DEVELOPMENT AND APPLICATION OF PROTEOMIC TECHNOLOGIES FOR THE ANALYSIS OF POST-TRANSLATIONAL MODIFICATIONS

APPROVED BY SUPERVISORY COMMITTEE

Yingming Zhao, Ph.D.

Mark Lehrman, Ph.D.

John Falck, Ph.D.

Melanie Cobb, Ph.D.

Michael White, Ph.D.

DEDICATION

I dedicate this dissertation to my dad, Robert W. Sprung, Sr., for encouraging me to pursue my education. I thank my family for their support, especially Katie, my wife of 5 years, and our son Colin who have helped keep life in prespective through trying times. I thank Dr. Yingming Zhao for providing me with the opportunity to pursue novel and challenging research. Finally, I thank my committee members for their support and guidance during the course of this dissertation work.

DEVELOPMENT AND APPLICATION OF PROTEOMIC TECHNOLOGIES FOR THE ANALYSIS OF POST-TRANSLATIONAL MODIFICATIONS

by

ROBERT WILLIAM SPRUNG, JR.

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2007

DEVELOPMENT AND APPLICATION OF PROTEOMIC TECHNOLOGIES FOR THE ANALYSIS OF POST-TRANSLATIONAL MODIFICATIONS

ROBERT WILLIAM SPRUNG, JR., M.S., PH.D.

The University of Texas Southwestern Medical Center at Dallas, 2007

YINGMING ZHAO, PH.D.

Post-translational modifications represent a rapid and dynamic means for diversifying the chemistry of the ~20 ribosomally coded amino acids. As such, they provide an ideal mechanism for promoting cellular adaptability by facilitating the tuning of protein interactions and functions in response to changing environmental conditions. Despite their fundamental importance in regulating cellular functions and their wide implications in physiology, efficient means for the detection, enrichment and identification of proteins bearing specific modifications are lacking for most modifications. The availability of such methods would constitute invaluable tools

supporting efforts to better understand the essential regulatory roles of modifications and the means by which aberrant modifications result in the onset and progression of disease. Towards this end, my dissertation describes the development and application of novel methods for the proteomic analysis of proteins bearing known modifications, including O-GlcNAc, lysine acetylation and methyl esterification. The identification of known targets of the modifications support some of the current ideas regarding their potential impact and serve as a means of validating the methods. More importantly, the identification of novel targets for the modifications challenges some currently held concepts, in particular regarding the relatively limited regulatory roles associated with lysine acetylation. In addition, the unparalleled power of proteomics as a screening strategy is demonstrated through compelling evidence of the existence of novel lysine acylations *in vivo* with respect to propionylation and butyrylation. Together, the methods described in this dissertation and the datasets generated embody powerful platforms and rich resources for the ongoing exploration of the fundamental contributions of post-translational modifications to the regulation of biological processes.

TABLE OF CONTENTS

CHAPTER ONE 1

Introduction 1

Proteins 1

CHAPTER TWO 25

The TAS Technology for the Proteomic Analysis of O-linked N-acetylglucosamine Modified Proteins 25

Summary 25

Introduction 27

Overview of the O-GlcNAc Modification 27

Results 31

Discussion 38

Methods 41

CHAPTER THREE 64

Global Characterization of Lysine ε-N-acetylation 64

Summary 64

Introduction 66

Results 72

Discussion 75

Concluding Remarks 90

Methods 91

CHAPTER FOUR 119
Prokaryotic Lysine Acetylation 119
Summary 119
Introduction 121
Results 122
Discussion 125
Concluding remarks 132
Methods 135
CHAPTER FIVE 150
Protein Propionylation and Butyrylation 150
Summary 150
Introduction 151
Results 153
Discussion 158
Concluding Remarks 160
Methods 161
CHAPTER SIX 170
Evidence for Abundant Aspartic and Gultamic Acid Methylation in Eukaryotes 170
Summary 170
Introduction 172

Results 173

Discussion 176

Concluding Remarks 180

Methods 185

CHAPTER SEVEN 193

Conclusions and Recommendations 193

BIBLIOGRAPHY 197

VITAE 210

PRIOR PUBLICATIONS

Edwards, J.; Sprung, R.; Sprague, R.; Spence, D., "Chemiluminescence detection of ATP release from red blood cells upon passage through microbore tubing." *Analyst* **2001,** 126, (8), 1257-60.

Sprung, R.; Sprague, R.; Spence, D., "Determination of ATP release from erythrocytes using microbore tubing as a model of resistance vessels in vivo." *Analytical Chemistry* **2002**, 74, (10), 2274-8.

Fischer, D.J.; Torrence, N.J.; Sprung, R.W.; Spence, D.M., "Determination of erythrocyte deformability and its correlation to cellular ATP release using microbore tubing with diameters that approximate resistance vessels in vivo." *Analyst* **2003**, 128, (9), 1163-8.

Sprague, R.S.; Olearczyk, J.J.; Spence, D.M.; Stephenson, A.H.; Sprung, R.W.; Lonigro, A.J., "Extracellular ATP signaling in the rabbit lung: erythrocytes as determinants of vascular resistance." *American Journal of Physiology - Heart and Circulatory Physiology* **2003,** 285, (2), H693-700.

Sprung, R.; Nandi, A.; Chen, Y.; Kim, S.C.; Barma, D.; Falck, J.R.; Zhao, Y., "Tagging-via-substrate strategy for probing O-GlcNAc modified proteins." *Journal of Proteome Research* **2005,** 4, (3), 950-7

Nandi, A.; Sprung, R.; Barma, D.K.; Zhao, Y.; Kim, S.C.; Falck, J.R.; Zhao, Y., "Global identification of O-GlcNAc-modified proteins." *Analytical Chemistry* **2006,** 78, (2), 452-8

Kim, S.C.; Sprung, R.; Chen, Y.; Xu, Y.; Ball, H.; Pei, J.; Cheng, T.; Kho, Y.; Xiao, H.; Xiao, L.; Grishin, N.V.; White, M.; Yang, X.-J.; Zhao, Y., "Substrate and Functional Diversity of Lysine Acetylation Revealed by a Proteomics Survey." *Molecular Cell* **2006**, 23, (4), 607-18

Lee, J.; Xu, Y.; Chen, Y.; Sprung, R.; Kim, S.C.; Xie, S.; Zhao, Y., "Mitochondrial Phosphoproteome Revealed by an Improved IMAC Method and MS/MS/MS." *Molecular and Cellular Proteomics* **2007**, 6, (4) 669-76

Xu, Y.; Sprung, R.; Kwon, S.W.; Kim, S.C.; Zhao, Y., "Isolation of Phosphopeptides by pI-difference-based Electrophoresis." *Journal of Proteome Research* **2007**, 6, (3), 1153-7

Chen, Y.; Sprung, R.; Tang, Y.; Ball, H.; Sangras, B.; Kim, S.C.; Falck, J.R.; Peng, J.; Gu, W.; Zhao, Y., "Lysine propionylation and butyrylation are novel post-translational modifications in histones." *Molecular and Cellular Proteomics* **2007**, 6, (5), 812-9

LIST OF FIGURES

FIGURE 1-1 21

FIGURE 1-2 22

FIGURE 1-3 23

FIGURE 2-1 47

FIGURE 2-2 48

FIGURE 2-3 49

FIGURE 2-4 60

FIGURE 2-5 61

FIGURE 2-6 62

FIGURE 3-1 99

FIGURE 3-2 100

FIGURE 3-3 101

FIGURE 3-4 103

FIGURE 3-5 104

FIGURE 3-6 117

FIGURE 4-1 140

FIGURE 4-2 141

FIGURE 4-3 142

FIGURE 5-1 164

FIGURE 5-2 166

FIGURE 5-3 167

FIGURE 6-1 189

FIGURE 6-2 190

LIST OF TABLES

TABLE 2-1 51

TABLE 2-2 53

TABLE 3-1 106

TABLE 3-2 110

TABLE 3-3 118

TABLE 4-1 143

TABLE 4-2 148

TABLE 5-1 169

TABLE 6-1 192

CHAPTER ONE

INTRODUCTION

Proteins

Proteins are the cellular workhorses and represent the main functional gene products encoded by DNA. They form the foundation for cellular architecture, catalyze the interconversion `of biomolecules necessary for efficient metabolism and cellular maintenance and even serve to recognize and eliminate foreign organisms. Proteinaceous enzymes are capable of catalyzing a remarkable array of chemical reactions, taking advantage of thermodynamic principles and precise recognition and spatial arrangement of reactants. Through billions of years of evolution, some enzymes have developed the capacity to selectively modify small molecules in regio- and stereospecific manners, feats envied to this day by organic chemists. Enzymes can be capable of carrying out these specific derivatizations even in the context of an enormous range of stereoisomers present in widely varying concentrations. Such specificity is required for the biosynthesis of steroid hormones and eicosanoids, in which the spatial configuration of appended chemical groups plays a significant and even antagonistic role. Other enzymes have evolved to catalyze reactions at rates limited only by diffusion of their substrate. Proteins are important from a socioeconomic perspective as well

and a great deal of industrial effort is expended on optimizing the performance of enzymes for use as catalysts.

Depending on their function, protein half-life can range from minutes or less up to many years. Given the constantly changing environment cells and organisms occupy, it is essential that cells possess the ability to regulate protein expression, turnover and subcellular localization in order to adapt. Indeed, the maintenance and precise spatio-temporal regulation of protein activity are indispensable to permit the intricate choreography necessary for cell viability and the health of an organism.

Cells take advantage of several strategies to regulate protein function from the time a protein's gene is transcribed until its translation and throughout its existence as an intact polypeptide. Transcriptionally, proteins are regulated by the interplay between various activators and repressors of gene expression. Once transcribed, the mRNA encoding the protein can also be subject to regulation through pathways affecting mRNA stability. Translationally, a protein may be regulated by the rate and number of ribosomal translation cycles. It is largely due to such regulation at the posttranscriptional and co-translational level that mRNA levels do not necessarily reflect the abundance of the encoded protein, as demonstrated in an analysis of Saccharomyces cerevisiae (Gygi et al., 1999).

While undeniably important for cellular maintenance, regulation of protein expression at the transcriptional and co-translational level is typically a long-term

event. Given the time scale at which transcription and translation operate, as well as their restricted compartmentalization, it would be impractical for these processes to be the primary response pathways to regulate protein activity in the far reaches of the cell or in response to acute, localized stimuli. Thus, the post-translational regulation of proteins is necessary to tune their functions according to cellular demands.

Following translation, proteolysis is one effective means by which to regulate the activity of a protein. The regulation of transcription factors, such as SREBP, proteolytic enzymes, including trypsin, and blood coagulation factors are examples of well-characterized proteolytic processing events. However, this regulatory mechanism is irreversible and would impose a tremendous load on the protein synthesis machinery if it were to be more generally applied to biosynthetic pathways.

As a more generally applicable regulatory mechanism, covalent post-translational modifications represent a dynamic means of diversifying the chemistry of the twenty ribosomally coded amino acids. The spatio-temporal patterning of modifications allows for precise control of protein function and interactions and integration of signals derived from multiple stimuli. In addition, the reversibility of many modifications allows the target protein's modification status to be reset, priming it to be responsive to additional rounds of stimuli without the need to resynthesize the protein.

Arising from the myriad biochemical reactions in the cell, the high energy molecules from which post-translational modifications originate occur at up to millimolar concentrations. Once appended onto proteins, modifications can have half-lives ranging from minutes to years, as seen with collagen. Some modifications occur spontaneously and without enzymatic intervention. Oxidation of proteins is an obvious example of this phenomenon. Another example is the spontaneous deamination of asparagine and its subsequent racemization to the unnatural D-stereoisomer. Unregulated reactions such as these can be important and have been implicated as potential aging mechanisms (Clarke, 2003). Most often, however, modifications themselves are regulated at many levels in order to fine tune the activity, localization and interactions of proteins in response to the immediate needs of the cell.

Over the past century, our understanding of how post-translational modifications alter protein structure and function has largely been the result of studies arising from a reductionist approach. Such works, often elegant in their design and execution, focus on specific modifications of individual proteins under well-defined conditions. The methodology used to carry out such characterizations include the use of antibodies raised against epitopes representing a defined post-translationally modified isoform of the target protein under consideration, followed by biochemical purification and reconstitution of the regulatory pathway *in vitro*. Thorough analysis of signaling cascades using these

methods, while time consuming, are invaluable for elucidation of the chronological order of modifications. What has arisen from these studies is a wide appreciation for the rich complexity of cellular signaling pathways and such studies have served to lay the foundation for further exploration into regulatory modifications. However, antibodies raised to a specific protein isoform suffer from epitope occlusion effects. As a result, variant isoforms of the proteins, including those bearing additional post-translational modifications within the antibody target region, may not be recognized by the antibody. Thus, such tools can overlook the combinatorial potential of multiple modifications, they provide no means of establishing the existence of a regulatory modification *de novo* and are generally not amenable to a discovery based approach. An alternative approach capable of characterizing diverse modifications at once and on a cell-wide scale has arisen out of the field of proteomics.

Proteomics

The proteome is defined as encompassing "all the proteins expressed by a genome, cell or tissue" (Bradshaw and Burlingame, 2005). With the proteome as its target, the field of proteomics has among its goals the comprehensive identification, quantitation and functional characterization of a given proteome. Achieving these goals would then allow for the comprehensive comparison of the changes in patterns of protein expression, modification, interactions and activity

that occur in response to a given stimulus or the onset and progression of disease. Many practitioners of proteomics envision the development of superior diagnostic and prognostic indicators of disease states as well as the identification of novel avenues for clinical intervention as a result of accomplishing such absolute coverage of the proteome. However, comprehensive coverage of the proteome remains a distant possibility and there is doubt as to whether it can ever be achieved (Bradshaw and Burlingame, 2005). Nevertheless, significant technological and methodological developments in the field have advanced us to a point where we are beginning to see applications for proteomics from basic science to translational research (Chung et al., 2007).

In attempting the proteomic analysis of a biological sample, a number of obstacles must be overcome. Biological samples are inherently complex. For example, current estimates predict the human genome to comprise between 20000 and 25000 protein-coding genes (Stein, 2004). The actual number expressed by a given cell varies over time and depends on such factors as the tissue type and the cellular environment. Additional complexity arises from alternative mRNA splicing and post-translational modifications, leading to estimates that the human proteome may be composed of over a million protein isoforms (Chung et al., 2007). Often, these variant forms are of functional significance and the subjects of particular scientific interest. Proteins also vary widely in terms of the dynamic range of their level of expression. Some housekeeping, structural and cellular

matrix proteins may be expressed at upwards of tens-of-millions of copies per cell, while certain transcription factors and signaling proteins might only exist at the level of a few copies per cell. An extreme example of the dynamic range problem is the challenge of identifying disease markers in a sample of blood plasma. A reservoir for proteins secreted from every organ, blood plasma can have a dynamic range of more than ten orders of magnitude (Qian et al., 2006), far exceeding our ability to detect those of lower abundance. Further complexity arises from the dynamic nature of the proteome resulting in a constantly changing cadre of cellular players, as well as their dynamic and functionally significant post-translational modification.

Unfortunately, there are no methods for the analysis of protein samples that allow for amplification of a target protein in a native sample. As a result, proteins present at low-to-medium abundance in a sample are often overlooked. Since many of these proteins are important in the regulation of gene expression and cell signaling processes, the development and application of strategies to overcome sample complexity is highly desirable. To address this complexity, samples are processed and analyzed using a combination of fractionation, enrichment, instrumentation and software optimized for the problem under consideration.

Reducing Sample Complexity: Separations

Among the most widely used techniques for the proteomic analysis of biological samples are multidimensional protein separation coupled with mass spectrometry. One of the first and highest-resolution methods used for multidimensional protein separation in proteomics was two-dimensional electrophoresis (Patterson, 2000). Two-dimensional electrophoresis has been successfully applied to comparative analysis of protein samples originating from diseased and normal cells as well as cells differentially treated with drugs or other stimuli. Nevertheless, the technique is still limited in that only about 2000 protein spots can be identified in a single run, detection of proteins is limited to those of greatest abundance so the dynamic range of proteins detected is limited and the method is time consuming, reducing throughput of downstream analyses.

Perhaps today's most frequently used methods for the analysis of complex protein samples involve proteolytic digestion of the sample, referred to as bottom-up proteomics, and subsequently using one or multiple dimensions of liquid chromatography separation coupled to mass spectrometric detection. The multidimensional protein identification technology (MudPIT), developed by the Yates group at Scripps (Link et al., 1999), is widely used for the on-line fractionation of complex samples. It typically uses two dimensions of chromatography: strong cation exchange (SCX) in the first dimension and reverse phase separation in the second dimension. The method has been demonstrated to be effective in increasing the number of protein and peptide identifications,

generating more than 5000 peptide identifications from nearly 1500 proteins in a single analysis of yeast lysate (Washburn et al., 2001). Despite its advantages, MudPIT requires extensive optimization and a desalting step prior to reverse phase chromatography. Peptides are usually eluted from the SCX column in a step-wise fashion and each elution is subjected to a full 1-2 hr reverse phase separation. Thus, the data acquisition time is extensive and downstream data processing and analysis requires abundant computational power and time.

Reducing Sample Complexity: Sub-proteome Enrichment

As an alternative to chromatographic and electrophoretic separations, the use of fractionation methods prior to digestion allows targeting of the analysis to a specific subcellular compartment and efficiently reduces sample complexity.

In addition, enrichment of the sample for a feature of interest, such as post-translational modifications, can help overcome the detection issues surrounding sub-stoichiometric modifications and proteins present at low abundance. More than two-hundred protein modifications have been described (Farriol-Mathis et al., 2004), yet detailed studies of their abundance and regulatory effects have been hampered in the past due to inefficient means of their analysis. The modifications comprise a wide range of molecular weight and diverse physicochemical properties and also vary in their stability towards

downstream analysis, often limiting the techniques that may be used in their analysis.

Various affinity capture techniques have been developed which allow isolation of proteins bearing specific post-translational modifications. By far, the most widely studied modification has been phosphorylation and a number of enrichment strategies have been developed for its analysis. These include immobilized metal affinity chromatography (IMAC) (Ficarro et al., 2002), β-elimination (McLachlin and Chait, 2003), TiO_2 columns (Larsen et al., 2005), and strong cation exchange (Beausoleil et al., 2004) as well as traditional methods employing antibodies raised against specific phosphorylation motifs (Rush et al., 2005). Some of these methods can be adapted to occur off-line (Lee et al., 2007), allowing facile customization of the techniques to the sample of interest. Furthermore, the methods can be coupled to reverse phase chromatographic separation to further reduce sample complexity during analysis.

Another approach to the targeted proteomic analysis of post-translationally modified proteins involves metabolic labeling of cells with chemically derivatized analogs of the modification group. This strategy has been successfully applied to the study of farnesylated (Kho et al., 2004), phosphorylated (Kwon et al., 2003) and glycosylated proteins (Nandi et al., 2006; Sprung et al., 2005).

Analysis of Proteins: Instrumentation

Today, the mass spectrometer is an indispensable tool for proteomics research. Fundamentally, the mass spectrometer is a detector. More specifically, it is a precise instrument for the measurement of the mass of ions. Ever since its original manifestation as a tool for the elucidation of the structure of the atom (Thomson, 1897), mass spectrometers have consisted of three components: a source of ions, a mass analyzer to separate ions by mass, and a detector to sense the ions (Figure 1-1). Today there are many variations of mass spectrometers at each component, which to the uninitiated can come across as an alphabet soup of tools. For simplicity and brevity, I will focus my discussion on those instruments which are directly used in this dissertation work.

Instrumentation: Ionization Sources

An absolute requirement for a sample to be analyzed by a mass spectrometer is that it be introduced into the mass analyzer as an ion. For most of the history of mass spectrometry, analysis was generally restricted to the analysis of small molecules and metal ions, which were relatively well-understood in terms of their ionization properties and gas phase stability. It wasn't until the 1980's with Dr. John Fenn's breakthrough development of electrospray ionization (ESI) that the analysis of high molecular weight polymers and biomolecules in solution became possible (Fenn et al., 1989).

The phenomenon of ESI occurs as a liquid sample, typically at acidic pH (~pH 3), elutes from an orifice and encounters a high voltage (~1500 V/cm) (Figure 1-2). The eluting liquid droplets acquire charge and are introduced into the mass analyzer by this high-voltage field and vacuum. A combination of solvent evaporation within a heated capillary and charge repulsion at the surface of the evaporating droplet leads to desolvation and ionization of the analyte (Mann, 1990; Smith et al., 1990). Within the mass spectrometer, pressures are maintained at levels in the range of 10^{-6} to 10^{-9} torr, further assisting in desolvation of the analyte and reducing energy-sapping collisions with gas molecules. ESI is considered a 'soft' ionization method, since the ionization process typically does not result in the fragmentation of the analyte. Among the advantages of ESI are the facts that it is directly applicable to samples in aqueous solution and ionization efficiency using this method is independent of the mass of the analyte (Mann, 1990). ESI also results in multiply-charged ion species, facilitating detection of fragments of the parent ion for elucidation of molecular structure. Thus, ESI is the ionization method of choice for the analysis of complex protein digests and even intact protein samples encountered during so-called top-down proteomics experiments.

Instrumentation: Mass Analyzers

Once an ionic species is generated, it is then necessary to separate the often heterogeneous molecules in order to obtain information on the sample composition. While it is possible to manipulate ions with combinations of electric and magnetic fields, modern mass spectrometers generally employ a combination of direct current (DC) potentials and alternating current (AC) potentials operating at radio frequencies.

Quadrupole mass analyzers are among the most widely used for proteomics purposes and are commonly coupled to ESI sources. The quadrupole analyzer consists of four rods to which time-varying electric potentials are applied (Aebersold and Mann, 2003). The spatial geometry of the quadrupole "cell", along with the amplitude and frequency of applied voltage, allow selection of ions at a particular m/z or range of m/z for further analysis. Refinements to the spatial arrangement and geometry of the quadrupole and the voltages applied to the electrodes can give rise to ion traps. In these cells, applied voltages can be selected such that ions of a selected m/z will follow a stable oscillatory trajectory and become captured within the cell. Superimposition of an AC potential at the resonant frequency of a captured ion will yield an increase in the ion's kinetic energy (Payne and Glish, 2005). In the presence of an inert gas, such as helium, collisions occurring during this resonant excitation process can result in fragmentation of the ion species in a process known as collisionally induced dissociation, or CID. For proteomics studies, the daughter ions generated from the

fragmentation of peptides are scanned and detected, allowing determination of the amino acid sequence (Figure 1-3). In commercial instrumentation, multiple quadrupoles are typically employed and are arranged in series. Such an arrangement permits selection and focusing of the ions of interest, fragmentation of molecules, and further focusing of the fragment ions for detection and determination of molecular structure. The advantages of quadrupole instruments are their sensitivity, due to the ability to trap ions over time, robustness, which facilitates their use for routine analyses, and relative low cost (Aebersold and Mann, 2003). However, space-charging effects that arise from trapping a number of ions within a finite volume can result in reduced mass accuracy.

Interpreting the Data

The advent of completed genomes for humans and other model organisms made it possible to predict the amino acid sequences of the proteins encoded by their genes. Thus, algorithms taking mass spectral fragmentation information and searching the deduced peptide sequence against the predicted protein sequences from a given genome represent powerful tools for the identification and characterization of proteins from a complex sample. Furthermore, with some basic knowledge of a specific protein of interest, any discrepancy between the predicted peptide mass and that actually observed is an indication of a post-translational modification event. Such database search algorithms are at the heart

of most proteomics analyses today. These algorithms incorporate user-defined parameters such as proteolytic enzyme used for processing, genome to be used for target search and post-translational modifications to be allowed, to generate a custom database of the masses of all the possible peptide fragments of proteins in a given genome that meet these criteria. The algorithm then searches all the experimentally generated mass spectra against this custom database and reports those 'hits' representing the best peptide match for a given fragmentation pattern (Figure 1-3).

While they are indispensable for facilitating the characterization and analysis of the enormous datasets generated by high-throughput proteomics experiments, the algorithms do have limitations requiring jurisprudence in the interpretation of results obtained from their use. For example, one must specify what post-translational modifications will be considered during a database search. Each additional search parameter adds to the complexity of the possible variations of the target database and significantly increases the computational time necessary to obtain the results. However, if the modification is not specified, it will not be identified regardless if it is present. Even with identifications reported by the search software, it is often necessary to manually inspect the spectra in order to validate its accuracy and eliminate false-positive matches. This is especially the case with post-translationally modified peptides where a labile modification can result in a fragmentation pattern with fewer peptide backbone cleavages and thus

less sequence information (Mann and Jensen, 2003). Furthermore, what is perhaps the most important limitation stems not from the algorithm itself, but from the database used to generate the results. Many repositories for genomic information contain significant redundancies with respect to predicted gene sequences. Among these redundant gene annotations there may also be slight differences in sequence. This situation is often the result of submission of data from a variety of sources using different experimental approaches and can lead to difficulty in assigning definitive protein identifications. Even among those repositories subjected to expert curation and maintenance, sequence differences derived from natural sequence heterogeneity or ribosomal translational frameshifting (Dinman, 2006) are largely undocumented, giving rise to unpredictable complexity. Such unpredictability cannot be easily taken into account during proteome wide analyses using current software. Nevertheless, such complexity is an integral part of biology and by overlooking it we are, by necessity, biasing our interpretations of proteomic results.

Despite the great promise of proteomic methods for clinical and research laboratories, the techniques remain largely underutilized by the wider biomedical community. Among the reasons for this are the expense of instrumentation and the requirement for experienced personnel for the operation and maintenance of the instrumentation and interpretation of data. Nevertheless, proteomic methods represent relatively unbiased experimental methods amenable to high-throughput,

discovery based analyses. As such, they are uniquely capable of fundamentally changing paradigms and revising long-held concepts regarding biological processes and their regulation.

Questions addressed by this dissertation research

When I started out on this dissertation research there were relatively few tools amenable to the proteomic analysis of post-translational modifications. This paucity of tools stems largely from the numerous obstacles preventing their facile detection and enrichment. For example, modified proteins are found in relatively low abundance amidst an ocean of unmodified species. This low stoichiometry of modification obviates an enrichment step in order to uncover these potential biological gems. Yet, despite its intuitive simplicity, execution of this enrichment is fraught with difficulty. While more than 200 protein modifications have been described, efficient tools for the detection and enrichment of proteins bearing the majority these are lacking. Antibodies to some of the better-characterized modifications, such as phosphorylation, have been generated. Often, however, these antibodies are specific to a modification only in the context of the surrounding protein structure and do not translate well into a study of pan-cellular protein modification patterns. In addition, suitable antibodies are simply not commercially available for many of the known modifications and efforts to produce specific antibodies often fail. In some instances, the unique

physicochemical properties of a modification, especially those carrying charge, can provide a convenient means of its enrichment. On the other hand, while heterogeneity among oligosaccharide structures provides a basis for differential sorting and localization of proteins within a cell, the similarity in the properties of the oligosaccharides make their biochemical separation difficult. The development of methods that permit efficient proteomic profiling of proteins bearing specific post-translational modifications will provide insight into the biochemical pathways subject to its regulation and enlighten us as to the molecular mechanisms by which diverse modifications exert their regulatory effects. This dissertation describes the development and application of proteomic methods for the global analysis of specific post-translational modifications including the O-GlcNAc modification, lysine acylation and methyl esterification. The goals of this research are to generate tools that allow the efficient enrichment and characterization of proteins bearing specific modifications and to illuminate the unexplored potential of the modifications to influence pathways beyond those with which they are canonically associated.

Following this introduction, my dissertation will consist of 5 chapters. In Chapter 2 I describe the development of a novel method for the proteomic analysis of O-linked N-acetylglucosamine modified proteins. Exploiting a metabolic labeling step and a selective, covalent chemical ligation strategy, I demonstrate the ability to detect, isolate and identify proteins bearing the O-

GlcNAc modification. The technique is applied to the detection of O-GlcNAc modified proteins among protein lysates derived from human, mouse, monkey and fruit fly cell lines and extended to the identification of the modified proteins in *Drosophila* S2 cells and in HeLa cells. In Chapter 3, acetyl-lysine specific antibodies are applied to a tryptic digest of cellular proteins in order to efficiently isolate those modified peptides specifically. This strategy proved successful in the identification of hundreds of lysine-acetylated proteins and demonstrates that DNA-templated processes are just one of many biological pathways subject to regulation by this dynamic and pervasive modification. Intriguingly, widespread lysine acetylation was observed in mouse liver mitochondria extracts, suggesting a fundamental and previously unappreciated role for the modification in the regulation of metabolic processes. Moreover, an argument for evolutionary conservation of the modification and its targets is made through an extension of this line of research to assess the extent of lysine acetylation in *E. coli* in Chapter 4. In Chapter 5, physicochemical and structural similarities between modifications are used in the identification of two novel lysine acylations: propionylation and butyrylation. Observation of the modifications among histones and demonstration of the ability of known acetyltransferases to utilize propionyl- and butyryl-CoA *in vitro*, suggest the potential for these modifications to play a role in the regulation of gene expression. Finally, in Chapter 6 the potential for unbiased proteomics data to serve as a platform for the elucidation of novel modifications is

demonstrated by the identification of aspartic acid and glutamic acid methyl esters from HeLa and yeast lysates, respectively. Identification of these methyl esters from unenriched samples suggests that the modifications may be more abundant that tyrosine phosphorylation.

The methodologies and results of this dissertation have served to enhance the power of proteomics research in the high-throughput, cell-wide characterization of post-translational modifications. In addition, the datasets presented offer a steppingstone toward the functional characterization of the modifications, confirm some of the existing ideas on their functional impact and reveal previously unappreciated roles for the modifications in the regulation of diverse biological activities.

Figure 1-1

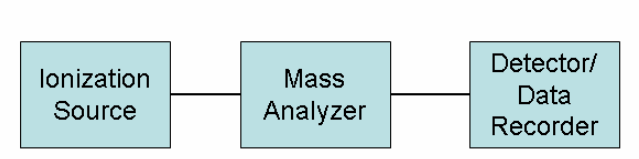

Figure 1-1. Schematic representation of the components of a mass spectrometer.

Figure 1-2

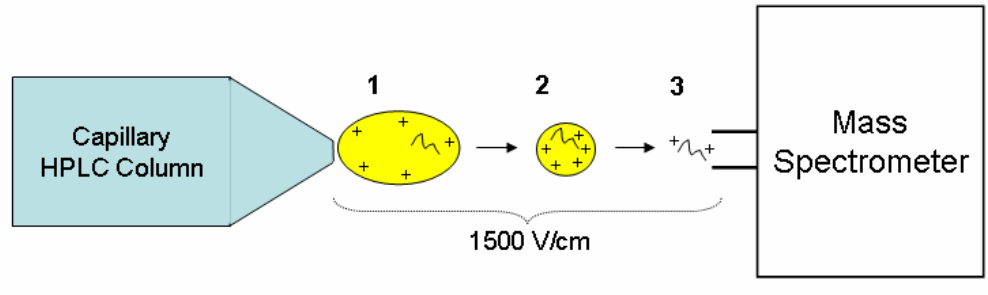

Figure 1-2. Schematic diagram of electrospray ionization. (1) Analyte-containing solvent eluting from HPLC column acquires charge from high potential field. (2) Droplets are accelerated by the electric field and solvent evaporates. (3) Solvent evaporation is further assisted upon entry into a heated capillary at the mass spectrometer interface.

Figure 1-3

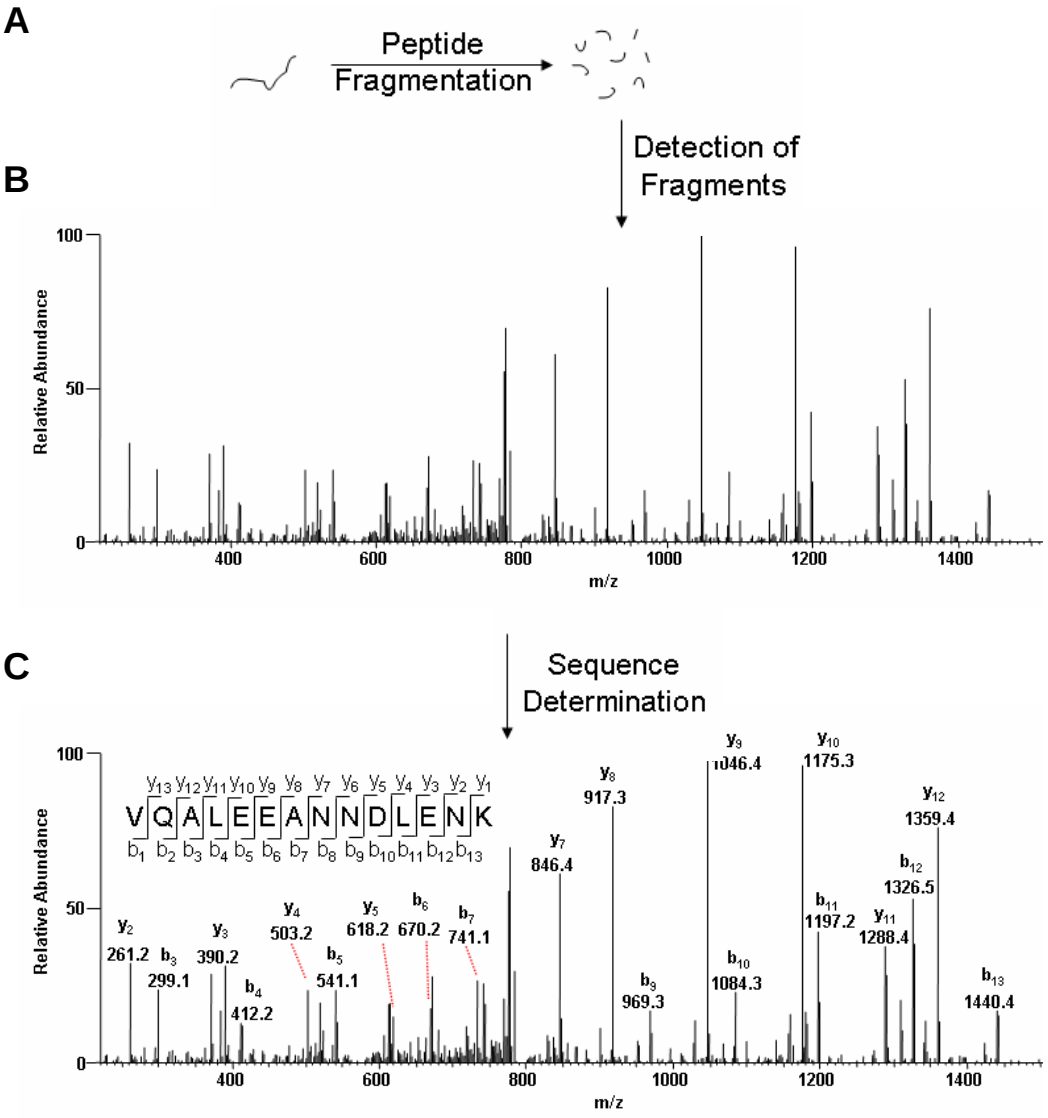
A
Peptide
Fragmentation
Detection of
Fragments
B
Relative Abundance
m/z
C
Sequence
Determination
VQALEEANNDLENK
y2 261.2
b3 299.1
y3 390.2
b4 412.2
y4 503.2
b5 541.1
y5 618.2
b6 670.2
b7 741.1
y7 846.4
y8 917.3
y9 1046.4
b9 969.3
b10 1084.3
y10 1175.3
b11 1197.2
y11 1288.4
b12 1326.5
y12 1359.4
b13 1440.4

Figure 1-3. Identification of peptide sequence through collisionally-induced dissociation (CID). (A) The peptide of interest is isolated in an ion trap mass spectrometer and subjected to CID. (B) The masses of the resulting fragments, comprised primarily of y- and b-ions (C-terminal and N-terminal fragments, respectively), are determined and a mass spectrum is obtained. (C) Database search algorithms compare the experimentally obtained spectrum to those predicted from the published genome of the target organism. Manual inspection of the returned 'hit' validates correct assignment of the mass spectrum as an unmodified peptide of sequence "VQALEEANNDLENK" from the human protein keratin 9.

CHAPTER TWO

THE TAS TECHNOLOGY FOR THE PROTEOMIC ANALYSIS OF O-LINKED N-ACETYLGLUCOSAMINE MODIFIED PROTEINS

SUMMARY

The identification of proteins bearing a specific post-translational modification can give insight into the functions of the modification. Proteomic analysis of post-translationally modified proteins is usually challenging due to the high complexity and wide dynamic range of biological samples, as well as the unavailability of efficient methods to enrich for proteins of interest. In this chapter, I demonstrate a strategy for the detection, isolation, and profiling of O-linked N-acetylglucosamine (O-GlcNAc) modified proteins. The strategy involves three steps: metabolic labeling of cells with an unnatural GlcNAc analog, peracetylated azido-GlcNAc; chemoselective conjugation of azido-GlcNAc modified proteins via the Staudinger ligation, which is specific between phosphine and azide, using a biotinylated phosphine capture reagent; and detection and affinity purification of the resulting conjugated O-GlcNAc modified proteins. Since the approach relies on a tag (azide) in the substrate, the technique

was dubbed the tagging-via-substrate (TAS) strategy. Using this approach, azido-GlcNAc modified proteins from the cytosolic lysates of HeLa, 3T3, COS-1, and S2 cell lines were able to be specifically labeled and subsequently detected, suggesting the azido-substrate could be tolerated by the enzymatic systems among these cells from diverse biological species. Azido-GlcNAc modified proteins were isolated from the cytosolic extract of S2 cells and led to the identification of 10 previously reported and 41 putative O-GlcNAc modified proteins, by nano-HPLC/MS/MS. Furthermore, extension of the technique to HeLa nucleocytoplasmic lysate resulted in the identification of 21 known and 178 novel O-GlcNAc modified proteins. This study demonstrates that the TAS approach is an efficient tool for the detection and proteomic analysis of O-GlcNAc modified proteins.

