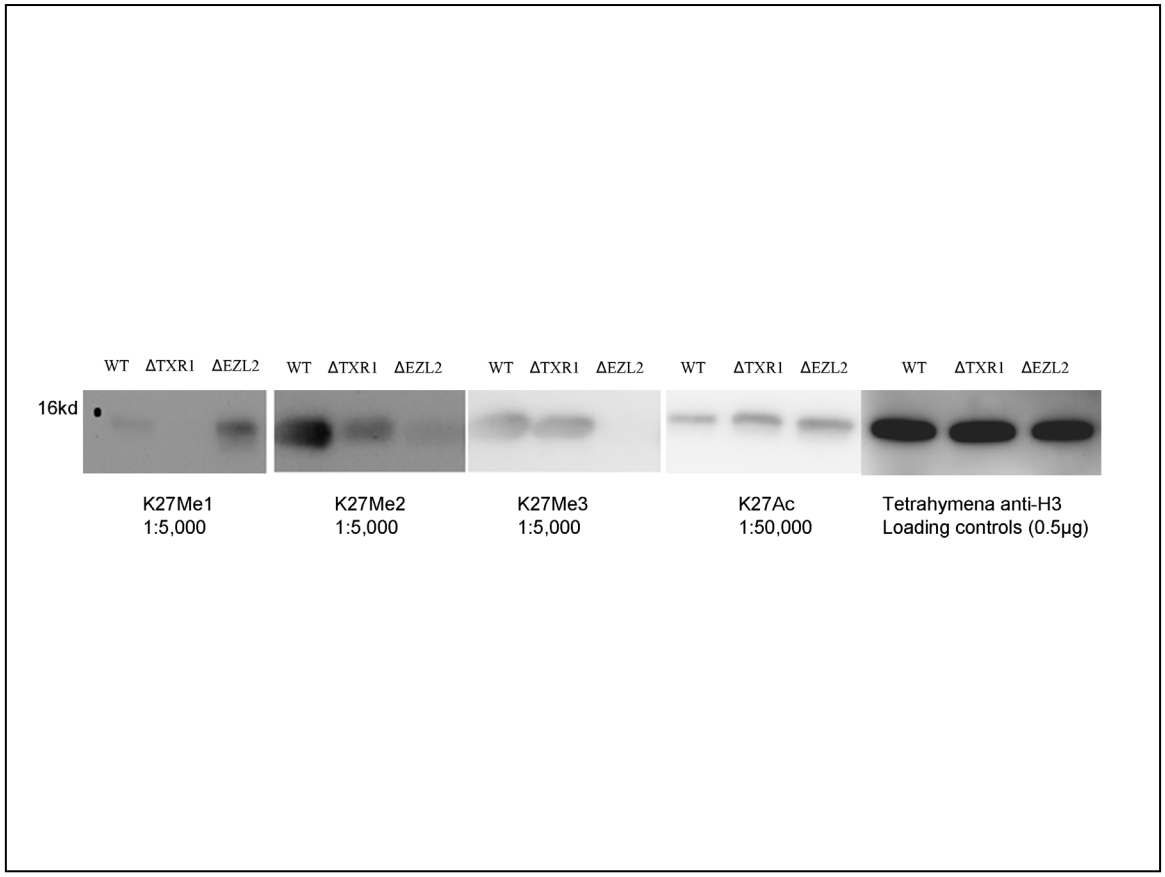


Figure 2.6 Validation of PTM expression by western blot
Rabbit polyclonal antibodies to H3K27Me1 (1:5000); H3K27Me2 (1:5000); H3K27Me3 (1:5000); and mouse monoclonal antibody to H3K27Ac (1:50000) were used to validate the PTM changes in wild-type and knockout cells with a 10 min exposure. Amount of H3 loaded: 0.5 μg. Elimination of *TXR1* shows a significant reduction of H3K27Me1/Me2 while knockout of *ELZ2* leads to reduction of H3K27Me2 and an undetectable signal in H3K27Me3. These results are consistent with the data from mass spectrometric quantification.

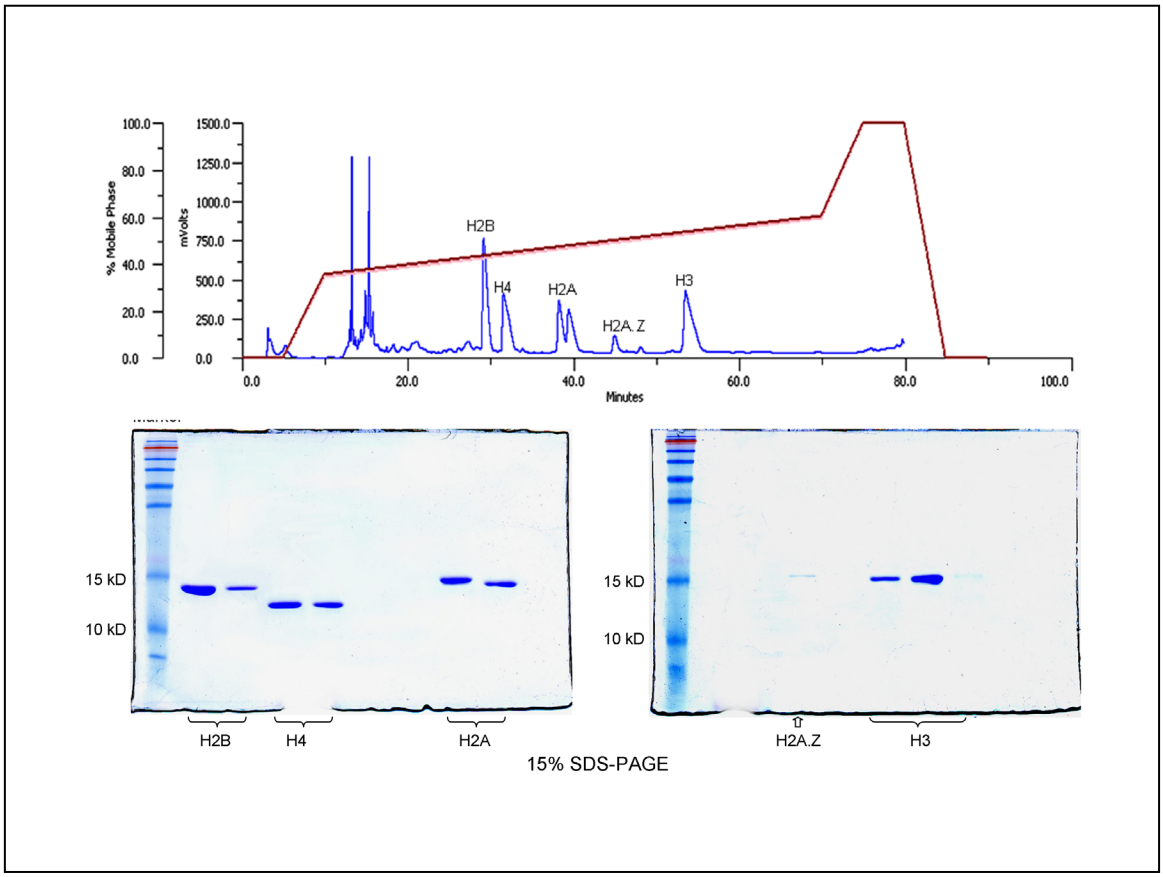


Figure S2.1 HPLC and SDS-PAGE separation of core histones

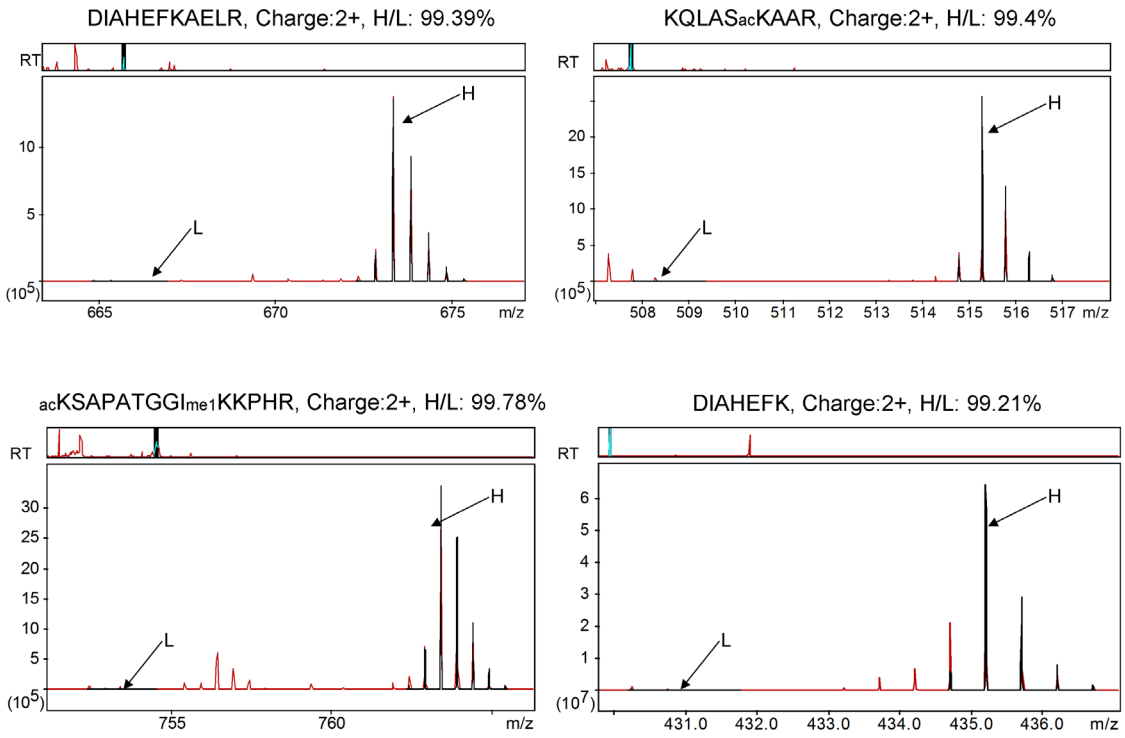


Figure S2.2 Labeling effeciencies of four representative ^{15}N-histone H3 peptides
To evaluate the incorporation of ^{15}N, unpropionylated ^{15}N-labeled H3 was digested with trypsin and ratios of H/L were quantified using Mascot Distiller. Analysis of raw MS spectra showed at least 99% incorporation suggesting that the labeling process was highly efficient.

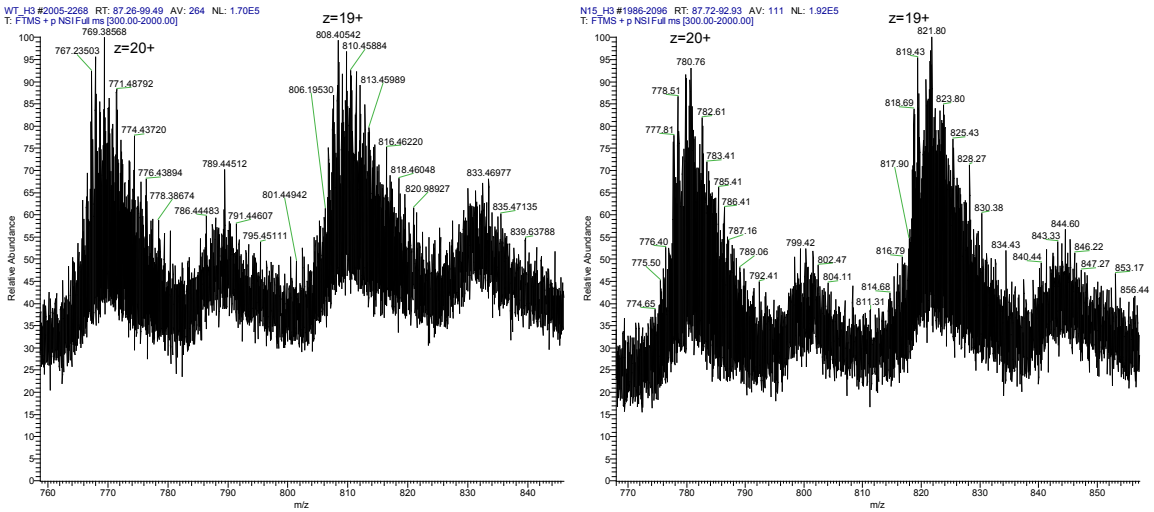


Figure S2.3 Mass analysis of WT and ^{15}N intact histone H3 illustrating similar profiles and complexity of PTMs

2.6 Appdendix-R code for PCA and biplot analysis

```
PTM=read.table("H3PTM.txt",header=T)
#PCA analysis
pca.ptm<-princomp(PTM[,-1], scale=TRUE, center=TRUE, cor=F)
summary(pca.ptm)
loadings(pca.ptm)

#Print cummulative proportion of variance explained
cumprop=cumsum(pca.ptm$sdev^2)/sum(pca.ptm$sdev^2)
print(cumprop,digits=2)

#Screeplot
plot(pca.ptm, main="Screeplot of histone PTM data")

# Data projected on the first two PCs
plot(pca.ptm$scores[,1:2],type="n",xlab="1st PC",ylab="2nd PC")
text(pca.ptm$scores[,1:2],rownames(PTM))
title("Projection of PTM data to the subspace spanned by the first 2 PCs")

#Biplot of histone PTM data
biplot(pca.ptm)
title("Biplot of histone PTM data")
```

2.7 References

[1] Kouzarides, T., Chromatin modifications and their function. *Cell* 2007, 128, 693–705.

[2] Campos, E.I., Reinberg, D., Histones: annotating chromatin. *Annual review of genetics* 2009, 43, 559–99.

[3] Allis, C.D., Jenuwein, T., Reinberg, D., Epigenetics. *Epigenetics* 2007, 2, 23–62.

[4] Jenuwein, T., Allis, C.D., Translating the Histone Code. *Science* 2001, 293, 1074–1080.

[5] Strahl, B.D., Allis, C.D., The language of covalent histone modifications. *Nature* 2000, 403, 41–45.

[6] Klose, R.J., Zhang, Y., Regulation of histone methylation by demethylimination and demethylation. *Nature Reviews Molecular Cell Biology* 2007, 8, 307–18.

[7] Zhang, Y., Reinberg, D., Transcription regulation by histone methylation interplay between different covalent modifications of the core histone tails. *Genes & Development* 2001, 2343–2360.

[8] Chi, P., Allis, C.D., Wang, G.G., Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nature Reviews Cancer* 2010, 10, 457–69.

[9] Bannister, A.J., Kouzarides, T., Reversing histone methylation. *Nature* 2005, 436, 1103–1106.

[10] Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., et al., High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* 2007, 129, 823–837.

[11] Loyola, A., Tagami, H., Bonaldi, T., Roche, D., et al., The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO reports* 2009, 10, 769–75.

[12] Beck, D.B., Oda, H., Shen, S.S., Reinberg, D., PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes & Development* 2012, 26, 325–337.

[13] Schotta, G., Lachner, M., Sarma, K., Ebert, A., et al., A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & Development* 2004, 18, 1251–1262.

[14] Zee, B.M., Levin, R.S., Xu, B., LeRoy, G., et al., In vivo residue-specific histone methylation dynamics. *The Journal of biological chemistry* 2010, 285, 3341–50.

[15] Albert, M., Helin, K., Histone methyltransferases in cancer. *Seminars in cell & developmental biology* 2010, 21, 209–20.

[16] Zhang, K., Dent, S.Y.R., Histone Modifying Enzymes and Cancer Going beyond Histones. *Journal of Cellular Biochemistry* 2005, 1148, 1137–1148.

[17] Raynaud, C., Sozzani, R., Glab, N., Domenichini, S., et al., Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in Arabidopsis. *The Plant journal* 2006, 47, 395–407.

[18] Sanchez, R., Zhou, M.-M., The PHD finger: a versatile epigenome reader. *Trends in biochemical sciences* 2011, 36, 364–72.

[19] Li, H., Ilin, S., Wang, W., Duncan, E.M., et al., Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 2006, 442, 91–5.

[20] Peña, P. V, Davrazou, F., Shi, X., Walter, K.L., et al., Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* 2006, 442, 100–3.

[21] Shi, X., Hong, T., Walter, K.L., Ewalt, M., et al., ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* 2006, 442, 96–9.

[22] Maga, G., Hubscher, U., Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *Journal of cell science* 2003, 116, 3051–60.

[23] Jacob, Y., Feng, S., LeBlanc, C. a, Bernatavichute, Y. V, et al., ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nature structural & molecular biology* 2009, 16, 763–8.

[24] Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., et al., RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. *Genes & Development* 2007, 21, 1530–1545.

[25] Chung, P.-H., Yao, M.-C., Tetrahymena JMJD3 homolog regulates H3K27 methylation and nuclear differentiation. *Eukaryotic cell* 2012, Epub ahead of print.

[26] Zhang, K., Tang, H., Analysis of core histones by liquid chromatography-mass spectrometry and peptide mapping. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* 2003, 783, 173–9.

[27] Britton, L.-M.P., Gonzales-Cope, M., Zee, B.M., Garcia, B.A., Breaking the histone code with quantitative mass spectrometry. *Expert review of proteomics* 2011, 8, 631–43.

[28] Garcia, B.A., Shabanowitz, J., Hunt, D.F., Characterization of histones and their post-translational modifications by mass spectrometry. *Current opinion in chemical biology* 2007, 11, 66–73.

[29] Ueberheide, B., Mollah, S., Deciphering the histone code using mass spectrometry. *International Journal of Mass Spectrometry* 2007, 259, 46–56.

[30] Su, X., Ren, C., Freitas, M.A., Mass spectrometry-based strategies for characterization of histones and their post-translational modifications. *Expert Rev Proteomics* 2007, 4, 211–225.

[31] Beck, H.C., Nielsen, E.C., Matthiesen, R., Jensen, L.H., et al., Quantitative proteomic analysis of post-translational modifications of human histones. *Molecular & cellular proteomics* 2006, 5, 1314–25.

[32] Plazas-Mayorca, M.D., Bloom, J.S., Zeissler, U., Leroy, G., et al., Quantitative proteomics reveals direct and indirect alterations in the histone code following methyltransferase knockdown. *Molecular bioSystems* 2010, 6, 1719–29.

[33] Jung, H.R., Pasini, D., Helin, K., Jensen, O.N., Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Molecular & cellular proteomics* 2010, 9, 838–50.

[34] Ong, S.-E., Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nature chemical biology* 2005, 1, 252–62.

[35] Garcia, B. a, Joshi, S., Thomas, C.E., Chitta, R.K., et al., Comprehensive phosphoprotein analysis of linker histone H1 from Tetrahymena thermophila. *Molecular & cellular proteomics* 2006, 5, 1593–609.

[36] Syka, J.E.P., Marto, J.A., Bai, D.L., Horning, S., et al., Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-translational Modifications research articles. *Journal of Proteome Research* 2004, 3, 621–626.

[37] Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S. a, et al., Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature protocols* 2007, 2, 933–8.

[38] Ishihama, Y., Sato, T., Tabata, T., Miyamoto, N., et al., Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. *Nature biotechnology* 2005, 23, 617–21.

[39] Wu, C.C., MacCoss, M.J., Howell, K.E., Matthews, D.E., Yates III, J.R., Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Analytical chemistry* 2004, 76, 4951–9.

[40] Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R., Mann, M., Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature methods* 2010, 7, 383–5.

[41] Loidl, J., Mochizuki, K., Tetrahymena Meiotic Nuclear Reorganization Is Induced by a Checkpoint Kinase-dependent Response to DNA Damage. *Molecular Biology of the Cell* 2009, 20, 2428–2437.

[42] Cassidy-Hanley, D., Bowen, J., Lee, J.H., Cole, E., et al., Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment. *Genetics* 1997, 146, 135–47.

[43] Gorovsky, M.A., Studies on nuclear structure and function in Tetrahymena pyriformis. II. Isolation of macro- and micronuclei. *The Journal of Cell Biology* 1970, 47, 619–630.

[44] Gorovsky, M.A., Studies on nuclear structure and function in Tetrahymena pyriformis. 3. Comparison of the histones of macro- and micronuclei by quantitative polyacrylamide gel electrophoresis. *The Journal of cell biology* 1970, 47, 631–6.

[45] Shechter, D., Dormann, H.L., Allis, C.D., Hake, S.B., Extraction, purification and analysis of histones. *Nature protocols* 2007, 2, 1445–57.

[46] Cui, B., Liu, Y., Gorovsky, M.A., Deposition and function of histone H3 variants in Tetrahymena thermophila. *Molecular and cellular biology* 2006, 26, 7719–30.

[47] Yu, L., Gorovsky, M.A., Constitutive expression, not a particular primary sequence, is the important feature of the H3 replacement variant hv2 in Tetrahymena thermophila. *Molecular and cellular biology* 1997, 17, 6303–10.

[48] Strahl, B.D., Ohba, R., Cook, R.G., Allis, C.D., Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proceedings of the National Academy of Sciences of the United States of America* 1999, 96, 14967–72.

[49] Beynon, R.J., Pratt, J.M., Metabolic labeling of proteins for proteomics. *Molecular & cellular proteomics* 2005, 4, 857–72.

[50] Jacob, Y., Stroud, H., Leblanc, C., Feng, S., et al., Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* 2010, 466, 987–991.

[51] Xu, M., Wang, W., Chen, S., Zhu, B., A model for mitotic inheritance of histone lysine methylation. *EMBO reports* 2011, 13, 60–7.

CHAPTER 3

Histone-wide investigation of chromatin crosstalk in H3K27-specific methyl-transferases knockout cells by quantitative proteomics

3.1 Summary

Histone methylation plays an important role in eukaryotic transcriptional regulation. The specific methylation at histone H3K27 is an important, evolutionarily conserved epigenetic mark generally linked to transcriptional repression. Here, we extend the success of a previously developed, MS-based quantitative proteomics method using ^{15}N metabolically labeled histones as the internal reference to map changes in global histone post-translational modifications (PTM) and search for potential cross-talk between H3K27 methylation and other PTMs across all major histones in three histone methyltransferase (HMT) knockout cells. We identified 66 PTMs and quantified 49 of them in *Tetrahymena* core histones. The most noticeable observation is the hyper-acetylation of histones H2A and H4 at their N-terminal domains in response to decreased methylation of H3K27. Decreased phosphorylation in the H2B C-terminal region is also observed in all HMT mutants. The results from quantitative analysis of histone PTMs in WT and HMT mutants suggest that the methylation of H3K27 has multiple pathways and the mechanisms through which heterochromatin is maintained are also disscussed in this chapter.

3.2 Introduction

In eukaryotic cells, the DNA duplex wraps around an octamer formed by two copies each of the core histones, H2A, H2B, H3, and H4, to form the nucleosome, the basic unit of eukaryotic chromatin. Histones play both structural and complex functional roles in chromosomes. The dynamic chromatin architecture, switching between compact and loose structures as well as interactions with functional complexes, is dictated by numerous histone covalent modifications deposited by histone modifying enzymes in a combinatorial way, also referred to as the 'histone code' [1,2] which is still being deciphered. The most common post-translational modifications (PTM) found in histones are: methylation, acetylation, phosphorylation, biotinylation, citrullination, ADP-ribosylation, and ubiquitination[3] which occur at numerous, specific sites within the histones. Reflecting its crucial role in central cell processes, the functionality of chromatin is relatively robust with relatively few single histone and histone modifying gene deletions being lethal[4]. Compensatory changes in chromatin composition and histone modifications occur in response to these mutations but details of these compensatory changes and their impacts on the health of organisms are incomplete[5].

Catalyzed by histone methyltransferases (HMT) and reversed by histone demethylases (HDMT)[6–9], histone methylation is an important dynamic epigenetic mark primarily observed on lysyl and arginyl residues. In histones, lysines may be mono-, di- or tri-methylated, whereas the arginines may be mono-methylated or dimethylated (symmetrically or asymmetrically)[9,10]. The first histone methyltransferase, SUV39, was identified in mammals[11] and histone methylation has been the target of considerable research efforts in recent decades reflecting its key role in modulating

histone function. Unlike acetylation, methylation does not directly modulate the interaction between histones and DNA by altering the net charge of histones, and their biological impact on gene transcription can be either repressive or active depending on the exact site of modification and its context[7,12].

Methylation of histone H3K27, the main target of this study, is evolutionarily conserved and generally associated with transcriptional repression[12,13]. Several methyl transferases have been implicated in H3K27 methylation and the control mechanism is complex[13–17]. Enhancer of zeste (E(Z)), one of these HMTs which converts Lys-9 and Lys-27 of histone H3 to their di-/trimethylation states, was first identified in *Drosophila*[17]. Its homologs known as MES-2 in *C. elegans* or EZH1/2 in human were subsequently identified and mutations or dysregulation of EZH2 is associated with cancers[12]. Recently, two functionally redundant proteins, ATXR5 and ATXR6, specific for monomethylation of H3K27 were identified in *Arabidopsis* and their biological roles have been defined in gene silencing and DNA replication[15,18].

Ezl1p, Ezl2p and Ezl3p are three E(Z)-like, H3K27-specific HMTs in *Tetrahymena* but only EZL2 is significantly expressed in vegetatively growing cells[19,20]. *Tetrahymena* is an attractive model organism because it is amenable to genetic manipulation and occupies an interesting phylogenetic position relative to other animals and plants. Txr1p is another H3K27-specific HMT recently identified in this species and has an evolutionary relationship with ATXR5 and ATXR6[21]. Our previous study showed that Txr1p and Ezl2p are the two major HMTs jointly regulating the methylation states of histone H3K27 providing an opportunity to examine the relationships between H3K27 methylation and other histone modifications. Txr1p preferentially

monomethylates H3K27 while Ezl2p is mainly responsible for di-/trimethylation of H3K27[21].

Although methylation of H3K27 has been relatively well studied, its relationship with other histone PTMs and their detailed pathways still remain unknown. In this study, we use a previously developed, MS-based quantitative proteomics method[21] to characterize the global histone PTM levels in WT and three HMT knockout cells to identify the likely PTM crosstalk for H3K27 methylation states across the major histones. We also propose two different pathways behind the different HMTs that control the formation of the heterochromatin.

3.3 Experimental procedures

3.3.1 Construction of HMT knockout strain

Three HMT mutants, ΔTXR1, ΔELZ2 and ΔTXR1:EZL2, were generated from the wild-type *Tetrahymena* cells according to the method we published previously[21]. Briefly, genomic sequences encoding TXR1 or ELZ2 were PCR amplified and fused with the neo4 cassette with paromomycin resistance. Standard biolistic transformations were adopted[22] and transformants were selected based on paromomycin resistance. The final replacements were verified by quantitative PCR.

3.3.2 Cell culture, core histone preparation, and HPLC purification

Media, procedures, and protocols used for cell culture, nuclear preparation, acid histone extraction, and histone purification were performed as previously[21]. The general experimental design is illustrated in Figure S3.1. Briefly, *Tetrahymena thermophila* wild-type, *TXR1*, *ELZ2*, and *TXR1:EZL2* knockout cells were grown in 1× SPP medium at 30°C with gentle shaking. Cell growth stopped at logarithmic-phase

(2×10^5/ml) and cells were collected for subsequent experiments; To prepare the [^{15}N] histones as internal standards, wild-type *Tetrahymena* cells were metabolically labeled by feeding on *Escherichia coli* BL21 cells grown in the [^{15}N] M9 minimal medium supplemented with ^{15}N-substituted Bioexpress (Cambridge Isotope Laboratories); To isolate the macronuclei from *Tetrahymena* cells, cells were collected by centrifugation, resuspended in 200 ml of medium A (0.1 M sucrose, 2 mM $MgCl_2$, 4% gum arabic, 10 mM Tris, 5mM EDTA, 10 mM butyric acid, 1mM iodoacetamide, and 1mM PMSF with a pH of 6.5), and disrupted by vigorous blending in the presence of 1-octanol (~0.7 ml); To prepare the crude histones, macronuclei were lysed in 1 ml of 0.4 N sulfuric acid after they were separated by differential centrifugation. The acid-extracted histone samples were then precipitated by 20% TCA, washed with acidified acetone (0.2% HCl), air-dried and dissolved in 500 µl of water; To further purify individual histones, the crude histone samples were fully resolved on a C8 reversed-phase HPLC column by applying a 60 min gradient with organic solvent (90% Acetonitrile in 0.1% TFA). HPLC fractions containing individual histones were combined after evaluation by 15% SDS-PAGE. Concentrations of the purified core histones from WT, [^{15}N] WT, and all mutants were determined by the Bradford method.

3.3.3 NanoLC-MS analysis of core histone PTMs in WT and HMT knockout cells

Histone samples were analyzed in biological duplicates(ΔTXR1:EZL2) or triplicates(WT, ΔTXR1, ΔELZ2). Chemical derivatization, trypsinization, and nanoLC-MS analysis of histone samples were performed as previously described[21,23]. Briefly, 3 µg of each core histone from WT, ΔTXR1, ΔELZ2, or ΔTXR1:EZL2 were equally mixed with the internal standards, the [^{15}N] core histones prepared from WT cells as

described above. Prior to mass spectrometry analysis, samples were chemically derivatized twice with propionic anhydride before and after trypsin digestion. Histones were in-solution digested with sequencing-grade trypsin and the fully propionylated peptide mixtures were vacuum-dried and dissolved in 0.1% formic acid before the LC-MS analysis. The peptide mixtures were separated with a C18 analytical column (3μm, 300Å, 150mm x 100μm; CVC Technologies), eluted, and introduced into a Thermo Fisher Orbitrap XL mass spectrometer with a 15μm PicoTip emitter (New Objective) at 200 nl/min flow rate. Mass spectrometer was connected with a manufacturer's nano-ESI source operated in both positive ion mode and data dependent mode at a resolution of 30,000 on MS1. The 8 strongest peaks were CID fragmented with 30% normalized collision energy. To achieve the best LC-MS/MS performance, a 2-minute dynamic exclusion and internal mass lock were also excuted.

Raw data processing, database searching, and peptide quantification were all done in Mascot distiller (Matrix Sciences, Version 2.4 for distiller and Version 2.2.07 for search engine). The NCBI *Tetrahymena* database was used in performing the database searching. Precursor ion tolerance was 10ppm and 0.8 Da for the fragment ions. N-terminal propionylation was set as fixed modification and the following PTMs were considered as variable modifications: Acetylation (Protein N-terminus, K), Methylation (Protein N-terminus, KR), Propionylation (K, unmodified or monomethylated), Phosphorylation (STY), Formylation (K), Citrullination (R), and Ubiquitination (K). Up to five missed cleavages were allowed for trypsin digestion as the propionylation and PTMs on lysyl residues blocks trypsin digestion. Several peptides from each of the core histones were selected for normalization by their averages: GKTASSKQVSR, GQASQDL, FLKHGR

from H2A; LLLPGELAR, IALESSKLVR, EVQTAVKLLLPGELAR from H2B; FRPGTVALR, VTIMTKDMQLAR, YQKSTDLLIR from H3; and SFLENVVR, QVLKSFLENVVR, TLYGFGG from H4. We assumed that the levels of those peptides were unchanged across different epigenetic features as their ratios (L/H) were close to 1 and quite consistent (Coefficient of Variation, CV<15%) in all experiments. Student t-test was done in Microsoft Excel and heatmaps were generated from R software. All spectra assigned with PTMs were manually validated based on the same criteria we published previously[21].

3.3.4 Analysis of N-terminal acetylation in histone H2A and H4 by acetic acid-urea polyacrylamide gel electrophoresis

To analyze the N-terminal acetylated isoforms of H2A and H4, intact proteins were separated by continous acetic acid-urea (AU) gel system as described previously[24]. Briefly, a resolving gel (approximately 18 cm) was made with 6M urea, 15% of acrylamide (acrylamide/bis-acrylamide ratio 37.5:1), and 10% acetic acid (all reagents were purchased from Sigma, St. Louis, MO). A 5% stacking gel was poured onto the top of the resolving gel after polymerization. To improve the analytical performance, the AU gel was pre-electrophoresed with constant voltage at 300 V overnight using 5% acetic acid as running buffer. Histone samples (H2A and H4) prepared from WT and mutants were dissolved in 10 μL of loading buffer containing 2 M urea, 2% β-mercaptoethanol, and 5% acetic acid before they were loaded onto the gel. A constant voltage at 100 V was first applied for 30 min to introduce the sample into the stacking gel and then the voltage was raised up to 400 V for 10 h at 5 °C. The gel was stained using the blue silver staining method for 30 min and visualized by destaining in deionized water for several times[25].

3.4 Results and discussion

While the method used in this study does not identify all possible post-translational modifications, it is quite extensive and in this study identified 66 PTMs in *Tetrahymena* core histones and quantified 49 of them with relatively small variations (CV<15%) after data normalization. The number of PTMs found in each histone were: 10 in H2A variants, 13 in H2B variants, 32 in H3 variants, and 11 in H4. The most common PTMs identified were acetylation and methylation while histone H3 had the most complex modifications. The detailed information is provided in Figure 3.1 and Table 3.1. At the Mascot identity threshold, the False Discovery Rate (FDR) is close to 1% on average as estimated from the decoy database. All detected tryptic peptides cover the sequences of H2A, H2B, H3, and H4 at 61%, 68%, 76%, and 97%, respectively (Figure 3.1). Sequence variants were identified for all histones except H4. They are: 2 H2A variants (H2A-1, H2A.X), 2 H2B variants (H2B-1, H2B-2), and 3 H3 variants (H3, H3.3, and H3.4).

3.4.1 Relative quantification of H3K27 methylation in 3 HMT mutants

We analyzed and compared all peptides associated with PTMs and their unmodified counterparts in WT and mutants of 3 HMTs that target H3K27. The methylation of H3K27 was significantly affected in all mutants as expected (see Figure 3.2, Table 3.2-H3). The monomethylation of H3K27 and H3K36, the major targets of Txr1p, were reduced at least 80% while the unmethylated form of the peptide increased significantly. The dimethylated form was also significantly reduced to 50% while its trimethylation state was only slightly affected. The knockout of Ezl2p primarily affected the di- and trimethylation of H3K27 with at least 80% decline in each state but monomethylation of H3K27 increased in this mutant as described previously[21], consistent with it being a

major precoursor to the di- and trimethylated forms. The double knockout of Txr1p and Ezl2p (ΔTXR1:EZL2), however, significantly down-regulated. All three methylation states for H3K27 and a correspondingly strong increase in the unmethylated form was observed (4.46X). Significant downregulation was also observed for H3K36Me1 for ΔTXR1 and for the double knockout. In contrast to the decreased H3K27 methylation, mild off-site hyper-monomethylation in H3 was observed at K4, K23, and particularly K56 in ΔTXR1 (Table 3.2, H3). Hyper-monomethylation of lysyl residues on histones other than H3 was less apparent at $p<0.05$ (Table 3.2, H2A, H2B, H4). Mild hypermethylation of arginyl residues was also observed at H3R83 and potentially at H3R128 ($p>=0.05$). It is possible that the roles of hypermethylation at these sites (K4, K23, K56, R83) are part of a compensatory mechanism to the hypomethylation at K27 in order to maintain the general state of chromatin function.

In WT actively growing cells, the major PTMs of H3 are H3K27Me1 and H3K27Me2, which constitute more than 55% of all modified states as indicated by the extracted ion chromatogram (XIC) analysis of the peptide (27-KSAPATGGIKKPHR-40) from the H3 major form (Table S3.1). However, all 3 methylation states were not completely eliminated even in the double knockout cells, suggesting the presence of other HMTs capable of H3K27 methylation, although less effectively, (probably Ezl1p and Ezl3p) in this species.

3.4.2 Hyper-acetylation is a ubiquitous, histone-wide response to undermethylation at H3K27

Controlled by histone acetyltransferases (HATs) and histone deacetylases (HDACs), histone acetylation is generally linked to transcriptional activation and associated

with euchromatin because acetylation neutralizes the positively charged histones, which in turn, attenuates electrostatic interactions with the negatively charged DNA backbone and thus contributes to a more open chromatin structure for gene transcription mediated by transcriptional binding factors and their accessory proteins[26,27].

As methylation of H3K27 is a repressive epigenetic mark, elimination of its modifying enzymes leads to on-site hypo-methylation, forming a decondensed chromatin structure, and eventually promoting DNA replication and activation of gene transcription in the cell. Elevated acetylation was extensively observed on many sites of all major core histones in all three K27-specific HMT mutants (Table 3.2). The hyper-acetylation of H2A and H4 in their N-terminal domains, in particular, is the most remarkable feature in the methylation-deficiency mutants (Figures 3.3 and 3.4). As inferred from the quantitative proteomics results, the N-terminal domains of H2A and H4 exhibit mono-, di-, tri-, and tetra-acetylated forms and their various permutations make the peptides quite complicated (Figures 3.3a and 3.4a). Interestingly, the levels of acetylation from low degree to high degree were gradually increased in all mutants relative to WT. The fully, tetra-acetylated peptide from the Txr1 knockout has a more than 2.5-fold increase in H4 and at least 3-fold in H2A (Figure 3.3b and 3.4b). This trend can also be clearly seen from the acetic acid-urea gel electrophoresis (Figure 3.5). All results suggested that the N-terminal tails of histone H2A and H4 are exhibit increased acetylation in all mutants and their fully acetylated forms may contribute transcriptionally active chromatin. In addition to the hyper-acetylation of H2A and H4 on their N-terminal domains, K9Ac, K14Ac, K18Ac, K23Ac, K27Ac, and K56Ac were also mildly elevated in H3.

