N^6-Formylation of Lysine: A Pathological Secondary Modification of Proteins
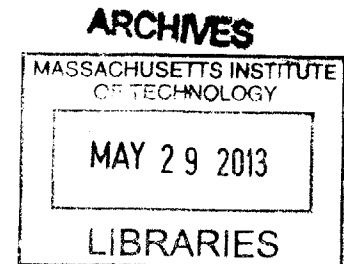
by

Bahar Edrissi

B.S. in Biomedical Engineering
University of Utah, 2003

SUBMITTED TO THE DEPARTMENT OF BIOLOGICAL ENGINEERING
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGICAL ENGINEERING

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2013

© 2012 Massachusetts Institute of Technology. All rights reserved.

Author...
Department of Biological Engineering
December 5, 2012

Certified by..
Peter C. Dedon
Professor of Toxicology and Biological Engineering
Thesis Supervisor

Accepted by..
Forest M. White
Associate Professor of Biological Engineering
Chairman, Department Committee on Graduate Theses

Committee members who voted in favor of thesis:

Advisor: Peter C. Dedon
Title: Professor of Toxicology and Biological Engineering

Chair: John M. Essigmann
Title: Professor of Chemistry and Biological Engineering

Other member: Steven R. Tannenbaum
Title: Professor of Chemistry and Biological Engineering

Other member: John S. Wishnok
Title: Senior Research Scientist

N^6-Formylation of Lysine: A Pathological Secondary Modification of Proteins

by

Bahar Edrissi

Abstract

There is increasing recognition that aberrant protein modifications play an important role in the pathophysiology of inflammation and oxidative stress in cells. We recently discovered that N^6-formylation of lysine is an abundant endogenous modification of histone and chromatin proteins. The high abundance of N^6-formyllysine in histone proteins and its chemical similarity to the biologically important N^6-acetyllysine has raised questions about its mechanisms of formation and biological consequences. Using novel ultrasensitive and specific liquid chromatography-coupled tandem mass spectrometry methods (LC-MS/MS) to quantify N^6-formyllysine lesions in proteins, we aimed to investigate the sources as well as the fate of this abundant endogenous protein modification. We present evidence that endogenous formaldehyde is a major source of N^6-formyllysine and that this adduct is widespread among proteins in all cellular compartments. We observed *in vitro* as well as *in vivo* that formaldehyde exposure leads to a dose-dependent increase in N^6-formyllysine protein adducts, with the use of isotopically-labeled formaldehyde to dissect endogenous from exogenous formaldehyde as sources of the adduct. Further, other isotope labeling studies revealed that lysine demethylation in histone proteins is not a source of N^6-formyllysine. With regard to N^6-formyllysine persistence in cells, our investigation of histone deacetylases revealed that despite chemical similarity of N^6-formyllysine to N^6-acetyllysine, the former is refractory to removal by histone deacetylases, which suggests that they will persist throughout the life of individual histone proteins. If not repaired, lysine formylation could accumulate to significant levels. The resemblance of N^6-formyllysine to N^6-acetyllysine, together with recent studies that mapped its location on many conserved lysine acetylation and methylation sites along histone proteins, support the idea that this abundant protein modification could interfere with normal regulation of gene expression, potentially leading to an epigenetic mechanism of disruption of cell function.

Thesis Supervisor: Peter C. Dedon
Title: Professor of Toxicology and Biological Engineering

Table of Contents

Chapter 1. Introduction

Chapter 2. Development of a liquid chromatography-coupled tandem mass spectrometry method for quantification of N^6-formyllysine adducts in proteins

Chapter 3. Quantitative analysis of histone modifications: Formaldehyde is a source of pathological N^6-formyllysine that is refractory to histone deacetylases

Chapter 4. Formaldehyde-induced formation of N^6-formyllysine protein adducts in rats

Chapter 5. Concluding remarks

Abbreviations

AGEs	advanced glycation end products
$CO_3^{\bullet-}$	carbonate radical anion
HATs	histone acetyl transferases
HDACs	histone deacetylases
HNE	4-hydroxy-2-nonenal
$HO^{\bullet}$	hydroxyl radical
H_2O_2	hydrogen peroxide
HOCl	hypochlorous acid
HPLC	high performance liquid chromatography
LC	liquid chromatography
LC-MS/MS	liquid chromatography-coupled tandem mass spectrometry
LSD1	lysine-specific demethylase 1
MDA	malondialdehyde
MRM	multiple reaction monitoring
MS	mass spectrometry
NO	nitric oxide
$NO_2^{\bullet}$	nitrogen dioxide radical
NO_2^{-}	nitrite
N_2O_3	nitrous anhydride
$^1O_2$	singlet oxygen
$O_2^{\bullet-}$	superoxide anion radical
O_3	ozone

ONOO$^-$	peroxynitrite
ONOOCO$_2^-$	nitrosoperoxycarbonate
PITC	phenylisothiocyanate
PTMs	post-translational modifications of proteins
QQQ	triple quadrupole mass spectrometer
SAHA	suberoylanilidehydroxamic acid
SOD	superoxide dismutase
SRM	selected reaction monitoring
TCE	trichloroethylene
TFA	trifluoroacetic acid

List of Figures

8

List of Tables

Acknowledgments

I consider myself very fortunate to have spent my graduate years at MIT, where I became part of a wonderful and vibrant graduate community and met many great individuals who guided and supported me both academically and personally.

First and foremost, I would like to thank my thesis advisor Professor Peter C. Dedon. I am truly grateful for his guidance and mentorship, his patience and understanding through numerous technical challenges of my project, and his support and generosity over the years, particularly during my health challenges. Pete's hard work, enthusiasm about science, and dedication to research is inspiring. It amazes me how he thinks up so many interesting ideas in even brief meetings. I learned a great deal from Pete and I feel I was very lucky to work with him and be part of his lab.

Next, I would like to thank my committee members, Professor Steve Tannenbaum, Professor John Essigmann, and Dr. John Wishnok for their very helpful discussions and suggestions. I could not have asked for a better committee and I am forever indebted to them for their support, advice, and encouragement.

I am very thankful to all the past and present members of Dedon Lab for their great friendship and help. They are all fantastic people and they made my everyday life in the lab very enjoyable. I also have many wonderful memories of Dedon Lab gatherings to cherish forever. I would like to extend a special thanks to Dr. Michael DeMott and Dr. Ramesh Indrakanti for keeping the laboratory running smoothly, and to our wonderful administrative staff, Olga Parkin and Kristine Marzilli, for their help and kindness over the years. I am also very grateful to Dr. Koli Taghizadeh for her tremendous help with analytical method development and for being very kind and caring. Koli has always believed in me and I am very thankful for her support and encouragement over the years. I'd also like to thank Laura Trudel for her help with establishing TK6 cell culture.

I am grateful to the BE Department, and particularly to Professor Douglas Lauffenburger for all his efforts, guidance, and support of all BE students. I also would like to thank the BE academic office, especially Dalia Fares and Aran Parillo for their great help and friendship.

Our collaboration with Professor James Swenberg's lab has truly been valuable for my dissertation. I really appreciate this opportunity, and I would like to thank them, especially Dr. Ben Moeller for his help and for promptly providing me with the animal tissues.

I am truly indebted to all my wonderful friends all over the world. I would like to especially thank my girlfriends Andrea Bryan and Anusuya Das for their great friendship that started during the first year at MIT and continues today. We have had many fun memories that I will cherish forever. I am also very grateful for the love and support of my extended family, in particular my aunts Shahpar Ghodsi and Nasrin Jalili. I would also like to thank my brother Afshin and his wife Melissa for their care and support. Of course, I could not have achieved this if it was not for the love, support, and sacrifices of

my parents Nastaran Jalili and Mansour Edrissi. I am truly at a loss for words to express how thankful I am for the sacrifices you have made to provide my brother and me with great educational opportunities. Thank you for all you have done!

In loving memory of my grandmother,

Mahin Shahabpour

Chapter 1

Introduction

Goals and summary of this thesis

This thesis project is focused on defining the mechanism of formation and biological significance of N^6-formyllysine, a recently discovered abundant endogenous secondary modification of proteins that resembles the physiologically important N^6-acetyllysine in histone proteins. This thesis is organized in 5 chapters. Chapter 1 presents a comprehensive review of protein secondary modifications with an emphasis on lysine acetylation, due to the resemblance of N^6-formyllysine to N^6-acetyllysine. The chapter subsequently discusses the role of reactive oxidants and electrophiles in generating pathological modifications of proteins and then focuses on N^6-formyllysine and its potential effects in terms of disrupting the epigenetic roles of histone modifications. Chapter 2 provides detailed development of an ultrasensitive and specific liquid chromatography-coupled tandem mass spectrometry method for quantifying this adduct in proteins. Topics such as distribution of this adduct among different histone classes, mechanisms for N^6-formyllysine generation in addition to the DNA oxidation pathway we originally reported, as well as its possible pathological consequences in cells are explored in Chapter 3, with the discovery of formaldehyde as a major endogenous source of N^6-formyllysine. Chapter 4 extends our studies to measurement of N^6-formyllysine protein adducts in tissues of rats exposed to formaldehyde by inhalation. In Chapter 5, we conclude this thesis by summarizing our results regarding sources as well as fate of N^6-formyllysine adducts and assess possible biological significance of this secondary protein modification, particularly in terms of interfering with regulatory roles of histones. N^6-Formylation of lysines in histones may represent an epigenetic mechanism of disruption of cell function leading to cancer and other diseases.

Post-translational modifications of proteins and their regulatory roles

In eukaryotes, the level of complexity of the proteome is several orders of magnitude greater than what is dictated by its genome sequence [1], with each gene effectively encoding several proteins. In addition to alternative RNA splicing that generates different mRNA transcripts from a single gene [2], the post-translational modification of proteins (PTMs) significantly enhances the complexity of the proteome [1,3]. This type of protein modification involves covalent alterations of the polypeptide backbone or side chains, leading to changes in a protein's structure and/or function [1,3,4]. In most cases, these modifications are reversible and under tight regulation of specific enzymes [5]. Eukaryotes dedicate a relatively high percentage of their genomes to enzymes modulating PTMs. For instance, in humans, greater than 5% of the genes encode such enzymes [3].

Protein modifications modulate protein-protein and protein-nucleic acid interactions and regulate the cellular environment by affecting a multitude of biological functions and processes such as protein activity, localization, turnover, signaling cascades, and cellular metabolism [1,5,6]. It is then of no surprise that dysregulation of PTMs leads to a plethora of disorders and diseases.

From more than 300 PTMs discovered to date [5], the most common are phosphorylation, acetylation, methylation, glycosylation, hydroxylation, palmitoylation, sulfation, and ubiquitination [1,2,5], as briefly discussed:

- **Phosphorylation.** Phosphorylation and dephosphorylation of serine, threonine, or tyrosine residues mediated by protein kinases (PKs) and phosphatases (PPs), respectively, is a key modification with many roles in regulating cellular

physiology and metabolism, mainly through modulating enzymatic activities and intracellular signaling pathways [2,7,8].

- **Methylation.** Another key PTM is methylation of lysine and arginine. Specific residues in histones as well as non-histone proteins such as p53, NF-κB, ERα are subject to methylation in cells [9]. In histones, lysine methylation is dynamically regulated by histone methyltransferases and histone demethylases, with important roles in gene expression and chromatin organization [10]. Protein methylation, in addition to DNA methylation/demethylation, has implications in embryonic and postnatal development [1,11,12].

- **Glycosylation.** Protein glycosylation, or the addition of sugar moiety to proteins, is an enzymatic mediated and reversible PTM [13], as opposed to the random non-enzymatic glycation that mostly renders proteins inactive [14]. Protein glycans formed by glycosylation have important structural and modulatory roles such as recognition elements for glycan binding proteins (GBPs), and therefore regulate a variety of processes such as development and survival [15].

- **Hydroxylation.** Hydroxylation of amino acids such as proline, asparagine, phenylalanine, tryptophan, and tyrosine is mediated by enzymes called hydroxylases [16,17]. Hydroxylation is critical in cellular detoxification reactions [1], as well as in the structural stability of proteins (e.g., hydroxyproline is an essential element of collagen) [18].

- **Palmitoylation.** Covalent attachment of fatty acids to proteins is another type of PTM. S-Palmitoylation of the cysteine thiol is a reversible modification with important physiological roles in G-protein-coupled receptor (GPCR) trafficking

17

and protein activity, as well as localization (as attachment of the hydrophobic moiety is important for membrane association) and T-cell signaling [19,20].

- **Sulfation.** The post-translational sulfation of tyrosine residues, a common modification of secreted and transmembrane proteins, has essential roles in chemokine signaling, leukocyte adhesion, immune response, extracellular protein–protein interactions, intracellular protein transport, protein activity and degradation [21-23].

- **Ubiquitination.** Ubiquitination is a covalent attachment of ubiquitin (Ub) to lysine residues of proteins in order to tag them for degradation and recycling. Attachment of Ub to proteins is a multi-step process, carried out in order with a set of 3 enzymes: Ub activating enzymes (E1), Ub conjugating enzymes (E2), and Ub ligases (E3) [24]. The Ub tag (mediated by mono-Ub or poly-Ub), is a crucial protein modification for proteasomal degradation, intracellular signaling cascades, protein binding, localization, and activity [25,26].

- **Acetylation.** Lysine acetylation is a key protein modification with inherent importance to this thesis due to the chemical similarity of lysine acetylation to lysine formylation and the critical role of lysine acetylation of histone proteins in regulating gene expression. To address this phenomenon, the next section of thesis introduction addresses histone protein structure and function followed by the biochemistry of histone acetylation.

Histone proteins and their post-translational modification

In eukaryotes, histone proteins are intimately associated with DNA and a plethora of other protein complexes forming chromatin [27,28]. The fundamental repeating unit of eukaryotic chromatin is nucleosome, an octamer of core histones (consisting of two H2A-H2B dimers and a H3-H4 tetramer) around which 147 base pairs of DNA are tightly wrapped in a left-handed super-helix 1.7 times around the core octamer [29-32]. The linker histone H1 binds the nucleosome and locks in place the DNA wrapped around the core histones [33-35]. Additionally, it binds to linker DNA (10-60 base pairs of DNA linking the nucleosomes) and allows for formation of the beads on a string structure and higher order assemblies, such as 30 nm solenoids and 100 nm fibers [32,34,36]. Chromatin is a dynamic structure, capable of unfolding and refolding [37], with two general states of tightly compact heterochromatin and the more open euchromatin that are in general associated with transcriptionally silent and active genomic regions, respectively [38,39].

In addition to serving as structural scaffolds for packaging DNA inside the nucleus, histones have regulatory roles affecting a variety of cellular processes such as chromatin organization and gene expression [33,40,41]. Besides the globular domains, all core histones have long (~20-35 residues) and highly conserved amino-terminal tails that protrude from the histone core domain [42] and compose ~25-30% of the mass of individual histones [43]. An exception is the histone H2A that has an additional long carboxyl terminal segment protruding from the nucleosome [37]. These tail domains may be structured or unstructured, may interact with nucleosomal or linker DNA, and participate in protein-protein interactions [44].

19

Histone tail segments, rich in basic amino acids, are subject to a variety of post translational modifications including acetylation, methylation, phosphorylation, ubiquitylation, and ADP ribosylation with some locations modified in more than one way [30,31,40,45,46]. For instance, acetylation can occur on lysines, methylation on lysines and arginines, phosphorylation on serines and threonines, and ubiquitylation and sumoylation on lysines. Just to add to the complexity, lysine can be mono-, di-, and tri-methylated, and arginine can be mono- and di-methylated. Most of these enormous secondary modifications identified to date occur on the long amino-terminal regions of histones, although there are some that occur on the main globular region [37]. Singly or in combination, these secondary modifications can act as "control switches" and play diverse regulatory roles in chromatin remodeling, regulation of transcription, DNA repair, and replication [40,47,48]. For instance, in higher eukaryotes, the combination of H3 K14 acetylation, H4 K8 acetylation, and H3 S10 phosphorylation usually signals transcriptional activation, and tri-methylation of H3 K9 and the lack of H3 and H4 acetylation are associated with transcriptional repression [37]. It is thought that through these covalent modifications the tail structure or contacts change, thus modulating histone:DNA interactions as well as creating new sites of dynamic interactions that can attract binding of regulatory proteins and affect the chromatin stability and DNA accessibility within the nucleosomes [40,44,49].

Most physiological histone modifications have been shown to be dynamic, reversible, and under regulation of histone modifying enzymes [30]. An exception is arginine demethylation during which arginine is changed into citrulline [30]. Lysine acetylation in histone proteins is among the best characterized PTMs [50], with inherent

Figure 1-1. Lysine formylation is a chemical homolog of lysine acetylation.
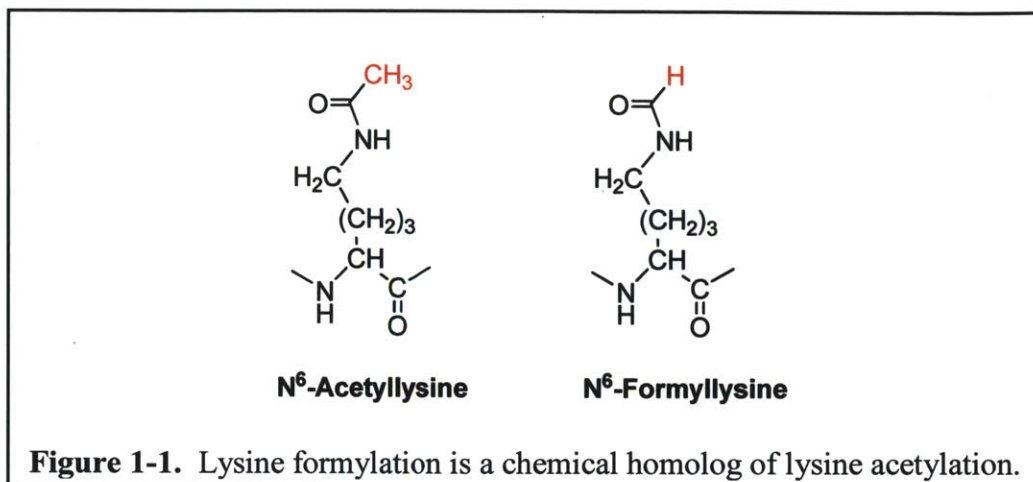
importance to this thesis due to chemical similarity of N^6-acetyllysine to N^6-formyllysine (Figure 1-1). The following section presents a detailed review of lysine acetylation.

Lysine acetylation is a key regulatory post-translational modification

Acetylation of the lysine ε-amino group is an abundant protein modification with key regulatory roles in transcription factors, histone proteins, and other chromatin proteins. Histone acetylation is among the best characterized PTMs [50] in its regulation of chromatin organization, gene expression, and DNA repair [36]. Lysine acetylation occurs at highly conserved sites along the tails of core histones and is under tight control of two sets of enzymes: histone acetyl transferases (HATs) and histone deacetylases (HDACs) (Figure 1-2) [51]. HATs transfer an acetyl group from acetyl coenzyme A to conserved lysine residues while HDACs catalyze the removal of the acetyl moiety [39,52]. In addition to disrupting the DNA/histone electrostatic interactions, acetyl lysine is recognized by the bromodomains of chromatin remodeling proteins, which in turn recruit other downstream factors involved in chromatin remodeling and DNA

Figure 1-2. Lysine acetylation is under tight control of two sets of enzymes: histone acetyl transferases (HATs) and histone deacetylases (HDACs).
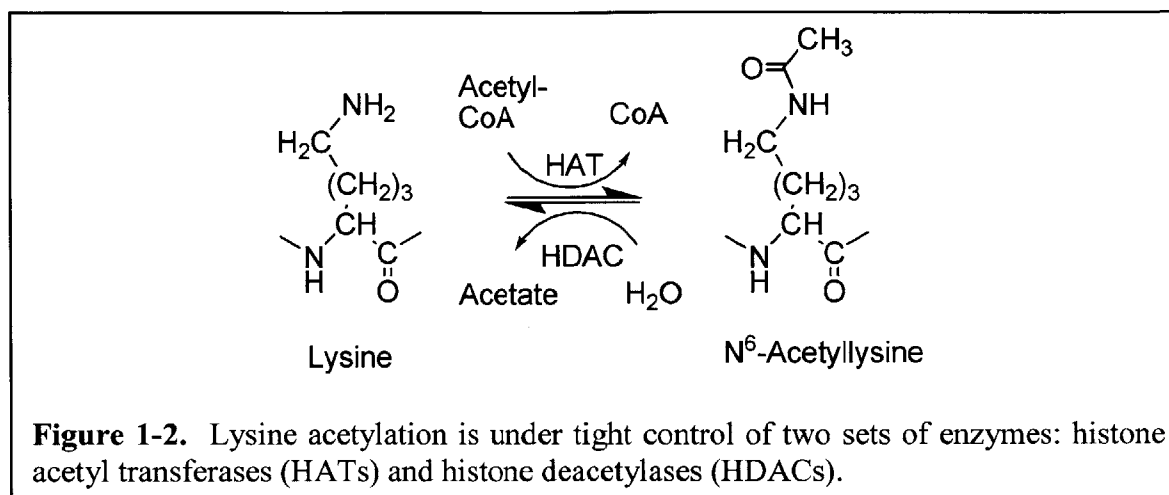
transcription [48,53]. Acetylation is almost exclusively associated with transcriptional activation [30]. The chemical similarities of N^6-formyllysine and N^6-acetyllysine (Figure 1-1) suggest a disruptive role for the former in signaling by histone acetylation. Indeed, N^6-formyllysine has been detected at conserved sites of lysine acetylation and methylation in histone proteins [35,54].

The state of acetylation in cells is also important in the pathophysiology of a variety of diseases including neurodegenerative diseases. For instance, it has been shown that β-amyloid hyperacetylation and acetylation of tau proteins are associated with Alzheimer's disease and dementia, respectively [1]. However, dysregulation of physiological and (in most cases) enzymatically controlled protein modifications are not the only ones affecting cellular physiology and disease. Protein modification by reactive oxidants and electrophiles constitute an important change to protein structure and/or function. The next sections examine some of the sources of these damaging agents and the protein modifications that they produce.

Production of reactive chemical species in cells

A wide range of physiological or pathophysiological processes in cells generate oxidants and electrophiles that directly damage biomolecules or lead to the formation of other reactive intermediates [55,56], thus affecting crucial cellular processes.

One source involves inflammation. As shown in Figure 1-3A, during inflammation, activated macrophages and neutrophils produce a variety of highly reactive oxygen and nitrogen species as part of the host defense system against pathogens. These highly reactive species alter cellular biomolecules through oxidation, nitrosation, nitration, and halogenations among other reactions [57]. The membrane-bound NADPH oxidase of these inflammatory cells catalyzes the transfer of an electron from NADPH to O_2, forming the superoxide anion radical ($O_2^{\bullet-}$) [58]. $O_2^{\bullet-}$ is not a strong oxidant, but it is a precursor to other reactive species [59]. Most cellular $O_2^{\bullet-}$ is converted to hydrogen peroxide (H_2O_2) via superoxide dismutase (SOD) [55,59]. A range of potent oxidants and oxidation/nitration intermediates are produced as H_2O_2 and $O_2^{\bullet-}$ react with other molecules (Figure 1-3A). H_2O_2 can be reduced to the highly reactive hydroxyl radical ($HO^{\bullet}$) by redox reactive metals (e.g., Fe^{2+} and Cu^{+1}) via Fenton chemistry. Reaction of $O_2^{\bullet-}$ with nitric oxide (NO) produced by activated macrophages in high concentration ($\leq 1\mu M$) [60,61] at sites of inflammation, yields peroxynitrite ($ONOO^-$). ONOOH (protonated form of $ONOO^-$) undergoes rapid homolysis ($t_{1/2}$ ~1 s) to produce $HO^{\bullet}$ and the weak oxidant, nitrogen dioxide radical ($NO_2^{\bullet}$) [55,57]. Reaction of $ONOO^-$ with carbon dioxide generates a potent nitrating agent, nitrosoperoxycarbonate ($ONOOCO_2^-$), that upon homolysis ($t_{1/2}$ ~ 50 ms) forms carbonate radical anion ($CO_3^{\bullet-}$) and $NO_2^{\bullet}$ [57].

Illustration by Jeff Dixon, copyright Peter C. Dedon.

$$O_2 \xrightarrow{1e^-} O_2^{\bullet-} \xrightarrow{SOD} H_2O_2$$

Figure 1-3. Generation of reactive oxygen and nitrogen species. Inflammation (A) and mitochondrial respiration (B), are two major sources for generation of reactive species. A and B are reproduced from [57] and [64], respectively.
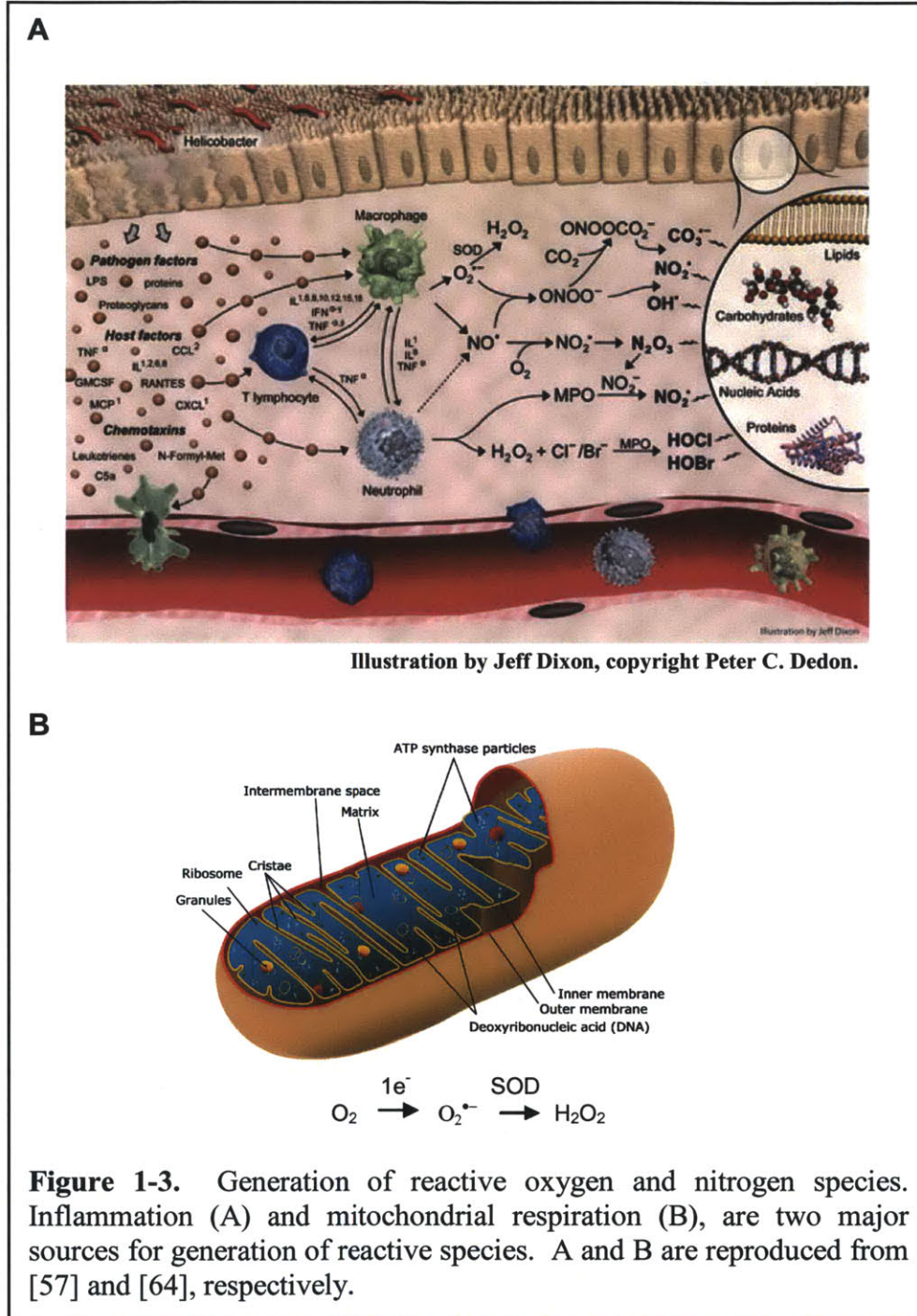
Another nitrosating agent, nitrous anhydride (N_2O_3), is derived from oxidation of NO. In activated neutrophils, myeloperoxidase-mediated reaction of H_2O_2 with Cl^- yields hypochlorous acid (HOCl), a strong oxidizing and halogenating agent [62].