This work came to fruition as the result of a collaborative effort between me and other members of the Zhao lab: Yue Chen, Drs. Animesh Nandi, Yingxin Zhao and Sungwon Kwon. The synthetic chemistry expertise of Dr. Deb Barma from the laboratory of Dr. John Falck in the synthesis of the capture reagents and azido-GlcNAc used for the TAS strategy also made this work possible.

INTRODUCTION

Overview of the O-GlcNAc Modification

The post-translational modification of serine and threonine residues by O-linked N-acetylglucosamine (O-GlcNAc) was first described by Dr. Gerald Hart more than two decades ago (Torres and Hart, 1984). While glycosylation is typically associated with stable, heterogeneous oligosaccharide modifications that take place in the endoplasmic reticulum and serve to direct protein and vesicle sorting, the O-GlcNAc modification occupies a unique niche as a dynamic monosaccharide modification among nuclear and cytosolic proteins and its modification patterns can be altered in response to extracellular stimuli. Addition and removal of the modification are each catalyzed by a single enzyme: O-GlcNAc transferase (OGT) and O-GlcNAcase, respectively. The modification is ubiquitous among human tissues and has been found in all metazoans analyzed from Cannahabidilis elegans to Drosophila to the rhesus monkey, mice and even plants. To date, the modification has not been observed in the fission yeast Schizosaccharomyces pombe, the budding yeast Saccharomyces cerevisiae or the bacterium Escherishia coli, despite investigation by me and other groups.

O-GlcNAc has been attributed a number of regulatory effects and the OGT enzyme has been demonstrated to be essential for viability at the single cell level (O'Donnell et al., 2004). Among the regulatory consequences of the

modification are the modulation of protein-protein interactions, effects on protein-DNA binding, as well as regulation of protein localization and stability (Zachara and Hart, 2004a). It has also been reported to be involved in the regulation of proteosome function (Zhang et al., 2003).

The O-GlcNAc modification is in a unique position to act as a cellular energy and nutrient sensor. The chemical components from which the molecule is made are derived from diverse aspects of cellular metabolism. A potentially important regulatory implication for O-GlcNAc modification is the fact that the intracellular level of UDP-GlcNAc, the source of GlcNAc for the transferase, is regulated by glucose availability. It is estimated that between 2-5% of the glucose taken up by a cell is diverted through the hexosamine biosynthetic pathway leading to formation of UDP-GlcNAc (Wells et al., 2003). Thus, the position of UDP-GlcNAc in the hexosamine biosynthetic pathway, coupled with the rapid turnover of O-GlcNAc, implicates the O-GlcNAc modification as a regulatory modification related to cellular energy status (Zachara and Hart, 2004a). Indeed, dysregulation of the O-GlcNAc modification has been shown to be involved in the development of insulin resistance and diabetes. Furthermore, O-GlcNAc levels have been experimentally demonstrated to be influenced by changes in extracellular levels of glucose and glucosamine. Yet, while the O-GlcNAc modification is known to be important for cellular regulation, the identification and confirmation of the modification remain daunting challenges.

A Chemical-biology Approach for Tagging Protein Modifications

Chemical reactions that are selective and specific to a single chemical moiety in the context of complex biological milieu represent powerful tools for chemical analysis of biologically interesting molecules. A prime example of one such reaction is the enhanced Staudinger ligation pioneered by the Bertozzi group as a bioorthoganal reaction between a phosphine and an azide (Saxon and Bertozzi, 2000). The reaction possesses several unique features that make it ideal for biological applications. First, the reaction can be carried out at room temperature in aqueous solution. Second, there is no significant cross-reaction with endogenous cellular compounds that is detectable during the experimental time window. Third, neither phosphines nor azides occur in any known biomolecule. Finally, the phosphine reagent can be engineered to incorporate a wide variety of tags including fluorescent probes and affinity tags such as biotin and the FLAG tag.

The cellular metabolism of azide-tagged sugars has been characterized by the Bertozzi group as it applies to sialic acid oligosaccharide modifications. Although inorganic azide is toxic to cells, the azido group appears much less so in the context of the monosaccharides and is tolerated at concentrations up to 250 μM. This chemistry has even been demonstrated to be a viable strategy for the labeling of cells *in vivo*, as indicated by a study involving the use of azide-tagged

N-acetyl mannosamine for the detection of sialic acid glycosylation of cell surface proteins in living mice (Prescher et al., 2004).

The Tagging-via-substrate (TAS) Approach for the O-GlcNAc Modification

To facilitate the proteomic analysis of O-GlcNAc modified proteins, the tagging-via-substrate, or TAS technology, was developed. The application of TAS involves three steps: the metabolic labeling of a cell with an azide-labeled analog of GlcNAc, the chemoselective conjugation of a biotin probe onto the labeled proteins via the Staudinger ligation, and an affinity purification step for the selective isolation of O-GlcNAc modified proteins from a complex biological sample (Figure 2-1).

For this dissertation work, the TAS strategy was extended to the detection and identification of O-GlcNAc modified proteins using peracetylated azido-GlcNAc as a substrate for O-GlcNAc modification. The results demonstrate that the substrate could be incorporated into proteins for O-GlcNAc modification in cell lines derived from diverse organisms including human, monkey, mouse and *Drosophila*, suggesting the unnatural azido-GlcNAc molecule is compatible with enzymatic systems in these cell lines. Chemical conjugation and subsequent affinity purification of azido-GlcNAc modified proteins from metabolically labeled S2 cell lysate led to the identification of 51 putative O-GlcNAc modified proteins, including 10 homologs of known O-GlcNAc modified proteins. The

strategy was further extended to the proteomic analysis of O-GlcNAc modified proteins in HeLa cells resulting in the identification of 199 putative O-GlcNAc modified proteins, 21 of which had been previously described as bearing the modification. Immunoprecipitation of specific proteins was used to confirm the modification status of 23 of the proteins identified in the HeLa screen, offering validation of the strategy. The results suggest that TAS technology is a useful tool for detection and proteomics of O-GlcNAc modified proteins in order to elucidate the functional consequences of the modification.

RESULTS

Specific Detection of O-GlcNAc Modified Proteins after Metabolic Labeling

To test the feasibility of metabolically labeling and detecting O-GlcNAc modified proteins, cells were incubated with peracetylated azido-GlcNAc or peracetylated GlcNAc as a control. The cytosolic proteins were extracted, and a capture reaction was performed to specifically biotinylate the azido-GlcNAc labeled proteins. The proteins were then precipitated to remove excess capture reagent. Fifteen μg of protein was separated on a gradient gel, transferred to a nitrocellulose membrane and probed with streptavidin-HRP (Figure 2-2). For HeLa, 3T3, COS-1 and S2 cell lines, the signal was specific for cells labeled with azido-GlcNAc. The signal was dependent on the capture reaction step and could

be competed out by performing the capture reaction in the presence of exogenous N_3-substrate or by probing the nitrocellulose membranes with streptavidin-HRP in the presence of exogenous biotin. These results demonstrate that cells can incorporate the exogenous azido-GlcNAc substrate into proteins. The results also suggest that we are able to specifically detect azido-GlcNAc modified proteins from cytosolic lysates of HeLa, 3T3, COS-1 and S2 cell lines. In addition, these results demonstrate that the TAS technology can be successfully applied to studies of diverse organisms including human, mouse, monkey and *Drosophila*.

Nano-HPLC/Mass Spectrometric Analysis of Affinity Purified Proteins

To identify those proteins bearing the O-GlcNAc modification, 10 mg each of control and azido-GlcNAc labeled cytosolic protein from S2 cells was subjected to a capture reaction. The resultant biotinylated proteins were affinity purified using streptavidin-conjugated beads. Non-specifically bound proteins were removed by stringent wash conditions (8 M urea, 3 M KCl, 1% SDS-PBS). Proteins remaining bound to the beads were digested with trypsin. Tryptic peptides were analyzed by LC-MS/MS for protein identification (Figure 2-3).

Putative O-GlcNAc modified proteins were identified by subtractive analysis of the nonspecifically binding proteins of the control sample from those proteins identified in azido-GlcNAc labeled cell lysates. Analysis of S2 cell cytosolic extract led to the identification of 52 proteins, 11 of which have been

previously identified as bearing the O-GlcNAc modification, providing a good positive control (Table 2-1).

Having demonstrated the viability of the TAS strategy for the proteomic analysis of O-GlcNAc modified proteins, the technique was extended to the nucleocytoplasmic extract of HeLa cells. One hundred ninety-eight azido-GlcNAc modified proteins were identified (Table 2-2) and 21 of these were previously reported to be O-GlcNAc modified (Zachara and Hart, 2004a). The number of proteins identified here represents a manifold increase in the number of O-GlcNAc modified proteins previously identified in a single study (Zachara and Hart, 2004a).

Summary of Identified Proteins

The identified proteins were manually classified according to their annotated activities described in the NCBI (www.ncbi.nlm.nih.gov) and FlyBase (http://flybase.bio.indiana.edu) protein databases (Figures 2-4 and 2-5, Tables 2-1 and 2-2).

The number of proteins identified in the two proteomic screens reinforces previous notions of the regulatory effects of the O-GlcNAc modification. In addition, the results offer new insight into a potential wider role for O-GlcNAc in metabolic regulation.

The O-GlcNAc modification has been implicated as a potential sensor of glucose availability and cellular nutritional status. It has also been suggested to play a role in stabilizing proteins towards degradation (Zachara and Hart, 2002). The peracetylated azido-GlcNAc used in this study would be expected to enter the hexosamine biosynthetic pathway at a point after the rate-limiting step of GlcNAc biosynthesis, thereby promoting an increase in the available pool of UDP-GlcNAc (Wells et al., 2003). One might predict that this cellular environment would mimic a situation of ample glucose availability. The expectation would be that proteins associated with the metabolism of glucose would be stabilized. Indeed, proteins involved in glucose metabolism were identified as being O-GlcNAc modified in the S2 sample (Figure 2-4 and Table 2-1), consistent with previously reported results from other species (Wells et al., 2002). O-GlcNAc was also found on many enzymes involved in amino acid, carbohydrate, lipid, cofactor and nucleotide metabolism from the HeLa sample (Figure 2-5 and Table 2-2). Such enzymes mark a primary interface between energy generation and energy use and represent ideal targets for a regulatory modification related to cellular energy status.

It has been estimated that as much as 25% of the energy used by a cell is for the purpose of protein synthesis and processing (Inoki et al., 2003). Thus, it would be reasonable to expect that the regulation of the transcription/translation machinery of the cell would be coupled to energy availability. The O-GlcNAc

modification, a good candidate for a regulatory modification related to glucose availability, is known to occur among transcription and translation factors (Zachara and Hart, 2004a). Using the TAS technology, transcription and translation factors that have been previously reported to be O-GlcNAc modified were identified. From the HeLa sample, more than a quarter of the proteins identified were annotated as being involved in protein synthesis and processing, representing the largest functional group among the identified proteins. Numerous factors involved in transcription and RNA processing were also identified. The prevalence of O-GlcNAc on these proteins provides a mechanism through which energy availability may be coupled directly to protein metabolism and gene expression patterns.

O-GlcNAc has also been suggested to play a role in the cellular stress response pathways (Zachara and Hart, 2004a; Zachara and Hart, 2004b). In support of this notion, 8 (15%) of the proteins identified in the S2 cell analysis were annotated as being involved in stress response pathways. These include heat shock proteins and proteins involved in the regulation of cellular redox homeostasis. Five proteins involved in the cellular response to DNA damage were identified from HeLa cells, suggesting that the O-GlcNAc modification may play a role in mediating cell survival pathways. It is already known that O-GlcNAcase is a target for caspase cleavage. Although the *in vivo* consequences of this cleavage are unknown, it does not affect the catalytic activities of the enzyme *in*

vitro (Wells et al., 2002). Cleavage may impact the enzyme's regulation or substrate recognition. In addition to the DNA-damage responsive proteins identified from HeLa cells, 16 proteins were annotated as playing a role in redox homeostasis. A number of chaperone proteins important in the unfolded protein response were also identified, further supporting a potential regulatory role for O-GlcNAc with respect to the stress response. An additional 17 HeLa proteins were identified for which a functional annotation was not available.

The identification of 21 previously known O-GlcNAc modified proteins represents almost 20 percent coverage of previously reported O-GlcNAc modified proteins. This serves as a good positive control and helps confirm that the TAS technology can be successfully applied to studies of the O-GlcNAc modification. In addition, the repertoire of putative O-GlcNAc modified proteins was greatly expanded by this work, now comprising >200 proteins (Tables 2-1 and 2-2)).

There was very little contamination of the cytosolic extract by subcellular organelles. This result is expected since the O-GlcNAc modification occurs primarily on soluble proteins of the nucleus and cytoplasm. Of the S2 cell proteins for which an unambiguous localization was annotated, 24 (70%) were localized to the cytoplasm, 5 (15%) were of nuclear origin, 3 (9%) were from the endoplasmic reticulum and a single protein (3%) was identified from both the peroxisome and the plasma membrane.

Verification of the O-GlcNAc Modification in HeLa Proteins

Identification of known O-GlcNAc modified proteins suggests that the strategy is amenable to the identification of proteins bearing the azido-GlcNAc analog. To further validate the presence of the azido-GlcNAc modification in the identified proteins, reciprocal immunoprecipitation was used in an effort to confirm some previously known and putative O-GlcNAc modified proteins from the list of identified proteins.

In the case of those proteins for which antibodies suitable only for Western blotting analysis, but not immunoprecipitation, were available, nucleocytoplasmic proteins were isolated from HeLa cells metabolically labeled with peracetylated GlcNAc or peracetylated azido-GlcNAc, and conjugated with biotinylated phosphine capture reagent. Biotinylated, azido-GlcNAc proteins were affinity purified using avidin-monomer agarose and resolved by SDS-PAGE. After transfer to a nitrocellulose membrane, proteins were probed with specific antibodies to verify their presence among the azido-GlcNAc modified proteins.

In the case of those proteins for which immunoprecipitation antibodies were available, the protein of interest was isolated from cells labeled with peracetylated GlcNAc or peracetylated azido-GlcNAc by immunoprecipitation first. The immunoprecipitated samples were then conjugated with capture reagent

and the resulting conjugate was subjected to Western blot analysis using streptavidin-HRP.

This analysis confirmed O-GlcNAc modification of all 23 proteins assayed including 10 known and 13 newly identified O-GlcNAc modified proteins (Figure 2-6). This result offers further evidence that the TAS technology can be applied to the accurate, selective identification of O-GlcNAc modified proteins.

DISCUSSION

Despite being implicated in myriad processes and pathways, the molecular mechanisms by which O-GlcNAc can potentiate these effects remain unknown. Global identification of the dynamically modified O-GlcNAc proteins would give insight into the pathways subject to its regulation and help in the determination of how O-GlcNAc exerts its regulatory effects. Unfortunately, as with most post-translational modifications, there has not been an efficient technology available for the selective isolation of O-GlcNAc modified proteins.

Previous methods used in the identification and characterization of O-GlcNAc modified proteins include the development of O-GlcNAc-specific antibodies (Comer et al., 2001), the use of wheat germ agglutinin columns (Montreuil and al., 1994), β-elimination coupled with Michael addition

(BEMAD) (Vosseller et al., 2005; Wells et al., 2002) and the attachment of chemically derivatized or radiolabeled galactose molecules onto O-GlcNAc modifications through enzymatic reactions (Tai et al., 2004; Torres and Hart, 1984). However, these methods have drawbacks that preclude their application to the global characterization of O-GlcNAc modified proteins.

Antibody-based methods using O-GlcNAc-specific monoclonal antibodies rely on the binding affinity between the antibody and its antigen. However, the antibodies available against O-GlcNAc demonstrate relatively low binding affinity, which precludes the use of harsh washing conditions. Any immunoaffinity purification using this tool would therefore not only isolate O-GlcNAc modified proteins, but their interaction partners as well. Sensitivity is also compromised due to the low binding affinity. Thus, while antibody-based methods represent a good starting point for the characterization of O-GlcNAc modified proteins, the issue of low binding affinity makes antibodies an inappropriate choice for large-scale proteomic analyses.

Wheat germ agglutinin columns provide a method for the enrichment of glycosylated proteins. However, such methods are not generally applicable for the proteomic analysis of monosaccharide modifications. Enrichment of glycosylated proteins using wheat germ agglutinin columns is not selective for a particular sugar and the binding of wheat germ agglutinin to monosaccharide modifications is weak (Montreuil and al., 1994). Such methods also suffer from contamination

of the sample by the non-specific binding of proteins and cannot employ the harsh washing conditions we use (8 M urea, 3 M KCl, 1% SDS/PBS) to reduce such non-specific interactions.

BEMAD involves the base-catalyzed elimination of chemical moieties at serine and threonine residues. The resulting α,β-unsaturated carbonyl is subsequently reacted with a nucleophilic agent, such as dithiothreitol, to tag the previously modified site. BEMAD has been described as a method allowing the selective isolation and site mapping of O-GlcNAc specifically (Wells et al., 2002), as well as serine and threonine post-translational modifications in general (Vosseller et al., 2005). However, the harsh conditions used in BEMAD strategy can lead to side reactions with unmodified serine and threonine residues and cause protein degradation (McLachlin and Chait, 2003).

Enzymatic tagging of GlcNAc residues with radiolabeled galactose was described in the seminal paper on O-GlcNAc (Torres and Hart, 1984). Unfortunately, this strategy suffers from low sensitivity and doesn't allow for the selective isolation of the labeled proteins for proteomic analysis. Subsequently, Tai et al. described a technique for the labeling of GlcNAc residues with a ketone analog of UDP-galactose (Tai et al., 2004). The method exploits a mutant of the enzyme β-1,4-galactosyl transferase containing an active site which accommodates the ketone moiety in the substrate. The ketone allows covalent coupling of aminooxy biotin as a tag. Using this method, 29 peptides

representing 25 proteins were identified as O-GlcNAc modified in samples of rat brain, including 2 previously known proteins (Khidekel et al., 2004).

The TAS technology offers significant advantages to the approaches outlined above. It can be used for both the selective detection and isolation of azido-GlcNAc modified proteins. The strong interaction between biotin and streptavidin allows efficient removal of non-specifically binding proteins by employing harsh washing solutions such as 8 M urea, 3 M KCl and 1% SDS, allowing selective enrichment. Finally, the Staudinger ligation is rapid and truly bioorthogonal, since neither azides nor phosphines have been identified as constituents of any known biomolecule.

METHODS

Materials. HPLC grade acetonitrile, water, and methanol were from EM Science (Gibbstown, NJ); trifluoroacetic acid (TFA) was from Fluka (Buchs, Switzerland); acetic acid was from Aldrich (St. Louis, MO); ammonium bicarbonate was from Fisher (Fair Lawn, NJ); trypsin was from Promega (Madison, WI). Peracetylated azido N-acetyl glucosamine (azido-GlcNAc), azido-farnesyl pyrophosphate and biotinylated phosphine capture reagent were synthesized in-house (Kho et al., 2004). DMEM and Drosophila SFM cell media were purchased from Gibco (Grand Island, NY). Protease inhibitor cocktail was

from Roche Molecular Biochemicals (Indianapolis, IN). Ammonium bicarbonate, SDS and salts used in buffer preparation were from Fisher Scientific (Fair Lawn, NJ). BSA, glucosamine, phosphate buffered saline, tween-20 and urea were purchased from Sigma (St. Louis, MO). Streptavidin polyacrylamide beads were from Pierce (Rockford, IL). 4-20% gradient minigels and nitrocellulose membrane were obtained from Bio-Rad Laboratories (Hercules, CA). Western Lightning Chemiluminescence Reagent Plus was from Perkin-Elmer Life Science (Boston, MA).

Cell Culture. *HeLa, 3T3 and Cos-1 cells.* Dishes (15 cm) of HeLa, 3T3 and Cos-1 cells were grown to 70% confluence in complete DMEM (4.5 g glucose/L) with 10% FBS and antibiotics in a 37º C, 5% CO_2 incubator. For labeling, media was aspirated and the cells were washed with PBS. Fifteen mL DMEM containing 250 μM peracetylated azido-GlcNAc and 1g glucose/ L was then added to each dish and cells were incubated for 24 h.

S2 cells. Dishes (15 cm) of S2 cells were grown to confluence in Drosophila SFM (2 g glucose/L, 2 g trehalose/L) with 10% FBS and antibiotics at 25º C. For labeling, media was aspirated and the cells were washed with PBS. Cells were labeled for 24 h with 500 μM peracetylated azido-GlcNAc in Drosophila SFM diluted 1:4 with S2 media prepared in-house containing no sugars (10% FBS, 0.5 g glucose/L, 0.5 g trehalose/L final).

Preparation of Cytosolic Extract. Three hours prior to harvest, 150 μL of 400 mM glucosamine (4 mM final) was added to plates to inhibit the action of O-GlcNAcase, the enzyme responsible for the removal of the O-GlcNAc modification. Cells were harvested by scraping in 2 mL PBS per plate and pelleted by centrifugation at 1,000 x *g* and 4º C for 10 min. The cell pellet was resuspended in an equal volume of ice-cold hypotonic lysis buffer (10 mM KCl, 1.5 mM $MgCl_2$, 1 M Hepes pH 7.4) with protease inhibitors and 4 mM glucosamine, incubated on ice for 30 min and homogenized using 50 passes in a Dounce homogenizer employing a B-rated pestle. The lysate was subjected to ultracentrifugation at 100,000 x *g* for 1 hr at 4º C to remove intact cells, cell membranes and organelles including nuclei. The supernatant was collected and considered the cytosolic extract.

Sample Preparation. Protein was precipitated from the cytosolic extract by acetone precipitation at –20º C overnight. Precipitated proteins were pelleted by centrifugation at 15,000 x *g* for 10 min, washed twice with 1 mL cold acetone and redissolved in PBS containing 1% SDS for a final concentration of 4 mg/mL.

Capture Reaction. In order to specifically biotinylate O-GlcNAc-N_3 modified proteins, a capture reaction using the biotinylated phosphine capture

reagent (bPPCR) was performed on cytosolic lysates. For HeLa, 3T3 and Cos-1 cell lysates, a 100 μL solution consisting of 150 μg protein and 50 μM biotinylated phosphine capture reagent in 1% SDS was reacted at room temperature overnight with shaking in the presence or absence of 100 μM azido-farnesyl pyrophosphate. For S2 cell lysate, a 100 μL solution consisting of 150 μg protein and 500 μM biotinylated phosphine capture reagent in 1% SDS was reacted at room temperature overnight with shaking in the presence or absence of 1 mM azido-farnesyl pyrophosphate.

Western Blot Analysis. Capture reaction samples were acetone precipitated to remove excess bPPCR and resuspended in SDS-PAGE sample buffer. Twenty-μg samples were run on a 4-20% gradient minigel. Proteins were then transferred to a nitrocellulose membrane. The membrane was blocked in 5% BSA TBS (25 mM Tris pH 7.5, 150 mM NaCl) with shaking at room temp for 1 h. The membrane was probed for 1 h with a 1:20,000 dilution of streptavidin-conjugated HRP in the presence and absence of 100 μM D-biotin. After probing, the membrane was washed 6 times with shaking for 15 min each time in TBST (25 mM Tris pH 7.5, 150 mM NaCl, 0.4% Tween-20). The blot was developed using the Western Lightning chemiluminescence reagent.

Affinity Purification. Excess bPPCR was removed from capture reaction samples by precipitating proteins in 9 volumes ice-cold acetone at –20º C for >2 h. Protein pellets were resuspended in 1% SDS PBS. Ten mg protein was added to a 20 µL volume of biotinylated polyacrylamide beads. Samples were rocked at room temp for 30 min. To remove the beads, samples were centrifuged at 1,000 x *g* for 5 min and the supernatant decanted. The beads were then washed three times each in 8 M urea, 3 M KCl, 1% SDS/PBS and 50 mM NH_4CO_3 pH 8.0.

Nano-HPLC/mass spectrometry for protein identification. After affinity purification, 50 ng trypsin was added directly to the beads. Tryptic peptides were sequentially extracted with 50% acetonitrile/ 50% water (v/v), and 75% acetonitrile/ 25% water (v/v) solutions. The peptide extracts were dried in a SpeedVac. The peptide samples were cleaned with ZipTip C_{18} (Millipore, Bedford, MA) prior to nano-HPLC/tandem mass spectrometry analysis.

Nano-HPLC/tandem mass spectrometry analysis was performed in an LCQ DECA XP ion-trap mass spectrometer (ThermoFinnigan, San Jose, CA) equipped with a nano ESI source (ThermoFinnigan, San Jose, CA). The electrospray source was coupled online with an Agilent 1100 series nano flow HPLC system (Agilent, Palo Alto, CA). Two µl of the peptide solution in buffer A (2% acetonitrile/ 97.9% water/ 0.1% acetic acid (v/v/v)) was manually loaded into a capillary HPLC column (50 mm length X 75 µm ID, 5 µm particle size, 300

Å pore diameter) packed in-house with Luna C18 resin (Phenomenex, St. Torrance, CA). The peptides were eluted from the column with a gradient of 5% to 80% buffer B (90% acetonitrile/ 9.9% water/ 0.1% acetic acid (v/v/v)) in buffer A over 30 min. The eluted peptides were electro sprayed directly into the LCQ mass spectrometer.

The MS/MS spectra were acquired in a data-dependent mode that determined the masses of the parent ions and fragments of the three strongest ions. All MS/MS spectra were searched against the NCBI non-redundant protein sequence database using the MASCOT search for protein identification. The following search parameters were used in all MASCOT searches: maximum of one missed trypsin cleavage and a maximum 4-Da error tolerance in the MS and 0.8-Da in MS/MS data. Those protein hits with three ions matched with a score higher than 35 were considered true protein identifications. All other hits were manually analyzed to ascertain the accuracy of protein identification. In the manual analysis, the criteria used for a true identification is that the masses of all the major peaks (typically more than 7 peaks) in a MS/MS spectrum had to match those of the theoretically calculated fragment ions and the mass errors were within 0.5 Da.

Figure 2-1

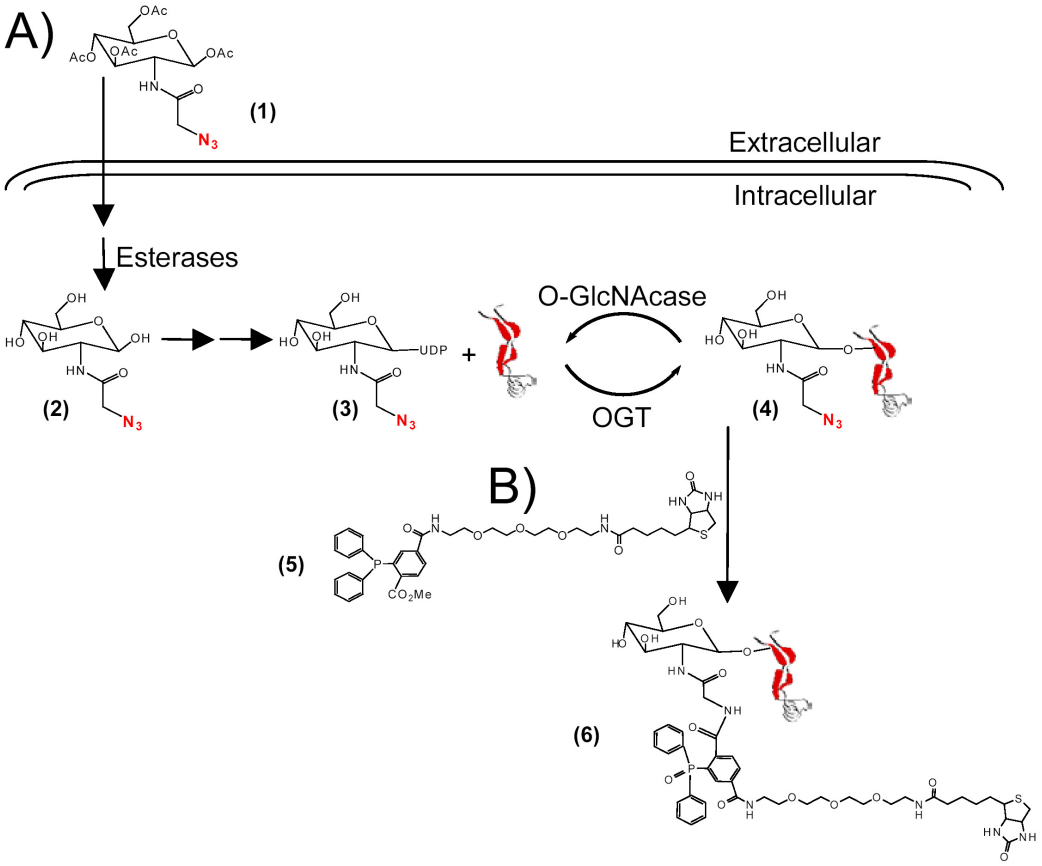

Figure 2-1. The TAS strategy for the global detection of O-GlcNAc modified proteins. A) Peracetylated N-(2-azidoacetyl)-glucosamine (1) is delivered into the cells through metabolic labeling and converted to UDP-azido-GlcNAc (3), which is used by O-GlcNAc transferase for O-GlcNAc modification of proteins (4). B) Biotin-conjugation reaction using biotinylated phosphine capture reagent (5) results in the selective biotinylation of azido-GlcNAc modified proteins (6).

Figure 2-2

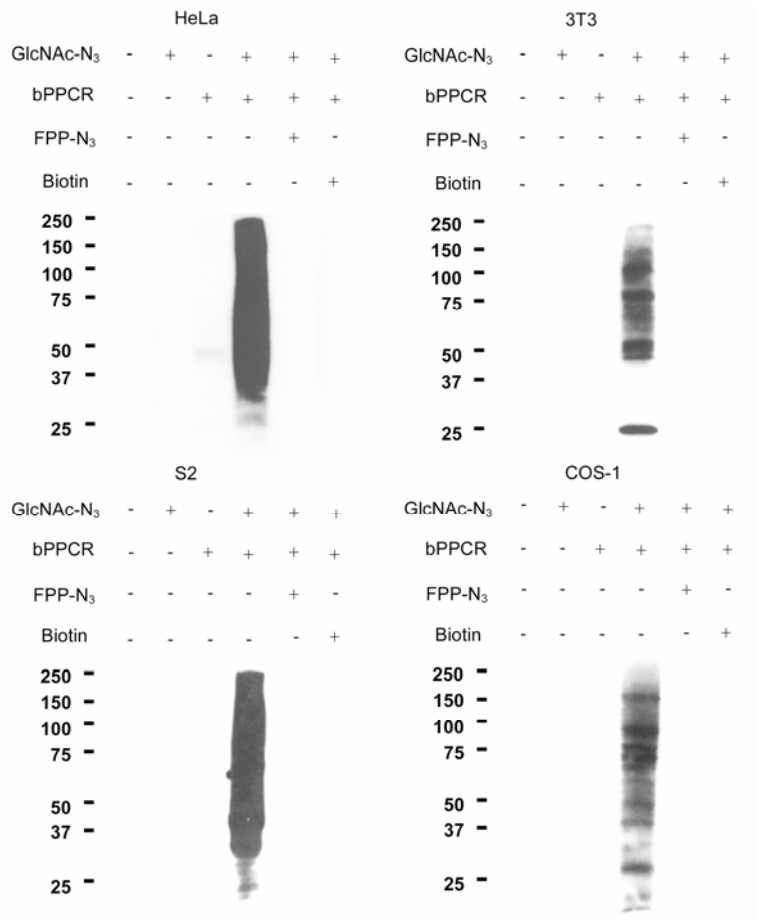

Figure 2-2. Detection of metabolically labeled O-GlcNAc-N_3 modified proteins in diverse organisms. Cytosolic proteins from cells in the presence and absence of metabolic labeling with azido-GlcNAc (GlcNAc-N_3) were conjugated to biotinylated phosphine capture reagent (5). The resulting biotinylated proteins were separated by SDS-PAGE and detected by western blot analysis, probing with HRP-conjugated streptavidin.

Figure 2-3

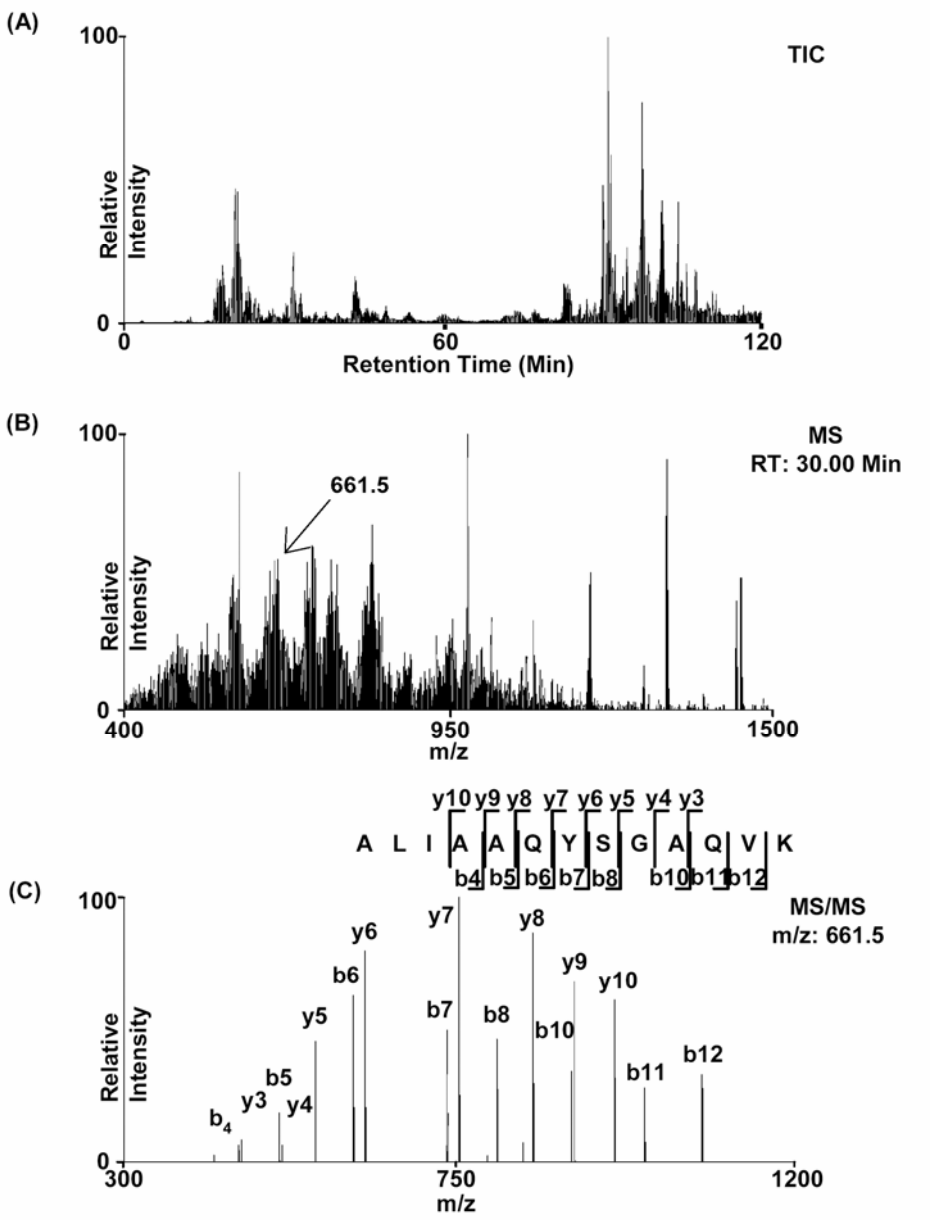
(A)
100
0
Relative Intensity
TIC
0
60
120
Retention Time (Min)
(B)
100
0
Relative Intensity
661.5
MS
RT: 30.00 Min
400
950
1500
m/z
y10 y9 y8 y7 y6 y5 y4 y3
A L I A A Q Y S G A Q V K
b4 b5 b6 b7 b8 b10 b11 b12
(C)
100
0
Relative Intensity
MS/MS
m/z: 661.5
b4
y3
b5
y4
y5
b6
y6
b7
y7
b8
y8
b10
y9
y10
b11
b12
300
750
1200
m/z

Figure 2-3. An example of nano-HPLC/MS/MS analysis for protein identification. (A) Total ion current chromatogram of a nano-HPLC/MS/MS of the on-bead digest of affinity purified S2 cell cytosolic extract. (B) MS spectrum at retention time of 30.00 min. (C) MS/MS spectrum of 661.5 m/z, which identified the peptide ALIAAQYSGAQVK, unique to translation elongation factor 1 gamma.