N-terminal acetylation of H4 is relatively conserved as compared with H2A. Previous studies have elucidated its role in regulating chromatin structure and function[28,29]. The tetra-acetylated form of H4 is highly correlated with transcription[28] and the most highly acetylated forms have the highest affinity for transcriptional factors[29]. Relative to acetylation at K5, K8, and K12, the acetylation of H4K16 is especially important and has major effects on chromatin structure in human cells[27]. Acetylation of histone H4 at K16 can prevent the formation of the inactive 30-nm coil in Hela cells[30]. Generally, our results of H4 N-terminal acetylation agreed well with these studies but provided a more detailed picture of the combinatorial response of these PTMs.

N-terminal acetylation of H2A, however, is less well studied and its biological roles are still not clearly understood. In *tetrahymena*, there are at least 4 H2A variants: H2A.1, H2A.X, H2AZ.hv1, and H2A.Y. This study only focused on H2A.1 and H2A.X as they are the most abundant isoforms in this species. Because the N-terminal sequences of H2A are more variable in different species, their modification patterns are less conserved and less easily compared among species. Yeast H2A is acetylated at K4 and K7 which is equivalent to acetylated K5 and K9 in human H2A[31,32]. The K7 acetylation in yeast, however, is present at a higher level and plays a more important role in regulating transcription[31]. Our study showed that, like histone H4, the distribution of mono-, di-, tri-, and tetra-acetylated forms (Figure 3.4) in the N-terminal domain of H2A are increased in the mutants. The patterns of acetylation were also very similar: the most highly acetylated forms of H2A were present at higher levels than the less modified forms and the fully acetylated peptide increased to the greatest degree. The results

suggested a similar, but previously unrecognized mechanism of modulating acetylation in both H2A and H4 in regulating chromatin structure and function.

Intriguingly, we found that H2A from *Tetrahymena* is also phosphorylated in its N-terminal domain. Although we cannot identify the specific residue that is phosphorylated due to incomplete ion fragmentation, S1 phosphorylation has been observed human H2A and its role in inhibiting transcription has been shown[33]. The opposing effect of this modification (relative to acetylation), on the other hand, may provide another compensatory or regulatory mechanism distinct from N-terminal domain acetylations toward regulating chromatin structure and function.

3.4.3 Other potential cross-talk related with H3K27 methylation

To date, all histones, including linker histone H1, have been found to be phosphorylated across species and their site-specific roles being involved in gene activation, chromosome condensation, DNA repair, apoptosis, and many other cellular functions[34,35]. In this study, we identified two *tetrahymena* phosphorylation sites and quantified one of them in H2A and H2B without further enrichment of phosphorylated peptides or proteins. Along with the hyper-acetylations in core histones, hypo-phosphorylation in the H2B C-terminal domain is another potential compensatory PTM associated with H3K27 methylation (Figure 3.6, Table 3.2-H2B). As with the H2A phosphorylation site, the H2B phosphorylation site cannot be precisely localized in a specific residue due to the lack of complete fragment ions. This modification was down-regulated by almost 80% in both H2B variants (H2B-1 and H2B-2) whose sequences only differ by one amino acid. Human H2B is phosphorylated at three sites (S113, T116, T120) near C-terminal domain but their detailed biological functions of this modification

are unclear[36]. Although we don't know whether these two types of modifications has a direct link with each other, previous studies may provide a clue between phosphorylation and acetylation: Proteome-wide investigation of protein phosphorylation and acetylation confirmed that they can reciprocally affect one another[37]; Phosphorylation at H3S10 promotes H4K16 acetylation and mediates transcriptional elongation[38].

We also identified several formylation sites in H2A and H3 and some of them (H2AK39 and H3K23) were slightly elevated in some of our mutants (Table 3.2). N-formylation of lysine is less well studied but is a relatively common PTM in histones, high mobility group proteins, and other non-histone proteins[39,40]. All histones, especially the linker histone H1, are formylated at multiple sites. Histone formylation plays a role in response to DNA damage [40] but their other biological functions remain unknown. Similar to acetylation in structure and chemical properties, formylation may share some common impacts with acetylation on chromatin structure and function.

3.4.4 Regulation of transcription by H3K27 methylation

The regulation of H3K27 methylation has been intensively studied and two major conserved pathways were defined in mammals and other systems[13,14,16,41,42]. One is E(Z)/EZH2-mediated pathway by controlling the di- and trimethylation levels of H3K27. Polycomb group (PcG) proteins are a key downstream effector sensing methylated histone H3 at K27 and thus keep chromatin at 'off-state'[14]. E(Z)-mediated methylation is essential for Hox gene silencing, Oct3/4 gene silencing and X-inactivation [14]. Independent of E(Z)/EZH2-mediated pathway, ATXR5/6 are two functionally redundant H3K27 mono-methyltransferases in plant which involves a previously unrecognized

pathway required for gene silencing[18,42]. Monomethylation of H3K27 built up by ATXR5 and ATXR6 regulates heterochromatic DNA replication in *Arabidopsis*[42].

The similar HMTs and homolog pathways related with H3K27 methylation were also identified in evoluationally distant species, *Tetrahymena thermophila*[21,41,43]. As disscussed in our previous publication[21], H3K27 methylation was differentailly regulated by two major HMTs (TXR1p and ELZ2p). TXR1p preferentially monomethylates H3K27 while the di- and trimethylation states were primarily maintained by ELZ2p. Principal Component Analysis (PCA) of all PTMs (Table S3.2, Data was log2 transformed in R) identified in this study clearly separated the monomethylation of H3K27 from its di- and trimethylation states(Figure 3.7). The monomethylated H3K27 and K36 were closely regulated by TXR1p and the di- and trimethylated states were co-regulated by ELZ2p. Tetra-acetylations of H2A and H4 were negatviely correlated with H3K27 methylaton while H2B phosphorylation at its C-terminus is positively correlated. The bioplot suggested TXR1p was less coupled with ELZ2p but had a closer relationship with the double knockout of *TXR1* and *ELZ2* (Figure 3.7). In *tetrahymena*, ELZ2p is a homolog of the canonical H3K27-specific HMT Enhancer of Zeste which is required for gene silencing via transcription. The H3K27 monomethyltransferase TXR1p, the homolog of ATXR5/6, probably maintains heterochromatin via DNA replication in this speicies (Unpublished data). The *TXR1*-mediated gene silencing pathway may be have a dominant effect over *ELZ2*-mediated pathway as suggested from bioplot analysis of three variables (Figure 3.7). The results combined with other studies in other systems delivered an important message that the

same type of PTM at same position may have a totally, different biological impacts on chromatin structure and functions.

3.5 Concluding remarks and future directions

The methylation of H3K27 is one of the most intensively studied PTMs in the epigenetics field due to its demonstrated biological importance in transcription regulation. The roles of H3K27 methylation have been elucidated in the regulation of large none-coding RNA expression[44], Polycomb-Group(PcG)-mediated gene silencing[14], heterochromatin formation[19], and heterochromatic DNA replication[15]. The balance of H3K27 methylation relative to other PTMs must be precisely maintained and deregulation of this modification may lead to cancers[5,13]. We demonstrated crosstalk between methylation and different PTMs on the same histone and across histones. All crosstalk related with methylation of H3K27 are illustrated in Figure 3.8. Aspects of these changes in histone PTM balance may reduce the robustness of chromosome functions and make the cell more susceptible to dysfunctions, including cancer in multicellular organisms. Our study first revealed the potential pathways involved in the methylation of H3K27 and the possible mechanism through which transcription is regulated by changing H3K27 methylation levels via enzymatic pathways. However, detailed information is still required in elucidating the whole biological network in future. This work also represents the most extensive investigation of core histone PTMs and H3K27-related cross-talk.

Table 3.1 All PTMs identified in core histones

H2A.X

Start	End	Sequence	Modifications	z	Observed	ΔM*	Ion Score
1	10	STTGKGGKAK.G	Ac (K5); Pr (N-term,K8,10)	2	572.8141	0	23
1	21	STTGKGGKAKGKTASSKQVSR.S	Ac (K5,K8,K10,K12); Pr (N-term,K17)	2	1172.628	1	54
				3	782.0877	1	41
1	21	STTGKGGKAKGKTASSKQVSR.S	Ac (K5,K8,K10); Pr (N-term,K12,K17)	3	786.7592	0	34
1	21	STTGKGGKAKGKTASSKQVSR.S	Ac (K5,K8,K12); Pr (N-term,K10,K17)	2	1179.6361	1	69
1	21	STTGKGGKAKGKTASSKQVSR.S	Ac (K5,K8); Pr (N-term,K10,K12,K17)	2	1186.6435	1	101
				3	791.4314	1	66
1	21	STTGKGGKAKGKTASSKQVSR.S	Ac (K5); Pr (N-term,K8,K10,K12,K17)	2	1193.6511	1	117
				3	796.1031	0	73
1	21	STTGKGGKAKGKTASSKQVSR.S	Phos (S1/T2/T3); Pr (K5,K8,K10,K12,K17)	2	1212.6277	1	43
6	21	K.GGKAKGKTASSKQVSR.S	Ac (K10,K12); Pr (N-term,K17)	2	893.4959	2	72
6	21	K.GGKAKGKTASSKQVSR.S	Ac (K10); Pr (N-term,K12,K17)	2	900.504	2	79
6	21	K.GGKAKGKTASSKQVSR.S	Ac (K12); Pr (N-term,K10,K17)	2	912.4688	2	58
9	21	K.AKGKTASSKQVSR.S	Ac (K12); Pr (N-term,K10,K17)	2	779.434	1	53
25	33	R.AGLQFPVGR.I	Me1(R33); Pr (N-term)	1	1027.5372	3	30
				2	514.2709	0	30
34	42	R.ISRFLKHGR.Y	Me2 (R36); Pr (N-term,K39)	2	627.3721	9	43
37	42	R.FLKHGR.Y	Me1 (R42); Pr (N-term,K39)	2	448.2434	0	39
37	42	R.FLKHGR.Y	Form (K39); Pr (N-term)	2	421.2374	0	36
86	92	R.HILLAIR.N	Me1 (R92); Pr (N-term)	1	917.5589	2	53

H2B-1

Start	End	Sequence	Modifications	z	Observed	ΔM*	Ion Score
1	12	APKKAPAAAAEK	Me1(A1,K3); Pr(K3, K4,K12)	2	682.373	0	52
1	12	APKKAPAAAAEK	Me2(K3);Pr(N-term,K4)	2	646.8825	0	34
1	12	APKKAPAAAAEK	Me3(K3); Pr(N-term,K4)	2	653.8902	0	56
1	13	APKKAPAAAAEKK	Me1(A1);Pr(K3,K4,K12)	2	731.9352	0	47
1	13	APKKAPAAAAEKK	Me3(K3);Ac(K4);Pr(N-term,K12)	2	738.9436	1	53
1	13	APKKAPAAAAEKK	Me3(K3);Pr(N-term,K3,K12)	2	745.951	1	52

1	15	APKKAPAAAAEKKVK	Me1(A1);Pr(K3,K4,K12,K13)	2	873.531	1	48
1	15	APKKAPAAAAEKKVK	Me1(A1,K3);Pr(K3,K4,K12,K13)	2	880.5388	1	56
1	15	APKKAPAAAAEKKVK	Me3(K3);Pr(N-term,K4,K12,K13)	2	887.5464	1	52
				3	592.0325	0	54
1	15	APKKAPAAAAEKKVK	Me3(K3);Ac(K4);Pr(N-term,K12,K13)	2	918.5218	1	42
				3	594.008	0	46
104	121	HAISEGTKAVTKFSSSTN	Pr(N-term,K111);Ac(K115)	1	1010.0059	1	40
104	121	HAISEGTKAVTKFSSSTN	Pr(N-term,K111);Phos(S117-T120)	2	1028.9841	1	38

H2B-2

Start	End	Sequence	Modifications	z	Observed	ΔM*	Ion Score
1	12	APKKAPAATTEK	Me3(K3);Pr(N-term,K4)	2	691.3801	2	43
1	12	APKKAPAATTEK	Me3(K3);Pr(N-term,K4,K12)	2	719.3925	1	44
1	13	APKKAPAATTEKK	Me1(A1);Pr(K3,K4,K12)	2	761.9465	1	54
1	13	APKKAPAATTEKK	Me2(K3); Pr(N-term,K4,K12)	2	768.9542	1	49
1	13	APKKAPAATTEKK	Me3(K3); Pr(N-term,K4,K12)	2	784.4372	1	58
1	13	APKKAPAATTEKK	Me3(K3);Ac(K4);Pr(N-term,K12)	2	777.4291	1	51
1	13	APKKAPAATTEKK	Me3(K3);Pr(N-term,K4,K12)	2	775.9615	0	50
1	15	APKKAPAATTEKKVK	Me1(A1,K3);Pr(K3,K4,K12,K13)	3	607.3677	0	65
1	15	APKKAPAATTEKKVK	Me2(K3);Pr(N-term,K4,K12,K13)	2	910.5493	1	53
1	15	APKKAPAATTEKKVK	Me3(K3);Pr(N-term,K4,K12,K13)	2	917.5566	1	53
				3	618.687	1	82
1	15	APKKAPAATTEKKVK	Me3(K3);Ac(K4);Pr(N-term,K12,13)	2	920.5201	2	50
				3	614.015	0	58
1	16	APKKAPAATTEKKVKK	Me3(K3);Pr(N-term,K4,K12,K13,K15)	2	1020.5847	1	60
				3	673.4137	0	53
104	121	HAISEGTKAVTKFSSSSN	Pr(N-term,K111);Phos(S117-S120)	2	1033.4508	9	31

H3

Start	End	Sequence	Modifications	z	Observed	ΔM*	Ion Score
3	8	TKQTAR	Me1(K4); Pr(N-term)	2	387.7271	0	33
3	8	TKQTAR	Me1(K4); Pr(N-term,K4)	2	421.2236	-1	28
3	8	TKQTAR	Me2(K4); Pr(N-term)	2	400.2186	0	23
3	8	TKQTAR	Me3(K4); Pr(N-term)	2	401.743	1	19

9	17	KSTGAKAPR	Ac(K14); Pr(N-term)	2	507.2902	-1	35
9	17	KSTGAKAPR	Ac(K9,K14); Pr(N-term)	1	1055.5856	1	23
				2	528.2957	0	60
9	17	KSTGAKAPR	Ac(K9); Me1(K14); Pr(N-term,K14)	2	549.2891	-3	31
9	17	KSTGAKAPR	Ac(K9); Pr(N-term,K14)	2	535.3029	-1	61
9	17	KSTGAKAPR	Ac(K14); Pr(N-term,K9)	2	542.2835	1	67
18	26	KQLASKAAR	Ac(K23); Pr(N-term)	2	535.8189	-1	24
18	26	KQLASKAAR	Ac(K23); Pr(N-term,K18)	2	563.8318	-1	54
18	26	KQLASKAAR	Ac(K18,K23); Pr(N-term)	2	556.8272	4	51
18	26	KQLASKAAR	Me1(K23); Pr(N-term,K18,K23)	1	1169.6461	1	22
				2	585.3263	1	55
18	26	KQLASKAAR	Me2(K23); Pr(N-term,K18)	2	556.842	-2	35
18	26	KQLASKAAR	Me3(K23); Pr(N-term,K18)	2	563.8512	1	22
18	26	KQLASKAAR	Form(K23); Pr(N-term,K18)	2	556.8272	4	25
19	26	QLASKAAR	Ac(K23); Pr(N-term)	1	942.5385	2	30
27	40	KSAPATGGIKKPHR	Ac(K27); Me2(K36); Pr(N-term,K37)	2	815.4768	2	75
27	40	KSAPATGGIKKPHR	Ac(K27); Pr(N-term,K36,K37)	2	829.4732	0	71
27	40	KSAPATGGIKKPHR	Me1(K27); Pr(N-term,K36,K37)	2	815.4768	3	51
27	40	KSAPATGGIKKPHR	Me2(K27); Pr(N-term,K36,K37)	2	822.4842	1	88
				3	548.6579	0	62
27	40	KSAPATGGIKKPHR	Me3(K27); Pr(N-term,K36,K37)	2	829.4902	-1	38
				3	553.3297	0	44
27	40	KSAPATGGIKKPHR	Me1(K27); Me1(36); Pr(N-term,K27,K36,K37)	2	850.4978	2	88
27	40	KSAPATGGIKKPHR	Me1(K27); Pr(N-term,K27,K36,K37)	2	843.489	1	90
27	40	KSAPATGGIKKPHR	Me1(K27); Me1(R40); Pr(N-term,K27,K36,K37)	2	850.4975	1	99
27	40	KSAPATGGIKKPHR	Me1(R40); Pr(N-term,K27,K36,K37)	2	843.4823	-7	65
41	49	FRPGTVALR	Me1(R49); Pr(N-term)	2	543.8246	0	36
53	63	KYQKSTDLLIR	Me1(K56); Pr(N-term,K56)	2	754.4066	-1	47
73	83	DIAHEFKAELR	Me1(K79); Pr(N-term,K79)	2	727.8854	0	41
73	83	DIAHEFKAELR	Me1(R83); Pr(N-term,K79)	2	727.8854	0	49
116	128	RVTIMTKDMQLAR	Me1(R128); Pr(N-term,K122)	2	844.964	1	46
H3.3							

Start	End	Sequence	Modifications	z	Observed	ΔM*	Ion Score
9	17	KSTGVKAPR	Ac(K9,K14); Pr(N-term)	2	549.2891	-3	44
9	17	KSTGVKAPR	Ac(K9); Pr(N-term,K14)	2	556.2984	0	20
9	17	KSTGVKAPR	Form(K14); Pr(N-term)	2	514.2981	0	51
9	17	KSTGVKAPR	Form(K14); Pr(N-term,K9)	2	542.3081	-6	64
18	26	KQLATKAAR	Ac(K18,K23); Pr(N-term)	1	1126.6537	-4	20
				2	563.8318	-1	59
18	26	KQLATKAAR	Ac(K23); Pr(N-term,K18)	1	1140.6729	-1	44
				2	570.8395	-2	63
18	26	KQLATKAAR	Ac(K23); Pr(N-term)	2	550.3057	1	43
18	26	KQLATKAAR	Form(K23); Pr(N-term)	2	535.8189	-1	24
18	26	KQLATKAAR	Form(K23); Pr(N-term,K18)	1	1126.6537	-4	25
				2	563.8318	-1	45
19	26	QLATKAAR	Form(K23); Pr(N-term)	1	942.5381	2	32
53	63	KYQKTTDLLIR	Ac(K56); Pr(N-term,K53)	2	766.9373	-1	41
53	63	KYQKTTDLLIR	Ac(K56); Pr(N-term)	2	738.9251	1	56
73	83	DIAMEMKSDIR	Me1(K79); Pr(N-term,K79)	2	725.3299	0	26

H4

Start	End	Sequence	Modifications	z	Observed	ΔM*	Ion Score
1	7	AGGKGGK	Ac(K4); Pr(N-term, K7)	1	728.3949	2	35
1	11	AGGKGGKGMGK	Ac(K4,K7); Pr(N-term)	1	1087.5635	6	53
				2	544.2822	1	55
1	11	AGGKGGKGMGK	Ac(K4,K7); Pr(N-term,K11)	1	1143.5856	2	64
				2	572.297	4	49
1	11	AGGKGGKGMGK	Pr(N-term,K4); Ac(K7)	1	1101.5749	3	51
1	11	AGGKGGKGMGK	Pr(N-term,K4,K11); Ac(K7)	1	1157.6016	3	35
				2	579.3038	2	71
1	16	AGGKGGKGMGKVGAKR	Ac(K4,K7,K11,K15); Pr(N-term)	1	1682.9073	4	51
				2	841.9559	2	51
1	16	AGGKGGKGMGKVGAKR	Ac(K4,K7, K11); Pr(N-term,K15)	1	1719.8554	4	57
				2	848.9645	3	67
				3	566.3106	2	38

1	16	AGGKGGKGMGKVGAKR	Ac(K4,K7); Pr(N-term,K11,K15)	2	867.4379	3	61
				3	570.9829	1	53
1	16	AGGKGGKGMGKVGAKR	Ac(K4,K11); Pr(N-term,K7,K15)	2	855.9704	1	68
1	16	AGGKGGKGMGKVGAKR	Ac(K7,K11); Pr(N-term,K4,K15)	2	855.9723	3	62
1	16	AGGKGGKGMGKVGAKR	Ac(K4); Pr(N-term,K7,K11,K15)	2	862.9794	2	62
1	16	AGGKGGKGMGKVGAKR	Ac(K11); Pr(N-term,K4,K7,K15)	2	862.9801	3	59
				3	575.6545	1	51
1	16	AGGKGGKGMGKVGAKR	Ac(K7); Pr(N-term,K4,K11,K15)	2	862.9779	1	46
1	16	AGGKGGKGMGKVGAKR	Ac(K4); Pr(N-term,K7,K11,K15)	2	862.9794	2	62
1	16	AGGKGGKGMGKVGAKR	Ac(K11); Pr(N-term,K4,K7,K15)	2	862.9801	3	59
1	16	AGGKGGKGMGKVGAKR	Ub(K7);Pr(N-term,K4,K15)	2	870.9785	-2	41
5	11	GGKGMGK	Ac(K7); Pr(N-term)	1	732.3727	2	29
5	11	GGKGMGK	Ac(K7); Pr(N-term,K11)	1	788.3993	3	30
5	16	GGKGMGKVGAKR	Ac(K7,K11,K15); Pr(N-term)	1	1327.7186	3	66
				2	664.365	6	50
5	16	GGKGMGKVGAKR	Ac(K7,K11); Pr(N-term,K15)	1	1359.6822	4	53
				2	671.3707	3	59
5	16	GGKGMGKVGAKR	Ac(K7); Pr(N-term,K11,K15)	1	1355.7502	3	44
				2	678.3777	1	50
8	15	GMGKVGAK	Ac(K11); Pr(N-term)	1	845.4576	3	38
8	16	GMGKVGAKR	Ac(K11,K15); Pr(N-term)	1	1043.5687	2	30
8	16	GMGKVGAKR	Ac(K11); Pr(N-term,K15)	1	1057.5854	3	34
				2	529.295	1	48
8	16	GMGKVGAKR	Ac(K15); Pr(N-term,K11)	2	529.2958	2	41
24	35	ASIEGITKPAIR	Me1(R35); Pr(N-term,K31)	1	1397.761	3	63
24	36	ASIEGITKPAIRR	Me1(R35); Pr(N-term,K31)	2	769.4585	2	43
46	55	ISSFIYDDSR	Me1(R55); Pr(N-term)	1	1285.5776	4	52
				2	643.291	2	48
56	67	QVLKSFLENVVR	Me1(R67) Pr(N-term,K59)	1	1575.852	3	54
68	77	DAVTYTEHAR	Me1(R77); Pr(N-term)	2	616.7969	-3	43
78	92	RKTVTAMDVVYALKR	Me1(K91); Pr(N-term,K79,K91)	2	967.0534	3	76
78	92	RKTVTAMDVVYALKR	Me1(R92); Pr(N-term,K79,K91)	3	645.0368	1	34
79	92	KTVTAMDVVYALKR	Me1(R92) ;Pr(N-term,K79,K91)	2	889.0012	1	43

Table 3.2 PTMs quantified in all histones

H2A.X/H2A.1					WT(L/H)		EZL2(L/H)		TXR1(L/H)		E2+T1(L/H)		Final Ratio		
start	end	z	Sequence	PTM	mean	sd	mean	sd	mean	sd	mean	sd	EZL2/WT	TXR1/WT	E2+T1/WT
1	21	2	STTGKGGKAKGKTASSKQVSR	K5AcK8AcK10AcK12Ac	0.47	0.16	1.75	0.14	1.54	0.71	1.78	0.32	3.7**	3.26	3.77
1	21	3	STTGKGGKAKGKTASSKQVSR	K5AcK8AcK10AcK12Ac	0.5	0.16	1.76	0.12	1.57	0.72	1.87	0.45	3.54**	3.15	3.77
1	21	2	STTGKGGKAKGKTASSKQVSR	K5AcK8AcK12Ac	0.64	0.11	1.55	0.3	1.24	0.33	1.3	0.21	2.41*	1.94	2.02
1	21	3	STTGKGGKAKGKTASSKQVSR	K5AcK8AcK12Ac	0.67	0.11	1.6	0.26	1.35	0.36	1.33	0.22	2.4*	2.01	1.99
1	21	2	STTGKGGKAKGKTASSKQVSR	K5AcK8Ac	0.65	0.05	1.15	0.07	1.02	0.13	0.93	0.11	1.78**	1.58*	1.44
1	21	3	STTGKGGKAKGKTASSKQVSR	K5AcK8Ac	0.7	0.05	1.23	0.11	1.12	0.12	0.96	0.09	1.76**	1.6*	1.38
1	21	2	STTGKGGKAKGKTASSKQVSR	K5AcK10Ac	0.68	0.11	1.15	0.07	1.02	0.13	1.02	0.01	1.69**	1.5*	1.5*
1	21	3	STTGKGGKAKGKTASSKQVSR	K5AcK10Ac	0.71	0.05	1.24	0.11	1.12	0.15	0.97	0.09	1.73**	1.57*	1.35
1	21	2	STTGKGGKAKGKTASSKQVSR	K5Ac	0.98	0.13	0.84	0.04	0.89	0.04	0.91	0.05	0.86	0.91	0.93
1	21	3	STTGKGGKAKGKTASSKQVSR	K5Ac	1.05	0.12	0.92	0.02	0.98	0.03	0.95	0.05	0.88	0.93	0.91
1	21	2	STTGKGGKAKGKTASSKQVSR	K5UnK8UnK10UnK12Un	1.35	0.05	1.18	0.38	1.09	0.41	1.05	0.08	0.87	0.81	0.78
1	21	3	STTGKGGKAKGKTASSKQVSR	K5UnK8UnK10UnK12Un	1.42	0.09	1.55	0.68	1.15	0.39	1.15	0.09	1.1	0.81	0.81
6	21	2	GGKAKGKTASSKQVSR	K10AcK12Ac	0.54	0.08	1.83	0.26	1.6	0.82	N/A	N/A	3.42**	2.98	N/A
6	21	2	GGKAKGKTASSKQVSR	K10Ac	0.83	0.05	1.52	0.24	1.28	0.06	1.61	0.03	1.83*	1.54*	1.93**
6	21	1	GGKAKGKTASSKQVSR	K10UnK12Un	1.15	0.2	0.88	0.03	0.89	0.1	1.02	0.03	0.76	0.77	0.89
6	21	2	GGKAKGKTASSKQVSR	K10UnK12Un	1.28	0.33	0.93	0.03	1.23	0.09	1.34	0.11	0.72	0.96	1.05
9	21	2	AKGKTASSKQVSR	K12Ac	0.7	0.12	1.23	0.3	1.35	0.35	1.24	0.41	1.76*	1.93*	1.77
9	21	1	AKGKTASSKQVSR	K12Un	1.08	0.18	0.88	0.03	0.95	0.06	1.12	0.13	0.81	0.88	1.03
37	42	2	FLKHGR	K39Fm	0.87	0.17	1.26	0.07	1.09	0.1	1.16	0.01	1.45*	1.25	1.33
37	42	2	FLKHGR	K39Un	0.88	0.05	1.09	0.09	1.02	0.02	1.02	0.09	1.24*	1.16*	1.16
86	92	1	HILLAIR	R92Me1	0.96	0.05	1.18	0.1	1.07	0.07	1.12	0.12	1.23*	1.12	1.16
86	92	2	HILLAIR	R92Me1	0.9	0.05	1.15	0.19	0.99	0.03	1.3	0.02	1.27	1.09	1.44**
86	92	1	HILLAIR	R92Un	0.88	0.08	0.98	0.04	0.96	0.08	0.94	0.1	1.11	1.09	1.06
86	92	2	HILLAIR	R92Un	0.9	0.08	0.99	0.03	0.99	0.04	0.93	0.11	1.1	1.1	1.03
H2B-1					WT(L/H)		EZL2(L/H)		TXR1(L/H)		E2+T1(L/H)		Final Ratio		
start	end	z	Sequence	PTM	mean	sd	mean	sd	mean	sd	mean	sd	EZL2/WT	TXR1/WT	E2+T1/WT
1	13	2	APKKAPAAAAEKK	K3Me3	0.42	0.03	0.4	0.03	0.36	0.02	0.46	0.04	0.93	0.84	1.07
1	13	2	APKKAPAAAAEKK	A1Me1K3Me1	0.43	0.04	0.33	0.07	0.41	0.14	0.4	0.01	0.76	0.94	0.93
1	15	2	APKKAPAAAAEKKVK	K3Me3K4Ac	0.68	0.09	0.84	0.18	0.76	0.21	N/A	N/A	1.25	1.12	N/A

1	15	3	APKKAPAAAAEKKVK	K3Me3K4Ac	0.5	0.04	0.67	0.23	0.48	0.08	N/A	N/A	1.33	0.95	N/A
1	15	2	APKKAPAAAAEKKVK	K3Me2	0.53	0.1	0.47	0.23	0.44	0.05	N/A	N/A	0.89	0.83	N/A
1	15	2	APKKAPAAAAEKKVK	K3Me3	0.5	0.02	0.48	0.17	0.44	0.01	0.63	0.01	0.95	0.87*	1.25**
104	121	2	HAISEGTKAVTKFSSSTN	K115Ac	0.97	0.04	1.01	0.05	1.01	0.07	1.03	0.08	1.05	1.05	1.07
104	121	3	HAISEGTKAVTKFSSSTN	K115Ac	0.97	0.04	1.01	0.04	1.02	0.07	1.09	0.07	1.04	1.05	1.12
104	121	2	HAISEGTKAVTKFSSSTN	K115Un	1.08	0.05	0.99	0.04	1.09	0.12	0.99	0.05	0.91	1	0.91
104	121	3	HAISEGTKAVTKFSSSTN	K115Un	1.08	0.04	0.99	0.05	1.09	0.12	1.02	0.05	0.92	1.01	0.94
104	121	2	HAISEGTKAVTKFSSSTN	C-term Phos	59.8	43.8	25.2	14.4	12.9	2.42	13.6	0.95	0.42	0.22	0.23
H2B-2															
1	13	2	APKKAPAATTEKK	K3Me3	0.45	0.07	0.55	0.12	0.56	0.22	0.55	0.03	1.22	1.24	1.24
1	15	2	APKKAPAATTEKKVK	K3Me3K4Ac	0.38	0.05	0.51	0.22	0.5	0.07	0.33	0.1	1.34	1.3	0.87
1	15	2	APKKAPAATTEKKVK	K3Me2	0.39	0.06	0.48	0.25	0.5	0.07	0.33	0.1	1.25	1.28	0.86
1	15	3	APKKAPAATTEKKVK	K3Me2	0.47	0.06	0.52	0.18	0.54	0.12	N/A	N/A	1.11	1.16	N/A
1	15	2	APKKAPAATTEKKVK	K3Me3	0.47	0.05	0.46	0.16	0.37	0.07	N/A	N/A	0.99	0.78	N/A
1	15	3	APKKAPAATTEKKVK	K3Me3	0.43	0.01	0.43	0.13	0.57	0.21	0.6	0.03	1.01	1.34	1.4
104	121	2	HAISEGTKAVTKFSSSSN	K115Un	0.97	0.05	1.01	0.05	1.01	0.07	1.03	0.08	1.04	1.04	1.06
104	121	3	HAISEGTKAVTKFSSSSN	K115Un	0.97	0.04	1.01	0.04	1.01	0.07	1.08	0.08	1.04	1.05	1.12
104	121	2	HAISEGTKAVTKFSSSSN	C-term Phos	57.7	16.3	11.4	2.46	12	2.09	12.1	2.57	0.2*	0.21*	0.21*
H3					WT(L/H)		EZL2(L/H)		TXR1(L/H)		E2+T1(L/H)		Final Ratio		
start	end	z	sequence	PTM	mean	sd	mean	sd	mean	sd	mean	sd	EZL2/WT	TXR1/WT	E2+T1/WT
3	8	2	TKQTAR	K4Me1	0.57	0.12	0.89	0.09	0.76	0.09	0.64	0	1.55*	1.33	1.11
3	8	2	TKQTAR	K4Me3	0.87	0.16	0.96	0.07	1.04	0.13	1.29	0.05	1.1	1.19	1.11
3	8	1	TKQTAR	K4Un	1.16	0.05	1.42	0.2	1.41	0.03	N/A	N/A	1.22	1.21**	N/A
9	17	1	KSTGAKAPR	K9AcK14Ac	1.37	0.16	1.85	0.2	2.16	0.28	1.7	0.02	1.34*	1.57*	1.24
9	17	2	KSTGAKAPR	K9AcK14Ac	1.35	0.13	1.75	0.17	2.19	0.27	1.65	0.03	1.3*	1.63*	1.22
9	17	1	KSTGAKAPR	K9Ac	1.05	0.09	1.32	0.06	1.24	0.14	1.14	0.01	1.25*	1.18	1.08
9	17	2	KSTGAKAPR	K9Ac	1.04	0.07	1.19	0.03	1.16	0.11	1.09	0.02	1.14*	1.12	1.05
9	17	1	KSTGAKAPR	K14Ac	1.05	0.08	1.34	0.12	1.25	0.14	1.45	0.52	1.27*	1.19	1.38
9	17	2	KSTGAKAPR	K14Ac	0.97	0.07	1.23	0.12	1.17	0.12	1.07	0.04	1.27*	1.2	1.1
9	17	1	KSTGAKAPR	K9UnK14Un	0.75	0.1	0.87	0.09	0.63	0.04	0.61	0.05	1.16	0.84	0.81
9	17	2	KSTGAKAPR	K9UnK14Un	0.73	0.1	0.85	0.09	0.61	0.04	0.64	0.1	1.16	0.83	0.88
18	26	1	KQLASKAAR	K18AcK23Ac	0.94	0.01	1.22	0.18	1.15	0.05	0.93	0.01	1.29	1.23*	0.99
18	26	2	KQLASKAAR	K18AcK23Ac	0.89	0.07	1.14	0.09	1.14	0.06	0.91	0.02	1.28*	1.28**	1.02
18	26	1	KQLASKAAR	K23Ac	0.8	0.04	1.02	0.07	0.91	0.03	0.77	0.02	1.28*	1.15*	0.97