Myeloperoxidase in these cells also mediates the conversion of nitrite (NO_2^-) to $NO_2^\bullet$ [63].

The other sources of reactive oxygen species include mitochondrial respiration (Figure 1-3B) and other O_2 metabolism pathways. The use of molecular O_2 for the cellular production of energy is responsible for the generation of many reactive oxygen species [65]. A variety of reactive molecules and free radicals are derived from O_2 such as $O_2^{\bullet-}$, $HO^\bullet$, H_2O_2, peroxyl radical ($ROO^\bullet$), ozone (O_3), and singlet oxygen ($^1O_2$). Reactive species such as $O_2^{\bullet-}$ and H_2O_2 are produced as a result of electron leakage from the mitochondrial electron transport chain during normal cellular metabolism [66,67]. Oxidoreductase enzymes such as NADPH oxidase, myeloperoxidase, and the cytochrome P450 enzymes also produce reactive species [68].

Reactive oxygen and nitrogen species have the capability of damaging molecules directly. However, damage caused by these reactive species (e.g., lipid peroxidation, carbohydrate oxidation, and direct amino acid oxidation), can yield electrophiles capable of causing further damage to biomolecules [55] (Figure 1-4). For instance, electrophilic species such as reactive aldehydes and epoxides are formed as a result of lipid peroxidation and can react with nucleophilic sites in DNA and proteins to yield adducts [55,69]. Among the products generated, the α,β-unsaturated aldehydes possessing two reactive sites are especially important since they can form cross-links or cyclic adducts with protein or nucleic acid [70-72]. Pathological modifications in proteins can be readily formed through reaction of nucleophilic amino acids lysine, histidine, and cysteine with reactive electrophiles generated from oxidation of polyunsaturated fatty

25

acids and carbohydrates such as malondialdehyde, 4-hydroxy-2-nonenal (HNE), 4-oxononenal, and glyoxal [73-77].

The next chapter covers in more detail these adventitious protein modifications due to reaction with primary and secondary reactive species. It should be noted that protein oxidative modification can result from other sources such as xenobiotics, cigarette smoke, acetaminophen, γ-irradiation, ultraviolet (UV) light, and ozone [68], among other agents. However, for the purpose of this thesis, we only focus on protein damage due to endogenously derived reactive oxidants and electrophiles.

Figure 1-4. Examples of reactive aldehydes generated from damage to biomolecules.
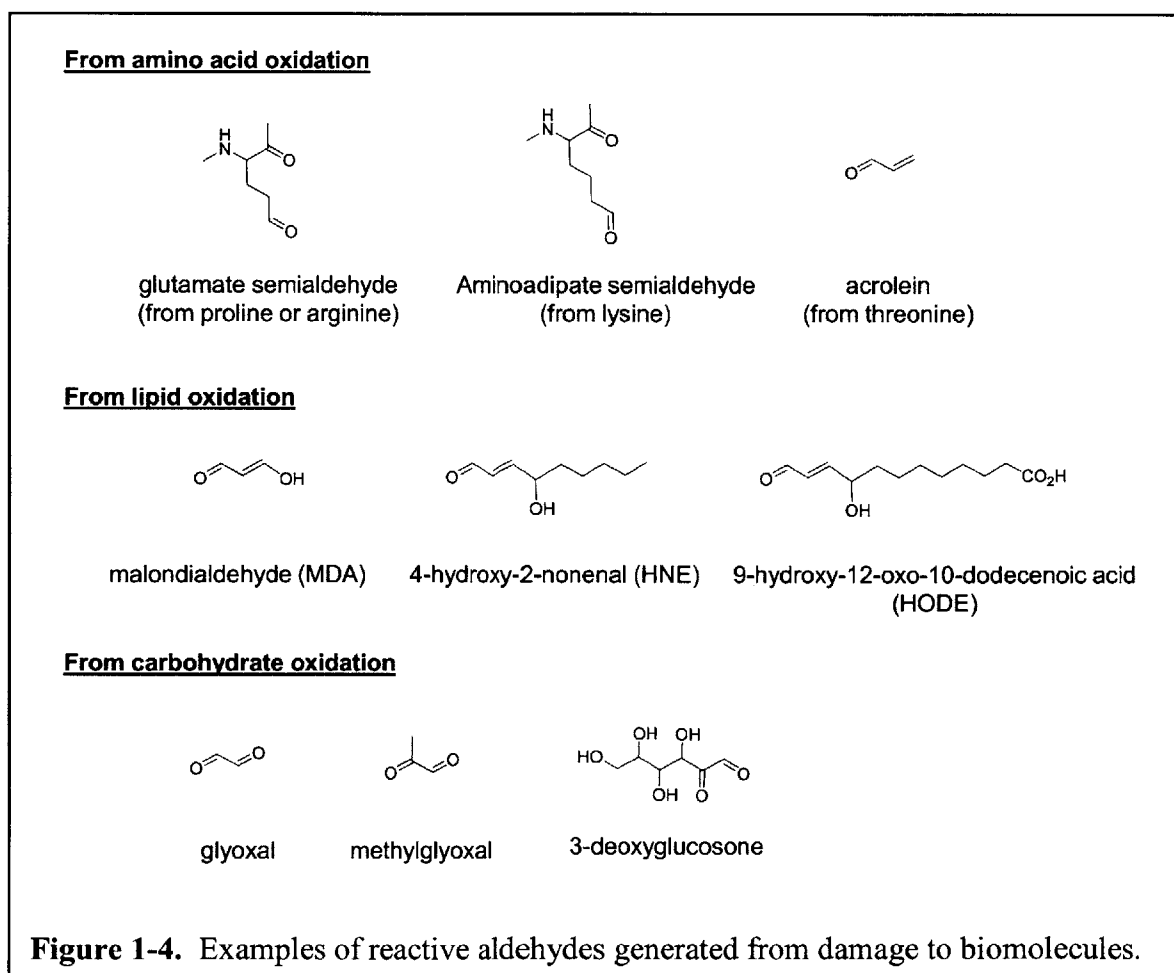
Protein damage due to reactive endogenous oxidants and electrophiles

In addition to physiological secondary modifications that are mostly enzymatically regulated, proteins are subjected to reactions with endogenous oxidants and electrophiles generated by oxidative stress, inflammation, and normal cell metabolic processes [77-80]. Protein targets of reactive species are vast and include structural and membrane proteins, metabolic and detoxification enzymes, or proteins involved in cell signaling and gene expression [55]. Protein modification caused by a single oxidant or electrophile, such as HNE can produce a diverse set of cellular responses with hundreds of genes upregulated and downregulated [58]. Here we review representative examples of the types of damage caused by these reactive species in cells.

Direct oxidation of amino acid side chains is one example of adventitious protein modification. Oxidative modification of proteins is usually specific to type and location of the residue and may occur on polypeptide backbone or the nucleophilic/redox-sensitive side chains. As mentioned previously, modifications of proteins affect their function in a positive or negative way, thus playing important roles in cellular physiology as well as progress of disease. Cysteine, tyrosine, and methionine are among the amino acids mostly modified by endogenous oxidants [55]. Cysteine sulfenic acid is formed via oxidation of the sulfhydryl group of cysteine, helping to absorb the oxidative insult [81]. Cysteine sulfenic acid is unstable and forms disulfide bonds with glutathione or other accessible thiols or undergoes further oxidation to cysteine sulfinic acid and cysteine sulfonic acid [82]. S-nitrosylation of cysteine and 3-nitration of tyrosine residues in proteins is another consequence of oxidative or nitrosative stress [83,84]. In fact, 3-nitrotyrosine is used as a biomarker of protein damage associated with inflammation and

variety of diseases [84]. Methionine oxidation to form methionine sulfoxide, reversed by methionine sulfoxide reductases, is speculated to serve as a defense mechanism against oxidative stress by preventing other residues from oxidative damage [85]. Elevated levels of oxidized aromatic amino acid residues (e.g., 3-nitrotyrosine, 3-chlorotyrosine, and 3,4-dihydroxylphenylalanine) are seen in age-related diseases such as neurodegenerative and cardiovascular diseases [55]. For instance, 3-nitrotyrosine adducts were detected in tyrosine hydroxylase isolated from brain tissues of Parkinson's disease model mice that exhibited reduced enzymatic activity [86]. Oxidation of proline, arginine, and lysine residues result in protein carbonyls such as glutamate and aminoadipate semialdehydes [73,78]. In general, protein carbonyls can be generated via direct damage to polypeptide backbone, oxidation of amino acids proline, lysine, arginine, threonine, glutamate, aspartic acid [1,55,73,78], or Michael addition of reactive α,β-unsaturated aldehydes to nucleophilic amino acids lysine, histidine, and cysteine [73-77]. Protein carbonylation, an irreversible modification, often leads to loss of protein function [87]. Indeed, total protein carbonyl content is regarded as a biomarker of oxidative stress and inflammation and elevated levels are associated with a variety of human diseases such as cardiovascular and neurodegenerative disorders, and processes such as aging [73,75,78].

Another adventitious protein modification involves reaction of amino acid side chains with physiological sugars and sugar oxidation products. Unlike the enzymatic mediated and site specific glycosylation of proteins, these advanced glycation end products (AGEs) are random events [1]. For instance, the reactive dicarbonyl sugar 3-deoxyglucosone, synthesized via the Maillard reaction and the polyol pathway [88],

results in formation of AGEs such as imidazolone, N^6-carboxymethyllysine, and pyrraline by reaction with amines [89] (Figure 1-5). AGEs have been shown to cause extracellular matrix dysfunction, to react with cell receptors and change cytokine and hormone levels, and to alter functions of a variety of intracellular proteins [90]. AGEs can also form crosslinks, especially in long lived proteins such as collagen and elastin [91]. The formation and accumulation of AGEs over time has been implicated in the development of diabetes as well as age-related degenerative processes by damaging blood vessels, connective tissue, ocular lens, and nerves [14,90].

Reactive aldehyde induced modifications of proteins typically arise by reaction of the nucleophilic side chains of lysine, cysteine, and histidine with reactive electrophiles such as malondialdehyde (MDA), 4-hydroxynonenal (HNE), acrolein, and glyoxal generated by oxidation of polyunsaturated fatty acids and carbohydrates, among other

Lipid peroxidation adducts

dihydropyridine-lysine

lysine-HNE
Schiff base

lysine-HNE
Michael adduct

cysteine -HNE
Michael adduct

AGEs

imidazolone

N^6-carboxymethyllysine

N^6-3-formyl-3,4-dehydropiperidino-lysine

Figure 1-5. Examples of protein adducts caused by reactive electrophiles.
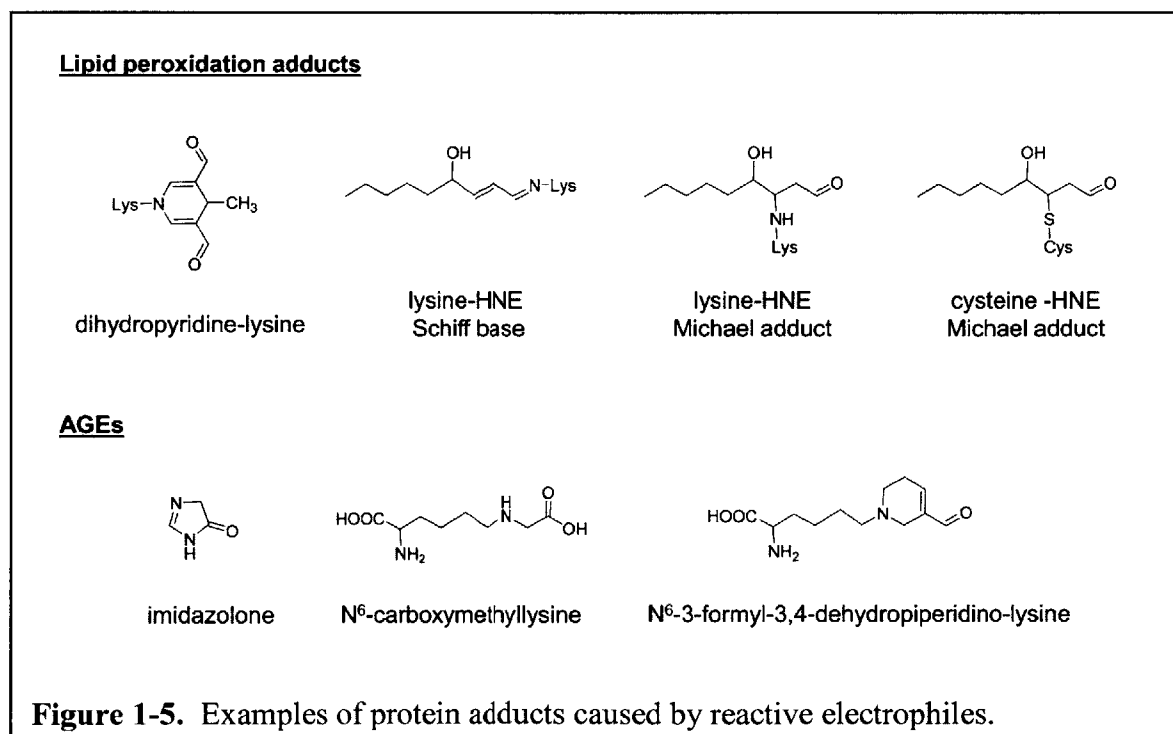
biomolecules [74,75,77,79,80]. Michael adducts are among the major adducts formed, for example, from the reaction of MDA with lysines residues (Figure 1-5) and 4-hydroxy-2-alkenals with lysine, cysteine, and histidine amino acids [55]. Sequential addition of two acrolein molecules to epsilon-amino group of lysine results in formation of cyclic N^6-3-formyl-3,4-dehydropiperidino-lysine adduct [92] (Figure 1-5). Schiff base adducts are formed from reaction of 4-hydroxy-2-alkenals [93] as well as formaldehyde with lysine residues [94]. Protein adducts generated by products of lipid peroxidation can alter protein function and thus affect the vast range of biological functions they regulate, leading to pathological processes associated with human diseases [77-80,95]. For instance, HNE induces cell toxicity by covalently modifying IκB kinase, thus inhibiting the subsequent transcription of NFκB-dependent genes needed for cell survival [58]. It was also shown that HNE results in loss of microtubule network in neuroblastoma cells by forming Michael adducts with tubulin isoforms [96].

Among adventitious protein adducts (Table 1-1 and Figure 1-5), N^6-formylation of lysine has recently emerged as an abundant protein modification [35,53,54,97]. N^6-Formyllysine is a chemical homolog of the biologically important N^6-acetyllysine and thus may interfere with acetylation signaling in cells. The following section is a more detailed introduction on this pathological protein modification that was originally observed in histones and chromatin proteins.

Table 1-1. Examples of oxidative modification of proteins. Adapted from [68].

Oxidative Modification	Amino Acids Residue
Disulfides, glutathiolation	Cys
S-Nitrosocysteine	Cys
3-Nitrotyrosine	Tyr
3-Chlorotyrosine	Tyr
Methionine sulfoxide	Met
Cross-links, aggregates	Several
Lipid peroxidation adducts	Lys, Cys, His
Carbonyls	All (Lys, Arg, Pro, Thr)
Amino acid oxidation	Lys, Cys, His

N^6-Formyllysine as an adventitious protein modification

Emerging evidence points to widespread modification of cellular proteins by N^6-formylation of lysine as a result of adventitious reactions with endogenous electrophiles [35,53,54,97]. These adducts were originally described in histones and other chromatin proteins [35,53,54], although they have since been identified as adducts arising in proteins subjected to nitrosative and oxidative stresses [97].

We have previously shown the generation of N^6-formyllysine adducts in histone proteins as a result of DNA oxidation [53]. Histones are rich in nucleophilic amino acids, such as lysine and arginine that can react with pathogenic electrophiles generated from oxidized DNA, proteins, and lipids [53]. Our previous study in which DNA was labeled with 5' tritiated thymidine followed by cell treatment with neocarzinostatin (an enediyne

31

Figure 1-6. Generation of N^6-formyllysine as a result of DNA oxidation. 3'-Formylphosphate residues are generated during 5'-oxidation of 2-deoxyribose in DNA (A). The formyl moiety can be transferred to lysines in histone proteins to form N^6-formyllysine (B) [53].
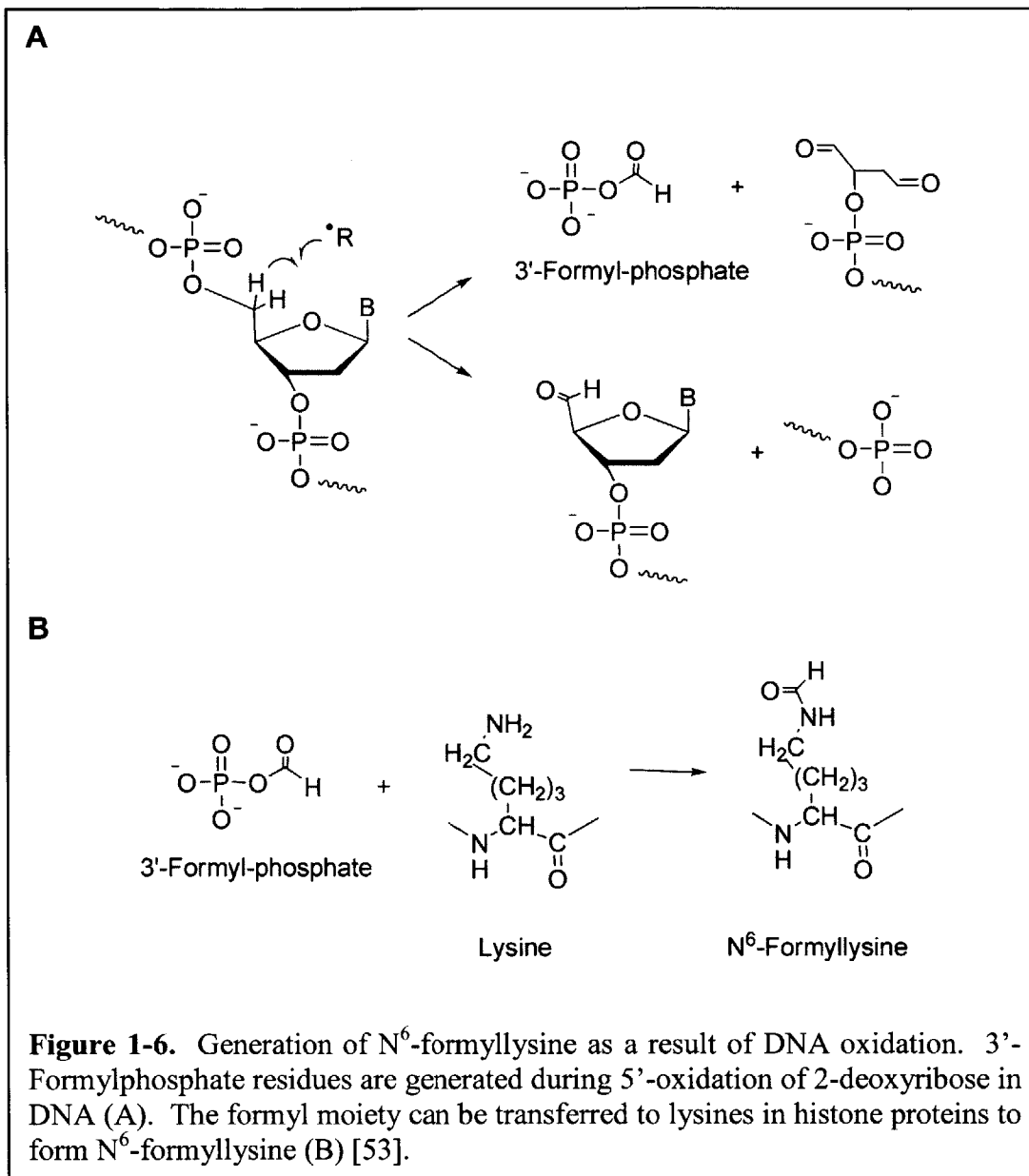
antibiotic that abstracts a 5'-H from 2-deoxyribose) showed that lysines in histones can react with 3'-formylphosphate residue, an electrophile generated from 5'-oxidation of 2-deoxyribose in DNA, to form N^6-formyllysine (Figure 1-6) [53].

N^6-Formyllysine was reported to be the major adduct formed when proteins were reacted with trichloroethylene (TCE) oxide *in vitro* [98], and a recent study points to its formation in proteins treated with biological oxidant, peroxynitrite [97], suggesting other

sources for its generation besides DNA oxidation. In 1985, N^6-formyllysine was reported to be formed by reaction of l-lysine with formaldehyde [99] and recently, it was shown it could result from silver staining procedures that involve the use of formaldehyde [100]. Considering that formaldehyde reacts with amines to give a carbinolamine intermediate that is only one oxidation state away from a formamide functional group, we hypothesized that endogenous formaldehyde could serve as a major source of N^6-formyllysine residues in histone and other proteins.

Formaldehyde is a reactive aldehyde forming adducts with nucleophilic sites in DNA and protein molecules resulting in products such as N^2-hydroxymethyl-dG DNA adducts [101] and formaldehyde induced Schiff bases on lysine residues of histones [94], in addition to protein-DNA crosslinks. It has been classified as a known human carcinogen by the International Agency for Research on Cancer in 2005 [102] and by the National Toxicology Program of the National Institute of Environmental Health Sciences in 2011 [103], although it is still considered a probable human carcinogen by the U.S. Environmental Protection Agency [104]. In addition to environmental and occupational sources [101,105,106], formaldehyde arises endogenously from cellular processes such as demethylation of histones and nucleic acids as well as biosynthesis of purines, thymidine, and some amino acids [10,105,107]. Formaldehyde is a relatively abundant metabolite at concentrations ranging from 13 to 97 µM in human plasma [105]. Thus, formaldehyde induced N^6-formyllysine generation in proteins could be the major source of this adduct compared to previous sources reported.

The relatively high abundance of N^6-formyllysine adducts in histones [35,53] suggests it could interfere with the regulatory roles of histones. The following section

examines the potential disruptive role of N^6-formyllysine in cellular epigenetics.

The biological implications of N^6-formyllysine modification of histones

N^6-Formyl lysine has been reported to be a major protein adduct in histones occurring on many conserved lysine residues involved in epigenetic regulations [35,53]. Even the DNA oxidation pathway in itself can result in high abundance of this adduct in histones. Indeed, there are thousands of oxidative events in a cell on a daily basis, resulting in hundreds of deoxyribose oxidations in DNA with each site producing different electrophilic products, leading to formation of 10's of N^6-formyllysines per cell per day [53,108]. As discussed later in this thesis, formaldehyde is found to be another major source giving rise to this adduct in histones. Therefore, if not removed, this pathological modification may accumulate to significant levels in cells.

The high rate of occurrence of N^6-formyllysine in histones raises the possibility of epigenetic interferences through affecting the regulatory roles of these proteins. In chromatin proteins, N^6-formyllysine has the potential to interfere with important physiological modifications that perform signaling functions. Indeed, N^6-formyllysine has been detected at conserved sites of lysine acetylation and methylation in histones [35,54]. In addition, the chemical similarities of N^6-formyllysine and N^6-acetyllysine suggest a disruptive role for the former in signaling by histone acetylation. Thus, this thesis project is focused on defining the mechanism of formation and biological significance of N^6-formyllysine. To that end, we developed an ultrasensitive and specific liquid chromatography-coupled tandem mass spectrometry method for quantifying this adduct in proteins (Chapter 2), and used it to answer questions about its mechanisms of

formation and pathological consequences in cells (Chapter 3). Chapter 4 extends our studies to measurement of N^6-formyllysine protein adducts in tissues of rats exposed to formaldehyde by inhalation, as we discovered formaldehyde to be a major endogenous source of N^6-formyllysine (Chapter 3). We conclude this thesis in Chapter 5 by summarizing our results regarding sources as well as fate of N^6-formyllysine adducts and assess possible biological significance of this secondary protein modification. The presence of N^6-formylation of lysines in histones may have important biological consequences leading to pathology and progress of variety of diseases.

References

1.   Karve TM, Cheema AK (2011) Small changes huge impact: the role of protein posttranslational modifications in cellular homeostasis and disease. J Amino Acids 2011: 207691.

2.   Seet BT, Dikic I, Zhou MM, Pawson T (2006) Reading protein modifications with interaction domains. Nat Rev Mol Cell Biol 7: 473-483.

3.   Walsh CT (2006) Posttranslational Modification Of Proteins: Expanding Nature's Inventory: Roberts and Company Publishers.

4.   Boeckmann B, Blatter MC, Famiglietti L, Hinz U, Lane L, et al. (2005) Protein variety and functional diversity: Swiss-Prot annotation in its biological context. C R Biol 328: 882-899.

5.   Jensen ON (2000) Modification-specific proteomics: systematic strategies for analysing post-translationally modified proteins. Trends Biotechnol 18: 36-42.

6.   Doyle HA, Mamula MJ (2001) Post-translational protein modifications in antigen recognition and autoimmunity. Trends Immunol 22: 443-449.

7.   Hunter T (1995) Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. Cell 80: 225-236.

8.   Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 294: 1351-1362.

9.   Yang XD, Lamb A, Chen LF (2009) Methylation, a new epigenetic mark for protein stability. Epigenetics 4: 429-433.

10.   Shi Y, Whetstine JR (2007) Dynamic regulation of histone lysine methylation by demethylases. Mol Cell 25: 1-14.

11.   Mann MR, Bartolomei MS (2002) Epigenetic reprogramming in the mammalian embryo: struggle of the clones. Genome Biol 3: REVIEWS1003.

12.   Litt MD, Simpson M, Gaszner M, Allis CD, Felsenfeld G (2001) Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. Science 293: 2453-2455.

13.   Varki A, Esko JD, Colley KJ (2009) Cellular Organization of Glycosylation. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY).

14.   Ulrich P, Cerami A (2001) Protein glycation, diabetes, and aging. Recent Prog Horm Res 56: 1-21.

15. Varki A, Lowe JB (2009) Biological Roles of Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY).

16. Lando D, Peet DJ, Whelan DA, Gorman JJ, Whitelaw ML (2002) Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. Science 295: 858-861.

17. Fitzpatrick PF (2003) Mechanism of aromatic amino acid hydroxylation. Biochemistry 42: 14083-14091.

18. Kivirikko KI, Myllyla R, Pihlajaniemi T (1989) Protein hydroxylation: prolyl 4-hydroxylase, an enzyme with four cosubstrates and a multifunctional subunit. FASEB J 3: 1609-1617.

19. Martin BR, Wang C, Adibekian A, Tully SE, Cravatt BF (2012) Global profiling of dynamic protein palmitoylation. Nat Methods 9: 84-89.

20. Smotrys JE, Linder ME (2004) Palmitoylation of intracellular signaling proteins: regulation and function. Annu Rev Biochem 73: 559-587.

21. Seibert C, Sakmar TP (2008) Toward a framework for sulfoproteomics: Synthesis and characterization of sulfotyrosine-containing peptides. Biopolymers 90: 459-477.

22. Huang SY, Shi SP, Qiu JD, Sun XY, Suo SB, et al. (2012) PredSulSite: prediction of protein tyrosine sulfation sites with multiple features and analysis. Anal Biochem 428: 16-23.

23. Zhang Y, Jiang H, Go EP, Desaire H (2006) Distinguishing phosphorylation and sulfation in carbohydrates and glycoproteins using ion-pairing and mass spectrometry. J Am Soc Mass Spectrom 17: 1282-1288.

24. Pickart CM (2001) Mechanisms underlying ubiquitination. Annu Rev Biochem 70: 503-533.

25. Komander D (2009) The emerging complexity of protein ubiquitination. Biochem Soc Trans 37: 937-953.

26. Xu G, Jaffrey SR (2011) The new landscape of protein ubiquitination. Nat Biotechnol 29: 1098-1100.

27. Verdone L, Agricola E, Caserta M, Di Mauro E (2006) Histone acetylation in gene regulation. Brief Funct Genomic Proteomic 5: 209-221.

28. Fierz B, Muir TW (2012) Chromatin as an expansive canvas for chemical biology. Nat Chem Biol 8: 417-427.

29. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389: 251-260.