Table 2-1

Identified Protein	gi#	Functional Classification	Molecular Function
Translation elongation factor 1 β[a,b]	13124189	Translation	Elongation factor
Translation elongation factor 1 γ[a,b]	6716514	Translation	Elongation factor
eIF-4a[a,b]	418925	Translation	Translation initiation factor
Nucleosome remodeling factor[b]	18447455	Transcription	Chromatin remodelling
Thioredoxin peroxidase 1	24641739	Stress Response	Peroxidase
Catalase[b]	17981717	Stress Response	Peroxidase
Slow superoxide dismutase	4572573	Stress Response	Superoxide dismutase
CG6776	21355779	Stress Response	Glutathione transferase
Glutathione-S-transferase 1-1	121694	Stress Response	Glutathione transferase
Capulet	28317136	Structural	Actin binding
Chd64	21355917	Structural	Actin binding
β tubulin 60D	21626738	Structural	Tubulin binding
CG10724	21356527	Structural	Actin polymerizing/depolymerizing
Cofilin/actin depolymerizing factor	1168731	Structural	Cofilin/actin depolymerizing
Hsc70-4[a]	24647034	Chaperone, Stress Response	Protein folding
Hsc70-3[a]	24641402	Chaperone, Stress Response	Protein folding
Hsc70Cb[a]	21357475	Chaperone, Stress Response	Protein folding
Hsp83	17647529	Chaperone, Stress Response	Protein folding
Proteosome alpha 6 subunit	24586400	Protein Degradation	Proteolysis
Proteosome alpha 4 subunit	12643270	Protein Degradation	Proteolysis
Prolyl oligopeptidase	24583414	Protein Degradation	Proteolysis
Puromycin sensitive aminopeptidase	24655260	Protein Degradation	Aminopeptidase
Ubiquitin activating enzyme 1	17861718	Protein Degradation	Ubiquitin activation
Protein disulfide isomerase[b]	17647799	Protein Modification	Disulfide isomerase
Rana	21356159	Protein Targeting	Small GTPase
Calreticulin[b]	24645441	Protein Transport	Calcium ion binding
Annexin X[a]	71777	Signaling	Calcium ion binding
Regucalcin	18860103	Signaling	Calcium mediated signaling
EGFR type 3	552091	Signaling	EGF receptor
CG10417	19921654	Signaling	Protein S/T phosphatase
CG6180	24583960	Signaling	Kinase inhibitor
14-3-3 ε	24647885	Cell Cycle	PKC inhibitor
Eb1	19921706	Cell Cycle	Microtubule binding
Pyruvate kinase[a]	3108349	Glucose Metabolism	Kinase
GAPDH[a]	84977	Glucose Metabolism	Dehydrogenase
Triosephosphate isomerase	267156	Glucose Metabolism	Isomerase
Myo-inositol-1-phosphate synthase	4106368	Carbohydrate Metabolism	
CG1129	21355443	Carbohydrate Metabolism	Mannose-1-phosphate guanylyltransferase
CG2767	24644950	Carbohydrate Metabolism	Alcohol dehydrogenase
CG6084	24662781	Carbohydrate Metabolism	Aldehyde reductase
Ade5	18860083	Nucleotide Metabolism	Purine base metabolism
CG32626	18859679	Nucleotide Metabolism	Purine base metabolism
Thymidylate synthase	13959716	Nucleotide Metabolism	Nucleotide synthase
Anon1A3	5616364	Unknown	
CG6259	21355253	Unknown	
CG5094	24584835	Unknown	
CG31305-PJ	24645977	Unknown	
CG2446	20129011	Unknown	
CG10630	24658349	Unknown	
Transitional ER ATPase[b]	3337433	Unknown	
CG6891	21483504	Unknown	

Table 2-1. List of identified O-GlcNAc-N_3 modified proteins from S2 cell extract and their functional classifications. Classifications are based on NCBI and FlyBase annotations. Proteins previously reported to be O-GlcNAc modified are indicated by [a]. Proteins with annotated cellular localization other than the cytoplasm are indicated by [b].

Table 2-2

Protein Name	gi#	Biological Process
Protein Synthesis/Folding/Turnover		
heat shock protein	4204880	Protein Folding
heat shock 27kDa protein 8	7657146	Protein Folding
heat shock protein 70 ***X,Z***	2135328	Protein Folding
heat shock 70kDa protein 4	6226869	Protein Folding
heat shock 70kda protein 5	16507237	Protein Folding
heat shock 70kda protein 8 isoform 1	5729877	Protein Folding
heat shock 70kDa protein 8 isoform 2	24234686	Protein Folding
heat shock 90kda protein 1, alpha ***X,Z***	13129150	Protein Folding
chaperone protein HSP90 beta ***X***	9082289	Protein Folding
heat shock 90kda protein 1, beta ***X***	20149594	Protein Folding
PRO1633	11493524	Protein Folding
Cell Growth/Maintenance		
TRF1 ***Z***	5032169	Telomere Maintenance
TRAF2 ***Z***	55959980	Cell Growth/Maintenance-Proliferation
calregulin ***Z***		Cell Growth/Maintenance
ezrin (p81) (Cytovillin) (Villin 2)	119717	Cell Growth/Maintenance
homologue to Drosophila photoreceptor protein calphotin	27544394	Signal Transduction
IQ motif containing GTPase activating protein 1	4506787	Signal Transduction
tyrosine 3/tryptophan 5-monooxygenase activation protein, beta polypeptide	4507849	Signal Transduction
platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit (45kD)	4557741	Signal Transduction
tubulin, beta, 2	5174735	Signal Transduction
Rho-GDI alpha ***X,Z***	4757768	Signal Transduction
cofilin 1 ***X***	5031635	Signal Transduction (Rho)
filamin 1 (actin-binding protein-280)	4503745	Signal Transduction (IkB/NFkB/MAPK Cascade)
filamin C (gamma)	4557597	Signal Transduction (MAPK Cascade)
heat shock 70kDa protein 9B	292059	Signal Transduction (MAPK Cascade)
actinin alpha 4 (ACTN4 protein)	33874637	Signal Transduction-Calcium/Calmodulin Binding
tyrosine phosphatase shp-2 ***Z***	4558223	Signal Transduction (Tyrosine Phosphatase)
paxillin gamma	1912057	Signal Transduction (Receptor Tyrosine Kinase Signaling)
galectin-3 ***Z***	3402185	Signal Transduction (Advanced Glycation End Product (AGE) Receptor)
G protein-coupled receptor 87	31542858	Signal Transduction (GPCR Signaling)
guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	5174447	Signal Transduction (GPCR Signaling)
kainate receptor subunit	17384624	Signal Transduction (GPCR Signaling)
ral guanine nucleotide dissociation stimulator	18204882	Signal Transduction (GPCR Signaling)
tyrosine 3/tryptophan 5-monooxgenase activation protein, gamma polypeptide	9507245	Signal Transduction (PKC Binding)
protein phosphatase 2 catalytic subunit, alpha isoform	4506017	Signal Transduction (Wnt, Jak/Stat, MAPK Cascade)
protein phosphatase 2 regulatory subunit A (PR65), alpha isoform	231443	Signal Transduction (Wnt, Jak/Stat, MAPK Cascade)
HMT1 hnRNP methyltransferase-like 2 isoform 3	38195087	Signal Transduction (Defense Response)

Protein Name	gi#	Biological Process
Protein Synthesis/Folding/Turnover		
poly(rC) binding protein 1	6754994	Translational Activator
Tu translation elongation factor, mitochondrial	34147630	Translation
HsGCN1	2282576	Translation
elongation factor 1-alpha ***X***	181967	Translation
eukaryotic translation elongation factor 1 alpha 1	4503471	Translation
eukaryotic translation elongation factor 1 gamma	4503481	Translation
eukaryotic translation elongation factor 2 ***X,Z***	4503483	Translation
eukaryotic translation initiation factor 2, subunit 3 gamma	4503507	Translation
eukaryotic translation initiation factor 3 subunit 9 eta	3123230	Translation
eukaryotic translation initiation factor 4A, isoform 1 ***X***	4503529	Translation
eukaryotic translation initiation factor 4 gamma, 1 isoform 4	38201619	Translation
eukaryotic translation initiation factor 5A ***Z***	4503545	Translation
eukaryotic translation initiation factor 5A2	9966867	Translation
eukaryotic translation initiation factor 5B	15451892	Translation
EPRS protein	37589132	Translation
asparaginyl-tRNA synthetase	4758762	Translation
glycyl tRNA synthetase	3845409	Translation
isoleucyl-tRNA synthetase	730870	Translation
leucyl-tRNA synthetase	24496789	Translation
threonyl-tRNA synthetase	14714853	Translation
tryptophanyl-tRNA synthetase isoform a	47419914	Translation
valyl-tRNA synthetase 2	5454158	Translation
nucleophosmin 1 (anaplastic lymphoma kinase Ki-1) ***X***	40353734	Translation
ribosomal protein S23	4506701	Translation
ribosomal protein L4	16579885	Translation
ribosomal protein L15	21040388	Translation
ubiquitin-activating enzyme E1	23510338	Protein Degradation
isopeptidase T (ubiquitin specific protease 5)	1732411	Protein Degradation
proteasome 26S ATPase subunit 2	4506209	Protein Degradation
proteosome beta 2 subunit	4506195	Protein Degradation
proteosome regulatory particle subunit p44S10	34304592	Protein Degradation
proteasome activator complex subunit 2 PA28 beta	18203506	Protein Degradation
proteasome 26S non-ATPase subunit 1	25777600	Protein Degradation
T-complex protein 1 (TCP1), alpha subunit (HSP60 family)	135538	Protein Folding
TCP1, subunit 3 (gamma) (HSP60 family)	31542292	Protein Folding
TCP1, subunit 4 (delta) (HSP60 family)	15928389	Protein Folding
TCP1, subunit 6A (zeta 1) (HSP60 family)	4502643	Protein Folding
TCP1, subunit 8 (theta) (HSP60 family)	42662316	Protein Folding

Protein Name	gi#	Biological Process
Cell Growth/Maintenance		
protein phosphatase methyltransferase-1	7022606	Phosphatase Inhibitor
myosin, light polypeptide 6, alkali, smooth muscle and non-muscle	33620739	Actin Binding
myosin heavy chain nonmuscle form A ***X***	625305	Actin Binding
fascin 1	13623415	Actin Binding
CAPZB ***Z***	56417843	Structural/Actin Binding
plastin 1 I isoform (PLS1 protein)	21411382	Structural/Actin Binding
actinin alpha 1	28723	Structural
actin, beta	14250401	Structural
actin, gamma 1 propeptide	4501887	Structural
dynein heavy chain	30581065	Structural
annexin 1 ***X,Z***		Structural
tubulin alpha 6 ***X***	21619816	Structural
tubulin, beta polypeptide 4, member Q	29788766	Structural
microtubule-associated protein 1 (MAP1)	6912494	Microtubule Binding
microtubule-associated protein 4 (MAP4) ***X***	539639	Microtubule Binding
RNA Processing/Transcription		
heterogeneous nuclear ribonucleoprotein A3	34740329	Transcription
heterogeneous nuclear ribonuclearprotein D	181914	Transcription
heterogeneous nuclear ribonucleoprotein K	13384620	Transcription
heterogeneous ribonuclear particle protein U	284156	Transcription
elongin C	5032161	Transcription Factor
stat5a ***X***	7513380	Transcription Factor
RNA Processing/Transcription		
Sp1 ***X,Z***	13638437	Transcription Factor
TBP-interacting protein 120A	16758920	Transcriptional Activator
tumor protein p53 binding protein, 1	5032189	Transcriptional Activator
EBNA-2 co-activator (100kD)	32879913	Transcription Regulator
NFkB ***X,Z***	34577122	Transcription Regulator
c-myc ***X,Z***	127619	Transcription Regulator
P105mcm (MCM6) ***Z***	1197636	Transcription Regulator
minichromosome maintenance protein 7 isoform 1	33469968	Nucleo-side/tide/acid/base Metabolism
DEAD (Asp-Glu-Ala-Asp) box polypeptide 1	4826686	RNA Processing
DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	11890755	RNA Processing
HLA-B associated transcript 1 (DEAD box motif)	4758112	RNA Processing
DEAH box polypeptide 15	13124667	RNA Processing
histidine triad nucleotide binding protein 1	10439439	RNA Processing
Far upstream element binding protein 2	37078468	RNA Processing
KHDRBS1 protein	33873325	RNA Processing

Protein Name	gi#	Biological Process
RNA Processing/Transcription		
poly(rC)-binding protein 2 isoform b	14141166	RNA Processing
TAR DNA binding protein	6678271	mRNA Processing
splicing factor 3b, subunit 3, 130 kDa	285999	RNA Splicing
Carbohydrate Metabolism		
glucose-6-phosphatase ***Z***	4557599	Glucose Transport
mitochondrial malate dehydrogenase precursor	21735621	TCA Cycle
malate dehydrogenase (EC 1.1.1.37), cytosolic	7431153	TCA Cycle
isocitrate dehydrogenase	11374664	TCA Cycle
NADP-dependent isocitrate dehydrogenase	3641398	TCA Cycle
ATP-citrate synthase	20141248	TCA Cycle
M2-type pyruvate kinase ***X***	189998	Glycolysis
pyruvate kinase 3 isoform 1 ***X***	33286418	Glycolysis
oxoglutarate (alpha-ketoglutarate) dehydrogenase protein	13436359	Glycolysis
glyceraldehyde 3-phosphate dehydrogenase ***X***	182979	Glycolysis
phosphofructokinase	227448	Glycolysis
phosphoglucomutase (EC 5.4.2.2) 1	284351	Glycolysis
phosphoglycerate kinase (EC 2.7.2.3) 2 ***X***	88344	Glycolysis
aldolase A ***Z***	28614	Glycolysis/Gluconeogenesis
phosphoglycerate mutase 1	38566176	Glycolysis/Gluconeogenesis
fructose bisphosphate aldolase	312137	Glycolysis/Gluconeogenesis
glucose phosphate isomerase	18201905	Glycolysis/Gluconeogenesis
neuroleukin	189238	Glycolysis/Gluconeogenesis
GPI protein	34784772	Glycolysis/Gluconeogenesis
alpha enolase ***X,Z***	2661039	Glycolysis/Gluconeogenesis
enolase 1 ***X***	4503571	Glycolysis/Gluconeogenesis
glucose 6 phosphate 1 dehydrogenase	120731	NADPH Metabolism
phosphogluconate dehydrogenase	40068518	NADPH Metabolism
Unknown Function		
ape1	6980832	Unknown
hypothetical protein DKFZp762N0610.1	11360295	Unknown
hypothetical protein FLJ32421	21389601	Unknown
probable protein 55.11 homolog p67	7513235	Unknown
similar to eukaryotic translation initiation factor 2, subunit 3 gamma, 52kda	22057930	Unknown
similar to eukaryotic translation initiation factor 3, subunit 5 epsilon, 47kDa	29731731	Unknown
similar to unactive progesterone receptor, 23 kD	27481323	Unknown
splicing factor homolog	543010	Unknown
moesin/anaplastic lymphoma kinase fusion protein ***X***	14625824	Unknown
tropomyosin 4-anaplastic lymphoma kinase fusion protein	13274400	Unknown

Protein Name	gi#	Biological Process
Unknown Function		
calpastatin, long form *X*	2134859	Unknown
profilin	3891601	Unknown
SP-H antigen	743447	Unknown
cysteine and glycine-rich protein 1	4758086	Unknown
glutathione transferase (EC 2.5.1.18) class mu, GSTM3	106129	Unknown
glypican-3	57283544	Unknown
Intracellular Transport		
coronin, actin binding protein, 1C (WD repeat-containing protein)	7656991	Intracellular Transport
GDP dissociation inhibitor 2	6598323	Intracellular Transport
coatomer protein complex, subunit alpha (COPA protein)	23512328	Intracellular Transport
coatomer protein complex, subunit beta	7705369	Intracellular Transport
clathrin heavy chain 1	4758012	Intracellular Transport
clathrin, heavy polypeptide-like 1 isoform b	9257202	Intracellular Transport
beta-coat protein	1082254	Intracellular Transport
archain 1	773575	Intracellular Transport
karyopherin (importin) beta 1	19923142	Intracellular Transport
karyopherin alpha 2 (importin alpha 1)	4504897	Intracellular Transport
importin 7	11544639	Intracellular Transport
CSE1 chromosome segregation 1-like protein isoform b	29029561	Intracellular Transport
cellular apoptosis susceptibility protein	3560557	Intracellular Transport
General Metabolism		
high density lipoprotein binding protein (vigilin)	4885409	Sterol/lipid Metabolism
leukotriene A4 hydrolase	4505029	Lipid Metabolism
methylenetetrahydrofolate dehydrogenase 1	30048109	Coenzyme Metabolism
pyridoxal kinase	32186842	Coenzyme Metabolism
carbonyl reductase 1	4502599	Secondary Metabolites Synthesis
S-adenosylhomocysteine hydrolase	9951915	Metabolism
phosphoglycerate mutase processed protein	20530939	Metabolism
endoplasmic reticulum-associated amyloid beta peptide-binding protein	20336350	Metabolism
peroxiredoxin 1	4505591	Redox Homeostasis
peroxiredoxin 2 isoform a	32189392	Redox Homeostasis
peroxiredoxin 4 (thioredoxin peroxidase)	5453549	Redox Homeostasis
Nucleotide Metabolism		
5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	20127454	Nucleotide Metabolism
Inosine-5'-monophosphate dehydrogenase 2	124419	Nucleotide Metabolism
phosphoribosylglycinamide formyltransferase	2425021	Nucleotide Metabolism
adenylate kinase 2 isoform b	7524346	Nucleotide Metabolism
phosphoribosyl pyrophosphate synthetase 1-like 1	28557709	Nucleotide Metabolism

Protein Name	gi#	Biological Process
Nucleotide Metabolism		
methylthioadenosine pyrophosphorylase	9082015	Nucleotide Metabolism
thymidine kinase ***X,Z***	23503074	Nucleotide Metabolism
ribonucleoside-diphosphate reductase M1 chain	4506749	Nucleotide Metabolism
Amino Acid Metabolism		
cystathionase (cystathionine gamma-lyase)	22027612	Amino Acid Metabolism
cystathionine beta-synthase minor isoform	3776243	Amino Acid Metabolism
3-phosphoglycerate dehydrogenase	5771521	Amino Acid Metabolism
phosphoglycerate dehydrogenase	23308577	Amino Acid Metabolism
argininosuccinate synthetase	16950633	Amino Acid Metabolism
carbamoyl-phosphate synthetase 1	21361331	Amino Acid Metabolism
carbamoylphosphate synthetase 2/aspartate transcarbamylase/dihydroorotase	18105007	Amino Acid Metabolism
DNA Damage Response		
mutS homolog 6	19548099	DNA Mismatch Repair
Ku70	15825665	DNA Damage Response
SLK protein	29126889	DNA Damage Response
poly(ADP-ribose) polymerase	190167	DNA Damage Response
DNA fragmentation factor 35	4926918	Signal Transduction (Caspase Cascade/Apoptosis)
DNA fragmentation factor 45 (DFF45) ***Z***	47132600	Signal Transduction (Caspase Cascade/Apoptosis)
Muscle development		
transgelin 2; SM22-alpha homolog	4507357	Structural

Table 2-2. List of all the putative O-GlcNAc modified proteins identified in a proteomic screen of HeLa nucleocytoplasmic lysates. Protein names, gi# and annotated biological process associated with each protein are reported. ***X*** indicates proteins previously reported to be O-GlcNAc modified. ***Z*** indicates those identifications that were validated by reciprocal immunoprecipitation or avidin chromatography/Western blotting analysis.

Figure 2-4

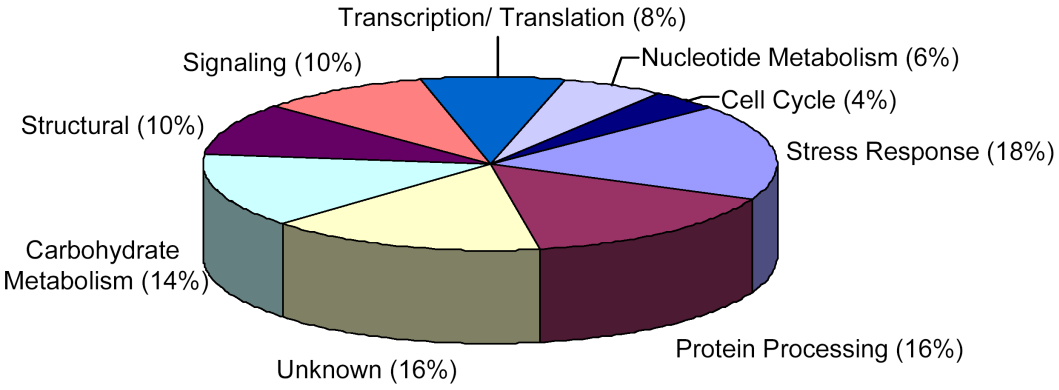

Figure 2-4. Pie diagram showing functional classes for putative O-GlcNAc modified proteins from S2 cell cytosolic extract. Classifications are based on NCBI and FlyBase annotations.

Figure 2-5

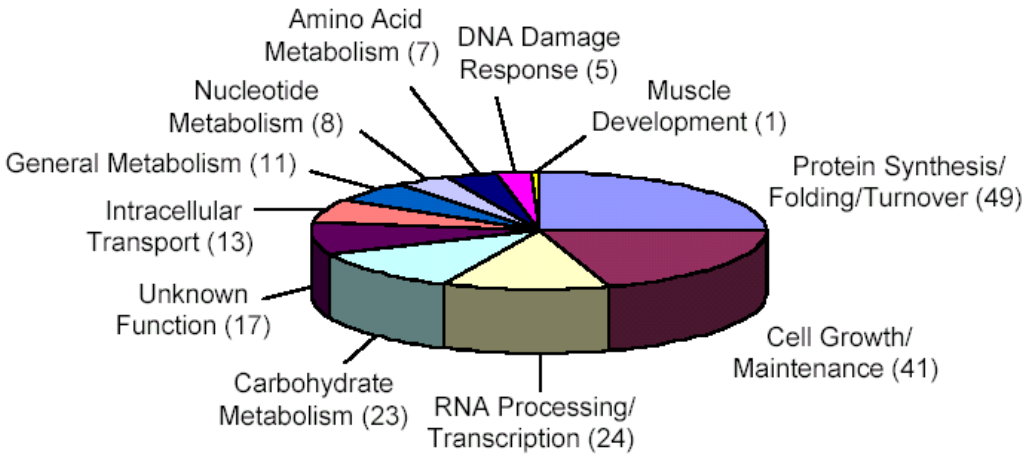

Figure 2-5. Pie diagram showing functional classes for putative O-GlcNAc modified proteins from HeLa nucleocytosolic extract. Classifications are based on NCBI annotations.

Figure 2-6

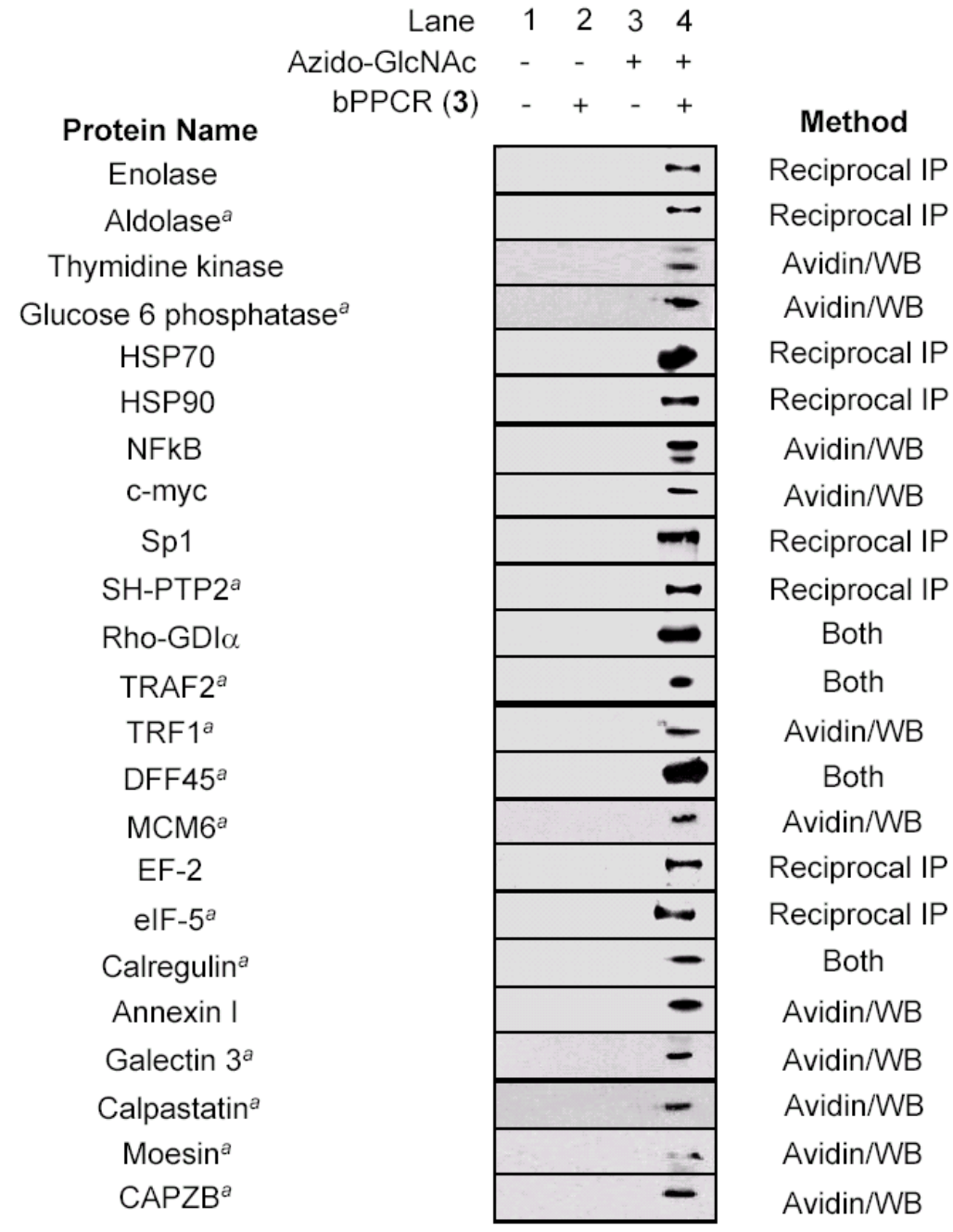
Lane 1 2 3 4
Azido-GlcNAc - - + +
bPPCR (3) - + - +
Protein Name
Method
Enolase
Reciprocal IP
Aldolase[a]
Reciprocal IP
Thymidine kinase
Avidin/WB
Glucose 6 phosphatase[a]
Avidin/WB
HSP70
Reciprocal IP
HSP90
Reciprocal IP
NFkB
Avidin/WB
c-myc
Avidin/WB
Sp1
Reciprocal IP
SH-PTP2[a]
Reciprocal IP
Rho-GDIα
Both
TRAF2[a]
Both
TRF1[a]
Avidin/WB
DFF45[a]
Both
MCM6[a]
Avidin/WB
EF-2
Reciprocal IP
eIF-5[a]
Reciprocal IP
Calregulin[a]
Both
Annexin I
Avidin/WB
Galectin 3[a]
Avidin/WB
Calpastatin[a]
Avidin/WB
Moesin[a]
Avidin/WB
CAPZB[a]
Avidin/WB

Figure 2.6. Verification of the azido-O-GlcNAc modification. Avidin/WB: proteins were conjugated with bPPCR, affinity-purified with avidin-monomer agarose and detected by Western blot analysis using specific antibody. Reciprocal IP: protein was immunoprecipitated with specific antibody, conjugated with bPPCR and detected by Western blot analysis using streptavidin HRP. *a* indicates novel O-GlcNAc modified proteins identified in this study. Experiments performed by Dr. Animesh Nandi.

CHAPTER THREE

GLOBAL CHARACTERIZATION OF LYSINE ε-N-ACETYLATION

SUMMARY

The reversible, post-translational acetylation of histones has been known for almost 40 years. Since its initial discovery, the suite of known targets of lysine acetylation has grown to a total pool of 80 proteins including various transcription factors. Acetylation as a regulatory modification has been gaining interest with the discovery of its role in the regulation of DNA-templated processes. Indeed, inhibitors of certain deacetylases show the potential to be chemotherapeutic interventions for various cancers, while homologs of the NAD-dependent Sir2 deacetylase from yeast have been shown to influence lifespan in C. elegans, yeast and mammals. However, the extent of lysine acetylation among proteins beyond those involved in DNA-templated processes remains largely unknown. The proteomic dissection of protein acetylation is a powerful tool for identifying the targets of acetylation and may aid in the elucidation of the functional consequences of acetylation while providing insight into the pathways regulated by the modification. Using a combination of immunoaffinity purification and

HPLC-tandem mass spectrometry, this dissertation work led to the identification of hundreds of lysine acetylated proteins from HeLa cell nuclear and cytosolic lysate, as well as in mouse liver mitochondria extract. The nuclear and cytosolic proteins identified in this screen include proteins with diverse biological functions including metabolic enzymes, stress response proteins and regulators of cytoskeletal dynamics. Furthermore, the novel observation of abundant mitochondrial lysine-acetylation offers compelling evidence of the regulatory potential of the modification toward metabolic processes in this organelle.

This work was the result of a collaborative effort between me and other members of the Zhao lab, including Yue Chen, Drs. Sung Chan Kim, Yoonjung Kho, Yingda Xu and Junmei Zhang. In addition, the laboratories of Drs. Michael White and Xiang-Jian Yang provided experimental evidence of the functional consequences of some of the novel acetylation events observed.

INTRODUCTION

Overview of Lysine Acetylation

Acetylation represents one of the most abundant of the more than 200 known protein modifications found in cells and falls into two categories: N-terminal and ε-lysine acetylation. However, post-translational lysine acetylation is fundamentally different from N-terminal acetylation. For example, N-terminal acetylation occurs co-translationally and has been estimated to modify around 85% of eukaryotic proteins (Polevoda and Sherman, 2002). Also, N-terminal acetylation is regarded as a static and permanent modification with a half-life on the order of that of the protein it modifies. Lysine acetylation, by contrast, has the distinction of being reversible, having the potential to be a regulatory modification correlating with extracellular or intracellular signaling events.

While the post-translational modification of lysine side chains was first described nearly four decades ago, through the ability of cells to incorporate C^{14}-labeled acetate onto histones (Vidali et al., 1968), the first non-histone protein to bear the modification, p53, was not revealed until 30 years later (Gu and Roeder, 1997). Since its initial discovery, fewer than 90 proteins have been described as targets for lysine acetylation (Kouzarides, 2000; Yang, 2004b). This lack of known targets for acetylation is one of the major bottlenecks restricting advances in our understanding of the biological impacts of lysine acetylation.

The reversibility of lysine acetylation is mediated by the opposing actions of a cell's particular complement of the ~20 known histone acetyltransferases (HATs) and 18 histone deacetylases (HDACs). Many of the HATs that have been identified are known to play dual roles as transcription factors. Indeed, even the nomenclature given to the HATs and HDACs betrays the general perception that the major function of these enzymes is to regulate chromatin dynamics and gene expression. This view is being challenged as several non-histone, non-nuclear targets for acetylation have been described and as deacetylases with subcellular localizations outside of the nucleus have been identified (Cohen and Yao, 2004; McKinsey and Olson, 2004; Michishita et al., 2005).

Consistent with the notion of a regulatory role for lysine acetylation, the modification has been found to modulate protein-DNA interactions, protein-protein interactions and protein stability. In addition, acetylation of a lysine side chain precludes modification by other moieties, such as ubiquitin and SUMO, thus there can be competition between modifications for a single site. This can result in differential function or processing of the substrate, as has been documented (Minucci and Pelicci, 2006).

Among the functional consequences that have been described as resulting from lysine acetylation, changes in gene expression are most prevalent. This is evidenced by the extensive body of work considering the impact of acetylation on chromatin structure, dynamics and DNA accessibility (Grunstein, 1997). DNA is

maintained in a condensed form known as chromatin through its interaction with histones. This interaction is believed to be mediated by electrostatic interactions between the negatively charged backbone of DNA and the lysine-rich N-terminal tails of histones. Acetylation of histone tails is generally associated with a relaxation of chromatin structure, allowing access to the underlying DNA by transcription factors.

Lysine Acetylation and Disease

There is also evidence that dysregulated acetylation may lead to disease. Inasmuch as cancer involves alterations in gene expression patterns that allow a cell to escape the regulatory influence of apoptotic programs and other cell death stimuli, regulatory modifications that are able to influence gene expression regimens represent prime pathways for cancer prevention and intervention. Post-translational lysine acetylation is of particular interest in this regard. Evidence that aberrant histone modifications, including lysine acetylation, may play a role in the onset and progression of cancer abounds (Hake et al., 2004). Changes in the modification status of transcription factors, histones, oncogene and tumor suppressors have been described and are correlated with HDAC activity (Zhu et al., 2004). These changes are found among cancers from a variety of origins including prostate, ovarian, breast, colorectal, lung cancers and lymphomas. However, the precise molecular mechanisms through which aberrant lysine

acetylation leads to cancer, and the particular HATs and HDACs that are responsible, are unclear at present and await further characterization.

Because of acetylation's known roles in the regulation of epigenetics and oncogenic processes, HDAC inhibitors have been developed and are currently undergoing more than 30 clinical trials as novel therapeutic agents for both hematologic malignancies and solid tumors (Egger et al., 2004). Suberoylanilide hydroxamic acid (SAHA), the most advanced clinical compound of HDAC inhibitors, affects cell growth, differentiation, and apoptosis preferentially on a broad spectrum of transformed cells, but not normal cells (Kelly et al., 2003). However, effects on histones cannot explain such specificity, as histones are ubiquitously expressed in both normal and transformed cells. Unfortunately, the downstream targets of SAHA, as well as the mechanism of its anti-tumor activity, are largely unknown.

Though its role in the onset and progression of cancer may come as no surprise, given the well-established role for lysine acetylation in the regulation of gene expression, cancer is not the only pathological state with which aberrant acetylation is associated. Aberrant acetylation has been implicated in such diverse disease states as neurodegeneration, cardiovascular disease and inflammatory lung disease as well (Blander and Guarente, 2004; Carrozza et al., 2003; McKinsey and Olson, 2004; Yang, 2004a).

Lysine Acetylation and Longevity

Another intriguing aspect of lysine acetylation is its putative role in the regulation of lifespan. This regulation has been proposed as being mediated through the activity of the class-III, NAD-dependent deacetylases (Sirt1-7 in humans). For example, one of the few laboratory perturbations able to efficiently extend the lifespan of organisms is to subject them to a calorie restricted diet. It has been experimentally demonstrated that the activity of Sirt1 is increased under such conditions (Bordone and Guarente, 2005). This affects the acetylation status and activity of its downstream targets including such diverse non-histone substrates as p53, FOXOs, Ku70, and tubulin, which regulate various cellular functions including the stress response, apoptosis, and energy metabolism. Although the mechanisms involved in conveying the longevity-extending properties of the pathway are unclear, the acetylation pathway, with respect to the class-III deacetylases, lies at a crossroads where energy availability and regulatory modifications intersect at the level of acetyl-CoA (the source for acetate in the transferase reaction) and NAD (cofactor for the class III HDACs).

Proteomics of Lysine Acetylation

The recent identification and functional characterization of acetyllysine residues among cytosolic proteins and signaling molecules (Cohen and Yao, 2004; Kaiser and James, 2004), and demonstration of the presence of HATs and

HDACs in the cytosol and mitochondria, support the notion that the modification may be ubiquitous among cellular compartments. Despite this evidence, the scope of lysine acetylation remains unclear. The application of proteomic approaches to identify the targets of acetylation represents a potentially powerful tool to enhance our understanding of the extent of lysine acetylation in cells and its role in the regulation of fundamental biological processes such as aging and disease.

The difficulties met with attempting a proteomic analysis of lysine acetylation are the same as those encountered in the analysis of other post-translational modifications. Proteins are often modified in sub-stoichiometric ratios. As a result, modified peptides are found amidst an excess of unmodified peptides following enzymatic digestion of a complex biological sample. This scenario precludes direct analysis of post-translational modifications except for the most abundant instances. In addition, many important regulatory proteins that are subject to the modification of interest are often of low-to-medium abundance. To analyze those proteins in the context of a sample of high complexity requires an efficient and selective isolation strategy.

In developing an efficient purification strategy for post-translationally modified proteins, one must exploit the unique features of modification in order to maximize the recovery of modified protein while eliminating most of the unmodified sample. Lysine acetylation poses a significant challenge in this regard, given the small size and relative lack of physical chemical properties that

readily distinguish lysine acetylated proteins from their unmodified counterparts. Thus the availability of antibodies directed toward acetyllysine represents a good tool for the proteomic characterization of the modification.

Here a proteomics survey of the acetyllysine sub-proteome from HeLa cells and mouse liver mitochondria is generated by combining immunoaffinity purification of lysine-acetylated peptides and exhaustive peptide identification by nano-HPLC/MS/MS analysis (Figure 3-1A). This screen led to the identification of 388 lysine acetylation sites from 195 proteins, dramatically extending the known inventory of *in vivo* acetylation sites and substrates. In contrast to previous studies of lysine acetylation, most lysine-acetylated proteins identified in this study do not have obvious roles in DNA-templated processes. Moreover, we identified 277 lysine acetylation sites in 133 mitochondrial proteins, thereby conclusively establishing that lysine acetylation is an abundant post-translational modification in the mitochondrion. These results clearly indicate that regulation of DNA-templated process is just one of many roles that lysine acetylation plays *in vivo*. The integration of these datasets offers a steppingstone toward understanding and characterizing functional consequences of lysine acetylation in diverse cellular regulatory schemes and in the pathophysiology of human diseases.