18	26	2	KQLASKAAR	K23Ac	0.79	0.06	0.97	0.03	0.92	0.02	0.78	0.02	1.24*	1.17*	0.99
18	26	1	KQLASKAAR	K23Me1	0.54	0.17	0.66	0.18	0.96	0.25	0.62	0	1.23	1.79	1.15
18	26	2	KQLASKAAR	K23Me2	0.35	0.05	0.21	0.05	0.3	0.05	N/A	N/A	0.6*	0.78*	N/A
18	26	2	KQLASKAAR	K23Me3	0.47	0.09	0.21	0.03	0.25	0.03	0.16	0.04	0.44*	0.53*	0.34*
18	26	2	KQLASKAAR	K23Fm	0.91	0.04	1.14	0.1	1.14	0.06	0.91	0.02	1.26*	1.26**	1
18	26	1	KQLASKAAR	K18UnK23Un	0.9	0.05	1.11	0.19	0.87	0.09	1.04	0.04	1.23	0.97	1.15
18	26	2	KQLASKAAR	K18UnK23Un	0.95	0.08	1.19	0.2	0.96	0.09	1.02	0.04	1.25	1	1.07
27	40	2	KSAPATGGIKKPHR	K27Ac	0.71	0.09	0.95	0.18	1.05	0.16	0.93	0.02	1.34	1.47	1.31*
27	40	2	KSAPATGGIKKPHR	K27Me1	1	0.07	2	0.08	0.18	0	0.05	0	1.99**	0.18**	0.05**
27	40	2	KSAPATGGIKKPHR	K27Me2	0.76	0.13	0.14	0.02	0.3	0.01	0.15	0.01	0.18*	0.4*	0.2*
27	40	3	KSAPATGGIKKPHR	K27Me2	0.7	0.09	0.05	0.04	0.33	0.11	0.06	0.06	0.07**	0.47*	0.09**
27	40	2	KSAPATGGIKKPHR	K27Me3	0.53	0.11	0.12	0.03	0.52	0.05	0.12	0.01	0.22*	0.99	0.23*
27	40	3	KSAPATGGIKKPHR	K27Me3	0.5	0.1	0.02	0.01	0.44	0.02	0.07	0.01	0.04*	0.89	0.14*
27	40	2	KSAPATGGIKKPHR	K27Me1K36Me1	0.62	0.12	0.8	0.1	0.06	0.01	0.01	0	1.28	0.1*	0.02*
27	40	2	KSAPATGGIKKPHR	K27UnK36Un	1.68	0.2	2.49	0.84	4.99	0.31	7.48	0.34	1.49	2.97**	4.46**
53	63	2	KYQKSTDLLIR	K56Me1	0.56	0.15	0.63	0.21	1.1	0.26	0.84	0	1.12	1.97*	1.5
53	63	2	KYQKSTDLLIR	K56Un	0.99	0.04	1.12	0.08	0.99	0.01	0.99	0.02	1.13	1	1
53	63	3	KYQKSTDLLIR	K56Un	1.09	0.05	1.2	0.12	1.09	0.03	1.15	0.17	1.1	1	1.06
73	83	2	DIAHEFKAELR	R83Un	0.92	0.13	1.13	0.1	1	0.01	0.97	0.02	1.23	1.09	1.06
73	83	2	DIAHEFKAELR	R83Me1	0.87	0.08	1.07	0.08	0.95	0.07	1.11	0.08	1.23*	1.09	1.27
116	128	2	RVTIMTKDMQLAR	R128Un	0.88	0.2	1.12	0.07	1	0.02	N/A	N/A	1.27	1.14	N/A
116	128	3	RVTIMTKDMQLAR	R128Un	0.83	0.21	1.16	0.1	0.96	0.02	1	0.01	1.39	1.15	1.14
116	128	2	RVTIMTKDMQLAR	R128Me1	0.75	0.2	1.07	0.21	1.17	0.19	0.97	0.02	1.43	1.57	1.16
H3.3															
9	17	2	KSTGVKAPR	K9AcK14Ac	0.73	0.11	0.84	0.09	0.61	0.04	0.63	0.08	1.16	0.84	0.87
9	17	1	KSTGVKAPR	K14Fm	0.72	0.11	0.88	0.08	0.63	0.05	0.61	0.05	1.23	0.88	0.83
9	17	2	KSTGVKAPR	K14Fm	0.72	0.1	0.86	0.09	0.62	0.03	0.64	0.1	1.18	0.85	0.94
9	17	2	KSTGVKAPR	K14Fm	0.8	0.15	0.89	0.08	0.68	0.03	0.66	0.05	1.12	0.86	1.04
9	17	1	KSTGVKAPR	K9UnK14Un	0.54	0.19	0.57	0.11	0.72	0.17	0.5	0.01	1.07	1.34	0.94
9	17	2	KSTGVKAPR	K9UnK14Un	0.49	0.16	0.56	0.11	0.72	0.08	0.51	0	1.14	1.46	1.04
10	17	1	STGVKAPR	K14Fm	0.67	0.09	0.75	0.09	0.48	0.06	N/A	N/A	1.12	0.71*	N/A
18	26	1	KQLATKAAR	K18AcK23Ac	0.8	0.04	1.02	0.07	0.91	0.03	0.76	0.03	1.28*	1.14*	0.95
18	26	2	KQLATKAAR	K18AcK23Ac	0.81	0.08	0.97	0.03	0.72	0.34	0.78	0.02	1.2	0.88	0.96
18	26	1	KQLATKAAR	K23Ac	0.9	0.05	1.09	0.31	0.9	0.07	1.04	0.07	1.2	1	1.15

18	26	2	KQLATKAAR	K23Ac	0.96	0.09	1.2	0.2	0.96	0.09	1.03	0.04	1.25	1	1.07
18	26	1	KQLATKAAR	K18UnK23Un	0.65	0.2	0.66	0.18	0.96	0.16	0.61	0	1.02	1.47	0.93
18	26	2	KQLATKAAR	K18UnK23Un	1.38	0.22	1.61	0.3	2.16	0.3	0.97	0.01	1.17	1.56*	0.7
18	26	1	KQLATKAAR	K23Fm	0.79	0.04	0.83	0.24	1.09	0.05	0.76	0.03	1.06	1.38**	0.96
18	26	2	KQLATKAAR	K23Fm	0.79	0.06	0.97	0.03	0.93	0.03	0.78	0.02	1.23*	1.18*	0.99
53	63	2	KYQKTTDLLIR	K56Ac	0.99	0.04	1.13	0.08	0.99	0.01	0.99	0.02	1.13	1	1
53	63	3	KYQKTTDLLIR	K56Ac	1.1	0.04	1.2	0.12	1.06	0.07	1.05	0.03	1.09	0.97	0.95
53	63	2	KYQKTTDLLIR	K56Un	0.54	0.16	0.61	0.17	1.07	0.28	0.84	0	1.14	2.01	1.57
53	63	2	KYQKTTDLLIR***	K56Ac	0.92	0.16	1.18	0.17	1.1	0.06	0.99	0.01	1.28	1.19	1.08
H4					WT(L/H)		EZL2(L/H)		TXR1(L/H)		E2+T1(L/H)		Final Ratio		
start	end	z	Sequence	PTM	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	EZL2/WT	TXR1/WT	E2+T1/WT
1	7	1	AGGKGGK	K4Ac	1.31	0.11	1.36	0.17	1.13	0.16	1.67	0.11	1.04	0.87	1.28
1	11	1	AGGKGGKGMGK	K4AcK7Ac	0.4	0.12	0.62	0.11	0.76	0.15	0.93	0.22	1.53	1.87*	2.3
1	11	1	AGGKGGKGMGK	K7Ac	0.86	0.14	1.01	0.02	1.11	0.11	1.17	0.15	1.17	1.29*	1.36
1	11	1	AGGKGGKGMGK	K4UnK7Un	1.96	0.24	1.44	0.13	1.06	0.17	0.89	0.16	0.74*	0.54**	0.46*
1	16	1	AGGKGGKGMGKVGAKR	K4AcK7AcK11AcK15Ac	0.4	0.11	1.09	0.11	2.06	0.4	1.16	N/A	2.72**	5.14*	2.9
1	16	2	AGGKGGKGMGKVGAKR	K4AcK7AcK11AcK15Ac	0.42	0.1	1.12	0.13	2.02	0.39	1.14	0.08	2.69**	4.85*	2.73**
1	16	1	AGGKGGKGMGKVGAKR	K4AcK7AcK11Ac	0.54	0.14	0.85	0.12	1.18	0.21	0.75	0.05	1.56*	2.17*	1.38
1	16	2	AGGKGGKGMGKVGAKR	K4AcK7AcK11Ac	0.54	0.15	0.86	0.1	1.2	0.21	0.73	0	1.59*	2.21*	1.35
1	16	1	AGGKGGKGMGKVGAKR	K4AcK7Ac	0.61	0.14	0.78	0.07	0.93	0.12	0.61	0.05	1.28	1.52*	1
1	16	2	AGGKGGKGMGKVGAKR	K4AcK7Ac	0.58	0.14	0.79	0.07	0.91	0.12	0.6	0.01	1.35	1.56*	1.03
1	16	1	AGGKGGKGMGKVGAKR	K7Ac	1.09	0.1	1.14	0.03	1.16	0.08	0.93	0.12	1.05	1.07	0.85
1	16	2	AGGKGGKGMGKVGAKR	K7Ac	1.07	0.07	1.14	0.05	1.14	0.07	0.9	0.04	1.07	1.07	0.84*
1	16	2	AGGKGGKGMGKVGAKR	K4UnK7UnK11UnK15Un	1.89	0.28	1.45	0.17	1.19	0.23	1.57	0.01	0.77	0.63*	0.83
5	16	1	GGKGMGKVGAKR	K7AcK11AcK15Ac	0.71	0.13	1.13	0.05	1.75	0.29	N/A	N/A	1.59*	2.46*	N/A
5	16	2	GGKGMGKVGAKR	K7AcK11AcK15Ac	0.56	0.18	1.13	0.11	1.65	0.33	N/A	N/A	2.01*	2.94*	N/A
5	16	1	GGKGMGKVGAKR	K7AcK11Ac	0.81	0.15	1.06	0.12	1.27	0.14	1.17	0.21	1.31	1.58*	1.46
5	16	2	GGKGMGKVGAKR	K7AcK11Ac	0.79	0.16	1.01	0.1	1.21	0.13	1.21	N/A	1.28	1.53*	1.53
5	16	1	GGKGMGKVGAKR	K7Ac	1.04	0.11	1.11	0.05	1.16	0.07	1.21	N/A	1.07	1.11	1.16
5	16	2	GGKGMGKVGAKR	K7Ac	1	0.06	1.07	0.1	1.12	0.06	1.43	0.55	1.07	1.12	1.42
5	16	1	GGKGMGKVGAKR	K7UnK11UnK15Un	1.94	0.24	1.4	0.2	1.14	0.2	1.64	0.15	0.72*	0.59*	0.85
8	16	1	GMGKVGAKR	K11AcK15Ac	1.7	0.24	2.3	0.21	2.37	0.99	1.2	0.34	1.35*	1.39	0.7
8	16	1	GMGKVGAKR	K11Ac	2.15	0.07	1.95	0.29	1.65	0.41	1.18	0.39	0.91	0.77	0.55
8	16	1	GMGKVGAKR	K11UnK15Un	1.83	0.23	1.33	0.16	1.08	0.19	1.69	0.21	0.73*	0.59*	0.92

46	55	1	ISSFIYDDSR	R55Me1	1.52	0.02	1.11	0.17	1.29	0.43	N/A	N/A	0.73	0.85	N/A
46	55	1	ISSFIYDDSR	R55Un	0.97	0.04	1.01	0.1	0.97	0.08	0.97	0.05	1.03	1	0.99
68	77	2	DAVTYTEHAR	R77Me1	1.1	0.13	1.12	0.04	1.1	0.07	1.02	0	1.02	1	0.93
68	77	2	DAVTYTEHAR	R77Un	1	0.03	0.98	0.01	0.99	0.03	0.98	0.05	0.97	0.98	0.97
78	92	2	RKTVTAMDVVYALKR	K91Un	1.03	0.05	1	0.02	0.98	0.01	1.02	0.04	0.97	0.96	0.99
78	92	2	RKTVTAMDVVYALKR	K91Me1	1.05	0.04	0.87	0.35	1.09	0.01	N/A	N/A	0.83	1.04	N/A

Note: *:p<0.05; **:p<0.01;***:incomplete propionlyation on the peptide N-ternimus.

Table S3.1 Percentages of histone H3 PTMs modified at K27 and K36

	WT		EZL2		TXR1		E2:T1	
Modifications	mean	sd	mean	sd	mean	sd	mean	sd
K27Ac	5.05%	1.17%	5.06%	3.03%	5.75%	0.77%	3.18%	0.13%
K27Me1K36Me1	1.43%	0.62%	2.11%	0.20%	<0.2%	0.00%	<0.2%	0.00%
K27Me1	35%	6.54%	57.50%	0.73%	6.68%	1.68%	4.13%	0.06%
K27Me2	20.70%	7.35%	1.27%	0.50%	8.73%	0.70%	2.04%	0.16%
K27Me3	3.73%	1.55%	0.38%	0.06%	3.05%	0.44%	0.76%	0.21%
Unmodified	34.20%	1.85%	33.70%	3.40%	75.80%	0.49%	89.90%	0.05%

Note: Relative quantification is based on the peak areas of a tryptic peptide (27-KSAPATGGIKKPHR-40) with different modification states.

Table S3.2 Final ratio of all PTMs converted from WT/^{15}N WT

PepID	Protein	z	Sequence	PTM	EZL2	TXR1	E2:T1
1	H2A	2	STTGKGGKAKGKTASSKQVSR	K5AcK8AcK10AcK12Ac	3.7	3.26	3.77
2	H2A	2	STTGKGGKAKGKTASSKQVSR	K5AcK8AcK12Ac	2.41	1.94	2.02
3	H2A	2	STTGKGGKAKGKTASSKQVSR	K5AcK8Ac	1.78	1.58	1.44
4	H2A	2	STTGKGGKAKGKTASSKQVSR	K5AcK10Ac	1.69	1.5	1.5
5	H2A	2	STTGKGGKAKGKTASSKQVSR	K5Ac	0.86	0.91	0.93
6	H2A	2	STTGKGGKAKGKTASSKQVSR	K5UnK8UnK10UnK12Un	0.87	0.81	0.78
7	H2A	2	GGKAKGKTASSKQVSR	K10Ac	1.83	1.54	1.93
8	H2A	2	GGKAKGKTASSKQVSR	K10UnK12Un	0.72	0.96	1.05
9	H2A	2	AKGKTASSKQVSR	K12Ac	1.76	1.93	1.77
10	H2A	1	AKGKTASSKQVSR	K12Un	0.81	0.88	1.03
11	H2A	2	FLKHGR	K39Fm	1.45	1.25	1.33
12	H2A	2	FLKHGR	K39Un	1.24	1.16	1.16
13	H2A	1	HILLAIR	R92Me1	1.23	1.12	1.16
14	H2A	1	HILLAIR	R92Un	1.11	1.09	1.06
15	H2B-1	2	APKKAPAAAAEKK	K3Me3	0.93	0.84	1.07
16	H2B-1	2	APKKAPAAAAEKK	A1Me1K3Me1	0.76	0.94	0.93
17	H2B-1	2	APKKAPAAAAEKKVK	K3Me3	0.95	0.87	1.25
18	H2B-1	2	HAISEGTKAVTKFSSSTN	K115Ac	1.05	1.05	1.07
19	H2B-1	2	HAISEGTKAVTKFSSSTN	K115Un	0.91	1	0.91
20	H2B-1	2	HAISEGTKAVTKFSSSTN	C-term Phos	0.42	0.22	0.23
21	H2B-2	2	APKKAPAATTEKK	K3Me3	1.22	1.24	1.24
22	H2B-2	2	APKKAPAATTEKKVK	K3Me3K4Ac	1.34	1.3	0.87
23	H2B-2	2	APKKAPAATTEKKVK	K3Me2	1.25	1.28	0.86
24	H2B-2	3	APKKAPAATTEKKVK	K3Me3	1.01	1.34	1.4
25	H2B-2	2	HAISEGTKAVTKFSSSSN	K115Un	1.04	1.04	1.06
26	H2B-2	3	HAISEGTKAVTKFSSSSN	K115Un	1.04	1.05	1.12
27	H2B-2	2	HAISEGTKAVTKFSSSSN	C-term Phos	0.2	0.21	0.21
28	H3	2	TKQTAR	K4Me1	1.55	1.33	1.11
29	H3	2	TKQTAR	K4Me3	1.1	1.19	1.11
30	H3	2	KSTGAKAPR	K9AcK14Ac	1.3	1.63	1.22
31	H3	2	KSTGAKAPR	K9Ac	1.14	1.12	1.05
32	H3	2	KSTGAKAPR	K14Ac	1.27	1.2	1.1
33	H3	2	KSTGAKAPR	K9UnK14Un	1.16	0.83	0.88
34	H3	2	KQLASKAAR	K18AcK23Ac	1.28	1.28	1.02
35	H3	2	KQLASKAAR	K23Ac	1.24	1.17	0.99
36	H3	1	KQLASKAAR	K23Me1	1.23	1.79	1.15
37	H3	2	KQLASKAAR	K18UnK23Un	1.25	1	1.07
38	H3	2	KSAPATGGIKKPHR	K27Ac	1.34	1.47	1.31
39	H3	2	KSAPATGGIKKPHR	K27Me1	1.99	0.18	0.05
40	H3	2	KSAPATGGIKKPHR	K27Me2	0.18	0.4	0.2

41	H3	2	KSAPATGGIKKPHR	K27Me3	0.22	0.99	0.23
42	H3	2	KSAPATGGIKKPHR	K27Me1K36Me1	1.28	0.1	0.02
43	H3	2	KSAPATGGIKKPHR	K27UnK36Un	1.49	2.97	4.46
44	H3	2	KYQKSTDLLIR	K56Me1	1.12	1.97	1.5
45	H3	2	KYQKSTDLLIR	K56Un	1.13	1	1
46	H3	2	DIAHEFKAELR	R83Un	1.23	1.09	1.06
47	H3	2	DIAHEFKAELR	R83Me1	1.23	1.09	1.27
48	H3	3	RVTIMTKDMQLAR	R128Un	1.39	1.15	1.14
49	H3	2	RVTIMTKDMQLAR	R128Me1	1.43	1.57	1.16
50	H3.3	2	KSTGVKAPR	K9AcK14Ac	1.16	0.84	0.87
51	H3.3	2	KSTGVKAPR	K9UnK14Un	1.14	1.46	1.04
52	H3.3	2	KQLATKAAR	K18AcK23Ac	1.2	0.88	0.96
53	H3.3	2	KQLATKAAR	K23Ac	1.25	1	1.07
54	H3.3	2	KQLATKAAR	K18UnK23Un	1.17	1.56	0.7
55	H3.3	2	KYQKTTDLLIR	K56Ac	1.13	1	1
56	H3.3	2	KYQKTTDLLIR	K56Un	1.14	2.01	1.57
57	H3.3	2	KQLATKAAR	K18UnK23Un	1.17	1.56	0.7
58	H3.3	2	KYQKTTDLLIR	K56Ac	1.13	1	1
59	H3.3	2	KYQKTTDLLIR	K56Un	1.14	2.01	1.57
60	H4	1	AGGKGGK	K4Ac	1.04	0.87	1.28
61	H4	1	AGGKGGKGMGK	K4UnK7Un	0.74	0.54	0.46
62	H4	2	AGGKGGKGMGKVGAKR	K4AcK7AcK11AcK15Ac	2.69	4.85	2.73
63	H4	2	AGGKGGKGMGKVGAKR	K4AcK7AcK11Ac	1.59	2.21	1.35
64	H4	2	AGGKGGKGMGKVGAKR	K4AcK7Ac	1.35	1.56	1.03
65	H4	2	AGGKGGKGMGKVGAKR	K4UnK7UnK11UnK15Un	0.77	0.63	0.83
66	H4	2	GGKGMGKVGAKR	K7AcK11Ac	1.28	1.53	1.53
67	H4	2	GGKGMGKVGAKR	K7Ac	1.07	1.12	1.42
68	H4	1	GGKGMGKVGAKR	K7UnK11UnK15Un	0.72	0.59	0.85
69	H4	1	GMGKVGAKR	K11AcK15Ac	1.35	1.39	0.7
70	H4	1	GMGKVGAKR	K11Ac	0.91	0.77	0.55
71	H4	1	GMGKVGAKR	K11UnK15Un	0.73	0.59	0.92
72	H4	2	DAVTYTEHAR	R77Me1	1.02	1	0.93
73	H4	2	DAVTYTEHAR	R77Un	0.97	0.98	0.97

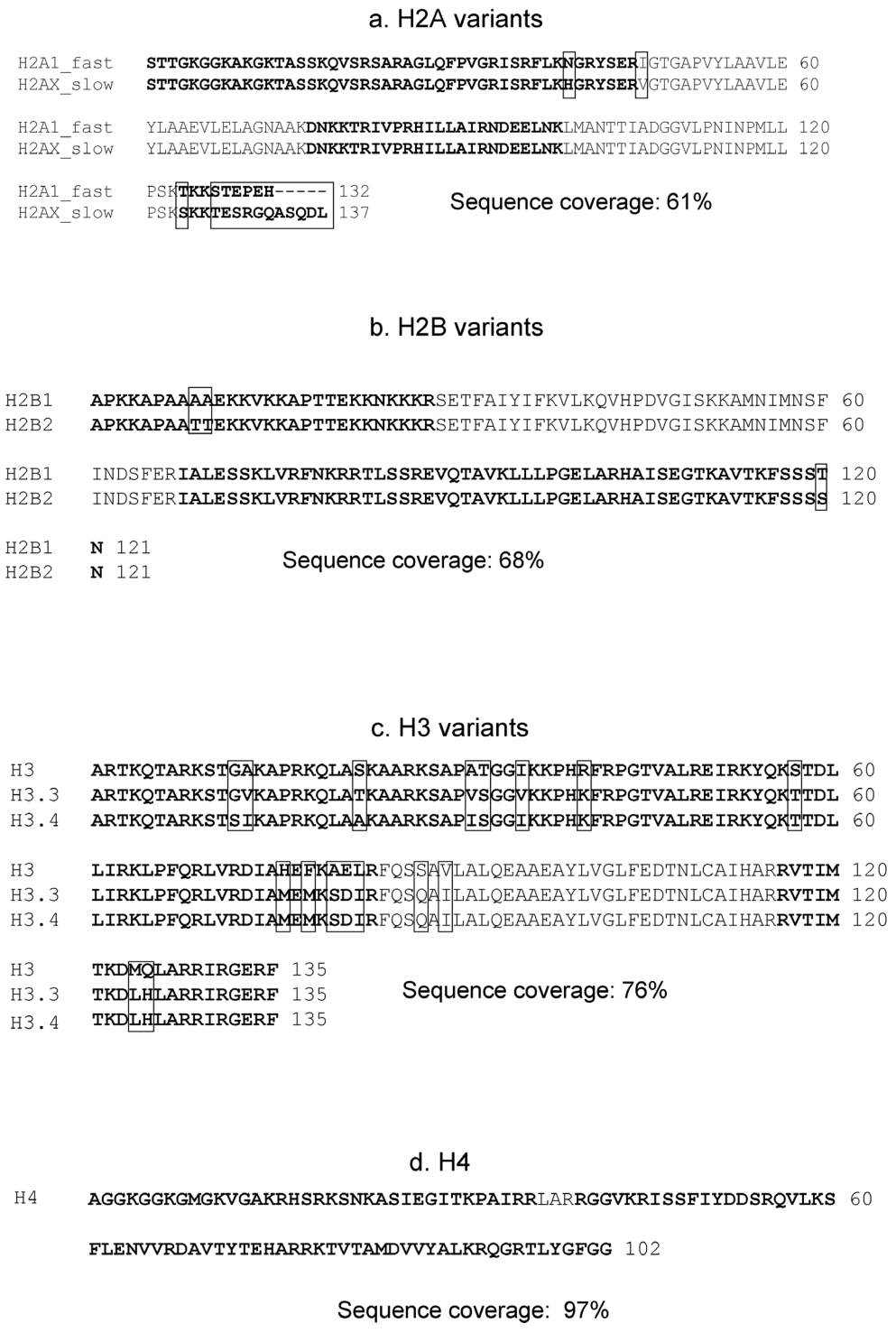


Figure 3.1 Histone, histone variants and sequence coverages

Pep ID	PTM	Z	EZL2/WT	TXR1/WT	E2:T1/WT
27-KSAPATGGIKKPHR-40					
1	K27Ac	2	1.34	1.47	1.31*
2	K27Me1	2	1.99**	0.18**	0.05**
3	K27Me2	2	0.18*	0.4*	0.2*
4	K27Me2	3	0.07**	0.47*	0.09**
5	K27Me3	2	0.22*	0.99	0.23*
6	K27Me3	3	0.04*	0.89	0.14*
7	K27Me1K36Me1	2	1.28	0.1*	0.02*
8	K27UnK36Un	2	1.49	2.97**	4.46**

*p<0.05; **p<0.01

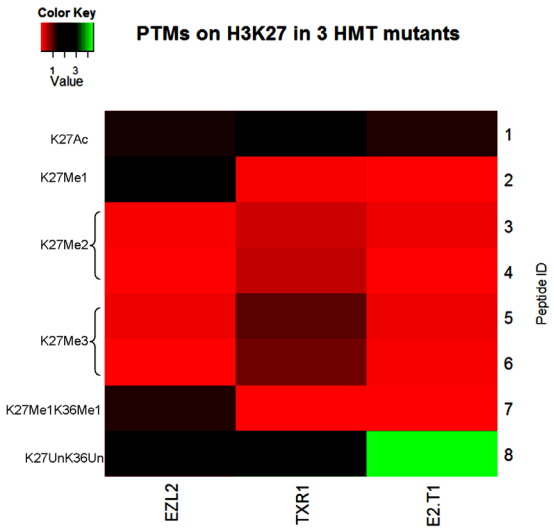
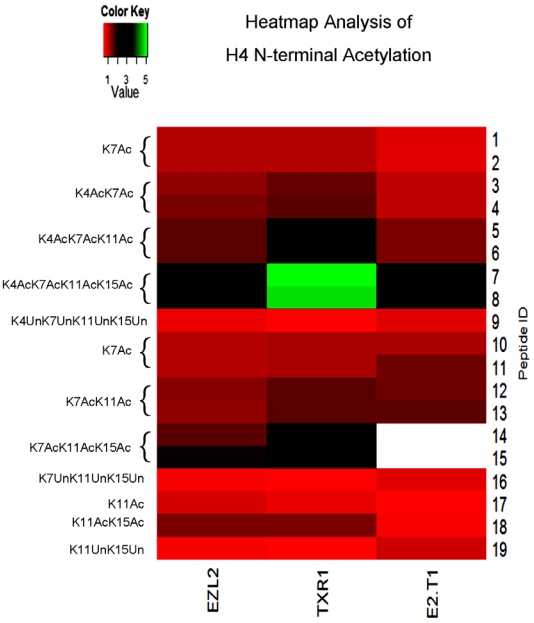


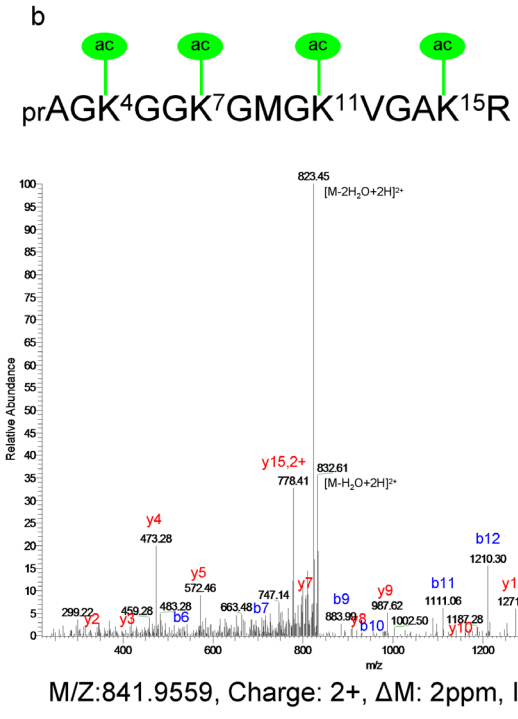
Figure 3.2 Methylation of H3K27 in 3 HTM mutants

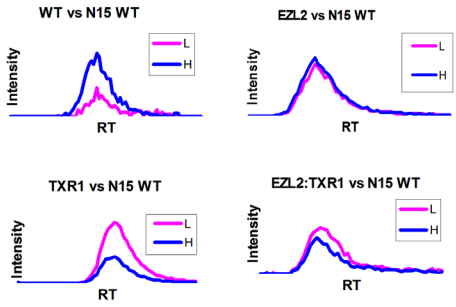
a

Pep ID	PTM	Z	EZL2/WT	TXR1/WT	E2:T1/WT
1-AGGKGGKGMGKVGAKR-16					
1	K7Ac	1	1.05	1.07	0.85
2	K7Ac	2	1.07	1.07	0.84*
3	K4AcK7Ac	1	1.28	1.52*	1
4	K4AcK7Ac	2	1.35	1.56*	1.03
5	K4AcK7AcK11Ac	1	1.56*	2.17*	1.38
6	K4AcK7AcK11Ac	2	1.59*	2.21*	1.35
7	K4AcK7AcK11AcK15Ac	1	2.72**	5.14*	2.9
8	K4AcK7AcK11AcK15Ac	2	2.69**	4.85*	2.73**
9	K4UnK7UnK11UnK15Un	2	0.77	0.63*	0.83
5-GGKGMGKVGAKR-16					
10	K7Ac	1	1.07	1.11	1.16
11	K7Ac	2	1.07	1.12	1.42
12	K7AcK11Ac	1	1.31	1.58*	1.46
13	K7AcK11Ac	2	1.28	1.53*	1.53
14	K7AcK11AcK15Ac	1	1.59*	2.46*	NA
15	K7AcK11AcK15Ac	2	2.01*	2.94*	NA
16	K7UnK11UnK15Un	1	0.72*	0.59*	0.85
8-GMGKVGAKR-16					
17	K11Ac	1	0.91	0.77	0.55
18	K11AcK15Ac	1	1.35*	1.39	0.7
19	K11UnK15Un	1	0.73*	0.59*	0.92



b





	L/H	s.d
AUC(WT)/AUC(N15 WT)	0.42	(0.1)
AUC(EZL2)/AUC(N15 WT)	1.12	(0.13)
AUC(TXR1)/AUC(N15 WT)	2.02	(0.39)
AUC(EZL2:TXR1)/AUC(N15 WT)	1.14	(0.08)

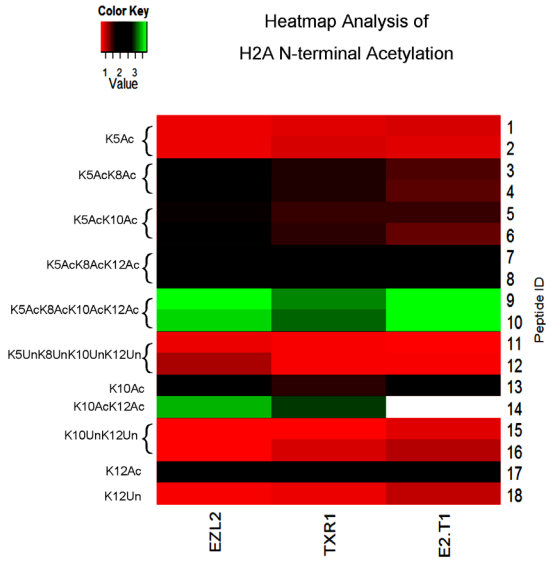
	Final ratio
EZL2/WT	2.69**
TXR1/WT	4.85*
EZL2:TXR1/WT	2.73**

*p<0.05; **p<0.01

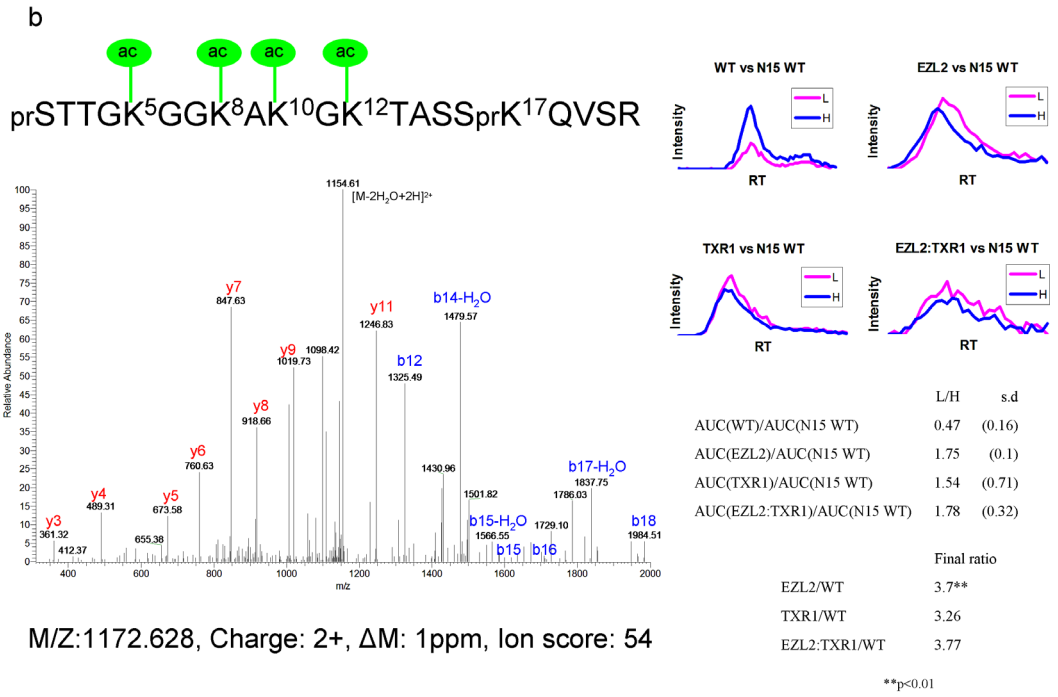
Figure 3.3 a) Combinatorial modifications in H4 N-terminal domains; b) Hyper-acetylation of H4 in its N-terminal domains

a

Pep ID	PTM	Z	EZL2/WT	TXR1/WT	E2:T1/WT
1-STTGKGGKAKGKTASSKQVSR-21					
1	K5Ac	2	0.86	0.91	0.93
2	K5Ac	3	0.88	0.93	0.91
3	K5AcK8Ac	2	1.78**	1.58*	1.44
4	K5AcK8Ac	3	1.76**	1.6*	1.38
5	K5AcK10Ac	2	1.69**	1.5*	1.5*
6	K5AcK10Ac	3	1.73**	1.57*	1.35
7	K5AcK8AcK12Ac	2	2.41*	1.94	2.02
8	K5AcK8AcK12Ac	3	2.4*	2.01	1.99
9	K5AcK8AcK10AcK12Ac	2	3.7**	3.26	3.77
10	K5AcK8AcK10AcK12Ac	3	3.54**	3.15	3.77
11	K5UnK8UnK10UnK12Un	2	0.87	0.81	0.78
12	K5UnK8UnK10UnK12Un	3	1.1	0.81	0.81
6-GGKAKGKTASSKQVSR-21					
13	K10Ac	2	1.83*	1.54*	1.93**
14	K10AcK12Ac	2	3.42**	2.98	NA
15	K10UnK12Un	1	0.76	0.77	0.89
16	K10UnK12Un	2	0.72	0.96	1.05
9-AKGKTASSKQVSR-21					
17	K12Ac	2	1.76*	1.93*	1.77
18	K12Un	1	0.81	0.88	1.03



b



	L/H	s.d
AUC(WT)/AUC(N15 WT)	0.47	(0.16)
AUC(EZL2)/AUC(N15 WT)	1.75	(0.1)
AUC(TXR1)/AUC(N15 WT)	1.54	(0.71)
AUC(EZL2:TXR1)/AUC(N15 WT)	1.78	(0.32)

	Final ratio
EZL2/WT	3.7**
TXR1/WT	3.26
EZL2:TXR1/WT	3.77

**p<0.01

Figure 3.4 a) Combinatorial modifications in H2A N-terminal domains; b) Hyper-acetylation of H2A in its N-terminal Domains

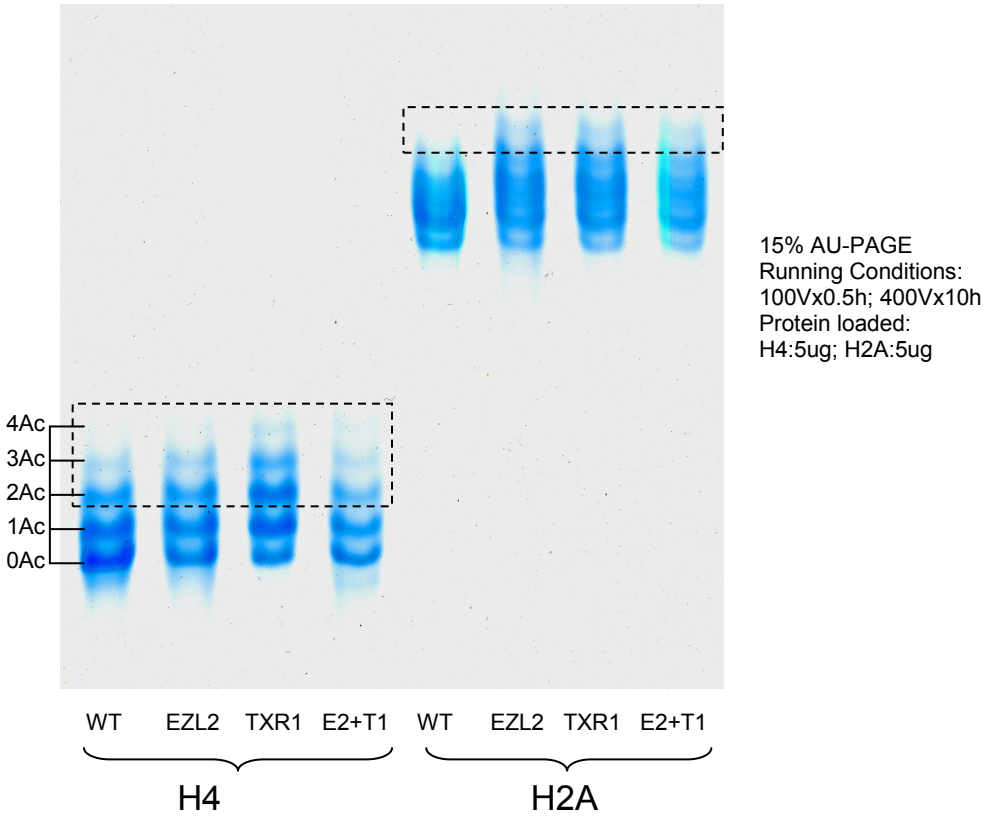
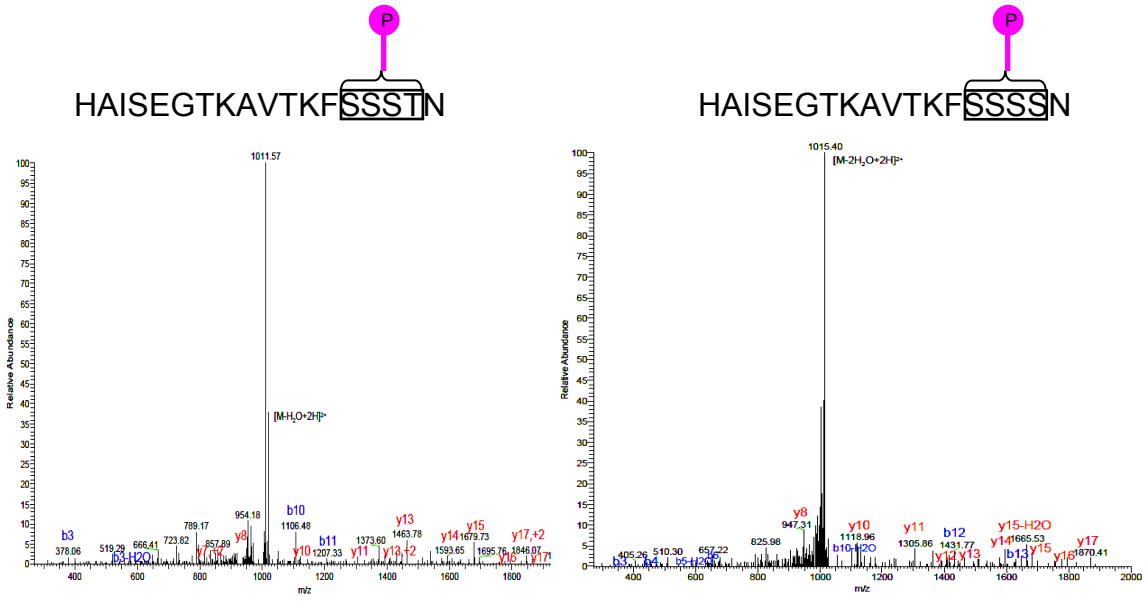