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

31. Garcia BA, Barber CM, Hake SB, Ptak C, Turner FB, et al. (2005) Modifications of human histone H3 variants during mitosis. Biochemistry 44: 13202-13213.

32. Horn PJ, Peterson CL (2002) Molecular biology. Chromatin higher order folding--wrapping up transcription. Science 297: 1824-1827.

33. Pesavento JJ, Kim YB, Taylor GK, Kelleher NL (2004) Shotgun annotation of histone modifications: a new approach for streamlined characterization of proteins by top down mass spectrometry. J Am Chem Soc 126: 3386-3387.

34. Hayes JJ (2002) Changing chromatin from the inside. Nature Struct Biol 9: 161-163.

35. Wisniewski JR, Zougman A, Mann M (2008) Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. Nucleic Acids Res 36: 570-577.

36. Recht J, Tsubota T, Tanny JC, Diaz RL, Berger JM, et al. (2006) Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis. Proc Natl Acad Sci U S A 103: 6988-6993.

37. Peterson CL, Laniel MA (2004) Histones and histone modifications. Curr Biol 14: R546-551.

38. Fischle W, Tseng BS, Dormann HL, Ueberheide BM, Garcia BA, et al. (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature 438: 1116-1122.

39. Hake SB, Xiao A, Allis CD (2004) Linking the epigenetic 'language' of covalent histone modifications to cancer. Br J Cancer 90: 761-769.

40. Siuti N, Roth MJ, Mizzen CA, Kelleher NL, Pesavento JJ (2006) Gene-specific characterization of human histone H2B by electron capture dissociation. J Proteome Res 5: 233-239.

41. Felsenfeld G, Groudine M (2003) Controlling the double helix. Nature 421: 448-453.

42. Munshi A, Shafi G, Aliya N, Jyothy A (2009) Histone modifications dictate specific biological readouts. J Genet Genomics 36: 75-88.

43. Pesavento JJ, Bullock CR, LeDuc RD, Mizzen CA, Kelleher NL (2008) Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. J Biol Chem 283: 14927-14937.

44. Zheng C, Hayes JJ (2003) Structures and interactions of the core histone tail domains. Biopolymers 68: 539-546.

45. Kao CF, Osley MA (2003) In vivo assays to study histone ubiquitylation. Methods 31: 59-66.

46. Galasinski SC, Resing KA, Ahn NG (2003) Protein mass analysis of histones. Methods 31: 3-11.

47. Wiley EA, Ohba R, Yao MC, Allis CD (2000) Developmentally regulated rpd3p homolog specific to the transcriptionally active macronucleus of vegetative Tetrahymena thermophila. Mol Cell Biol 20: 8319-8328.

48. Boyne MT, 2nd, Pesavento JJ, Mizzen CA, Kelleher NL (2006) Precise characterization of human histones in the H2A gene family by top down mass spectrometry. J Proteome Res 5: 248-253.

49. Angelov D, Vitolo JM, Mutskov V, Dimitrov S, Hayes JJ (2001) Preferential interaction of the core histone tail domains with linker DNA. Proc Natl Acad Sci U S A 98: 6599-6604.

50. Grunstein M (1997) Histone acetylation in chromatin structure and transcription. Nature 389: 349-352.

51. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403: 41-45.

52. Hildmann C, Riester D, Schwienhorst A (2007) Histone deacetylases--an important class of cellular regulators with a variety of functions. Appl Microbiol Biotechnol 75: 487-497.

53. Jiang T, Zhou X, Taghizadeh K, Dong M, Dedon PC (2007) N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. Proc Natl Acad Sci U S A 104: 60-65.

54. LeRoy G, Weston JT, Zee BM, Young NL, Plazas-Mayorca MD, et al. (2009) Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications. Mol Cell Proteomics 8: 2432-2442.

55.   Marnett LJ, Riggins JN, West JD (2003) Endogenous generation of reactive oxidants and electrophiles and their reactions with DNA and protein. J Clin Invest 111: 583-593.

56.   Winterbourn CC (2008) Reconciling the chemistry and biology of reactive oxygen species. Nat Chem Biol 4: 278-286.

57.   Lonkar P, Dedon PC (2011) Reactive species and DNA damage in chronic inflammation: reconciling chemical mechanisms and biological fates. Int J Cancer 128: 1999-2009.

58.   Marnett LJ (2012) Inflammation and cancer: chemical approaches to mechanisms, imaging, and treatment. J Org Chem 77: 5224-5238.

59.   Turrens JF (2003) Mitochondrial formation of reactive oxygen species. J Physiol 552: 335-344.

60.   Lewis RS, Tamir S, Tannenbaum SR, Deen WM (1995) Kinetic analysis of the fate of nitric oxide synthesized by macrophages in vitro. J Biol Chem 270: 29350-29355.

61.   Mancardi D, Ridnour LA, Thomas DD, Katori T, Tocchetti CG, et al. (2004) The chemical dynamics of NO and reactive nitrogen oxides: a practical guide. Curr Mol Med 4: 723-740.

62.   Hazen SL, d'Avignon A, Anderson MM, Hsu FF, Heinecke JW (1998) Human neutrophils employ the myeloperoxidase-hydrogen peroxide-chloride system to oxidize alpha-amino acids to a family of reactive aldehydes. Mechanistic studies identifying labile intermediates along the reaction pathway. J Biol Chem 273: 4997-5005.

63.   Eiserich JP, Hristova M, Cross CE, Jones AD, Freeman BA, et al. (1998) Formation of nitric oxide-derived inflammatory oxidants by myeloperoxidase in neutrophils. Nature 391: 393-397.

64.   Retrieved from http://en.wikipedia.org/wiki/Mitochondrion, November 18, 2012

65.   Davies KJ (1995) Oxidative stress: the paradox of aerobic life. Biochem Soc Symp 61: 1-31.

66.   Imlay JA (2003) Pathways of oxidative damage. Annu Rev Microbiol 57: 395-418.

67.   Trachootham D, Alexandre J, Huang P (2009) Targeting cancer cells by ROS-mediated mechanisms: a radical therapeutic approach? Nat Rev Drug Discov 8: 579-591.

68. Shacter E (2000) Quantification and significance of protein oxidation in biological samples. Drug Metab Rev 32: 307-326.

69. Esterbauer H, Schaur RJ, Zollner H (1991) Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. Free Radic Biol Med 11: 81-128.

70. Stone MP, Cho YJ, Huang H, Kim HY, Kozekov ID, et al. (2008) Interstrand DNA cross-links induced by alpha,beta-unsaturated aldehydes derived from lipid peroxidation and environmental sources. Acc Chem Res 41: 793-804.

71. Minko IG, Yamanaka K, Kozekov ID, Kozekova A, Indiani C, et al. (2008) Replication bypass of the acrolein-mediated deoxyguanine DNA-peptide cross-links by DNA polymerases of the DinB family. Chem Res Toxicol 21: 1983-1990.

72. Minko IG, Kozekov ID, Harris TM, Rizzo CJ, Lloyd RS, et al. (2009) Chemistry and biology of DNA containing 1,N(2)-deoxyguanosine adducts of the alpha,beta-unsaturated aldehydes acrolein, crotonaldehyde, and 4-hydroxynonenal. Chem Res Toxicol 22: 759-778.

73. Slade PG, Williams MV, Chiang A, Iffrig E, Tannenbaum SR, et al. (2011) A filtered database search algorithm for endogenous serum protein carbonyl modifications in a mouse model of inflammation. Mol Cell Proteomics 10: M111 007658.

74. Codreanu SG, Zhang B, Sobecki SM, Billheimer DD, Liebler DC (2009) Global analysis of protein damage by the lipid electrophile 4-hydroxy-2-nonenal. Mol Cell Proteomics 8: 670-680.

75. Tallman KA, Kim HY, Ji JX, Szapacs ME, Yin H, et al. (2007) Phospholipid-protein adducts of lipid peroxidation: synthesis and study of new biotinylated phosphatidylcholines. Chem Res Toxicol 20: 227-234.

76. Connor RE, Marnett LJ, Liebler DC (2011) Protein-selective capture to analyze electrophile adduction of hsp90 by 4-hydroxynonenal. Chem Res Toxicol 24: 1275-1282.

77. Jacobs AT, Marnett LJ (2010) Systems analysis of protein modification and cellular responses induced by electrophile stress. Acc Chem Res 43: 673-683.

78. Levine RL (2002) Carbonyl modified proteins in cellular regulation, aging, and disease. Free Radic Biol Med 32: 790-796.

79. Thornalley PJ (2008) Protein and nucleotide damage by glyoxal and methylglyoxal in physiological systems--role in ageing and disease. Drug Metabol Drug Interact 23: 125-150.

41

80.   Dedon PC (2008) The chemical toxicology of 2-deoxyribose oxidation in DNA. Chem Res Toxicol 21: 206-219.

81.   Rehder DS, Borges CR (2010) Cysteine sulfenic acid as an intermediate in disulfide bond formation and nonenzymatic protein folding. Biochemistry 49: 7748-7755.

82.   Wang Y, Vivekananda S, Men L, Zhang Q (2004) Fragmentation of protonated ions of peptides containing cysteine, cysteine sulfinic acid, and cysteine sulfonic acid. J Am Soc Mass Spectrom 15: 697-702.

83.   Crabtree M, Hao G, Gross SS (2003) Detection of cysteine S-nitrosylation and tyrosine 3-nitration in kidney proteins. Methods Mol Med 86: 373-384.

84.   Jones LH (2012) Chemistry and biology of biomolecule nitration. Chem Biol 19: 1086-1092.

85.   Levine RL, Mosoni L, Berlett BS, Stadtman ER (1996) Methionine residues as endogenous antioxidants in proteins. Proc Natl Acad Sci U S A 93: 15036-15040.

86.   Ara J, Przedborski S, Naini AB, Jackson-Lewis V, Trifiletti RR, et al. (1998) Inactivation of tyrosine hydroxylase by nitration following exposure to peroxynitrite and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). Proc Natl Acad Sci U S A 95: 7659-7663.

87.   Madian AG, Regnier FE (2010) Proteomic identification of carbonylated proteins and their oxidation sites. J Proteome Res 9: 3766-3780.

88.   Niwa T (1999) 3-Deoxyglucosone: metabolism, analysis, biological activity, and clinical implication. J Chromatogr B Biomed Sci Appl 731: 23-36.

89.   Jono T, Nagai R, Lin X, Ahmed N, Thornalley PJ, et al. (2004) Nepsilon-(Carboxymethyl)lysine and 3-DG-imidazolone are major AGE structures in protein modification by 3-deoxyglucosone. J Biochem 136: 351-358.

90.   Brownlee M (1995) Advanced protein glycosylation in diabetes and aging. Annu Rev Med 46: 223-234.

91.   Bakris GL, Bank AJ, Kass DA, Neutel JM, Preston RA, et al. (2004) Advanced glycation end-product cross-link breakers. A novel approach to cardiovascular pathologies related to the aging process. Am J Hypertens 17: 23S-30S.

92.   Kaminskas LM, Pyke SM, Burcham PC (2005) Differences in lysine adduction by acrolein and methyl vinyl ketone: implications for cytotoxicity in cultured hepatocytes. Chem Res Toxicol 18: 1627-1633.

93.   Surh J, Kwon H (2005) Estimation of daily exposure to 4-hydroxy-2-alkenals in Korean foods containing n-3 and n-6 polyunsaturated fatty acids. Food Addit Contam 22: 701-708.

94.   Lu K, Boysen G, Gao L, Collins LB, Swenberg JA (2008) Formaldehyde-induced histone modifications in vitro. Chem Res Toxicol 21: 1586-1593.

95.   Prasad A, Bekker P, Tsimikas S (2012) Advanced Glycation Endproducts and Diabetic Cardiovascular Disease. Cardiol Rev.

96.   Neely MD, Sidell KR, Graham DG, Montine TJ (1999) The lipid peroxidation product 4-hydroxynonenal inhibits neurite outgrowth, disrupts neuronal microtubules, and modifies cellular tubulin. J Neurochem 72: 2323-2333.

97.   Vana L, Kanaan NM, Hakala K, Weintraub ST, Binder LI (2011) Peroxynitrite-induced nitrative and oxidative modifications alter tau filament formation. Biochemistry 50: 1203-1212.

98.   Cai H, Guengerich FP (2000) Acylation of protein lysines by trichloroethylene oxide. Chem Res Toxicol 13: 327-335.

99.   Tyihak E, Trezl L, Kolonits P (1985) The isolation of Nepsilon-formyl-L-lysine from the reaction between formaldehyde and L-lysine and its identification by OPLC and NMR spectroscopy. J Pharm Biomed Anal 3: 343-349.

100.   Oses-Prieto JA, Zhang X, Burlingame AL (2007) Formation of epsilon-formyllysine on silver-stained proteins: implications for assignment of isobaric dimethylation sites by tandem mass spectrometry. Mol Cell Proteomics 6: 181-192.

101.   Lu K, Moeller B, Doyle-Eisele M, McDonald J, Swenberg JA (2011) Molecular dosimetry of N2-hydroxymethyl-dG DNA adducts in rats exposed to formaldehyde. Chem Res Toxicol 24: 159-161.

102.   Cogliano VJ, Grosse Y, Baan RA, Straif K, Secretan MB, et al. (2005) Meeting report: summary of IARC monographs on formaldehyde, 2-butoxyethanol, and 1-tert-butoxy-2-propanol. Environ Health Perspect 113: 1205-1208.

103.   (2011) NTP 12th Report on Carcinogens. Report on carcinogens : carcinogen profiles / US Dept of Health and Human Services, Public Health Service, National Toxicology Program: iii-499.

104.   (2006) Integrated Risk Information System (IRIS): Formaldehyde. US Environmental Protection Agency.

105.   Zhang L, Freeman LE, Nakamura J, Hecht SS, Vandenberg JJ, et al. (2010) Formaldehyde and leukemia: epidemiology, potential mechanisms, and implications for risk assessment. Environ Mol Mutagen 51: 181-191.

106. Le Curieux F, Pluskota D, Munter T, Sjoholm R, Kronberg L (2000) Identification of fluorescent 2'-deoxyadenosine adducts formed in reactions of conjugates of malonaldehyde and acetaldehyde, and of malonaldehyde and formaldehyde. Chem Res Toxicol 13: 1228-1234.

107. Begley TJ, Samson LD (2003) AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. Trends Biochem Sci 28: 2-5.

108. Dedon PC, Jiang ZW, Goldberg IH (1992) Neocarzinostatin-mediated DNA damage in a model AGT.ACT site: mechanistic studies of thiol-sensitive partitioning of C4' DNA damage products. Biochemistry 31: 1917-1927.

Chapter 2

Development of a liquid chromatography-coupled tandem mass spectrometry method for quantification of N^6-formyllysine adducts in proteins

Abstract

The initial step in investigating the mechanism of formation of N^6-formyllysine and assessing its possible pathological consequences in cells involved developing an ultrasensitive and specific liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) method for quantifying this adduct in proteins. The quantification method we developed previously (Jiang *et al. PNAS* **104**: 60-5, 2007) proved to be relatively insensitive and biased, so a new method was developed, which involved proteolytic hydrolysis of proteins to individual amino acids, normal phase chromatographic resolution of the amino acids followed by tandem electrospray mass spectrometry to quantify them. N^6-Formyllysine is reported as a percentage of total lysines in order to normalize it across samples with different amounts and types of proteins. The use of *Streptomyces griseus* protease greatly diminished the background levels of lysine and formyllysine observed previously with proteinase K. Further, the direct analysis of amino acids, as opposed to phenylisothiocyanate (PITC) derivatization, improved both the sensitivity and specificity of the assay. In addition to lysine and N^6-formyllysine species, N^6-acetyl, mono-, di-, and tri-methyllysine modifications are monitored, with limit of detection of 1-50 fmol for all species. Measurement of N^6-formyllysine in histones of TK6 cells indicates that this adduct is an abundant endogenous protein modification (~1 per 10^4 lysines). In addition, the dilution rate of the absolute quantity of N^6-formyllysine in histone proteins of TK6 cells (grown in media containing isotopically labeled lysine) reveals that N^6-formyllysine is a rather stable modification of histones, with a half-life closely matching TK6 cell doubling rate.

Introduction

The first step in defining the biological relevance of a new molecule is to quantify its presence in cells and tissues. With this in mind, a new method was developed to quantify the N^6-formyllysine protein adducts. The method workflow shown in Figure 2-1 involves hydrolysis of proteins to generate individual amino acids that are then analyzed using liquid chromatography-coupled tandem mass spectrometry. Each step in the method is critical to the precision and accuracy of the analytical data.

Proteolytic digestion using proteinase K has been utilized previously for generating individual amino acids containing N^6-formyllysine moiety [1,2] in order to avoid removal of the formyl group by traditional strong acid-mediated protein hydrolysis. However, we found that proteinase K undergoes extensive autolysis and, when used in small quantities to avoid self digestion, it was not an efficient protease. The next step of chromatographic resolution of amino acids was previously accomplished with PITC derivatization and reversed phase HPLC. However, this approach decreased the sensitivity of the method and added an extra step. The final step of tandem mass spectrometric quantification of the amino acids is relatively straight forward.

This chapter covers the new enzyme discovered as well as other advancements to our previous method to make it more reliable and sensitive and presents an example of its immediate applications (i.e., determining the rate of dilution of this adduct in cells). We extend our quantitative analysis of proteins to include other lysine modifications such as acetylation and methylation besides formylation. We report the development of a highly robust and sensitive LC-MS/MS method for measuring all N^6-formyl, -acetyl, and -methyllysine modifications, with limits of detection of 1-50 fmol for all species. In
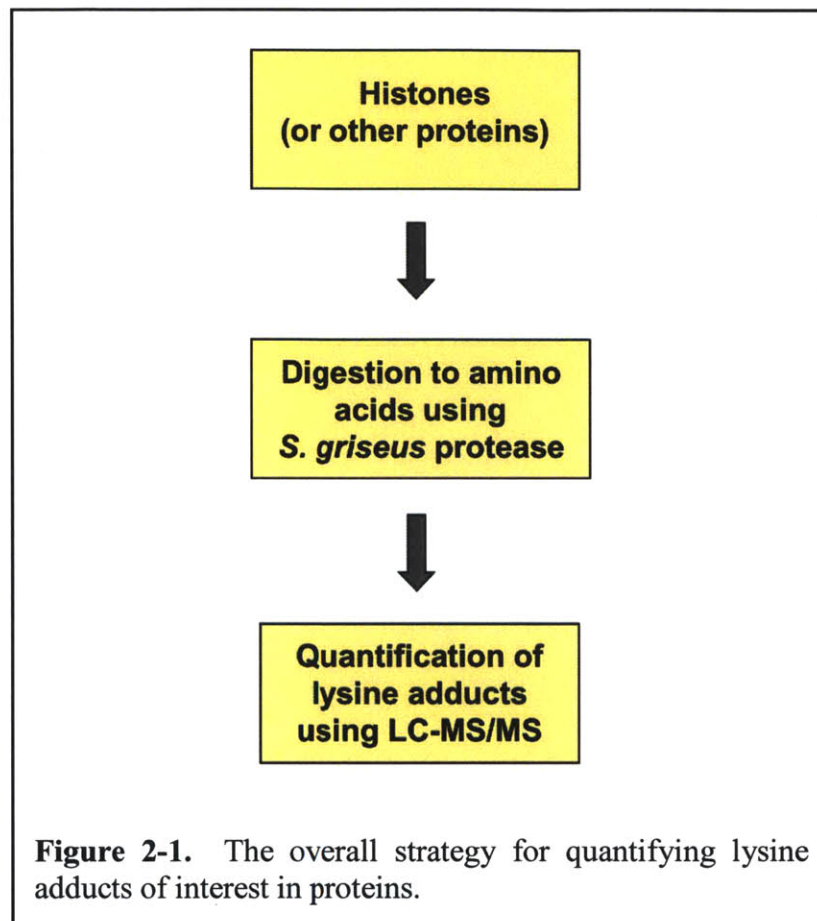
**Histones
(or other proteins)**

↓

**Digestion to amino
acids using
S. griseus protease**

↓

**Quantification of
lysine adducts
using LC-MS/MS**

Figure 2-1. The overall strategy for quantifying lysine adducts of interest in proteins.

addition to providing insights into N^6-formyllysine content of proteins, there are numerous applications of this method for characterizing the chemical biology of this protein modification as presented in the subsequent chapters.

Materials and Methods

Materials. Lysine internal standard, 4,4,5,5,-[^{2}H]-Lysine, was purchased from Cambridge Isotope Laboratories (Andover, MA). N^6-Formyllysine internal standard, 4,4,5,5-[^{2}H]-N^6-formyllysine, was synthesized from 4,4,5,5-[^{2}H]-lysine according to Jiang *et al.* [1]. N^6-Acetyllysine internal standard, 3,3,4,4,5,5,6,6-[^{2}H]-N^6-acetyllysine

was obtained from CDN Isotopes (Pointe-Claire, Quebec, Canada). Lysine, N^6-formyllysine, N^6-acetyllysine, *Streptomyces griseus* protease, and protease inhibitor cocktail (for use with mammalian cell and tissue extracts) were obtained from Sigma-Aldrich (St. Louis, MO). N^6-Mono-methyllysine, N^6-di-methyllysine, and N^6-tri-methyllysine were purchased from Bachem Bioscience Inc. (King of Prussia, PA). Nonidet P-40 was from Roche Diagnostic Corporation (Indianapolis, IN). Human lymphoblastoid TK6 cell line was a generous gift of Wogan Lab at Massachusetts Institute of Technology.

TK6 cell culture and labeling. TK6 cells were cultured in RPMI 1640 medium (Cellgro, Manassas, VA) supplemented with 10% heat-inactivated horse serum (Atlanta Biologicals, Lawrenceville, GA), 10,000 U penicillin/ml and 10,000 μg streptomycin/ml (Lonza, Walkersville, MD), and 2mM L-glutamine (Lonza, Walkersville, MD) at 37° C in a 5% CO_2 atmosphere. For labeling proteins with endogenous isotopically labeled N^6-formyllysine, TK6 cells were grown in a customized RPMI-1640 medium, with everything identical to traditional RPMI 1640 medium (supplemented with horse serum, antibiotics, and L-glutamine), except for the presence of deuterated lysine (3,3,4,4,5,5,6,6-[^{2}H]-Lysine) instead of non-labeled lysine. After one week of growth (roughly 9 doubling times) in the customized medium, cells were washed and resuspended in non-labeled RPMI medium. Histones were extracted from ~10 million cells at 0h, 18h, 36h, and 54h, and the quantity of isotopically labeled formyllysine, as a percentage of total labeled and unlabeled lysines, was measured at each of those time points.

Histone extraction from TK6 cells and bovine liver tissue. Extraction of histones was performed according to Boyne *et al.* [3], with modifications. Cells (~10^7 per sample) were pelleted by centrifugation at 1000 x g for 5 min at 4 °C and the pellet was washed once with PBS. Cell pellets were then lysed by resuspension in ice-cold lysis buffer (15 mM Tris-HCl, pH 7.5, 15 mM NaCl, 60 mM KCl, 1 mM CaCl$_2$, 5 mM MgCl$_2$, 250 mM sucrose, 1 mM dithiothreitol, 10 mM sodium butyrate) containing a 100:1 v:v dilution of protease inhibitor cocktail in the presence of 0.03% Nonidet P-40 and incubation on ice for 5 min with occasional gentle mixing. Nuclei were pelleted by centrifugation at 600 x g for 5 min at 4 °C, and the pellet was washed twice with ice-cold lysis buffer without Nonidet P-40. Histones were extracted by addition of ice-cold 0.4N H$_2$SO$_4$ and incubation overnight on ice. The solution was centrifuged at 3000 x g for 5 min and proteins in supernatant were precipitated by addition of 20% v/v trichloroacetic acid and overnight incubation at 4 °C. Samples were then centrifuged at 14000 x g for 10 min at 4°C, washed once with ice-cold acetone containing 0.1% HCl, and once with ice-cold acetone. The extracts were air-dried and stored at -20 ° C until use. For extracting histones from bovine liver tissue, 20 mg of tissue was cut into small pieces and washed with PBS. Chromatin bound fraction was obtained using the Subcellular Protein Fractionation Kit from Thermo Scientific (Waltham, MA) and a Kontes all-glass Dounce homogenizer (10 strokes with a type B pestle). Proteins in chromatin bound fraction were precipitated by addition of 20% v/v trichloroacetic acid and overnight incubation at 4 °C. Samples were then centrifuged at 14000 x g for 10 min at 4 °C, washed once with ice-cold acetone containing 0.1% HCl, and once with ice-cold acetone. The extracts were air-dried and stored at -20 ° C until use.

Purification of individual histones. HPLC purification of total histones was performed according to Boyne *et al.* [3], with modifications. Total histones ($\leq$ 50 μg) were dissolved in 0.1% trifluoroacetic acid (TFA) in water and fractionated by HPLC on an Agilent 1100 series instrument (Agilent Technologies, Santa Clara, CA), using a Source 5RPC C_{18} reversed-phase column (4.6 x 150 mm, 5 μm particles; GE Healthcare Life Sciences). The mobile phase flow rate was 1 mL/min and the solvent system was 0.1% TFA in water (A) and 0.094% TFA in acetonitrile (B) with the elution starting at 0% B, linearly increasing to 28% B over 28 min, reaching 37% B at 70 min, 38% B at 100 min, 60% B at 150 min, and finally 100% B at 151 min, before the column was re-equilibrated to 0% B for 10 min. Protein elution was monitored by UV absorbance at 214 nm and histones in each fraction were tentatively identified by resolution on a 13% SDS-PAGE gel with Coomassie Blue staining (see Figure 3-S1).

Enzymatic hydrolysis of proteins. Histones extracted from TK6 cells and other protein samples were dissolved in 50 μL of 100 mM ammonium bicarbonate buffer (pH 8.5), 4,4,5,5-[^{2}H]-lysine (2 nmol),4,4,5,5-[^{2}H]-N^6-formyllysine (1 pmol), and 3,3,4,4,5,5,6,6-[^{2}H]-N^6-acetyl lysine (10 pmol) were added as internal standards, and the proteins hydrolyzed by addition of *Streptomyces griseus* protease (freshly prepared solution each time) with incubation at 37 °C for $\geq$ 16 h. *S. griseus* was used at a ratio of 1 μg of enzyme per 10 μg of protein. Samples were then dried under vacuum and resuspended in 50 μL of water before mass spectrometry analysis.