RESULTS

Specific Detection of Lysine-acetylated Proteins

To evaluate the specificity of anti-acetyllysine antibody, immunofluorescent staining of HeLa cells was performed by Dr. Tzuling Cheng in the laboratory of Dr. Michael White. The analysis revealed that the antibody produced strong positive signals in the nucleus and the midbody of the cells (Figure 3-1A). This result suggested that lysine acetylation was prevalent among these two subcellular regions. Western blotting analysis with nuclear and cytosolic protein extracts of HeLa cells resulted in many distinct protein bands spanning a wide mass range in both extracts (Figure 3-1B). In addition, these signals could be competed away efficiently in the presence of acetylated BSA at a concentration of 160 μg/mL (Figure 3-1C). Similarly, the antibody detected numerous proteins in mitochondrial fractions from either fed or fasted mice (Figure 3-1D, E). These results indicate that the antibody is specific and that many lysine-acetylated proteins from diverse subcellular compartments remain to be identified.

Immunoaffinity Purification and Identification of Lysine-acetylated Peptides

To identify lysine-acetylated peptides and map acetylation sites, agarose beads bearing immobilized anti-acetyllysine antibody were used to affinity purify peptides from a tryptic digest of a protein fraction of interest (Figure 3-2A). The

enriched peptides were analyzed by nano-HPLC/MS/MS in an LTQ mass spectrometer (Figure 3-2B, C). The MASCOT search algorithm used the resulting MS/MS spectra to identify peptide candidates from a human protein database. To ensure the quality of analysis, positive identifications were verified by manual inspection of the MS/MS spectra.

The analysis yielded 388 lysine acetylation sites in 195 proteins, including 67 lysine acetylation sites in 37 proteins from the cytosolic fraction, 81 lysine acetylation sites in 38 proteins from the nuclear fraction, and 277 lysine acetylation sites in 133 proteins from mitochondria (Figure 3-4, Tables 3-1 and 3-2). Only 13 of the 195 proteins were previously reported to be acetylated. Thus, this survey greatly expands the inventory of lysine-acetylated proteins.

Verification of Lysine-acetylated Peptides

To ascertain the accuracy of the proteomics data, three independent experimental approaches were taken. First, a fraction of the newly identified lysine-acetylated proteins was selected for confirmation by immunoprecipitation and Western blotting analysis (Figure 3-3A). The candidate proteins were selected based on commercial availability of antibodies against them. Second, a subset of lysine-acetylated peptides was confirmed by MS/MS spectra of synthetic peptides, the standard for verification of peptide identification (Figure 3-3B, C). Confirmation of lysine acetylation among all the candidate peptides, especially

those with low MASCOT scores, demonstrates that the peptide identification algorithm and manual verification procedure were reliable. As a third approach, the expertise of the laboratories of Drs. Michael White and Xiang-Jiao Yang was enlisted to analyze how lysine acetylation affects the functions of three novel lysine-acetylated proteins: p120 catenin, Rho GDI, and actin. The results indicate that the acetylation is important for regulating the function of each of these proteins as discussed below. Furthermore, identification of known acetylation sites on core histones, p300, and CBP offer further evidence regarding the validity of the datasets (Tables 3-1 and 3-2).

DISCUSSION

Lysine Acetylation Observed in Diverse Groups of Proteins

Lysine acetylation substrates with diverse cellular functions were identified, including the four core histones, regulators of transcription and chromatin structure, splicing and translational proteins, chaperones, cytoskeletal proteins, signaling proteins, and metabolic enzymes (Figure 3-4, Tables 3-1 and 3-2). Most notably, 133 mitochondrial proteins were found to be acetylated on lysine, conclusively demonstrating that lysine acetylation is an abundant post-translational modification in that organelle.

RNA splicing and translation factors

No reports have been made of lysine acetylation in splicing or translation factors. Acetylated peptides from hnRNP A1, an uncharacterized protein with a potential role in splicing, and two translation factors (EF1α and eIF-5A) were identified in this screen, suggesting that acetylation may play a role in regulating RNA splicing and translation. Discovery of acetyllysine residues in these splicing and translation factors suggests regulation of the two cellular processes by previously unappreciated processes that await further characterization.

Chaperones

This proteomics analysis identified lysine acetylation in chaperone proteins, including Hsp70, Hsp27, Hsp90, two subunits of the TCP1 ring complex TriC, cyclophilin A, and FK506-binding protein 4. Of these, only Hsp90 was previously known to be lysine-acetylated, but its acetylation sites have not been mapped. Acetylation of Hsp90 influences its ability to regulate the glucocorticoid receptor (Kovacs et al., 2005). Identification of the acetylation site should help further investigation on how the modification affects activities of the receptor. Because the acetylation site of Hsp70 is located at its substrate-binding region, the modification may affect substrate recognition.

Cyclophilin, a peptidyl-prolyl isomerase, binds the immunosuppressive drug cyclosporin A and modulates T-cell activation by inhibiting the phosphatase

calcineurin. Cyclophilin is acetylated at K125, a residue located in the loop surrounding the drug pocket and a potential site of ionic interaction with calcineurin. Mutation of K125 to either an anionic or neutral residue (K125E or K125Q) has minimal effects on isomerase activity and cyclosporin A binding, but renders the mutant protein unable to associate with calcineurin or inhibit its phosphatase activity (Etzkorn et al., 1994). Acetylation of K125 could affect the interaction of cyclophilin with calcineurin.

Structural proteins

Remarkably, a broad cadre of proteins participating in microfilament formation and dynamics were identified. Actin itself is acetylated at lysine 61, a residue immediately adjacent to a critical arginine required for actin polymerization (Posern et al., 2004). Acetylation was also observed within cofilin, thymosin beta 10, profilin, moesin, and tropomyosin. To begin to examine the consequences of lysine acetylation on actin dynamics, epithelial cells were incubated briefly in the presence of pharmacological HDAC antagonists or agonists. As shown in Figure 3-5, exposure to the class I/II HDAC inhibitor, tricostatin A, resulted in a dramatic accumulation of filamentous actin, suggesting stabilized actin stress fibers. In contrast, the agonist of class III HDACs, resveratrol, potently inhibited stress fiber formation and/or stability (Figure 3-5A).

The Rho-family GTPases are central regulators of actin-cytoskeletal dynamics. In addition to the structural proteins described above, RhoGDI and p120 catenin (p120), two proteins known to modulate Rho GTPase function, were identified. RhoGDI selectively inhibits Rac, Rho or CDC42 activation and membrane localization depending upon the phosphorylation of key regulatory residues in RhoGDI. In an attempt to examine the potential consequences of RhoGDI acetylation, mutational analysis was carried out by Dr. Tzuling Cheng of Dr. Michael White's lab. The lysine residue identified as being acetylated, K141, was replaced with arginine to block its acetylation, but retain the native charge status, or glutamine, which was intended to mimic acetylation of the site. Expression of RhoGDI in epithelial cells inhibited stress fiber accumulation as previously reported, and the K141R substitution did not grossly alter this behavior. In contrast, the replacement of the putative site of acetylation with glutamine resulted in the accumulation of both thickened stress fibers and numerous filopodia (Figure 3-5B). This result suggests that the neutralization of charge which would accompany the acetylation of K141 may inhibit RhoGDI function and/or alter its G-protein specificity. Furthermore, p120 catenin (p120), a protein associated with E-cadherin in cell-cell adhesions, is acetylated at three lysine residues, two of which are known to be important for inhibiting Rho GTPase activation thus promoting development of a dendritic morphology in fibroblasts (Anastasiadis et al., 2000). The three acetylatable lysine residues are

located in a 7-residue lysine-rich stretch that is disrupted by a 6-residue insert in a rare alternative-spliced isoform (Figure 3-5C). Mutational analysis of this protein was carried out in the laboratory of Dr. Xiang-Jiao Yang. Swapping the acetylated lysines with arginine modestly decreased p120-induced "dendrite" formation (Figure 3-5D). On the other hand, glutamine substitutions completely abrogated this phenotype. These results indicate that the acetylatable lysine residues are important for p120's functions in cells and suggest that acetylation may represent a new mechanism for regulation of its function.

In addition to actin and microfilament regulators, two intermediate filament proteins (lamins A and C) were also found acetylated. The observations that the pharmacological HDAC inhibitor and activator regulate cytoskeleton structure and that lysine acetylation is present among all the three classes of cytoskeletal proteins further suggest that this modification could represent a common mechanism for modulation of the cytoskeleton structures.

Cytosolic metabolic enzymes

No acetylation has been previously identified among metabolic enzymes in mammalian eukaryotes. In this study, five cytosolic metabolic enzymes were found to be lysine-acetylated: aldolase, enolase, triosephosphate isomerase 1, phosphoglycerate mutase 1, and transketolase. Lysine acetylation is likely to affect activities of some of these enzymes. One good candidate is the glycolytic

enzyme aldolase (or fructose 1,6-bisphosphate aldolase), which is acetylated at K12, K41, and K146. K146 is located at the active site of the enzyme. Mutation of the conserved K146 among orthologs dramatically decreased enzymatic activity, by a factor of over 500 for K146R and a factor of over 1×10^5 for K146Q (Morris and Tolan, 1994). Acetylation of K146 is likely to have an effect similar to that of the K146Q mutation. In addition to serving as an enzyme, aldolase has other cellular activities, including actin binding. The actin-binding residues are located around the catalytic pocket of the enzyme (Wang et al., 1996). Thus, in addition to modulating enzymatic activity, lysine acetylation of aldolase might also affect cytoplasmic structure of actin filaments or the assembly of protein complexes. Similar to cytosolic metabolic enzymes, many metabolic enzymes from the mitochondrion may also be subject to dynamic regulation by lysine acetylation.

Signaling proteins

This analysis revealed lysine acetylation in three signaling molecules: Rho GDI, annexin V, and phospholipase C beta 1. Phospholipase C beta 1 catalyzes a key reaction in intracellular signal transduction, generating two second messengers, inositol 1,4,5-trisphosphate and diacylglycerol, from phosphatidylinositol 4, 5-bisphosphate. The protein is activated by diverse signals through G-protein-coupled receptors and receptor tyrosine kinases. Annexin V is

a calcium-dependent phospholipid-binding protein that can inhibit phospholipase A2 (Buhl, 1992; Mira et al., 1997) and protein kinase C (Schlaepfer et al., 1992). The identification of lysine-acetylated signaling proteins implies that lysine acetylation might be involved in cross-talk between signal transduction pathways.

Chromatin and transcription regulators

Chromatin-binding proteins, such as high-mobility-group (HMG) chromosomal proteins are known to be acetylated. This analysis identified K30 of HMG-1 as a new acetylation site. Because K30 is located within the DNA-binding domain, acetylation at this site is likely to affect the DNA-binding affinity or DNA-bending ability of HMG-1. Among chromatin-modifying enzymes, histone acetyltransferases are known to be autoacetylated (Yang, 2004b). Consistent with this, acetylated peptides from p300, CBP, and MOZ were identified (Table 3-1). Autoacetylation of p300 is important for stimulating its acetyltransferase activity (Thompson et al., 2004). For p300, the four sites identified here correspond to the previously mapped autoacetylation sites (Thompson et al., 2004). These four lysine residues are conserved in CBP and were also found to be acetylated, providing further evidence of the high quality of the datasets.

Acetylation was also observed on other chromatin-modifying enzymes. First, RbAp48, a core subunit of the Sin3a and NuRD deacetylase complexes, is

acetylated at a lysine residue close to several WD40 repeats. This WD40-repeat protein is also a key subunit of three other complexes: chromatin assembly factor, HAT1 complex, and NuRD. Second, ING4 is acetylated at three lysine residues within a known nuclear localization signal (NLS). The ING family of proteins has growth-suppressing ability and associates with HAT or HDAC activity. Importantly, the NLS of ING4 is subject to frequent mutations in cancer cell lines (Kim et al., 2004). ING4 is part of a complex containing the acetyltransferase HBO1 and PHD finger protein 15 (also known as Jade2) (Doyon et al., 2006). Like ING4, Jade2 is acetylated. Third, two methyltransferases are acetylated. The histone methyltransferase MLL3 is acetylated at K1869, while the DNA methyltransferase DNMT1 is acetylated at three lysines on a GK repeat linker. These repeats link the methyltransferase domain to the N-terminal regulatory domain and may functionally resemble the N-terminal flexible tails of core histones. Finally, lysine acetylation is present in proteins involved in assembling, compacting, and remodeling chromatin. These acetylated proteins include SET (a histone-binding protein) and SMARCA2 (an SNF2 homolog). Together, these results support the notion that components of chromatin, as well as chromatin-utilizing proteins, are subjects of dynamic lysine acetylation.

The proteomics survey presented here also identified 22 acetylation sites in core histones, including 7 new and 15 known sites. The new sites are K11, K16, and K23 in H2B; K36 in H3; and K4, K7 and K11 in an H2A isoform (these

three lysines are not conserved in the canonical H2A sequence). Although functional roles of acetylation at these newly identified sites remain to be established, acetylation of K36 on H3 suggests a competition with methylation on the same residue. Methylation of K36 is a key modification event during transcriptional elongation (Margueron et al., 2005; Martin and Zhang, 2005). This potential interplay is evocative of competitive acetylation and methylation of K9 on H3.

A significant fraction of known lysine-acetylated proteins, especially transcription factors, were not identified in this experiment. This may have been caused by low expression levels of these proteins in HeLa cells (*e.g.*, p53) or by the inhibitory effects of abundant, lysine-acetylated proteins such as histones. An initial fractionation of the cell lysate with high resolution methods would help to alleviate the histone competition problem in the future.

Lysine Acetylation is Abundant in the Mitochondrion

The proteomic screen described here demonstrates that lysine acetylation is a common post-translational modification in the mitochondrion. Proteomics analysis identified 277 unique acetylation sites in 133 proteins from two pools of mouse liver mitochondria, one from fed mice and the other from fasted mice (Figure 3-4). To my knowledge, in none of these peptides and proteins had lysine acetylation been previously observed. Among the acetylated proteins, 62% were

identified in both fractions, 14% were specific to fed mice, and 24% were specific to fasted mice. Most lysine-acetylated proteins from mitochondrial fractions were metabolic enzymes (91 proteins).

The results suggest that ~20% of mitochondrial proteins are lysine-acetylated, if we assume that this analysis is as sensitive as previous proteomics studies that identified ~600 mitochondrial proteins (Mootha et al., 2003; Taylor et al., 2003). Identification of such a large number of lysine-acetylated mitochondrial proteins is significant and opens the door to exiting new research avenues for future study. For example, which lysine acetyltransferases and deacetylases are responsible for controlling mitochondrial acetylation status? The class III HDACs Sirt3, 4 and 5 have been localized to mitochondria by immunofluorescent methods (Michishita et al., 2005) and are good candidates for the regulatory deacetylases, but no mitochondrial acetyltransferase has been described. Also, it is unknown whether acetylation reactions really occur within the mitochondrion or whether they are part of a mitochondrial targeting and transport process.

Metabolic proteins

Of the lysine-acetylated mitochondrial proteins that are annotated with possible functions, over half are involved in some aspect of energy metabolism (Table 3-2). Lysine-acetylated proteins are present among all the major metabolic

pathways in mitochondria, including 6 TCA cycle proteins, 26 proteins involved in oxidative phosphorylation, 27 beta-oxidation or lipid metabolism proteins, 8 associated with amino acid metabolism, 10 with carbohydrate metabolism, 3 with nucleotide metabolism, and 2 with the urea cycle. Transporter and channel proteins also play an important role in regulating energy metabolism. In this study, 15 transporters or channel proteins were found to be lysine-acetylated. Although it remains unclear how lysine acetylation of metabolic enzymes regulates their functions, an attractive hypothesis is that the acetylation state of these proteins influences the mode or rate of energy production, or other mitochondrial functions.

Dehydrogenases

Notable among the acetylated metabolic enzymes are the dehydrogenases, a group of enzymes that oxidize a substrate by transferring hydrogen to an acceptor, usually NAD, NADP, or a flavin coenzyme. Forty-eight dehydrogenases were identified in previous proteomics studies of mitochondrial proteins (Mootha et al., 2003; Taylor et al., 2003). This proteomics screening found that at least 44% of mitochondrial dehydrogenases (21 protein complexes) were lysine-acetylated. Among them, 14 use NAD as the electron acceptor to catalyze the reactions in oxidative, catabolic routes.

Enzymatic activities of some of these proteins (*e.g*., the PDH complex and isocitrate dehydrogenase) are modulated by the cellular energy charge (*e.g*., NAD^+/NADH ratio or acetyl-CoA level). On the other hand, acetyl-CoA is used as a substrate for HATs, while NAD is a cofactor for class III HDACs. NAD concentration has been suggested as a means to modulate the enzymatic activities of class III HDACs (Blander and Guarente, 2004). Thus, it is tempting to propose that lysine acetylation serves as a feedback mechanism for regulation of dehydrogenase activities in mitochondria, in response to cellular energy charge.

Acetylated mitochondrial dehydrogenases include dihydrolipoyl dehydrogenase, nine subunits of NADH dehydrogenase (oxidative phosphorylation complex I), four TCA cycle-related dehydrogenases, (fumarate hydratase, succinate dehydrogenase, isocitrate dehydrogenase, and malate dehydrogenase), five proteins involved in beta-oxidation, and 10 proteins involved in other metabolic reactions. For most of these enzymes, aberrant changes of their expression or functions lead to one of several metabolic diseases, suggesting a potential role of lysine acetylation in these diseases.

Longevity-related mitochondrial proteins are lysine acetylated

Proteomics analysis identified acetyllysine in 16 proteins believed to influence longevity (Table 3-3), including three oxidative phosphorylation proteins (ATP synthase g subunit, NADH dehydrogenase, NADH-ubiquinone

oxidoreductase), two proteins in the TCA cycle (fumarate hydratase, isocitrate dehydrogenase), and two antioxidant enzymes (superoxide dismutase 1, superoxide dismutase 2). In addition to the mitochondrial proteins, four lysine-acetylated proteins from cytosolic or nuclear fractions are also involved in the regulation of aging. These proteins include lamin A/C (an alternately spliced nuclear matrix protein), profilin 1 (an actin-binding protein), phosphoglycerate mutase (a glycolytic protein), and a eukaryotic initiation factor (elF-5A). The roles of these proteins in aging were established by either genetics analysis or RNAi screening (Browner et al., 2004; Hamilton et al., 2005).

Regulatory roles of Sir2, Sir2 activators (*e.g.*, resveratrol), and HDAC inhibitors (*e.g.*, phenylbutyrate) in aging or longevity pathways imply a link between lysine acetylation and aging. It remains unclear, however, how this modification affects aging and longevity. The lysine-acetylated proteins related to aging and longevity identified in this study provides important leads for further characterization of the roles of lysine acetylation in aging.

Implications for the Mechanistic Impact of Lysine Acetylation

As discussed above, analysis of the identified acetylation sites indicates that this modification exerts its impact via alteration of protein-protein interactions, nuclear localization, enzymatic activity, and interplay with other lysine modifications (*e.g.*, methylation and sumoylation). One striking feature of

the group of acetylated proteins identified here is that many are metabolic enzymes, which implies that alteration of enzymatic activity may emerge as a major consequence of acetylation. Lysine acetylation was also observed to occur in known or putative NLS sequences. Therefore, change of subcellular localization may be another effect of this modification. In addition to acetylation, a lysine residue can be subject to other post-translational modifications, such as methylation, ubiquitination, and sumoylation. On a single lysine residue, these modifications are mutually exclusive. Acetylation sites in 11 proteins conform to the sumoylation consensus sequence, suggesting that acetylation may compete with sumoylation. Consistent with this hypothesis, acetylation of p300 is known to inhibit its sumoylation (Bouras et al., 2005). However, unlike sumoylation, consensus sequences for methylation and ubiquitination have not been defined. Further studies of the acetylated proteins identified here will certainly reveal interplay of acetylation with these two modifications, as has been nicely shown for histones and the p53 tumor suppressor. Thus, the large datasets compiled here (Tables 3-1 and 3-2) present a golden opportunity for further study of interplay among different lysine modifications.

Analysis of Lysine Acetylation Sites

Identification of the large set of acetylation sites here made possible, for the first time, a preliminary analysis of motif preference (Figure 3-6). The residue

preference for acetylated peptides from non-histone S100 proteins is asparagine at the -1 position, while histidine at the +1 position is favored for non-histone proteins from nuclear extract. For histone proteins, a propensity for lysine or acetylated lysine at the -4 or +4 position was observed. Lysine-acetylated peptides from mitochondria show different preferences in their flanking sequences, namely histidine or tyrosine at the +1 position (Figure 3-6).

The differences in preference of amino acid residues flanking acetyllysine residues suggest that the modification might be catalyzed by a subset of acetyltransferases with substrate preferences unique to the subcellular organelles. It is highly likely that linear lysine acetylation motifs exist among lysine-acetylated proteins similar to motifs observed for phosphorylation.

Local Structural Properties of Acetylated Lysines

Acetylated lysine shows a different preference for secondary structure than non-acetylated lysine. Acetylated lysines are found in a helical conformation about 9 percent more frequently than the average lysine, and in a coil about 8 percent less frequently (both comparisons are statistically significant with $P<0.03$). Thus acetylation appears to be favored in the context of specific secondary structural characteristics. Acetylated lysines have an average relative side-chain solvent accessibility of 61.5, which is higher than that of all lysines (50.5), suggesting that acetylated lysines prefer more exposed environments. Of

all the lysines in these proteins, about 14 percent are predicted to be in disordered regions (at a false positive rate threshold of 5%). In contrast, only 7.4 percent of acetyllysines are predicted to be in disordered regions. This result is consistent with the secondary structure type analysis. Acetylation sites appear to have local structural preferences that are quite different from those of protein phosphorylation sites; phosphorylation sites have a strong correlation with disordered regions (Iakoucheva et al., 2004).

CONCLUDING REMARKS

Non-histone targets of acetylation have only been identified relatively recently. This group is rapidly expanding, however, with the advent of techniques and analyses targeted specifically to their identification. As demonstrated in this chapter, lysine acetylation has the potential to impact a diverse range of biological processes, especially with regard to cellular metabolism. Emerging evidence also suggests that non-histone acetylation substrates represent a rich and unexploited source for potential drug targets. However, the identities and characterization of the relevant targets await further study. In particular, proteomic approaches can make a significant contribution to our understanding of the molecular mechanisms through which certain HDAC inhibitors exhibit apparent selective cytotoxicity toward tumor cells. For example, combining stable isotope labeling strategies,

such as stable isotope labeling with amino acids in cell culture (SILAC), with the proteomic analysis of acetylated proteins outlined in this chapter, can provide a wealth of quantitative information regarding the changes in protein expression and modification patterns that accompany treatment with HDAC inhibitors. The results of such a study could provide valuable insight into the mechanisms of cytotoxicity and inform next-generation interventions.

METHODS

Cell culture, transfection, and immunocytochemisty. Suspension HeLa S3 cells were grown in DMEM supplemented with 5% newborn calf serum, non-essential amino acids, 100 units/ml penicillin, and 100 μg/ml streptomycin in 2 L round-bottom side-arm spinner flasks. Cells were treated with 0.3 μM tricostatin A (TSA; class I/II HDAC inhibitor) and 50 μM of sirtinol (class III HDAC inhibitor) at a density of 5 x 10^5 cells/ml for 24 hr before harvesting, for preparation of cytosolic and nuclear extracts.

HeLa adhesion cells and NIH 3T3 cells were grown in DMEM supplemented with 10% fetal bovine serum and 10% newborn calf serum, respectively. Transfections were carried out using Superfect transfection reagent for NIH 3T3 cells and Lipofectamine 2000 for HeLa cells.

For subcellular localization of acetyllysine proteins, HeLa cells were fixed and stained with anti-acetyllysine polyclonal antibody (0.65 mg/ml) using a 1:50 dilution with or without competition with 50 μg of acetylated-BSA. To assess the effects of acetylation on stress fiber formation, HeLa cells were treated with 1 μM TSA, or 0.2 mM resveratrol (class III HDAC agonist) or left untreated for 2 hrs. Rhodamine-conjugated phalloidin was used to visualize filamentous actin.

Preparation of cytosolic and nuclear extract from HeLa cells and mitochondrial fraction from mouse livers. Cytosolic (S-100) and nuclear extracts from HeLa S3 cells were prepared as previously described (Dignam et al., 1983). Cells treated with TSA and sirtinol were collected and washed with cold PBS. The cell pellet was resuspended in 10 packed-cell-pellet volumes of ice-cold hypotonic buffer (10 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-Cl, pH 7.4) with 10 mM 2-mercaptoethanol and 0.2 mM PMSF, incubated at 4 °C for 10 min. The cells were collected by centrifugation at 1000 x g for 10 min. The swollen cells were resuspended again in 1/4 packed-cell-pellet volumes of the hypotonic buffer, incubated on ice for additional 10 min before homogenizing with 15 strokes using a Dounce homogenizer (B type pestle). The lysate was subjected to centrifugation at 4,000 x g for 15 min at 4 °C. The supernatant was carefully collected and subjected to ultracentrifugation at 100,000 × *g* for 1 h at 4 °C. The supernatant was collected and considered as the cytosolic extract (S-100). The nuclei fraction

was resuspended in 1/4 packed pellet volumes of low-salt buffer (20mM Tris-Cl, pH 7.4, 25% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 20 mM KCl) and homogenized using 15 strokes with a Dounce homogenizer (B type pestle). The resulting suspension was added with 1/2 volume of the high salt buffer (20mM Tris-Cl, pH 7.4, 25% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 1.2 M KCl) at a rate of 1 drop per second and incubated at 4 °C for 30 min. The suspension solution was centrifuged at 14,000 x g for 30 min. The supernatant was collected and considered the nuclear extract.

The mouse mitochondrial protein lysates were prepared by a procedure reported previously (Luo et al., 1998). Mouse were either fed at normal diet or starved for 12 hr. The mouse livers were removed after sacrifice and Dounce-homogenized in ice-cold mitochondria isolation buffer (MIB) containing 250 mM mannitol, 0.5 mM EGTA, 5 mM HEPES, and 0.1% (w/v) BSA (pH 7.2) supplemented with the protease inhibitors leupeptin (1 mg/ml), pepstatin A (1 mg/ml), antipain (50 mg/ml), and PMSF (0.1 mM). Unbroken cells and nuclei were pelleted by centrifugation at 600 x g for 5 min at 4 °C. The supernatants were further centrifuged at 10,000 x g for 10 min at 4 °C to pellet the mitochondria. The mitochondria pellet was resuspended in 4 ml MIB and loaded onto a continuous Percoll gradient consisted of 30% v/v) Percoll, 225 mM mannitol, 25 mM HEPES, 0.5 mM EGTA, and 0.1% (w/v) BSA (pH 7.2). The suspension/gradient was centrifuged at 40,000 x g for 1 hr. The mitochondria

were removed from the brownish band at 1.10 g/ml with a transfer pipette. The mitochondrial pellets were washed with MIB by centrifuging for 10 min at 6300 x g at 4 °C. The mitochondria were then resuspended gently in mitochondria resuspension buffer containing 400 mM mannitol, 10 mM KH_2PO_4, and 50 mM Tris-Cl (pH 7.2) with 5 mg/ml BSA and stored on ice.

In-solution tryptic digestion. Ten milligrams of proteins of interest were precipitated with acetone, followed by centrifugation at 22,000 x *g* for 10 min. The resulting pellet was rinsed twice with cold acetone to remove residual salts, resuspended in 50 mM NH_4HCO_3 (pH 8.5), and digested with trypsin at an enzyme-to-substrate ratio of 1:50 for 16 h at 37°C. The tryptic peptides were reduced with 5 mM DTT at 50°C for 30 min and then alkylated using 15 mM iodoacetamide at RT for 30 min in darkness. The reaction was quenched with 15 mM cysteine at RT for 30 min. To ensure complete digestion, additional trypsin at an enzyme-to-substrate ratio of 1:100 was added to the peptide mixture and incubated for an additional 3 h. The peptides were dried in a SpeedVac.

Affinity purification of lysine-acetylated peptides. The affinity-purified anti-acetyllysine antibody (ImmunoChem Pharmaceuticals Inc., Burnaby, British Columbia, Canada) was immobilized to protein A-conjugated agarose beads at 3-4 mg/ml by incubation at 4°C for 4 h. The supernatant was removed and the beads

were washed three times with NETN buffer (50 mM Tris-Cl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% NP40).

The tryptic peptides obtained above were resolubilized in NETN buffer. Insoluble particles were removed by centrifugation at 10,000 x *g* for 5 min. Affinity purification was carried out by incubating the peptides with 20 μl of anti-acetyllysine antibody-protein A-immobilized agarose beads at 4°C for 6 h with gentle shaking. The beads were washed three times with 1 ml of NETN buffer and twice with ETN (50 mM Tris-Cl, pH 8.0, 100 mM NaCl, 1 mM EDTA). The bound peptides were eluted from the beads by washing three times with 50 μl of 0.1% TFA. The eluates were combined and dried in a SpeedVac. The resulting peptides were cleaned in μC18 ZipTips (Millipore, Bedford, MA) according to the manufacturer's instructions, prior to nano-HPLC/mass spectrometric analysis.

Protein identification by Nano-HPLC-MS/MS and manual verification of peptide identification. The peptides from immunoaffinity purification were cleaned with ZipTip μ-C_{18} (Millipore, Bedford, MA) prior to nano-HPLC/tandem mass spectrometry analysis. Resulting peptides were resolubilized in buffer A solution (2% acetonitrile/ 97.9% water/ 0.1% acetic acid (v/v/v)). Nano-HPLC/tandem mass spectrometry analysis was performed in LTQ 2-D ion-trap mass spectrometer (ThermoFinnigan, San Jose, CA) equipped with a nano-electrospray ionization source. The source was coupled online to an Agilent

1100 series nano flow LC system (Agilent, Palo Alto, CA). Two µl of the peptide solution in buffer A was manually loaded onto a capillary HPLC column (10 cm length X 75 µm ID, 5 µm particle size, 300 Å pore diameter) packed in-house with Luna C18 resin (Phenomenex, St. Torrance, CA). The peptides were eluted from the column with a gradient of 5% to 90% buffer B (90% acetonitrile/ 9.9% water/ 0.1% acetic acid (v/v/v)) in buffer A over 30 min. The eluted peptides were electrosprayed directly into the LTQ 2-D ion trap mass spectrometer. The MS/MS spectra were acquired in a data-dependent mode that determined the masses of the parent ions and fragments of the ten strongest ions.

Protein sequence database search. Tandem mass spectra were searched against NCBI-nr database with MASCOT search engine (Matrix Science, London, UK) allowing for the potential incorporation of 42 m/z onto lysine residues, indicative of an acetylation site. Thus, both non-acetylated lysine and acetylated lysine will be considered during the protein sequence database search. Enzyme was specified as trypsin with 8 missing cleavages. Mass error for parent ion mass was set as +/- 3.5 Da and for fragment ion as +/- 0.6 Da. Spectra with +1, +2, +3 charge states were considered. If more than one spectrum were assigned to one peptide, each spectrum was given a Mascot score and only the spectrum with the highest score was used for fragmentation analysis. Peptides identified with a Mascot score higher than 25 were considered as potential

positive identification and each of them was manually verified by the method specified as following.

Manual verification of protein identification. The raw spectrum of each acetylated peptide was printed; a-, b-, y-type ions, their corresponding water/amine loss ions, and charge status were manually assigned. Strict manual analysis was applied to validate protein identification results using a procedure similar to the one previously described.(Chen et al., 2005) The following criteria were used for manual verification. Only a, b, and y ions as well as their water loss or amine loss peaks were considered. All the major isotope-resolved peaks should match fragment masses of the identified peptide. The isotope-resolved peaks were emphasized because a single peak could come from electronic spark and is less likely to be relevant to peptide fragments. The major isotope-resolved peaks are defined as (1) those isotope-resolved daughter ions with m/z higher than parent m/z and intensity higher than 5% of the maximum intensity or (2) those isotopically resolved peaks with intensities higher than 20% of the maximum intensity and m/*z* values between one-third of the parent m/*z* and the parent m/*z*. Typically >7 isotope-resolved peaks were matched to theoretical masses of the peptide fragments.

Analysis of local structural properties of acetylated lysines. To access the structural features of acetylated lysine residues in proteins, we searched a database of proteins with known structures using BLAST for the 136 mitochondrial proteins (2 histone proteins ignored) with a total of 270 experimentally characterized acetylated lysines. Eighty eight proteins have at least one close homolog in the Protein Data Bank (PDB) database (e-value < 0.001). For these proteins, we mapped the positions of acetylated lysine residues to the model structure using the BLAST local alignment. The secondary structure types were determined for positions with acetylated lysines using the program DSSP (Kabsch and Sander, 1983) (H and G are considered to be helix conformation, E and B to be strand conformation, and the rest of the letters to be coil conformation). The fraction of each secondary structure type of acetylated lysines was compared to that of all the lysine residues. The relative sidechain solvent accessibilities were also calculated using program NACESS (Hubbard et al., 1991). Predictions of disordered regions for the 136 mitochondrial proteins were made using the program DISOPRED2 (Jones and Ward, 2003).

Figure 3-1

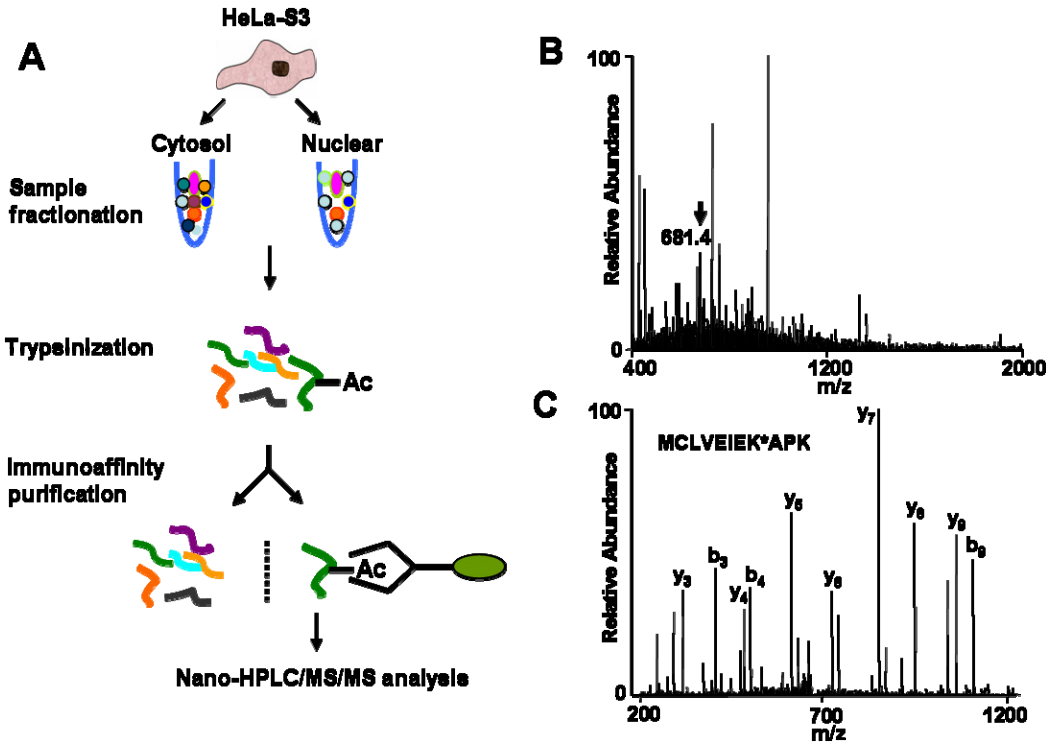

Figure 3-1. **Strategy for proteomics analysis of lysine acetylated proteins**. (A) Cytosolic and nuclear protein fractions from HeLa cells were digested with trypsin. The resulting peptides were subjected to immunoaffinity purification using an anti-acetyllysine antibody. The isolated peptides were analyzed by HPLC/MS/MS in an LTQ 2-D ion-trap mass spectrometer for peptide identification. (B) An example MS spectrum, with a peptide targeted for MS/MS analysis marked by an arrow. (C) MS/MS analysis of the peptide marked in Panel B led to identification of an acetylated peptide with sequence MCLVEIE**K***APK (**K*** indicates an acetylated lysine), unique to the NADH dehydrogenase 75 kDa subunit. The labels *b* and *y* designate the N- and C-terminal fragments, respectively, of the peptide produced by breakage at the peptide bond in the mass spectrometer. The number represents the number of N- or C-terminal residues present in the peptide fragment.