Figure 3.5 Acetic Acid-Urea gel electrophoresis analysis of acetylation of H2A and H4 in their N-terminal domains



M/Z:1028.9874, Charge: 2+, ΔM: 1ppm, Ion score: 38 M/Z:1033.451, Charge: 2+, ΔM: 9ppm, Ion score: 31

Figure 3.6 Down-regulation of H2B C-terminal phosphorylation

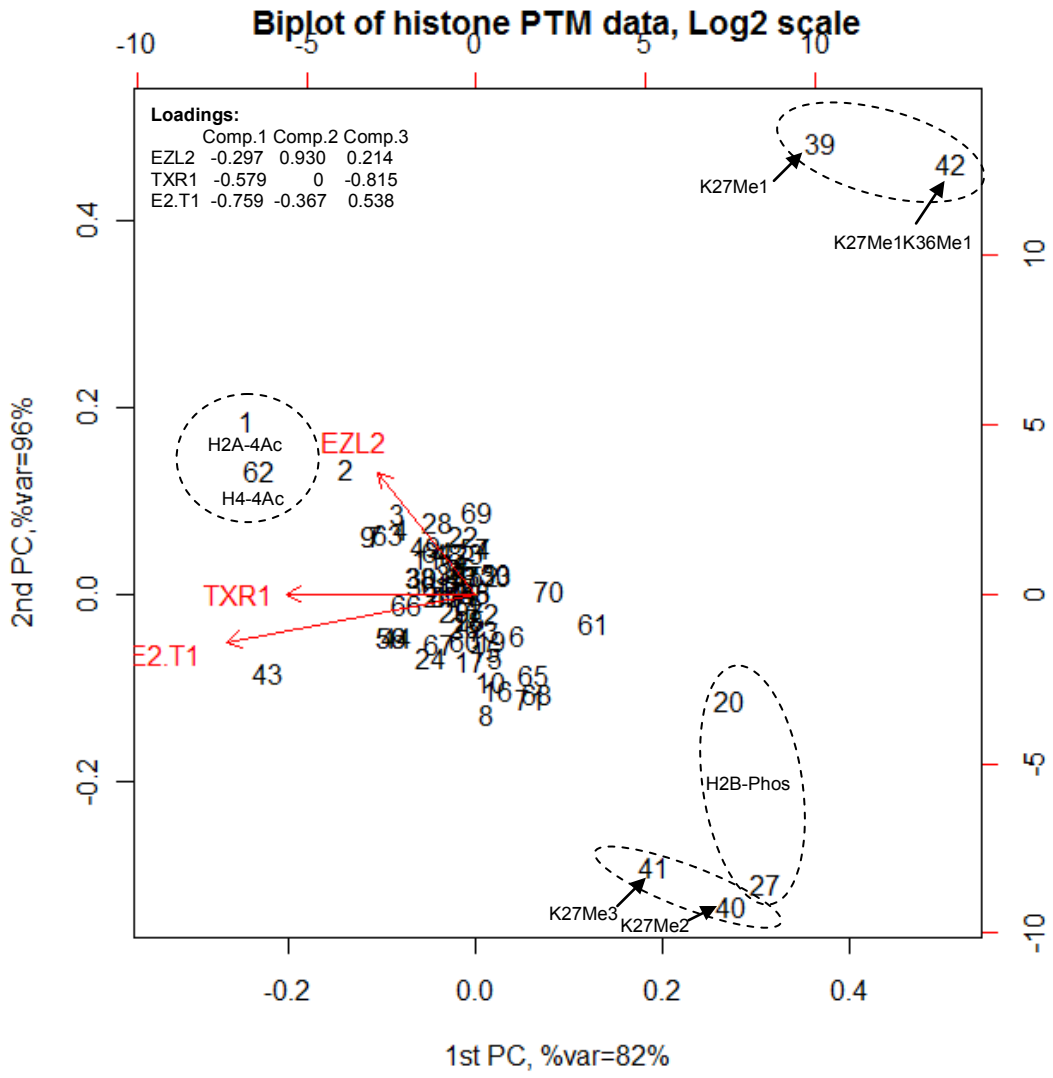


Figure 3.7 PCA bioplot analysis of histone PTMs

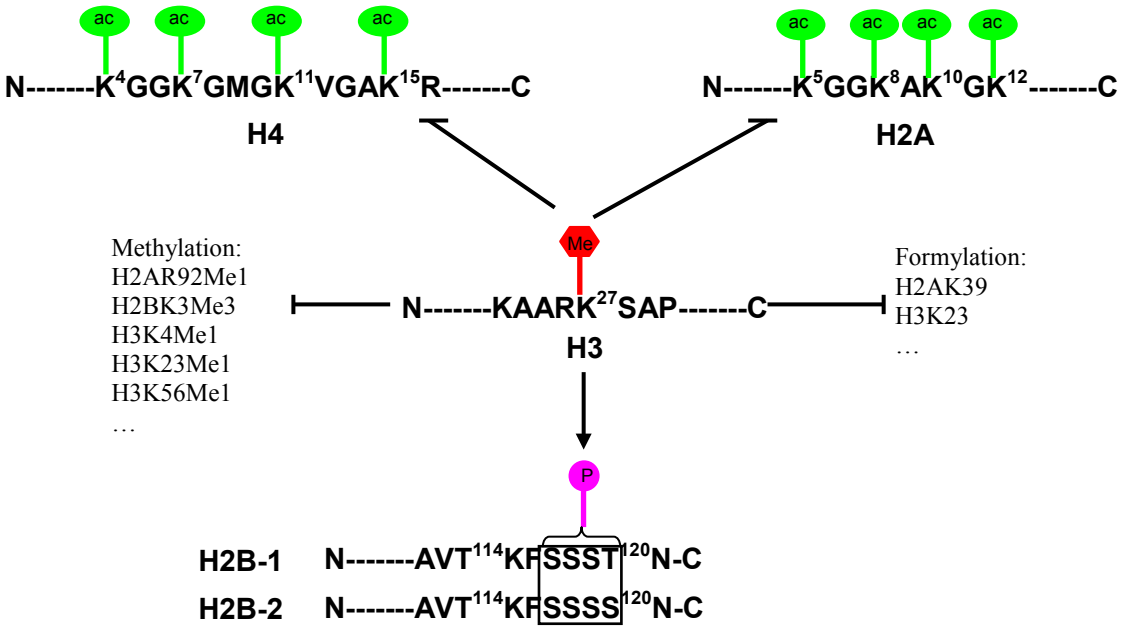


Figure 3.8 Crosstalk related with H3K27 methylation

Generally, methylation of H3K27 inhibits the acetylation of H4 and H2A in their N-terminal domians but promotes the H2B C-terminal phosphorylation. Hyper-methylation of H3K27 also prevents some off-site methylations and formylations in the core histones.

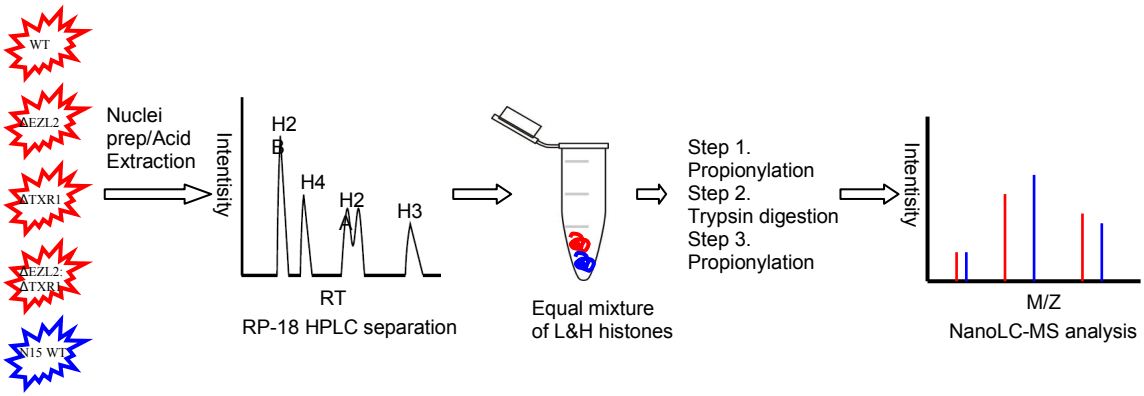


Figure S3.1 General experimental procedure

3.6 Appdendix-R code for PCA and heatmap

```
#PCA
PTM=read.table("ALLPTMLOG.txt",header=T,sep="\t")
print(summary(PTM),digits=2)

#PCA analysis
pca.ptm<-princomp(PTM[,6:8], scale=TRUE, center=TRUE, cor=F)
summary(pca.ptm)
loadings(pca.ptm)

#Print cummulative proportion of variance explained
cumprop=cumsum(pca.ptm$sdev^2)/sum(pca.ptm$sdev^2)
print(cumprop,digits=2)

#Biplot of histone PTM data
biplot(pca.ptm, xlab="1st PC, %var=82%",ylab="2nd PC,%var=96%")
title("Biplot of histone PTM data, Log2 scale")

#Heatmap
install.packages("gplots")
install.packages("RColorBrewer")
library(RColorBrewer)
library(gplots)

PTM=read.table("H4Ac.txt",header=T,sep="\t")
row.names(PTM) <- PTM$PepID
PTM <- PTM[,1:6]
PTM_matrix <- data.matrix(PTM)
PTM_heatmap <- heatmap.2(PTM_matrix[,4:6], Rowv=NA, Colv=NA, col=redgreen(75),
scale="none", key=TRUE, keysize=1, density.info="none", trace="none", cexRow=1.5,
cexCol=1.5, margins=c(6,4), ylab= "Peptide ID", main = " H4 acetylation of N-terminus")

PTM=read.table("H2Aac.txt",header=T,sep="\t")
row.names(PTM) <- PTM$PepID
PTM <- PTM[,1:5]
PTM_matrix <- data.matrix(PTM)
PTM_heatmap <- heatmap.2(PTM_matrix[,3:5], Rowv=NA, Colv=NA, col=redgreen(75),
scale="none", key=TRUE, keysize=1, density.info="none", trace="none", cexRow=1.5,
cexCol=1.5, margins=c(6,4), ylab= "Peptide ID", main = " H2A acetylation of N-
terminus")

PTM=read.table("H3K27me.txt",header=T,sep="\t")
PTM <- PTM[,1:5]
PTM_matrix <- data.matrix(PTM)
```

```
PTM_heatmap <- heatmap.2(PTM_matrix[,3:5], Rowv=NA, Colv=NA, col=redgreen(75),
scale="none", key=TRUE, keysize=1, density.info="none", trace="none", cexRow=1.5,
cexCol=1.5, margins=c(6,4), ylab= "Peptide ID", main = "PTMs on H3K27 in 3 HMT
mutants ")

PTM=read.table("ALLPTM.txt",header=T,sep="\t")
PTM <- PTM[,1:8]
PTM_matrix <- data.matrix(PTM)
PTM_heatmap <- heatmap.2(PTM_matrix[,6:8], Rowv=NA, Colv=NA, col=redgreen(75),
scale="none", key=TRUE, keysize=1, density.info="none", trace="none", cexRow=1.5,
cexCol=1.5, margins=c(6,4), ylab= "Peptide ID", main = "All PTMs in 3 HMT mutants")
```

3.7 References

[1] Jenuwein, T., Allis, C.D., Translating the Histone Code. *Science* 2001, 293, 1074–1080.

[2] Strahl, B.D., Allis, C.D., The language of covalent histone modifications. *Nature* 2000, 403, 41–45.

[3] Sidoli, S., Cheng, L., Jensen, O.N., Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *Journal of proteomics* 2012, Epub ahead.

[4] Roth, S.Y., Denu, J.M., Allis, C.D., Histone acetyltransferases. *Annual review of biochemistry* 2001, 70, 81–120.

[5] Martinez-Garcia, E., Licht, J.D., Deregulation of H3K27 methylation in cancer. *Nature genetics* 2010, 42, 100–1.

[6] Klose, R.J., Zhang, Y., Regulation of histone methylation by demethylimination and demethylation. *Nature Reviews Molecular Cell Biology* 2007, 8, 307–18.

[7] Zhang, Y., Reinberg, D., Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & Development* 2001, 2343–2360.

[8] Chi, P., Allis, C.D., Wang, G.G., Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nature Reviews Cancer* 2010, 10, 457–69.

[9] Bannister, A.J., Kouzarides, T., Reversing histone methylation. *Nature* 2005, 436, 1103–1106.

[10] Mosammaparast, N., Shi, Y., Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annual review of biochemistry* 2010, 79, 155–79.

[11] Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., et al., Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000, 406, 593–9.

[12] Greer, E.L., Shi, Y., Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics* 2012, 13, 343–57.

[13] Cao, R., Zhang, Y., The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Current opinion in genetics & development* 2004, 14, 155–64.

[14] Cao, R., Wang, L., Wang, H., Xia, L., et al., Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 2002, 298, 1039–43.

[15] Jacob, Y., Stroud, H., Leblanc, C., Feng, S., et al., Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* 2010, 466, 987–91.

[16] Yuan, W., Wu, T., Fu, H., Dai, C., et al., Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* 2012, 337, 971–5.

[17] Czermin, B., Melfi, R., McCabe, D., Seitz, V., et al., Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* 2002, 111, 185–96.

[18] Jacob, Y., Feng, S., LeBlanc, C. a, Bernatavichute, Y. V, et al., ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nature structural & molecular biology* 2009, 16, 763–8.

[19] Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., et al., RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. *Genes & Development* 2007, 21, 1530–1545.

[20] Chung, P.-H., Yao, M.-C., Tetrahymena JMJD3 homolog regulates H3K27 methylation and nuclear differentiation. *Eukaryotic cell* 2012, Epub ahead of print.

[21] Zhang, C., Molascon, A.J., Gao, S., Liu, Y., Andrews, P.C., Quantitative proteomics reveals that the specific methyltransferases Txr1p and Ezl2p differentially affect the mono-, di- and tri-methylation states of histone H3 lysine 27. *Molecular & cellular proteomics* n.d., Epub ahead of print.

[22] Cassidy-Hanley, D., Bowen, J., Lee, J.H., Cole, E., et al., Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment. *Genetics* 1997, 146, 135–47.

[23] Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S. a, et al., Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature protocols* 2007, 2, 933–8.

[24] Zhang, L., Su, X., Liu, S., Knapp, A.R., et al., Histone H4 N-Terminal Acetylation in Kasumi-1 Cells Treated with Depsipeptide Determined by Acetic Acid-Urea Polyacrylamide Gel Electrophoresis , Amino Acid Coded Mass Tagging , and Mass Spectrometry research articles. *Journal of Proteome Research* 2007, 6, 81–88.

[25] Candiano, G., Bruschi, M., Musante, L., Santucci, L., et al., Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* 2004, 25, 1327–33.

[26] Grunstein, M., Histone acetylation in chromatin structure and transcription. *Nature* 1997, 389, 349–52.

[27] Shahbazian, M.D., Grunstein, M., Functions of site-specific histone acetylation and deacetylation. *Annual review of biochemistry* 2007, 76, 75–100.

[28] Chahal, S.., Matthew, H.., Bradbury, E.., Acetylation of histone H4 and its role in chromatin structure and function. *Nature* 1980, 287, 76–9.

[29] Vettese-Dadey, M., Grant, P. a, Hebbes, T.R., Crane- Robinson, C., et al., Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *The EMBO journal* 1996, 15, 2508–18.

[30] Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., et al., Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 2006, 311, 844–7.

[31] Wyrick, J.J., Parra, M. a, The role of histone H2A and H2B post-translational modifications in transcription: a genomic perspective. *Biochimica et biophysica acta* 2009, 1789, 37–44.

[32] Bonenfant, D., Coulot, M., Towbin, H., Schindler, P., Van Oostrum, J., Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry. *Molecular & cellular proteomics* 2006, 5, 541–52.

[33] Zhang, Y., Griffin, K., Mondal, N., Parvin, J.D., Phosphorylation of Histone H2A Inhibits Transcription on Chromatin Templates. *The Journal of biological chemistry* 2004, 279, 21866–21872.

[34] Loury, R., Histone phosphorylation: how to proceed. *Methods* 2003, 31, 40–48.

[35] Oki, M., Aihara, H., Ito, T., Role of histone phosphorylation in chromatin dynamics and its implications in diseases. *Subcellular Biochemistry* 2007, 41, 319–336.

[36] Olsen, J. V, Vermeulen, M., Santamaria, A., Kumar, C., et al., Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science signaling* 2010, 3, ra3.

[37] Van Noort, V., Seebacher, J., Bader, S., Mohammed, S., et al., Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Molecular systems biology* 2012, 8, 571.

[38] Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., et al., Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* 2009, 138, 1122–36.

[39] Wisniewski, J.R., Zougman, A., Mann, M., Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. *Nucleic acids research* 2008, 36, 570–7.

[40] Jiang, T., Zhou, X., Taghizadeh, K., Dong, M., Dedon, P.C., N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proceedings of the National Academy of Sciences of the United States of America* 2007, 104, 60–5.

[41] Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., et al., RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. *Genes & Development* 2007, 1530–1545.

[42] Jacob, Y., Stroud, H., Leblanc, C., Feng, S., et al., Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* 2010, 466, 987–91.

[43] Chung, P.-H., Yao, M.-C., Tetrahymena thermophila JMJD3 homolog regulates H3K27 methylation and nuclear differentiation. *Eukaryotic cell* 2012, 11, 601–14.

[44] Wu, S.C., Kallin, E.M., Zhang, Y., Role of H3K27 methylation in the regulation of lncRNA expression. *Cell research* 2010, 20, 1109–16.

CHAPTER 4

Systematic analysis of histone modifications across 15 epigenetic features

4.1 Summary

Histones, especially their N-terminal domains, are adorned with the numerous epigenetic marks also known as post-translational modifications (PTMs). Recent studies indicate that the crosstalk between these different types of posttranslational modifications is key to systematically understand the combinatorial histone code. Here we performed quantitative proteomics analysis of histone PTMs across 15 epigenetic features (growing, conjugating, or starving cells from WT and 7 histone modifying enzyme knockout cells) in the model organism, *Tetrahymena thermophila*, to reveal global relationships in PTM patterns in the core histones based on my newly developed ^{15}N metabolic labeling method. In total, we quantified 53 combinations of histone PTMs and their unmodified counterparts, which covers 40 PTMs in the core histones and their variants. Multivariate statistical analysis of histone PTM data revealed 5 functionally-related subgroups of these chemical marks in response to the changes in epigenetic, physiological, or nutritional conditions. We also proposed a 5-factor model which reveals the hidden features underlying the perturbed chromatin profiles suggesting that PTMs should not be simply classified as a binary code (‘active’ or ‘repressive’ mark) as proposed in the traditional

model. This study greatly expands our understanding on the 'histone code' and provides a new way to think about chromatin modifications and classifications.

4.2 Introduction

A large number of proteins, especially in eukaryotic cells, undergo various site-specific covalent modifications. These posttranslational modifications (PTM) are widely spread over the whole cellular proteome often act in concert to control biogical outcomes in the cell[1]. Histones, in particular, are a protein class whose numerous epigenetic marks, especially in their N-terminal domains, are thought to be cross-regulated in a combinatorial way as proposed in the histone code hypothesis[2,3]. These different chemical modifications are written or erased by histone modifying enzymes sometimes referred to as histone code writers or erasers, and subsequently recognized by effectors or chromatin binding proteins acting, in effect, as histone code readers[4].

A growing number of examples of PTM crosstalk has been found in both histone and non-histone proteins[1,5–7]. Some residues like lysine are highly modified with more than two different types of PTMs (e.g., acetylation, formylation, methylation, ubiquitinylation at lysyl residues) at the same site leading to mutually exclusive modifications at each lysyl residue crosstalk[1,5]. A well-studied, non-histone, example is the homologous crosstalk between phosphorylation and O-GlcNAcylation which is critical for many cellular functions[8]. This is one example of two different PTMs affecting the same sites, however, crosstalk is not just limited to direct interactions between two or more PTMs at the same site. Crosstalk can be indirect or heterologous, that is, two different PTMs may affect one another even as they affect different sites within a protein or even across different target proteins[1]. For example, phosphorylation

and acetylation have been confirmed to affect one another reciprocally in bacteria[9]. Lysine acetylation in vertebrates, as another example, has the potential for 'crosstalk' with many other different types of PTMs such as methylation, phosphorylation, sumoylation, and ubiquitination homologously or heterologously[10,11].

Chromatin crosstalk across histones is currently a very active field. The numerous chemical modifications of these important proteins can affect each other either positively or negatively via several mechanisms[1,5–7,12]. Chromatin *cis* crosstalk can exist at the level of a single histone tail. For example, acetylation at H3K18 and K23 promotes the methylation of H3R17 by the CARM1 methyl-transferase in a human cell line[13]. Crosstalk may also present in *trans* actions where PTMs on one histone affect the modifications on other histones. Well-known examples include: 1) phosphorylation at H3S10 promotes H4K16 acetylation and coordinately mediates transcription elongation[14]; 2) H2B monoubiquitination is required for H3 methylation at multiple sites[15,16]. Sometimes, crosstalk may even go beyond histones or proteins. Examples are the cross-talk between specific DNA elements and histones[7].

Although our understanding of cross-regulation of histone modifications is becoming increasingly clear, the limited examples of crosstalk still hamper us in systematically revealing and modeling the global interactions and regulations underlying the dynamic chromatin structure that is thought to be necessary to elucidate the entire histone code. Using our previously developed ^{15}N metabolic labeling method[17], we were able to quantify most of the epigenetically important histone modifications in order to search for functionally related clusters and reveal the underlying features that dictate the levels of

these chemical marks by perturbation of chromatin modification patterns across 15 epigenetic features.

4.3 Experimental procedures

Most of the methods used in this study were already described in the previous chapters. Histone modifying enzyme knockout strains were constructed from wild-type *Tetrahymena thermophila* and provided by Dr. Liu's lab at UM pathology department. Briefly, ΔTXR1, ΔELZ2 and ΔTXR1:EZL2 are HMT mutants; JMJ1 and JMJ2 are two hypothetical HDMTs while RIN1 is a hypothetical ubiquitin E3 ligase as suggested from sequence comparison with their known corresponding genes in other species. Their potential modification sites are listed table S5.1. All knockouts were confirmed and verified by quantitative PCR.

4.3.1 Cell culture, ^{15}N metabolic labeling, and pure histone preparation

All media, procedures, and protocols used for cell growth, ^{15}N uniform Labeling, and histone separation and purification were performed as previously chapters except for different growing conditions[17]. For preparing growing cells, all strains were grown in 1× SPP medium at 30°C with gentle shaking, stopped at logarithmic-phase (2×10^5/ml), and were collected with centrifugation; Cell starvation was performed according to the procedure by Allis *et al*[18]. Briefly, growing cells were collected and washed once with 10mM Tris, pH 7.4. Cells were then starved in Tris buffer for 12h before they were collected for the subsequent experiments; The procedure for cell conjugation was adopted from Pfeffer *et al*[19]. In more details, growing cells were collected by centrifugation, washed three times with 10mM Tris, pH 7.4 and finally resuspended and diluted in the same buffer with an optical density of 0.2 at 600 nm. After ~24 h starvation at 30°C,

equal volumes of the two strains were mixed for appropriate time until more than 60% conjugated cells were visualized under a microscopy.

Crude histones were acid-extracted and prepared from macronuclei prior to HPLC purification. A C8 reversed-phase column was used to separate individual histones. LC fractions were evaluated by 15% SDS-PAGE and same histone fractions were subsequently combined. Concentrations of core histones were determined by the Bradford method. The general strategy and experimental design were already proposed in the chapter 3.

4.3.2 Quantitative analysis of histone PTMs by mass spectrometry

Histone samples were analyzed in biological duplicates or triplicates. Detailed sample information is listed in the table 5.1. Propionylation, trypsinization, and nanoLC-MS analysis of histone samples were all done as previously[17,20]. Raw data processing, database searching, and peptide quantification were all performed in Mascot Distiller (Matrix Sciences, Version 2.4 for distiller and Version 2.2.07 for search engine). In performing the database searching, ion tolerance was 10ppm for MS1 and 0.8 Da for MS2. N-terminal propionylation was considered as fixed modification. Variable modifications were: Acetylation (Protein N-terminus, K), Methylation (Protein N-terminus, KR), Propionylation (K, unmodified or monomethylated). Five missed cleavages were allowed for trypsin digestion. The following peptides from core histones were selected for normalization based on their stable ratios and small variations in all strains: TASSKQVSR, GQASQDL, FLKHGR in H2A; IALESSKLVR, RTLSSR in H2B; FRPGTVALR, VTIMTKDMQLAR, YQKSTDLLIR in H3; and ISSFIYDDSR, RKTVTAMDVVYALKR in H4. We assumed that the levels of those peptides were

consistently unchanged and their ratios (L/H) were very close to 1 in all experiments. All spectra assigned with PTMs were manually validated based on the same criteria discussed in previous chapters[17].

4.3.3 Multivariate statistical analysis of histone PTM data

Heatmap, clustering analysis, and factor analysis (FA) were performed with routines written in the statistical programming language R (http://www.r-project.org/). All normalized PTM ratios were base-2 log transformed. For clustering analysis, the Partitioning Around Medoids (PAM) algorithm-a robust version of K-means clustering statistical technique or Ward's hierarchical clustering method was used to search for the functionally related histone modifications. More details are provided in the appendix-"R code". Factor analysis is a model based version of Principal Component Analysis whose essential purpose is to describe the relationships between variables based on a data covariance matrix. The primary concern in the FA model is whether the observed variables can be reduced to a lower number of unobserved underlying variables called common factors based on the data correlation structure. In detail, assuming X is a p-variate random vector and each observation satisfies the following equation: $X_j = \sum_{k=1}^{K} \lambda_{jk} F_k + \mu_j$, where λ's are factor loadings, F_k are common factors, and μ_j are errors. Alternatively in matrix notation, $X = \Lambda F + U$. In this model, F and U are independent and are multivariate normal both with expectations equal to zero. The number of factors was estimated by PCA screeplot according to the following rules: 1) number of eigenvalues greater than one; 2) % of variance explained by factors.

4.4 Results and discussion

4.4.1 Global analysis of histone PTMs across 15 epigenetic features

Overall, we quantified 53 peptides and 40 histone PTMs in tetrahymena core histones with comparable sequence coverages (H2A: 61%, H2B: 68%, H3: 76%, and H4: 97%) and False Discovery Rate (FDR<1%) observed in previous chapters after data normalization. The histone variants were found as follows: H2A: H2A-1, H2A.X; H2B: H2B-1, H2B-2; and H3: H3 major, H3.3, and H3.4. The detailed PTM information is provided in table 5.1. All biological and technical replicates were combined when we did database searching and peptide quantification in order to minimize the missing data. Thus, the final ratio is the average ratio of all replicates. Furthermore, PTMs were all averaged over the same combination found in multiple charge states, different propionylation degrees, or multiple peptides. Generally, the average ratio of these combinations in each strain is fairly close to zero after data normalization and log2 transformation showing the data processing and quantification is very sucessful (Figure 5.1a). Some samples are strongly positively correlated while others are less correlated as suggested from the bivariate scatterplot of histone PTM data (Figure 5.1b). The primary sites affected by EZL2p were H3K27Me2/3 while TXR1p primarily affected the monomethylation of H3K27. These results have been described in detail in the previous chapters. The JMJ1p and JMJ2p are two hypothetical HDMTs but their effects on histone methylation were minor (see Table 5.1).

4.4.2 Histone modifications and chromatin: binary classification is NOT enough

To explore the common factors between variables (15 epigenetic features) that dictate the levels of histone PTMs, we performed factor analysis of all 40 histone PTMs and their unmodified counterparts based on the data summarized in table 5.1. The ratios were

log2 transformed in the FA model. We first attempted to fit the data with a 2-factor model since chromatin has classically been classified as having two forms: euchromatin and heterochromatin. The euchromatin is generally associated with transcriptional activation while heterochromatin is considered to be transcriptionally silent[21]. The 2-factor model did not fit the data well as suggested from the R output shown in supplementary Figure S5.1. The 2-factor model only captured 44.6% of data variance (the variance shared with other variables via common factors) which is insufficient for approximating the covariance matrix. Many variables have large errors or uniquenesses (i.e. the variance not shared with other variables). Heatmap analysis of histone PTM data also didn't show a dichotomous mosaic pattern across all strains (Figure S5.2). These results suggested that strains or PTMs cannot be simply explained by a 2 factor model: repressive or active forms.

We performed a PCA screeplot of the data correlation matrix to determine how many factors are required for the FA model based on the following rules: 1) number of eigenvalues greater than one; 2) % of variance explained by first several factors. We found that a 5-factor model fit the data quite well: The model captured about 84.5% of the cumulative variance and is sufficient for data reduction, and most of the variables have fairly small uniqueness values (Figure 5.2). This is not to say that the five state model is a true representation of histone crosstalk, but based on our current data it provides the best fit and the actual number of features is likely to be higher. Factor loadings represent how well variables are correlated with each of the factors and those loadings with large numbers usually suggest meaningful interpretations. The first factor captures associations between growing strains (WT_G, RIN1_G, JMJ1_G, JMJ2_G, and

JMJ1:JMJ2_G) which may be explained as a factor for cell growth; the second factor is a mixture of starving and conjugating cells (WT_S, WT_C, RIN1_S, and JMJ2_C); The third factor captures the growing or starving ELZ mutants which likely reflects gene transcription based on our understanding of *Tetrahymena*; The fourth factor may represent DNA replication as it captures the growing or starving TXR1 mutants; The last factor is derived from 2 conjugating cells (JMJ1_C and JMJ1:JMJ2_C). The significance of the biological backgrounds are provided below (see Section 4.5).

4.4.3 Functionally-related histone modifications

The histone code model proposes that the dynamic chromatin structure and functions are dictated by various histone modifications via changing levels and patterns of these epigenetic marks. To identify functionally-related subgroups of PTMs, I performed clustering analysis of perturbed histone modifications in 15 epigenetically different strains. A top-down clustering method-Partitioning Around Medoids (PAM) algorithm and a bottom-up approach-Ward's hierarchical clustering algorithm were used to group PTMs that have similar biological functionality according to their levels under different conditions. Both methods partitioned the data into 5 clusters with relatively acceptable silhouette coefficients-a statistical method which measures cluster quality according to cluster homogeneity and separation (Figure 5.3). Here we list common PTMs and clusters identified by both methods (Figure 5.4). The first cluster is the N-terminal acetylation of H2A, H3 and H4 (Figure 5.4a). They were up-regulated in most of the growing cells and down-regulated in starvation and conjugation. Interestingly, changing levels of these PTMs were not always consistent in all strains. For example, the most up-regulated PTMs were H4:K4AcK7AcK11AcK15Ac and H3:K9AcK14Ac in ΔTXR1 growing cells

while the N-terminal triple- or tetra-acetylation in H2A had the largest ratio increases in ΔJMJ1:JMJ2 growing cells(Figure 5.4a). Taken together, these results indicate that changes in H3K27 methylation affects acetylation on histones H4 and H2A and that this crosstalk is modulated by other factors. We cannot determine from these experiments whether the coordinate changes observed are direct (i.e., demethylation being required for increased acetylation at these sites) or indirect as might be expected for a compensatory mechanism. The second major cluster reflects the unchanged levels of some PTMs and unmodified forms (Figure 5.4b). Most of the PTMs in this group are mono-acetylation or C-terminal acetylations suggesting that mono-acetylated or C-terminally acetylated histone has weaker effects on chromatin functions than the N-terminal or triple-/tetra-acetylated forms. This is consistent with the current understanding of histone modifications[22]. PTMs grouped in the third cluster are H2B-2:K3Me3K4Ac, H3:K23Me1, and H3.3:K9Ac but their biological roles are still unknown (Figure 5.4c). The fourth and fifth clusters include the methylation of K27 and K36 in H3 and its variant (Figure 5.4d and e). Monomethylation of H3K27 and K36 were elevated in ΔEZL2 cells and decreased in ΔTXR1 cells. Di- and tri-methylation of H3K27, however, significantly decreased in ΔEZL2 cells but increased in most starving and conjugating cells. As discussed in Chapters 2 and 3, TXR1p and ELZ2p are involved in two different pathways via regulation of H3K27 methylation. TXR1p monomethylates H3K27 and K36 while EZL2p is primarily respsible for di- and trimetylation of H3K27. Although both HMTs create a repressive chromatin form, the elevated di- and tri-methylated forms in both starving and conjugating cells implied that these marks cannot be simply classified as two-state marks as also suggested from the FA model.

4.5 Discussion and concluding remarks

The binary classification of chromatin as euchromatin and heterochromatin is a general classification of chromatin and has a long history[21]. The euchromatin is considered the transcriptionally-active form and the heterochromatin is the transcriptionally-silent form as proposed in the traditional chromatin model. This concept was recently challenged in a paper published in 2010[23] by Filion *et al.* which identified five principal chromatin classes that they coded in 5 colors by genome-wide mapping the distributions and locations of 53 chromatin-associated proteins in *Drosophila* cells[23]. In this model, the heterochromatin can be further dissected into three subtypes: two of them (designated as Green and Blue in their figures) are consistent with the 'classic' heterochromatin marked by heterochromatin protein 1 (HP1), HP1-interacting proteins, and polycomb group (PcG)-associated proteins. These genomic regions are rich in the histone modifications, H3K9me2/me3 and H3K27me3, which are well-studied repressive marks. A new subtype of heterochromatin (Black in their figures), however, surprisingly occupies nearly half of the non-repetitive genome containing >4,000 genes that might regulate cell development. The Yellow and Red chromatin in their model are two subtypes of euchromatin containing proteins and histone modifications that are typically associated with transcriptional activation. In another study, Kharchenko and his colleagues derived 9 combinatorial patterns in the same species[24]. A recent study performed by Ernst *et al.* used the same multivariate statistical model (Hidden Markov Model, HMM) identified 51 distinct chromatin states in human lymphocytes based on the maps of 38 core histone marks[25]. Four chromatin states were also observed in *Arabidopsis* based on the epigenomic maps for eight histone modifications[26]. The modENCODE project,

however, generated a more conservative, three combinatorial states based on genome-wide maps of 19 histone modifcations and 2 key histone variants in *C. elegans*[27]. All these studies suggested a more refined, but a limited number of principal chromatin types. The discrepancy of the number of subtypes in these various studies may be derived from differences between species[21], the particular parameters measured, or the statistical methods used. The bottom line is that multiple studies have demonstrated more that two chromatin states and there is good evidence for multiple subtypes.