Quantification of amino acids. N^6-Formyllysine and other amino acids were quantified

as a percentage of the total quantity of lysine, by liquid chromatography-coupled mass spectrometry (LC-MS/MS). HPLC was performed with an Agilent 1100 series instrument. Adducts of interest (Figure 2-2) in the resuspended protein hydrolysates were separated using an aqueous normal phase Cogent diamond hydride column (2.1 x 150 mm, 4 μm) from MicroSolv Technology Corporation (Eatontown, NJ). The mobile phase flow rate was 400 μL/min, and the column temperature was maintained at 20 °C. The solvent system was 0.1% acetic acid in water (A) and 0.1% acetic acid in acetonitrile (B), with the elution starting at 100% B, the gradient linearly decreased to 25% B over 30

Figure 2-2. The structure and MW of lysine species monitored.
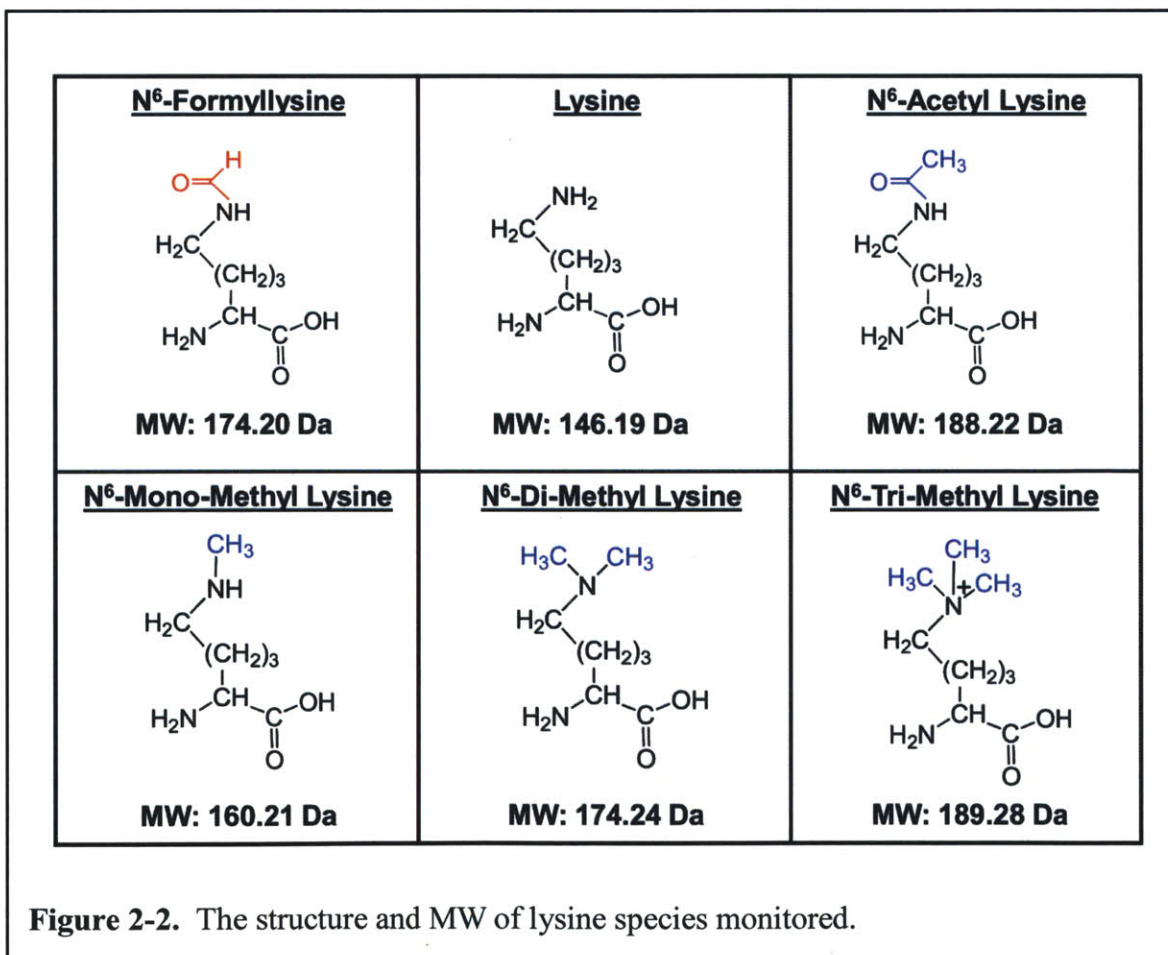
min, stayed at 25% B for 3 additional min before the column was re-equilibrated at 100%

B for 7 min. In order to separate the co-eluting M+1 citrulline signal from M+2 ion of

N^6-formyllysine, an extended chromatography run time was used, with the elution

starting at 100% B, the gradient linearly decreased to 75% B over 75 min, further

decreased to 25% B over the next 3 min, reached 15% by 83 min before the column was

re-equilibrated at 100% B for 7 min. The species of interest were then analyzed using an

Agilent 6410 triple quadrupole mass spectrometer (MS/MS) equipped with an

electrospray ionization (ESI) source operated in positive ion mode. The operating

parameters were as follows: ESI capillary voltage, 4000 V; gas temperature, 350 °C;

drying gas flow, 12 L/min; and nebulizer pressure, 30 psi. Selected reaction monitoring

(SRM) transitions are summarized in Table 2-1. Note that in addition to chromatographic

separation and presence of internal standards, the unique product ions of m/z 112 and m/z

126 for formyl and acetyl lysines, respectively, distinguish them from their isobaric

compounds di- and tri-methyl lysines (Table 2-1 and Figure 2-3). The fragmentor

voltage and collision energy were optimized in order to maximize the signal of each

product ion monitored (Figure 2-4) and are summarized in Table 2-1. 4,4,5,5-[^{2}H]-

lysine, 4,4,5,5-[^{2}H]-N^6-formyllysine, and 3,3,4,4,5,5,6,6-[^{2}H]-N^6-acetyl lysine were used

as internal standards and external calibration curves for methyl species were prepared

using deuterated acetyl lysine (Figure 2-5).

Table 2-1. Summary of mass spectrometry parameters for each species.

Species	SRM Transition (*m/z*)	Internal Standard Transition (*m/z*)	Qualifier Ion Transition (*m/z*)	Fragmentor Voltage (V)	Collision Energy (V)
Lysine	147 → 130	151 → 134	NA	100	8
N^6-Formyllysine	175 → 112	179 → 116	NA	105	10
N^6-Acetyllysine	189 → 126	197 → 134	NA	105	10
N^6-Mono-methyllysine	161 → 84	NA	161 → 130	100	10
N^6-Di-methyllysine	175 → 84	NA	175 → 130	100	14
N^6-Tri-methyllysine	189 → 84	NA	189 → 130	100	14
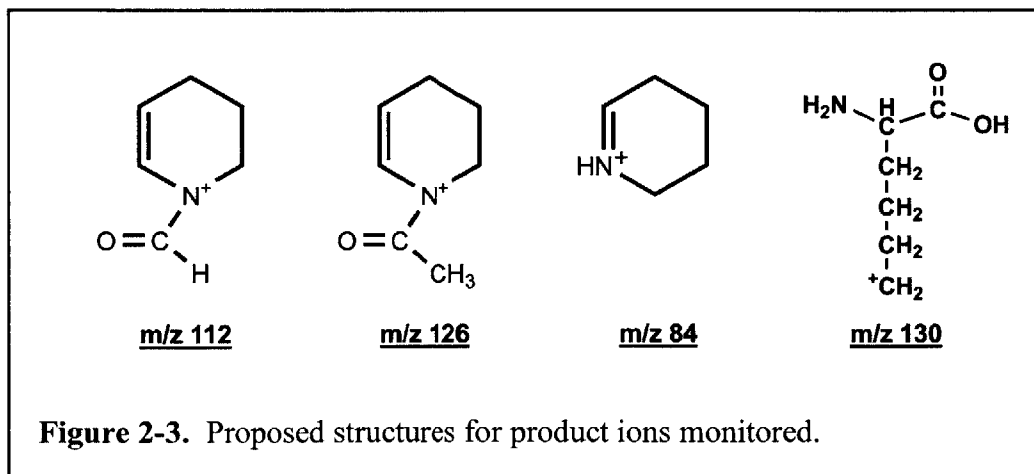
Figure 2-3. Proposed structures for product ions monitored.
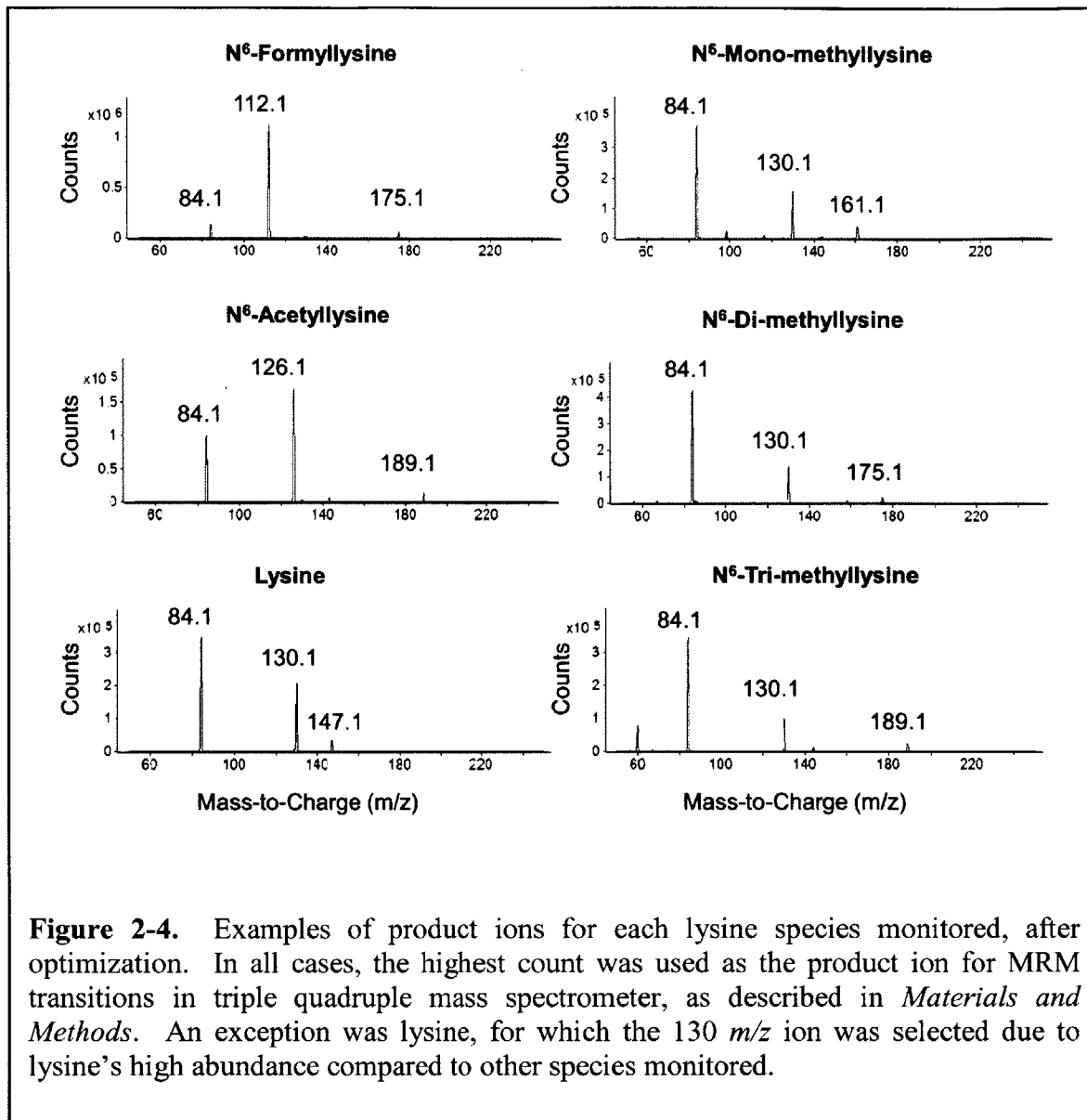
Figure 2-4. Examples of product ions for each lysine species monitored, after optimization. In all cases, the highest count was used as the product ion for MRM transitions in triple quadruple mass spectrometer, as described in *Materials and Methods*. An exception was lysine, for which the 130 *m/z* ion was selected due to lysine's high abundance compared to other species monitored.

Figure 2-5. Examples of calibration curves for the isotope dilution LC-MS/MS analysis of lysine species, as described in *Materials and Methods*. Abbreviations: FK, N^6-formyllysine; DFK, deuterium-labeled N^6-formyllysine; AK, N^6-acetyllysine; DAK, deuterium-labeled N^6-acetyllysine; K, lysine; DK, deuterium-labeled lysine; MK, N^6-mono-methyllysine; M_2K, N^6-di-methyllysine; M_3K, N^6-tri-methyllysine.

Results

An improved method to quantify N^6-formyllysine in proteins. Our previous method for quantifying N^6-formyllysine in proteins involved proteinase K-mediated hydrolysis of proteins, derivatization of the resulting amino acids with PITC, HPLC pre-purification of amino acid derivatives, and final LC-MS/MS analysis of the derivatized amino acids [1]. This method proved to be relatively insensitive and biased as a result of using proteinase K, which produced only partial hydrolysis of some proteins when used in small quantities to minimize background autolysis (Figure 2-6). To overcome these problems, we used *Streptomyces griseus* protease at a ratio of 1 µg enzyme per 10 µg proteins, which resulted in efficient and complete digestion of all proteins as judged by the total amount
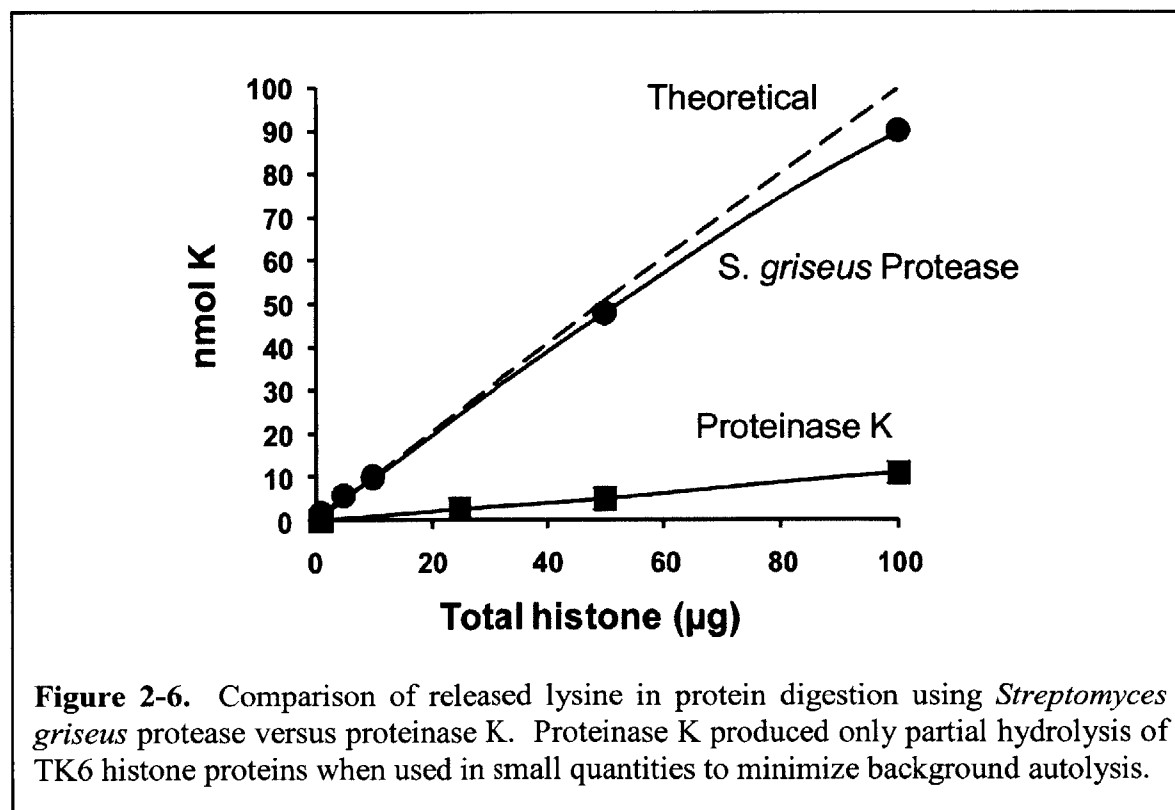
Figure 2-6. Comparison of released lysine in protein digestion using *Streptomyces griseus* protease versus proteinase K. Proteinase K produced only partial hydrolysis of TK6 histone proteins when used in small quantities to minimize background autolysis.

of lysine released per μg of purified histone proteins (Figure 2-6). In addition, the method was optimized to eliminate the HPLC pre-purification step and the need for PITC derivatization to achieve chromatographic resolution of amino acids. This was achieved using aqueous normal phase HPLC with a diamond-hydride column to resolve individual amino acids. This chromatographic system resolved N^6-acetyllysine, mono-, di-, and tri-N^6-methyllysines, as well as N^6-formyllysine and lysine, as shown in Figure 2-7. With isotopically labeled internal standards added prior to protease digestion, identification and quantification of amino acids were accomplished by isotope-dilution LC-MS/MS in positive ion mode, using multiple reaction monitoring (MRM) transitions. The limits of detection were found to be 1 fmol for N^6-formyl- and N^6-acetyllysine, 10 fmol for lysine, and 50 fmol for each of N^6-mono-, di-, and tri-methyl lysine. Data for the various lysine modifications are expressed here as proportions of the total number of lysines in the sample.

Measurement of N^6-formyllysine adducts in total histone proteins. The first application of the new method involved quantification of the various modified lysines in a mixture of total histone proteins extracted from TK6 lymphoblastoid cells (Table 2-2). Data show that N^6-formylation of lysine occurs at a low frequency (e.g., 1 per 10^4 lysines) in total histones compared to higher frequencies of acetylation and methylation (40 – 230 per 10^4 lysines). Further, an analysis of total histone proteins, either extracted from TK6 cells or from commercial sources, show a similar background level of N^6-formyllysine in all samples (Table 2-3). The next chapter covers the distribution of various lysine adducts in different classes of histones separated by HPLC.

Figure 2-7. An example of different lysine species detected in purified histone H4 from TK6 cells. Lysine adducts were monitored by tandem mass spectrometry, as described in *Materials and Methods*.
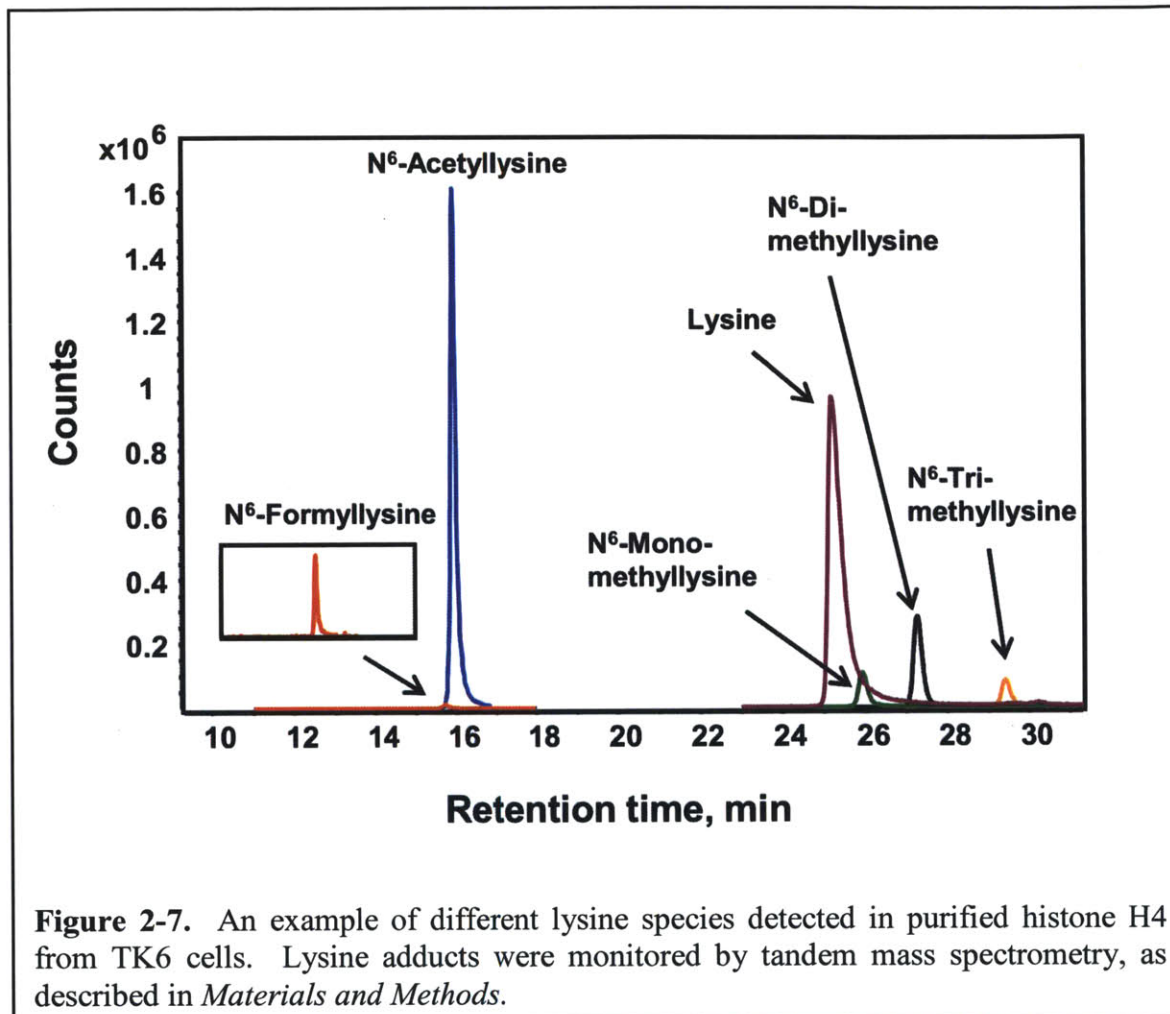
Table 2-2. Quantification of lysine modifications in a sample of total histone proteins.

	Formyl[1]	Acetyl	Methyl	Dimethyl	Trimethyl
TK6 Cells	1.0 ± 0.1[2]	40 ± 10	230 ± 50	130 ± 40	25 ± 5

[1]Column titles denote different N^6-modifications of lysine.

[2]Data are expressed as modifications per 10^4 total lysines and represent mean $\pm$ SD for 3 biological replicates.

Table 2-3. Quantification of N^6-Formyllysine in total histones.

Histone source	N^6-Formyllysine (per 10^4Lys)
TK6 cells	1.0 ± 0.1
Calf thymus	2 ± 0.2
Bovine liver	2 ± 0.6

Resolving co-elution problem of a contaminant with M+2 ion of N^6-formyllysine. In addition to monitoring N^6-formyllysine ion (*m/z* transition of 175 → 112), some applications of the quantification method, as will be discussed in Chapters 3 and 4, involve monitoring the heavy isotope of N^6-formyllysine (*m/z* transition of 177 → 114). However, we encountered the presence of 177 *m/z* ion in control samples (without the heavy isotope) that was well beyond the background abundance of M+2 ion of N^6-formyllysine (~0.7%), suggesting the co-elution of a contaminant peak with the analyte. We have determined the contaminant to be the M+1 ion of citrulline (Figures 2-8 and 2-9) and thus substantially extended the chromatography to separate the two.
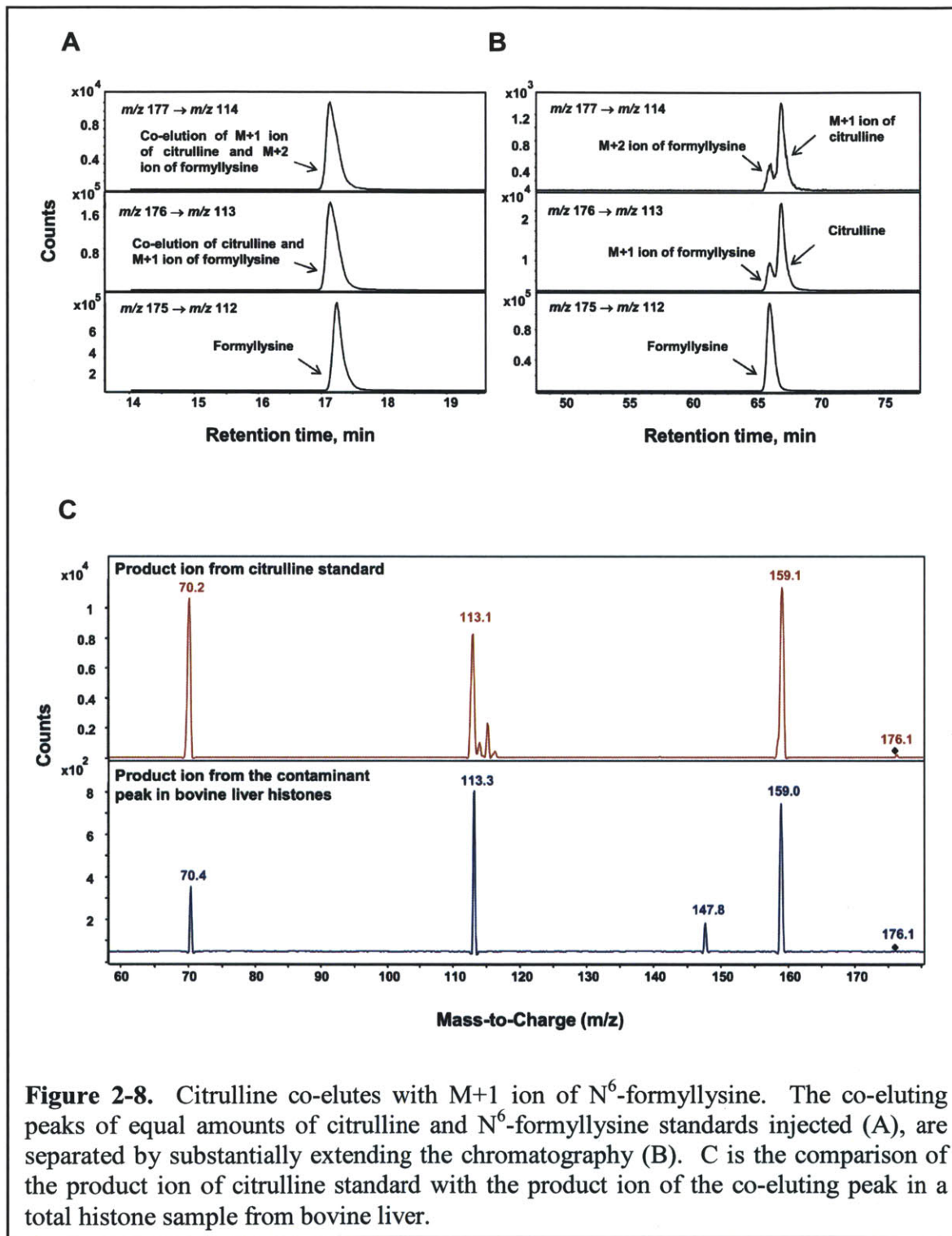
Figure 2-8. Citrulline co-elutes with M+1 ion of N^6-formyllysine. The co-eluting peaks of equal amounts of citrulline and N^6-formyllysine standards injected (A), are separated by substantially extending the chromatography (B). C is the comparison of the product ion of citrulline standard with the product ion of the co-eluting peak in a total histone sample from bovine liver.

61
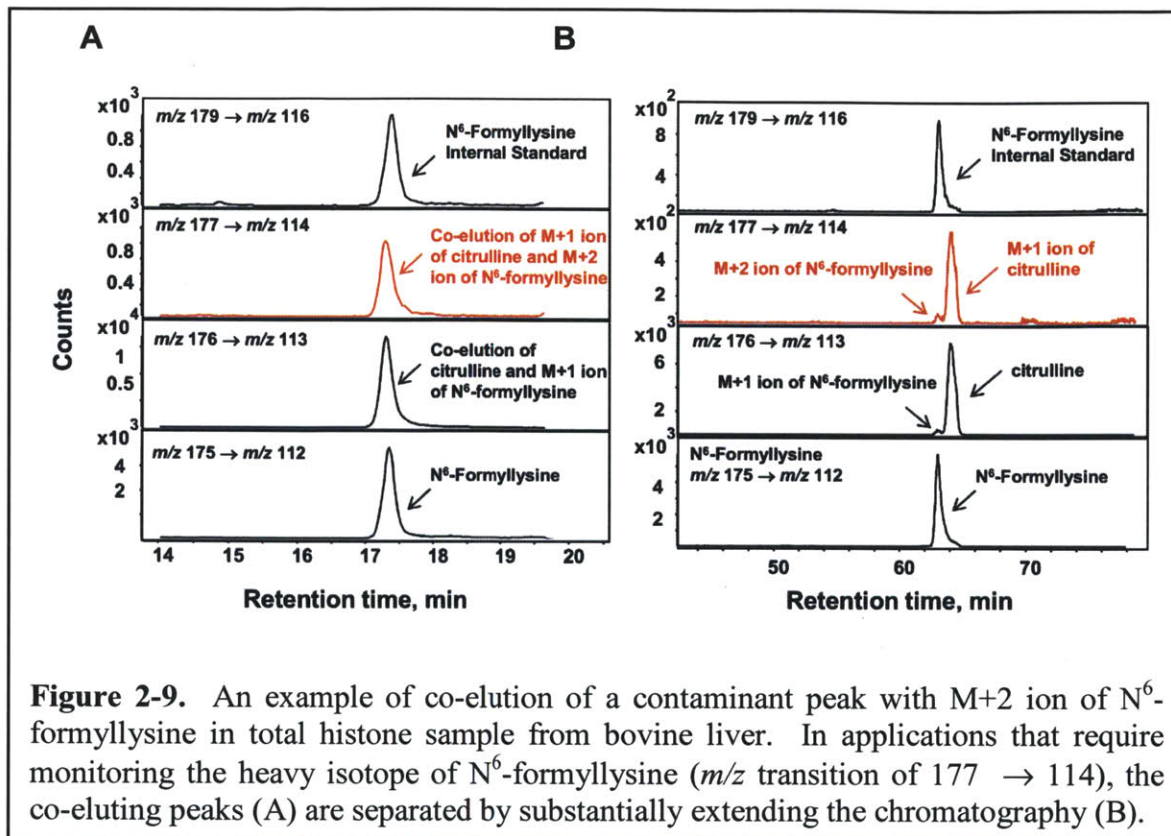
Figure 2-9. An example of co-elution of a contaminant peak with M+2 ion of N^6-formyllysine in total histone sample from bovine liver. In applications that require monitoring the heavy isotope of N^6-formyllysine (m/z transition of 177 → 114), the co-eluting peaks (A) are separated by substantially extending the chromatography (B).