Figure 3-2

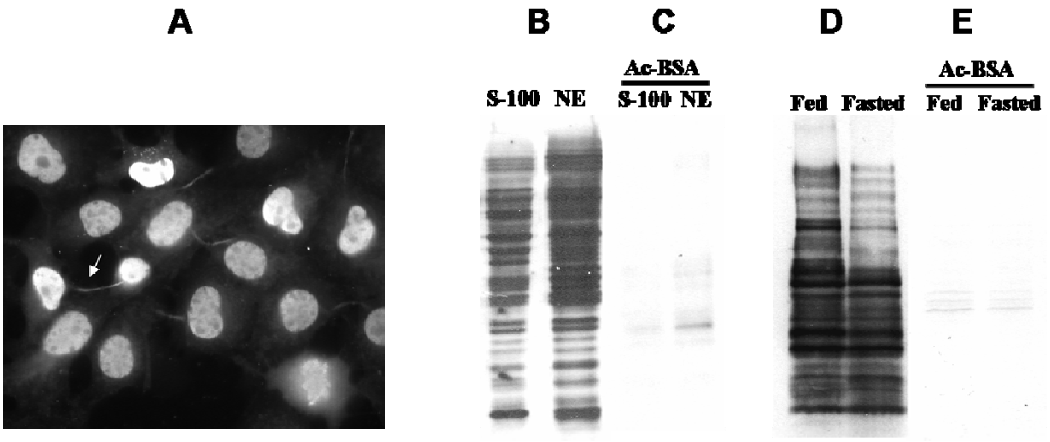

Figure 3-2. **Detection of lysine-acetylated proteins using anti-acetyllysine antibody**. (A) Immunofluorescent image of HeLa cells using anti-acetyllysine antibody, showing strong lysine acetylation in nuclear and midbody localization of acetylated proteins (indicated by an arrow). (B, C) Western blotting analysis of two protein fractions (S-100 = cytosolic, NE = nuclear) without (B) or with (C) competition with lysine-acetylated BSA. (D, E) Western blotting analysis of two mitochondrial fractions from livers of either fed or fasted mice, without (D) or with (E) competition with lysine-acetylated BSA. BSA was included at 5% by weight with an additional 150 μg BSA in buffers for B and D or 150 μg acetylated BSA for C and E. Immunofluorescence microscopy was carried out by Dr. Tzuling Cheng of Dr. Michael White's laboratory. Western blotting was performed by Dr. Sung Chan Kim.

Figure 3-3

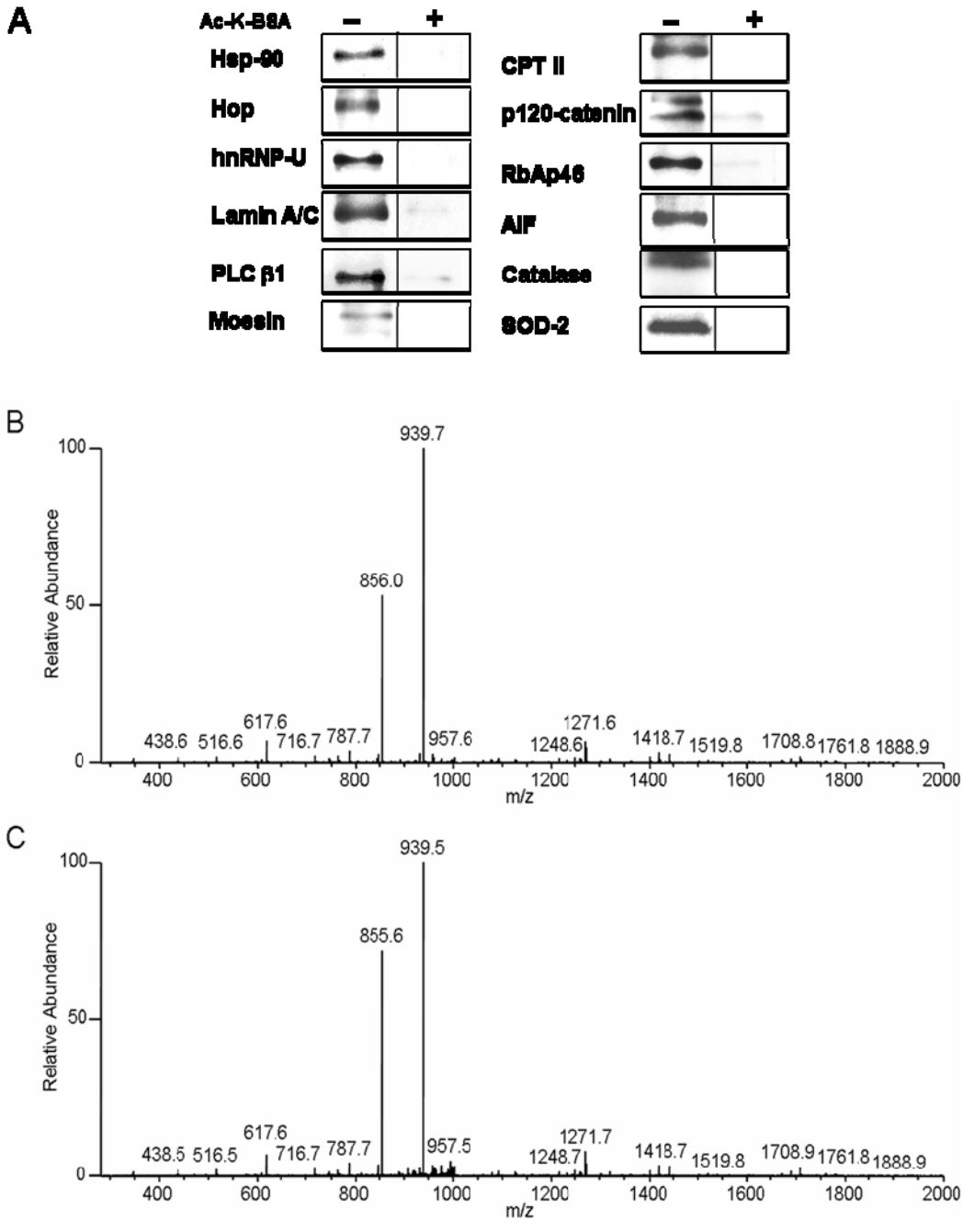
A
Ac-K-BSA
–
+
Hsp-90
Hop
hnRNP-U
Lamin A/C
PLC β1
Moesin
CPT II
p120-catenin
RbAp46
AIF
Catalase
SOD-2
B
939.7
856.0
617.6
438.6
516.6
716.7
787.7
957.6
1271.6
1248.6
1418.7
1519.8
1708.8
1761.8
1888.9
Relative Abundance
m/z
C
939.5
855.6
617.6
438.5
516.5
716.7
787.7
957.5
1271.7
1248.7
1418.7
1519.8
1708.9
1761.8
1888.9
Relative Abundance
m/z

Figure 3-3. Experimental verification of lysine-acetylated peptides by MS/MS of synthetic peptides and reciprocal immunoprecipitation. (A) Verification of lysine acetylation by Western blotting analysis. The protein of interest was immunoprecipitated with an antibody and probed with an anti-acetyllysine antibody by Western blotting analysis, with or without lysine-acetylated BSA (Ac-K-BSA). (B) Tandem mass spectrum of a tryptic peptide from HeLa nuclear extracts, which resulted in identification of acetylated peptide SAPAP**K*K***GS**K*K***AVT**K***AQK (**K*** indicates an acetylated lysine) from histone 2B. (C) Tandem mass spectrum of the synthetic peptide corresponding to the sequence identified in Panel B, showing similar ion intensity.

Figure 3-4

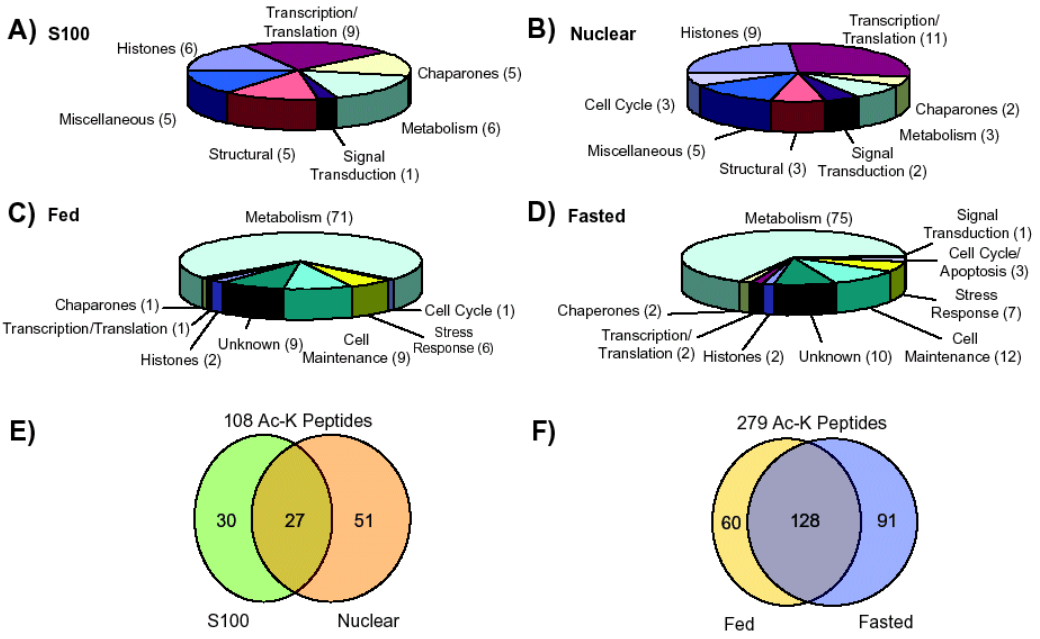

Figure 3-4. Summary of lysine-acetylated peptides and proteins. Functional classification of lysine-acetylated proteins from HeLa cytosolic (S-100) fraction (A), HeLa nuclear fraction (B), and mitochondrial fractions from either fed (C) or fasted (D) mouse liver, with the number of acetylated proteins in each functional group. Venn diagrams show the number of lysine-acetylated peptides unique to and common between S100 and nuclear fractions (E) or fed mouse and fasted mouse liver mitochondria (F).

Figure 3-5

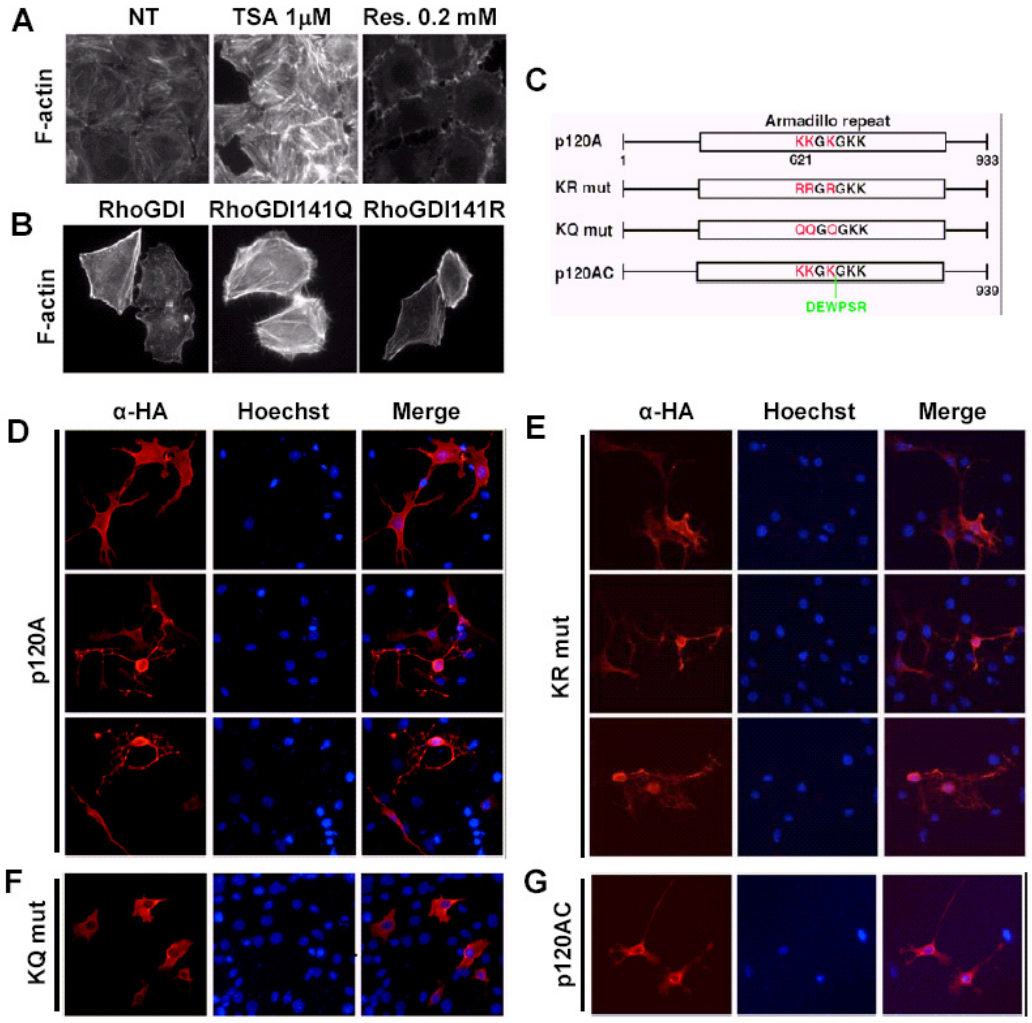
A
NT
TSA 1μM
Res. 0.2 mM
F-actin
B
RhoGDI
RhoGDI141Q
RhoGDI141R
F-actin
C
Armadillo repeat
p120A
KKGKGKK
1
621
933
KR mut
RRGRGKK
KQ mut
QQGQGKK
p120AC
KKGKGKK
939
DEWPSR
D
α-HA
Hoechst
Merge
p120A
E
α-HA
Hoechst
Merge
KR mut
F
KQ mut
G
p120AC

Figure 3-5. Acetylatable lysine residues of Rho GDI and p120 catenin impact actin/cytoskeleton dynamics. (A) HeLa cells were treated for 2 hours with the indicated compounds. Filamentous actin was visualized with rhodamine-conjugated phalloidin (NT- no treatment; TSA- tricostatin A; res- resveratrol). (B) HeLa cells transiently expressing the indicated RhoGDI variants were stained with rhodamine-phalloidin to visualize F-actin as in (A). Myc epitope-tagged RhoGDI expression was verified by immunostaining (not shown). (C) Schematic representation of p120 isoforms and point mutants. Rectangles represent Armadillo-repeat domains. The sequence of residues 621-627, a lysine-rich stretch located between repeats 6 and 7, is shown with the three acetylatable residues shown in red. The p120 gene contains three exons (A, B and C) for alternative splicing. p120A, a major isoform, contains a peptide encoded by exon A, whereas p120AC possesses this as well as a hexapeptide, encoded by the rare C exon (green) (Anastasiadis et al., 2000) inserted after K624. The KR mutant (KR mut) contains arginine substitution of the three acetylatable lysines, and in the KQ mutant, these residues are replaced with glutamine. (D-G) NIH3T3 fibroblasts were transfected with expression plasmids for HA-tagged isoforms and point mutants of p120 as indicated. Transfected cells were subjected to indirect fluorescence microscopy by immunostaining with anti-HA antibody and Cy3-labelled secondary antibody. The images for p120 subcellular localization (α-HA staining) and nuclei (Hoechst staining), as well as their merged images are shown. Unlike p120 AC and the KQ mutant, p120A and the KR mutant display three different types of subcellular distribution whose representative images are shown. Mutational analysis and microscopy for RhoGDI characterization carried out by the laboratory of Dr. Michael White. Mutational analysis and microscopy for p120 catenin characterization carried out by the laboratory of Dr. Xiang-Jiao Yang.

Table 3-1

Protein Name	gi	Peptide Sequence	No. Sites	Known Acetylated Proteins	Nuclear Extract	S100	Biological Process	Molecular Function
Histones								
histone H2A	gi|20357599	AGGK*AGK*DSGK	2 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H2A	gi|20357599	AGGK*AGK*DSGK*AK	3 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H2A	gi|20357599	AGGK*AGK*DSGK*AK*TK	4 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H2A	gi|20357599	AGK*DSGK	Acetyl (K)	*		*	Chromosome Organization	DNA Binding
histone H2A	gi|20357599	AGK*DSGK*AK	2 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H2A	gi|20357599	AGK*DSGK*AK*TK	3 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H2A	gi|20357599	DSGK*AK	Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H2A	gi|48257076	AGK*DSGK*AK*AK	4 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	K*GSK*K*AVTK*AQK	4 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H2B	gi|22770651	K*GSK*K*AVTK*AQK*K	5 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	KGSK*K*AVTK*AQK	3 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	LLLPGELAK*HAVSEGTK	Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	PDPAK*SAPAPK*K*GSK	3 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	SAPAPK*K*GSK*K	3 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	SAPAPK*K*GSK*K*AVTK	4 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	SAPAPK*K*GSK*K*AVTK*AQK	5 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	SAPAPK*K*GSK*K*AVTK*AQK*K	6 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	AVTK*AQK*K	2 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	GSK*K*AVTK	2 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	GSK*K*AVTK*AQK	3 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H2B	gi|22770651	K*AVTK*AQK	2 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	K*AVTK*AQK*K	3 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|22770651	K*GSK*K*AVTK	3 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|55960983	K*GSK*K*AVTK*VQK	4 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|55960983	AVTK*VQK	Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|55960983	K*AVTK*VQK	2 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H2B	gi|7768756	PEPAK*SAPAPK*K	2 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H3	gi|45768281	K*QLATK*AAR	2 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H3	gi|45768281	K*SAPATGGVK	Acetyl (K)	*			Chromosome Organization	DNA Binding
histone H3	gi|45768281	K*SAPSTGGVK*KPHR	2 Acetyl (K)	*		*	Chromosome Organization	DNA Binding
histone H3	gi|45768281	K*STGGK*APR	2 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H3	gi|45768281	STGGK*APR	Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H3	gi|386772	K*SAPATGGVK*KPHR	2 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H3	gi|45768840	GGK*APR	Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H3	gi|51959376	TKQTALKSTGGK*APR	Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H4	gi|21954750	GGK*GLGK*GGAK*R	3 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H4	gi|21954750	GK*GGK*GLGK	2 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H4	gi|21954750	GK*GGK*GLGK*GGAK	3 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H4	gi|21954750	GK*GGK*GLGK*GGAK*R	4 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H4	gi|21954750	GLGK*GGAK	Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H4	gi|21954750	GLGK*GGAK*R	2 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H4	gi|21954750	GGK*GLGK	Acetyl (K)	*		*	Chromosome Organization	DNA Binding
histone H4	gi|21954750	GGK*GLGK*GGAK	2 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H4	gi|22724936	AGK*GLGK*GGAK	2 Acetyl (K)	*	*		Chromosome Organization	DNA Binding
histone H4	gi|22724936	GK*AGK*GLGK*GGAK	3 Acetyl (K)	*	*	*	Chromosome Organization	DNA Binding
histone H4	gi|22724936	SVRGK*AGK*GLGK*GGAK	3 Acetyl (K)	*	*		Chromosome Organization	DNA Binding

Protein Name	gi	Peptide Sequence	No. Sites	Known Acetylated Proteins	Nuclear Extract	S100	Biological Process	Molecular Function
Transcription/ Translation								
DNA (cytosine-5-)methyltransferase 1 variant	gi|62088406	GK*GK*GK*GK	3 Acetyl (K)			*		
SET translocation (myeloid leukemia-associated)	gi|4506891	K*ELNSNHDGADETSEK	Acetyl (K)			*	Nucleosome Assembly	Phosphatase Inhibitor
hnRNP A1 isoform b	gi|14043070	SSGPYGGGGQYFAK*PR	Acetyl (K)		*		RNA Processing	RNA Binding
p29ING4	gi|18873723	QIESSDYDSSSSK*GK*K	2 Acetyl (K)		*	*	Transcription	DNA Binding
p29ING4	gi|18873723	SK*GK*NSDEEAPK	2 Acetyl (K)		*		Transcription	DNA Binding
SON DNA binding protein	gi|5737751	LTDLDK*AQLLEIAK	Acetyl (K)		*		Transcription	DNA Binding
HMG-1	gi|968888	K*HPDASVNFSEFSK	Acetyl (K)	*	*		Transcription	DNA Binding
MLL3	gi|10864041	TLVLSDK*HSPQK	Acetyl (K)		*		Transcription	Methyltransferase
E1A-associated protein p300	gi|3024341	NNK*K*TSK*NK*SSLSR	4 Acetyl (K)	*	*	*	Transcription	Histone Acetyltransferase
E1A-associated protein p300	gi|3024341	NNK*K*TSK*NK	3 Acetyl (K)	*	*		Transcription	Histone Acetyltransferase
MYST histone acetyltransferase 3	gi|5803098	ELEISVGK*SVSHENK	Acetyl (K)		*		Transcription	Transcription Regulator
SMARCA2 (SWI/SNF family member)	gi|55958983	K*GK*GK*KR	3 Acetyl (K)		*	*	Transcription	Transcription Coactivator
CREB-binding protein	gi|631119	K*TNK*NK*SSISR	3 Acetyl (K)	*	*	*	Transcription	Transcription Coactivator
CREB-binding protein	gi|631119	K*NNK*K*TNK*NK	4 Acetyl (K)	*	*		Transcription	Transcription Coactivator
CREB-binding protein	gi|631119	KK*NNK*K*TNK*NK*SSISR	5 Acetyl (K)	*		*	Transcription	Transcription Coactivator
CREB-binding protein	gi|631119	NNK*K*TNK*NK	3 Acetyl (K)	*		*	Transcription	Transcription Coactivator
CREB-binding protein	gi|631119	NNK*K*TNK*NK*SSISR	4 Acetyl (K)	*	*	*	Transcription	Transcription Coactivator
PHD finger protein 15	gi|18676594	CSK*LPSSTK*SGWPR	2 Acetyl (K)		*		Transcription	Transcription Coactivator
unactive progesterone receptor	gi|12804293	DVNVNFEK*SK	Acetyl (K)			*	Transcription	Transcription Regulator
EF1 alpha	gi|62896589	NVSVK*DVR	Acetyl (K)			*	Translation	Elongation Factor
eIF-5A	gi|42659529	K*YEDICPSTHNMDVPNIK	Acetyl (K)		*		Translation	Translational Initiation Factor
Chaparones								
heat shock 27 kDa protein 1	gi|4504517	DGVVEITGK*HEER	Acetyl (K)			*	Protein Folding	Chaparone
chaparonin containing TCP 1, subunit 8 (theta)	gi|1136741	EGAK*HFSGLEEAVYR	Acetyl (K)			*	Protein Folding	Chaparone
heat shock 90 kDa protein 1 beta	gi|34304590	K*HLEINPDHPIVETLR	Acetyl (K)		*	*	Protein Folding	Chaparone
t-complex 1	gi|30582777	SLHDALCVVK*R	Acetyl (K)			*	Protein Folding	Chaparone
Cyclophilin A (FKBP)	gi|10863927	TEWLDGK*HVVFGK	Acetyl (K)		*		Protein Folding	Chaparone
FK506 binding protein 4	gi|4503729	LEQSTIVK*ER	Acetyl (K)			*	Protein Folding	Chaparone
Hop	gi|54696884	DFDTALK*HYDK	Acetyl (K)			*	Protein Folding	Chaparone
Structural								
cofilin 1 (non-muscle)	gi|30582531	LTGIK*HELQANCYEEVK	Acetyl (K)			*	Structural	Actin Binding
thymosin beta 10	gi|16876901	ADK*PDMGEIASFDK	Acetyl (K)		*	*	Structural	Actin Binding
profilin I	gi|30582841	STGGAPTFNVTVTK*TDK	Acetyl (K)			*	Structural	Actin Binding
tropomyosin 3	gi|55665780	CTK*EEHLCTQR	Acetyl (K)			*	Structural	Actin Binding
actin, gamma 1	gi|16924319	DSYVGDEAQSK*R	Acetyl (K)			*	Structural	
lamin A/C	gi|21619981	ASSHSSQTQGGGSVTK*K	Acetyl (K)		*		Structural	
moesin	gi|4505257	K*VTAQDVR	Acetyl (K)		*		Structural	
Signal Transduction								
Annexin V	gi|12655149	HALK*GAGTNEK	Acetyl (K)			*	Signal Transduction	Calcium Channel
Rho GDI alpha	gi|30582607	IDK*TDYMVGSYGPR	Acetyl (K)		*		Signal Transduction	
phospholipase C beta 1	gi|9438229	K*DSKK*K	2 Acetyl (K)		*		Signal Transduction	Liberates inositol 1,4,5-triphosphate
Metabolism								
triosephosphate isomerase 1	gi|16877874	TATPQQAQEVHEK*LR	Acetyl (K)			*	Carbohydrate Metabolism	Glycolysis, NADPH Metabolism
alpha enolase	gi|693933	AVEHINK*TIAPALVSK	Acetyl (K)			*	Carbohydrate Metabolism	Glycolysis
aldolase A protein	gi|28595	GILAADESTGSIAK*R	Acetyl (K)			*	Carbohydrate Metabolism	Glycolysis
aldolase A protein	gi|28595	PYQYPALTPEQK*K	Acetyl (K)			*	Carbohydrate Metabolism	Glycolysis
aldolase A	gi|28614	DGADFAK*WR	Acetyl (K)		*		Carbohydrate Metabolism	Glycolysis
transketolase (Wernicke-Korsakoff syndrome)	gi|31417921	AFGQAK*HQPTAIIAK	Acetyl (K)		*	*	Carbohydrate Metabolism	
transketolase (Wernicke-Korsakoff syndrome)	gi|31417921	GQAK*HQPTAIIAK	Acetyl (K)		*		Carbohydrate Metabolism	
phosphoglycerate mutase 1	gi|44890768	AETAAK*HGEAQVK	Acetyl (K)		*	*	Carbohydrate Metabolism	Hydrolase
hemoglobin, beta	gi|40886941	VVAGVANALAHK*YH	Acetyl (K)			*	Metabolism	Oxygen Transport

Protein Name	gi	Peptide Sequence	No. Sites	Known Acetylated Proteins	Nuclear Extract	S100	Biological Process	Molecular Function
Cell Cycle								
PCNP	gi|9966827	GK*HGFSDNQK	Acetyl (K)		*		Cell Cycle	
cyclin T1	gi|2981196	VHAAADK*HNSVEDSVTK	Acetyl (K)		*		Cell Cycle/ Proliferation	
VEGF	gi|3712671	GK*GK*GQK*R	3 Acetyl (K)		*		Cell Cycle/ Proliferation	Growth Factor
Miscellaneous								
B23 nucleophosmin	gi|825671	SNQNGK*DSKPSSTPR	Acetyl (K)		*		Cell Maintenance	Protein Transport
B23 nucleophosmin	gi|825671	SAPGGGSK*VPQK*K	2 Acetyl (K)		*		Cell Maintenance	Protein Transport
guanylate binding protein 4	gi|15558943	KLVDTIEK*K	Acetyl (K)			*	Defense Response	Hydrolase
metallothionein MT IIpg	gi|224278	CAQGCICK*GASDK	Acetyl (K)			*	Redox	Heavy Metal Binding
thioredoxin	gi|55957946	VGEFSGANK*EK	Acetyl (K)			*	Redox	Electron Transport
B-cell CLL/lymphoma 9-like	gi|32698936	GHCPPAPAK*PMHPENK	Acetyl (K)			*	unknown	
Chain G, Twinning In Crystals Of Human Skeletal Muscle D-Glyceraldehyde-3-Phosphate Dehydrogenase	gi|230868	FHGTVK*AEDGK	Acetyl (K)			*	unknown	
Hypothetical protein LOC84081	gi|62020992	NQEKPSNSESSLGAK*HR	Acetyl (K)		*		Unknown	
Huntingtin interacting protein K	gi|27692116	K*HDSGAADLER	Acetyl (K)		*		Unknown	Protein Binding
retinoblastoma binding protein 7 (RbAp46)	gi|57209887	IECEIK*INHEGEVNR	Acetyl (K)		*		Unknown	
ubiquitin-like 4	gi|57284174	GK*ALADGK*R	2 Acetyl (K)		*		Unknown	

Table 3-1. List of lysine-acetylated proteins identified from a proteomic screen of HeLa nuclear and cytosolic extracts. Protein name, gi#, peptide sequence and annotated biological process associated with the protein with its annotated molecular function are indicated. Source of modified protein from nuclear or cytosolic (S100) sample is also indicated. A K followed by an asterisk designates the lysine residue identified as being acetylated.

Table 3-2

Protein Name	gi	Peptide Sequence	Fed	Fasted	Biological Process	Molecular Function
Metabolism						
Aldh2 protein	gi|13529609	TEQGPQVDETQFK*K	*		Glycolysis	Oxidoreductase
aconitase 2	gi|18079339	YDLLEK*NINIVR		*	TCA Cycle	Hydro-lyase
malate dehydrogenase 2	gi|26324299	ITPFEEK*NIAEAIPELK		*	TCA Cycle	Oxidoreductase
malate dehydrogenase 2	gi|26324299	IQEAGTEVVK*AK		*	TCA Cycle	Oxidoreductase
malate dehydrogenase 2	gi|26324299	K*HGVYNPNK	*	*	TCA Cycle	Oxidoreductase
Is ocitrate dehydrogenase 2	gi|37748684	DIFQEIFDK*HYK	*		TCA Cycle	Oxidoreductase
Is ocitrate dehydrogenase 2	gi|37748684	DQTNDQVTIDSALATQIK*YSVAVK	*	*	TCA Cycle	Oxidoreductase
fumarate hydratase 1	gi|33859564	AAAEVNQEYGLDPK*IASAIMK	*	*	TCA Cycle	Hydro-lyase
fumarate hydratase 1	gi|33859564	VPTDIK*YYGAQTVR	*	*	TCA Cycle	Hydro-lyase
fumarate hydratase 1	gi|33859564	TAHK*NGSTLK		*	TCA Cycle	Hydro-lyase
pyruvate carboxylase	gi|200246	LTSDSVKLAKQVGYENAGTVEFLVDK*HGK		*	TCA Cycle	Carboxylase
pyruvate carboxylase	gi|200246	QVGYENAGTVEFLVDK*HGK	*	*	TCA Cycle	Carboxylase
dihydrolipoamide dehydrogenase	gi|20078522	MMEQIK*HSAVK		*	TCA Cycle	Oxidoreductase
NADH dehydrogenase (ubiquinone) flavoprotein 1	gi|19526814	GPDWILGEMIK*TSGLR	*		Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) flavoprotein 1	gi|19526814	WSFMNK*PSDGRPK	*	*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) flavoprotein 1	gi|19526814	LIEFYK*HESCGQCTPCR		*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) Fe-S protein 1	gi|21704020	MCLVEIEIK*APK		*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) Fe-S protein 6	gi|56711244	TGTCGYCGLDFK*QHHH	*		Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2	gi|13905064	YAFGQEIK*TVSLNNLSADEVTR		*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2	gi|13905064	ECSEVQPK*LWAR		*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4	gi|33563266	NNPEPWNK*LGPNEQYK	*	*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5	gi|20306389	WK*PWEPLVEEPPANQWK		*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5	gi|20306389	TLDILK*HFPK	*		Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5	gi|20306389	K*YTEQITNEK	*	*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9	gi|37231685	WLSSEIEETK*PAK	*	*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 10	gi|13196624	FYDDPK*SNDGNSYR	*		Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 10	gi|13196624	K*QCVDHYNEIK	*		Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 10	gi|13196624	YAPGYNAEVGDK*WMMLK		*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 10	gi|13196624	VTSAYLQDIENAYK*K		*	Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 beta subcomplex 3	gi|12851039	IEGTPLETVQK*K	*		Oxidative Phosphorylation	Complex I
NADH dehydrogenase (ubiquinone) 1 beta subcomplex 3	gi|12851039	QWK*IEGTPLETVQK	*	*	Oxidative Phosphorylation	Complex I
NADH-ubiquinone oxidoreductase subunit B17.2 , complex I	gi|26325655	IGTLVGEDK*YGNK	*	*	Oxidative Phosphorylation	Complex I
NADH-ubiquinone oxidoreductase subunit B17.2 , complex I	gi|26328555	YGNK*YYEDNK	*	*	Oxidative Phosphorylation	Complex I
succinate dehydrogenase Fp subunit	gi|3851614	AFGGQSLK*FGK	*		Oxidative Phosphorylation	Complex II
succinate dehydrogenase Fp subunit	gi|3851614	ISQLYGDLK*HLK	*		Oxidative Phosphorylation	Complex II
succinate dehydrogenase Fp subunit	gi|3851614	GQVLK*HVNGQDQIVPGLYACGEAACASVHGANR	*	*	Oxidative Phosphorylation	Complex II
succinate dehydrogenase Fp subunit	gi|3851614	VPSIK*ANAGEEESVMNLDK	*	*	Oxidative Phosphorylation	Complex II
succinate dehydrogenase Fp subunit	gi|3851614	ANAGEEESVMNLDK*LR		*	Oxidative Phosphorylation	Complex II
succinate dehydrogenase Fp subunit	gi|3851614	VGSVLQEGCEK*ISQLYGDLK		*	Oxidative Phosphorylation	Complex II
Ubiquinol-cytochrome c reductase complex 11k Da protein	gi|51831145	EHCEQLEK*CVK	*	*	Oxidative Phosphorylation	Complex III
Ubiquinol-cytochrome c reductase complex 14k Da protein	gi|17380333	YEEDK*FYLEPYLK	*	*	Oxidative Phosphorylation	Complex III
Ubiquinol-cytochrome c reductase complex 14k Da protein	gi|17380333	SAVSASSK*WLDGFR	*	*	Oxidative Phosphorylation	Complex III
Ubiquinol-cytochrome-c reductase complex core protein I	gi|14548301	MVLAAAGGVEHGQLLDLAQK*HLSSVSR	*		Oxidative Phosphorylation	Complex III
Ubiquinol-cytochrome-c reductase complex core protein I	gi|14548301	ALSIK*DLPK		*	Oxidative Phosphorylation	Complex III
ubiquinol cytochrome c reductase core protein 2	gi|22267442	SQLK*DIK		*	Oxidative Phosphorylation	Complex III
ubiquinol cytochrome c reductase core protein 2	gi|22267442	GGLGLAGAK*AK		*	Oxidative Phosphorylation	Complex III
cytochrome c oxidase, subunit VIc	gi|6716343	NYDSMK*DFEEMR		*	Oxidative Phosphorylation	Complex IV
ATP synthase B chain, mitochondrial precursor	gi|204648228	SISVQQEK*ETIAK	*		Oxidative Phosphorylation	Complex V

Protein Name	gi	Peptide Sequence	Fed	Fasted	Biological Process	Molecular Function
Metabolism						
ATP synthase B chain, mitochondrial precursor	gi|20454828	K*YGASFGEFIDK	*		Oxidative Phosphorylation	Complex V
ATP synthase B chain, mitochondrial precursor	gi|20454828	LHK*AYK	*		Oxidative Phosphorylation	Complex V
ATP synthase B chain, mitochondrial precursor	gi|20454828	EEEHMIDWVEK*HVVK	*		Oxidative Phosphorylation	Complex V
ATP synthase B chain, mitochondrial precursor	gi|20454828	AQQALVQK*R	*	*	Oxidative Phosphorylation	Complex V
ATP synthase B chain, mitochondrial precursor	gi|20454828	HVVK*SISVQQEK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase B chain, mitochondrial precursor	gi|20454828	KEEEHMIDWVEK*HVVK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase B chain, mitochondrial precursor	gi|20454828	LNEEK*IAQLEEVK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase F0 subunit 8	gi|62184373	TPWELK*WTK	*		Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	gi|21313679	IPVPEDK*YTALVDQEEKEDVK	*		Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	gi|21313679	ANVAK*PGLVDDFEK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	gi|21313679	EDVK*SCAEFVSGSQLR	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	gi|21313679	IPVPEDK*YTALVDQEEK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	gi|21313679	IQEYEK*QLEK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	gi|21313679	YNALK*IPVPEDK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	gi|21313679	K*YPYWPHQPIENL	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F	gi|14789694	FDDPK*FEVIDKPQS	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F	gi|14789694	QMYGK*GEMDTFPTFK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit g	gi|31980744	VELVPPTPAEIPTAIQSVK*K	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit g	gi|31980744	TGSFK*HLTVK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	gi|3680748	ISEQSDAK*LK	*		Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	gi|3680748	STVAQLVK*R	*		Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	gi|3680748	FNDGTDEK*K	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	gi|3680748	SDGK*ISEQSDAK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	gi|3680748	AMK*QVAGTMK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	gi|3680748	SDGK*ISEQSDAK*LK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	gi|3680748	DNGK*HALIIYDDLSK	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	gi|3680748	EGDVVK*R		*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	gi|3680748	TSIAIDTIINQK*R		*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	gi|3680748	GYLDK*LEPSK		*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit	gi|31980648	EGNDLYHEMIESGVINLK*DATSK	*		Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit	gi|31980648	VLDSGAPIK*IPVGPETLGR		*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit	gi|31980648	ADK*LAEEHGS	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	gi|11602916	VYGTGSLALYEK*ADIK	*		Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	gi|11602916	K*HLIIGVSSDR	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	gi|11602916	QMK*NEVAALTAAGK		*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	gi|20070412	TVLK*SFLSPNQILK	*		Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	gi|20070412	IGEK*YVDMSAK	*		Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	gi|20070412	KLDQVEK*ELLR	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	gi|20070412	LDQVEK*ELLR	*		Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	gi|20070412	LEIK*TDPSIMGGMIVR	*	*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	gi|20070412	GEVPCTVTTASPLDDAVLSELK*TVLK		*	Oxidative Phosphorylation	Complex V
ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	gi|20070412	VGQLLK*DPK		*	Oxidative Phosphorylation	Complex V
PREDICTED: similar to ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	gi|51766932	GQVPCTVTTASPLDDAVLSELK*TVLK	*		Oxidative Phosphorylation	Complex V
PREDICTED: similar to ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	gi|51766932	SFLSPNQILK*LEIK	*		Oxidative Phosphorylation	Complex V
lactamase, beta	gi|13507666	EGK*YAMAWGVVEK		*	Carbohydrate Metabolism	Hydrolase
lactamase, beta 2	gi|21703764	NINNDTTYCIK*K	*	*	Carbohydrate Metabolism	Hydrolase
aspartate transaminase (EC 2.6.1.1) precursor	gi|90311	NLDK*EYLPIGGLAEFCK	*		Amino Acid Metabolism	Transaminase