In this study, we performed quantitative proteomics analysis of 40 histone modifications individually or combinatorially across 15 epigenetic features to generate a high-thoughput dataset reflecting the perturbed histone modification patterns. We assume that the knockout of key histone-modifying enzmes and changes in other key epigenetic features in response to growth conditions could lead to the changes in chromatin architecture both temporally and spatially, but the number of chromatin types is limited and is determined by certain key hidden features. Our FA model revealed that, at chromatin modification levels, PTMs are dominated by at least 5 chromatin states we have labeled as growth, replication, transcription, and two other mixed states. Therefore, the binary classification of histone modifications of chromatin, as proposed in traditional model, is insufficient to explain our results. We also identified five functionally-related PTM subgroups based on 2 distinct clustering methods. The results brought new specific insights on histone modifications:

1) Acetylations and particularly a high-degree of acetylation (triple- or tetra-acetylated forms) in the N-terminal domains play more important roles in

chromatin functions than the single or multiple acetylations at the histone C-termini.

2) Changes in levels and combinations of PTMs (i.e acetylations) within the same peptide can have different biological outcomes;

3) The methylation of H3K27 in different states represents two different repressive chromatin states as suggested from both a FA model and clustering analysis. This information is consistent with our knowledge in epigenetic studies that the regulation of H3K27 methylation has two major conserved pathways as intensively discussed in Chapter 3, Section 3.4.4. One is the E(Z)/EZH2-mediated pathway associated with transcription by controlling the di- and trimethylation levels of H3K27[28]. Monomethylation of H3K27 determined by ATXR5 and ATXR6 in plants, however, is involved in DNA replication-a novel biological pathway independent of E(Z)/EZH2-mediated transcription[29]. The homologous pathways responsible for H3K27 methylation were also identified in *Tetrahymena thermophila*, a more evolutionarily distant species (Unpublished data).

Our study represents the most comprehensive exploration of histone modifications in *Tetrahymena thermophila*, a model organism useful for histone studies. Compared to previous studies, we provide a complementary way to study histone modifications and chromatin features. We validated some of the key concepts as proposed in the 'histone code' hypothesis, verifying that the complexity of the code is achieved in a combinatorial and dynamic way. It is worthwhile to note that the number of epigenetic features identified in this paper may be underestimated as only a limited number of epigenetic

profiles was explored. The detailed downstream pathways and regulatory networks will be the major focus in our future studies.

Table 4.1 All PTM ratios across 15 epigenetic features

PepID	PTMs	WT_G n=3	WT_S n=2	WT_C n=2	EZL2_G n=3	EZL2_S n=2	RIN1_G n=3	RIN1_S n=2	TXR1_G n=3	TXR1_S n=2	JMJ1_G n=2	JMJ1_C n=2	JMJ2_G n=3	JMJ2_C n=2	J1+J2_G n=2	J1+J2_C n=2
1	H2A:K5AcK8AcK10AcK12Ac	0.62	0.65	0.69	1.68	0.45	1.16	0.72	1.58	0.73	1.56	0.87	1.38	0.13	2.42	0.84
2	H2A:K5AcK8AcK12Ac	0.76	0.78	0.79	1.45	0.6	0.94	1.08	1.26	0.89	1.09	0.83	1.31	0.23	1.8	0.88
3	H2A:K5AcK8Ac	0.7	0.88	0.93	1.04	0.72	0.93	0.91	0.95	0.91	0.95	0.98	1.06	0.53	1.08	1.02
4	H2A:K5AcK10Ac	0.77	0.88	0.94	1.05	0.77	0.93	0.93	0.97	0.97	0.98	1	0.9	0.57	1.12	1.04
5	H2A:K5Ac	1.11	0.98	1.05	0.81	1.13	0.94	1.02	0.82	1.08	0.96	1.05	1.04	1.35	0.81	1.06
6	H2A:K5UnK8UnK10UnK12Un	1.38	1.98	0.98	1.71	0.91	0.77	0.93	1.26	0.8	1.28	0.72	0.86	0.96	1.1	0.86
7	H2A:K8UnK10UnK12Un	1.35	1.45	1.03	1.32	1.01	0.99	1.08	1.62	1.07	1	1.1	1.02	1.2	0.89	1.07
8	H2A:K12Ac	0.7	1.17	1.24	1.3	0.81	0.99	0.94	1.38	0.97	1.05	1.02	1.13	0.87	1.18	1.47
9	H2A:K12Un	1.12	1.04	0.97	1.01	1.02	0.98	1.04	1.13	0.99	0.97	0.95	1.04	1.17	1.03	1.02
10	H2A:R92Me1	1.03	1.09	0.99	0.95	0.94	1.03	0.96	0.99	1.05	0.97	1	1.01	1.06	0.98	1.05
11	H2A:R92Un	0.99	0.96	0.98	0.96	0.98	0.92	0.94	1	1	1	1	0.99	1.05	0.94	1.06
12	H2B1:K111AcK115Ac	1.29	1.26	1.32	1.05	1.03	1.08	1.04	1.45	1.09	1.27	1.27	1.06	1.16	1.1	1.11
13	H2B1:K115Ac	0.98	0.96	0.96	0.98	0.96	1.02	0.86	1.19	0.94	1.3	1.05	1.04	0.95	1.01	0.91
14	H2B1:K115Un	1.04	1.05	1.18	0.96	1.02	1.07	0.98	1.27	1	1.15	0.99	0.97	1.05	1.09	1.04
15	H2B2:K3Me3K4Ac	0.21	0.68	0.28	0.25	0.49	0.43	0.53	0.55	0.42	0.39	0.72	0.4	0.24	0.45	0.41
16	H2B2:K3Me3	0.38	0.96	1.19	0.42	1.25	0.8	1.04	0.54	0.82	0.75	1.1	0.72	1.31	0.62	0.82
17	H2B2:K111AcK115Ac	1.12	1	0.97	1.01	1.05	1.1	0.87	1.34	0.91	1.02	1.01	1.05	0.97	0.96	0.9
18	H2B2:K115Un	0.98	0.92	0.96	1.06	1	1.12	0.84	1.13	0.96	1.1	0.99	1.04	0.96	1.04	1.01
19	H3:K9AcK14Ac	1.28	0.96	0.84	1.75	0.7	1.69	0.78	2.31	1.42	1.36	1.41	1.13	0.3	1.45	1.17
20	H3:K14Ac	1.15	1	0.84	1.21	0.79	1.18	0.92	1.33	1.27	1.16	0.99	0.79	0.58	1.1	0.83
21	H3:K9UnK14Un	0.84	1.25	1.02	0.83	1.12	0.86	1	0.71	1.02	0.84	1.03	0.78	1.11	0.82	0.95
22	H3:K18AcK23Ac	1	0.8	0.52	1.19	0.76	1.01	0.77	1.14	1.02	0.84	0.9	0.99	0.32	0.98	0.8
23	H3:K23Ac	0.84	0.85	0.85	0.91	0.73	0.79	0.8	0.88	0.82	0.81	1.03	0.73	0.79	0.8	0.99
24	H3:K23Me1	0.36	0.9	1.05	0.66	1.2	0.55	0.73	0.89	1.43	0.38	1.26	0.31	0.79	1.06	0.89
25	H3:K18UnK23Un	1.12	1.25	1.12	1.02	1.17	1.04	1.11	1.05	1.04	1.15	1.11	0.98	1.25	1.09	0.91
26	H3:K27Ac	0.81	0.53	0.41	1.13	0.66	0.76	0.76	1.05	1.14	0.58	0.61	0.42	0.16	0.62	0.25
27	H3:K27Me1	1.08	1.17	1.12	2.1	2.06	1.2	1.04	0.16	0.32	1.11	1.07	0.99	1.03	1.26	1.02
28	H3:K27Me1K36Me1	0.73	1.07	1.2	0.96	1.65	0.72	0.99	0.07	0.15	0.61	1.06	0.45	0.78	0.55	0.77
29	H3:K27Me2	0.79	1.2	1.15	0.07	0.08	0.76	1.02	0.39	0.5	0.77	1.04	0.62	0.94	0.77	0.92
30	H3:K27Me3	0.55	1.09	0.86	0.01	0.02	0.49	0.78	0.42	0.78	0.46	1.07	0.52	0.92	0.51	1.22

31	H3:K27UnK36UnK40Un	1.77	1.51	1.01	1.93	1.91	2.55	1.44	6.6	5.2	2.33	0.88	2.34	0.66	2.45	0.53
32	H3:K79Me1	1.1	1.15	0.96	1.07	1.06	0.97	0.99	1.05	1.04	1.02	1.01	1.1	1.11	1.08	0.97
33	H3:K79Un	1.12	1.07	0.96	1.05	0.99	1.06	1.01	1.05	1.02	1.03	1.04	0.95	0.98	1.05	1
34	H3.3:K9AcK14Ac	0.81	1.51	1.01	0.9	1.14	0.84	1.06	0.77	1.06	0.86	1.13	1.07	1.28	0.88	1.06
35	H3.3:K9Ac	0.84	0.71	0.9	1.94	0.83	0.42	0.76	2.83	2.13	0.39	1.07	0.32	0.44	0.64	0.92
36	H3.3:K9UnK14Un	0.64	1.03	1.03	0.54	1.25	0.46	0.75	0.61	1.13	0.35	0.94	0.39	0.87	0.34	0.84
37	H3.3:K18AcK23Ac	0.8	1.04	0.93	0.93	0.77	0.85	0.8	0.91	0.82	0.82	1.1	0.75	0.78	0.81	0.99
38	H3.3:K23Ac	1.1	1.25	1.18	0.97	1.18	1.26	1.11	1.01	1.16	1.11	1.08	0.98	1.15	1.09	0.91
39	H3.3:K18UnK23Un	0.43	0.91	0.98	0.68	1.06	0.46	0.67	0.87	1.3	0.31	1.17	0.32	0.68	0.32	0.77
40	H3.3:K27Me1	0.79	0.7	1.41	1.44	1.56	0.54	0.83	0.21	0.38	0.41	1.07	0.49	0.77	0.59	1.03
41	H3.3:K27Un	0.74	1.01	1.33	0.9	1.04	0.7	1.85	1.82	4.79	0.82	0.92	1.24	0.68	0.61	0.84
42	H3.3:K56Ac	1.03	1.13	1.02	0.99	1.01	0.98	1.01	0.96	1.04	1.04	1.03	0.88	0.91	1.05	1
43	H3.3:K56Un	0.53	0.77	1.08	1	1.2	0.64	0.75	0.93	1.51	0.33	0.89	0.45	0.7	0.5	0.79
44	H4:K4AcK7AcK11AcK15Ac	0.43	0.57	0.44	1.13	0.82	1.37	0.6	2.47	0.81	0.87	2.3	0.99	0.25	1.15	1.53
45	H4:K4AcK7AcK11Ac	0.51	0.66	0.58	0.82	0.76	1	0.6	1.29	0.76	0.73	1.43	0.75	0.29	0.77	1.24
46	H4:K4AcK7Ac	0.53	0.76	0.63	0.75	0.72	0.85	0.69	0.85	0.71	0.69	1.12	0.68	0.39	0.66	1.03
47	H4:K7Ac	0.97	0.94	0.78	1	0.94	0.97	0.89	1.09	0.98	1.03	0.97	0.96	0.8	0.88	0.96
48	H4:K4UnK7UnK11UnK15Un	2.02	1.33	1.06	1.39	1.54	1.25	1.34	1.17	1.5	1.59	0.7	1.5	2.33	1.43	0.99
49	H4:K7UnK11UnK15Un	3.02	1.4	1.12	1.51	1.55	1.22	1.25	1.17	1.75	1.63	0.7	1.48	2.36	1.46	0.92
50	H4:K20Me1	1.07	1.05	1.04	1.22	1.11	0.95	1.13	1.1	1.07	1.04	0.97	0.91	1.04	1.05	0.91
51	H4:K20UnK23UnK31Un	1.04	0.99	1	1.14	1.04	0.99	0.95	1.04	1.02	1.02	1.04	0.98	1.01	1.03	1.04
52	H4:R77Me1	1.08	1.06	1	1.09	1.16	1.01	1.05	1.06	1.11	1.04	1.06	1.03	1.09	1.14	1.09
53	H4:R77Un	0.98	0.94	0.94	0.98	0.97	0.91	0.9	0.97	0.98	0.96	0.95	0.91	0.97	0.96	0.98

Note:Ratios were averaged over different charge states, propionylation degrees, and tryptic cleavage sites.
n: biological replicates; G: Growing cells, S: Starving cells, C: Conjugating cells; J1+J2: double knockout of JMJ1 and JMJ2

Table S4.1 Genes and their potential biological roles in chromatin functions

Histone lysine methyl-transferases		
Genes	Sites affected	Function
TXR1	H3K27me1	Replication
EZL2	H3K27me2/me3	Transcription repression
Histone lysine demethylases		
Genes	Sites affected	Function
JMJ1	H3K27	heterochromatin formation
JMJ2	H3K9/H3K36?	heterochromatin formation?
Histone ubiquitin E3 ligase		
Genes	Sites affected	Function
RIN1	H2A K123ub1?	transcription repression

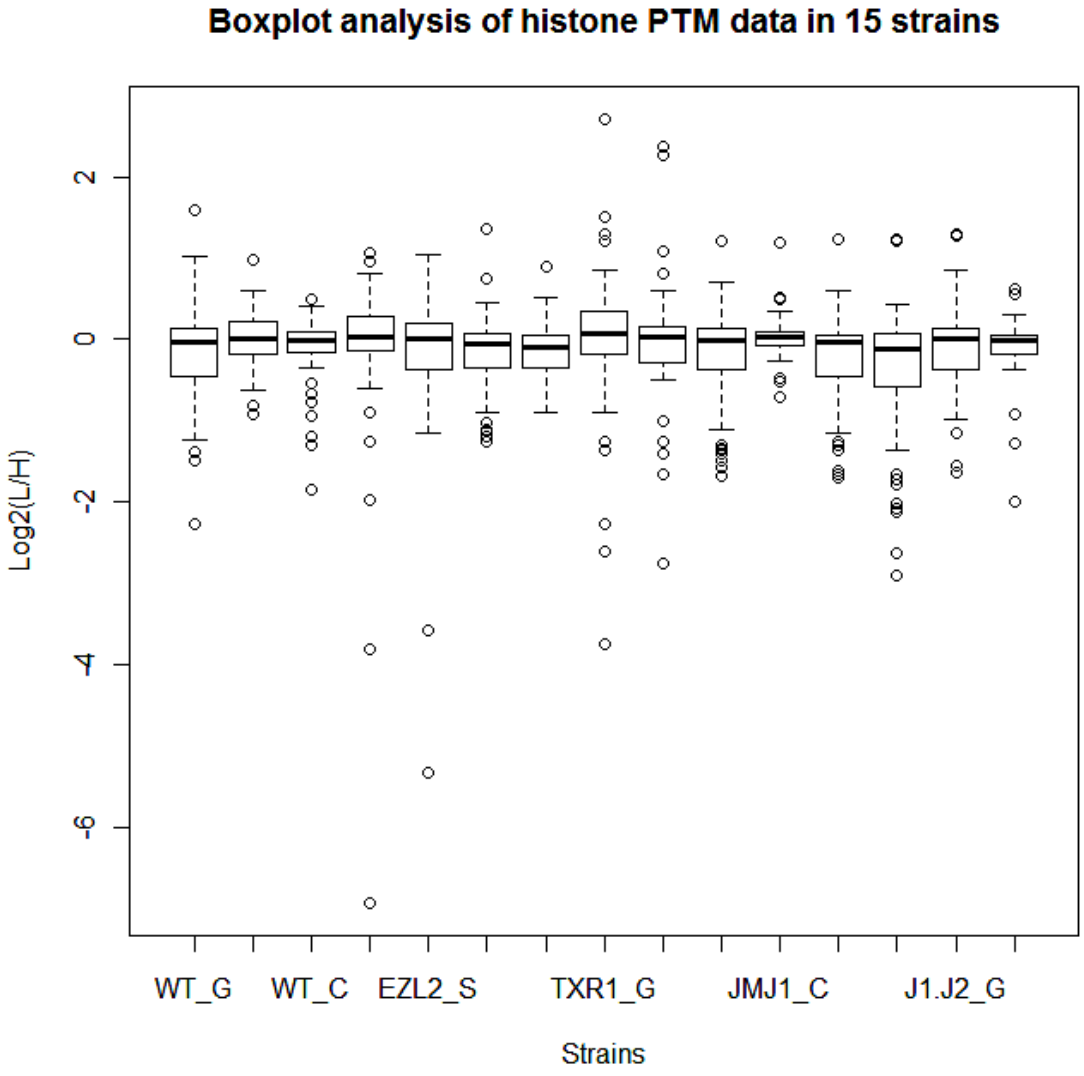

Boxplot analysis of histone PTM data in 15 strains
Log2(L/H)
2
0
-2
-4
-6
WT_G
WT_C
EZL2_S
TXR1_G
JMJ1_C
J1.J2_G
Strains

Figure 4.1a

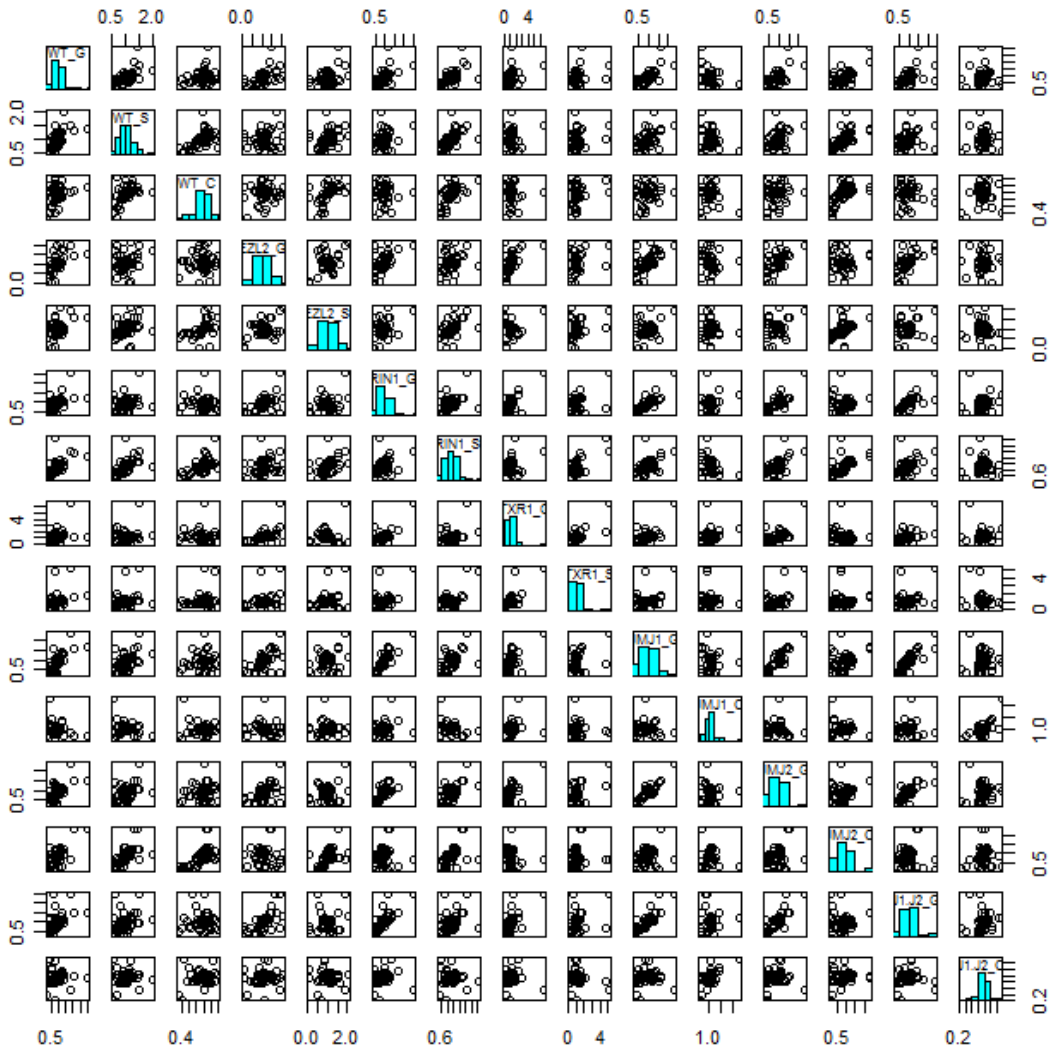


Figure 4.1b

Figure 4.1 Boxplot and bivariate scatterplot of PMT data

a)Boxplot analysis of histone PTM data across 15 epigenetic features. The average peptide log2 ratio in all strains is very close to one. Outliers are shown in small circles; b)Bivariate scatterplot of Histone PTM data.

```
> ptm.FA

Call:
factanal(x = mydata, factors = 5, scores = "Bartlett", rotation = "promax")

Uniquenesses:
   WT_G    WT_S    WT_C  EZL2_G  EZL2_S  RIN1_G  RIN1_S  TXR1_G  TXR1_S  JMJ1_G  JMJ1_C  JMJ2_G  JMJ2_C
  0.295   0.318   0.223   0.005   0.158   0.168   0.191   0.088   0.005   0.005   0.453   0.106   0.189
J1.J2_G J1.J2_C
  0.198   0.005
```

'Growth' 'Transcription' 'Replication'

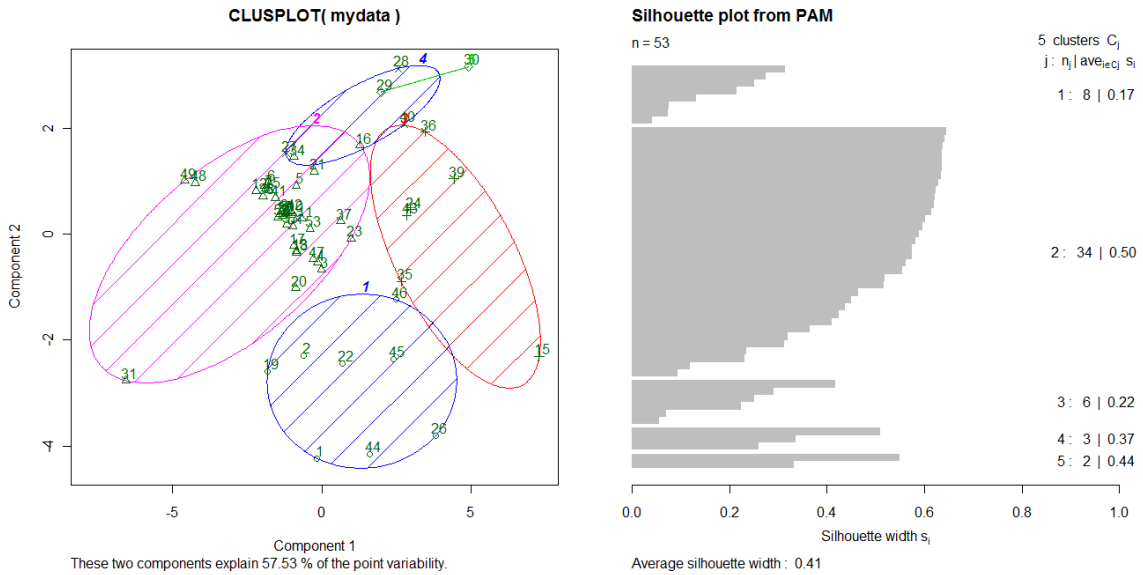
```
Loadings:
        Factor1 Factor2 Factor3 Factor4 Factor5
WT_G     0.501   0.423   0.144
WT_S     0.223   0.748  -0.141
WT_C    -0.218   0.893   0.141           0.245
EZL2_G   0.161  -0.161   0.952
EZL2_S  -0.153   0.181   0.917
RIN1_G   0.936
RIN1_S   0.225   0.726           0.150  -0.174
TXR1_G   0.237  -0.272           0.830
TXR1_S  -0.209   0.204           1.042
JMJ1_G   1.032
JMJ1_C          -0.118                   0.721
JMJ2_G   0.908
JMJ2_C  -0.120   0.939                   0.138
J1.J2_G  0.922  -0.154
J1.J2_C  0.143   0.283                   0.952

               Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings      4.204   3.258   1.824   1.817   1.570
Proportion Var   0.280   0.217   0.122   0.121   0.105
Cumulative Var   0.280   0.497   0.619   0.740   0.845
```

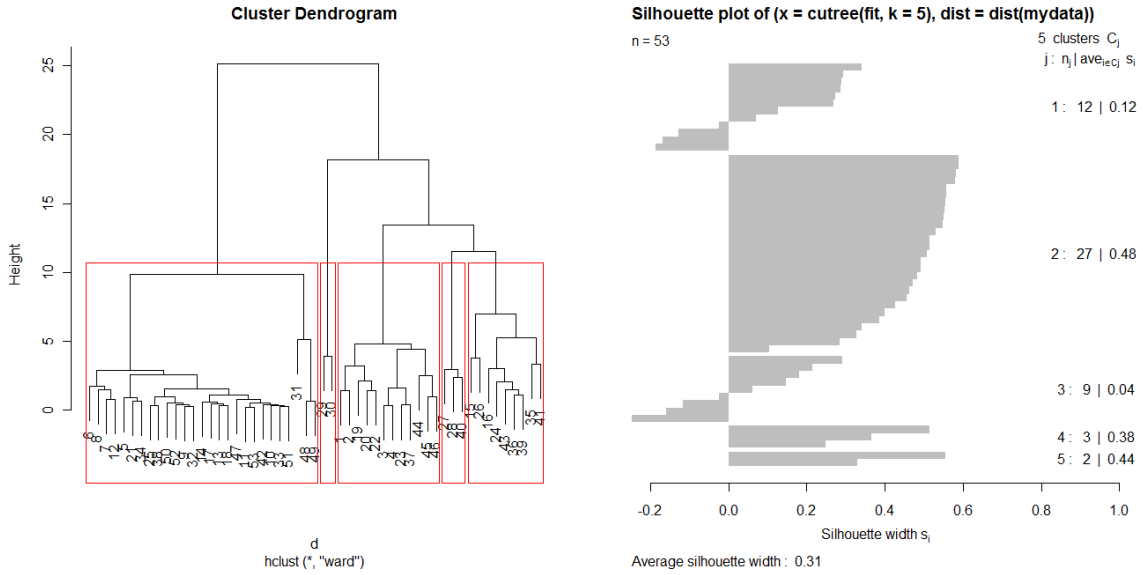
Figure 4.2 Output of 5-factor model from R software

The first 5 factors accounts up to 84.5% of variance and most of factors has small uniqueness. The loadings with large numbers are identified with red oval circles.

Note: Factor loadings are very similar to regrssion coefficents in Generalized Linear Model. They represent how well variables are correlated with each of the factors. The loadings with large numbers usually provide meaningful interpretations of factors.



a



b

Figure 4.3 Clustering analysis of PTM data

a)PAM Clustering; b) Ward's hierarchical clustering. Dissimiliarity matrix as input in PAM and euclidean distance measured in hierarchical clustering.

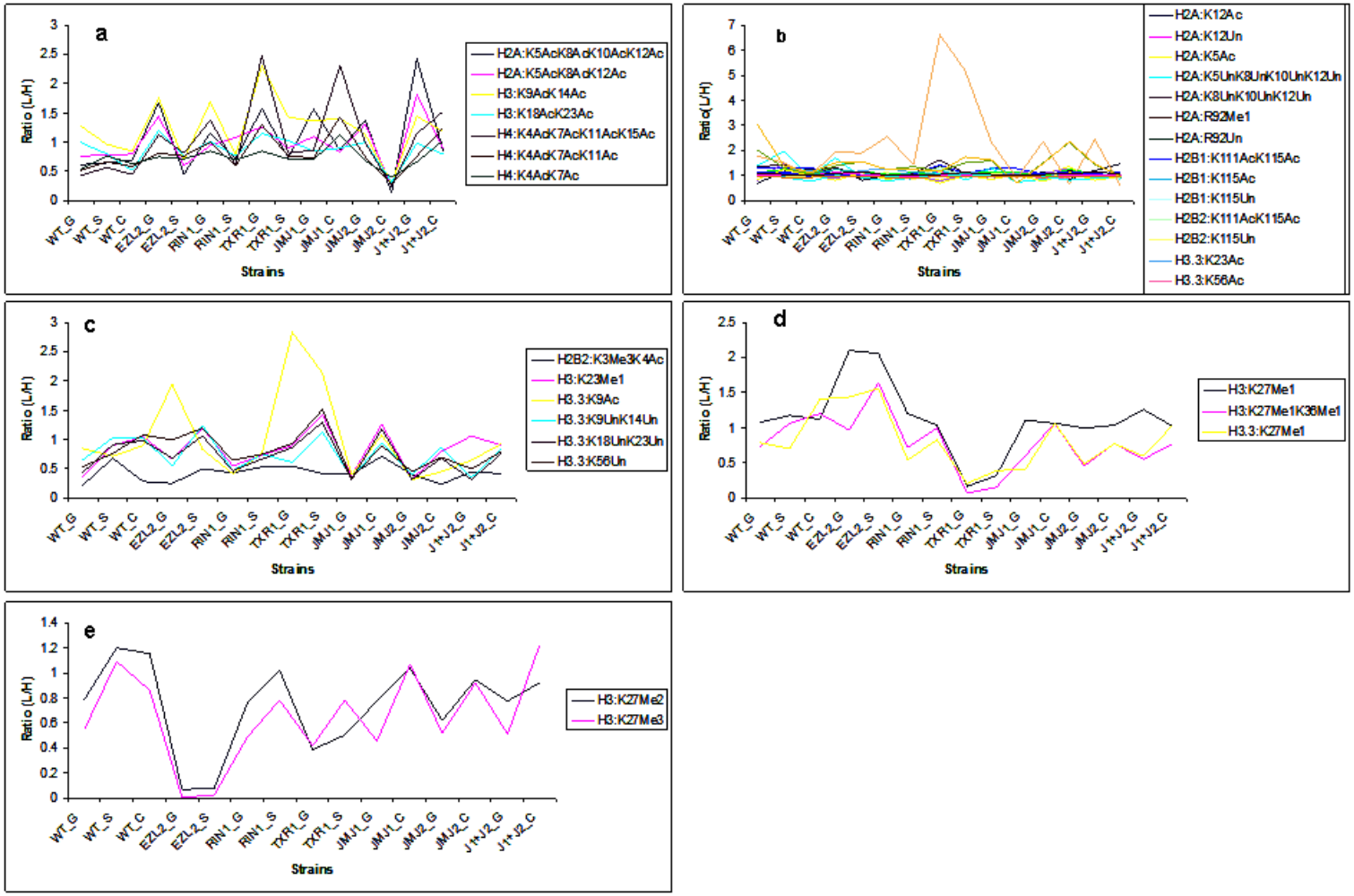


Figure 4.4 Five functionally-related histone PTM subgroups

The common PTMs and clusters identifed by both PAM and Ward's hierarchical clustering algorithm. a) Cluster 1: N-terminal acetylation of H2A, H3 and H4; b) Cluster 2: Some of unchanged levels of PTMs and unmodified forms; c) Cluster 3: PTMs in this group are H2B-2:K3Me3K4Ac, H3:K23Me1, and H3.3:K9Ac; Cluster 4: Monomethylation of H3K27 and K36; and e) Cluster 5: Di-/tri-methylation of K27.

```
> ptm.FA

Call:
factanal(x = mydata, factors = 2, scores = "Bartlett", rotation = "promax")

Uniquenesses:
   WT_G    WT_S    WT_C  EZL2_G  EZL2_S  RIN1_G  RIN1_S  TXR1_G  TXR1_S  JMJ1_G  JMJ1_C  JMJ2_G  JMJ2_C
  0.458   0.748   0.961   0.005   0.273   0.179   0.673   0.778   0.923   0.017   0.991   0.117   0.922
J1.J2_G J1.J2_C
  0.231   0.934

Loadings:
       Factor1 Factor2
WT_G     0.682   0.133
WT_S     0.529  -0.209
WT_C     0.192
EZL2_G           0.961
EZL2_S           0.871
RIN1_G   0.872
RIN1_S   0.585
TXR1_G   0.371   0.194
TXR1_S   0.213   0.123
JMJ1_G   0.993
JMJ1_C
JMJ2_G   0.946
JMJ2_C   0.281  -0.177
J1.J2_G  0.831   0.120
J1.J2_C  0.267  -0.132

               Factor1 Factor2
SS loadings      4.810   1.873
Proportion Var   0.321   0.125
Cumulative Var   0.321   0.446
```

Figure S4.1 Output of 2-factor model from R software

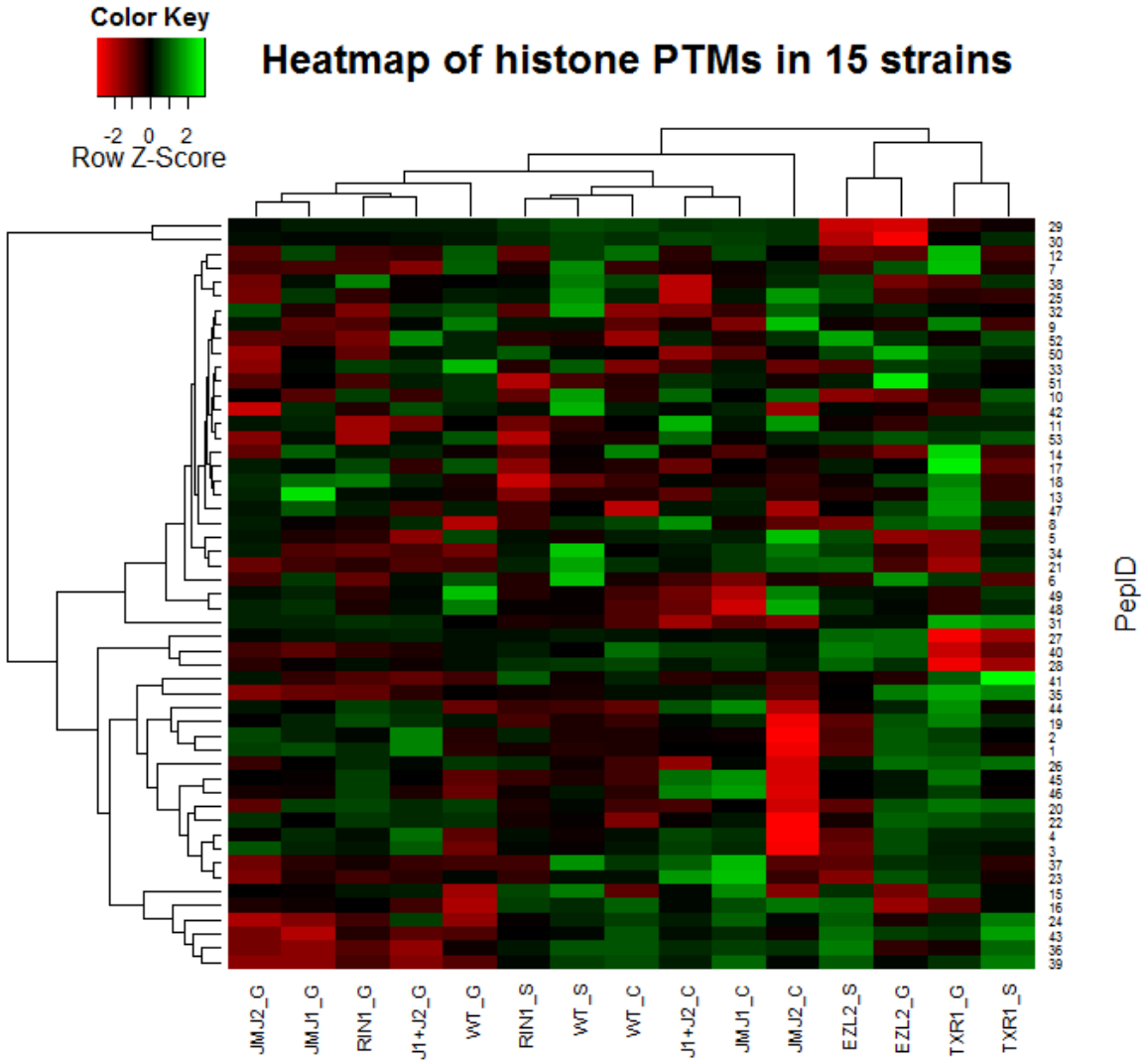


Figure S4.2 Heatmap analysis of all histone modifications in 15 strains

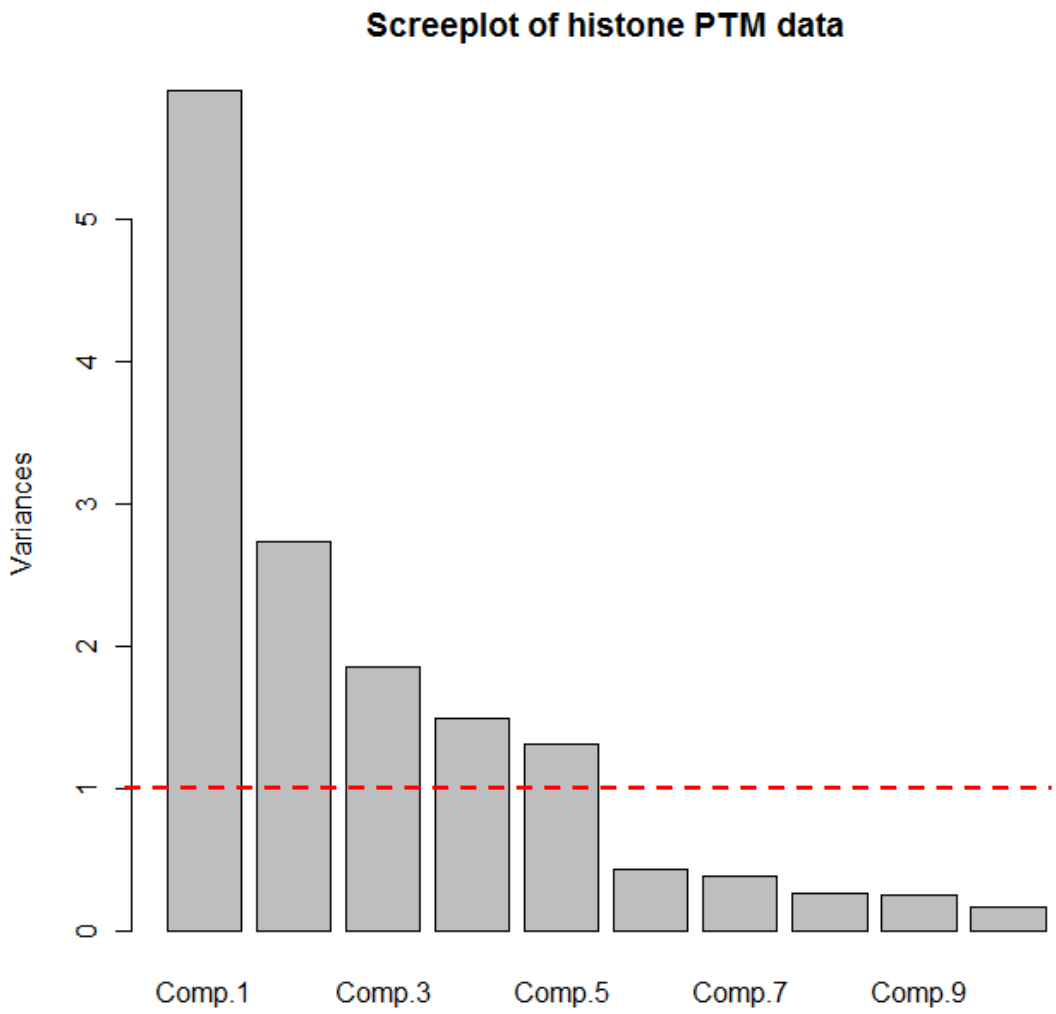


Figure S4.3 Choosing number of factors by screeplot
The number of factors were determined by PCA screeplot based on the rules: 1) number of eigenvalues> 1; 2) % of variance explained by first several factors.