Measuring the rate of dilution of N^6-formyllysine. The next application of the method involved determining the rate of loss of N^6-formyllysine in TK6 cells containing isotopically-labeled lysine in proteins. The chemical similarity of N^6-formyllysine and N^6-acetyllysine (Figure 1-1) suggested that the former might be subject to removal by histone deacetylases [4] or other protein repair or degradation processes. The dilution experiment was based on labeling TK6 cell proteins with [$^2H_8$]-lysine by growing the cells in customized RPMI medium containing 3,3,4,4,5,5,6,6-[^2H]-lysine for one week prior to resuspending them in non-labeled medium. Due to the average doubling time of 18 h for TK6 cells, histones were extracted at 0 h, 18 h, 36 h, and 54 h, corresponding to 0-3 replication cycles. The quantity of [$^2H_8$]-N^6-formyllysine, as a percentage of total labeled and unlabeled lysines, was then measured at each time point (Figure 2-10A). A

comparison of labeled and unlabeled lysine signals revealed > 90% labeling of all lysines and modified lysines, including N^6-formyllysine. Upon removal of the $[^2H_8]$-lysine, the rate of loss of $[^2H_8]$-N^6-formyllysine in histones was detected by mass spectrometry as ions that are 8 mass units heavier than the unlabeled counterparts. The results show a dilution rate of 15 ± 0.5h for $[^2H_8]$-N^6-formyllysine, which closely matches the cell doubling time of 16 ± 0.9h for this experiment (calculated based on $[^2H_8]$-lysine dilution curve as shown in Figure 2-10B).

Figure 2-10. Dilution rate of N^6-formyllysine in TK6 cells. The rate of dilution of $[^2H_8]$-N^6-formyllysine (A) closely matches the exact cell doubling time calculated based on $[^2H_8]$-lysine rate of dilution (B).
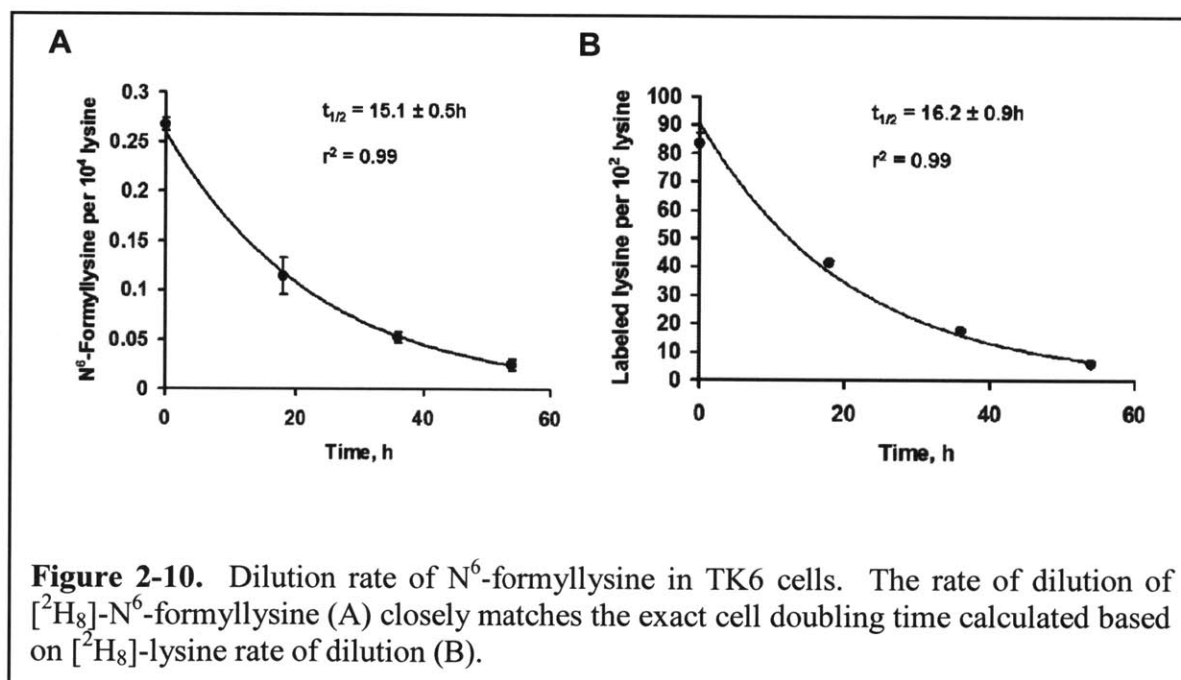
Discussion

We set out to develop a highly sensitive and specific LC-MS/MS method to accurately quantify N^6-formyllysine lesions in proteins extracted from cells and tissues, as a first step towards defining the biological relevance of lysine N^6-formylation. The quantification method we developed previously [1] proved to be relatively insensitive and biased, and thus major improvements were required. We were able to implement significant improvement to increase the efficiency as well as sensitivity and specificity of our previous N^6-formyllysine quantification method.

First, we discovered and optimized a new proteolytic enzyme, *Streptomyces griseus* protease, that proved to be highly efficient at digesting proteins to individual amino acids compared to proteinase K previously used (Figure 2-6). Using proteinase K at concentrations low enough to prevent the autolysis of enzyme resulted in only 10% digestion compared to complete digestion performed by comparable amount of *S. griseus* enzyme. In addition, we extended our quantitative analysis of proteins to include other lysine modifications such as acetylation and methylation besides formylation, and utilization of the new highly efficient enzyme allowed for detection of all lysine species (Table 2-2).

The HPLC step was further optimized to make the method more robust. The pre-purification step was eliminated and the need for PITC derivatization for chromatographic resolution of amino acids with reversed phase HPLC, was obviated by switching to normal phase HPLC. We were able to achieve baseline separation of analytes of interest in the resuspended protein hydrolysates (Figure 2-7). The elimination

of the derivatization step removed an extra step and further enhanced the sensitivity of the method.

The identification and quantification of amino acids were accomplished by isotope-dilution LC-MS/MS in positive ion mode, using MRM transitions. The discovery of the unique product ions of m/z 112 and m/z 126 for formyl and acetyl lysines, respectively, distinguished them from their isobaric compounds di- and tri-methyl lysines (Table 2-1 and Figures 2-3 and 2-4) while lowering the limits of detection by more than 10 fold (10 fmol to 1 fmol). The optimization of fragmentor voltage and collision energies (Table 2-1 and Figure 2-4) further decreased the limits of detection for all species. The new sensitive and specific method allows for robust quantification of all N^6-methyl-, -acetyl- and -formyl-lysine modifications, with limits of detection of 1 fmol for N^6-formyl- and N^6-acetyllysine, 10 fmol for lysine, and 50 fmol for each of N^6-mono-, di-, and tri-methyl lysine.

In addition to providing insights into N^6-formyllysine content of proteins, there are numerous applications of our quantification method for characterizing the chemical biology of this protein modification, as presented in the subsequent chapters. Here, we provided two examples of its immediate application. The first example was quantification of the background levels of lysine adducts in a sample of total histones extracted from TK6 cells. As shown in Table 2-2, N^6-formyllysine adducts occur at a low frequency (e.g., 1 per 10^4 lysines) in total histones compared to higher frequencies of acetylation and methylation (40 − 230 per 10^4 lysines). The next chapter covers the distribution of N^6-formyllysine adducts in different classes of histones separated by HPLC. Our quantification data is consistent with published studies of relative quantities

of histone modifications using immunologic and radiolabeling techniques [5-8]. Moreover, using our sensitive methods, we report ~1 lysine formylation per 10^4 lysines in total histones extracted from TK6 cells compared to 4-10 modification per 10^4 previously reported [1].

The other application of the quantification method presented here was to determine the rate of dilution of $[^2H_8]$-N^6-formyllysine in TK6 cells. Data showed that the dilution rate of 15 h for $[^2H_8]$-N^6-formyllysine closely matches the cell doubling time of 16 h (Figure 2-10), suggesting that N^6-formyllysine is either a stable modification of histones or is formed and removed at the same frequency, resulting in no overall change in its steady state levels. The observation that cell replication did not result in an apparent repair or removal of N^6-formyllysine adducts in histone proteins is consistent with a previous study of conservation of histone carcinogen adducts during replication [9].

Our robust and sensitive analytical method could readily be extended for quantification of other amino acid modifications in proteins, with the advantage of using a highly efficient yet mild proteolytic digestion as opposed to the traditionally used strong hydrochloric acid for protein hydrolysis that could remove the adduct of interest. The absolute quantification of a specific protein modification, in addition to mapping studies, could provide insights on various biological pathways in cells. For instance, mapping studies of N^6-formyllysine sites reveal that this adduct occurs on conserved lysines involved in epigenetic regulation [5], and through absolute quantification, we showed that this modification is rather stable or is formed and removed at the same frequency, resulting in no overall change in its steady state levels. Furthermore, this

method is used in subsequent chapters to quantify various lysine modifications in individual classes of histones. Histone modification-based signaling involves the location and number of specific modification targets within a histone protein as well as the frequency of modification of a target among all copies of a particular histone protein. Our quantification data provide some insight into this issue. Further, N^6-formyllysine quantification is used to prove possible mechanisms for its formation and shed light on its fate.

References

1. Jiang T, Zhou X, Taghizadeh K, Dong M, Dedon PC (2007) N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. Proc Natl Acad Sci U S A 104: 60-65.

2. Cai H, Guengerich FP (2000) Acylation of protein lysines by trichloroethylene oxide. Chem Res Toxicol 13: 327-335.

3. Boyne MT, 2nd, Pesavento JJ, Mizzen CA, Kelleher NL (2006) Precise characterization of human histones in the H2A gene family by top down mass spectrometry. J Proteome Res 5: 248-253.

4. Hake SB, Xiao A, Allis CD (2004) Linking the epigenetic 'language' of covalent histone modifications to cancer. Br J Cancer 90: 761-769.

5. Wisniewski JR, Zougman A, Mann M (2008) Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. Nucleic Acids Res 36: 570-577.

6. Pasqualini JR, Mercat P, Giambiagi N (1989) Histone acetylation decreased by estradiol in the MCF-7 human mammary cancer cell line. Breast Cancer Res Treat 14: 101-105.

7. Byvoet P, Barber M, Amidei K, Lowell N, Trudeau W (1986) Effect of exogenous histone H5 on integration of histone H1 in rat liver chromatin. Correlations with aberrant epsilon-N-methylation of histone H1. Biochim Biophys Acta 867: 163-175.

8. Wu M, Allis CD, Richman R, Cook RG, Gorovsky MA (1986) An intervening sequence in an unusual histone H1 gene of Tetrahymena thermophila. Proc Natl Acad Sci U S A 83: 8674-8678.

9. Ozbal CC, Velic I, SooHoo CK, Skipper PL, Tannenbaum SR (1994) Conservation of histone carcinogen adducts during replication: implications for long-term molecular dosimetry. Cancer Res 54: 5599-5601.

Chapter 3

Quantitative analysis of histone modifications: Formaldehyde is a source of pathological N^6-formyllysine that is refractory to histone deacetylases

Abstract

Aberrant protein modifications play an important role in the pathophysiology of many human diseases, in terms of both dysfunction of physiological modifications and the formation of pathological modifications by reaction of proteins with endogenous electrophiles. Recent studies have identified a chemical homolog of lysine acetylation, N^6-formyllysine, as an abundant modification of histone and chromatin proteins, one possible source of which is the reaction of lysine with 3'-formylphosphate residues from DNA oxidation. Using a new liquid chromatography-coupled to tandem mass spectrometry method to quantify all N^6-methyl-, -acetyl- and -formyl-lysine modifications, we now report that endogenous formaldehyde is a major source of N^6-formyllysine and that this adduct is widespread among cellular proteins in all compartments. N^6-Formyllysine was evenly distributed among different classes of histone proteins from human TK6 cells at 1-4 modifications per 10^4 lysines, which contrasted strongly with lysine acetylation and mono-, di- and tri-methylation levels of 1.5-380, 5-870, 0-1400 and 0-390 per 10^4 lysines, respectively. While isotope labeling studies revealed that lysine demethylation is not a source of N^6-formyllysine in histones, formaldehyde exposure was observed to cause a dose-dependent increase in N^6-formyllysine, with use of $[^{13}C,^2H_2]$-formaldehyde revealing unchanged levels of adducts derived from endogenous sources. Inhibitors of class I and class II histone deacetylases did not affect the levels of N^6-formyllysine in TK6 cells and the class III histone deacetylase, SIRT1, had minimal activity ($<10\%$) with a peptide substrate containing the formyl adduct. These data suggest that N^6-formyllysine is refractory to removal by histone deacetylases, which supports the idea that this abundant protein modification could interfere with normal regulation of gene expression if it arises at conserved sites of physiological protein secondary modification.

Introduction

In addition to physiological secondary modifications, proteins are subjected to reactions with endogenous electrophiles generated by oxidative stress, inflammation, and normal cell metabolic processes [1-5]. These adventitious or pathological modifications typically arise by reaction of the nucleophilic side chains of lysine, histidine, and cysteine with reactive electrophiles such as malondialdehyde, 4-hydroxynonenal (HNE), and glyoxal generated by oxidation of polyunsaturated fatty acids and carbohydrates, among other biomolecules [2-4,6,7]. The resulting adducts, which can alter protein function and lead to protein degradation, are associated with a variety of pathological processes and human diseases [1-5,8]. Among these pathological adducts, N^6-formylation of lysine has recently emerged as an abundant protein modification [5,9-11]. While originally described in chromatin proteins [9-11], it has since been identified as an adduct arising in proteins subjected to nitrosative and oxidative stresses [5]. In chromatin proteins, N^6-formyllysine has the potential to interfere with the functions of other post-translational modifications that perform signaling functions [12-15], such as acetylation, methylation, phosphorylation, ubiquitylation, and ADP ribosylation, with some locations modified in more than one way (e.g., refs. [16-18]). The chemical similarities of N^6-formyllysine and N^6-acetyllysine suggest a disruptive role for the former in signaling by histone acetylation. Indeed, N^6-formyllysine has been detected at conserved sites of lysine acetylation and methylation in histones [10,11].

While N^6-formyllysine adducts are now well recognized as abundant protein modifications in cells, the source of these pathological adducts remains unclear. We recently showed that some portion of N^6-formyllysine arises in chromatin proteins by

reaction of lysine side chains with the 3'-formylphosphate residue derived from 5'-oxidation of 2-deoxyribose in DNA in cells (Figure 3-1) [9]. However, the observation of this adduct in proteins treated with the biological oxidant, peroxynitrite, suggests other sources for the formyl species [5]. Considering that formaldehyde reacts with amines to give a carbinolamine (N^6-hydroxymethyllysine) intermediate that is only one oxidation state away from a formamide functional group (Figure 3-1), we hypothesized that endogenous formaldehyde could serve as a source of N^6-formyllysine residues in histone and other proteins. In addition to environmental and occupational sources [19-21], formaldehyde arises from cellular processes such as oxidative demethylation of nucleic acid and histone proteins, as well as biosynthesis of purines, thymidine, and some amino acids [20,22,23], making it a relatively abundant metabolite at concentrations ranging from 13 to 97 μM in human plasma [20]. To test this hypothesis and to clarify the cellular locations and quantities of N^6-formyllysine relative to other major histone modifications, we developed a novel liquid chromatography-coupled to electrospray tandem mass spectrometry (LC-MS/MS) method to quantify all N^6-methyl-, -acetyl-, and -formyl-lysine modifications. Application of this method reveals that endogenous formaldehyde is a major source of N^6-formyllysine, that this adduct is widespread among proteins in all cellular compartments, and that, in chromatin proteins, it is refractory to removal by histone deacetylases.
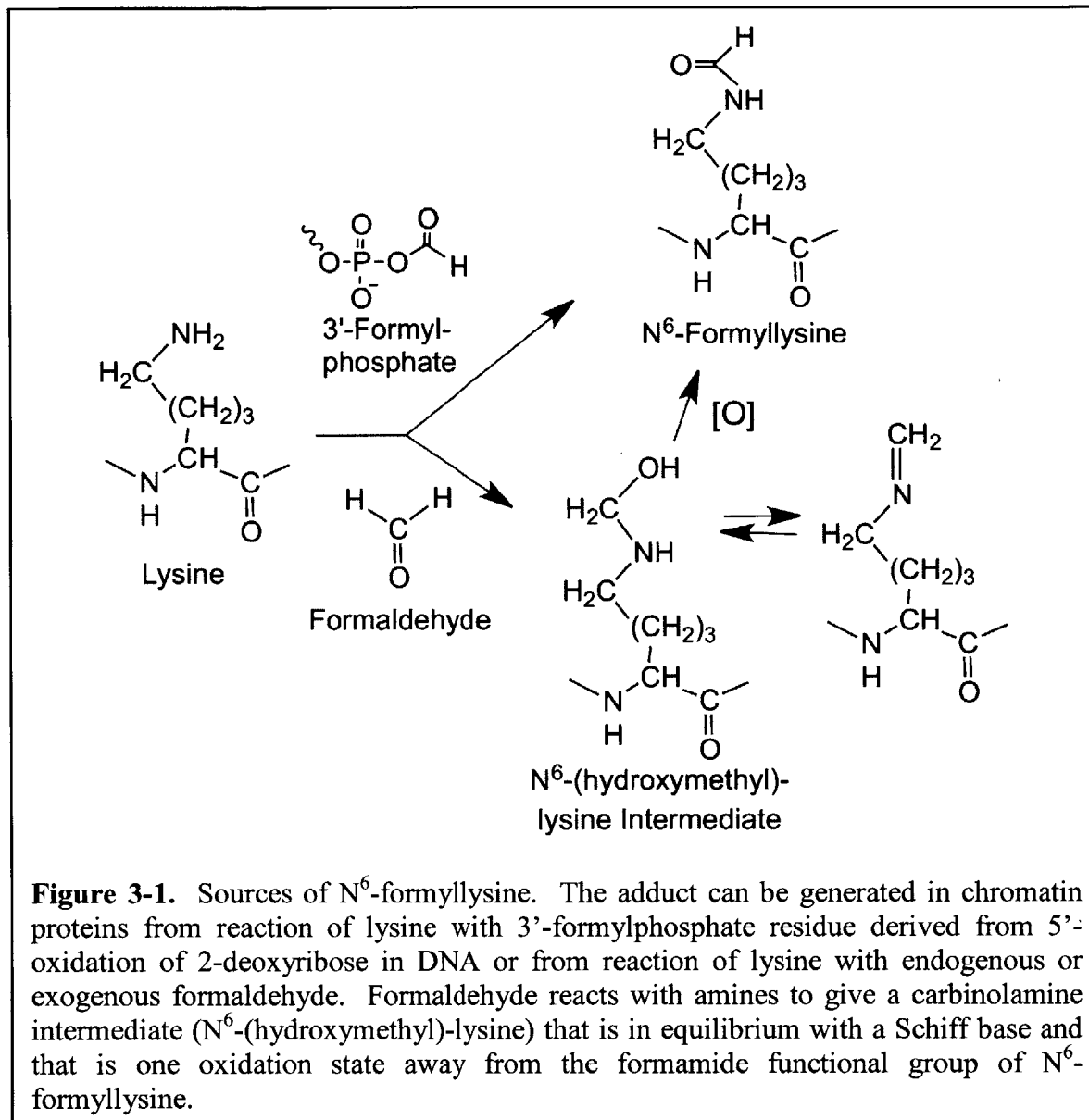
Figure 3-1. Sources of N^6-formyllysine. The adduct can be generated in chromatin proteins from reaction of lysine with 3'-formylphosphate residue derived from 5'-oxidation of 2-deoxyribose in DNA or from reaction of lysine with endogenous or exogenous formaldehyde. Formaldehyde reacts with amines to give a carbinolamine intermediate (N^6-(hydroxymethyl)-lysine) that is in equilibrium with a Schiff base and that is one oxidation state away from the formamide functional group of N^6-formyllysine.

Materials and Methods

Materials. Unlabeled and [^{13}C,^{2}H$_2$]-labeled formaldehyde were purchased as 37% and 20% aqueous solutions from Amresco (Solon, OH) and Isotec (Miamisburg, OH), respectively. 4,4,5,5-[^{2}H]-Lysine was purchased from Cambridge Isotope Laboratories

(Andover, MA). 4,4,5,5-[^{2}H]-N^6-Formyllysine was synthesized from 4,4,5,5-[^{2}H]-lysine according to Jiang *et al.* [9]. 3,3,4,4,5,5,6,6-[^{2}H]-N^6-acetyllysine were obtained from CDN Isotopes (Pointe-Claire, Quebec, Canada). L-Methionine-([^{13}C,^{2}H$_3$]-methyl) was obtained from Isotec (Miamisburg, OH). Lysine, N^6-formyllysine, N^6-acetyllysine, bovine serum albumin, human recombinant HMG-1, human IgG, *Streptomyces griseus* protease, and protease inhibitor cocktail (for use with mammalian cell and tissue extracts) were obtained from Sigma-Aldrich (St. Louis, MO). N^6-Mono-methyllysine, N^6-di-methyllysine, and N^6-tri-methyllysine were purchased from Bachem Bioscience Inc. (King of Prussia, PA). Nonidet P-40 was from Roche Diagnostic Corporation (Indianapolis, IN). Suberoylanilidehydroxamic acid (SAHA) and SIRT1 (human recombinant) enzyme were purchased from Cayman chemical (Ann Arbor, MI). Peptide substrates for SIRT1 (GGAKRHR and its lysine-acetylated and -formylated forms) were synthesized at Massachusetts Institute of Technology Biopolymers Laboratory. The human lymphoblastoid TK6 cell line was a generous gift of Prof. Gerald Wogan (Massachusetts Institute of Technology).

TK6 cell culture, exposure, and labeling. TK6 cells were cultured in RPMI 1640 medium (Cellgro, Manassas, VA) supplemented with 10% heat-inactivated horse serum (Atlanta Biologicals, Lawrenceville, GA), 10,000 U penicillin/ml and 10,000 µg streptomycin/ml (Lonza, Walkersville, MD), and 2 mM L-glutamine (Lonza, Walkersville, MD) at 37 °C in a 5% CO$_2$ atmosphere. For formaldehyde exposure studies, TK6 cells were pelleted, washed, and resuspended in RPMI 1640 medium without any supplements, prior to addition of formaldehyde to the medium. Following

addition of formaldehyde, cells were incubated at 37 °C for 2 h with occasional mixing

prior to extracting chromatin proteins. Histones were extracted from ~10^7 cells after

exposure and the quantity of formyllysine, as a percentage of total lysine, was measured

as described below. For lysine demethylation studies, TK6 cells were grown in a

customized RPMI-1640 medium identical to the traditional medium (e.g., supplemented

with horse serum, antibiotics, and L-glutamine), except for the presence of labeled

methionine (L-methionine-([$^{13}C,^2H_3$]-methyl)) instead of non-labeled methionine.

Histones (from ~10^7 cells) were extracted every 2 d for 20 d in order to investigate the

formation of N^6-[$^{13}C,^2H$]-formyllysine. For histone deacetylase studies, TK6 cells were

incubated with the histone deacetylase inhibitor, SAHA, for 20 h at 37 °C in a 5% CO_2

atmosphere prior to histone extraction. SAHA was dissolved in a 50:50 solution of

DMSO:PBS prior to addition to cell media. Control cells (~10^7) were treated with the

DMSO:PBS vehicle.

Histone extraction from TK6 cells and subcellular protein fractionation from tissue.

For detailed procedure regarding histone extraction, refer to *Materials and Methods*

section of Chapter 2. For collecting membrane, cytosolic, and nuclear fractions from

tissues, 20 mg of bovine liver tissue was cut into small pieces and washed with PBS, and

proteins were fractionated using the Subcellular Protein Fractionation Kit from Thermo

Scientific (Waltham, MA) and a Kontes all-glass Dounce homogenizer (10 strokes with a

type B pestle). Proteins in subcellular extracts were precipitated by addition of 20% v/v

trichloroacetic acid and overnight incubation at 4 °C. Samples were then centrifuged at

14000 x g for 10 min at 4 °C, washed once with ice-cold acetone containing 0.1% HCl,

and once with ice-cold acetone. The extracts were air-dried and stored at -20 ° C until use.

Purification of individual histones. HPLC purification of total histones was performed according to Boyne *et al.* [24] with modifications. For detailed procedure of histone purification into individual amino acids, refer to *Materials and Methods* section of Chapter 2.

Enzymatic hydrolysis of proteins. Proteins were hydrolyzed to individual amino acids using *S. griseus* protease (freshly prepared solution each time) with incubation at 37 °C for ≥ 16 h. For detailed procedure regarding protein digestion into individual amino acids, refer to *Materials and Methods* section of Chapter 2.

Quantification of amino acids. N^6-Formyllysine and other amino acids of interest were quantified as a percentage of the total quantity of lysine, by liquid chromatography-coupled mass spectrometry (LC-MS/MS). For detailed procedure regarding LC-MS/MS based quantification of individual amino acids, refer to *Materials and Methods* section of Chapter 2.

SIRT1 peptide experiment. SIRT1 peptide substrates (GGAKRHR, GGAK$_{acetyl}$RHR, and GGAK$_{formyl}$RHR) were HPLC purified on an Agilent 1100 series instrument using Vydac 218TP52 C18 reverse-phase column (2.1 x 250 mm, 5 μm) from Grace Vydac (Hesperia, CA). The mobile phase flow rate was 200 μL/min, and the column

temperature was maintained at 30 °C. The solvent system was 0.05% trifluoroacetic acid in water (A) and 0.05% trifluoroacetic acid in 80% acetonitrile (B), with the isocratic elution of 5% B for the first 5 min, then a linear increase to 42% B over 25 min, reaching 100% B at 31 min followed by column re-equilibration at 5% B for 10 min. Each purified SIRT1 peptide substrate (100 pmol) was incubated overnight (12 h) at 25 °C with 1 μg SIRT1, in 50 mM Tris-HCl (pH 8) buffer containing 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$, and 6 mM NAD^+. The removal of acetyl and formyl groups from SIRT1 peptide substrates was determined using liquid chromatography-coupled mass spectrometry. HPLC was performed on an Agilent 1100 series instrument using Agilent Exclipse XDB C18 reverse-phase column (4.6 x 150 mm, 5 μm). The mobile phase flow rate was 200 μL/min, and the column temperature was maintained at 40 °C. The solvent system was 0.1% acetic acid in water (A) and 0.1% acetic acid in acetonitrile (B), with the elution starting at 20% B, the gradient linearly increased to 50% B over 5 min, reached 100% B at 6 min, and kept at 100% B for 9 minutes before the column was re-equilibrated at 20% B for 10 min. The species of interest were then analyzed using the Agilent 6410 MS/MS system equipped with an electrospray ionization (ESI) source operated in positive ion mode. The operating parameters were as follows: ESI capillary voltage, 3500 V; gas temperature, 345 °C; drying gas flow, 8 L/min; and nebulizer pressure, 30 psi. Multiple reaction monitoring (MRM) transitions were as follow: GGAKRHR peptide, m/z 781.1 → 625.2; GGAK$_{formyl}$RHR peptide, m/z 809.4 → 516.3; and GGAK$_{acetyl}$RHR peptide, m/z 823.4 → 530.4. The fragmentor voltage and collision energy were 200 V and 40 V for GGAKRHR peptide, respectively; 100 V and 46 V for GGAK$_{formyl}$RHR peptide; and 100 V and 52 V for GGAK$_{acetyl}$RHR peptide.