Protein Name	gi	Peptide Sequence	Fed	Fasted	Biological Process	Molecular Function
Metabolism						
aspartate transaminase (EC 2.6.1.1) precursor	gi|90311	YYDPK*TCGFDFSGALEDISK	*	*	Amino Acid Metabolism	Transaminase
aspartate transaminase (EC 2.6.1.1) precursor	gi|90311	DVFLPK*PSWGNHTPIFR	*	*	Amino Acid Metabolism	Transaminase
aspartate transaminase (EC 2.6.1.1) precursor	gi|90311	VGAFTVVCK*DAEEAK		*	Amino Acid Metabolism	Transaminase
aspartate transaminase (EC 2.6.1.1) precursor	gi|90311	TQLVSNLK*K		*	Amino Acid Metabolism	Transaminase
aspartate transaminase (EC 2.6.1.1) precursor	gi|90311	QWLQEVK*GMADR		*	Amino Acid Metabolism	Transaminase
isovaleryl coenzyme A dehydrogenase	gi|12836655	AQEIDQTNDFK*NLR		*	Amino Acid Metabolism	Oxidoreductase
aldehyde dehydrogenase family 4, member A1	gi|34328415	FCYADK*ALLNR		*	Amino Acid Metabolism	Oxidoreductase
aldehyde dehydrogenase family 4, member A1	gi|34328415	K*WLEHAR		*	Amino Acid Metabolism	Oxidoreductase
aldehyde dehydrogenase family 4, member A1	gi|34328415	K*EWDLKPMADR		*	Amino Acid Metabolism	Oxidoreductase
Serine hydroxymethyl transferase 2	gi|30802174	SFLLK*DPETSQR		*	Amino Acid Metabolism	Methyltransferase
3-Hydroxy-3-methylglutaryl coenzyme A lyase	gi|409499	DGLQNEK*SIVPTPVK	*	*	Amino Acid/ Carbohydrate Metabolism	Lyase
propionyl Coenzyme A carboxylase, beta polypeptide	gi|13385310	IDAQHK*R	*	*	Amino Acid/ Carbohydrate Metabolism	Carboxyl Transferase
ornithine transcarbamylase	gi|6679184	GYEPDPNIVK*LAEQYAK	*		Amino Acid Metabolism/ Urea Cycle	Carbamoyltransferase
ornithine transcarbamylase	gi|6679184	LAEQYAK*ENGTK	*	*	Amino Acid Metabolism/ Urea Cycle	Carbamoyltransferase
ornithine transcarbamylase	gi|6679184	SLGMIFEK*R		*	Amino Acid Metabolism/ Urea Cycle	Carbamoyltransferase
glutamate dehydrogenase 1	gi|6680027	ISGASEK*DIVHSGLAYTMER	*	*	Amino Acid Metabolism/ Urea Cycle	Oxidoreductase
glutamate dehydrogenase 1	gi|6680027	TAMK*YNLGLDLR	*	*	Amino Acid Metabolism/ Urea Cycle	Oxidoreductase
glutamate dehydrogenase 1	gi|6680027	GASIVEDK*LVEDLK		*	Amino Acid Metabolism/ Urea Cycle	Oxidoreductase
hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase	gi|33859811	AGLEQGSDAGYLAESQK*FGELALTK	*	*	Amino Acid/ Fatty Acid/ Lipid/ Coenzyme Metabolism	Oxidoreductase
hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase	gi|33859811	ALMGLYNGQVLCK*K	*	*	Amino Acid/ Fatty Acid/ Lipid/ Coenzyme Metabolism	Oxidoreductase
hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase	gi|33859811	ILQEGVDPK*K	*	*	Amino Acid/ Fatty Acid/ Lipid/ Coenzyme Metabolism	Oxidoreductase
hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase alpha subunit	gi|33859811	MFEK*LEK		*	Amino Acid/Fatty Acid/Lipid/Coenzyme Metabolism	Oxidoreductase
hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase alpha subunit	gi|33859811	FVDLYGAQK*VVDR		*	Amino Acid/Fatty Acid/Lipid/Coenzyme Metabolism	Oxidoreductase
Hydroxyacyl-Coenzyme A dehydrogenase beta subunit	gi|13542763	LSLLSK*FR		*	Amino Acid/Fatty Acid/Lipid/Coenzyme Metabolism	Oxidoreductase
Hydroxyacyl-Coenzyme A dehydrogenase beta subunit	gi|13542763	DFIYVSQDPK*DQLLLGPTYATPK	*	*	Amino Acid/Fatty Acid/Lipid/Coenzyme Metabolism	Oxidoreductase
PREDICTED: carbamoyl-phosphate synthetase 1	gi|51705066	EWPANLDLK*K	*		Urea Cycle	
PREDICTED: carbamoyl-phosphate synthetase 1	gi|51705066	DELGLNK*YMESDGIK	*	*	Urea Cycle	
PREDICTED: carbamoyl-phosphate synthetase 1	gi|51705066	GVLK*EYGVK	*	*	Urea Cycle	
PREDICTED: carbamoyl-phosphate synthetase 1	gi|51705066	TAHIVLEDGTK*MK	*	*	Urea Cycle	
PREDICTED: carbamoyl-phosphate synthetase 1	gi|51705066	TLK*GLNSDSVTEETLR	*	*	Urea Cycle	
PREDICTED: carbamoyl-phosphate synthetase 1	gi|51705066	K*EPLFGISTGNIITGLAAGAK		*	Urea Cycle	
PREDICTED: carbamoyl-phosphate synthetase 1	gi|51705066	K*ELSEPSSTR		*	Urea Cycle	
PREDICTED: carbamoyl-phosphate synthetase 1	gi|51705066	VMIGESIDEK*R		*	Urea Cycle	
PREDICTED: carbamoyl-phosphate synthetase 1	gi|51705066	SAYALGGLGSGICPNK*ETLIDLGTK		*	Urea Cycle	
adenylate kinase 3 alpha-like 1	gi|12836369	HFELK*HLSSGDLLR	*	*	Nucleotide Metabolism	Adenylate Kinase
adenylate kinase isozyme 2	gi|4760598	NLETPSCK*NGFLLDGFPR	*		Nucleotide Metabolism	Adenylate Kinase
urate oxidase	gi|6678509	K*DYLHGDNSDIIPTDTIK	*		Nucleotide Metabolism	Oxidoreductase
urate oxidase	gi|6678509	DIVLQK*FAGPYDK	*	*	Nucleotide Metabolism	Oxidoreductase
urate oxidase	gi|6678509	DQFTTLPEVK*DR	*	*	Nucleotide Metabolism	Oxidoreductase
urate oxidase	gi|6678509	FEK*NGIK	*	*	Nucleotide Metabolism	Oxidoreductase
urate oxidase	gi|6678509	MGLINK*EEVLLPLDNPYGK	*	*	Nucleotide Metabolism	Oxidoreductase
urate oxidase	gi|6678509	FAGPYDK*GEYSPSVQK	*	*	Nucleotide Metabolism	Oxidoreductase
urate oxidase	gi|6678509	DGK*YHSIK	*	*	Nucleotide Metabolism	Oxidoreductase
urate oxidase	gi|6678509	EEVLLPLDNPYGK*ITGTVK		*	Nucleotide Metabolism	Oxidoreductase
urate oxidase	gi|6678509	AHVYVEEVPWK*R		*	Nucleotide Metabolism	Oxidoreductase
urate oxidase	gi|6678509	VLK*TTQSGFEGFLK		*	Nucleotide Metabolism	Oxidoreductase
peroxisomal delta3, delta2-enoyl-Coenzyme A isomerase	gi|26364942	KDPGNEVK*LR		*	Lipid/ Fatty Acid Metabolism	Isomerase
very-long-chain acyl-CoA dehydrogenase	gi|2388724	DAATGAVK*EK	*	*	Lipid/ Fatty Acid Metabolism	Acyl CoA dehydrogenase

Protein Name	gi	Peptide Sequence	Fed	Fasted	Biological Process	Molecular Function
Metabolism						
very-long-chain acyl-CoA dehydrogenase	gi|2388724	SSAIPSPCGK*YYTLNGSK		*	Lipid/ Fatty Acid Metabolism	Acyl CoA dehydrogenase
very-long-chain acyl-CoA synthetase	gi|3087820	SK*FSASQFWDDCR	*	*	Lipid/ Fatty Acid Metabolism	Ligase
very-long-chain acyl-CoA synthetase	gi|3087820	YLCNTPQK*PNDR		*	Lipid/ Fatty Acid Metabolism	Ligase
Long-chain-fatty-acid-CoA ligase 1	gi|729927	DGWLHTGDIGK*WLPNGTLK	*	*	Lipid/ Fatty Acid Metabolism	Ligase
acetyl-Coenzyme A acetyltransferase 1 precursor	gi|21450129	VLK*YAGLK	*	*	Lipid/ Fatty Acid Metabolism	Acetyltransferase
acetyl-Coenzyme A acetyltransferase 1 precursor	gi|21450129	EAWDAGK*FASEITPITISVK	*	*	Lipid/ Fatty Acid Metabolism	Acetyltransferase
acetyl-Coenzyme A acetyltransferase 1 precursor	gi|21450129	DGLTDVYNK*IHMGNCAENTAK	*	*	Lipid/ Fatty Acid Metabolism	Acetyltransferase
acetyl-Coenzyme A acetyltransferase 1 precursor	gi|21450129	GATPYGGVK*LEDLIVK	*	*	Lipid/ Fatty Acid Metabolism	Acetyltransferase
Acetyl-Coenzyme A acyltransferase 2	gi|20810027	FGTK*FGLDLK	*	*	Lipid/ Fatty Acid Metabolism	Acyltransferase
Acetyl-Coenzyme A acyltransferase 2	gi|20810027	K*HNFTPLAR	*	*	Lipid/ Fatty Acid Metabolism	Acyltransferase
Carnitine O-octanoyltransferase	gi|15126777	MMK*DCSHGK	*	*	Lipid/ Fatty Acid Metabolism	Acyltransferase
carnitine palmitoyltransferase I	gi|2388722	FGK*HLR	*		Lipid/ Fatty Acid Metabolism	Acyltransferase
carnitine palmitoyltransferase 2	gi|6753514	YHGQLTK*EAAMGQGFDR	*	*	Lipid/ Fatty Acid Metabolism	Acyltransferase
cytochrome b-5	gi|13385268	YYTLEEIQK*HK	*	*	Lipid/ Fatty Acid Metabolism	Oxidoreductase
17beta-hydroxysteroid dehydrogenase IV	gi|1213008	GGTVAAK*WTIDLK	*		Lipid/ Fatty Acid Metabolism	Oxidoreductase
unnamed protein product	gi|26324826	YAQVKPDGTYVK*PLSNK	*	*	Lipid/ Fatty Acid Metabolism	Oxidoreductase
unnamed protein product	gi|26324826	SPLNK*TEVHESYYK	*		Lipid/ Fatty Acid Metabolism	Oxidoreductase
L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain	gi|6680163	TFESLVDFCK*TLGK	*		Lipid/ Fatty Acid Metabolism	Oxidoreductase
KE6a	gi|3811386	GK*HAAFQADVSQGPAAR	*		Lipid/ Fatty Acid Metabolism	Oxidoreductase
hydroxysteroid dehydrogenase-5, delta<5>-3-beta	gi|6680293	VLK*GDILDAQCLK	*	*	Lipid/ Fatty Acid Metabolism	Steroid Hormone Metabolism
hydroxysteroid dehydrogenase-5, delta<5>-3-beta	gi|6680293	K*HEEELSK	*	*	Lipid/ Fatty Acid Metabolism	Steroid Hormone Metabolism
hydroxysteroid dehydrogenase-5, delta<5>-3-beta	gi|6680293	TSEWIGTLVK*QHR		*	Lipid/ Fatty Acid Metabolism	Oxidoreductase
Cyp27a1 protein	gi|12805421	LVQK*YEIALAPGMGEVK	*		Lipid/ Fatty Acid Metabolism	Monooxygenase
Cyp27a1 protein	gi|12805421	YLNGWDNIFSFGK*K		*	Lipid/ Fatty Acid Metabolism	Monooxygenase
Cyp27a1 protein	gi|12805421	QEGK*YPIR		*	Lipid/ Fatty Acid Metabolism	Monooxygenase
acyl-Coenzyme A oxidase 2, branched chain	gi|16716411	SPANTQENPAYK*K		*	Lipid/ Fatty Acid Metabolism	Oxidoreductase
Acyl-Coenzyme A dehydrogenase, short/branched chain	gi|32450737	IGHGYK*YAIGSLNEGR	*		Lipid/ Fatty Acid/ Bile Acid Metabolism	Oxidoreductase
Peroxisomal bifunctional enzyme	gi|17366737	SVQASVK*HPYEVAIK	*		Lipid/ Fatty Acid/ Amino Acid Metabolism	Oxidoreductase
Peroxisomal bifunctional enzyme	gi|17366737	VGIPVVAVESDPK*QLDTAK	*	*	Lipid/ Fatty Acid/ Amino Acid Metabolism	Oxidoreductase
Echs1 protein	gi|12805413	AFAAGADIK*EMQNR		*	Lipid/ Fatty Acid/ Amino Acid Metabolism	Hydro-lyase
HMG CoA synthase 2	gi|555835	DASPGSPLEK*LVSSVSDLPK	*	*	Lipid/ Cholesterol/ Ketone Metabolism	Acyltransferase
HMG CoA synthase 2	gi|555835	FNNVEAGK*YTVGLGQTR	*	*	Lipid/ Cholesterol/ Ketone Metabolism	Acyltransferase
HMG CoA synthase 2	gi|555835	VSK*DASPGSPLEK		*	Lipid/ Cholesterol/ Ketone Metabolism	Acyltransferase
sterol-carrier protein X	gi|293794	KLEEEGEQFVK*K	*		Lipid/ Fatty Acid/ Cofactor Metabolism	Thiolase
sterol-carrier protein X	gi|293794	EHMEK*YGTK	*		Lipid/ Fatty Acid/ Cofactor Metabolism	Thiolase
sterol-carrier protein X	gi|293794	GDNTYGGK*WVINPSGGLISK	*	*	Lipid/ Fatty Acid/ Cofactor Metabolism	Thiolase
sterol-carrier protein X	gi|293794	TTPTDK*HIEVLIDK	*	*	Lipid/ Fatty Acid/ Cofactor Metabolism	Thiolase
sterol-carrier protein X	gi|293794	VVGYDMSK*EAAR		*	Lipid/ Fatty Acid/ Cofactor Metabolism	Thiolase
Glycine-N-acyltransferase	gi|19353840	K*YLPESLK	*	*	Hormone Metabolism	Acyltransferase
4-aminobutyrate aminotransferase	gi|37202121	K*HGCAFLVDEVQTGGGCTGK	*		Neurotransmitter Metabolism	Transaminase
microsomal glutathione S-transferase	gi|5762309	ITNK*VFANPEDCAGFGK		*	Coenzyme Metabolism	Glutathione Transferase
microsomal glutathione S-transferase	gi|5762309	VFANPEDCAGFGK*GENAK		*	Coenzyme Metabolism	Glutathione Transferase
class kappa glutathione S-transferase	gi|30314454	DFFGETVK*K		*	Coenzyme Metabolism	Peroxidase
choline dehydrogenase	gi|26335131	ELQPGSHVQSDK*EIDAFVR		*	Metabolism	Oxidoreductase
sarcosine dehydrogenase	gi|26352359	DPSGGPVSLDFVK*NGEYALER	*	*	Metabolism	Aminomethyl Transferase
2,4-dienoyl CoA reductase 1	gi|12832971	TGNK*VHAIR	*		Metabolism	Oxidoreductase
2,4-dienoyl CoA reductase 1	gi|12832971	ATAEEISSK*TGNK	*	*	Metabolism	Oxidoreductase
monoamine oxidase B	gi|26327655	NVK*YVDLGGSYVGPTQNR	*		Metabolism	Oxidoreductase

Protein Name	gi	Peptide Sequence	Fed	Fasted	Biological Process	Molecular Function
Metabolism						
electron transferring flavoprotein, dehydrogenase	gi|34784638	WK*HHPSIQPTLEGGK	*		Metabolism	Oxidoreductase
electron transferring flavoprotein, dehydrogenase	gi|34784638	FAILTEK*HR	*	*	Metabolism	Oxidoreductase
electron transferring flavoprotein, dehydrogenase	gi|34784638	QLAAEQGK*DIR		*	Metabolism	Oxidoreductase
electron transferring flavoprotein, dehydrogenase	gi|34784638	GIATNDVGIQK*DGAPK		*	Metabolism	Oxidoreductase
3-hydroxybutyrate dehydrogenase	gi|20071589	K*YFDEK	*	*	Metabolism	Oxidoreductase
3-hydroxybutyrate dehydrogenase	gi|20071589	SGLK*DPEK		*	Metabolism	Oxidoreductase
diaphorase 1 (NADH)	gi|12846591	DTHPK*FPAGGK	*	*	Metabolism	Oxidoreductase
diaphorase 1 (NADH)	gi|12846591	STPAITLENPDIK*YPLR	*	*	Metabolism	Oxidoreductase
hemoglobin alpha, adult chain 1	gi|12846963	AAWGK*IGGHGAEYGAEALER	*		Metabolism	Oxygen Transport
5730439E10Rik protein	gi|28280023	LDK*YGTGVGYCK	*	*	Metabolism	Peptidase
Aldehyde dehydrogenase family 6, subfamily A1	gi|23271115	K*WLPELVDR		*	Metabolism	Oxidoreductase
Aldehyde dehydrogenase family 6, subfamily A1	gi|23271115	SDK*WIDIHNPATNEVVGR		*	Metabolism	Oxidoreductase
Aldehyde dehydrogenase family 6, subfamily A1	gi|23271115	ENLK*EIAR		*	Metabolism	Oxidoreductase
coproporphyrinogen oxidase	gi|436577	K*HCDDSYTPR		*	Metabolism	Oxidoreductase
bilirubin/phenol UDP-glucuronosyltransferase	gi|436187	NTILVK*WLPQNDLIGHPK	*	*	Metabolism	Glycosyltransferase
cytochrome P450, family 1, subfamily a, polypeptide 2	gi|309205	NSIQDITSALFK*HSENYK	*		Metabolism	Monooxygenase
cytochrome P450, family 2, subfamily c, polypeptide 29	gi|473717	EHK*ESLDVTNPR	*		Metabolism	Monooxygenase
cytochrome P450, family 3, subfamily a, polypeptide 11	gi|6681113	K*HELFK	*	*	Metabolism	Monooxygenase
Stress Response						
Sod1	gi|201006	K*HGGPADEER	*	*	Stress Response	Redox
Sod2 protein	gi|17390379	HHAAYVNNLNATEEK*YHEALAK	*	*	Stress Response	Redox
Msra protein	gi|15928529	VLSK*HNFGPITTDIR	*	*	Stress Response	Redox
AFG3l2	gi|29144877	HLSDAINEK*HFEQAIER	*		Stress Response	Protease/ ATPase
AFG3l2	gi|51770961	FK*DVAGCEEAK	*	*	Stress Response	Protease/ ATPase
major urinary protein	gi|199869	FAQLCEK*HGILR	*	*	Stress Response	Lipid Binding
stress-70 protein (PBP74/CSA)	gi|903309	AQFEGIVTDLIK*R	*	*	Stress Response	Protein Folding
stress-70 protein (PBP74/CSA)	gi|903309	ETGVDLTK*DNMALQR		*	Stress Response	Protein Folding
stress-70 protein (PBP74/CSA)	gi|903309	HIVK*EFK		*	Stress Response	Protein Folding
Glutathione peroxidase	gi|121666	NALPTPSDDPTALMTDPK*YIIWSPVCR		*	Stress Response	Peroxidase
Cell Cycle/ Apoptosis						
Kidney expressed gene 1	gi|14789838	DPK*HCQEFLGSSEVINWK	*	*	Cell Cycle	Tubulin Binding
Kidney expressed gene 1	gi|14789838	K*HLPESLK		*	Cell Cycle	Tubulin Binding
programmed cell death 8 (AIF)	gi|6755004	IIK*DGEQHEDLNEVAK		*	Apoptosis	Oxidoreductase
cytochrome c	gi|26353026	K*IFVQK		*	Apoptosis	Electron Transport
Histones						
histone H2B	gi|15030326	K*GSK*K*AVTK*AQK	*	*	Chromosome Organization	Structural
histone H2B	gi|15030326	SAPAPK*K*GSK*K*AVTK*AQK	*		Chromosome Organization	Structural
histone H4	gi|27692935	GK*GGK*GLGK*GGAK	*	*	Chromosome Organization	Structural
histone H4	gi|27692935	GLGK*GGAK*R	*	*	Chromosome Organization	Structural
histone H4	gi|27692935	GGK*GLGK*GGAK*R	*	*	Chromosome Organization	Structural
histone H4	gi|27692935	GK*GGK*GLGK*GGAK*R	*	*	Chromosome Organization	Structural

Protein Name	gi	Peptide Sequence	Fed	Fasted	Biological Process	Molecular Function
Cell Maintenance						
diazepam binding inhibitor	gi|22135648	AK*WDSWNK	*		Cell Maintenance	Lipid, Cofactor, Fatty Acid Binding
VDAC-1	gi|10720404	GYGFGLIK*LDLK	*		Cell Maintenance	Anion Transport
VDAC-1	gi|10720404	FGIAAK*YQVDPDACFSAK	*	*	Cell Maintenance	Anion Transport
VDAC-1	gi|10720404	VNGSLETK*YR	*	*	Cell Maintenance	Anion Transport
VDAC-1	gi|10720404	DVFTK*GYGFGLIK	*	*	Cell Maintenance	Anion Transport

Protein Name	gi	Peptide Sequence	Fed	Fasted	Biological Process	Molecular Function
Cell Maintenance						
VDAC-2	gi|37514858	DIFNK*GFGFGLVK	*		Cell Maintenance	Anion Transport
VDAC-2	gi|37514858	YK*WCEYGLTFTEK	*	*	Cell Maintenance	Anion Transport
VDAC-3	gi|6755967	DVFNK*GYGFGMVK		*	Cell Maintenance	Anion Transport
VDAC-3	gi|6755967	YK*LDCR		*	Cell Maintenance	Anion Transport
VDAC-3	gi|6755967	ASGNLETK*YK		*	Cell Maintenance	Anion Transport
solute carrier family 25, member 1	gi|23943838	ILK*NEGPK	*		Cell Maintenance	Adenine Transport
solute carrier family 25, member 1	gi|23943838	FIHDQTSSNPK*YR	*	*	Cell Maintenance	Adenine Transport
solute carrier family 25, member 1	gi|23943838	GTYQGLTATVLK*QGSNQAIR	*	*	Cell Maintenance	Adenine Transport
solute carrier family 25, member 3	gi|19526818	MQVDPQK*YK	*	*	Cell Maintenance	Transport
solute carrier family 25, member 3	gi|19526818	MYK*EEGLNAFYK		*	Cell Maintenance	Transport
solute carrier family 25, member 5	gi|22094075	EFK*GLGDCLVK	*	*	Cell Maintenance	Transporter
solute carrier family 25, member 5	gi|22094075	GLGDCLVK*IYK	*	*	Cell Maintenance	Transporter
solute carrier family 25, member 5	gi|22094075	IYK*SDGIK	*	*	Cell Maintenance	Transporter
solute carrier family 25, member 5	gi|22094075	YFPTQALNFAFK*DK	*	*	Cell Maintenance	Transporter
solute carrier family 25, member 5	gi|22094075	DFLAGGVAAAISK*TAVAPIER	*	*	Cell Maintenance	Transporter
solute carrier family 25, member 5	gi|22094075	AAYFGIYDTAK*GMLPDPK		*	Cell Maintenance	Transporter
solute carrier family 25, member 13	gi|12833101	ADPAELK*AIFLK		*	Cell Maintenance	Transporter
solute carrier family 25, member 13	gi|12833101	DLGFFGIYK*GAK		*	Cell Maintenance	Transporter
solute carrier family 25, member 20	gi|10048462	CLLQIQASSGENK*YSGTLDCAK	*		Cell Maintenance	Adenine Transport
solute carrier family 25, member 20	gi|10048462	FQTAPPGK*YPNGFR	*	*	Cell Maintenance	Adenine Transport
mitochondrial carrier homolog 2	gi|9790055	ESK*YCGLCDSIVTIYR		*	Cell Maintenance	Transport
ATP-binding cassette, sub-family D (ALD), member 3	gi|32484187	MTIMEQK*YEGEYR	*		Cell Maintenance	Transporter
ATP-binding cassette, sub-family B, member 7	gi|38086485	YFNNEK*YEAQR	*	*	Cell Maintenance	
muscle protein 637	gi|10178975	VMQLEK*HR		*	Cell Proliferation	Protein synthesis
thiosulfate sulfurtransferase	gi|6678449	NWLK*EGHPVTSEPSRPEPAVFK		*	Cell Maintenance	Thiosulfate Sulfurtransferase
Chaperones						
Trap1 protein	gi|13879408	GSVSK*HEFQAETK		*	Chaperone	Protein Folding
AAA-ATPase TOB3	gi|30725845	VAQFDYGK*K		*	Chaparone	ATPase
AAA-ATPase TOB3	gi|30725845	MYFDK*YVLKPATEGK	*		Chaparone	ATPase
Transcription/ Translation						
C. elegans ceh-10 homeo domain containing homolog	gi|6671750	AGEALSK*PK		*	Transcription	Transcription Factor
ribosomal protein L10A	gi|6755350	K*YDAFLASESLIK	*	*	Translation	Ribosome Structure
Signal Transduction						
src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristylation sit	gi|54287682	VIK*SADMK		*	Signal Transduction	Tyrosine Kinase
Unknown						
chromogranin B	gi|12860285	DASHPQFK*EIQK	*			
chromogranin B	gi|12860285	EAK*IYHSEERVGKER		*		
es1 protein	gi|20070420	ALGAK*HCVK		*		
inner membrane protein, mitochondrial	gi|26328849	VQEQELK*YEFEQGLSEK	*	*		
glycoprotein, synaptic 2	gi|12846015	K*YDFTSSR		*		
Similar to arylacetyl acyl-CoA N-acyltransferase	gi|58476803	DPK*HCLEFLGTPDVINWK	*	*		
Ociad1 protein	gi|13905276	SSPPGHYTQK*PK	*			
brain protein 44	gi|21312594	LMDK*VELLLPK	*			
Usmg5 protein	gi|19353381	K*YFNSYTLTGR		*		
Usmg5 protein	gi|19353381	AGAESDGQFQFTGIK*K	*	*		
PREDICTED: similar to RIKEN cDNA C730036D15	gi|38078346	K*HLLQSGSK	*	*		
hypothetical protein LOC52637	gi|19527228	HNEETGDNVGPLIIK*K		*		
hypothetical protein LOC69893	gi|28076933	FNCEK*YGHGDSCYK	*	*		
Hypothetical LOC235043	gi|14789776	DKDPK*YSALR	*	*		

Table 3-2. List of lysine-acetylated proteins identified from a proteomic screen of fed and fasted mouse liver mitochondria. Protein name, gi#, peptide sequence and annotated biological process associated with the protein with its annotated molecular function are indicated. Source of modified protein from fed or fasted mouse liver mitochondria is also indicated. A K followed by an asterisk designates the lysine residue identified as being acetylated.

Figure 3-6

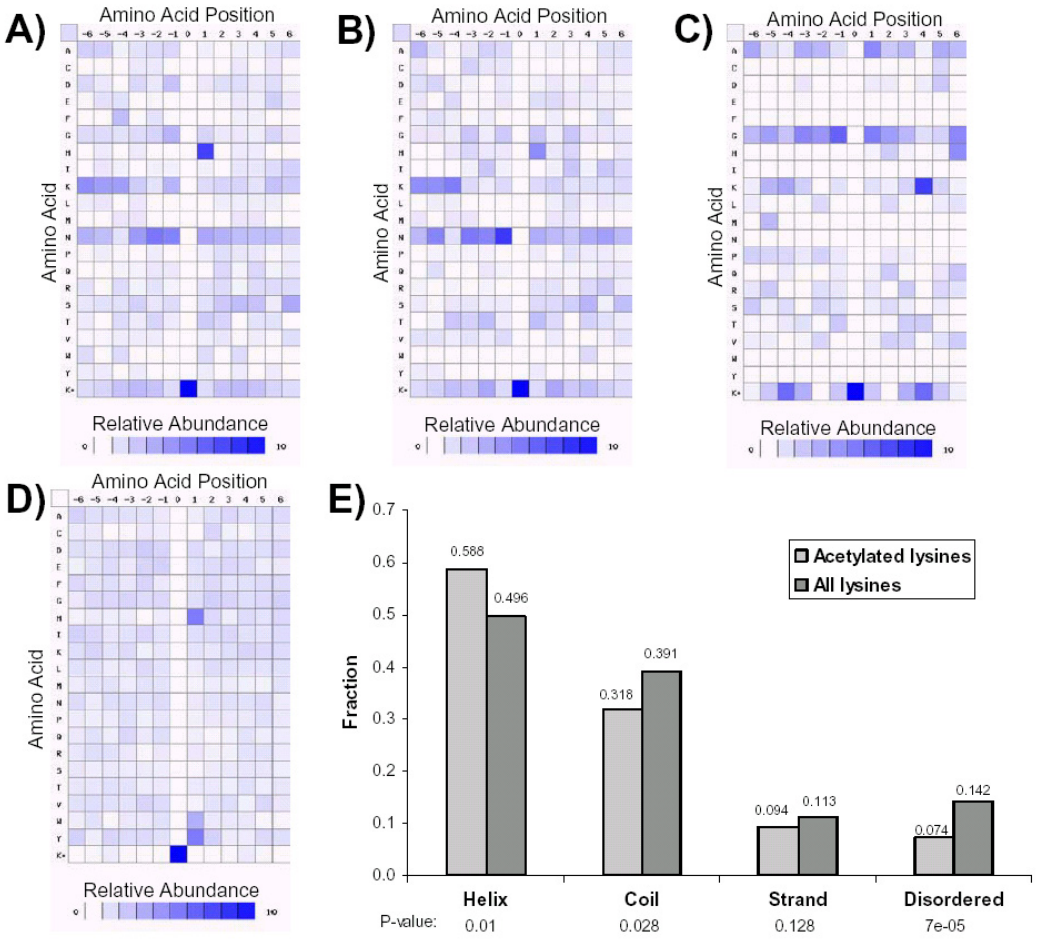

Figure 3-6. Bioinformatics analysis of the lysine acetylated sites. The relative abundance of each amino acid residue (including acetylated lysine) surrounding sites of lysine acetylation was calculated and schematically represented by density maps. Prevalence of specific amino acids at positions surrounding lysine acetylation sites are shown for (A) S100 (cytosolic) fraction (excluding histones), (B) nuclear extract (excluding histones), (C) histones, and (D) mitochondrial fractions from both fed- and starved-mouse mitochondria. (E) Distribution of acetylated and non-acetylated lysine residues in protein secondary structures. Local structure analysis carried out by Dr. Jimin Peng of Dr. Nick Grishin's laboratory.

Table 3-3

Protein Name	gi	Functional group	# Sites
Fumarate hydratase 1*	gi|33859554	TCA Cycle	3
Malate dehydrogenase 2	gi|31982186	TCA Cycle	3
Isocitrate dehydrogenase 2*	gi|37748684	TCA Cycle	2
Dihydrolipoamide dehydrogenase	gi|2078522	TCA Cycle	1
NADH dehydrogenase (ubiquinone) flavoprotein 1*	gi|19526814	Oxidative Phosphorylation	3
NADH dehydrogenase (ubiquinone) Fe-S protein 1*	gi|21704020	Oxidative Phosphorylation	1
NADH dehydrogenase (ubiquinone) Fe-S protein 6*	gi|58711244	Oxidative Phosphorylation	1
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2*	gi|13905064	Oxidative Phosphorylation	2
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4*	gi|33563286	Oxidative Phosphorylation	1
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5*	gi|20306389	Oxidative Phosphorylation	3
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9*	gi|37231695	Oxidative Phosphorylation	1
NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 10*	gi|13195624	Oxidative Phosphorylation	4
NADH dehydrogenase (ubiquinone) 1 beta subcomplex 3*	gi|12851039	Oxidative Phosphorylation	2
NADH-ubiquinone oxidoreductase subunit B17.2 , complex I*	gi|28328555	Oxidative Phosphorylation	2
Ubiquinol-cytochrome c reductase complex 11 kDa protein*	gi|51831145	Oxidative Phosphorylation	1
Succinate dehydrogenase Fp subunit	gi|3851814	Oxidative Phosphorylation	6
ATP synthase, H+ transporting, mitochondrial F0 complex, subunit g*	gi|31980744	Oxidative Phosphorylation	2
ATP synthase, H+ transporting, mitochondrial F1 complex, gamma 1*	gi|11602916	Oxidative Phosphorylation	3
Acyl-Coenzyme A dehydrogenase, short/branched chain	gi|32450737	Fatty Acid Metabolism	1
Very-long-chain acyl-CoA dehydrogenase	gi|2388724	Fatty Acid Metabolism	2
Electron transferring flavoprotein, dehydrogenase	gi|34784638	Fatty Acid Metabolism	4
L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain	gi|6680163	Fatty Acid Metabolism	1
Hydroxyacyl-Coenzyme A dehydrogenase	gi|33859911	Fatty Acid Metabolism	5
Hydroxyacyl-Coenzyme A dehydrogenase beta subunit	gi|13542763	Fatty Acid Metabolism	2
17beta-hydroxysteroid dehydrogenase IV	gi|1213008	Fatty Acid Metabolism	1
Isovaleryl coenzyme A dehydrogenase	gi|7669117	Amino Acid Metabolism	1
Sarcosine dehydrogenase	gi|20149748	Amino Acid Metabolism	1
Choline dehydrogenase	gi|30425036	Amino Acid Metabolism	1
Aldehyde dehydrogenase family 6, subfamily A1	gi|23271115	Amino Acid Metabolism	3
Aldehyde dehydrogenase family 4, member A1	gi|28336725	Amino Acid Metabolism	3
Glutamate dehydrogenase 1	gi|6660027	Amino Acid Metabolism	3
3-hydroxybutyrate dehydrogenase	gi|20071599	Ketone Body Metabolism	2
Phosphoglycerate mutase 1*	gi|44890768	Carbohydrate Metabolism	1
Triosephosphate isomerase 1*	gi|16877874	Carbohydrate Metabolism	1
Aldehyde dehydrogenase 2	gi|13529509	Alcohol Metabolism	1
Hydroxysteroid dehydrogenase-5, delta<5>-3-beta	gi|6680283	Steroid Hormone Metabolism	1
Sod1*	gi|201008	Stress Response	1
Sod2*	gi|17390379	Stress Response	1
Profilin I*	gi|30582841	Structural	1
Lamin A/C*	gi|21619981	Structural	1
Thioredoxin*	gi|55957946	Redox	1
PHD finger protein 15*	gi|19878594	Transcription	2
eIF-5A*	gi|42659529	Translation	1

Table 3-3. A list of longevity related and mitochondrial dehydrogenase enzymes identified as lysine acetylated. Longevity related proteins are marked with an asterisk. Protein name, gene index number, functional classification, and number of lysine acetylation sites are specified.

CHAPTER FOUR

PROKARYOTIC LYSINE ACETYLATION

SUMMARY

Lysine acetylation and its regulatory enzymes are known to have pivotal functions in mammalian cellular physiology. However, the extent and functions of this modification in prokaryotic cells remains largely unexplored, presenting a hurdle to further functional study of the modification in prokaryotic cells. In this chapter, I describe the first global screening of prokaryotic lysine acetylation, identifying 138 lysine acetylation sites in 91 proteins in *E. coli*, none of which has been connected with the modification before. In addition to transcriptional regulators, proteins with diverse functions were found to be subject to lysine acetylation. Interestingly, more than 70% of substrate proteins identified in this proteomics screening are metabolic enzymes and translation-related proteins, suggesting an intimate link between the modification and energy metabolism. The dataset suggests that the post-translational modification could be abundant in prokaryotic cells. In addition, the results imply that DNA-independent functions of lysine acetylation are evolutionarily conserved from bacteria to mammalian cells.

This work was the result of a collaborative effort between me and other members of the Zhao lab, including Drs. Sung Chan Kim and Junmei Zhang.

INTRODUCTION

The high abundance of lysine acetylation in mammalian mitochondrial proteins implies the possible wide existence of the modification in prokaryotes, given the evolutionary lineage of eukaryotic mitochondria from bacteria (Gray et al., 1999). Acetyl-coenzyme A (CoA) synthetase (Acs) and CheY are the only proteins in prokaryotes known to be lysine-acetylated (Barak and Eisenbach, 2001; Starai et al., 2002). In *S. enterica,* the lysine-acetylation status of Acs is regulated by CobB deacetylase, a Sir2 homolog in bacteria, as well as Pat acetyltransferase (Starai and Escalante-Semerena, 2004). Acs activation via its deacetylation by CobB is required for bacterial growth on short-chain fatty acids such as acetate and propionate. Interestingly, both human Sirt2 and yeast Sir2 NAD-dependent deacetylases are capable of restoring growth of CobB sirtuin-deficient strains of *S. enterica* on short-chain fatty acids, suggesting that the sirtuins may have evolutionarily conserved roles in cellular metabolism (Starai et al., 2003). Despite evidence of the presence and roles of lysine acetylation in prokaryotes, the extent of lysine acetylation among these microorganisms has not been carefully examined before.