4.6 Appdendix-R code for heatmap, clustering and factor analysis

```
# Data input
PTM=read.table("ALLPTMLOGTRANS.txt",header=T)
print(summary(PTM),digits=2)

# Bivariate scatterplot matrix of the Histone PTM data
panel.hist <- function(x, ...)
{
   usr <- par("usr"); on.exit(par(usr))
   par(usr = c(usr[1:2], 0, 1.5) )
   h <- hist(x, plot = FALSE)
   breaks <- h$breaks; nB <- length(breaks)
   y <- h$counts; y <- y/max(y)
   rect(breaks[-nB], 0, breaks[-1], y, col="cyan", ...)
}
pairs(PTM[,3:17], diag.panel=panel.hist)

#Boxplot of PTM data
boxplot(PTM[,3:17], cexCol=0.8, xlab="Strains", ylab="Log2(L/H)", main = "Boxplot
analysis of histone PTM data in 15 strains")

#Heatmap
install.packages("gplots")
install.packages("RColorBrewer")
library(RColorBrewer)
library(gplots)
PTM=read.table("ALLPTMLOGTRANS.txt",header=T)
PTM_matrix <- data.matrix(PTM[,3:17])
PTM_heatmap <- heatmap.2(PTM_matrix, Rowv=T, Colv=T, labCol = colnames(PTM),
col=redgreen(75), scale="row", key=TRUE, keysize=1, density.info="none",
trace="none", cexRow=0.6, cexCol=0.8, margins=c(6,4), ylab="PepID", main = "
Heatmap of histone PTMs in 15 strains")

#Clustering
PTM=read.table("ALLPTMLOGTRANS.txt",header=T)
mydata=PTM[,3:17]
library(stats)
library(cluster)

#PAM clustering
pam.ptm = pam(mydata,k=5,diss=F)
# The silhouette plot
diss.ptm = daisy(mydata)
sil.pam = silhouette(pam.ptm$cluster, diss.ptm)
plot(sil.pam, main = 'Silhouette plot from PAM')
```

```
#Projection onto the first two PCs
clusplot(mydata, pam.ptm$cluster, color=T, shade=T, labels=2, lines=0)

# Ward's hierarchical clustering
library(stats)
library(cluster)
PTM=read.table("ALLPTMLOGTRANS.txt",header=T)
mydata=PTM[,3:17]
# distance matrix
d <- dist(mydata, method = "euclidean")
fit <- hclust(d, method="ward")
# display dendogram
plot(fit)
# cut tree into 5 clusters
groups <- cutree(fit, k=5)
# draw dendogram with red borders around the 5 clusters
rect.hclust(fit, k=5, border="red")
#plot the silhouette values by using 5 clusters
plot(silhouette(cutree(fit,k=5),dist(mydata)))

#Factor analysis
# Determine number of factors by PCA Screeplot
library(stats)
pca.ptm<-princomp(PTM[,3:17], scale=TRUE, center=TRUE, cor=TRUE)
summary(pca.ptm)
loadings(pca.ptm)

#Print cummulative proportion of variance explained
cumprop=cumsum(pca.ptm$sdev^2)/sum(pca.ptm$sdev^2)
print(cumprop,digits=2)

#Screeplot
plot(pca.ptm, main="Screeplot of histone PTM data")

#FA with 2 factors and promax rotation
ptm.FA=factanal(mydata, factors=2, scores="Bartlett", rotation="promax")
ptm.FA
plot(ptm.FA$scores[,1:2], type="n", main="Factor Analysis, rotation=promax ")
text(ptm.FA$scores[,1:2])

#FA with 5 factors and promax rotation
ptm.FA=factanal(mydata, factors=9, scores="Bartlett", rotation="promax")
ptm.FA
plot(ptm.FA$scores[,1:2], type="n", main="Factor Analysis, rotation=promax ")
text(ptm.FA$scores[,1:2])
```

4.7 References

[1] Hunter, T., The Age of Crosstalk Phosphorylation, Ubiquitination, and Beyond. *Molecular Cell* 2007, 730–738.

[2] Strahl, B.D., Allis, C.D., The language of covalent histone modifications. *Nature* 2000, 403, 41–45.

[3] Jenuwein, T., Allis, C.D., Translating the Histone Code. *Science* 2001, 293, 1074–1080.

[4] Wang, Y., Fischle, W., Cheung, W., Jacobs, S., et al., Beyond the double helix: writing and reading the histone code. *Novartis Foundation symposium* 2004, 259, 3–17; discussion 17–21, 163–9.

[5] Latham, J. a, Dent, S.Y.R., Cross-regulation of histone modifications. *Nature structural & molecular biology* 2007, 14, 1017–24.

[6] Suganuma, T., Workman, J.L., Crosstalk among Histone Modifications. *Cell* 2008, 135, 604–7.

[7] Fischle, W., Yanming Wang, Allis, C.D., Histone and chromatin cross-talk. *Current Opinion in Cell Biology* 2003, 172–183.

[8] Hart, G.W., Slawson, C., Ramirez-Correa, G., Lagerlof, O., Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annual review of biochemistry* 2011, 80, 825–58.

[9] Van Noort, V., Seebacher, J., Bader, S., Mohammed, S., et al., Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Molecular systems biology* 2012, 8, 571.

[10] Lu, Z., Cheng, Z., Zhao, Y., Volchenboum, S.L., Bioinformatic Analysis and Post-Translational Modification Crosstalk Prediction of Lysine Acetylation. *PLoS ONE* 2011, 6, e28228.

[11] Yang, X., Seto, E., Lysine acetylation: codified crosstalk with other posttranslational modifications. *Oncology* 2009, 31, 449–461.

[12] Kouzarides, T., Chromatin modifications and their function. *Cell* 2007, 128, 693–705.

[13] Daujat, S., Bauer, U.-M., Shah, V., Turner, B., et al., Crosstalk between CARM1 methylation and CBP acetylation on histone H3. *Current biology* 2002, 12, 2090–7.

[14] Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., et al., Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* 2009, 138, 1122–36.

[15] Sridhar, V. V, Kapoor, A., Zhang, K., Zhu, J., et al., Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. *Nature* 2007, 447, 735–8.

[16] Lee, J., Shukla, A., Schneider, J., Swanson, S.K., et al., Histone Crosstalk between H2B Monoubiquitination and H3 Methylation Mediated by COMPASS. *Cell* 2007, 1084–1096.

[17] Zhang, C., Molascon, A.J., Gao, S., Liu, Y., Andrews, P.C., Quantitative proteomics reveals that the specific methyltransferases Txr1p and Ezl2p differentially affect the mono-, di- and tri-methylation states of histone H3 lysine 27. *Molecular & cellular proteomics* n.d., Epub ahead of print.

[18] Allis, C.D., Gorovsky, M.A., Cell, L.L., Regulation, C., Histone Phosphorylation in Macro- and Micronuclei of Tetrahymena thermophila. *Biochemistry* 1981, 20, 3828–3833.

[19] Pfeffer, U., Ferrari, N., Tosetti, F., Vidali, G., Histone Acetylation in Conjugating Tetrahymena thermophila. *The Journal of cell biology* 1989, 109, 1007–14.

[20] Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S. a, et al., Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature protocols* 2007, 2, 933–8.

[21] Van Steensel, B., Chromatin: constructing the big picture. *The EMBO journal* 2011, 30, 1885–95.

[22] Chahal, S.., Matthew, H.., Bradbury, E.., Acetylation of histone H4 and its role in chromatin structure and function. *Nature* 1980, 287, 76–9.

[23] Filion, G.J., Van Bemmel, J.G., Braunschweig, U., Talhout, W., et al., Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* 2010, 143, 212–24.

[24] Kharchenko, P. V, Alekseyenko, A. a, Schwartz, Y.B., Minoda, A., et al., Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* 2011, 471, 480–5.

[25] Ernst, J., Kellis, M., Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature biotechnology* 2010, 28, 817–25.

[26] Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., et al., Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *The EMBO journal* 2011, 30, 1928–38.

[27] Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., et al., Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* 2010, 330, 1775–87.

[28] Cao, R., Zhang, Y., The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Current opinion in genetics & development* 2004, 14, 155–64.

[29] Jacob, Y., Stroud, H., Leblanc, C., Feng, S., et al., Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* 2010, 466, 987–91.

CHAPTER 5

Myelin basic protein: implications from its post-translational modifications

5.1 Summary

Myelin basic protein (MBP) is an important component of the myelin sheath surrounding neurons and it is directly affected in demyelinating diseases. MBP contains a relatively large number of post-translational modifications (PTMs) which have been reported to play a role in multiple sclerosis while MBPs from lower vertebrates have been reported to be incapable of inducing multiple sclerosis or allergic encephalitis. This study reveals the extent of differences in PTM patterns for mammalian and non-mammalian MBPs. This included intact mass and *de novo* sequence analysis of approximately 85% of rattlesnake MBP, the first reptile MBP to be characterized and of bovine MBP. We identified 12 PTM sites in the five bovine MBP charge components which includes both previously reported and novel modifications. The most notable modification is an acetylation of lysine 121. Other modifications found in bovine MBP include N-terminal acetylation in components C1, C2, and C3; Oxidation of methionine 19 in all five components; All charge isomers had both a mono- and di-methylated (symmetric) arginine at position 106; Deimination in arginines 23 and 47 was found only in component C8b; Deimination of arginine 96 and deamidation in glutamine 102 was found in components C2, C3, C8a, and C8b; Phosphorylation in threonine 97 was

restricted to charge components C2 and C3; Deimination in arginine 161 was only found in component C3; Deamidation of glutamine 120 was only observed in C1. All four deiminated arginines and one acetylated lysine were first experimentally revealed in this study for bovine MBP. Mascot database searching combined with *de novo* sequence analysis of rattlesnake MBP provided more than 85% sequence coverage. A few PTMs were also revealed in rattlesnake MBP: Mono- and dimethylated Arg, Protein N-terminal acetylation, and deiminated Arg. Overall, snake MBP was found to undergo less modification than bovine MBP based on the mass heterogeneity of the intact protein, the bottom-up structure analysis, and the limited complexity of rattlesnake MBP chromatography. The combined data from this study and information from previous studies extend the known MBP PTMs and PTMs unique to higher vertebrates are proposed.

5.2 Introduction

Myelin basic protein, MBP, is a major component of the myelin membrane that envelops the nerve axons of mammalian and sub-mammalian species. It plays a crucial role in maintenance of the tight spiral wrappings of the myelin sheath required for proper functioning of nerves. The sequence of MBP is highly conserved in higher animals and it is a very basic protein (pI=11.65 for Bovine MBP, 18.3 kDa isoform) due to the large number of arginyl and lysyl residues present in the sequence. In bovine preparations, the 18.3 kDa form represents over 95% of the MBP. The primary structure of MBP undergoes extensive post-translational modifications (PTMs) which include phosphorylations, methylation of arginyl residues, citrullination of arginyl residues, deamidated glutaminyl residues, and N-terminal acetylation[1–9]. The number and extent

of these modifications in mammals appears to be greater than in non-mammalian species[5,6].

MBP has been implicated in demyelinating diseases, in particular, multiple sclerosis[10] and allergic encephalitis[11,12]. Characterized as a human autoantigen, several studies have shown a role for autoantibodies against MBP in the pathogenesis of multiple sclerosis[10,13]. Genomic studies have yet to provide direct insight into the mechanism underlying this phenomenon and interest in the relationship between MBP and multiple sclerosis has centered on the potential role of these MBP post-translational modifications. The purified mammalian protein containing extensive post-translational modifications differs significantly from the sub-mammalian proteins in its ability to elicit a mammalian demyelinating pathology in experimental allergic encephalomyelitis (EAE) which has been developed as an animal model of multiple sclerosis[14]. In this model, specific MBP forms can elicit an immunologic response in some species but fail to elicit a pathological response in other species[15,16].

The myriad PTMs in MBP such as methylation on arginyl residues and the addition of negatively charged phosphate groups on seryl and threonyl residues could trigger protein conformational changes, alter interactions between MBP and myelin membranes or between MBP and other myelin associated proteins. One hypothesis proposes that PTMs in MBP could be one of the key factors responsible for MBP antigenicity and the development of multiple sclerosis in mammals. The first inference comes from a quantitative study of the changes of PTMs in human MBP from normal and multiple sclerosis tissues[8]. This study demonstrated that MBP from multiple sclerosis patients was less phosphorylated but more methylated and more deiminated compared with the

protein purified from normal tissues. Other studies also showed an aberrant pattern of PTMs that may contribute to demyelination, the protein antigenicity, and its association with a phospholipid membrane[17,18]. These correlations are consistent with this hypothesis but a strict causal relationship has not yet been established. Which specific PTMs might be unique to higher animals, what differences in modification patterns exist between mammalian and non-mammalian species, and what types of PTMs are present in more evolutionarily distant animals such as reptiles, remain to be addressed and have a direct bearing on this hypothesis in light of the differences in antigenicity observed for lower vertebrates[8,19–22].

The extensive and variable PTMs of MBP lead to a complex mixture of isoforms which must be at least partially resolved before characterization. Most previous work was based on resolution of MBP charge isoforms by ion exchange chromatography prior to characterization. To maintain the ability to correlate previous studies, this study compares the extent and nature of PTMs for these charge isoforms in bovine MBP with rattlesnake MBP and adds to the current limited information on structure, isoforms, and antigenic properties of MBP from mammalian and non-mammalian species. The objective of this study was to determine whether the less antigenic MBP from lower vertebrates exhibit fewer post-translational modifications and to further characterize the PTMs of the antigenic bovine MBP. We extend the characterization of the bovine MBP PTMs across different charge components to identify additional modification sites in each charge component and we identify the probable presence of acetyl-lysine in MBP for the first time. We also reveal the partial sequence of rattlesnake MBP for the first

time, identify its PTMs, and document its relatively low level of post-translational modification relative to mammalian MBP.

5.3 Experimental procedures

5.3.1 Extraction, isolation and purification of MBP charge components

The protocol described here for bovine brain was adjusted appropriately for rattlesnake brains. Bovine MBP was isolated from the white matter of beef brains and the rattlesnake MBP were extracted from whole brains of Eastern Diamondback Rattlesnakes (*crotalus adamanteus*) from Georgia using a minor modification of the procedure of Martenson, *et al*.[19]. White matter (36-40 gm) from large brains or whole small brains were placed in a Waring blendor and 750 ml of cold 2:1 (v/v) chloroform-methanol added plus 1 ml of 0.1 M phenylmethanesulfonyl fluoride and then homogenized at low speed for 1 minute, then at high speed for 3 minutes. The extract containing dispersed tissue was poured into a large beaker and stirred overnight at 4°C. The slurry was then filtered on Whatman No 1 filter paper in a Buchner funnel to form a pad of tissue. The tissue pad was macerated in acetone overnight, with stirring and filtered to give another filter cake. This second filter cake was stirred with 0.2 N H_2SO_4 overnight with phenylmethanesulphonyl fluoride and the extract filtered and transferred to dialysis bags. The extract was dialyzed against deionized and distilled water with at least 4 to 6 changes of water at 4 degrees until the pH of the dialysate is approximately 7. When the dialysis tested neutral for pH (pH 6.8-7.1) the solution was mixed with an equal volume of 95% ethanol and placed in a freezer at -20°C. The resultant precipitate was then collected by centrifugation at 10.000 rpm and the resulting precipitate was washed three times with cold ethanol, and allowed to dry at room temperature. The dried white powder was then stored at -80°C.

5.3.2 Isolation of charge components

Using the method of Martenson, *et al*.[19], approximately 100mg (by O.D. 280) of the unfractionated charge component mixture was dissolved in a glycine-urea buffer at pH 9.5 room temperature and applied to a carboxymethyl cellulose column (Whatman CM 52, Pharmacia K9/30) equilibrated with the same buffer. The column was washed with glycine-urea buffer pH 10.5 and the charge component eluted with a linear gradient of 0 to 0.15M sodium chloride contained in 0.08 M glycine, 2M urea, pH 9.5 buffer. Fractions of 3 ml were collected. The charge components elute in order of increasing positive charge. Each fraction containing a major charge component was dialyzed at 4°C, against deionized water overnight, lyophilized and stored at -20°C. To remove minor contaminants in the rattlesnake MBP, the fraction was further separated by a Vydac C8 reverse phase column (Part No. 208TP54, 250mm x 4.6mm) with a linear gradient from 5% to 70% acetonitrile in 0.1% formic acid/water over 90 min. Each component was subjected to SDS gel electrophoresis under basic pH conditions to ascertain purity. Purity of the rattlesnake and bovine MBPs were estimated at 95% and 98% respectively by Coomassie Blue staining of SDS gels.

5.3.3 Protein digestion, mass spectrometer analysis, and PTM assignment

For each charge component, about 20μg of MBP was digested with sequencing-grade trypsin (Promega, Madison, WI) or Glu-C (Roche Applied Science) at a ratio of 1:30 enzyme:protein at 37 degrees, 6h in 100mM ammonium bicarbonate buffer. In addition to trypsin or Glu-C digestion, rattlesnake MBP was also digested with chymotrypsin (Roche Applied Science) at a ratio of 1:20 enzyme:protein at 37 degrees, 3h in the same digestion buffer. The digests were further vacuum dried and reconstituted in 0.1% formic acid for

mass spectrometry analysis. For bottom up analysis of myelin basic protein, peptides after trypsin or Glu-C digestion were resolved with a 12cm, in-house made, C18 capillary column (5µm, 300Å, column ID: 99µm) and eluted into a Thermo Fisher Orbitrap XL mass spectrometer with a linear gradient from 5% to 90% acetonitrile in 0.1% formic acid/water at 200nl/min flow rate over 60 min. The mass spectrometer was run in data dependent mode with 5 CID or CID/HCD hybrid microscans for the most abundant peaks with dynamic exclusion set for 120 seconds. The full scan (MS1) covers the range of m/z from 400 to 2,000 m/z at resolution=60,000. The normalized collision energy was 27% for both CID and HCD. A neutral loss mass list with 32.67, 49, and 98 m/z was set to enhance the detection of phosphorylated peptides. Internal mass calibration was achieved by mass lock of 2 polydimethylcyclosiloxane (PCM) ions (m/z =445.12002 and 429.088735); For intact mass analysis, proteins were directly infused into or separated on a 15cm, in-house made, C8 capillary column (5µm, 300Å, column ID: 99µm) and eluted into a Thermo Fisher Orbitrap XL mass spectrometer under data dependent mode with 5 CID or HCD events. Digests of rattlesnake MBP were also analyzed on a Sciex model 4800 MALDI TOFTOF to extend the sequence coverage and to aid *de novo* analysis. Proteolytic digest were modified using sulfophenylisothiocyanate (SPITC) according to the method of Joss, *et al.*[23]. Manual *de novo* analysis was performed by an experienced mass spectrometrist using MSExpedite, and in-house tool to aid *de novo* sequencing.

Spectra collected from bovine MBP were searched against the Uniprot bovine protein database using the Mascot search engine (Matrix Sciences, version 2.2.02) including its recently available top-down algorithm. Spectra from rattlesnake MBP were error-tolerant searched against NCBI protein database including all species. A false discovery rate was

estimated and calculated from the bovine decoy database. Mass error tolerance for the precursor ion was 10ppm and was 0.8Da for the fragment ion. Up to two missed cleavages were allowed for enzyme digestion. The following variable modifications have been considered for both bottom-up and top-down methods: protein N-terminal acetylation, acetylation(K), deamidation(NQ), deimination(R), methylation(KR), phosphorylation(STY) and oxidation(M). Spectra from intact mass analysis were deconvoluted by the Xtract module of BioWorks software from ThermoFisher (Version 2.0.7), to get the monoisotopic mass and the isotope pattern of the intact proteins. *De novo* sequence analysis of rattlesnake MBP was achieved with PEAKS Studio 5.3. All significant spectra assigned with PTMs were considered candidates for further manual validation. Rules for those peptides assigned with PTMs based on *de novo* sequence analysis are: 1) Candidate must have a high-quality spectrum; 2) Most abundant ions should be assigned as b or y ions; 3) They should have more than 2 spectra observed; 4) There were at least 3 consecutive peaks covered; 5) Local confidence on a PTM should be at least more than 50%; and 6) Precursor has a mass error less than 5ppm. The rattlesnake *de novo* sequence analysis was performed manually from MALDI TOFTOF spectra using the MSExpedite software.

Bovine MBP isoforms or rattlesnake MBP resolved by SDS gel electrophoresis were verified from an in-gel trypsin digest of Coomassie Blue-stained gels after direct application to a MALDI plate in alpha-cyano-4-hydroxycinnamic acid matrix and analysis on a Sciex model 4800 MALDI TOFTOF. The spectra were identified by a MASCOT search against the protein database.

5.3.4 Western blot analysis of protein acetylation in bovine MBP

Bovine MBP C1-C3 components with Tetrahymena histone H3 and bovine carbonic anhydrase II as positive and negative controls, respectively, were separated by SDS-PAGE under reducing conditions and transferred to nitrocellulose membranes. Membrane was washed and blocked with non-fat dry milk at room temperature and then incubated overnight with acetylated-lysine rabbit monoclonal antibody (Ac-K^2-100, 1:500) purchased from Cell Signaling Technology. The membrane was washed and then incubated with peroxidase-labeled Anti-Rabbit IgG (1:1000) for 60 min and visualized using the ECL chemiluminescent reagent (GE Healthcare/Amersham) with a 10 min exposure. Finally, the film was imaged and analyzed with the Bio-Rad imaging system and associated software.

5.4 Results and discussion

5.4.1 PTMs found in bovine MBP and those found in other species

Generally, the sequence coverage generated from trypsin digestion was more than 80% and it was more than 60% from Glu-C digestion for a total coverage of more than 90% for the combined methods. A typical false discovery rate for tryptic peptides at the identity threshold was around 1.5% and it was under 3% for those peptides cleaved from Glu-C digestion. At least three variants of bovine MBP appeared to be present in the preparation by SDS gel electrophoresis, a major form corresponding to approximately 18 kDa and minor forms corresponding to approximately 16 kDa, 23 kDa, and 37 kDa. Some of these minor forms were verified by MALDI peptide mass fingerprinting of the gel bands analysis of tryptic digests and altogether were estimated to correspond to less

than 2% of the total MBP and thus did not interefere with the intact mass analyses. Rattlesnake MBP was estimated to be 19 kDa by SDS gel electrophoresis.

We identified a total of 12 PTMs in all 5 charge components from bovine MBP: C1-3, C8a and C8b based on LC-MSMS of proteolytic digests (Figure 5.1, Table 5.1). The PTMs identified are: acetylated Ala at the protein N-terminus; oxidized M19; deimination at R23, R47, R96 and R161; deamidation at Q102 and Q120; phosphorylated T97; mono- and dimethylated R106; and acetylated K121.

Among these modifications of bovine MBP, the putative presence of acetylation of a lysyl residue has not been previously reported. The modified peptide identified in this study was consistent with the presence of acetyl-lysine (Figure 5.2) at position 121, adjacent to a deamidated glutamine at 120. The total mass shift of 43 Daltons is localized to the QKP tripeptide based on the observed fragment ions and was initially interpreted as a potential modification of the lysyl residue in the deamidated peptide. The observed mass (1842.8370 Da) at 2.4 ppm is consistent with the calculated mass for the acetyl-lysine peptide (1842.8326 Da) rather than trimethylysine (1842.7966 Da) at 22 ppm but indistinguishable from the mass of the carbamylated and non-deamidated peptide (1842.8438) at 3.7 ppm. The average experimental mass accuracy for peptides identified in this study is 5 ppm.

Inspection of the MSMS spectrum identified a weak fragment ion at 901.18 Da which might correspond to the neutral loss reported for carbamylated peptides[24,25], but the intensity is much lower than expected and the mass assignment is limited by mass accuracy in the MSMS mode. To distinguish between the presence of acetylation or carbamylation on Lys 121, a western blot of bovine MBP (Figure 5.3) using anti-

acetyllysine antibody was performed. The immunoblot is consistent with the presence of acetyllysine in MBP, with the highest level in component C3. The staining intensity observed for acetyl-lysine immunoreactivity in MBP is consistent with the low spectral count obtained for the putative acetylated form of this peptide (5 spectra were detected in C3, none were found in other components). This modified lysine appears to be a rare modification based on spectral counts of the modified versus non-modified peptide. Note that the observation of a probable acetyl-lysine only in the deamidated peptide may simply be a sampling issue and we cannot exclude the possibility that the parent peptide is not also acetylated.

We identified both mono- and dimethylated forms of R106 in all five MBP components (Figures 5.4a and b) which verified a previous study[1]. The strong neutral losses of 31 Da and 49 Da observed in the MSMS spectra correspond respectively to monomethylamine (MMA, H2N-CH3) and MMA+H2O[26], confirming that R106 was symmetrically dimethylated (Figure 5.4b). This is consistent with previous reports for other species which used non-mass spectrometric methods[1]. The level of methylation observed varied in the different charge components. By intact mass analysis, the ratio of monomethylation vs dimethylation decreased in components C1 through C3 (Figure 5.5a-c) but it was not possible to verify this in components C8a and C8b because intact mass analysis failed to produce high-quality spectra due to impurities present in these two components. Brostoff, S. *et al*., postulated that the biological function of the methylated arginine residue was to stabilize the double-chain structure for the protein induced by the triproline site[1]. In human MBP, more highly methylated Arg was observed in multiple sclerosis patients compared with normal tissue and has been suggested to play an

important role in the pathogenesis of multiple sclerosis[8]. It is noteworthy that this modification has been found in many species such as human, bovine, rabbit, and chicken but was not found in spiny dogfish. We also found clear evidence for this modification in rattlesnake based on intact mass analysis of rattlesnake MBP (Figure 5.5d). Spectra from MSMS also confirmed this modification in a peptide with the sequence G**R**GLSFSR (Figures 5.6a and b) as this sequence matched well with known species and the fragmentation pattern of this peptide is very similar to that of the methylated peptide from bovine MBP (Figures 5.4a and b). Absence of this modification in some lower animals such as dogfish may be due to lack of an arginine in this position (Figure 5.7), suggesting methylated arginine is widely distributed from mammals to reptiles but absent in MBP from lower vertebrates which lack this residue.

Several sites were observed in bovine MBP for conversion of arginine to citrulline. Catalysed by peptidylarginine deiminases (PADs), deimination (citrullination) of arginine residues can have significant consequences for the structure and function of proteins, and has also been implicated in the pathogenesis of multiple sclerosis[27–29]. Native MBP contains several non-deiminated arginines and forms tighter, more compact myelin sheaths. Deimination of MBP affects the stability of the myelin sheath by conversion of positively charged arginine into uncharged citrulline, which increases the hydrophobicity of the protein and reduces the interaction with the negatively charged phosphatidylserine in the membrane, leading to a more open structure susceptible to proteolysis by cathepsin D[29,30]. The ratio of deiminated MBP/total MBP was found to be crucial in the physiological function of CNS and the degree of deimination of MBP correlated well with the severity of multiple sclerosis[28,29]. MBP was highly deiminated in multiple

sclerosis patients and in infants during normal CNS development and was less deiminated in healthy adults[8,28,29]. More interestingly, possible crosstalk between deimination and methylation in MBP has been revealed by Pritzker, *et al*.[31] who observed that methylation of R106 in bovine MBP (R107 in human MBP) had a significant correlation to arginine deimination. In human MBP, increased deimination of arginyl residues accompanied with a decreased methylation of R107 has been observed. In this study, we found a total of 4 arginines deiminated at positions 23, 47, 96, and 161 respectively (Figure S5.1-4). Although deimination of arginine in MBP has been found in other species, deimination has not previously been reported in bovine MBP. The residue corresponding to R23 in bovine MBP has been reported to be deiminated in human and chicken MBP[6,8] and it was also observed in rattlesnake MBP in this study (Figure S5.5) indicating that deimination is highly conserved at this site. The residues corresponding to R47 and R161 in bovine MBP have only been previously observed to be deiminated in human MBP, while deiminated R96 found in this study has not been found in other species thus far. The conserved Arg in rattlesnake corresponding to bovine R161 was not found to be deiminated but the nearby R (in peptide sequence GY**R**YDGQGTLSK, Figure S5.6) was found to be partially deiminated (both forms observed by MS). There is no homologous Arg residue in mammalian MBPs. The degree of conservation of deimination in mammalian MBPs is still incomplete as deimination of Arg has not yet been examined in rabbit or other mammalian MBPs.

Acetylated alanine at the protein N-terminus was found in components C1, C2, and C3 in this study. This modification was not found in the previous study of bovine MBP PTMs by capillary electrophoresis-mass spectroscopy[2]. And it was also not found in

the tryptic peptides in this study possibly because the peptide with N-acetylalanine (AAQK) was too short to be effectively detected by the mass spectrometry method or the peptide did not ionize well. When we digested MBP with Glu-C protease, the N-terminally acetylated peptide was clearly detected and the PTM was unambiguously assigned by tandem mass spectrometry (Figure S5.7). Intact protein mass analysis also agreed with our Glu-C bottom-up study (Figure 5.5a-c). Bovine MBP in C1-3 appears to be predominantly acetylated since no spectra corresponding to the unmodified peptide or protein were observed in both bottom-up and intact mass analyses. Acetylation of the N-terminal alpha-amino group of proteins is a common modification in eukaryotes and about 85% of eukaryotic proteins are N-terminally acetylated[32], with N-terminal Ala being a common acetylation site. As a conserved and widespread modification, the biological role of this N-terminal acetylation is unclear, although it has been suggested that one biological function is as a signal for protein degradation[33]. This modification was also found in other MBPs from mammals such as human, bovine, and rabbit, to non-mammals including chicken, dogfish, and rattlesnake in this study (Figure S5.8).

A wide range of MBP phosphorylation sites have been reported in previous studies[1–9], however, this study did not focus on phosphorylation so no effort was made to enrich for phosphopeptides. Despite the absence of enrichment methods, we identified a phosphorylation site at T97 which was present in bovine MBP components C2 and C3 (Figure S5.9).

Phosphorylation of T97 has been previously reported for human, bovine, rabbit, and chicken MBPs but not in spiny dogfish because the sequence (PR<u>T</u>PPP) associated with this PTM is highly conserved in higher animals but is missing in dogfish. Other

conserved phosphorylated sites are: S7 found in human, bovine, rabbit and chicken; Phosphorylated Ser in sequence RGSGK corresponding to S54 in bovine, and S56 in human and rabbit; Phosphorylated Ser at protein C-terminus (SGSPXARR, X is M or V) in human, bovine, rabbit, and chicken. None of these phosphorylated sites were detected in spiny dogfish and appear to be absent also in rattlesnake as no clear evidence of phosphorylation was found from the intact mass analysis and *de novo* sequencing study (see below).

Deamidated Q102 was found to be present in components C2, C3, C8a, and C8b (Figure S5.10); A deamidated glutamine at position 120 is also presented in C3 (Figure 5.2). Deamidation of Q102 and Q120 was also observed in human MBP but has not been examined for rabbit MBP[8,9]. Overall, the C1 fraction was least modified while the less basic C3 and C8b fractions were more highly modified. The analysis of proteolyzed MBP fractions were also consistent with our intact mass analysis of MBP proteins that the deconvoluted spectra showed C3 had the most complicated spectra as compared with C1 and C2 fractions(Figure 5.5a-c). We have no unambiguous evidence that rattlesnake MBP is deamidated and note that at least some of the Gln and Asn residues reported to be deamidated in mammals are substituted in rattlesnake by other residues. However, the MALDI MSMS sequence for the peptide isolated from a GluC digest having the sequence HKYAHQ/KGHQ/KGYRYDGE appears to end in a Glu residue despite strong evidence for a Gln residue in other peptides spanning this region.

The oxidized methionine in bovine MBP at position 19 has not been reported before and was found in all five components in this study (Figure S5.11). While it is quite likely that the methionine oxidation observed in this study resulted after the protein was

extracted from the cell, we cannot rule out its presence *in vivo*. It is worth noting that several studies in other proteins have demonstrated that methionine oxidation has a significant effect on protein function, stability, aggregation and folding[34–38]. More generally, oxidation of methionine can decrease protein stability. One explanation for the effect is that the extra oxygen atom introduced by methionine oxidation changes the protein hydrophobicity at that site[39].

5.4.2 Intact mass study of bovine MBP charge components and unfractionated rattlesnake MBP

The presence and distribution of several PTMs in three charge components of bovine MBP, C1-3, have been successfully characterized by intact mass analysis(Figure 5.5a-c). In C1, the first major component had a calculated monoisotopic mass of 18,354.44 Da while the theoretical monoisotopic mass calculated from the protein sequence was 18312.39 Da. The mass difference was +42.05 Da which is consistent with the N-terminal acetylation observed at the peptide level. The mass of the second major peak was further increased by 14.04 Da which is consistent with monomethylation of the N-terminally acetylated protein. The third major peak had a mass difference of +28.05 Da from the first peak suggesting that MBP in the C1 fraction was also modified with a dimethylation. The most abundant peak (Mw=18,368.48 Da) in the C1 fraction corresponds to MBP modified with both N-terminal acetylation and monomethylation. All the ions found in the C1 fraction were also present in the C2 fraction. However, the most abundant peak in C2 corresponds to the protein modified with N-terminal acetylation and dimethylation. An additional species mass shifted by +0.99 Da from 18,368.48 Da was found in C2, which is consistent with either deamidation or

citrullination. Evidence consistent with the presence of an oxidized Met in C2 came from a peak with mass increased by +16 Da from 18,382.48 Da. The C3 component had the most complex spectrum arising from the combination of different PTMs. In addition to those PTMs found in C1 and C2, deamidated or citrullinated residues were suggested by mass increases of 1 or 2 Da. Although all three components were modified with a mono- and di-methylated Arg, the distribution of non-methylated, mono- and dimethylated forms changed in different charge isomers. For instance, the ratio of non-methylated: monomethylated: dimethylated was 1 :1.18 :0.80 in C1 while this ratio was changed to 1 :0.45:1.14 in C2. Generally, the results from intact mass analysis were consistent with those from analysis of the proteolytic digests: C1 was least modified while C3 was most modified.