Results

Quantification of N⁶-lysine modifications in proteins. In order to quantify N^6-formyllysine adducts, proteins are first subjected to proteolytic digestion to generate individual amino acids that are then analyzed using our improved LC-MS/MS based method, as outlined in detail in Chapter 2. To validate the new analytical method for lysine modifications, we compared the frequency of N^6-formyllysine among different classes of histone proteins extracted from TK6 cells and resolved by reversed-phase HPLC. As shown in Figure 3-S1A, all of the major histone classes were separated, with further resolution of variant forms of histones H1 and H3 (SDS-PAGE verification in Figure 3-S1B), which is consistent with previous observations in cultured human cells [10]. N^6-Formyllysine was detected in all histone classes at a frequency of 1-4 modifications per 10^4 lysines. This 3- to 4-fold variation among histone classes stands in contrast to the 10- to 100-fold variation in the frequency of other modifications (Table 3-1). The data in Table 3-1 represent the first absolute quantification of the various lysine acetyl and methyl modifications in histone proteins, and are consistent with published studies of relative quantities of histone modifications using immunologic and radiolabeling techniques [10,25-27]. Histone modification-based signaling involves the location and number of specific modification targets within a histone protein, as well as the frequency of modification of a target among all copies of a particular histone protein. Our data provide some insight into this issue. For example, we observed low-level acetylation and methylation in histone H1, which is consistent with studies using radiolabeled acetate [25], while this low level of modification maps to specific sites in the globular domain and N-terminal tail of histone H1 [10]. This low-level of acetylation

and methylation in histone H1 stands in contrast to the high level of acetylation of H2, H3 and H4 (Table 3-1), which is again supported by studies using radiolabeled acetate [25].

Table 3-1. Quantification of lysine modifications in HPLC-purified histone proteins.[1]

	Formyl[2]	Acetyl	Methyl	Dimethyl	Trimethyl
H1 (1)[1]	1.5 ± 0.3[3]	1.6 ± 0.2	8.0 ± 2.0	ND	ND
H1 (2)[1]	1.2 ± 0.2	1.5 ± 0.3	7.0 ± 2.0	ND	ND
H2A	2.7 ± 0.2	26 ± 20	20 ± 8	60 ± 30	ND
H2B	1.7 ± 0.7	66 ± 20	5.0 ± 7.0	ND	85 ± 30
H3(1)[1]	3.9 ± 0.5	280 ± 130	680 ± 80	1100 ± 220	220 ± 60
H3(2)[1]	3.6 ± 1.3	380 ± 100	870 ± 30	1400 ± 270	390 ± 140
H4	2.6 ± 0.4	73 ± 60	260 ± 80	740 ± 40	3.0 ± 6.0

[1]Classes of histone proteins resolved by reversed-phase HPLC, with putative isoforms denoted in parentheses.

[2]Column titles denote different N^6-modifications of lysine.

[3]Data are expressed as modifications per 10^4 total lysines and represent mean $\pm$ SD for 3 biological replicates.

The new analytical method was next applied to quantify N^6-formyllysine in non-histone proteins. We had previously observed N^6-formyllysine mainly in histone proteins [9], perhaps as a result of biased proteolysis or subsequent steps in the technique. However, using the new method, we are now able to detect N^6-formyllysine modifications in a variety of different proteins, as shown in Table 3-2. Further, an analysis of proteins in nuclear, cytosolic, and membrane compartments in bovine liver revealed the presence of N^6-formyllysine in all three locations (Table 3-2). These observations are consistent with a source for N^6-formyllysine other than the 3'-formylphosphate residues of DNA oxidation previously identified for histone proteins [9].

Table 3-2. Quantification of N^6-formyllysine in different proteins.

Identity of Protein	N^6-Formyllysine per 10^4Lys
BSA	5 ± 0.5
IgG	2 ± 0.4
Collagen	2 ± 0.5
HMG-1	< 0.2 ± 0.04
Bovine liver nuclear proteins	2 ± 0.6
Bovine liver membrane proteins	4 ± 1.0
Bovine liver cytosolic proteins	4 ± 0.6

Formaldehyde as a source of N^6-formyllysine. One alternative to 3'-formylphosphate residues as a source of N^6-formyllysine is oxidation of the carbinolamine (N^6-hydroxymethyllysine) intermediate in the reaction of formaldehyde with side chain amine of lysine (Figure 3-1). To test this hypothesis, we performed a series of experiments, starting with an *in vitro* reaction of L-lysine with different concentrations of formaldehyde and quantification of N^6-formyllysine. As shown in Figure 3-2A, there was a concentration-dependent formation of N^6-formyllysine in reactions with formaldehyde, presumably as a result of oxidation of the carbinolamine adduct by the background of reactive oxygen species generated by trace metals and dissolved oxygen in the solution. This dose-response relationship for N^6-formyllysine formation was also observed in histone proteins extracted from TK6 cells exposed to formaldehyde for 2 h at 37 °C (Figure 3-2B), with 10 mM formaldehyde producing roughly the same frequency

Figure 3-2. Formaldehyde is a source of N^6-formyllysine. Formation of N^6-formyllysine in (A) *in vitro* reactions of 1 mM L-lysine with formaldehyde for 2 h at 37 °C, and in (B) TK6 cells exposed to formaldehyde, as described in *Materials and Methods*. Data represent mean ± SD for N=3, with asterisks denoting statistically significant differences by Students t-test ($p < 0.05$).
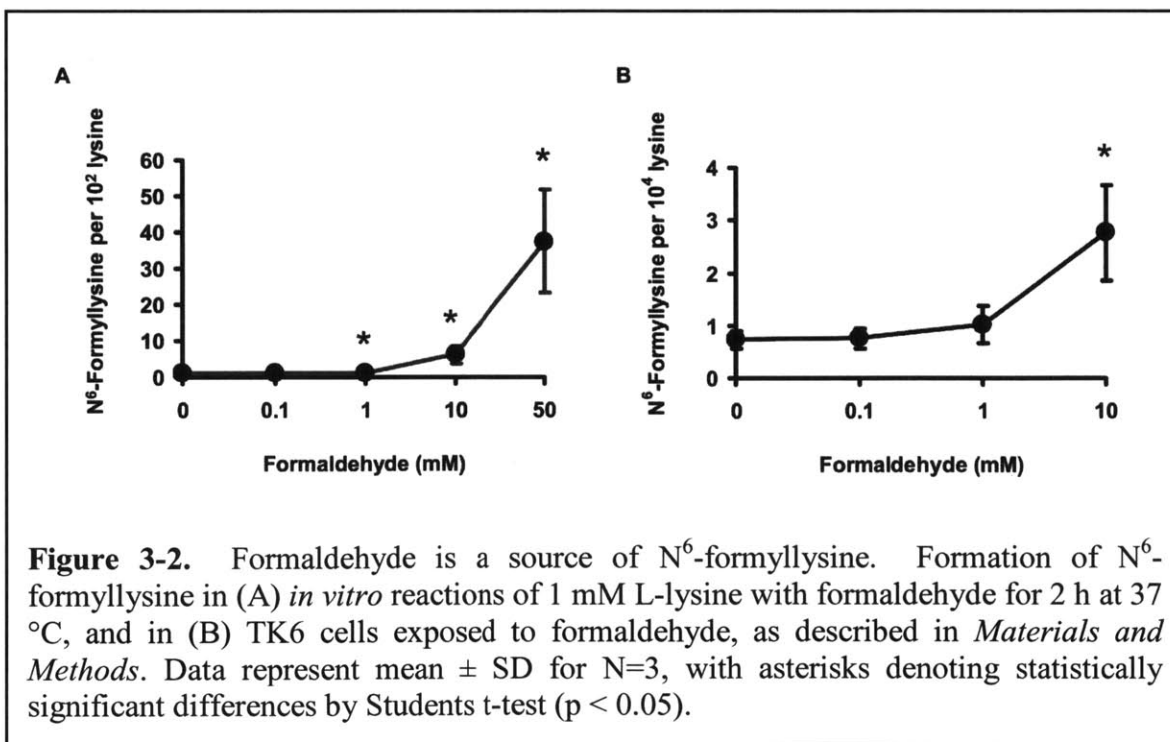
of N^6-formyllysine in both *in vitro* and cellular studies (i.e., 6 per 10^4 lysine *versus* 3 per 10^4 lysine, respectively).

The relatively high endogenous levels of N^6-formyllysine in histone and other proteins (Tables 3-1 and 3-2) raised the question of the contribution of exogenous formaldehyde exposures to the total load of N^6-formyllysine in the cells. To address this issue, we exposed TK6 cells to $[^{13}C,^2H_2]$-labeled formaldehyde, which led to the formation of N^6-$[^{13}C,^2H]$-formyllysine that is 2 mass units heavier than the endogenous adducts (Figure 3-3A). Following extraction of the histone proteins from formaldehyde-treated TK6 cells (2 h, 37 °C), both endogenous and exogenous N^6-formyllysine were quantified by monitoring the transitions *m/z* 175 → 112 and *m/z* 177 → 114, respectively (Figure 3-3A), with a third transition (*m/z* 179 → 116) for the 4,4,5,5-$[^2H]$-N^6-formyllysine internal standard. As shown in Figure 3-3B, levels of endogenous (unlabeled) N^6-formyllysine remained constant at all concentrations of $[^{13}C,^2H_2]$-formaldehyde, while N^6-$[^{13}C,^2H]$-formyllysine increased as a function of the concentration of labeled formaldehyde.

Lysine demethylation as a source of N^6-formyllysine. The enzymatic demethylation of N^6-methyllysine modifications represents another possible source of N^6-formyllysine in histone proteins, given both the carbinolamine intermediate known to form during the process of lysine demethylation and the ultimate release of the methyl group as formaldehyde [23]. Adventitious oxidation of the carbinolamine intermediate or secondary reaction of the released formaldehyde could result in the formation of N^6-formyllysine locally. To test these hypotheses, TK6 cells were grown in customized

RPMI medium containing L-methionine with a $[^{13}C,^2H_3]$-methyl group for 20 days to label all methyl groups in N^6-methyllysine species, and histone proteins were extracted for analysis every 2 days. If N^6-formyllysine is a product of disrupted lysine

Figure 3-3. Addition of $[^{13}C,^2H_2]$-formaldehyde to TK6 cells distinguishes exogenous from endogenous sources of N^6-formyllysine. (A) LC-MS/MS analysis showing signals for the three isotopomeric N^6-formyllysine species, as described in *Materials and Methods*. (B) Plot of N^6-formyllysine levels as a function of exposure to $[^{13}C,^2H_2]$-formaldehyde. Data represent mean ± SD for N=3.
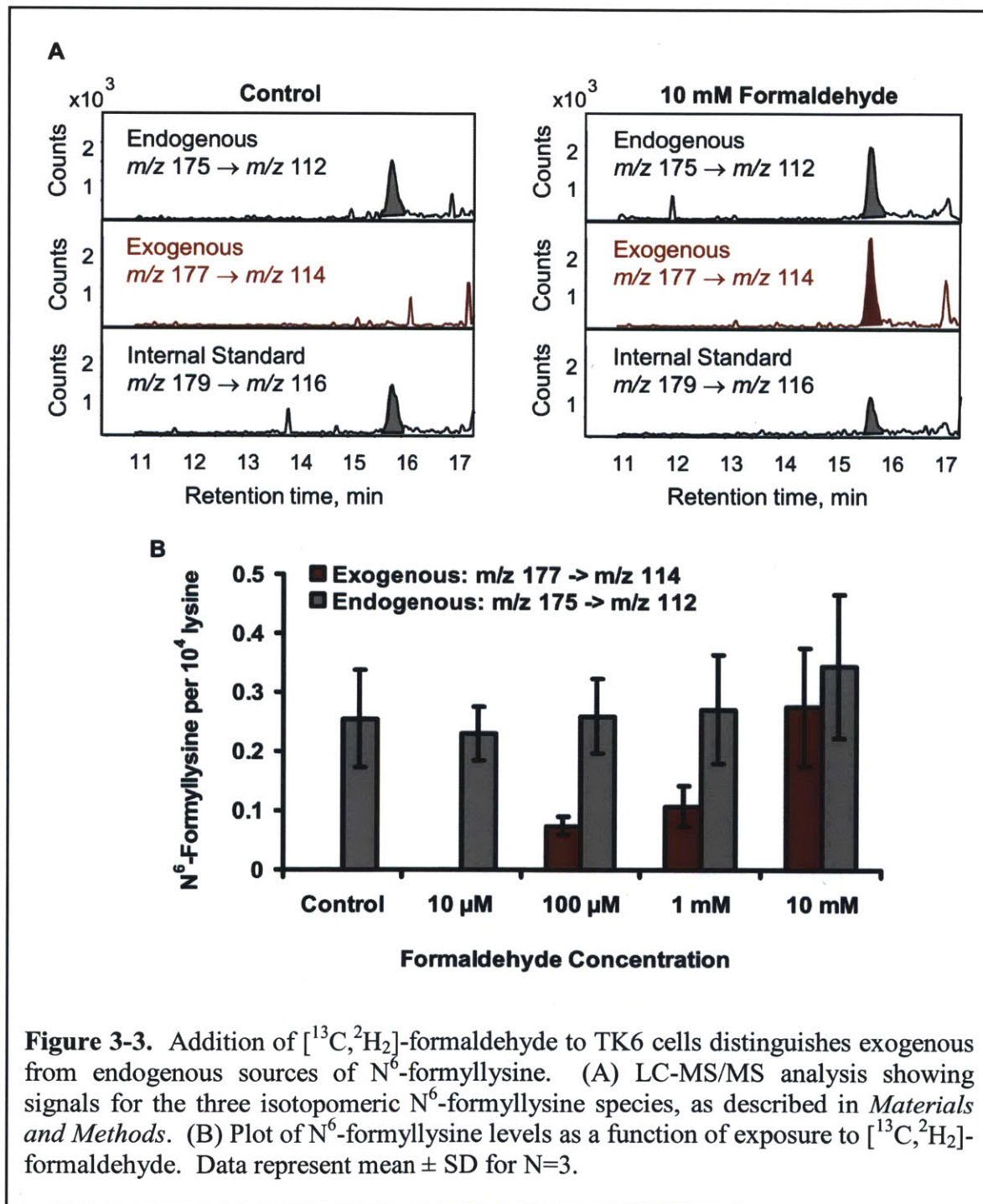
demethylation in histones and is formed via oxidation of the carbinolamine intermediate known to form during the process of lysine demethylation [23], or by reaction of lysine with the formaldehyde released at the last step of successful lysine demethylation [23], then one would expect to see an increase of 2 mass units corresponding to formation of N^6-[^{13}C, ^{2}H]-formyllysine (*m/z* 177 → 114 transition). In order to increase the signal-to-noise ratio for N^6-[^{13}C,^{2}H]-formyllysine, N^6-formyllysine was HPLC-pre-purified in all samples before LC-MS/MS analysis. Figure 3-4 depicts an example of the analysis using the day 6 sample. As shown in Figure 3-4, N^6-mono-methyllysine and N^6-di-methyllysine are predominately labeled (>90%) with heavy isotope methyl groups (i.e., [^{13}C,^{2}H$_3$]-methyl). In contrast to methyllysines, the level of N^6-[^{13}C,^{2}H]-formyllysine did not increase beyond the natural isotope abundance level of ~0.7% for the [M+2] ion of N^6-formyllysine (Figure 3-4C and Figure 3-S2). Note that the HPLC gradient was changed here to fully resolve a contaminant signal from the TK6 cells (identified as the [M+1] ion of citrulline) that otherwise co-eluted with N^6-formyllysine and produced an *m/z* value similar to the [M+2] isotopomer of N^6-formyllysine.

Figure 3-4. Analysis of lysine demethylation as a source of N^6-formyllysine. Methyl groups in N^6-methyllysine species in TK6 cells were labeled using L-methionine-($[^{13}C,^2H_3]$-methyl) and N^6-formyllysine and N^6-methyllysine species were quantified by LC-MS/MS as described in *Materials and Methods*. Panels A and B: N^6-mono-methyllysine and N^6-di-methyllysine are predominately labeled (>90%) with heavy isotope methyl groups (mass increase of 4 *m/z* and 8 *m/z*, respectively), with <10% of the modifications containing unlabeled methyl groups. Panel C: the level of N^6-$[^{13}C,^2H]$-formyllysine (*m/z* 177 → 114 transition) in histones did not show an increase beyond the natural isotope abundance level of ~0.7% for [M+2] ion of N^6-formyllysine.

Persistence of N^6-formyllysine in cells. The chemical similarity of N^6-formyllysine and N^6-acetyllysine suggested that the former might be subject to removal by lysine deacetylases that recognize and remove N^6-acetyllysine from histone and other proteins [16,28-31]. Lysine deacetylases fall into several classes, including classes I and II that share a common hydrolytic mechanism and are all inhibited by suberoylanilidehydroxamic acid (SAHA), and the class III enzymes (i.e., sirtuins) that are NAD^+-dependent deacetylases [32,33]. In order to assess the activity of lysine deacetylases with N^6-formyllysine substrates, TK6 cells were treated with SAHA and the levels of N^6-acetyllysine and N^6-formyllysine were quantified. The results shown in Figure 3-5A reveal that, while SAHA caused a 3-fold increase in the level of N^6-acetyllysine (4 to 14 per 10^3 lysines), lysine formylation was not affected. To assess sirtuin activity against N^6-formyllysine, we performed *in vitro* reactions of SIRT1 with a consensus peptide (GGAKRHR) containing N^6-formyllysine or N^6-acetyllysine, and the quantities of the modified and unmodified peptides were analyzed by LC-MS/MS. As shown in Figure 3-5B, SIRT1 removed the acetyl modification completely to generate the unmodified peptide, while only ~10% ($\pm$ 4%) of N^6-formyllysine was removed.

Figure 3-5. Effect of lysine deacetylases on N^6-formyllysine. (A) TK6 cells were treated with the class I and class II histone deacetylase inhibitor, SAHA, as described in *Materials and Methods*. Data represent mean ± SD for N=3, with asterisks denoting statistically significant differences by Students t-test (p < 0.05). (B) Treatment of a peptide substrate containing N^6-formyllysine with the class III histone deacetylase, SIRT1.

Discussion

N^6-Formyllysine was first reported in 1985 in reactions of lysine with formaldehyde *in vitro* [34] and, more recently, it was shown to form during *in vitro* silver-staining procedures that involve the use of formaldehyde [35]. The recent discovery of N^6-formyllysine as a relatively abundant endogenous posttranslational modification of histones and other nuclear proteins in cells [9-11] has raised questions about its mechanism of formation and its potential for interfering with the regulatory function of lysine N^6-acetylation. With respect to formation, we previously presented evidence that N^6-formyllysine in histones could arise from reactions with 3'-formylphosphate residues derived from DNA oxidation [9]. However, formaldehyde emerged as an alternative source given the high reactivity of formaldehyde toward primary amines, such as the side chain of lysine, and the potential for endogenous oxidation to convert a formaldehyde-derived carbinolamine to a stable formamide (Figure 3-1). We have now demonstrated *in vitro* and in cells that formaldehyde exposure leads to formation of N^6-formyllysine residues in proteins. The fact that this modification arises in proteins other than chromatin proteins and in cellular compartments other than the nucleus (Table 3-2) suggests that 3'-formylphosphate residues in oxidized DNA do not account for all N^6-formyllysine adducts. This is consistent with recent studies in which N^6-formyllysine was detected in a protein oxidized with peroxynitrite *in vitro* [5]. The absence of detectable N^6-formyllysine arising from demethylation of N^6-methyllysine species (Figure 3-4C and Figure 3-S2) suggests that interruption of histone demethylation reactions to form the carbinolamine precursor of N^6-formyllysine occurs at low frequency, or that the formaldehyde produced by complete lysine demethylation [23]

does not occur at concentrations high enough to drive formylation of lysine or cause substantial changes in N^6-formyllysine levels detected by our current analytical method. Furthermore, there is the possibility that the formaldehyde released during lysine demethylation may be scavenged before it could react with lysines in histones. A recent study reports that lysine-specific demethylase 1 (LSD1) is a folate binding protein [36], which led the authors to hypothesize that the biological function of folate is to trap the formaldehyde that is generated during lysine demethylation [36]. In addition, the observation that the formaldehyde equivalent derived from histone demethylation might not account for a significant portion of formyllysine residues is not surprising in light of the abundance of formaldehyde from other cellular processes [20,22,23]. This is clear from the high steady-state concentrations of formaldehyde in human plasma (13-97 µM) [20].

The relative abundance of N^6-formyllysine in histone and other proteins (Tables 3-1 and 3-2) [9-11] and the persistence of these adducts in histone proteins provides insights into both their source and their potential effects on cell function. The N^6-formyllysine residues are relatively evenly distributed among different classes of histone proteins (Table 3-1), while the other functional modifications show very biased distributions over a large frequency range, which is consistent with the known function and conserved locations for lysine methylation and acetylation [16-18]. This random distribution of formyllysine adducts in histone proteins suggests that they are adventitious and not physiological. The fact that N^6-formyllysine levels are similar in histone and non-nuclear proteins and in all cell compartments also suggests that the sources of this protein modification are equally balanced in the various compartments and proteins, or

that there is a single dominant source that distributes uniformly throughout the cell. With regard to their persistence in cells, the N^6-formyllysine adducts appear to be refractory to removal by histone deacetylase enzymes, which suggests that they will persist throughout the life of individual histone proteins. Figure 3-6 summarizes our findings presented here on N^6-formyllysine adducts.

The isotope labeling studies revealed that lysine demethylation is not a source for N^6-formyllysine in histones. Furthermore, our data suggest that N^6-formyllysine is refractory to removal by histone deacetylases, which is consistent with the persistence of this pathological adduct throughout the life of individual histone proteins. That the N^6-formyllysine adducts have been observed at many of the known conserved functional locations for lysine acetylation and methylation in histones [10,11] suggests that N^6-formyllysine could interfere with signaling processes associated with physiological histone modifications [16,28]. The association of N^6-formyllysine with a variety of different cell and organismal processes, including metabolism, and the oxidative and nitrosative stresses of inflammation [5,9], suggest that this adduct may play a role in many pathophysiological processes in humans.
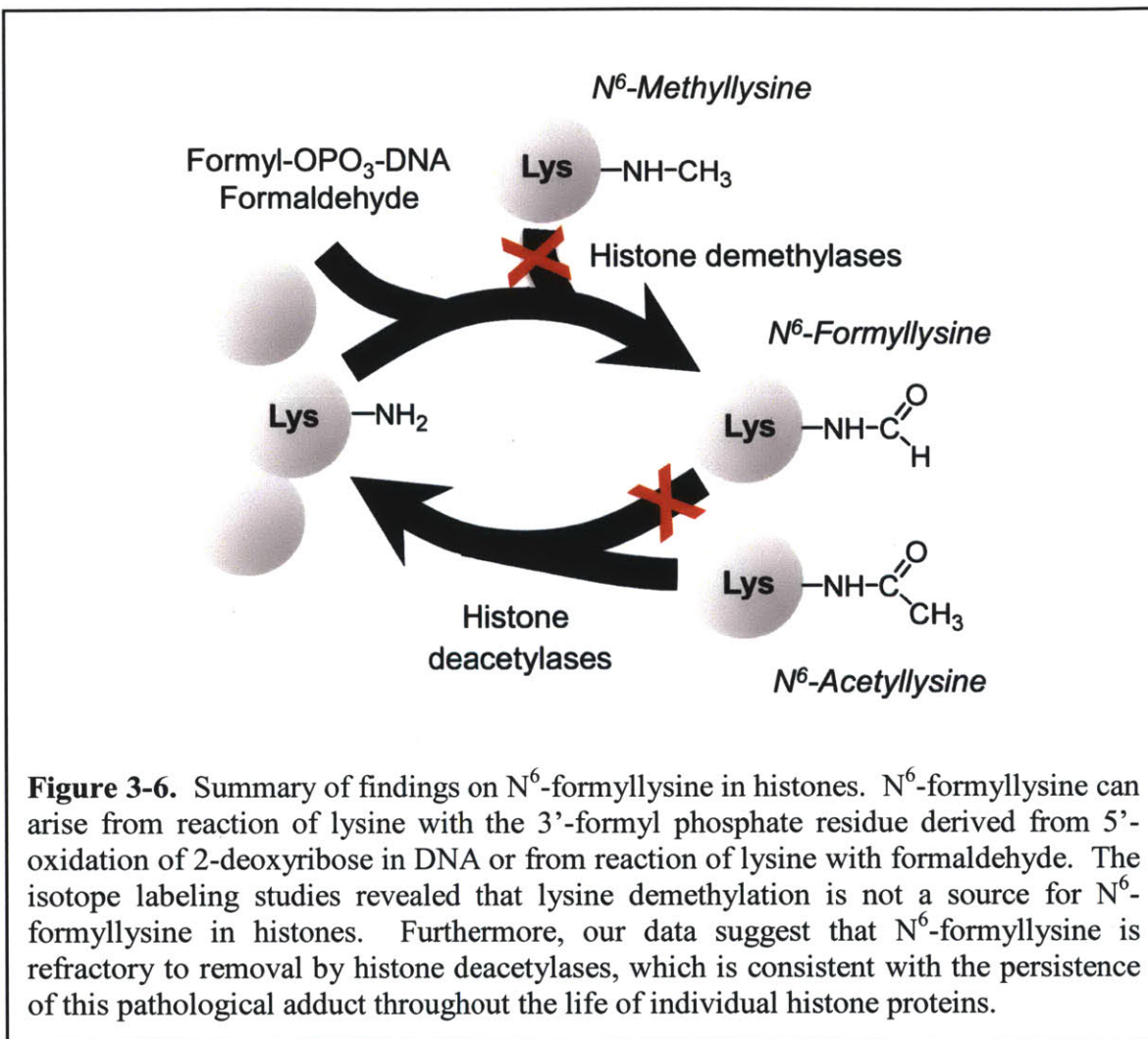
Figure 3-6. Summary of findings on N^6-formyllysine in histones. N^6-formyllysine can arise from reaction of lysine with the 3'-formyl phosphate residue derived from 5'-oxidation of 2-deoxyribose in DNA or from reaction of lysine with formaldehyde. The isotope labeling studies revealed that lysine demethylation is not a source for N^6-formyllysine in histones. Furthermore, our data suggest that N^6-formyllysine is refractory to removal by histone deacetylases, which is consistent with the persistence of this pathological adduct throughout the life of individual histone proteins.

References

1. Levine RL (2002) Carbonyl modified proteins in cellular regulation, aging, and disease. Free Radic Biol Med 32: 790-796.

2. Jacobs AT, Marnett LJ (2010) Systems analysis of protein modification and cellular responses induced by electrophile stress. Acc Chem Res 43: 673-683.

3. Thornalley PJ (2008) Protein and nucleotide damage by glyoxal and methylglyoxal in physiological systems--role in ageing and disease. Drug Metabol Drug Interact 23: 125-150.

4. Dedon PC (2008) The chemical toxicology of 2-deoxyribose oxidation in DNA. Chem Res Toxicol 21: 206-219.

5. Vana L, Kanaan NM, Hakala K, Weintraub ST, Binder LI (2011) Peroxynitrite-induced nitrative and oxidative modifications alter tau filament formation. Biochemistry 50: 1203-1212.