Here, the first proteomics screening of lysine acetylation in *E. coli* is presented. This proteomics study involves efficient affinity isolation of lysine acetylated, tryptic peptides with anti-acetyllysine antibodies and subsequent

peptide identification by nano-HPLC/mass spectrometric analysis. The screening identified 138 lysine acetylation sites in 91 proteins in *E. coli*, 25% of which have mammalian sequence homologs. The results suggest that diverse groups of bacterial proteins are the substrates of lysine acetylation, including metabolic enzymes, stress response proteins, as well as transcription and translation factors. The lysine acetylation substrates in *E. coli* are highly enriched in metabolic enzymes (~53%) and proteins involved in translation (~22%), two processes intimately linked with cellular energy status. These data therefore reveal previously unappreciated roles for lysine acetylation in the regulation of prokaryotic biochemical pathways and imply that DNA-independent functions of lysine acetylation are evolutionarily conserved from prokaryotic cells to eukaryotic cells.

RESULTS

Proteomics Screening of Lysine Acetylation

Lysine acetylation is more difficult to identify by a candidate approach than protein phosphorylation due to the low radioactivity of [^{14}C]-acetyl-CoA and the weak binding affinity of anti-acetyllysine antibody. A lack of protein substrates represents one of the major bottlenecks for characterization of its biological functions. To begin the systematic study of lysine acetylation in

prokaryotes, we carried out the first proteomics screening of lysine-acetylated substrates in bacteria. The goals of this study were (i) to determine the spectrum and extent of lysine acetylation in bacteria, (ii) to identify novel lysine-acetylation substrates and lysine-acetylation sites that could provide candidate proteins for further functional studies, and (iii) to define the molecular pathways that are likely to be affected by lysine acetylation.

The proteomics analysis of lysine acetylation was carried out as described in Chapter 3, and consisted of four steps. First, *E. coli* protein lysate was proteolytically digested with trypsin. Next, the resulting tryptic peptides were subjected to affinity purification by anti-acetyllysine antibody. Third, the affinity purified, lysine-acetylated peptides were analyzed by nano-HPLC/MS/MS for peptide identification and precise localization of lysine-acetylation sites. Finally, the peptide candidates were further manually evaluated to ensure the accuracy of the identifications (Figure 4-1A, C).

Western blotting analysis demonstrates that proteins of a wide molecular weight range can be lysine-acetylated (Figure 4-1B). Subsequent proteomics screening identified 138 lysine-acetylation sites among 91 proteins (Table 4-1). To the best of my knowledge, none of these bacterial proteins have been associated with the modification before.

Lysine-acetylated Protein Groups in *E. coli*

An unbiased screen of a large set of lysine acetylation substrates can provide insight into cellular pathways that would not be apparent through single-protein analysis. This has been exemplified by a recent study on proteomics of lysine acetylation in mammalian cells (Kim et al., 2006b). We therefore attempted to assign each lysine-acetylation substrate to a functional group based on Gene Ontology (GO) molecular functions or biochemical process groups, or previously published literature (Figure 4-2A).

Several of the protein groups defined in this way are of particular interest. For example 48 of the 91 lysine-acetylation substrates (~53%) are metabolic enzymes, including 4 TCA cycle proteins, 7 glycolytic enzymes as well as several enzymes involved in the metabolism of nucleotides, amino acids, and fatty acids (Table 4-2). A few translational regulators are found to be lysine acetylation substrates in mammalian cells. However, 22% of the lysine-acetylation substrates identified here are proteins that are either subunits of translational machinery or are enzymes associated with translational processes, such as the tRNA synthetases. Proteins involved in stress response represented ~5% of lysine acetylation substrates. Transcription factors are well-known examples of lysine-acetylation targets in eukaryotic cells. This proteomics screening identified 2 bacterial transcriptional regulators and one RNA polymerase subunit as substrates of lysine acetylation.

DISCUSSION

Lysine-acetylated Substrate Proteins in Bacteria

The identification of a large number of metabolic enzymes and TCA proteins is reminiscent of lysine acetylation substrates identified in mitochondria (Table 3-4) (Kim et al., 2006b), and provides additional evidence supporting the importance of the modification in the regulation of energy metabolism.

Glycolytic enzymes

We detected 7 of 9 glycolytic enzymes, catalyzing the key reactions in the degradation of glucose to pyruvate, as being subjects of lysine acetylation (Figure 4-3). These proteins include phosphoglucose isomerase, aldolase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, enolase, and pyruvate kinase. Three of these enzymes were also found to be acetylated in mammalian samples, suggesting potentially conserved functional consequences for the modification in the regulation of glycolytic flux.

Pyruvate dehydrogenases

Pyruvate is one of the major products of glycolysis. It can be cleaved by either pyruvate dehydrogenase or pyruvate formate-lyase, both of which were found to be lysine acetylated. The oxidative decarboxylation of pyruvate by

pyruvate dehydrogenase generates acetyl-CoA and NADH, and is therefore functionally restricted to respiratory metabolism under aerobic conditions. Two of the three subunits of pyruvate dehydrogenase are subjects of lysine acetylation. On the other hand, anaerobic cleavage of pyruvate to acetyl CoA and formate is catalyzed by pyruvate formate-lyase. The activities of these two enzymes are thus mutually exclusive. Pyruvate formate-lyase is a homodimer that catalyzes the conversion of pyruvate to acetyl-CoA, which is in turn used to produce ATP through acetyl phosphate. The ATP produced in this manner is the single ATP source in *E. coli* cells under anaerobic conditions with pyruvate as the sole carbon and energy source (Pascal et al., 1981). The activity of pyruvate formate-lyase is tightly regulated at both the transcriptional level and by post-translational modification.

TCA cycle proteins

Four of the eight TCA cycle proteins, citrate synthase, isocitrate dehydrogenase, succinate dehydrogenase and 2-ketoglutarate dehydrogenase are lysine acetylated (Figure 4-3). Among these four proteins, citrate synthase and isocitrate dehydrogenase are known to be regulated at the level of transcription and protein modification. The synthesis of citrate synthase is subject to catabolite repression. It is repressed by glucose and anaerobic conditions, and induced by acetate and oxygen. Isocitrate dehydrogenase catalyzes the conversion of

isocitrate to α-ketoglutarate and CO_2 and is also involved in the regulation of carbon flux at the branch point between the TCA and glyoxylate cycles. Protein phophorylation is also known to regulate the protein's functions (Borthwick et al., 1984; Thorsness and Koshland, 1987).

Enzymes involved in the metabolism of amino acids and nucleotides are also lysine acetylated. These proteins include 7 proteins involved in amino acid metabolism and 7 proteins in nucleotide metabolism.

One enzyme involved in nucleotide metabolism, deoxyribosephosphate aldolase, is of particular interest. The enzyme catalyzes the reversible reaction between glyceraldehyde-3-phosphate and acetaldehyde to form 2-deoxyribose-5-phosphate. We identified this enzyme as being acetylated at lysine 167. This residue has been implicated in the enzyme's catalytic mechanism, forming a Schiff base with the substrate (Heine et al., 2004). Acetylation of this active site lysine would therefore be expected to block enzymatic activity. Such acetylation of active site lysine may represent a common regulatory mechanism among the aldolase family of proteins. Previously, a survey of lysine acetylation among mammalian proteins identified acetylation of lysine 146 of the glycolytic enzyme fructose-1, 6-bisphosphate aldolase (Kim et al., 2006b), which is also the key active site catalytic residue involved in Schiff base formation.

Transcription and translation factors

Transcription factors and histones are founding members of lysine acetylation substrates in mammalian cells. Reminiscent of this, two transcription factors (cAMP receptor protein and trp repressor binding protein) and one subunit of RNA polymerase (rpoB) are lysine-acetylation substrates. In addition to transcriptional regulators, 20 proteins involved in translational regulation are subjects of lysine acetylation. These proteins include translational elongation factors, ribosomal proteins as well as transfer RNA synthetases. Lysine acetylation has not been detected in ribosomal proteins and transfer RNA synthetases in mammalian proteins before, suggesting that lysine acetylation in ribosomal proteins might have been lost during evolution.

A large portion of cellular energy is used in protein synthesis. Therefore, translational machinery needs to be synchronized with energy availability. It is not surprising that the protein expression level for about half of the aminoacyl-tRNA synthetases and for translational factors are under metabolic control (Neidhardt, 1996). This screening identified lysine acetylation at 5 tRNA synthetases, 11 ribosomal subunits, and three translational machinery proteins, implying that post-translational modifications might provide an alternative avenue to modulate the activity of cellular translational machinery.

Stress response proteins

In eukaryotic cells, activation of the sirtuin family of deacetylases promotes cell survival and resistance to stress, implying that the lysine acetylation status of proteins may be altered in response to stress. For example, small molecule activators of sirtuins have been demonstrated to extend *Saccharomyces cerevisiae* lifespan (Howitz et al., 2003). Lysine acetylation has previously been identified in eukaryotes among heat shock proteins as well as ROS regulators such as SOD (Kim et al., 2006b). Like eukaryotic cells, a significant portion of stress response proteins in *E. coli* are also substrates of lysine acetylation, including chaperones and proteins involved in regulation of free radical reduction (such as superoxide dismutase, alkyl hydroperoxide reductase and thioredoxin). In *E. coli*, loss of alkyl hydroperoxide reductase and SOD leads to an increased sensitivity to hydrogen peroxide and alkyl hydroperoxide oxidative stress.

Several respiratory components of the mammalian mitochondrion are closely related to those in the *E. coli*. Mitochondrion NADH dehydrogenase (complex I), succinate dehydrogenase (complex II), and cytochrome c oxidase (complex IV) have homologs in *E. coli*, whose *E. coli* proteins are lysine acetylated

Evolutionary Conservation of Lysine Acetylation between *E. coli* and Mammalian Cells

The mitochondrion is hypothesized to have evolved from α-proteobacteria. Phylogenetic studies suggest that about one tenth of yeast mitochondrial proteins may have evolved from ancestral free-living α-proteobacteria with the remaining 90% being recruited from the nuclear genome of the eukaryotic host (Karlberg et al., 2000).

The results of this screen for lysine acetylated proteins in *E. coli* suggest that regulatory programs tuning the activities of metabolic enzymes may be conserved between *E. coli* and mammalian cells. This concept is supported by the overlap in protein identities between the mammalian and *E. coli* datasets and the evidence of a common regulatory modification observed among the active site lysine of aldolase family members. A comparison, between lysine-acetylated *E. coli* proteins and those in mammalian cells identified in the previous proteomics screening (Kim et al., 2006b), indicates that 22 of the 91 *E. coli* lysine-acetylated proteins (~25%) have mammalian homologs in which the lysine residue is conserved, based on the BLAST sequence alignment with 25% sequence identity as the cutoff score (Table 4-1). This number is likely to be significantly underestimated due to the low percentage of lysine-acetylated proteins identified in mammalian cells that was caused by the wide dynamic range and limited sensitivity of the previous proteomics screening (Kim et al., 2006b). Identification of a large number of lysine-acetylated homologs between *E. coli* and mammalian cells, the abundance of lysine acetylation in both mitochondria and *E. coli*, and

the evolutionary linkage between the mitochondrion and *E. coli* suggest that the modification is likely to be evolutionarily conserved from bacteria to mammalian cells.

Lysine Acetylation Motifs in Bacterial Substrate Proteins and Bioinformatics Analysis

Conserved protein sequence motifs are associated with some post-translational modifications such as protein phosphorylation. Structural studies of GCN5 HAT and H3 tail peptides suggest a recognition site of G-K14-X-P (Rojas et al., 1999), nevertheless, the motifs for lysine acetylation remains largely known. The dataset of acetylation sites identified from *E. coli* proteins allows us to conduct a preliminary analysis of motif preference (Figure 4-2B). Such analysis identified histidine and tyrosine as preferred amino acid resides at the +1 position.

In contrast to the *E. coli* lysine-acetylation motifs, mammalian lysine-acetylation substrates in general show less preference for residues surrounding acetyllysine residues (Kim et al., 2006b). This could be caused by averaged motif information from diverse acetyltransferases. Alternatively, substrates may be recognized through the three dimensional structure of the substrate proteins rather than their linear sequence. However, when considering the mitochondrial subset of lysine acetylation substrates identified in the mammalian dataset, a preference for histidine and tyrosine at the +1 position was also observed (Kim et al., 2006b).

The preference of amino acid residues flanking acetyllysine residues in *E. coli* proteins suggests that the modification might be catalyzed by a limited subset of acetyltransferases with unique substrate preferences. In addition, related acetyltransferases may be at work in the case of mitochondrial acetylation. These results suggest the possibility that linear lysine acetylation motifs exist among lysine-acetylated proteins similar to motifs observed for phosphorylation.

To evaluate the structural features of acetyllysines in the substrate proteins, those proteins for which a homologous crystal structure was available were first identified. BLAST local alignment was used to map the position of the acetylated lysine onto the crystal structure. Visual inspection of the available crystal structures using the Pymol program suggested that most of the acetylation sites (95%) identified occur on the surface of proteins, rather than at active site or ligand binding residues. However, in some instances, modification sites were found within oligomerization interfaces, such as in serine hydroxymethyltransferase, isocitrate dehydrogenase and IMP cyclohydrolase. This suggests that acetylation may affect the oligomerization of proteins or the formation of protein complexes. Such effects of lysine acetylation are known to occur in eukaryotes through the specific binding of acetylated lysine by bromodomains.

CONCLUDING REMARKS

Several important points have emerged from this most comprehensive protein screening of protein lysine acetylation. First, a large number of previously unknown lysine-acetylation substrates exist in prokaryotic species. Given the key roles of many of these substrate proteins in cellular functions and high sequence conservation among prokaryotic species, it is highly likely the modification is conserved and shares similar functions among the prokaryotic species other than *E. coli*. Second, two preferred residues, histidine and tyrosine, at the +1 position of acetyllysine were identified, implying that linear sequence might be of importance for substrate binding and/or activity of acetyltransferase(s) in bacteria. This finding is intriguing given that a similar preference was observed among lysine-acetylated substrates identified in a screen of mitochondrial proteins (Kim et al., 2006b), suggesting the existence of a related set of acetyltransferases in mitochondria. Third, the spectrum of lysine-acetylation substrates is evolutionarily well-conserved. A comparison of lysine-acetylation substrates from *E. coli* and mammalian cell lines suggest that three groups of substrate proteins are enriched in lysine acetylation, including metabolic enzymes, transcriptional/translational regulators, and stress response proteins. Identification of a high number of ribosomal proteins and transfer RNA synthetases was a surprise to us as these proteins are abundant in eukaryotic cells and lysine acetylation was not detected among these proteins in mammalian cells. Lysine acetylation of these proteins might have been lost during evolution.

As with the previous initial studies of lysine acetylation in mammalian cells, this study has limitations. First, only protein substrates present in medium to high abundance were likely to be identified. A significant portion of lysine acetylation substrates that are either of low abundance or modified at a low stoichiometry are expected to be missed, due to the relatively weak binding affinity of the anti-acetyllysine antibody. Second, this study only represents an initial screen and dynamic analyses have not been carried out under diverse genetic backgrounds and nutrient sources. Third, some documented substrates of known acetyltransferases and deacetylases were not identified. To address these questions, future research using quantitative proteomics of lysine acetylation and improved sensitivity will be required.

It is of interest to note that only one acetyltransferase and one deacetylase have been described in prokaryotes. The identification of such a large number of lysine acetylation substrate proteins raises the possibility of additional enzymes in bacteria that regulate the modification status. The recent discovery of a yeast acetyltransferase sharing no significant sequence homology with existing acetyltransferases further supports such a possibility (Driscoll et al., 2007).

The results presented here provide a large number of future research opportunities in prokaryotic biology. The large datasets will provide protein leads for further genetic and molecular biology studies to test roles of lysine acetylation sites in cellular physiology. Given the essential roles of lysine acetylation in

diverse metabolic pathways and importance of acetyl Co-A and NAD (the co-factors for lysine acetylation regulatory enzymes) in cellular metabolism, the modification is likely to play an important role in bacterial physiology. Therefore, understanding, capturing, and optimization of the modification landscape might provide a novel approach for the engineering of industrial bacterial species with high efficiency. The bioterror threat posed from certain bacteria (such as anthrax) and the emergence of drug resistant bacteria remind us that the biology of prokaryotes remains to be further studied. Unfortunately, post-translational modifications in prokaryotes have been overlooked in the past. Generation of lysine-acetylation datasets reminds us of the need for further studies of modification pathways in prokaryotes, some of which might provide good avenues for therapeutic intervention.

METHODS

Materials. The reagents used in this work include the affinity-purified anti-acetyllysine polyclonal antibody from ImmuneChem Pharmaceuticals Inc. (Burnaby, British Columbia, Canada) and anti-acetyllysine monoclonal antibody from Cell Signaling Technology (Boston, MA); Protein A conjugated agarose beads from Amersham Biosciences (Uppsla, Sweden); Luria-Bertani (LB) medium from Life Technologies (Carlsbad, CA); iodoacetamide, C_{18} ZipTips

from Millipore Corp. (Bedford, MA); Luna C_{18} resin from Phenomenex (Torrance, CA).

Preparation of cell lysate from *E. coli*. *E. coli* DH5 was grown aerobically in LB medium at 37 °C. The cultured cells were harvested at log phase by centrifugation at 4500 × *g* for 10 min and washed twice by resuspension of the pellet in ice-cold PBS buffer (0.1 M Na_2HPO_4, 0.15 M NaCl, pH 7.2). The cells were resuspended in chilled lysis buffer (50 mM Tris-Cl, pH 7.5, 100 mM NaCl, 5 mM DTT) and then sonicated with 12 short bursts of 10 s followed by intervals of 30 s for cooling. Unbroken cells and debris were removed by centrifugation at 4 °C for 30 min at 21,000 × *g*. The supernatant was divided into aliquots and stored at -80 °C until use.

In-solution tryptic digestion. Five milligrams of proteins were precipitated with acetone, followed by centrifugation at 22,000 × *g* for 10 min. The resulting pellet was digested according to a procedure previously described (Kim et al., 2006a). Briefly, the protein pellet was rinsed twice with cold acetone to remove residual salts, resuspended in 50 mM NH_4HCO_3 (pH 8.5) (the protein pellet was not completely resolved but suspected in small particles), and digested with trypsin (Promega, Madison, WI) at an enzyme-to-substrate ratio of 1:50 for 16 h at 37 °C. The tryptic peptides were reduced with 5 mM DTT at 50°C for 30

min and then alkylated using 15 mM iodoacetamide at RT for 30 min in darkness. The reaction was quenched with 15 mM cysteine at RT for 30 min. To ensure complete digestion, additional trypsin at an enzyme-to-substrate ratio of 1:100 was added to the peptide mixture and incubated for an additional 3 h.

Affinity purification of lysine-acetylated peptides. The anti-acetyllysine antibodies from ImmuneChem Pharmaceuticals Inc. and Cell Signaling Technology were mixed at a ratio of 1:1 and then immobilized to protein A-conjugated agarose beads at 4-6 mg/ml by incubation at 4°C for 4 h. The supernatant was removed and the beads were washed three times with NETN buffer (50 mM Tris-Cl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% NP40).

The tryptic peptides obtained from in-solution digestion were reconstituted in NETN buffer. Insoluble particles were removed by centrifugation. Affinity purification was carried out by incubating the peptides with 20 µl of anti-acetyllysine antibody-protein A-immobilized agarose beads at 4 °C for 6 h with gentle shaking. The beads were washed three times with 1 ml of NETN buffer and twice with ETN (50 mM Tris-Cl, pH 8.0, 100 mM NaCl, 1 mM EDTA). The bound peptides were eluted from the beads by washing three times with 50 µl of 1% TFA. The eluates were combined and dried in a SpeedVac. The resulting peptides were cleaned with C18 ZipTips (Millipore, Bedford, MA) according to the manufacturer's instructions, prior to nano-HPLC/mass spectrometric analysis.

HPLC/MS/MS analysis. HPLC/MS/MS analysis was performed in an integrated system that includes an Agilent 1100 series nanoflow LC system (Agilent, Palo Alto, CA) and a LTQ 2D trap mass spectrometer (Thermo Electron, Waltham, MA) equipped with a nanoelectrospray ionization (NSI) source. 1 μl of the above tryptic peptides in buffer A (97.95% water/2% acetonitrile/0.05% acetic acid) was manually injected and separated in a capillary HPLC column (11 mm length × 75 μm I.D., 5 μm particle size, 100 Å pore diameter) packed in-house with Luna C18 resin (Phenomenex, Torrance, CA). Peptides were eluted from the column with a gradient of 6.0% to 90% buffer B (90% acetonitrile/9.95% water/0.05% acetic acid) in a 2 h duty cycle. The eluted peptides were electrosprayed directly into the LTQ ion trap mass spectrometer. LC/MS/MS was operated in a data-dependent mode such that the ten strongest ions in each MS scan were subjected to collisionally activated dissociation (CAD) with a normalized CAD energy of 35%.

Protein sequence database search and manual verification. Tandem mass spectra were used to search the NCBI-nr database with the Mascot search engine (version 2.1, Matrix Science, London, U.K.). Trypsin was specified as the proteolytic enzyme and up to 6 missed cleavage sites per peptide were allowed. Carbamidomethylation of cysteine was set as fixed modification and oxidation of

methionine and acetylation of lysine as variable modifications. Charge states of +1, +2 or +3 were considered for parent ions. Mass tolerance was set to ±3.0 Da for parent ion masses and ± 0.6 Da for fragment ion masses. Acetylated lysine containing peptides identified with a Mascot score of 25 were manually verified by the method previously described (Chen et al., 2005).

Local structure analysis of *E. coli* lysine acetylation sites. For each acetylated *E. coli* protein, BLAST was run against a database of domain sequences with known structures from the SCOP90 representative set of ASTRAL compendium (version 1.71). Database size of BLAST was set to the size of the protein nr database as of April 6th (39,280,211,952 letters) to impose a stringent e-value cutoff. Hits with an e-value less than 0.001 were analyzed. We identified homologous structures for 69 of the 91 acetylated proteins (~76%). For these proteins, we mapped the positions of acetylated lysines to the model structures using BLAST local alignments and visually inspected the crystal structures to determine the role of the conserved lysine in substrate and protein binding or catalytic activity.

Figure 4-1

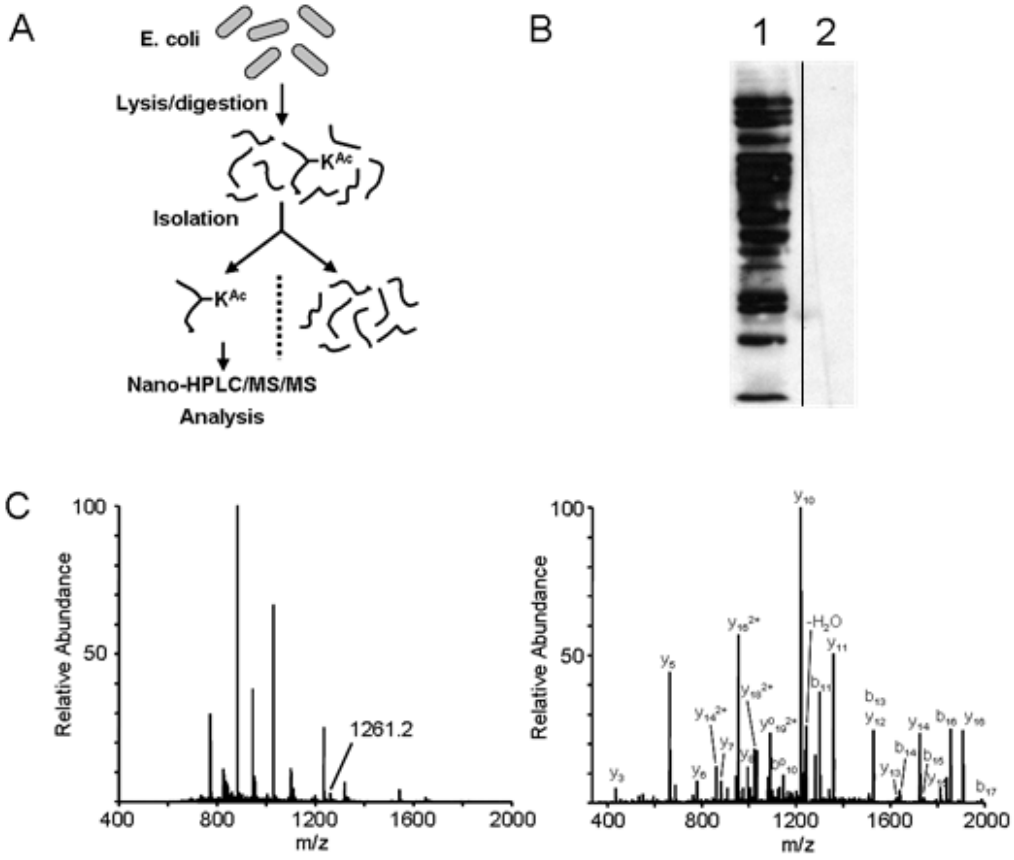

Figure 4-1. **Synopsis of lysine acetylation proteomics.** (A) Schematic representation of the sequential steps used for global profiling of lysine acetylation in *E. coli*. (B) Western blotting analysis of protein lysate from *E. coli* with anti-acetyllysine antibody from ImmunoChem Inc. Lane 1: 30 μg of whole cell lysate from *E. coli* probed with anti-acetyllysine antibody; Lane 2: competition with acetylated BSA (300 μg/ml). (C) An example of MS and MS/MS analysis of a lysine-acetylated peptide for peptide identification and mapping of acetyllysine site in *E. coli*. Left: Full MS spectrum at a retention time of 63.69 min. Right: MS/MS spectrum of m/z 1261.2, which identifies acetylated peptide YYQGTPSPVK*HPELTDMVIFR in isocitrate dehydrogenase, a metabolic protein in *E. coli*. Western blotting performed by Dr. Sung Chan Kim.

Figure 4-2

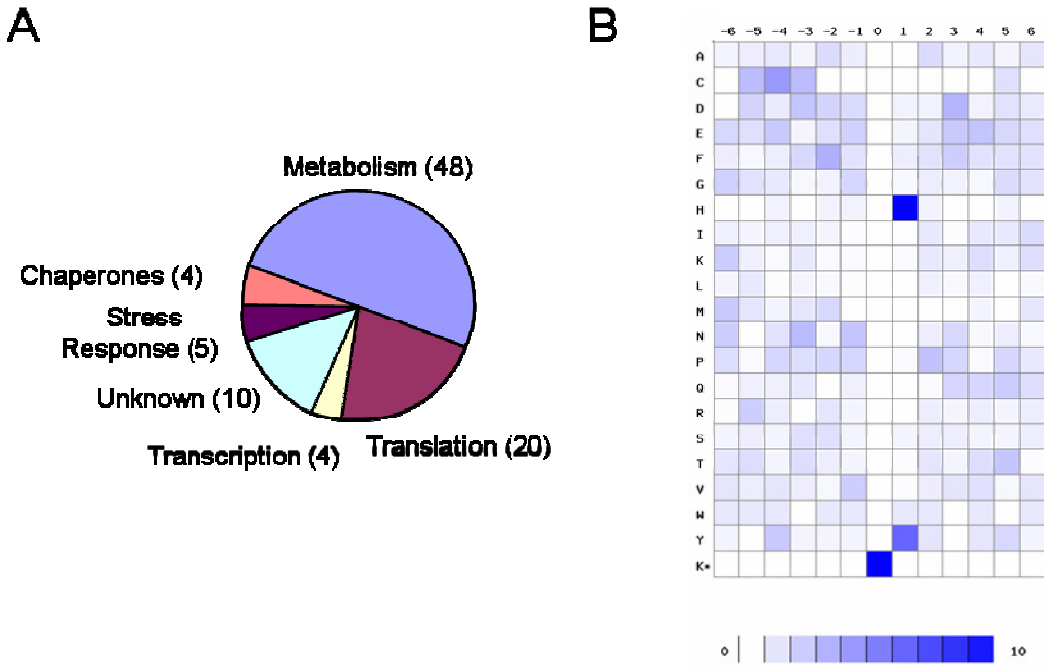

Figure 4-2. **Lysine acetylation in E. coli.** (A) Pie chart of functionally annotated protein groups that are lysine acetylated. (B) Density map of lysine-acetylated peptides. The frequency of occurrence of amino acid residues surrounding sites of lysine acetylation was calculated, relative to the frequency of the residue within the entire *E. coli* genome, and schematically represented by a density map using a method previously described (11). Prevalence of specific amino acids at positions surrounding lysine acetylation sites is shown.

Figure 4-3

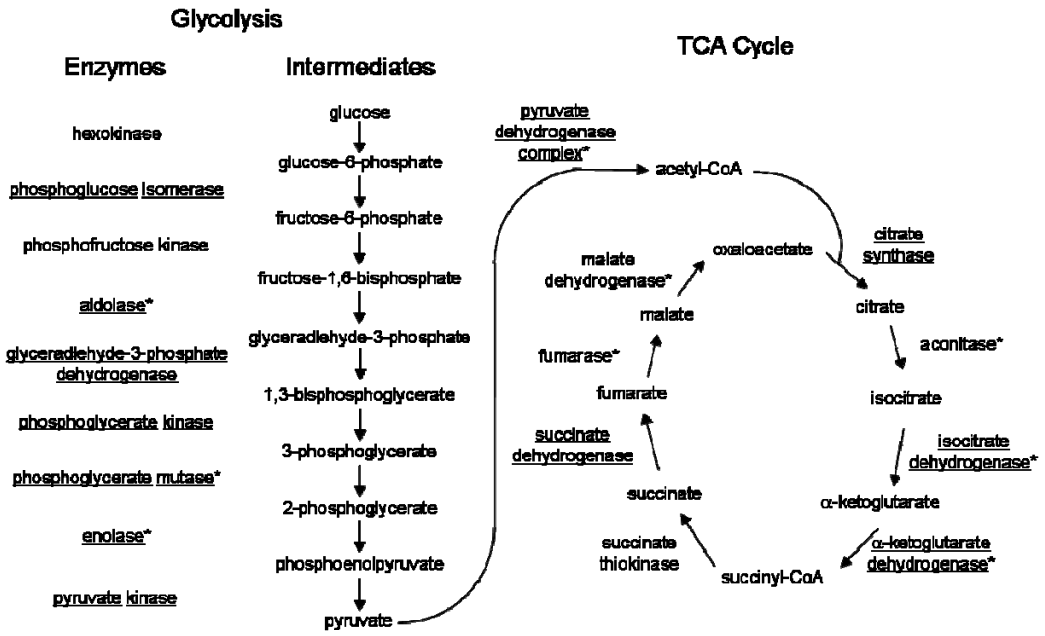

Figure 4-3. Lysine acetylated proteins involved in glucose degradation and TCA cycle. Proteins identified as lysine-acetylated in the *E. coli* screen are underlined. Those identified in a screen of mammalian cells are marked with an asterisk.

Table 4-1

Protein Name	GI No.	Functional Classification	Peptide Sequence	Orthologs in Human
Carbohydrate metabolism				
citrate synthase	gi|16128695	TCA cycle	MLEEISSVKHIPEFVR	*
isocitrate dehydrogenase	gi|2618886	TCA cycle	YYQGTPSPVKHPELTDMWFR	
pyruvate dehydrogenase complex, dehydrogenase component (E1)	gi|83584560	TCA cycle	GIYKLETIEGSK	
dihydrolipoamide acetyltransferase (E2)	gi|26246694	TCA cycle	EDVEKHLAK	*
dihydrolipoamide acyltransferase (E2)	gi|75258892	TCA cycle	VDFSKFGEIEEVELGR	*
dihydrolipoamide dehydrogenase (E3)	gi|26106453	TCA cycle	DIVKVFTK	*
succinate dehydrogenase catalytic subunit	gi|2624691	TCA cycle	GEGGYLLNKHGER	
phosphoglucose isomerase	gi|9664438	glycolysis	AAGDEKHVAK	
phosphoglucose isomerase	gi|9664438	glycolysis	DWFLKAAGDEK	
pgi	gi|6830411 0	glycolysis	ECDLAGAIKSMFSGEK	
fructose-bisphosphate aldolase class II	gi|11064306 9	glycolysis	IFDFVKPGVITGDDVQK	
dehydrin (fructose-bisphosphate aldolase class I)	gi|1658028	glycolysis	NSAFKK	
dehydrin (fructose-bisphosphate aldolase class I)	gi|1658028	glycolysis	VYSKLTSENPDLVR	
enolase	gi|563868	glycolysis	DGKYVLAGEGNK	*
phosphoglycerate kinase	gi|26249339	glycolysis	DKLSNPVR	
phosphoglyceromutase	gi|13360242	glycolysis	AETAEKYGDEQVK	*
phosphoglyceromutase	gi|13360242	glycolysis	HGESQWNKENR	*
phosphoglyceromutase	gi|13360242	glycolysis	HYGALQGLNKAETAEK	*
glyceraldehyde-3-phosphate dehydrogenase	gi|37699654	glycolysis	DNTPMFVKGANFDK	*
glyceraldehyde-3-phosphate dehydrogenase	gi|37699654	glycolysis	GANFDKYAGQDIVSNASCTTNCLAPLAK	*
glyceraldehyde-3-phosphate dehydrogenase	gi|37699654	glycolysis	LEKAATYEQIK	*
glyceraldehyde 3-phosphate dehydrogenase	gi|37699654	glycolysis	TVDGPSHKDWR	*
pyruvate kinase	gi|26248120	glycolysis	IPSINVSKHR	
pyruvate kinase I	gi|147276	glycolysis	GKYPLEAVSIMATICER	*
pyruvate kinase I	gi|147276	glycolysis	TMKLEGGNDVSLK	*
pyruvate formate lyase subunit	gi|16130504	anerobic glycolysis	HPEKYPQLTIR	
pyruvate formate lyase subunit	gi|16130504	anerobic glycolysis	AGYAEDEWAVSKLGDIEYR	
pyruvate formate lyase subunit	gi|16130504	anaerobic glycolysis	ETLEDAVKHPEK	
formate acetyltransferase 1	gi|15800764	anaerobic glycolysis	ALIPFGGIKMIEGSCK	
formate acetyltransferase 1	gi|15800764	anaerobic glycolysis	IQKLHTYR	
formate acetyltransferase 1	gi|15800764	anaerobic glycolysis	TMLYAINGGVDEKLK	
formate acetyltransferase 1	gi|15800764	anaerobic glycolysis	VMEGVKLENR	
formate acetyltransferase 1	gi|15800764	anaerobic glycolysis	VALYGIDYLMKDK	
formate acetyltransferase 1	gi|15800764	anaerobic glycolysis	TMACGIAGLSVAADSLSAIKYAK	
anaerobic Class I fumarate hydratase	gi|146048	anaerobic glycolysis	GGGSANKTYLYQETK	
phosphoenolpyruvate carboxykinase	gi|147113	gluconeogenesis	DTFWMVADKGK	
phosphoenolpyruvate carboxykinase	gi|147113	gluconeogenesis	LFIDNFDKYTDTPAGAALVAAGPK	
transketolase	gi|75227108	carbohydrate metaoolism	DFLKHNPQNPSWADR	*