Unlike the spectra from bovine MBP fractions C1-C3, intact mass analysis of rattlesnake MBP revealed a much simpler spectrum (Figure 5.5d), suggesting rattlesnake MBP has considerably fewer PTMs than bovine MBP although the protein sequences are fairly well conserved in many species. Note that the complete protein sequence for rattlesnake MBP is unknown. The intact monoisotopic mass of rattlesnake MBP was approximately 19,564 Da. A form having a mass shift of 1 Da was observed for intact rattlesnake MBP suggesting rattlesnake MBP might be deamidated or citrullinated. This is consistent with the evidence with citrullination observed from the bottom-up analysis discussed below. No evidence was observed for deamidation in the bottom-up analysis. A mass shift of 14 Da is consistent with the bottom-up analysis which identified methylation of Arg in the peptide, G**R**GLSFR. Spectra of intact rattlesnake MBP are generally less post-translationally modified compared with spectra of intact bovine MBP.

5.4.3 Partial protein sequence of rattlesnake MBP and its PTMs

We performed Mascot database searching against the entire protein sequence database with error tolerance as well as *de novo* sequence analysis of peptides from rattlesnake MBP to search for conserved sequences and possible PTMs present in the protein. The *de novo* sequencing software, PEAKS 5.3, was employed in this study for Orbitrap with manual confirmation. The error tolerance was 10ppm for precursor ions and 0.8 Da for fragment ions. Total Local Confidence (TLC) and Average Local Confidence (ALC%) scores have been used to determine the quality of the predicted sequence for particular MS/MS spectra[40]. In addition to these procedures, independent manual *de novo* sequencing was performed on MALDI TOFTOF MSMS spectra by an expert (Walker). Mascot database searching together with the *de novo* interpretations identified more than 20 unique peptides (or a total of 150 AAs) and 4 PTMs (acetylation, monomethylation, dimethylation, deimination) at 4 sites in rattlesnake MBP (Table S5.1, Figure S5.12), which covers 85% of gekko and 89% of anole MBP sequences respectively (Figure 5.8). One additional potential modification was observed for deamidation of Gln to generate the C-terminal Glu residue in the peptide HKYAHKGHKGYRYDGE. The residue assignments for I/L and K/Q in Figure 5.8 were based on the Anole and Gekko sequences except when v and w fragment ions were able to distinguish I/L (See Figure S5.13 for an annotated example spectrum and Table S5.1 under MALDI TOFTOF peptides. The sequence of rattlesnake MBP exhibited good sequence identity with most species but is closest to reptile MBP sequences (gekko and anole) (Figure 5.8). Mono- and dimethylated Arg were also identified in a peptide with the sequence G**R**GLSFSR and this peptide exhibited a very similar fragmentation pattern to the homologous peptide

from bovine MBP having only a single amino acid difference (Figures 5.6a and b). Other PTMs found in this species are: Protein N-terminal acetylation in a chymotryptic peptide (Figure S5.8) and a deiminated Arg in two tryptic peptides (LATASTIDHA**R**HGSPR and GY**R**YDGQGTLSK) (Figures S5.5 and S5.6). Both of these sites are partially deiminated as evidenced by observation of peptides containing arginine as well as peptides containing citrulline. All MBPs from other species including rattlesnake, which have been previously investigated or addressed in this study, have been found to have acetylated alanine at the N-terminus indicating this modification is conserved across different species (Figure 5.7).

5.5 Conclusions

Similarly to histone proteins, myelin basic protein is a very basic, sequence-conserved protein exposed to various post-translational modifications. Extensive post-translational modifications including acetylation, deamidation, deimination, methylation, and phosphorylation are found in both proteins. Interestingly, the dynamic transitions between transcriptionally-active and transcriptionally-silent chromatin states is determined by a combinatorial set of these PTMs on one or more histones, a well-known 'histone code' hypothesis, which determines the compaction of eukaryotic chromatins[41,42]. Some evidence has also emerged that the compaction of the myelin sheath in CNS could also be affected by dynamic protein post-translational modifications of MBP[8,18] and the limited number or absence of these modifications in lower vertebrates may have implications for myelin function in these species.

The total number of conserved PTMs found in our study might be underestimated as the different methodologies employed to identify PTMs in different studies may have

some biases and low level modifications may be missed. Our intact MS analyses, however, strongly suggest that the bulk of rattlesnake MBP is much less modified than bovine MBP. This observation is consistent with the bottom-up analysis and with the single fraction observed on ion exchange chromatography.

Our study represents the most extensive MS/MS analysis of bovine myelin basic protein PTMs and leads to identification of new PTMs in bovine MBP and the relative levels of the PTM among isoforms has been revealed. This also represents the first analysis of PTMs in MBP from a reptile through tandem mass spectrometry and demonstration that citrullination and methylation of specific homologous Arg residues are conserved from snake to primates. We conclude that the results of this study are consistent with the general correlation between MBP PTMs and the antigenicity of MBP in demyelinating diseases such as multiple sclerosis and suggest specific modifications to evaluate.

Table 5.1 Peptides with PTMs found in bovine MBP isomers

Start	End	Sequence	PTM site information	Mascot ion score	ΔM*	m/z ratio	Charge	Present in charge isomers
1	20	AAQKRPSQRSKYLASASTMD	Acetyl (N-term)	41	4	746.7175	+3	C1,C2,C3
10	23	SKYLASASTMDHAR	Oxidation (M19)	93	2	777.3734	+2	C1,C2,C3,C8a,C8b
10	23	SKYLASASTMDHAR	Oxidation (M19)	60	2	518.5853	+3	C1,C2,C3,C8a
12	23	YLASASTMDHAR	Oxidation (M19)	50	4	669.8116	+2	C1,C2,C3,C8b
12	23	YLASASTMDHAR	Oxidation (M19)	40	0	446.8751	+3	C1,C2,C3
12	29	YLASASTMDHARHGFLPR	Deimination (R23)	40	1	677.6674	+3	C8b
12	29	YLASASTMDHARHGFLPR	Oxidation(M19), Deimination(R23)	38	1	682.9996	+3	C8b
42	51	FFGSDRGAPK	Deimination (R47)	55	1	541.7674	+2	C8b
91	104	NIVTPRTPPPSQGK	Deimination (R96)	56	3	746.9118	+2	C2,C3,C8a,C8b
91	104	NIVTPRTPPPSQGK	Phosphorylation (T97)	32	3	786.4043	+2	C2,C3
91	104	NIVTPRTPPPSQGK	Deamidation (Q102)	31	1	746.9102	+2	C2,C3,C8a,C8b
105	112	GRGLSLSR	Methylation (R106)	39	1	430.2590	+2	C1,C2,C3,C8a,C8b
105	112	GRGLSLSR	Dimethylation (R106)	31	1	437.2669	+2	C1,C2,C3,C8a,C8b
113	129	FSWGAEGQKPGFGYGGR	Deamidation(Q120), Acetyl(K121)	31	6	922.4236	+2	C3
159	168	DSRSGSPMAR	Deimination(R161)	24	-1	532.7426	+2	C3

*Note: Mass error is measured as ppm.

Table S5.1 List of peptides and PTMs from rattlesnake MBP identified by MALDI TOFTOF or Orbitrap mass spectrometer

a. Mascot database search with error tolerance

From 4800

Trypsin

Sequence	M/Z	ppm	mascot score	Matched Species	Modification/Subsitution
K.LATASTIDHAR.H	1155.6149	3	70	Gekko japonicus	
R.YREPSLLDSLGR.F	1405.7516	6	65	Gekko japonicus	
R.VSHVSSLPHPR.S	1215.6614	2	78	Gekko japonicus	
R.FSWGGETHKPGYGSGK.F	1694.8066	9	86	Taeniopygia guttata	

From Orbitrap

Trypsin

Sequence	M/Z	ppm	mascot score	Matched Species	Modification/Subsitution
K.LATASTIDHAR.H	578.3096	0	94	Gekko japonicus	
K.LATASTIDHARHGSPR.R	845.9372	3	57	Gekko japonicus	Deimination (R11)
R.YREPSLLDSLGR.F	703.376	1	61	Gekko japonicus	
R.VSHVSSLPHPR.S	405.8913	0	65	Gekko japonicus	
R.FSWGGETHKPGYGSGK.F	847.901	2	56	Taeniopygia guttata	
R.GKDYHHAAR.V	535.7601	0	63	Gekko japonicus	Ala->Ser (A8)
K.NIVSPR.T	685.4007	2	35	Gekko japonicus	
R.EPSLLDSLGR.F	543.7941	2	34	Gekko japonicus	
K.GRGLSFSR.F	440.2438	1	40	Gekko japonicus	
K.GRGLSFSR.F	447.2523	2	30	Gekko japonicus	Methyl(R2)
K.GRGLSFSR.F	454.2591	0	23	Gekko japonicus	Dimethyl(R2)
K.FYEHK.Y	362.1768	0	25	Gekko japonicus/ Taeniopygia guttata	
K.GYHYDGQGTLSK.L	672.8304	-1	71	Gekko japonicus	His->Arg (H3)
H.YDGQGTLSK.L	484.7382	1	58	Gekko japonicus	
K.LGGSGSRPGSR.H	530.7793	2	51	Anolis carolinensis	

Glu-c

Sequence	M/Z	ppm	mascot score	Matched Species	Modification/Subsitution
E.THKPGYG.S	759.3798	2	42	Taeniopygia guttata	
E.THKPGYGSGKFYE.H	735.8574	4	58	Taeniopygia guttata	
K.PGYGSGKFYE.H	1104.5053	5	43	Taeniopygia guttata	

Chymotrypsin

Sequence	M/Z	ppm	mascot score	Matched Species	Modification/Subsitution
M.ASQKRSSF.R	476.7465	1	37	Gekko japonicus	Acetyl (Protein N-term)
F.RHGSKL.A	697.4107	0	24	Gekko japonicus	
L.DSLGRF.F	694.3534	2	18	Gekko japonicus	
F.SWGGETHKPGY.G	609.7809	1	44	Gekko japonicus	
W.GGETHKPGYGSGKF.Y	711.3457	2	51	Gekko japonicus	
K.PGYGSGKFY.E	975.463	6	52	Gekko japonicus	
Y.GSGKFY.E	658.3187	-1	18	Gekko japonicus	
-.ENPVVHFF.K	988.491	2	48	Gekko japonicus	

b. *De novo* interpretation by PEAKS

From Orbitrap							Modification
Glu-c							
Sequence	TLC	ALC (%)	m/z	z	ppm	local confidence (%)	
THKPGYGSGKFYE	8.6	66	490.907	3	3.7	77 79 86 53 41 52 92 88 62 53 64 54 55	
THKPGYGSGKFYE	10	77	735.8578	2	4.9	81 86 96 93 90 74 53 55 75 79 81 64 67	
THKPGYGSGKFYE	9.7	74	1470.711	1	6.7	72 76 89 94 93 91 66 53 48 63 74 70 72	
YVDENPVVR*KSSE*	8.2	63	761.3824	2	2.1	52 57 78 81 70 82 85 95 57 63 45 28 21	
SKLATASTLD*AFKPD*	10.2	68	782.9169	2	3	40 47 69 92 95 93 93 94 94 93 61 32 37 48 21	
*GP*YGSGKFYE	8.3	83	1104.506	1	5.4	84 82 82 92 92 89 88 81 71 67	
Trypsin							
Sequence	TLC	ALC (%)	m/z	z	ppm	local confidence (%)	
YDGQGTLSK	7.4	82	484.738	2	0.4	96 95 92 77 80 80 80 82 50	
TASTLDHAR	7.1	79	486.2496	2	1.5	74 76 90 94 90 89 86 88 21	
ATASTLDHAR	7.6	76	521.7682	2	1.5	55 61 81 91 94 86 84 89 91 21	
LATASTLDHAR	9.1	83	578.3113	2	3.3	90 90 94 93 93 91 82 81 84 85 21	
EPSLLDSLGR	6.8	68	543.7962	2	5.8	35 37 40 66 87 86 85 84 89 63	
VSHVSSLPHPR	7.8	70	608.3347	2	2.5	27 42 62 93 94 91 91 92 87 71 21	
FFGGDKHYPR	8.2	82	612.303	2	2.5	87 88 78 77 82 82 90 70 91 72	
DENPVVHFFR	7.5	75	630.3106	2	-2.1	88 84 79 85 88 93 89 60 41 41	
DENPVVHFFR	7.1	71	1259.622	1	4.1	94 91 82 86 88 78 56 49 37 43	
VDENPVVHFFR	8	73	679.8484	2	3.3	42 48 63 68 89 91 91 91 85 62 62	
YVDENPVVHFFR	9.8	82	761.3809	2	4	62 65 83 90 86 92 93 94 94 87 66 66	
Chymotrypsin							
A(+42.01)SQKRSSF	5.5	69	476.7461	2	0.2	75 75 48 57 35 77 89 91	N-Acetyl
FSRF	3.7	93	556.2883	1	0.9	92 91 94 93	
EHKY	3.6	90	576.2784	1	1.5	91 91 95 82	
AKGRGL	3.7	61	601.3789	1	1.5	53 57 77 80 83 16	
YEHKY	3.7	75	739.3406	1	-0.4	58 90 60 94 70	
DSLGRFF	5.4	78	841.4224	1	2.6	93 87 84 84 85 88 21	
NPVVHFF	5.4	78	859.4479	1	2.1	72 84 85 16 94 94 94	
ENPVVHFF	5.9	74	988.4893	1	0.6	57 53 84 87 16 97 97 97	
DENPVVHFF	7.7	85	1103.519	1	3	88 84 72 70 73 89 96 96 96	
VDENPVVHFF	7	70	1202.588	1	2.9	37 41 81 73 58 60 92 89 91 76	

Note: Red, italic residues are misinterpreted by PEAKS as confirmed by manual de novo sequence analysis.

c. Manual *De novo*

From 4800

Trypsin

m/z	Sequence	Note
1155.61	[I/L]ATAST[I/L]DHAR	
1215.67	VSHVSSLPHPR	
1345.63	GY**R**YDG[K/Q]GTLSK	Arg --> Citrulline
1521.79	YVDENPVVHFFR	
1561.86	RYREPSLLDSLGR	
1694.81	FSWGGETH[K/Q]PGYGSGK	

GluC

m/z	Sequence	Note
1302.78	Acetyl-AS[K/Q][K/Q]RSSFRHG	
1470.81	TH[K/Q]PGYGSG[K/Q]FYE	
1564.91	S[K/Q]LATASTIDHARHG	
1945.9	H[K/Q]YAH[K/Q]GH[K/Q]GYRYDGE	no citrulline at R

Trypsin SPITC

m/z	Sequence	Note
1094.6	GRGLSFSR	
1108.4	G**R**GLSFSR	methyl Arg
1122.5	G**R**GLSFSR	dimethyl Arg
1275.6	LSGSGSRPGSR	
1370.6	LATAST[I/L]DHAR	
1430.6	V[SH]VSSLPHPR	
1438.6	FFGGD[K/Q]HYPR	
1620.7	YEPSLLDSLGR	
1736.8	YVDENPVVHFFR	
1905.9	……TIDHA**R**HGSPR	Arg --> Citrulline

From orbitrap

Trypsin

m/z	Sequence	Note
673.33	GY**R**YDGQGTLSK	Arg --> Citrulline

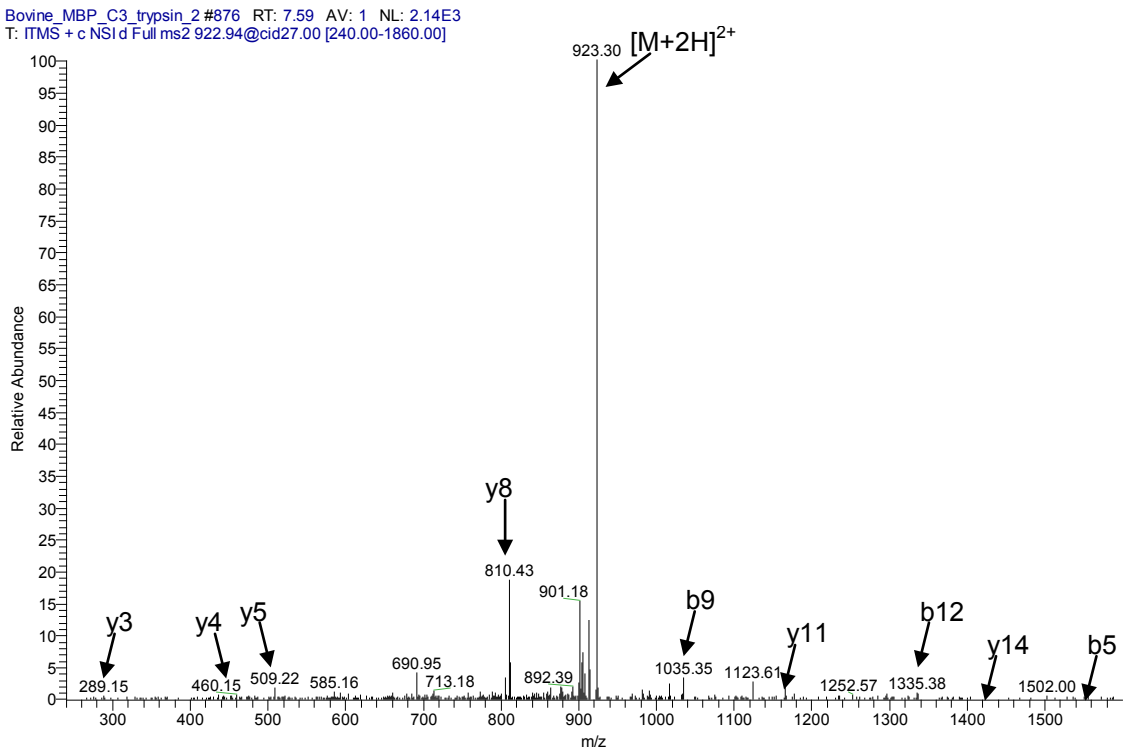
Chymotrypsin

m/z	Sequence	Note
952.4999	**A**SQKRSSF	Acetyl (Protein N-term)

>sp|P02687|MBP_BOVIN; pI=11.65; Average Mass:18323.48; Monoisotopic Mass:18312.39

1-50 AAQKRPSQRS KYLASASTMD HARHGFLPRH RDTGILDSLG RFFGSDRGAP
51-100 KRGSGKDGHH AARTTHYGSL PQKAQGHRPQ DENPVVHFFK NIVTPRTPPP
101-150 SQGKGRGLSL SRFSWGAEGQ KPGFGYGGRA SDYKSAHKGL KGHDAQGTLS
151-169 KIFKLGGRDS RSGSPMARR

Figure 5.1 All PTMs found in this study indicated in the bovine MBP primary sequence
A total of 11 PTMs were found in bovine MBP and their positions are underlined in the figure. Modifications include: acetylated Ala at the protein N-terminus; oxidized M19; deimination at R23, R47, R96 and R161; deamidation at Q102 and Q120; phosphorylated T97; mono- and dimethylated R106; and acetylated K121.



M/Z: 922.4236, Charge: +2, Deam (Q120), Acetyl(K121), Ion Score: 31, Expect:0.8

Figure 5.2 MS/MS Fragmentation of 113-FSWGAEG**QK**PGFGYGGR-129

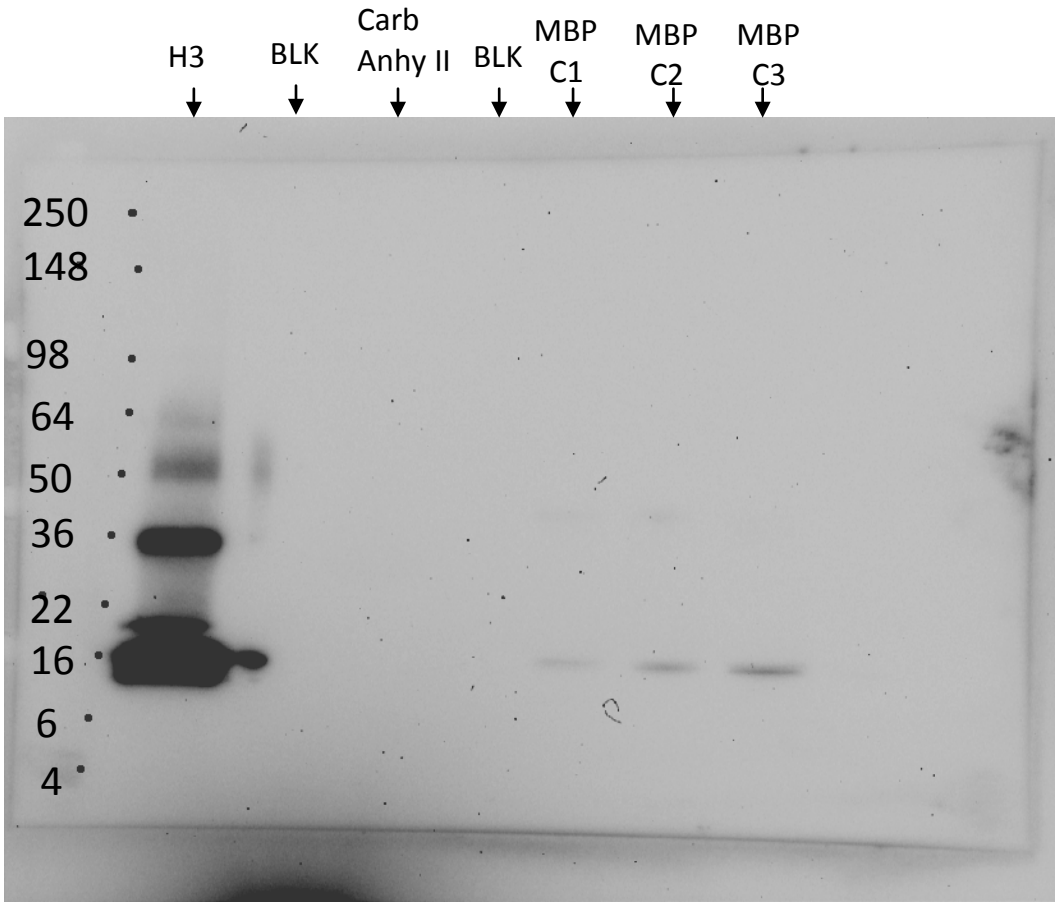


Figure 5.3 Confirmation of the acetylated-Lysine in MBP fraction C3 by western blot
Acetylated-Lysine Rabbit monoclonal antibody: Ac-K^2-100 (1:500); Anti_Rabbit IgG (1:1000); 10 minute exposure; Amount of protein loaded from Left to Right: histone H3 (tetrahymena thermophila): 3 µg; carbonic anhydrase II (bovine): 30 µg; MBP C1-C3 (bovine): 40 µg. Immunoreactivity was were detected in all 3 charge components (C1, C2, C3) and in the positive control (histone H3) while no band was found in the negative control (bovine carbonic anhydrase II). The strongest signal was observed for fraction C3 in which Acetyl-lysine was identified by MSMS.

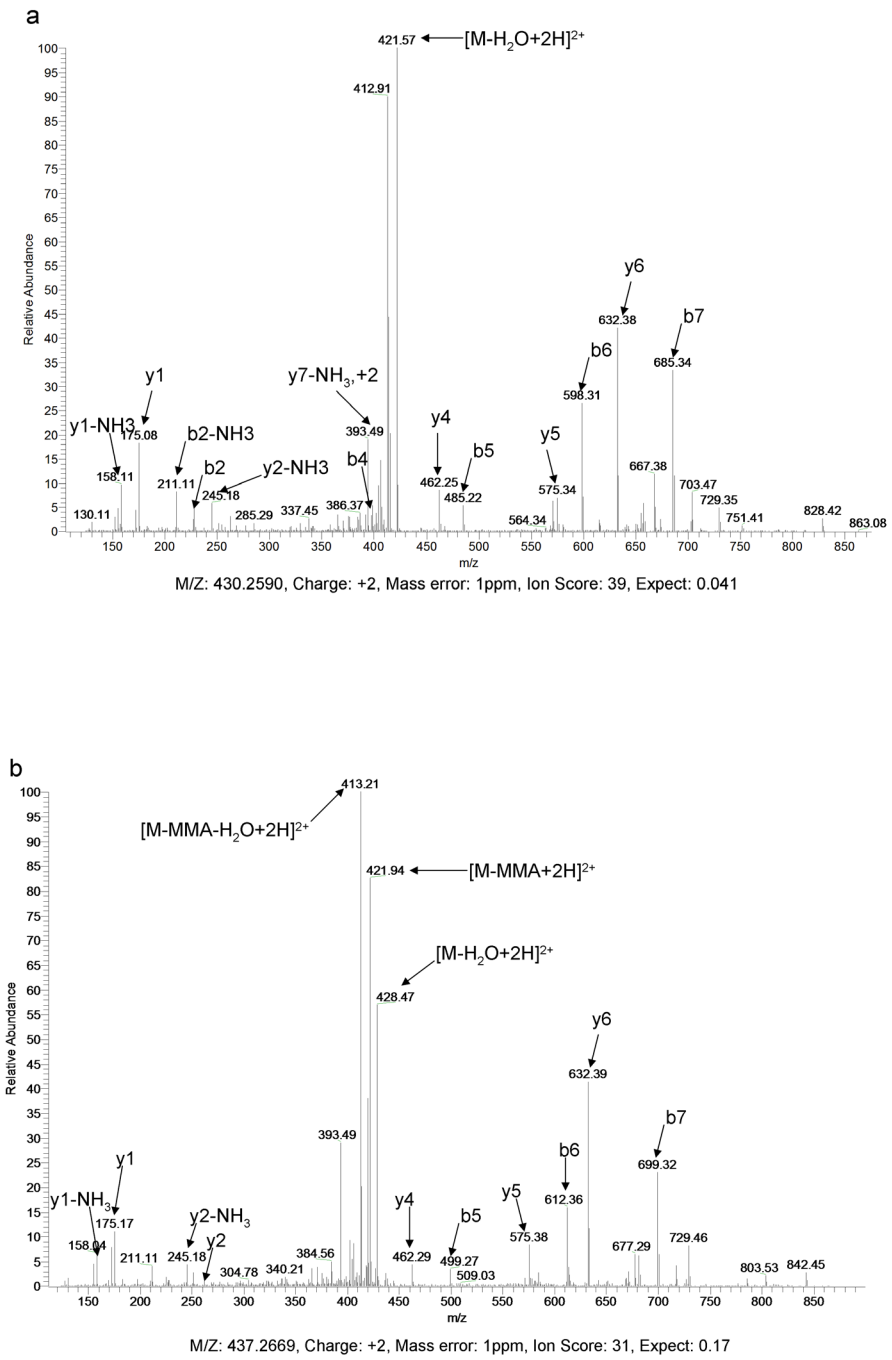
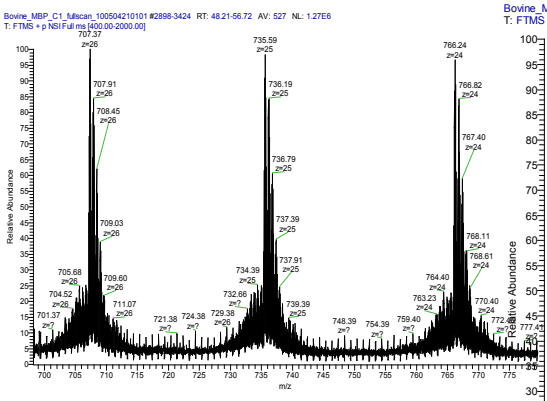
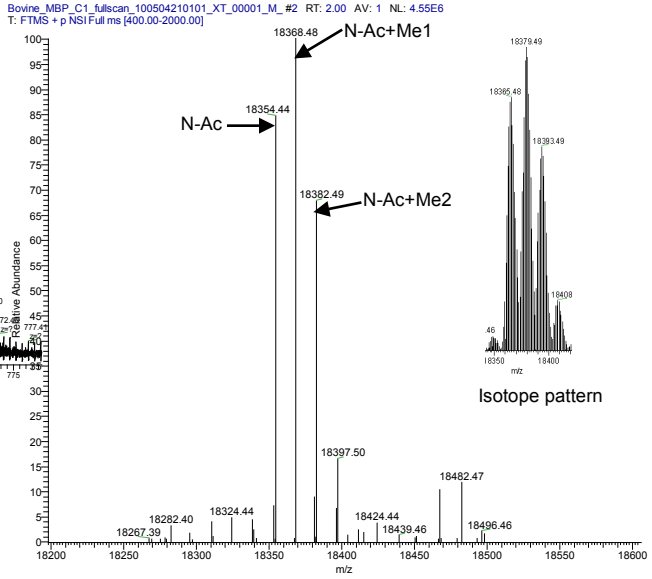


Figure 5.4 Identification of arginine methylation at position 106 in all five components
a) Monomethylated R106 in peptide sequence 105-G**R**GLSLSR-112. M/Z: 430.2590, Charge: +2, Mass error: 1ppm, Ion Score: 39, Expect: 0.041; b) Dimethylated R106 at the same position. The strong neutral losses of 31 Da (monomethylamine) and 49 Da (monomethylamine + water) correspond to peaks labeled [M-MMA +2H]+2 and [M-MMA -H2O +2H]+2 respectively, confirmed that R106 was a symmetrical dimethylation. M/Z: 437.2669, Charge: +2, Mass error: 1ppm, Ion Score: 31, Expect: 0.17.

a) C1 component

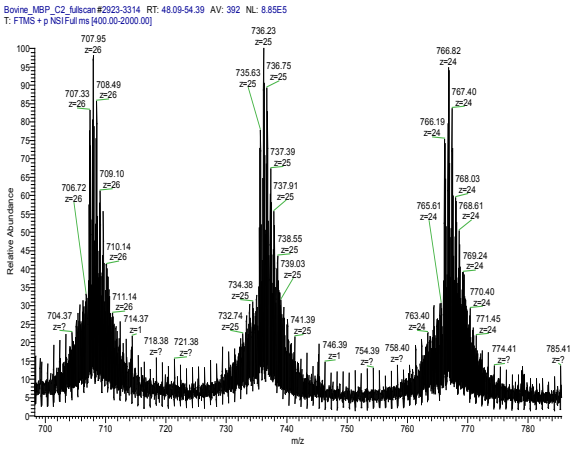


Unmod :Me1 :Me2=1 :1.18 :0.80

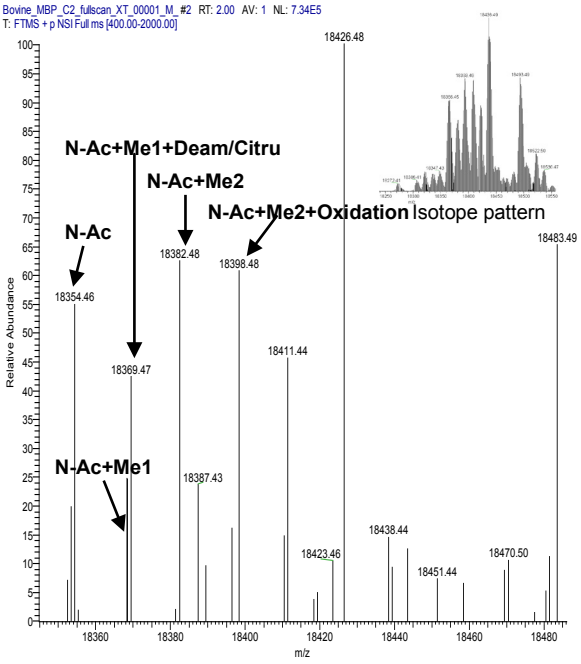


Monoisotopic mass

b) C2 component

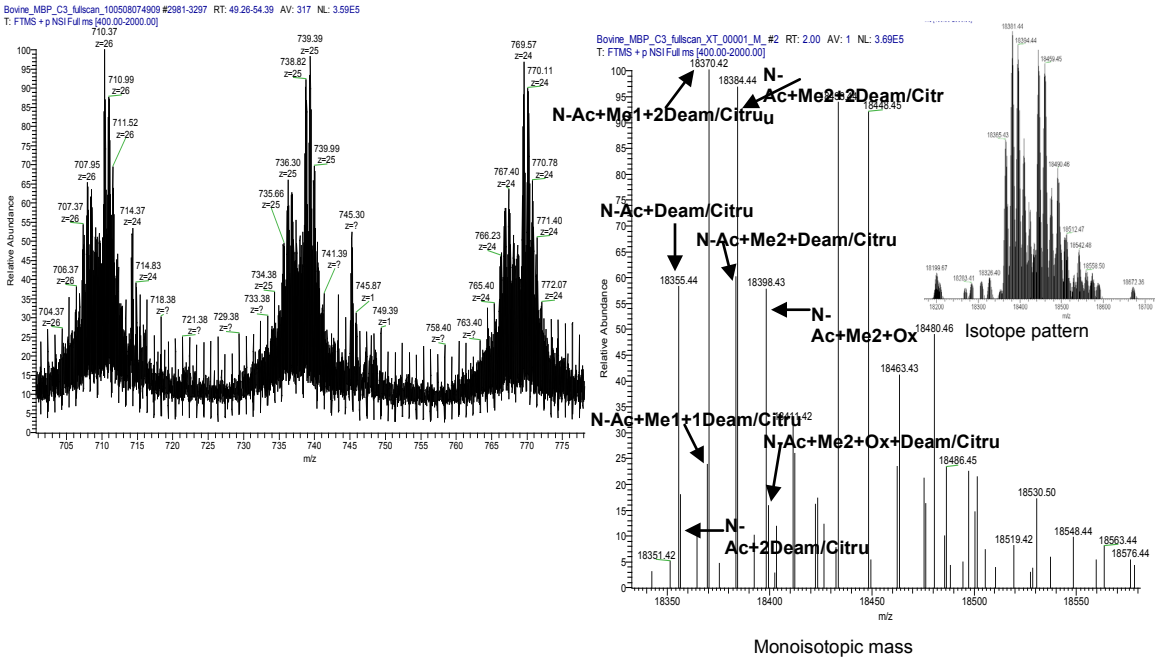


Unmod :Me1 :Me2= 1 :0.45 :1.14



Monoisotopic mass

c) C3 component



d) Rattlesnake MBP

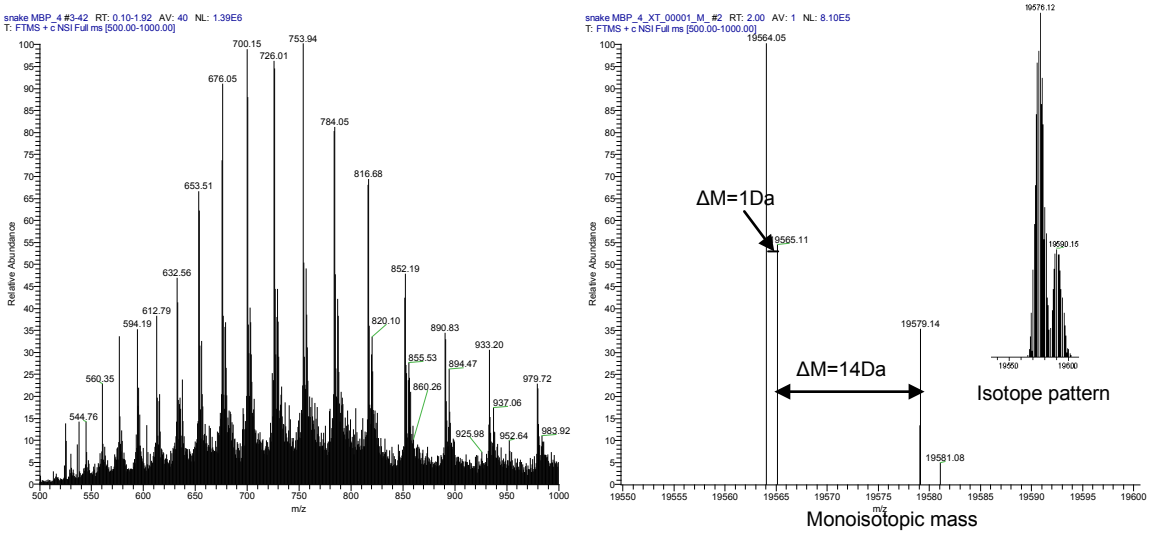


Figure 5.5 Intact mass analysis of 3 bovine MBP components and unfractionated rattlesnake MBP

a) C1 component: The mass difference between the observed monoisotopic mass of 18,354.44 Da and the theoretical monoisotopic mass of 18312.39Da is +42.05 Da which is consistent with the N-terminal acetylation observed at the peptide level. The mass increases of 14.04 Da and 28.05 Da are also consistent with the masses of mono- and dimethylated forms. The most abundant peak (Mw=18,368.48Da) in the C1 fraction corresponds to MBP modified with both N-terminal acetylation and monomethylation. The ratio of unmethylated: monomethylated: dimethylated form was 1:1.18 :0.80 in C1. Me1 and Me2 indicate the monomethylated and dimethylated forms respoectively. N-Ac, Deam, Citru, and Oxidation refer to N-terminal acetylation, deamidation of glutamine or asparagine, conversion of arginine to citrulline, and oxidation of methionine, respectively. b) C2 component: all the ions found in the C1 fraction were also present in C2 fraction. However, the most abundant peak in C2 corresponds to the protein modified with N-terminal acetylation and dimethylation. Deamidation or citrullination was found in C2 as +0.99 Da mass shift was observed. An oxidized Met in C2 came from a peak with mass increased by +16 Da from 18,382.48Da. The ratio of the unmethylated: monomethylated: dimethylated forms in C2 was 1:0.45:1.14. c) C3 component: C3 component had the most complex spectrum arising from the combination of different PTMs. More deamidated/citrullinated forms were observed by mass increases of 1 or 2 Da. d) Rattlesnake MBP: intact mass analysis of rattlesnake MBP revealed a much simpler spectrum as compared with spectra of bovine MBP charge components. Rattlesnake MBP has an intact monoisotopic mass roughly around 19,564 Da and forms having masses consistent with deamidation, citrullination, and methylation were observed.