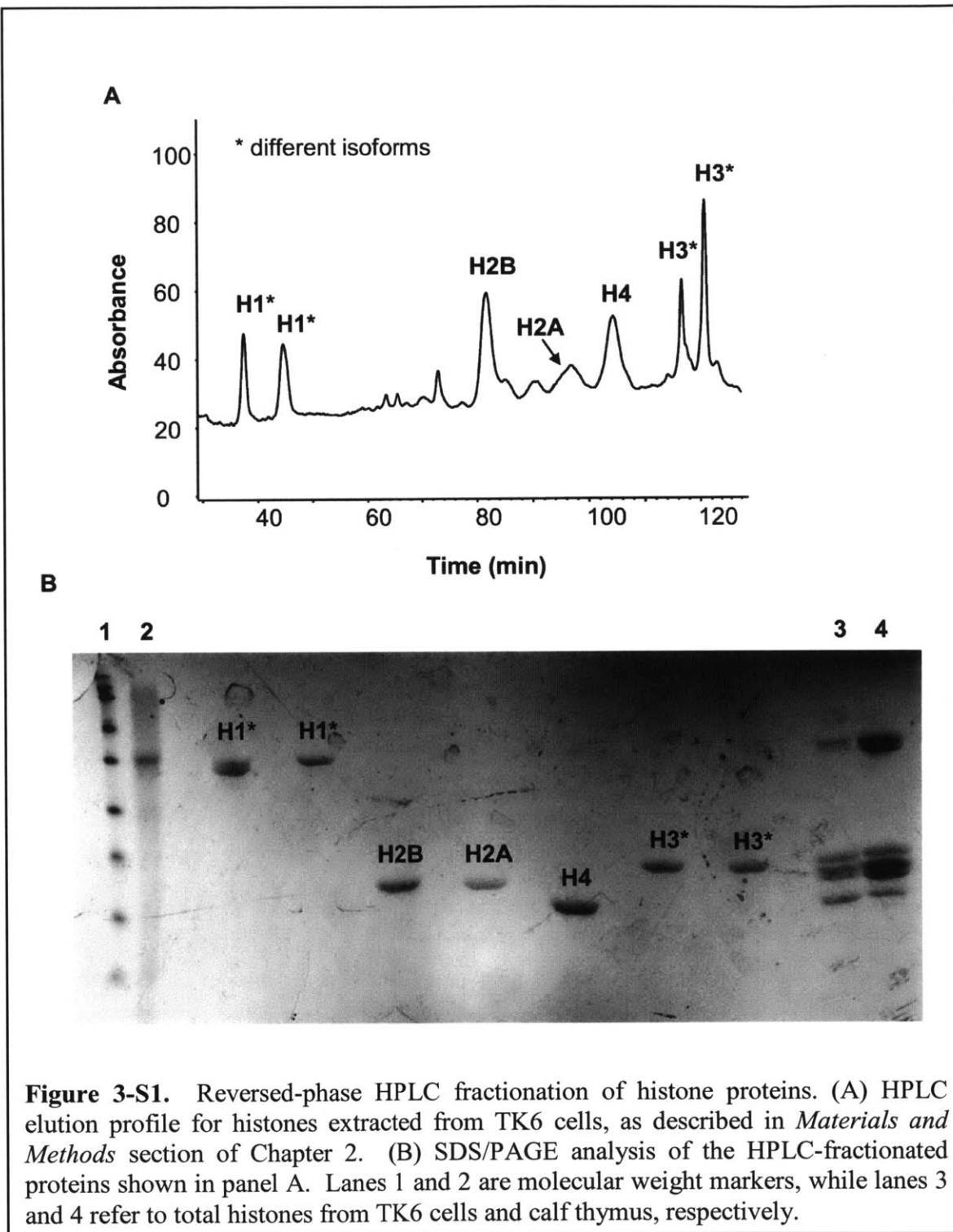6. Codreanu SG, Zhang B, Sobecki SM, Billheimer DD, Liebler DC (2009) Global analysis of protein damage by the lipid electrophile 4-hydroxy-2-nonenal. Mol Cell Proteomics 8: 670-680.

7. Tallman KA, Kim HY, Ji JX, Szapacs ME, Yin H, et al. (2007) Phospholipid-protein adducts of lipid peroxidation: synthesis and study of new biotinylated phosphatidylcholines. Chem Res Toxicol 20: 227-234.

8. Prasad A, Bekker P, Tsimikas S (2012) Advanced Glycation Endproducts and Diabetic Cardiovascular Disease. Cardiol Rev.

9. Jiang T, Zhou X, Taghizadeh K, Dong M, Dedon PC (2007) N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. Proc Natl Acad Sci U S A 104: 60-65.

10. Wisniewski JR, Zougman A, Mann M (2008) Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. Nucleic Acids Res 36: 570-577.

11. LeRoy G, Weston JT, Zee BM, Young NL, Plazas-Mayorca MD, et al. (2009) Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications. Mol Cell Proteomics 8: 2432-2442.

12. Pesavento JJ, Kim YB, Taylor GK, Kelleher NL (2004) Shotgun annotation of histone modifications: a new approach for streamlined characterization of proteins by top down mass spectrometry. J Am Chem Soc 126: 3386-3387.

13. Felsenfeld G, Groudine M (2003) Controlling the double helix. Nature 421: 448-453.

14. Chi P, Allis CD, Wang GG (2010) Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer 10: 457-469.

15. Rando OJ (2012) Combinatorial complexity in chromatin structure and function: revisiting the histone code. Curr Opin Genet Dev.

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

17. Siuti N, Roth MJ, Mizzen CA, Kelleher NL, Pesavento JJ (2006) Gene-specific characterization of human histone H2B by electron capture dissociation. J Proteome Res 5: 233-239.

18. Garcia BA, Barber CM, Hake SB, Ptak C, Turner FB, et al. (2005) Modifications of human histone H3 variants during mitosis. Biochemistry 44: 13202-13213.

19. Lu K, Moeller B, Doyle-Eisele M, McDonald J, Swenberg JA (2011) Molecular dosimetry of N2-hydroxymethyl-dG DNA adducts in rats exposed to formaldehyde. Chem Res Toxicol 24: 159-161.

20. Zhang L, Freeman LE, Nakamura J, Hecht SS, Vandenberg JJ, et al. (2010) Formaldehyde and leukemia: epidemiology, potential mechanisms, and implications for risk assessment. Environ Mol Mutagen 51: 181-191.

21. Le Curieux F, Pluskota D, Munter T, Sjoholm R, Kronberg L (2000) Identification of fluorescent 2'-deoxyadenosine adducts formed in reactions of conjugates of malonaldehyde and acetaldehyde, and of malonaldehyde and formaldehyde. Chem Res Toxicol 13: 1228-1234.

22. Begley TJ, Samson LD (2003) AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. Trends Biochem Sci 28: 2-5.

23. Shi Y, Whetstine JR (2007) Dynamic regulation of histone lysine methylation by demethylases. Mol Cell 25: 1-14.

24. Boyne MT, 2nd, Pesavento JJ, Mizzen CA, Kelleher NL (2006) Precise characterization of human histones in the H2A gene family by top down mass spectrometry. J Proteome Res 5: 248-253.

25. Pasqualini JR, Mercat P, Giambiagi N (1989) Histone acetylation decreased by estradiol in the MCF-7 human mammary cancer cell line. Breast Cancer Res Treat 14: 101-105.

26. Byvoet P, Barber M, Amidei K, Lowell N, Trudeau W (1986) Effect of exogenous histone H5 on integration of histone H1 in rat liver chromatin. Correlations with aberrant epsilon-N-methylation of histone H1. Biochim Biophys Acta 867: 163-175.

27. Wu M, Allis CD, Richman R, Cook RG, Gorovsky MA (1986) An intervening sequence in an unusual histone H1 gene of Tetrahymena thermophila. Proc Natl Acad Sci U S A 83: 8674-8678.

28. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403: 41-45.

29. de Ruijter AJ, van Gennip AH, Caron HN, Kemp S, van Kuilenburg AB (2003) Histone deacetylases (HDACs): characterization of the classical HDAC family. Biochem J 370: 737-749.

30. Hildmann C, Riester D, Schwienhorst A (2007) Histone deacetylases--an important class of cellular regulators with a variety of functions. Appl Microbiol Biotechnol 75: 487-497.

31. Hake SB, Xiao A, Allis CD (2004) Linking the epigenetic 'language' of covalent histone modifications to cancer. Br J Cancer 90: 761-769.

32. Chen L (2011) Medicinal chemistry of sirtuin inhibitors. Curr Med Chem 18: 1936-1946.

33. Dokmanovic M, Clarke C, Marks PA (2007) Histone deacetylase inhibitors: overview and perspectives. Mol Cancer Res 5: 981-989.

34. Tyihak E, Trezl L, Kolonits P (1985) The isolation of Nepsilon-formyl-L-lysine from the reaction between formaldehyde and L-lysine and its identification by OPLC and NMR spectroscopy. J Pharm Biomed Anal 3: 343-349.

35. Oses-Prieto JA, Zhang X, Burlingame AL (2007) Formation of epsilon-formyllysine on silver-stained proteins: implications for assignment of isobaric dimethylation sites by tandem mass spectrometry. Mol Cell Proteomics 6: 181-192.

36. Luka Z, Moss F, Loukachevitch LV, Bornhop DJ, Wagner C (2011) Histone demethylase LSD1 is a folate-binding protein. Biochemistry 50: 4750-4756.

Supplementary Figures

Figure 3-S1. Reversed-phase HPLC fractionation of histone proteins. (A) HPLC elution profile for histones extracted from TK6 cells, as described in *Materials and Methods* section of Chapter 2. (B) SDS/PAGE analysis of the HPLC-fractionated proteins shown in panel A. Lanes 1 and 2 are molecular weight markers, while lanes 3 and 4 refer to total histones from TK6 cells and calf thymus, respectively.

Figure 3-S2. Lysine demethylation is not a source of N^6-formyllysine in histones. By culturing TK6 cells in customized RPMI medium containing L-Methionine-([^{13}C,^{2}H$_3$]-methyl) for 20 days, it was shown that in contrast to predominant heavy isotope labeling of mono-methyllysines (>90%), even as early as day 2, the level of N^6-[^{13}C, ^{2}H]-formyllysine did not show an increase beyond the natural isotope abundance level (~0.7% for [M+2] ion of N^6-formyllysine), for any day.

Chapter 4

Formaldehyde-induced formation of N^6-formyllysine protein adducts in rats

Abstract

The emergence of lysine N^6-formylation as an abundant endogenous modification of proteins has raised questions about its mechanism of formation. In addition to DNA oxidation pathway previously reported (Jiang *et al,. PNAS* **104**: 60-5, 2007), we demonstrated *in vitro* and in cells that formaldehyde exposure leads to formation of N^6-formyllysine residues in histone proteins (Chapter 3). Here, we extend our studies to quantify N^6-formyllysine protein adducts in rats exposed to isotope-labeled formaldehyde by inhalation in order to calculate the contribution of exogenous versus endogenous sources of formaldehyde to the total load of N^6-formyllysine in cells. Moreover, our studies complement previous studies of formaldehyde-induced N^2-hydroxymethyl-dG DNA adducts in exposed rats (Lu *et al., Chem. Res. Toxicol.* **24**: 159-61, 2011), where we aim to correlate the protein and DNA adducts given the carcinogenic potency of formaldehyde. Using our ultrasensitive and specific liquid chromatography-coupled tandem mass spectrometry method, we quantified endogenous and exogenous N^6-fromyllysine protein adducts in total and fractionated proteins (cytoplasmic, membrane, and nuclear) extracted from nasal epithelium, lung, and liver tissues of rats exposed to single-dose, nose-only inhalation of 0.7, 2, 6, and 10 ppm $[^{13}C^2H_2]$-formaldehyde for 6 h. Similar to the DNA study, exogenous N^6-formyllysine adducts were only detectable in nasal epithelium and not in distant tissues, with a dose-dependent formation in total as well as in fractionated proteins. At 10 ppm exposure, nearly one-third of the total load of N^6-formyllysine was derived from exogenous formaldehyde (0.9 ± 0.1 exogenous versus 2 ± 0.1 endogenous N^6-formyllysine adducts per 10^4 lysines in total proteins). Moreover, there was a general decrease in exogenous adducts observed in nuclear proteins compared to proteins obtained from other compartments, consistent with expected decrease of inhaled formaldehyde concentration as it reacts with other biomolecules on its way to nucleus. In histone proteins, the exogenous to endogenous adduct ratios for N^6-formyllysine are observed to occur about one third of ratios observed for DNA adduct, although they follow a similar pattern of increase in response to formaldehyde dosage (e.g., an increase in formaldehyde dosage from 6 ppm to 10 ppm, resulted in ~3-fold increase in exogenous versus endogenous ratios for both protein and DNA adducts). Our formaldehyde studies shed light on yet another potential pathway associated with formaldehyde toxicity and carcinogenicity, that is through epigenetic disruption of histone proteins in cells by permanently formylating conserved sites of important physiological modifications.

Introduction

Aberrant protein modifications due to oxidants and electrophiles generated during physiological or pathophysiological processes can lead to altered cellular processes and progression of disease[1-5]. N^6-Formylation of lysines has recently emerged as an abundant adventitious protein modification [6-9], raising questions about its mechanism of formation and biological consequences in cells. In histone proteins, the chemical similarities of N^6-formyllysine to N^6-acetyllysine suggest a disruptive role for the former in signaling by histone acetylation, and the occurrence at many conserved lysine acetylation and methylation sites along histone proteins [7] suggests possible interference with the regulatory roles of post-translational modifications of these proteins. Studies to date point to formaldehyde exposure as a major source of this adduct (Chapter 3) [10,11] in addition to oxidative and nitrosative stresses of inflammation [6,9].

Formaldehyde exposure from environmental and occupational sources[12-14] as well as endogenous cellular processes [13,15,16] make it an abundant electrophile in the human body, as indicated by its concentration range of 13 to 97 µM in plasma [13,17]. Formaldehyde, classified as a known human carcinogen according to IARC [18,19], is reported to cause squamous cell carcinoma in rats [20,21] with sufficient epidemiological evidence for causing nasopharyngeal cancer but limited evidence for leukemia in humans [13,18,19,22].

As a reactive aldehyde, formaldehyde readily forms adducts with nucleophilic sites in DNA and proteins such as N^2-hydroxymethyl-dG DNA adducts [12] and Schiff bases on the side chain amine of lysines in proteins [23], in addition to extensive DNA-protein and DNA-DNA crosslinks [24]. Our recent studies revealed N^6-formyllysine

formation in histone proteins of TK6 cells from the reaction of formaldehyde with the ε-amino group of lysine (Chapter 3). Our data showed a dose-dependent formation of N^6-formyllysine residues as a function of formaldehyde concentration, with use of $[^{13}C,^{2}H_2]$-formaldehyde revealing unchanged levels of N^6-formyllysinespecies due to endogenous sources.

Here, we extend our cell studies to an *in vivo* assessment of formaldehyde-induced lysine N^6-formylation in rats exposed by inhalation, with the use of $[^{13}C,^{2}H_2]$-formaldehyde to differentiate endogenous from exogenous adducts. Using our ultrasensitive and specific liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) method, we quantified the extent to which exposure to $[^{13}C,^{2}H_2]$-formaldehyde by inhalation induces N^6-$[^{13}C,^{2}H]$-formyllysine. Endogenous N^6-$[^{12}C,^{1}H]$-formyllysine and exposure-induced N^6-$[^{13}C,^{2}H]$-formyllysine adducts were quantified in total and fractionated (cytoplasmic, membrane, and nuclear) proteins extracted from nasal epithelium, lung, and liver tissues of rats exposed to single-dose, nose-only inhalation of 0.7, 2, 6, and 10 ppm $[^{13}C^{2}H_2]$-formaldehyde for 6 h. These studies complement previous studies of $[^{13}C,^{2}H_2]$-formaldehyde-induced N^2-hydroxymethyl-dG DNA adducts in rats from the same formaldehyde exposure studies [12], which allows us to correlate the protein and DNA adducts arising with the same formaldehyde dose.

Application of our sensitive analytical methods revealed that exposure to $[^{13}C,^{2}H_2]$-formaldehyde vapor leads to the formation of exogenous N^6-$[^{13}C,^{2}H]$-formyllysine in proteins of nasal epithelium, and that the exogenous adducts are formed in proteins extracted from all cellular compartments in a dose-dependent manner, with endogenous adducts dominating at all exposure conditions. Similar to the DNA study,

the exogenous N^6-formyllysine adducts were only detected in nasal epithelium and not in distant tissues, with exogenous to endogenous adduct ratios in total proteins closely matching increases seen in DNA adducts.

Materials and Methods

Materials. Rat tissues were provided by Prof. James Swenberg's Lab at University of North Carolina at Chapel Hill [12]. Subcellular Protein Fractionation Kit was purchased from Thermo Scientific (Waltham, MA). Lysine internal standard, 4,4,5,5,-[^{2}H]-Lysine, was purchased from Cambridge Isotope Laboratories (Andover, MA). N^6-Formyllysine internal standard, 4,4,5,5-[^{2}H]-N^6-formyllysine, was synthesized from 4,4,5,5-[^{2}H]-lysine according to Jiang *et al.*[6]. *Streptomyces griseus* protease was purchased from Sigma-Aldrich (St. Louis, MO).

Rat exposure to formaldehyde vapor. Rat exposure to [^{13}C^{2}H$_2$]-formaldehyde was performed at the Lovelace Respiratory Research Institute, Albuquerque, NM, by our collaborator, Prof. James Swenberg, according to appropriate and approved protocols for the use of vertebrate animals in experiments [12]. Fischer rats (6 w old, male) were exposed to formaldehyde vapor by single-dose, nose-only inhalation of [^{13}C^{2}H$_2$]-formaldehyde for 6 h to produce final target exposure concentrations of 0.7, 2, 6, and 10 ppm. Rats were euthanized using an intraperitoneal barbiturate injection. For nasal mucosa collection, the skull was split with a slight bias to one side (to preserve septal mucosa) and maxilloturbinates collected, wrapped in aluminum foil, immediately snap

frozen in liquid nitrogen and held at -80 °C pending analyses. Other tissue samples were processed in a similar fashion.

Subcellular protein fractionation from tissue. For extracting proteins from tissues, ~10 mg of nasal epithelium, lung, or liver tissue was cut into small pieces and washed with PBS. Tissues were homogenized using a Kontes all-glass Dounce homogenizer, with 10 strokes of a type B pestle, and proteins were separated into cytoplasmic, membrane, soluble nuclear, and chromatin bound fractions using the Subcellular Protein Fractionation Kit from Thermo Scientific. Total protein was obtained by centrifuging tissue lysate, prior to fractionation, for 14000 x g for 15 min at 4 °C and collecting the supernatant. Protein precipitation from each fraction was accomplished by adding 20% v/v trichloroacetic acid and incubating overnight at 4 °C. Samples were centrifuged at 14000 x g for 10 min at 4 °C, washed once with ice-cold acetone containing 0.1% HCl, and once with ice-cold acetone before being air-dried and stored at -20 ° C until use. N^6-Formyllysine adducts were found to be stable under these conditions.

Enzymatic hydrolysis of proteins. Extracted proteins from tissues were dissolved in 50 μL of 100 mM ammonium bicarbonate buffer (pH 8.5), 4,4,5,5-[^{2}H]-lysine (2 nmol), and 4,4,5,5-[^{2}H]-N^6-formyllysine (1 pmol) were added as internal standards, and the proteins were hydrolyzed by addition of 2 μl of freshly prepared *S. griseus* protease solution (1μg/μl) with incubation at 37°C for 16 h. Samples were dried under vacuum and resuspended in 50 μL of water prior to mass spectrometry analysis.

Quantification of amino acids. N^6-Formyllysine amino acids in the resuspended protein

hydrolysates were quantified as a percentage of lysine content by liquid chromatography-

coupled mass spectrometry (LC-MS/MS). Analytes in total, cytoplasmic, and membrane

protein hydrolysates were analyzed without a pre-purification step. Nuclear protein

hydrolysates were first HPLC pre-purified to enhance the signal for the M+2 ion of N^6-

formyllysine. HPLC pre-purification was done using an Agilent 1100 series system, with

an aqueous normal-phase Cogent diamond hydride column (2.1 x 150mm, 4μm) from

MicroSolv Technology Corporation (Eatontown, NJ), with mobile phase flow rate of 400

μL/min and column temperature of 20 °C. The solvent system was 0.1% acetic acid in

water (A) and 0.1% acetic acid in acetonitrile (B), with the elution starting at 100% B, the

gradient linearly decreased to 25% B over 30 min and held at 25% B for 3 additional min

before the column was re-equilibrated with 100% B for 7 min. For N^6-formyllysine

quantification, an extended chromatography was used, in order to separate the M+1 ion

of citrulline from M+2 ion of N^6-formyllysine (Figure 2-8). The same solvent system of

0.1% acetic acid in water (A) and 0.1% acetic acid in acetonitrile (B) was used, with the

elution starting at 100% B, the gradient linearly decreased to 75% B over 75 min,

reaching 25% B over the next 3 min and 15% B by 83 min. The column was then re-

equilibrated with 100% B for 7 min. N^6-Formyllysine and lysine were then analyzed

using an Agilent 6410 triple quadrupole mass spectrometer (MS/MS) equipped with an

electrospray ionization (ESI) source operated in positive ion mode. The operating

parameters were as follows: ESI capillary voltage, 4000 V; gas temperature, 350 °C;

drying gas flow, 12 L/min; and nebulizer pressure, 30 psi, with selected reaction

monitoring (SRM) transitions of *m/z* 175→112 and for N^6-formyllysine (with internal

standard transition of m/z 179→116) and m/z 147→130 for lysine (with internal standard transition of m/z 151→134). The fragmentor voltage and collision energies were 105 V and 10 V for N^6-formyllysine and 100 V and 8 V for lysine, respectively.

Results

Quantification of background endogenous N^6-formyllysine in tissues. In addition to monitoring the formaldehyde-induced (exogenous) N^6-formyllysine modifications, it is important to quantify the background (endogenous) levels of this protein adduct to parse the contribution of exogenous formaldehyde sources to the total amount of lysine formylation in cells. To this end, total proteins as well as fractionated proteins from cytosolic, membrane, and nuclear compartments were extracted from nasal epithelium, lung, and liver of exposed rats, and endogenous adducts were quantified using our sensitive and specific LC-MS/MS method with the limits of detection of 1 fmol for N^6-formyllysine and 10 fmol for lysine. Table 4-1 summarizes the background levels of N^6-formyllysine in all proteins analyzed. There are similar levels of endogenous adducts among different types of tissue analyzed with a range of 2-4 N^6-formyllysine residues per 10^4 lysines. Each tissue shows similar levels in control rats and rats exposed to the highest dose of 10 ppm [$^{13}C,^2H_2$]-formaldehyde, which indicates that exposure to formaldehyde did not affect the endogenous adduct levels.
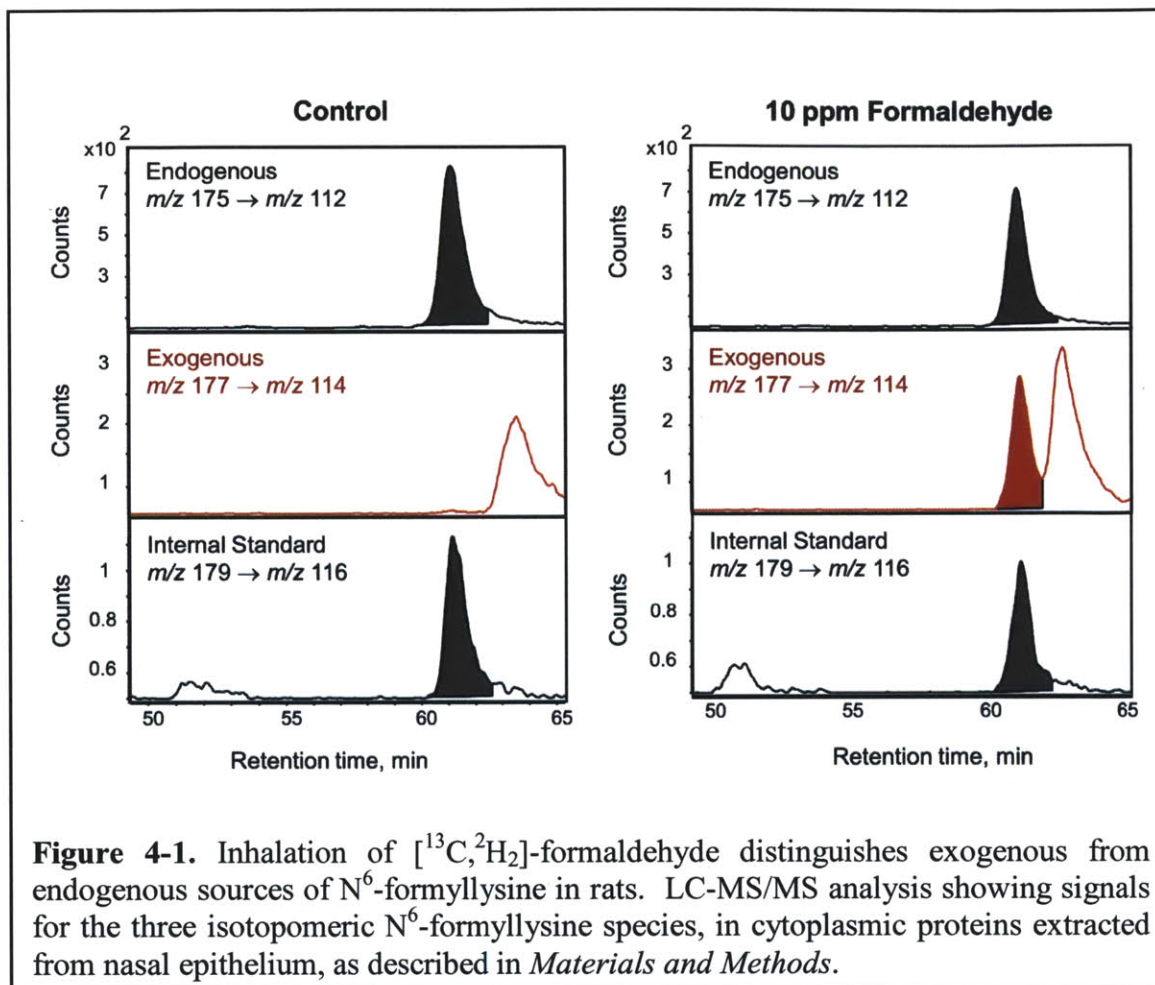
Table 4-1. Quantification of endogenous N^6-formyllysine protein adducts in total and fractionated proteins.[1]

Protein source	Nasal Epithelium		Lung		Liver	
Dose	0 ppm	9.7 ppm	0 ppm	9.7 ppm	0 ppm	9.7 ppm
Total protein	2 ± 0.1^2	2 ± 0.1	4 ± 0.2	3 ± 0.4	3 ± 0.1	3 ± 0.5
Cytoplasmic	2 ± 0.4	2 ± 0.4	4 ± 0.6	4 ± 0.6	4 ± 0.5	4 ± 0.1
Membrane	3 ± 0.8	2 ± 0.4	3 ± 0.5	3 ± 0.4	3 ± 0.3	3 ± 0.2
Soluble nuclear	2 ± 0.3	2 ± 1.0	4 ± 1.0	4 ± 0.3	4 ± 0.8	4 ± 0.7
Chromatin bound	2 ± 0.1	2 ± 0.4	3 ± 1.0	3 ± 0.2	4 ± 0.5	3 ± 0.3

[1]Column titles denote endogenous N^6-formyllysine modifications in each tissue analyzed (for control versus 9.7 ppm [^{13}C,^{2}H$_2$]-formaldehyde exposed rats).

[2]Data are expressed as adducts per 10^4 total lysines and represent mean $\pm$ SD for 3 rats.

Formaldehyde-induced formation of exogenous N^6-formyllysine adducts. The formation of N^6-formyllysine from inhaled [^{13}C,^{2}H$_2$]-formaldehyde was monitored using the m/z 177→114 transition for N^6-[^{13}C,^{2}H]-formyllysine that is 2 mass units heavier than the endogenous adduct, as noted in Chapter 3, with the 4,4,5,5-[^{2}H]-N^6-formyllysine internal standard monitored at m/z 179 →116 (Figure 4-1). Exogenously-derived N^6-formyllysine adducts were only detected in proteins extracted from nasal epithelium and not in distant tissues of lung and liver. In all samples analyzed from lung and liver tissues, the exogenous adducts did not show an increase beyond the natural isotope abundance level of ~0.7% for [M+2] ion of N^6-formyllysine. In addition to total proteins, an analysis of proteins in cytosolic, membrane, and nuclear compartments revealed the dose-dependent formation of exogenous N^6-formyllysine species in all locations (Table 4-2 and Figure 4-2). Data show that, at all exposures, the endogenous adducts dominate.

Figure 4-1. Inhalation of $[^{13}C,^2H_2]$-formaldehyde distinguishes exogenous from endogenous sources of N^6-formyllysine in rats. LC-MS/MS analysis showing signals for the three isotopomeric N^6-formyllysine species, in cytoplasmic proteins extracted from nasal epithelium, as described in *Materials and Methods*.