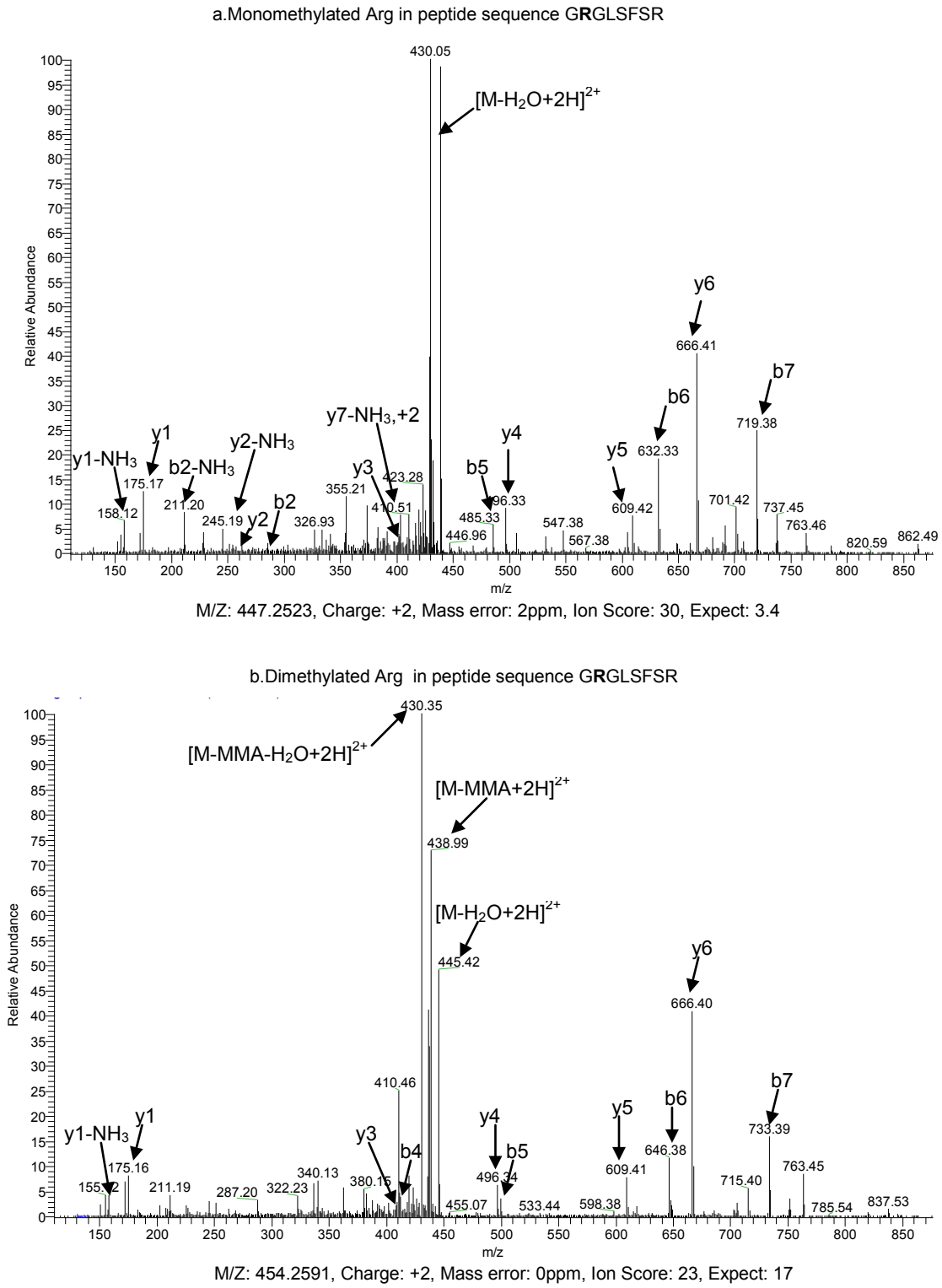


Figure 5.6 Identification of arginine methylation in rattlesnake MBP

a) Monomethylated arginine in a peptide with sequence G**R**GLSFSR. M/Z: 447.2523, Charge: +2, Mass error: 2ppm, Ion Score: 30, Expect: 3.4; b) Dimethylated Arginine at the same position. M/Z: 454.2591, Charge: +2, Mass error: 0ppm, Ion Score: 23, Expect: 17.

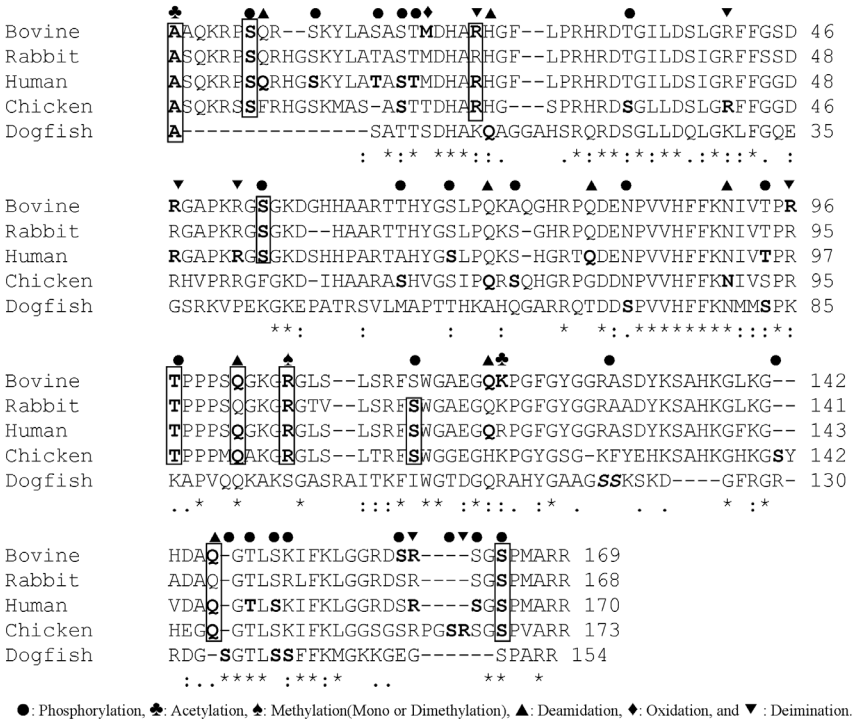


Figure 5.7 Comparison of MBP PTMs in different species

PTM information collected from previous studies and including our present data (human[Ref:12-14], bovine[Ref:6, 7, and present data], rabbit[Ref:8, 9], chicken[Ref:11], dogfish[Ref:10]). Sequences of mammalian MBPs were taken from the major, 18.5kd isoforms. Sequence alignment is based on CLUSTAL 2.1 multiple sequence alignment. Asterisks below the sequences indicate highly conserved residues in the primary structures. For dogfish S120 and S121 (italic), it was not possible to distinguish which residue was phosphorylated.

```
Bovine      AAQKRPSQR--SKYLASASTMDHARH--GFLPRHRDTGILDSLGRFFG-- 44
Rabbit      ASQKRPSQRHGSKYLATASTMDHARH--GFLPRHRDTGILDSIGRFFS-- 46
Human       ASQKRPSQRHGSKYLATASTMDHARH--GFLPRHRDTGILDSIGRFFG-- 46
Snake       ASQKRSSFRHGSK-LATASTIDHARH--GSPR-YREPSLLDSLGRFFG--
Anole       ASQKRSSFRHGSK-LASASIIDHARH--GSPRRYREPGLLDSLGRFFG-- 45
Gekko       ASQKRSSFRHGSK-LATASTIDHARH--GSPRRYREPSLLDSLGRFFG-- 45
Chicken     ASQKRSSFRHGSK-MASASTTDHARH--GSPR-HRDSGLLDSLGRFFG-- 44
Dogfish     ---------------ASATTSDHAKQAGGAHSRQRDSGLLDQLGKLFGQE 35

Bovine      SDRGAPKRGSGKDGHHAARTTHYGSLPQKAQGH----RPQDENPVVHFFK 90
Rabbit      SDRGAPKRGSGKD--HAARTTHYGSLPQKS-GH----RPQDENPVVHFFK 89
Human       GDRGAPKRGSGKDSHHPARTAHYGSLPQKS-HG----RTQDENPVVHFFK 91
Snake       GDKHYPR---GKDYHHASRVSHVSSLPHPR--------YVDENPVVHFFR
Anole       GDKHVPRR--GKDYHHAARASHVSSLPHQRSQR---GHYVDDNPVVHFFK 90
Gekko       GDKHVPRR--GKDYHHAARVSHVSSLPHPRSQHGRYGRYVDDNPVVHFFK 93
Chicken     GDRHVPRRGFGKDIH-AARASHVGSIPQ-RSQH---GRPGDDNPVVHFFK 89
Dogfish     GSRKVPEK--GKE--PATRSVLMAPTTHKAHQG--ARRQTDDSPVVHFFK 79

Bovine      NIVTPRTPPPSQGKGRGLS--LSRFSWGAEGQKPGFGYGGRASDYKSAHK 138
Rabbit      NIVTPRTPPPSQGKGRGTV--LSRFSWGAEGQKPGFGYGGRAADYKSAHK 137
Human       NIVTPRTPPPSQGKGRGLS--LSRFSWGAEGQRPGFGYGGRASDYKSAHK 139
Snake       NIVSPR------AKGRGLS--FSRFSWGGETHKPGYGSG-KFYEHKYAHK
Anole       SIVSPRSTPPMQAKGRGSP--FSRFSWGGETHKPGYGSG-KFFER---QK 134
Gekko       NIVSPRTPPPMQAKGRGLS--FSRFSWGGETPKHGYGSG-KFYEHKYAHK 140
Chicken     NIVSPRTPPPMQAKGRGLS--LTRFSWGGEGHKPGYGSG-KFYEHKSAHK 136
Dogfish     NMMSPKKAPVQQKAKSGASRAITKFIWGTDGQRAHYGAAGSSKSK----D 125

Bovine      GLKG--HDAQGTLSKIFKLGGRDSR----SGSPMARR 169
Rabbit      GLKG--ADAQGTLSRLFKLGGRDSR----SGSPMARR 168
Human       GFKG--VDAQGTLSKIFKLGGRDSR----SGSPMARR 170
Snake       GHKGYRYDGQGTLSK---LGSSGSRPGSR--------
Anole       GQKG--YDAQGTLSKIFKLGSSGSRPGSRSGSPVARR 169
Gekko       GHKGYHYDGQGTLSKLFKLGGSGSRPGSRHGSPIARR 177
Chicken     GHKGSYHEGQGTLSKIFKLGGSGSRPGSRSGSPVARR 173
Dogfish     GFRG-RRDGSGTLSSFFKMGKKGEG------SPARR- 154
```

Sequence Coverage:

150/177=85%(Gekko)

150/169=89%(Anole)

Figure 5.8 Rattlesnake MBP partial sequence and PTMs matched to other species

The rattlesnake MBP sequence matches best with gekko or anole MBP sequence with only a few amino acid substitutions. A total of 150 AAs has been interpreted which covers 85% of gekko and 89% of anole MBP sequence respectively. PTMs identified in rattlesnake are shown in bold.

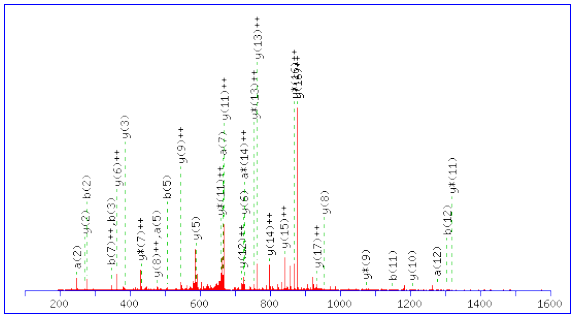


Figure S5.1 MS/MS Fragmentation of 12-YLASASTMDHA**R**HGFLPR-29
m/z: 677.6674, charge: +3, Deimination (R23), Ion Score: 40, Expect: 0.13

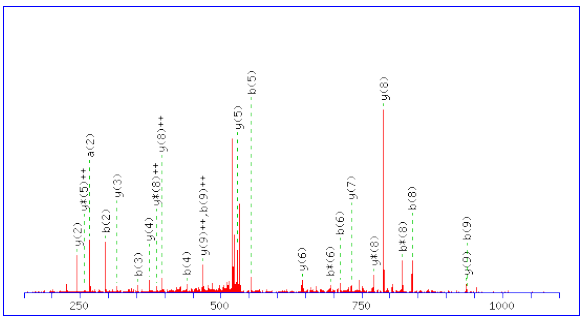


Figure S5.2 MS/MS Fragmentation of 42-FFGSD**R**GAPK-51
m/z: 541.7674, charge: +2, Deimination (R47); Ion Score: 55, Expect: 0.0014

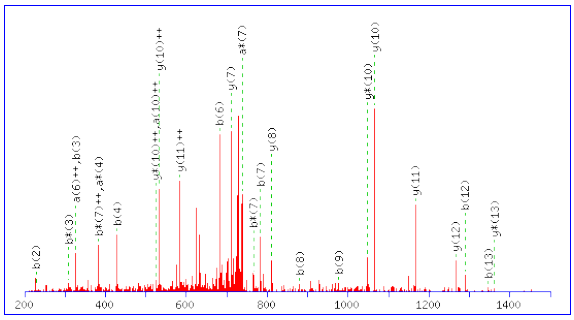


Figure S5.3 MS/MS Fragmentation of 91-NIVTP**R**TPPPSQGK-104
m/z: 746.9118, charge: +2, Deimination (R96), Ion Score: 56, Expect: 0.0012

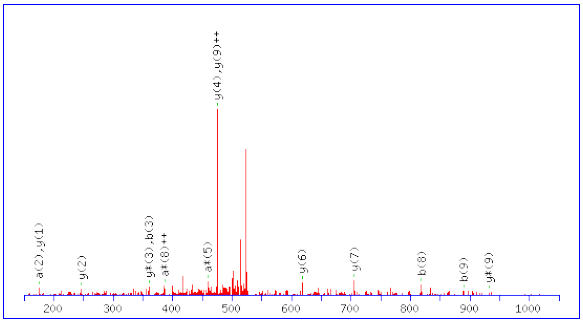


Figure S5.4 MS/MS Fragmentation of 159-DS**R**SGSPMAR-168
m/z: 532.7426, charge: +2, Deimination (R161) ,Ion Score: 24, Expect:0.77

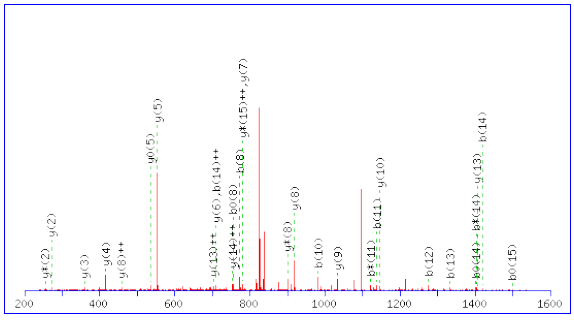


Figure S5.5 MS/MS Fragmentation of LATASTIDHA**R**HGSPR
m/z: 845.9372, charge: +2, Deimination (R11), Ion Score: 57, Expect: 0.0061

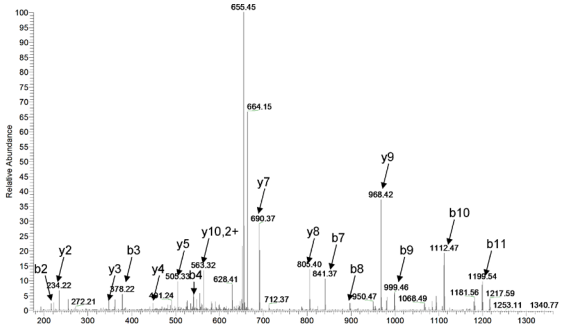


Figure S5.6 MS/MS Fragmentation of GY**R**YDGQGTLSK
m/z: 673.33, charge: +2, Deimination (R3)

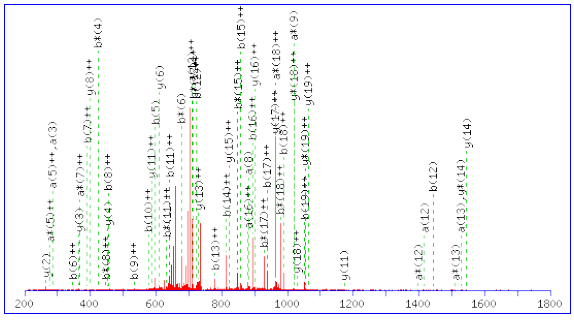


Figure S5.7 MS/MS Fragmentation of 1-**A**AQKRPSQRSKYLASASTMD-20
m/z: 746.7175, charge: +3, Acetyl (Protein N-term), Ion Score: 41, Expect: 0.022

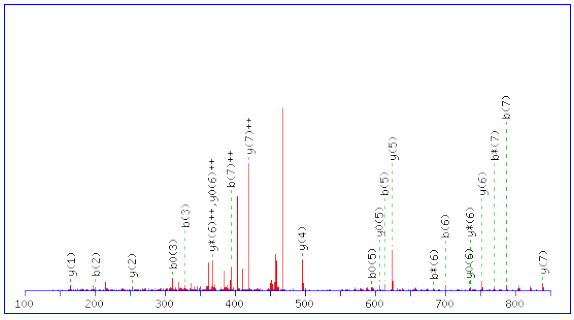


Figure S5.8 MS/MS Fragmentation of **A**SQKRSSF
m/z: 476.7465, charge: +2, Acetyl (Protein N-term), Ion Score: 37, Expect: 0.69

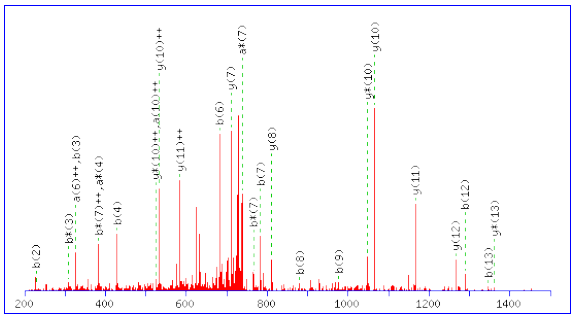


Figure S5.9 MS/MS Fragmentation of 91-NIVTPR**T**PPPSQGK-104
m/z: 786.4043, charge: +2, Phospho (T97), Ion Score: 32, Expect: 0.44

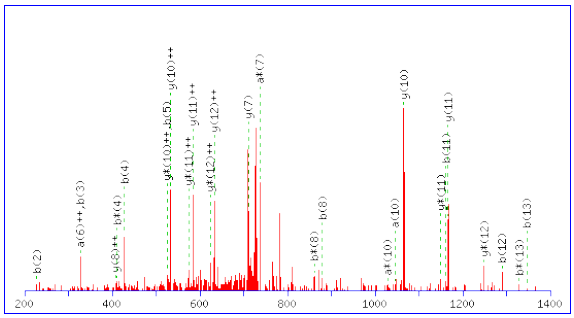


Figure S5.10 MS/MS Fragmentation of 91-NIVTPRTPPPS**Q**GK-104
m/z: 746.9102, charge: +2, Deamidation (Q102), Ion Score: 31, Expect: 0.41

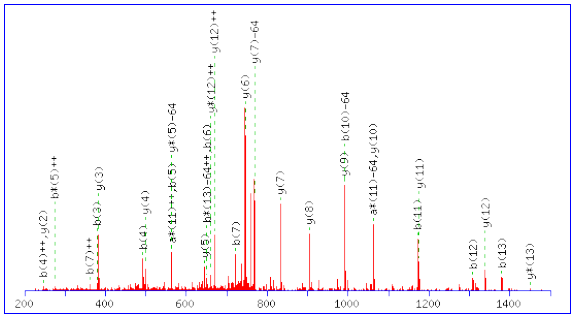


Figure S5.11 MS/MS Fragmentation of 10-SKYLASAST**M**DHAR-23
m/z: 777.3734, charge: +2, Oxidation (M19), Ion Score: 93, Expect: 3.6e-07

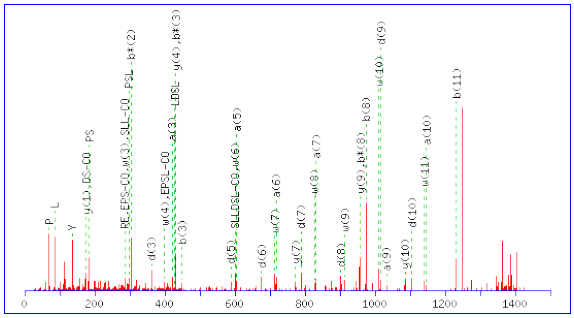


Figure S5.12 MS/MS Fragmentation of YREPSLLDSLGR
M/Z: 1404.7443, Ion Score: 65, Expect: 0.0085

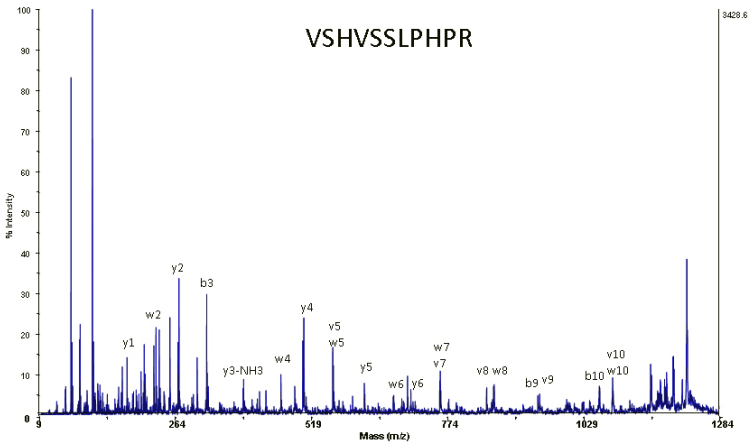


Figure S5.13 Differentiation of L/I from high-energy CID TOF/TOF
Note presence of v- and w- ions. W and V ions confirm L is present and not I

Figure S5.1-13 List of MS2 of peptides assigned with PTMs

5.6 References

[1] Eylar, H., Localization of methylated arginine in the A1 Protein from myelin. *Proceedings of the National Academy of Sciences of the United States of America* 1971, 68, 765–769.

[2] Zand, R., Li, M.X., Jin, X., Lubman, D., Determination of the sites of posttranslational modifications in the charge componentss of bovine myelin basic protein by capillary electrophoresis-mass spectroscopy. *Biochemistry* 1998, 37, 2441–9.

[3] Brostoffl, S.W., The proposed amino acid sequence of the P1 protein of rabbit sciatic nerve myelin. *Archives of Biochemistry and Biophysics* 1972, 153, 590–8.

[4] Martenson, R.E., Law, M.J., Deibler, G.E., Identification of multiple in vivo phosphorylation sites in rabbit myelin basic protein. *The Journal of biological chemistry* 1983, 258, 930–7.

[5] Zand, R., Jin, X., Kim, J., Wall, D.B., et al., Studies of posttranslational modifications in spiny dogfish myelin basic protein. *Neurochemical research* 2001, 26, 539–47.

[6] Kim, J., Zhang, R., Strittmatter, E.F., Smith, R.D., Zand, R., Post-translational modifications of chicken myelin basic protein charge components. *Neurochemical research* 2009, 34, 360–72.

[7] Baldwin, G.S., Carnegie, P.R., Isolation and partial characterization of methylated arginines from the encephalitogenic basic protein of myelin. *The Biochemical journal* 1971, 123, 69–74.

[8] Kim, J.K., Mastronardi, F.G., Wood, D.D., Lubman, D.M., et al., Multiple sclerosis: an important role for post-translational modifications of myelin basic protein in pathogenesis. *Molecular & Cellular Proteomics* 2003, 2, 453–62.

[9] Kim, J., Zand, R., Lubman, D.M., Electrophoretic mobility for peptides with post-translational modifications in capillary electrophoresis. *Electrophoresis* 2003, 24, 782–93.

[10] Berger, T., Rubner, P., Schautzer, F., Egg, R., et al., Antimyelin antibodies as a predictor of clinically definite multiple sclerosis after a first demyelinating event. *The New England journal of medicine* 2003, 349, 139–45.

[11] Mannie, M.D., Paterson, P.Y., U'Prichard, D.C., Flouret, G., Induction of experimental allergic encephalomyelitis in Lewis rats with purified synthetic peptides: delineation of antigenic determinants for encephalitogenicity, in vitro

activation of cellular transfer, and proliferation of lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America* 1985, 82, 5515–9.

[12] Gupta, M.K., Myelin basic protein and demyelinating diseases. *Critical reviews in clinical laboratory sciences* 1987, 24, 287–314.

[13] Meinl, E., Hohlfeld, R., Immunopathogenesis of multiple sclerosis: MBP and beyond. *Clinical & Experimental Immunology* 2002, 128, 395–397.

[14] Steinman, L., Assessment of animal models for MS and demyelinating disease in the design of rational therapy. *Neuron* 1999, 24, 511–4.

[15] Driscoll, B.F., Kira, J., Kies, M.W., Alvord, E.C., Mechanism of demyelination in the guinea pig. *Neurochemical Pathology* 1986, 4, 11–22.

[16] Agrawal, H.C., Banik, N.L., Bone, A.H., Cuzner, M.L., et al., The chemical composition of dogfish myelin. *The Biochemical journal* 1971, 124, 70P.

[17] Harauz, G., Libich, D.S., The classic basic protein of myelin--conserved structural motifs and the dynamic molecular barcode involved in membrane adhesion and protein-protein interactions. *Current protein & peptide science* 2009, 10, 196–215.

[18] Harauz, G., Ladizhansky, V., Boggs, J.M., Structural polymorphism and multifunctionality of myelin basic protein. *Biochemistry* 2009, 48, 8094–104.

[19] Martenson, R.E., Deibler, G.E., Kies, M.W., Levine, S., Alvord, E.C., Myelin Basic Proteins of Mammalian and Submammalian Vertebrates: Encephalitogenic Activities in Guinea Pigs and Rats. *The journal of immunology* 1972, 109, 262–70.

[20] Agrawal, H.C., O'Connell, K., Randle, C.L., Agrawal, D., Phosphorylation in vivo of four basic proteins of rat brain myelin. *The Biochemical journal* 1982, 201, 39–47.

[21] Paterson, P.Y., A Study of Experimental Encephalomyelitis Employing Mammalian and Nonmammalian Nervous Tissues. *The journal of immunology* 1957, 78, 472–5.

[22] Zand, R., *In Myelin Basic Protein*, Nova Science Publishers, Inc, New York 2008.

[23] Joss, J.L., Molloy, M.P., Hinds, L.A., Deane, E.M., Evaluation of Chemical Derivatisation Methods for Protein Identification using MALDI MS/MS. *International Journal of Peptide Research and Therapeutics* 2006, 12, 225–35.

[24] Park, Z., Sadygov, R., Clark, J.M., Clark, J.I., Yates, J.R., Assigning in vivo carbamylation and acetylation in human lens proteins using tandem mass

spectrometry and database searching. *International Journal of Mass Spectrometry* 2007, 259, 161–73.

[25] Lapko, V.N., Smith, D.L., Smith, J.B., Methylation and carbamylation of human γ-crystallins. *Protein Science* 2003, 12, 1762–74.

[26] Brame, C.J., Moran, M.F., McBroom-Cerajewski, L.D.B., A mass spectrometry based method for distinguishing between symmetrically and asymmetrically dimethylated arginine residues. *Rapid communications in mass spectrometry* 2004, 18, 877–81.

[27] Ishigami, A., Maruyama, N., Importance of research on peptidylarginine deiminase and citrullinated proteins in age-related disease. *Geriatrics & gerontology international* 2010, 10 Suppl 1, S53–8.

[28] György, B., Tóth, E., Tarcsa, E., Falus, A., Buzás, E.I., Citrullination: a posttranslational modification in health and disease. *The international journal of biochemistry & cell biology* 2006, 38, 1662–77.

[29] Harauz, G., Musse, A. a, A tale of two citrullines--structural and functional aspects of myelin basic protein deimination in health and disease. *Neurochemical research* 2007, 32, 137–58.

[30] Pritzker, L.B., Joshi, S., Gowan, J.J., Harauz, G., Moscarello, M. a, Deimination of myelin basic protein. 1. Effect of deimination of arginyl residues of myelin basic protein on its structure and susceptibility to digestion by cathepsin D. *Biochemistry* 2000, 39, 5374–81.

[31] Pritzker, L.B., Joshi, S., Harauz, G., Moscarello, M. a, Deimination of myelin basic protein. 2. Effect of methylation of MBP on its deimination by peptidylarginine deiminase. *Biochemistry* 2000, 39, 5382–8.

[32] Glozak, M. a, Sengupta, N., Zhang, X., Seto, E., Acetylation and deacetylation of non-histone proteins. *Gene* 2005, 363, 15–23.

[33] Hwang, C.-S., Shemorry, A., Varshavsky, A., N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* 2010, 327, 973–7.

[34] Kim, Y.H., Berry, a H., Spencer, D.S., Stites, W.E., Comparing the effect on protein stability of methionine oxidation versus mutagenesis: steps toward engineering oxidative resistance in proteins. *Protein engineering* 2001, 14, 343–7.

[35] Wood, M.J., Helena Prieto, J., Komives, E. a, Structural and functional consequences of methionine oxidation in thrombomodulin. *Biochimica et biophysica acta* 2005, 1703, 141–7.

[36] Wang, Z.Y., Shimonaga, M., Muraoka, Y., Kobayashi, M., Nozawa, T., Methionine oxidation and its effect on the stability of a reconstituted subunit of the light-harvesting complex from Rhodospirillum rubrum. *European journal of biochemistry* 2001, 268, 3375–82.

[37] Wong, Y.Q., Binger, K.J., Howlett, G.J., Griffin, M.D.W., Methionine oxidation induces amyloid fibril formation by full-length apolipoprotein A-I. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107, 1977–82.

[38] Sigalov, a B., Stern, L.J., Oxidation of methionine residues affects the structure and stability of apolipoprotein A-I in reconstituted high density lipoprotein particles. *Chemistry and physics of lipids* 2001, 113, 133–46.

[39] Hoshi, T., Heinemann, S.H., Faculty, M., Schiller, F., et al., Regulation of cell function by methionine oxidation and reduction. *The Journal of physiology* 2001, 531, 1–11.

[40] Ma, B., Lajoie, G., De novo interpretation of tandem mass spectra. *Current protocols in bioinformatics* 2009, Chapter 13, Unit 13.10.

[41] Strahl, B.D., Allis, C.D., The language of covalent histone modifications. *Nature* 2000, 403, 41–45.

[42] Jenuwein, T., Allis, C.D., Translating the Histone Code. *Science* 2001, 293, 1074–1080.

CHAPTER 6

Concluding remarks and perspectives

This dissertation studies the biologically important covalent chemical modifications of proteins called post-translational modifications (PTMs) that specifically modify most key proteins via enzymatic mechanisms. Histones and myelin basic protein, two classes of small, basic proteins that are rich in PTMs, are extensively explored in my thesis.

In the histone project, I present a robust MS-based quantitative proteomics method for studying histone PTMs in the model organism, *Tetrahymena thermophila*, that uses ^{15}N metabolically labeled histones as an internal reference. The construction and application of this method led us to successfully identify Txr1p as a novel putative histone methyl transferase (HMT) which preferentially monomethylates histone H3K27. I also elucidated its relationship with Ezl2p, another HMT in the same organism, which facilitates the build-up of H3K27 di- and tri-methylation. The *in vivo* uniform labeling with chemical derivatization (propionylation) allowed us to identify a large number of post-translational modifications with fairly small statistical variations. The application of this technique in WT and 3 HMT knockout cells (Δ*TXR1*, Δ*EZL2*, Δ*TXR1:EZL2*) further identified 66 PTMs in core histones, and revealed several important PTMs that exhibit cross-talk with methylation of H3K27. Hyper-acetylation of histones H2A and H4 in their N-terminal domains is one of the most noticeable consequences of decreased

methylation of H3K27. Hypo-phosphorylation in the H2B C-terminal region is yet another potentially important crosstalk related to H3K27 methylation.

The relationships between PTMs in the four core histones, however, is a key step to understanding the dynamic, combinatorial histone code. By virtue of the same analytical method, I was also able to quantify 40 PTMs and their unmodified counterparts across multiple epigenetic features in core histones and their variants. Clustering analysis of histone PTM data identified 5 functionally related subgroups, some of which were discussed in detail. I also proposed a 5-factor statistical model which aims to reveal the hidden features underlying the perturbed chromatin profiles suggesting that PTMs or chromatin should not be simpled classified as a binary code: 'active' or 'repressive' mark as proposed in the traditional model. These studies taken together greatly expand our understanding of the 'histone code' hypothesis and provide clues to the mechanisms behind the robustness and complexity of the histone code.

Myelin basic protein (MBP) shares many similarities with histones. Both are relatively small, strongly basic, and evolutionally conserved. They also share many of the same PTMs, including methylation, acetylation, phosphorylation, and citrullination. I demonstrated the first evidence for the acetylation of the epsilon-amino group of lysine, providing another example of the similarities between the PTMs of MBP and histones. A detailed cross-species analysis of the MBP PTMs from bovine and rattlesnake brain tissue revealed the major differences in MBP modification patterns between mammals and reptiles. As a general rule, myelin basic protein undergoes a broader range of modifications in mammals than in lower vertebrates as suggested from the mass heterogeneity of the intact protein and bottom-up structure analysis. Some of the

functionally conserved PTMs in mammals were revealed and proposed in the thesis. The MBP project extended our knowledge about relationships between MBP PTMs and demyelinating diseases.

The complicated, combinatorial, and dynamic 'histone code' is directly linked to the compaction of eukaryotic chromatin, which determines its accessibility to transcription and replication factors. MBP may also develop a 'PTM' code which determines the formation and compaction of the myelin sheath-an insulating layer wrapping around the axon and required for proper functioning of the nervous system.

Histones and myelin basic protein are, nonetheless, not the only proteins adorned with a large number of covalent modifications in the cell. The increasing examples of such proteins and the growing number of PTMs led to the protein barcode hypothesis-a universal language build on a broad range of protein chemical modifications. Although verification and elucidation of the protein code is extremely challenging due to the limited number of quantitative data sets of protein PTMs, the extensive and intensive exploration of protein functions by taking advantage of rapid development of system biology will eventually make the code clearer in the future and by doing so will allow more complete predictive models of cell regulation.

LIST of RELEVANT PUBLICATIONS

Peer-reviewed articles

1. **Chunchao Zhang**, Angela K. Walker, Robert Zand, Mario A Moscarello, Jerry Mingtao Yan, Philip C Andrews. Myelin basic protein undergoes a broader range of modifications in mammals than in lower vertebrates. ***J Proteome Res***. 2012, 11, 4791-4802.
2. **Chunchao Zhang**, Anthony J Molascon, Shan Gao, Yifan Liu, Philip C Andrews. Quantitative proteomics reveals that the specific methyltransferases TXR1 and EZL2 differentially affect the mono-, di- and tri-methylation states of histone H3 lysine 27. ***Mol Cell Proteomics*** (Epub ahead of print)
3. **Chunchao Zhang**, Anthony J Molascon, Yifan Liu, Philip C Andrews. Histone-wide investigation of chromatin crosstalk in H3K27-specific methyl-transferases knockout cells by quantitative proteomics. (In preparation)
4. **Chunchao Zhang,** Anthony J Molascon, Yifan Liu, Philip C Andrews. Systematic analysis of histone modifications across 15 epigenetic features: crosstalk and beyond. (In preparation)
5. Shan Gao, Jie Xiong, **Chunchao Zhang**, Brian Berquist, Rendong Yang, *et al*. Deficiency in histone H3 lysine 27 mono-methylation leads to severe replication stress. (Collaborative work, Submitted)

Conference abstract

1. **Chunchao Zhang**, Robert Zand, and Philip Andrews. Bottom-up and Top-down Analysis of Post-translational Modifications in Myelin Basic Protein. **59th ASMS Conference**

Book chapters and reviews

1. **Chunchao Zhang,** Yifan Liu, Philip C Andrews. Quantification of histone modifications using uniform labeling. ***Methods*** (Invited Review, In Process)