There is a clear dose-response relationship for lysine N^6-formylation across the range of inhaled formaldehyde doses (0.7, 2, 6, 10 ppm) (Figure 4-2), with exogenous adducts in total proteins rising from < 3% of endogenous adducts to > 40% for a 14-fold increase in formaldehyde dosage (0.7 to 10 ppm). At all exposure levels, the ratio of exogenous to endogenous N^6-formyllysine adducts in all proteins, except nuclear proteins, closely correlate with ratios of exogenous and endogenous N^2-hydroxymethyl-dG DNA adducts determined by Swenberg and coworkers [12]. For instance, N^6-formyllysine ratios of 0.15 and 0.4 in total protein closely match the DNA adduct ratios of 0.2 and 0.6, for 6 and 10 ppm exposures, respectively.
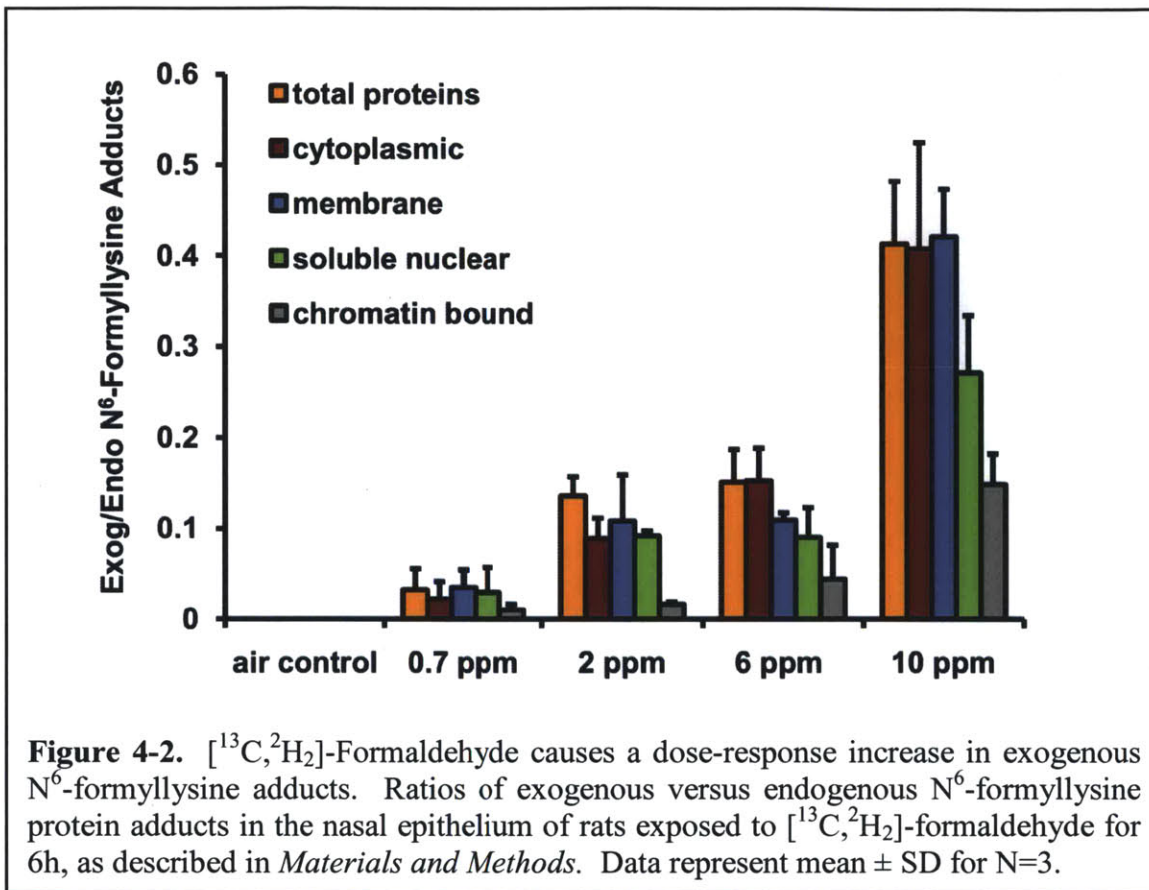
Table 4-2. Quantification of endogenous and exogenous N^6-formyllysine protein adducts in nasal epithelium from rats exposed to $[^{13}C,^2H_2]$-formaldehyde.

Dose	0.7 ppm		2 ppm		5.8 ppm		9.7 ppm	
Adduct type	Endo[1]	Exog	Endo	Exog	Endo	Exog	Endo	Exog
Total Protein	2 ± 0.1[2]	0.06 ± 0.04	2 ± 0.2	0.2 ± 0.02	2 ± 0.4	0.3 ± 0.04	2 ± 0.1	0.9 ± 0.1
Cytoplasmic	2 ± 0.3	0.05 ± 0.04	3 ± 0.2	0.2 ± 0.07	2 ± 0.8	0.4 ± 0.2	2 ± 0.4	0.8 ± 0.1
Membrane	2 ± 0.3	0.06 ± 0.02	2 ± 0.8	0.2 ± 0.03	3 ± 0.2	0.3 ± 0.02	2 ± 0.4	0.7 ± 0.2
Soluble nuclear	2 ± 0.3	0.05 ± 0.05	2 ± 1.0	0.2 ± 0.1	4 ± 0.4	0.4 ± 0.2	2 ± 1.0	0.5 ± 0.2
Chromatin bound	2 ± 0.4	0.02 ± 0.02	2 ± 0.8	0.03 ± 0.01	2 ± 0.1	0.05 ± 0.05	2 ± 0.4	0.2 ± 0.01

[1]Column titles denote endogenous or exogenous N^6-formyllysine modifications for each $[^{13}C,^2H_2]$-formaldehyde exposure dose.

[2]Data are expressed as modifications per 10^4 total lysines and represent mean ± SD for 3 rats.

As shown in Table 4-2 and Figure 4-2, there is a general decrease in adduct levels in nuclear proteins compared to proteins from other cell compartments. For example, a 10 ppm formaldehyde exposure produced 0.2 exogenous N^6-formyllysine adducts per 10^4 lysines in chromatin proteins compared to 0.7 and 0.8 residues in cytoplasmic and membrane fractions, respectively ($p < 0.05$).

Figure 4-2. [^{13}C,^{2}H$_2$]-Formaldehyde causes a dose-response increase in exogenous N^6-formyllysine adducts. Ratios of exogenous versus endogenous N^6-formyllysine protein adducts in the nasal epithelium of rats exposed to [^{13}C,^{2}H$_2$]-formaldehyde for 6h, as described in *Materials and Methods*. Data represent mean ± SD for N=3.
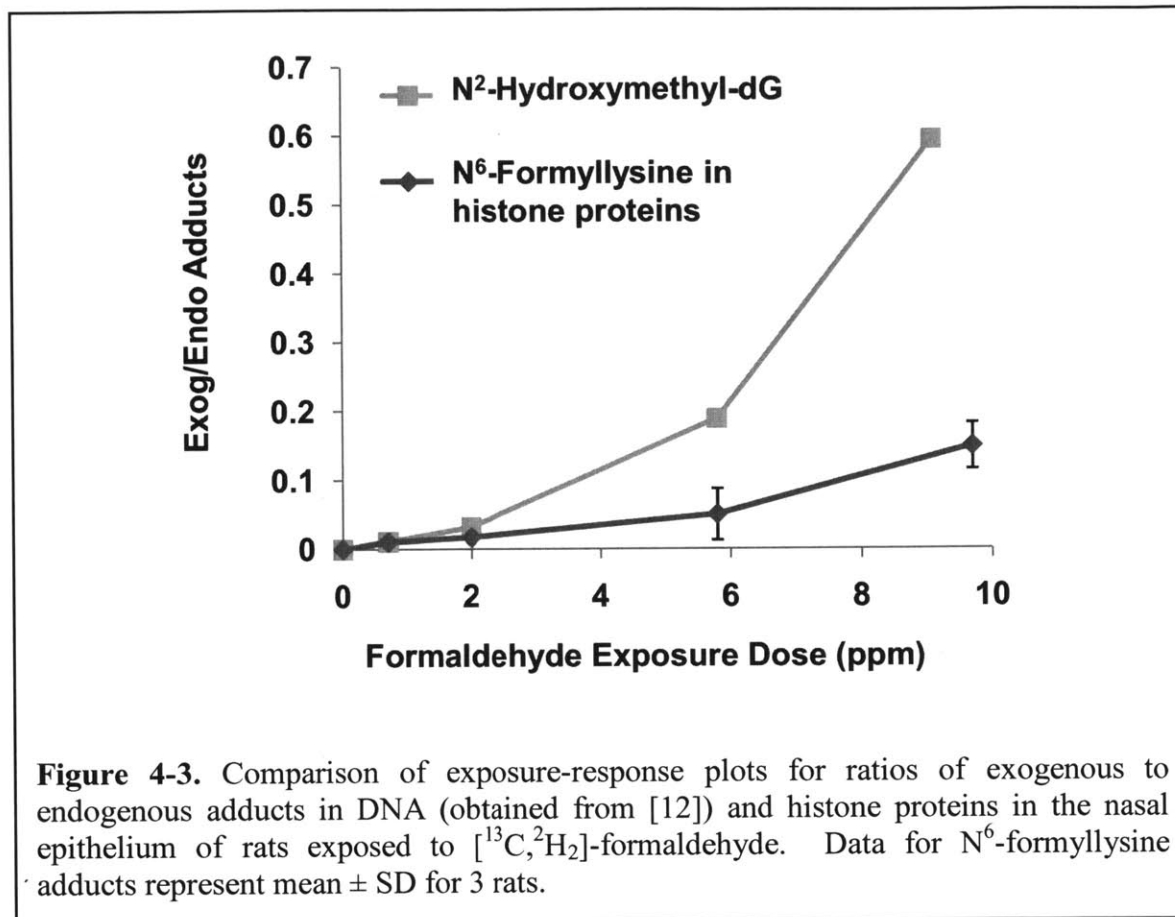
Discussion

The results of these *in vivo* studies point to several important features of N^6-formyllysine formation and formaldehyde toxicity. While originally described as an adventitious protein modification of histone and other nuclear proteins [6-8], N^6-formylation of lysine has been shown to occur globally in proteins from different cellular compartments as well as plasma proteins (Chapter 3). These observations, together with the *in vitro* formaldehyde studies of Chapter 3 and the relatively high levels of formaldehyde exposure from environmental [12-14] and endogenous [13,15,16] sources, point to formaldehyde as a major source of N^6-formyllysine adducts in cells, as compared to DNA

oxidation pathway previously reported. Here, our *in vivo* assessment of N^6-formyllysine formation in rats exposed to formaldehyde by inhalation is consistent with our previous conclusion of formaldehyde as a source of lysine N^6-formylation. Interestingly, the background levels of N^6-formyllysine species due to endogenous sources were unaffected even at the highest formaldehyde exposure dosage, which suggests that the formaldehyde doses used for exposure do not alter the cellular production of formaldehyde.

The observation that background N^6-formyllysine levels are similar in proteins from all cellular compartments suggests that the sources of this protein modification are equally balanced in the various compartments, consistent with cellular abundance of formaldehyde due to environmental and endogenous sources [12-16]. The dose-dependent formation of exogenous N^6-$[^{13}C,^2H]$-formyllysine adducts in all compartments, with decreased levels in nucleus, is consistent with exogenous stable isotope-labeled formaldehyde being the source, as inhaled formaldehyde concentration is expected to decrease on the way to nucleus by reacting with other biomolecules.

To further correlate protein and DNA adducts given the carcinogenic potency of formaldehyde, exogenous/endogenous ratios of N^6-formyllysine residues in histone proteins are plotted against the published values of N^2-hydroxymethyl-dG DNA adducts [12], as shown in Figure 4-3. Data reveal that DNA adducts dominate histone adducts for the same concentration of formaldehyde. For instance, for the case of 10 ppm exposure, the exogenous to endogenous N^2-hydroxymethyl-dGadduct ratio is more than 3-fold of that for N^6-formyllysine (i.e., ~0.6 versus less than 0.2). However, they both follow the same pattern of increase as a response to formaldehyde exposure. For example, going from 2 ppm to 10 ppm, and from 6 ppm to 10 ppm formaldehyde exposure doses result in

Figure 4-3. Comparison of exposure-response plots for ratios of exogenous to endogenous adducts in DNA (obtained from [12]) and histone proteins in the nasal epithelium of rats exposed to $[^{13}C,^{2}H_2]$-formaldehyde. Data for N^6-formyllysine adducts represent mean ± SD for 3 rats.

same 20-fold and 3-fold ratio increases, respectively, in both DNA and protein adducts (Figure 4-3).

The analysis of N^6-formyllysine adducts sheds light on the mechanisms of formaldehyde toxicity. The protein adduct data complement previous studies of formaldehyde-induced DNA adducts in rats[12], with our results showing strong correlations between protein and DNA adduct formation during formaldehyde exposure. Our results show that, similar to N^2-hydroxymethyl-dG DNA adducts, the exogenously-derived N^6-formyllysine residues were only detected in nasal epithelium and not in distant tissues of lung and liver, with a dose-dependent formation of exogenous adducts

in total proteins as well as proteins from specific cellular compartments (Figure 4-2 and Table 4-2). With the exception of nuclear proteins, the levels of N^6-formyllysine formation in proteins closely correlates with N^2-hydroxymethyl-dG adduct formation in DNA. However, as shown in Figure 4-3,the proportion of DNA adducts derived from exogenous formaldehyde was significantly higher than that for the histone adducts for the same concentration of formaldehyde, even though theoretically the same number of exogenous formaldehyde molecules should reach DNA and histones. This difference in ratios could be due to several factors such as DNA guanine content compared to histone lysine content, faster kinetics of N^2-hydroxymethyl-dGformation versus N^6-formyllysine formation, as well as possible greater solvent exposure of guanine N^2-positions compared to lysine amino side chains in histone proteins. Nonetheless, both N^2-hydroxymethyl-dG and N^6-formyllysine adducts follow the same pattern of increase as a response to formaldehyde dosage (Figure 4-3).

There have been extensive studies to date on possible mechanisms for formaldehyde toxicity and carcinogenicity. For instance, some studies found a non-linear dose-dependent formation of DNA damage products in rats and non-human primates exposed to formaldehyde inhalation [12,17,25,26] with other studies showing long term formaldehyde exposure of higher than 6 ppm dosage resulting in a substantial increases in squamous cell carcinomas in rats [20,21,27]. On the path to understanding formaldehyde's biological consequences, our results shed light on yet another pathway, that is through formation of N^6-formyllysine adducts in histones. N^6-Formyllysine adducts have been mapped on many conserved lysine acetylation and methylation sites along histones [7,8]. This observation along with chemical similarity of lysine N^6-

formylation to lysine N^6-acetylation, as well as our results showing N^6-formyllysine is refractory to removal by histone deacetylases (Chapter 3) suggests that N^6-formyllysine could interfere with the epigenetic roles associated with physiological histone modifications [28,29]. In other words, our studies suggest that prolonged exposure to formaldehyde from environmental and occupational sources can drive substantial and permanent formylation of conserved sites of important physiological modifications in histone proteins, thus interfering with their regulatory roles and further contributing to toxicity and carcinogenicity associated with formaldehyde exposure.

References

1. Marnett LJ, Riggins JN, West JD (2003) Endogenous generation of reactive oxidants and electrophiles and their reactions with DNA and protein. J Clin Invest 111: 583-593.

2. Levine RL (2002) Carbonyl modified proteins in cellular regulation, aging, and disease. Free Radic Biol Med 32: 790-796.

3. Stadtman ER, Berlett BS (1997) Reactive oxygen-mediated protein oxidation in aging and disease. Chem Res Toxicol 10: 485-494.

4. Jacobs AT, Marnett LJ (2010) Systems analysis of protein modification and cellular responses induced by electrophile stress. Acc Chem Res 43: 673-683.

5. Dedon PC (2008) The chemical toxicology of 2-deoxyribose oxidation in DNA. Chem Res Toxicol 21: 206-219.

6. Jiang T, Zhou X, Taghizadeh K, Dong M, Dedon PC (2007) N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. Proc Natl Acad Sci U S A 104: 60-65.

7. Wisniewski JR, Zougman A, Mann M (2008) Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. Nucleic Acids Res 36: 570-577.

8. LeRoy G, Weston JT, Zee BM, Young NL, Plazas-Mayorca MD, et al. (2009) Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications. Mol Cell Proteomics 8: 2432-2442.

9. Vana L, Kanaan NM, Hakala K, Weintraub ST, Binder LI (2011) Peroxynitrite-induced nitrative and oxidative modifications alter tau filament formation. Biochemistry 50: 1203-1212.

10. Tyihak E, Trezl L, Kolonits P (1985) The isolation of Nepsilon-formyl-L-lysine from the reaction between formaldehyde and L-lysine and its identification by OPLC and NMR spectroscopy. J Pharm Biomed Anal 3: 343-349.

11. Oses-Prieto JA, Zhang X, Burlingame AL (2007) Formation of epsilon-formyllysine on silver-stained proteins: implications for assignment of isobaric dimethylation sites by tandem mass spectrometry. Mol Cell Proteomics 6: 181-192.

12. Lu K, Moeller B, Doyle-Eisele M, McDonald J, Swenberg JA (2011) Molecular dosimetry of N2-hydroxymethyl-dG DNA adducts in rats exposed to formaldehyde. Chem Res Toxicol 24: 159-161.

13. Zhang L, Freeman LE, Nakamura J, Hecht SS, Vandenberg JJ, et al. (2010) Formaldehyde and leukemia: epidemiology, potential mechanisms, and implications for risk assessment. Environ Mol Mutagen 51: 181-191.

14. Le Curieux F, Pluskota D, Munter T, Sjoholm R, Kronberg L (2000) Identification of fluorescent 2'-deoxyadenosine adducts formed in reactions of conjugates of malonaldehyde and acetaldehyde, and of malonaldehyde and formaldehyde. Chem Res Toxicol 13: 1228-1234.

15. Begley TJ, Samson LD (2003) AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. Trends Biochem Sci 28: 2-5.

16. Shi Y, Whetstine JR (2007) Dynamic regulation of histone lysine methylation by demethylases. Mol Cell 25: 1-14.

17. Moeller BC, Lu K, Doyle-Eisele M, McDonald J, Gigliotti A, et al. (2011) Determination of N2-hydroxymethyl-dG adducts in the nasal epithelium and bone marrow of nonhuman primates following 13CD2-formaldehyde inhalation exposure. Chem Res Toxicol 24: 162-164.

18. Cogliano VJ, Grosse Y, Baan RA, Straif K, Secretan MB, et al. (2005) Meeting report: summary of IARC monographs on formaldehyde, 2-butoxyethanol, and 1-tert-butoxy-2-propanol. Environ Health Perspect 113: 1205-1208.

19. Humans IWGotEoCRt (2006) Formaldehyde, 2-butoxyethanol and 1-tert-butoxypropan-2-ol. IARC Monogr Eval Carcinog Risks Hum 88: 1-478.

20. Monticello TM, Swenberg JA, Gross EA, Leininger JR, Kimbell JS, et al. (1996) Correlation of regional and nonlinear formaldehyde-induced nasal cancer with proliferating populations of cells. Cancer Res 56: 1012-1022.

21. Swenberg JA, Kerns WD, Mitchell RI, Gralla EJ, Pavkov KL (1980) Induction of squamous cell carcinomas of the rat nasal cavity by inhalation exposure to formaldehyde vapor. Cancer Res 40: 3398-3402.

22. Zhang L, Tang X, Rothman N, Vermeulen R, Ji Z, et al. (2010) Occupational exposure to formaldehyde, hematotoxicity, and leukemia-specific chromosome changes in cultured myeloid progenitor cells. Cancer Epidemiol Biomarkers Prev 19: 80-88.

23. Lu K, Boysen G, Gao L, Collins LB, Swenberg JA (2008) Formaldehyde-induced histone modifications in vitro. Chem Res Toxicol 21: 1586-1593.

24. Lu K, Collins LB, Ru H, Bermudez E, Swenberg JA (2010) Distribution of DNA adducts caused by inhaled formaldehyde is consistent with induction of nasal carcinoma but not leukemia. Toxicol Sci 116: 441-451.

25. Casanova M, Morgan KT, Steinhagen WH, Everitt JI, Popp JA, et al. (1991) Covalent binding of inhaled formaldehyde to DNA in the respiratory tract of rhesus monkeys: pharmacokinetics, rat-to-monkey interspecies scaling, and extrapolation to man. Fundam Appl Toxicol 17: 409-428.

26. (2006) Integrated Risk Information System (IRIS): Formaldehyde. US Environmental Protection Agency.

27. Kerns WD, Pavkov KL, Donofrio DJ, Gralla EJ, Swenberg JA (1983) Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure. Cancer Res 43: 4382-4392.

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

29. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403: 41-45.

Chapter 5

Concluding remarks

The recent discovery of N^6-formyllysine as a relatively abundant endogenous secondary modification of histones and other nuclear proteins [1-3] has raised questions about its mechanism of formation and its potential for interfering with the regulatory roles of these proteins, particularly in terms of disrupting histone acetylation signaling [4,5] due to chemical similarity of N^6-formyllysine to N^6-acetyllysine (Figure 1-1). With respect to formation, we considered formaldehyde as a source because of its high reactivity toward primary amines, such as the side chain of lysine, and the potential for endogenous oxidation converting the formaldehyde-derived carbinolamine to a formamide bond (Figure 3-1). Using our ultrasensitive and specific liquid chromatography-coupled tandem mass spectrometry method to quantify N^6-formyllysine lesions in proteins, we presented evidence for formaldehyde being a major source of this pathological protein modification in addition to products of oxidative DNA damage previously reported [1]. More specifically, the work presented in this thesis demonstrated *in vitro* as well as *in vivo* (Chapters 3 and 4) that formaldehyde exposure leads to a dose-dependent lysine N^6-formylation in all cellular proteins, with the use of $[^{13}C,^2H_2]$-formaldehyde to differentiate endogenous from exogenous adducts. Indeed, our study of formaldehyde exposure in TK6 cells showed that the exogenous formaldehyde contributed to half of the total load of formyllysine in histone proteins at 10 mM concentration (Chapter 3), while in rats exposed to 10 ppm formaldehyde by inhalation, nearly one-third of the total load of N^6-formyllysine was derived from exogenous sources in total proteins extracted from nasal epithelium (Chapter 4).

The formaldehyde studies of Chapters 3 and 4 and the high incident of formaldehyde exposure due to environmental [6-8] and endogenous [7,9,10] sources

together with our discovery of global lysine formylation occurring in plasma proteins and proteins from all cellular compartments (Tables 3-2 and 4-1), point to formaldehyde as the major source of N^6-formyllysine adducts as compared to DNA oxidation pathway [1]. The fact that N^6-formyllysine levels are similar in histone and non-nuclear proteins and in all cell compartments also suggests that the sources of this protein modification are equally balanced in the various compartments and proteins, or that there is a single dominant source that distributes uniformly throughout the cell. Though it is not yet possible to distinguish N^6-formyllysine adducts arising from endogenous formaldehyde versus other sources, the fact that human serum concentration of formaldehyde occur in the range of 10 to 100 μM [7,11,12] is consistent with formylation of lysine *in vivo*.

In terms of other sources of this adduct, we considered disrupted histone demethylation that would leave the carbinolamine precursor of N^6-formyllysine for subsequent oxidation and the formaldehyde produced by complete lysine demethylation [10]. However, we presented evidence in Chapter 3 that lysine demethylation process does not serve as a source of lysine N^6-formylation. The absence of detectable N^6-formyllysine arising from N^6-methyllysine species (Figure 3-4C) suggests that disrupted histone demethylation occurs at low frequency, or that the formaldehyde produced at the last step of lysine demethylation does not occur at concentrations high enough to drive formylation of lysine or cause substantial changes in N^6-formyllysine levels detected by our current analytical methods. In addition, there is the possibility that the formaldehyde released during lysine demethylation may be scavenged before it could react with lysines in histone proteins. Indeed, a recent study reports lysine-specific demethylase 1 (LSD1) as a folate binding protein and suggests folate serves as a trap for the formaldehyde

119

generated during lysine demethylation [13].

With regard to N^6-formyllysine persistence in cells, there is still no evidence supporting the enzymatic removal of N^6-formyllysine. Our investigation of histone deacetylases (Chapter 3) revealed that despite chemical similarity of N^6-formyllysine to N^6-acetyllysine, the former is not removed by histone deacetylases, suggesting that they will persist throughout the life of individual histone proteins. Further, the $[^2H_8]$-N^6-formyllysine dilution studies (Chapter 2) indicate that at one extreme N^6-formyllysine persists in cells for the life of histone proteins and at the other extreme, its steady state levels doesn't change as it is formed and removed at the same frequency. Although the studies presented in this thesis point to the stability of lysine N^6-formylation in proteins, we cannot rule out the possibility of an enzyme that would remove this modification from selected conserved lysine sites in histone proteins, resulting in small change in the quantity of formyllysine that is simply not detectable by our current analytical method.

With respect to pathological consequences of N^6-formyllysine in cells, the abundance of lysine N^6-formylation in histone proteins points to its potential for epigenetic alterations in gene expression. The observation that N^6-formyllysine adducts are randomly and relatively evenly distributed among different classes of histone proteins, as opposed to the very biased distributions of other functional modifications (Table 3-1), suggests that they are adventitious and not physiological. The high abundance of lysine N^6-formylation in histone proteins (Table 2-3) as well as its occurrence on many conserved functional locations for lysine acetylation and methylation [2,3] suggest a disruptive role for N^6-formyllysine with regards to signaling processes associated with physiological histone modifications [4,5]. In other words, if N^6-

formyllysine is not removed, it could accumulate to significant levels in histone proteins and interfere with their epigenetic regulatory processes through sabotaging the lysine acetylation signaling or blocking the conserved lysines sites and preventing them from undergoing important physiological modifications needed for proper functions of cells. This potential pathological consequence of N^6-formyllysine in cells is consistent with toxicity and carcinogenicity associated with formaldehyde. Indeed, on the path to understanding formaldehyde's biological consequences, our results shed light on yet another pathway, that is through formation of N^6-formyllysine adducts in histone proteins. This potential epigenetic mechanism of disruption of cell function together with the association of N^6-formyllysine with oxidative and nitrosative stresses of inflammation [1,14], suggest that this adduct may play a role in many pathophysiological processes in humans.

References

1. Jiang T, Zhou X, Taghizadeh K, Dong M, Dedon PC (2007) N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. Proc Natl Acad Sci U S A 104: 60-65.

2. Wisniewski JR, Zougman A, Mann M (2008) Nepsilon-formylation of lysine is a widespread post-translational modification of nuclear proteins occurring at residues involved in regulation of chromatin function. Nucleic Acids Res 36: 570-577.

3. LeRoy G, Weston JT, Zee BM, Young NL, Plazas-Mayorca MD, et al. (2009) Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications. Mol Cell Proteomics 8: 2432-2442.

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

5. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403: 41-45.

6. Lu K, Moeller B, Doyle-Eisele M, McDonald J, Swenberg JA (2011) Molecular dosimetry of N2-hydroxymethyl-dG DNA adducts in rats exposed to formaldehyde. Chem Res Toxicol 24: 159-161.

7. Zhang L, Freeman LE, Nakamura J, Hecht SS, Vandenberg JJ, et al. (2010) Formaldehyde and leukemia: epidemiology, potential mechanisms, and implications for risk assessment. Environ Mol Mutagen 51: 181-191.

8. Le Curieux F, Pluskota D, Munter T, Sjoholm R, Kronberg L (2000) Identification of fluorescent 2'-deoxyadenosine adducts formed in reactions of conjugates of malonaldehyde and acetaldehyde, and of malonaldehyde and formaldehyde. Chem Res Toxicol 13: 1228-1234.

9. Begley TJ, Samson LD (2003) AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. Trends Biochem Sci 28: 2-5.

10. Shi Y, Whetstine JR (2007) Dynamic regulation of histone lysine methylation by demethylases. Mol Cell 25: 1-14.

11. Heck HD, Casanova-Schmitz M, Dodd PB, Schachter EN, Witek TJ, et al. (1985) Formaldehyde (CH2O) concentrations in the blood of humans and Fischer-344 rats exposed to CH2O under controlled conditions. Am Ind Hyg Assoc J 46: 1-3.

12. Heck HD, White EL, Casanova-Schmitz M (1982) Determination of formaldehyde in biological tissues by gas chromatography/mass spectrometry. Biomed Mass Spectrom 9: 347-353.

13. Luka Z, Moss F, Loukachevitch LV, Bornhop DJ, Wagner C (2011) Histone demethylase LSD1 is a folate-binding protein. Biochemistry 50: 4750-4756.

14. Vana L, Kanaan NM, Hakala K, Weintraub ST, Binder LI (2011) Peroxynitrite-induced nitrative and oxidative modifications alter tau filament formation. Biochemistry 50: 1203-1212.