Protein Name	GI No.	Functional Classification	Peptide Sequence	Orthologs in Human
Carbohydrate metabolism				
NADP-dependent malic enzyme	gi|26109236	carbohydrate metsbolism	DPLKAYK	
transketolase 2, thiamin-binding	gi|16130390	carbohydrate metsbolism	TTQKYINELQANPAK	
phosphopentomutase	gi|1790843	carbohydrate metsbolism	ATGLDALFDATIKEMK	
NAD-dependent aldehyde dehydrogenase	gi|75515146	carbohydrate metsbolism	AEKLVAMGGIGHTSCLYTDQDNQPAR	
mannitol-1-phosphate dehydrogenase	gi|46095211	carbohydrate metsbolism	YGFDADKHAAYIQK	
6-phosphogluconolactonase	gi|16128735	carbohydrate metsbolism	GFNVDHSGKYLIAAGQK	
phosphorylase, maltodextrin	gi|224195	carbohydrate metsbolism	AEQQGINAEKLTK	
mannose-6-phosphate isomerase	gi|1742663	carbohydrate metsbolism	AGLTPKYIDIPELVANVK	*
fused mannose-specific PTS enzymes: IIA component/IIB component	gi|1788120	carbohydrate metabolism (phosphotransferase system)	VYNNPKYAGER	
fused mannose-specific PTS enzymes: IIA component/IIB component	gi|1788120	carbohydrate metabolism (phosphotransferase system)	YNAQLAKLDTTK	
nucleoside-diphosphate-sugar epimerase	gi|75229016	carbohydrate metabolism, outer membrane biosynthesis	GQIEYIPFPDKLK	
ADP-heptose synthase	gi|26249631	lipopolysaccharide biosynthesis	AGVPVLIDPKGTDFER	
Translation				
alanyl-tRNA synthetase	gi|110642817	tRNA synthesis	AGGKHNDLENVGYTAR	*
isoleucyl-tRNA synthetase	gi|75189130	tRNA synthesis	GAKPVHWCVDCR	*
lysyl-tRNA synthetase (lysU)	gi|146689	tRNA synthesis	ALRPLPDKFHGLQDQEVR	
lysyl-tRNA synthetase (class II)	gi|75231274	tRNA synthesis	DSLPEGVYNDQFKK	
lysyl-tRNA synthetase (class II)	gi|75231274	tRNA synthesis	LKEYQYQEVK	
Tyrosyl-tRNA synthetase	gi|83587972	tRNA synthesis	DIGKHFSVNQMINK	
ribosomal protein L5	gi|146574	ribosome	AKLHDYYK	
ribosomal protein L6	gi|229631	ribosome	TLNDAVEVKHADNTLTFGPR	
protein L9	gi|223572	ribosome	AGDEGKLFGSIGTR	
ribosomal protein L10	gi|223035	ribosome	ANAKFEVK	
ribosomal protein L10	gi|223035	ribosome	GVTVDKMTELR	
50S ribosomal subunit protein L2	gi|85676724	ribosome	NFGKHPVTPWGVQTK	*
50S ribosomal subunit protein L31	gi|85676124	ribosome	DIHPKYEEITASCSCGNVMK	
50S ribosomal subunit protein L36	gi|85676742	ribosome	VICSAEPKHK	
S1 ribosomal protein	gi|42900	ribosome	ANPWQQFAETHNKGDR	
S1 ribosomal protein	gi|42900	ribosome	VKHPSEIVNVGDEITVK	
S1 ribosomal protein	gi|42900	ribosome	YPEGTKLTGR	
30S ribosomal protein S6	gi|133976	ribosome	TKHAVTEASPMVK	
ribosomal protein S12	gi|11120595	ribosome	GALDCSGVKDR	*
ribosome-associated protein Y (PSrp-1)	gi|75513204	ribosome	EPQGFVADATINTPNGVLVASGKHEDMYTAINELINK	
ribosome recycling factor	gi|67472290	translation	SQDDVQKLTDAAIK	
elongation factor G	gi|110643581	translation	GMLKGQESEVTGVK	
elongation factor G	gi|110643581	translation	VTDVEGKHAK	
elongation factor Tu	gi|26249935	translation	FESEVYILSKDEGGR	*
translation initiation factor IF2-3	gi|12053635	translation	SPKFGAIAGCMVTEGVVK	*

Protein Name	GI No.	Functional Classification	Peptide Sequence	Orthologs in Human
Amino acid metabolism				
glutamate decarboxylase isozyme	gi|15804061	amino acid metabolism	ASLKYLSDHPK	
glutamate decarboxylase isozyme	gi|15804061	amino acid metabolism	LQGIAQQNSFKHT	
glutamate decarboxylase isozyme	gi|15804061	amino acid metabolism	YLSDHPKLQGIAQQNSFK	
serine hydroxymethyltransferase	gi|26248915	amino acid metabolism	EAMEPEFKTYQQQVAK	*
serine hydroxymethyltransferase	gi|26248915	amino acid metabolism	GFKEAEAK	*
serine hydroxymethyltransferase	gi|26248915	amino acid metabolism	GGSEELYKK	*
serine hydroxymethyltransferase	gi|26248915	amino acid metabolism	NSVPNDPKSPFVTSGIR	*
serine hydroxymethyltransferase	gi|26248915	amino acid metabolism	RGFKEAEAK	*
serine hydroxymethyltransferase	gi|26248915	amino acid metabolism	VMQAQGSQLTNKYAEGYPGK	*
tryptophanase	gi|41936	amino acid metabolism	ENAANIKGLTFTYEPK	
tryptophanase	gi|41936	amino acid metabolism	ENFKHLPEPFR	
tryptophanase	gi|41936	amino acid metabolism	GAEQIYIPVLIKK	
tryptophanase	gi|41936	amino acid metabolism	MENFKHLPEPFR	
tryptophanase	gi|41936	amino acid metabolism	NVYIKEAFDTGVR	
3-deoxy-D-arabino-heptulosonate-7-phosphate synthase	gi|1651339	amino acid metabolism	EPNYSAKHVAEVK	
glutaminase	gi|26246500	amino acid metabolism	GQKMVASVAK	*
aminoacyl-histidine dipeptidase	gi|26246281	protein metabolism	EKNLALLLDSVANDK	
S-adenosylmethionine synthetase II	gi|146851	coenzyme metabolism	AKHLFTSESVSEGHPDK	
Stress Response				
iron-containing superoxide dismutase	gi|5902908	antioxidant	GTAFEGKSLEEIIR	
alkyl hydroperoxide reductase, C22 subunit	gi|1786822	oxidative stress response	AAQYVASHPGEVCPAKWK	*
alkyl hydroperoxide reductase, C22 subunit	gi|1786822	oxidative stress response	IKAAQYVASHPGEVCPAK	*
alkyl hydroperoxide reductase, C22 subunit	gi|1786822	oxidative stress response	IKYAMIGDPTGALTR	*
alkyl hydroperoxide reductase, C22 subunit	gi|1786822	oxidative stress response	NQAFKNGEFIEITEK	*
alkyl hydroperoxide reductase, C22 subunit	gi|1786822	oxidative stress response	WKEGEATLAPSLDLVGK	*
alkyl hydroperoxide reductase, large subunit	gi|75189646	NADH-dependent, ROS regulator	GVTYCPHCDGPLFKGK	
alkyl hydroperoxide reductase, large subunit	gi|75189646	NADH-dependent, ROS regulator	VTFKEDNSLPVR	
thioredoxin	gi|148071	redox homeostasis	LNIDQNPGTAPKYGIR	*
hypothetical protein yibN	gi|26110687	rhodanese-like	ANNVGELEKHK	
Chaperones				
GroEL	gi|99867128	chaperone	AVAAGMNPMDLKR	
osmC	gi|42181	chaperone	GQAHWEGDIKR	
	gi|110642755	chaperone	VLNLCDKLAQK	
ClpB protein	gi|110642755	chaperone	AEQGKLDPVIGR	
ClpB protein	gi|110642755	chaperone	IDMSEFMEKHSVSR	
molecular chaperone	gi|75235743	HSP70 family	AKLESLVEDLVNR	*
molecular chaperone	gi|75235743	HSP70 family	GQKMAPPQISAEVLK	*
molecular chaperone	gi|75235743	HSP70 family	LINYLVEEFKK	*

Protein Name	GI No.	Functional Classification	Peptide Sequence	Orthologs in Human
Chaperones				
molecular chaperone	gi|75235743	HSP70 family	NTTIPTKHSQVFSTAEDNQSAVTIHVLQGER	*
molecular chaperone	gi|75235743	HSP70 family	QVEEAGDKLPADDK	*
Nucleotide Metabolism				
purine nucleoside phosphorylase	gi|15804956	nucleotide salvage pathway	AKYIAETFLEDAR	
Adenylate kinase	gi|1773156	nucleotide metabolism	EAEAGNTKYAK	*
deoxyribosephosphate aldolase	gi|537221	nucleotide metabolism	AGADFIKTSTGK	
IMP dehydrogenase	gi|146275	nucleotide metabolism	AKYPDLQIIGGNVATAAGAR	*
IMP dehydrogenase	gi|146275	nucleotide metabolism	LKEIIHQQMGGLR	*
IMP cyclohydrolase	gi|26250778	nucleotide metabolism	EFAEPACVIVKHANPCGVAIGNSILDAYDR	*
ribonucleotide reductase, alpha subunit	gi|75259700	nucleotide metabolism	GGEAFHTGCIPFYKHFQTAVK	
glycinamide ribonucleotide transformylase	gi|16129802	purine biosynthesis	SAEQLAQAWKYAQQGGR	
adenylosuccinate lyase	gi|145203	purine biosynthesis	GVSKLEVNR	
adenylosuccinate lyase	gi|145203	purine biosynthesis	TTNHDVKAVEYFLK	
Transcription				
rpoB	gi|42818	RNA polymerase beta	QNQLEQLAEQYDELKHEFEK	
rpoB	gi|42818	RNA polymerase beta	ELLKLGDLPTSGQIR	
polymerase beta,RNA	gi|223476	RNA polymerase beta	NTELKLIDEFGR	
trp repressor binding protein	gi|148264	transcriptional corepressor	VPETMPPQLFEKAGGK	
Crp variant	gi|73671324	transcription factor	TACEVAEISYKK	
Unknown				
uncharacterized protein	gi|83586399		VTKYEQR	
beta-galactosidase, beta subunit	gi|75198101		GKHDIEGNR	
hypothetical protein ycjX	gi|26108059		LAQADKHTNAGMLR	
hypothetical oxidoreductase yghA	gi|26109978		MTPVPDCGEKTYVGSGR	
hypothetical protein Ecol5_01000089	gi|75516030		NPQKNLYTFK	
hypothetical protein yiiU	gi|26111125		SLEVFEKLEAK	
exoribonuclease R	gi|75242432		SMKQAIYDPENR	
conserved hypothetical protein	gi|1799623		SSGDPADQKYVELK	
hypothetical protein	gi|12517823		SYLTVAIGCTGGKHR	
o137	gi|148172		YFDNKHYPR	

Table 4-1. List of lysine-acetylated proteins identified from a proteomic screen of E. coli. Protein name, gi#, peptide sequence and annotated biological process associated with the protein are indicated. The presence of an orthologous protein in humans, possessing >25% sequence identity, is also indicated. K designates the lysine residue identified as being acetylated.

Table 4-2

Protein Name	gi	Functional Group	# Sites
citrate synthase	gi|16128695	TCA cycle	1
isocitrate dehydrogenase	gi|2618886	TCA cycle	1
pyruvate dehydrogenase complex, dehydrogenase component	gi|83584560	TCA cycle	1
dihydrolipoamide dehydrogenase	gi|26106453	TCA cycle	1
dihydrolipoamide acetyltransferase	gi|26246694	TCA cycle	1
dihydrolipoamide acyltransferase (E2)	gi|75258892	TCA cycle	1
succinate dehydrogenase catalytic subunit	gi|26246691	TCA cycle	1
phosphoglucose isomerase	gi|9664438	glycolysis	2
pgi	gi|68304110	glycolysis	1
fructose-bisphosphate aldolase class II	gi|110643069	glycolysis	1
dehydrin (fructose-bisphosphate aldolase class I)	gi|1658028	glycolysis	2
enolase	gi|563868	glycolysis	1
phosphoglycerate kinase	gi|26249339	glycolysis	1
phosphoglyceromutase	gi|13360242	glycolysis	3
glyceraldehyde-3-phosphate dehydrogenase	gi|37699654	glycolysis	4
pyruvate kinase	gi|26248120	glycolysis	1
pyruvate kinase I	gi|147276	glycolysis	2
pyruvate formate lyase subunit	gi|16130504	anerobic glycolysis	3
formate acetyltransferase 1	gi|15800764	anaerobic glycolysis	6
anaerobic class I fumarate hydratase	gi|146048	anaerobic glycolysis	1
phosphoenolpyruvate carboxykinase	gi|147113	gluconeogenesis	2
transketolase	gi|75227108	carbohydrate metabolism	1
NADP-dependent malic enzyme	gi|26109236	carbohydrate metabolism	1
transketolase 2, thiamin-binding	gi|16130390	carbohydrate metabolism	1
phosphopentomutase	gi|1790843	carbohydrate metabolism	1
NAD-dependent aldehyde dehydrogenase	gi|75515146	carbohydrate metabolism	1
mannitol-1-phosphate dehydrogenase	gi|46095211	carbohydrate metabolism	1
6-phosphogluconolactonase	gi|16128735	carbohydrate metabolism	1
phosphorylase,maltodextrin	gi|224195	carbohydrate metabolism	1
mannose-6-phosphate isomerase	gi|1742663	carbohydrate metabolism	1
fused mannose-specific PTS enzymes: IIA component/IIB component	gi|1788120	carbohydrate metabolism	2
nucleoside-diphosphate-sugar epimerase	gi|75229016	carbohydrate metabolism	1
ADP-heptose synthase	gi|26249631	lipopolysaccharide biosynthesis	1
glutamate decarboxylase isozyme	gi|15804061	amino acid metabolism	3
serine hydroxymethyltransferase	gi|26248915	amino acid metabolism	6
tryptophanase	gi|41936	amino acid metabolism	5
3-deoxy-D-arabino-heptulosonate-7-phosphate synthase	gi|1651339	amino acid metabolism	1
glutaminase	gi|26246500	amino acid metabolism	1
aminoacyl-histidine dipeptidase	gi|26246281	protein metabolism	1
S-adenosylmethionine synthetase II	gi|146851	coenzyme metabolism	1
purine nucleoside phosphorylase	gi|15804956	nucleotide salvage pathway	1
adenylate kinase	gi|1773156	nucleotide metabolism	1
IMP dehydrogenase	gi|146275	nucleotide metabolism	2
IMP cyclohydrolase	gi|26250778	nucleotide metabolism	1
ribonucleotide reductase, alpha subunit	gi|75259700	nucleotide metabolism	1
deoxyribosephosphate aldolase	gi|537221	nucleotide metabolism	1
glycinamide ribonucleotide transformylase	gi|16129802	purine biosynthesis	1
adenylosuccinate lyase	gi|145203	purine biosynthesis	2

Table 4-2. A list of metabolic proteins identified as acetylated in *E. coli*.
Protein name, gi number, functional classification and number of acetylation sites identified are indicated.

CHAPTER FIVE

PROTEIN PROPIONYLATION AND BUTYRYLATION

SUMMARY

The positively charged lysine residue plays an important role in protein folding and functions. Neutralization of the charge often has a profound impact on the substrate proteins. Accordingly, all the known post-translational modifications at lysine have pivotal roles in cell physiology and pathology. This chapter concerns the discovery of two novel, *in vivo* lysine modifications among histones, lysine propionylation and butyrylation. In addition, the ability of two previously identified acetyltransferases, p300 and CBP, to catalyze *in vitro* lysine propionylation and lysine butyrylation in histones was demonstrated. Finally, p300 and CBP could carry out autopropionylation and autobutyrylation *in vitro*. Taken together, these results conclusively establish that lysine propionylation and lysine butyrylation are novel post-translational modifications. Given the unique roles of propionyl-CoA and butyryl-CoA in energy metabolism and the significant structural changes induced by the modifications, the two modifications are likely to have important, but distinct functions in the regulation of biological processes.

INTRODUCTION

Dissection of the molecular networks of post-translational modifications that regulate cellular processes and disease progression stands as one of the major goals of post-genomic biological research. To date, more than 200 post-translational modifications have been described, providing an efficient way to diversify a protein's primary structure and possibly its functions. The remarkable complexity of these molecular networks is exemplified by modifications at the side chain of lysine. The electron-rich and nucleophilic nature of the lysine side chain makes it suitable for undergoing covalent post-translational modification reactions with diverse substrates. The residue can be potentially modulated by several post-translational modifications including methylation, acetylation, biotinylation, ubiquitination, and sumoylation, which have pivotal roles in cell physiology and pathology.

Acetyl-CoA, a member of the high-energy CoA compounds, is the substrate used by acetyltransferases to catalyze the lysine-acetylation reaction. However, it remains unknown whether cells can use other short-chain CoAs, such as propionyl- and butyryl-CoA (which are structurally similar to acetyl-CoA), to carry out similar post-translational modifications at lysine. Nevertheless, several lines of evidence suggest such a possibility. First, like acetyl-CoA, propionyl-CoA and butyryl-CoA are high-energy molecules, making a reaction with a lysine

side chain thermodynamically feasible. Second, propionyl-CoA and butyryl-CoA are structurally similar to acetyl-CoA, with a difference of only one or two CH_2. Third, propionyl-CoA and butyryl-CoA are present at high concentrations in cells. In the case of starved mouse liver, the two CoAs' concentrations are only 1-3 times less than acetyl-CoA (King and Reiss, 1985). Finally, it appears, from structural studies on some HATs (such as Hat1), that the enzyme has ample space within the cofactor binding pocket to accept propionyl-CoA without steric interference (Dutnall et al., 1998). Despite such evidence, the short-chain CoAs with the exception of acetyl-CoA have not been described as a substrate for protein modification.

The following work describes the identification and validation of two novel post-translational protein modifications, propionylation and butyrylation at lysine residues, by a proteomics study. The unbiased global screening involved exhaustive peptide identification by nano-HPLC/MS/MS analysis, protein sequence database search, and manual verification. The resulting propionylated and butyrylated peptides were verified by MS/MS of their corresponding synthetic peptides. Using *in vitro* labeling with isotopic propionyl-CoA and butyryl-CoA as well as mass spectrometry, it was verified that the acetyltransferases p300 and CBP could perform lysine modifications on histones using these substrates. Furthermore, it was demonstrated that p300 and CBP could carry out autopropionylation and autobutyrylation at lysine residues in a similar

fashion as autoacetylation. Taken together, these results reveal that lysine propionylation and butyrylation are novel lysine modifications that can be catalyzed by acetyltransferases. Given the unique roles of propionyl-CoA and butyryl-CoA in energy metabolism (Nelson and Cox, 2005), their distinct structure, and significant structural changes induced by the modifications, it is anticipated that lysine propionylation and butyrylation will have important, but likely distinct functions in the regulation of biological processes. The identification of lysine-propionylated and lysine-butyrylated substrates described here provides an entry point for future functional studies of the two modifications in cellular physiology and pathology.

RESULTS

Initial Identification of Lysine-propionylated and -butyrylated Peptides in Histone H4

Given the close structural similarity between acetyllysine residue and the proposed propionyllysine residue it was hypothesized that propionyl-CoA could be used by acetyltransferases for lysine propionylation (Figure 5-1). In addition, it was further assumed that some propionylated, tryptic peptides could be affinity-purified by anti-acetyllysine antibody due to this structural similarity. To identify lysine-propionylated peptides, we searched the MS/MS datasets of affinity-

enriched acetyllysine-containing tryptic peptides acquired by nano-HPLC/LTQ mass spectrometry (Kim et al., 2006b). During the protein sequence database search, the lysine was considered as unmodified, acetylated, or propionylated. The database search and manual verification of peptide hits led to the identification of eleven lysine-propionylated histone H4 peptides (Table 5-1). The successful identification of putative lysine-propionylated peptides motivated a subsequent search for lysine-butyrylated peptides. The same datasets were searched again, with the lysine considered as unmodified, acetylated, or butyrylated. The analysis identified two additional histone H4 peptides with lysine butyrylation sites (Table 5-1).

A representative MS/MS spectrum demonstrates the identification of lysine-propionylated and lysine-butyrylated peptides (Figure 5-2). In the initial peptide identification by protein sequence database search, the spectrum in Fig. 4-2A led to the identification of a lysine-propionylated peptide 1 (Table 5-1). Careful inspection of the spectrum revealed that there were a number of fragment ions with strong intensities that could not be assigned to the fragment ions derived from the predicted peptide sequence. This result suggested the presence of an additional peptide(s). Three lines of evidence suggest that the spectrum was from peptide isomers: (i) multiple peak-pairs with mass difference of 14 Da were observed; (ii) the peptides possess identical molecular weights; and (iii) the peptides co-eluted. Indeed, the remaining peaks in spectrum Figure 5-2A could be

accounted for by two additional lysine-butyrylated peptides: Peptide 12 and Peptide 13 (Table 5-1).

The chemical nature of an identified peptide can be confirmed by MS/MS of their corresponding synthetic peptides, the standard for verification of peptide identification and chemical identity. To validate the identities of the propionylated and butyrylated peptides in Figure 5-2A, MS/MS of 3 synthetic peptides with sequences and modification patterns corresponding to Peptides 1, 12 and 13 in Table 5-1 were analyzed. In order to validate the interpretation of the fragmentation spectrum observed in Figure 5-2A, which was obtained from the cell-derived peptide mixture, three synthetic peptides were mixed at a molar ratio of 4:2:1 (Peptide 1: Peptide 12: Peptide 13). MS/MS analysis of the peptide mixture resulted in a fragmentation pattern (Figure 5-2B) highly similar to the *in vivo* spectrum in Figure 5-2A, verifying the identification of one lysine-propionylated peptide and two lysine-butyrylated peptides.

Together, *in vivo* lysine propionylation was identified at K5, K8, and K12, while *in vivo* lysine butyrylation was found at K5 and K12 of histone H4 (Table 5-1). K5, K8, and K12 of histone H4 are known to be acetylated, while K12 is also subject to lysine methylation (www.histone.com). Lysine acetylation at the four H4 lysine residues is associated with transcriptional activation, transcriptional silencing, chromatin high-order structure, and DNA repair (Peterson and Laniel, 2004; Shia et al., 2006). Some of the acetyllysine residues

(e.g., K8 of histone H4) provide a docking site to recruit a bromodomain-containing chromatin remodeling enzyme SWI/SNF (Agalioti et al., 2002). While biological functions of lysine propionylation and butyrylation in histones remain unknown, it is intriguing to speculate that propionyllysine or butyryllysine would be involved in the interaction/recruiting of a distinct set of proteins/enzymes to control chromatin structure and transcriptional activities.

Propionylation and Butyrylation of Core Histones can be Catalyzed by p300/CBP

Since histone H4 can be propionylated and butyrylated *in vivo*, it was next tested whether core histones could be propionylated and butyrylated *in vitro* by acetyltransferases, using either ^{14}C-propionyl CoA or ^{14}C-butyryl CoA. Five acetyltransferases were tested, CBP, p300, Tip60, MOF and PCAF. CBP and p300 are known acetyltransferases for K5, K8, K12, and K16 of histone H4.

The core histones were incubated with one of the CoAs and the acetyltransferase of interest. The protein mixture was then resolved in SDS-PAGE and visualized by autoradiography. CBP and p300 showed significant activities in catalyzing modifications on histone H4 (Figure 5-3A). On the other hand, no significant propionylation or butyrylation products were detected for the other three acetyltransferases, Tip60, MOF, and PCAF.

To corroborate evidence of an *in vitro* modification reaction at lysine residues, nano-HPLC/mass spectrometric analysis was used to map the CBP-catalyzed, lysine-modified residues in histone H4. K5, K8, K12, K16, K31, K44, K77, K79 and K91 were found to be both propionylated and butyrylated by CBP. Together, these data establish that histone H4 can be lysine-propionylated and -butyrylated directly by CBP and p300 *in vitro*.

p53 can be Propionylated and Butyrylated by p300/CBP *in vitro*

CBP and p300, co-activators of p53, augment its transcriptional activity and modulate its biological functions (Avantaggiati et al., 1997; Gu et al., 1997). Multiple lysine residues in p53 can be acetylated, some of which are known to be modified by CBP/p300 (Gu et al., 1997; Sakaguchi et al., 1998; Tang et al., 2006). Given the fact that CBP/p300 are the acetyltransferases for p53 and that they have enzymatic activities for lysine propionylation and butyrylation in histones, it was hypothesized that the HATs could catalyze similar reactions in p53. To test this hypothesis, *in vitro* propionylation and butyrylation reactions were carried out using p53 as the target, employing the procedure described above. Again, only two of the five acetyltransferases, CBP and p300, could carry out propionylation and butyrylation reactions at p53 at a significant reaction rate under our experimental conditions (Figure 5-3B). Interestingly, p300 shows an

apparently higher catalytic activity than CBP for p53. In contrast, the two enzymes have comparable activities toward histones.

CBP and p300 are acetyltransferases that can catalyze autoacetylation reactions. To test if the proteins can carry out autopropionylation and autobutyrylation reactions, the modification sites at p300 and CBP were mapped by mass spectrometry. Twenty-one lysine-propionylation sites and eleven lysine-butyrylation sites were localized in p300, while twelve lysine-propionylation sites and seven lysine-butyrylation sites were mapped in CBP (Table 5-1). Identification of propionylated and butyrylated peptides in non-histone proteins, p53, CBP, and p300, suggests the possibility that the two modifications are not restricted to histones.

DISCUSSION

In summary, this work demonstrated the identification of *in vivo* lysine propionylation and lysine butyrylation in histones. The modifications were validated by synthetic peptides and *in vitro* enzymatic reactions. Testing of five acetyltransferases suggested that two previously known acetyltransferases, CBP and p300, could catalyze lysine modification reactions using either propionyl-CoA or butyryl-CoA, among histones and p53, at a significant reaction rate *in vitro*.

Lysine is one of the two major ribosomally-coded amino acid residues with a positive-charged side chain at physiological pH, playing an important role in protein folding and functions. We do not know yet what will be the functional consequences of the two modifications described here. Nevertheless, given the regulatory functions of the known lysine modifications and the significant structural changes induced by lysine propionylation and butyrylation, it is anticipated that the two modifications are likely to have biological functions in the same fashion as other lysine modifications.

Histones are known to be modified by an array of post-translational modifications, including methylation, acetylation, ubiquitination, small ubiquitin-like modification, and ribosylation (Fischle et al., 2003). A combinatorial array of post-translational modifications in histones, termed the "histone code", dictates the proteins' functions in gene expression and chromatin dynamics (Jenuwein and Allis, 2001). Post-translational modifications of histones have been elegantly studied by both biochemistry (Jenuwein and Allis, 2001) and mass spectrometry. Nevertheless, lysine propionylation and butyrylation at histones have not been reported before. Our results suggest that the complexity of histone modifications remains to be explored.

Acetyl-CoA can arise during the catabolism of sugars, fatty acids and amino acids. Propionyl-CoA derives only from odd-chain fatty acid and amino acid catabolism, while butyryl-CoA is a metabolic intermediate formed during the

β-oxidation of fatty acids as well as a substrate for fatty acid elongation. The concentration of the short-chain CoAs fluctuates depending on diet and cellular physiological conditions (King and Reiss, 1985). If the rate of the modifications depends on the concentration of the short-chain CoAs, directly or indirectly, it would be appealing to speculate that lysine propionylation and butyrylation may regulate cellular metabolic pathways in response to cellular physiological conditions. Such a scenario then opens up the potential for the biochemical intermediates thus produced to lead to tissue-specific and environmentally-responsive regulatory programs.

The activities and specificities of regulatory enzymes responsible for propionylation and butyrylation might be influenced by other factors. For some acyltransferases (e.g., acetyltransferases), the dimensions of the CoA-binding pocket might determine the binding affinity to the short-chain CoA. Alternatively, it is also possible that the binding affinity might be modulated by other regulatory molecules, such as association of co-factors or interaction proteins. Finally, identification of lysine propionylation and lysine butyrylation further suggests the possible existence of novel enzymes specific for the modifications (other than the known acetyltransferases).

CONCLUDING REMARKS

Discovery of lysine propionylation and butyrylation raises many interesting questions. What are the enzymes responsible for controlling the status of lysine-propionylation and -butyrylation? Given the close structural similarity, some of the HATs and HDACs might be able to catalyze the reactions. What determines the specificities of a transferase among three types of lysine modifications: acetylation, propionylation, and butyrylation? What are the biological pathways regulated by lysine propionylation and butyrylation? The study described here represents the beginning of future research into addressing the fundamental questions of the novel lysine modifications.

METHODS

In-gel digestion. Protein in-gel digestion, peptide extraction, and peptide cleaning using a μ-C18 ZipTip were carried out as previously reported (Zhao et al., 2004).

HPLC/MS/MS analysis. HPLC/MS/MS analysis for mapping propionylation and butyrylation sites in histone H4, p53, p300, and CBP were carried out by nano-HPLC/LTQ mass spectrometry as previously described (Kim et al., 2006b). Briefly, each tryptic digest was dissolved in 10 μl HPLC buffer A (0.1% formic acid in water (v/v)) and 2 μl were injected into an Agilent HPLC

system (Agilent, Palo Alto, CA). Peptides were separated on a home-made capillary HPLC column (50 mm length X 75 μm ID, 4 μm particle size, 90 Å pore diameter) with Jupiter C12 resin (Phenomenex, St. Torrance, CA) and directly electrosprayed into the mass spectrometer using a nano-spray source. The LTQ mass spectrometer was operated in the data-dependant mode acquiring fragmentation spectra of the ten strongest ions.

Protein sequence database search and manual verification. All MS/MS spectra were searched against NCBI-nr protein sequence database specifying lysine modifications using the MASCOT database search engine. A low cutoff peptide score of 20.0 was selected to maximize the identification of lysine-modified peptides. For each Mascot search, the peptide mass error was set to +/- 4 Da, fragment ion mass error was set to +/- 0.6 Da and six missing cleavages were allowed. All lysine-propionylated or -butyrylated peptides identified with MASCOT score > 20.0 were manually examined with the rules previously described (Chen et al., 2005) and all lysine propionylation or butyrylation sites had to be identified by consecutive b- or y- ions so that the possibility that propionylation (+56 Da) or butyrylation (+70 Da) occurred on adjacent residues was eliminated.

***In vitro* propionylation and butyrylation assay.** *In vitro* propionylation and butyrylation assays were carried out essentially as previously described (Gu and Roeder, 1997) with some modifications. Tagged human FLAG-p300, HA-CBP, FLAG-MOF, and FLAG-PCAF proteins from transfected 293 cells and tagged human GST-Tip60 and GST-p53 expressed in bacteria were purified to homogeneity under stringent conditions (500 mM NaCl + 1% Triton X-100). Ten-microliter reactions contained 50 mM Tris pH7.9, 10% glycerol, 1 mM DTT, 10 mM sodium butyrate, 1 μl of (^{14}C)-acyl-CoA (55 mCi/mmol; acetyl-CoA from Amersham and propionyl-CoA and butyryl-CoA from ARC, Inc.). Two and a half μg of substrates (core histones, recombinant human Histone H4 or GST-p53) and 20 – 100 ng of the enzyme protein, as indicated, were incubated at 30^{o}C for 1 hour. The reaction mixture was then subjected to electrophoresis on SDS-PAGE gels, followed by either autoradiography or Coomassie Blue staining.

Figure 5-1

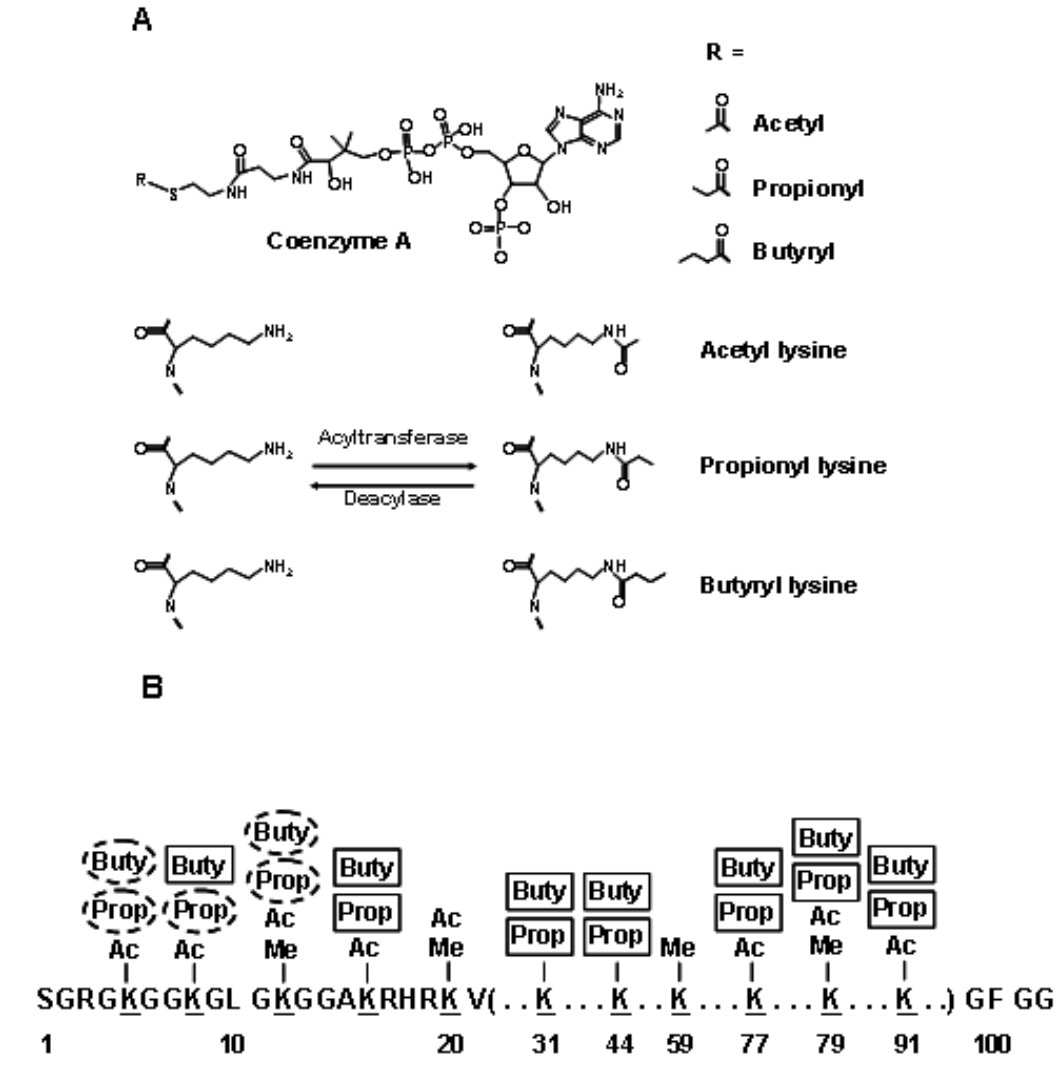
A
R =
Acetyl
Propionyl
Butyryl
Coenzyme A
Acetyl lysine
Acyltransferase
Deacylase
Propionyl lysine
Butyryl lysine
B
SGRGKGGKGL GKGGAKRHRK V(. . K . . . K . . K . . . K . . . K . . . K . . .) GF GG
1
10
20
31
44
59
77
79
91
100

Figure 5-1. (A) Structures of three short-chain CoAs: the acetyl-CoA, propionyl-CoA, and butyryl-CoA, as well as the three modified lysines: acetyllysine, propionyllysine, and butyryllysine. (B) An illustration of novel lysine propionylation and butyrylation sites in histone H4 (unmarked labels: lysine-acetylation and -methylation sites identified previously; circled labels: novel, *in vivo* lysine-modification sites identified in this study; boxed labels: additional novel, *in vitro* lysine-modification sites identified in this study). The known sites of lysine acetylation and methylation were obtained from a histone website (www.histone.com).

Figure 5-2

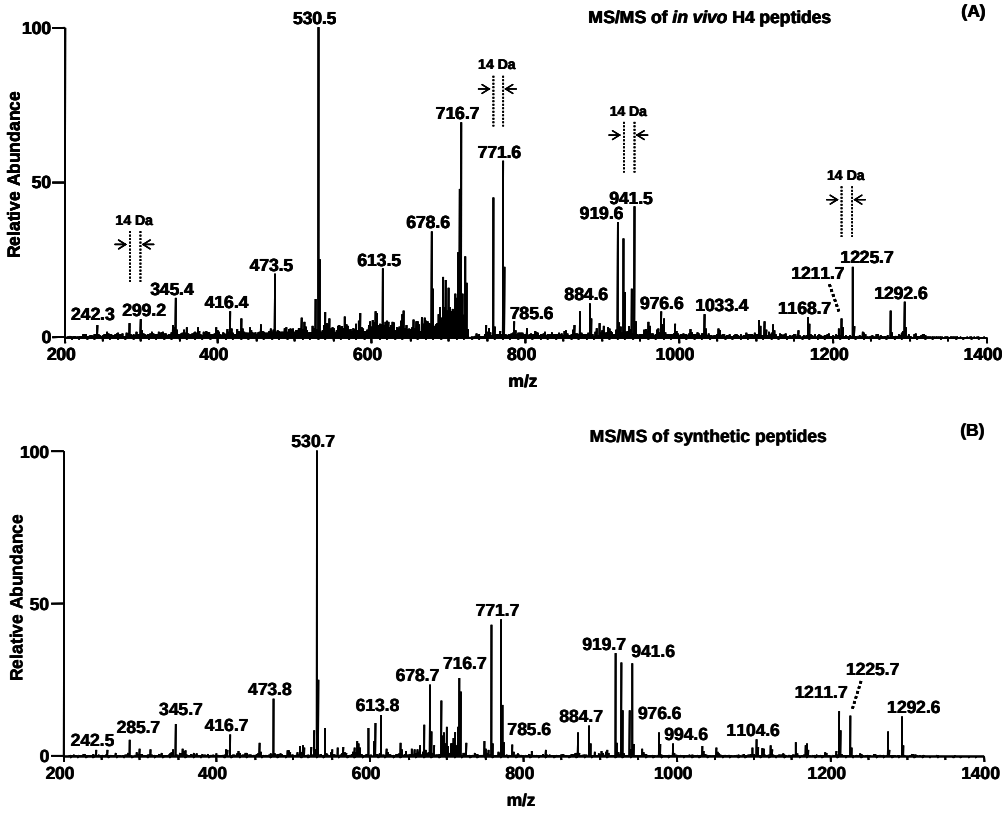

Figure 5-2. Identification and verification of lysine-propionylated and lysine-butyrylated histone H4 peptides. (A) The tandem mass spectrum (MS/MS) of a tryptic peptide ion from a peptide mixture that was affinity-enriched with an anti-acetyllysine antibody from tryptic peptides of HeLa nuclear extracts. The spectrum was used to identify one lysine-propionylated and two lysine-butyrylated peptides: Peptides 1, 12 and 13 (Table 5-1) from histone H4. (B) Tandem mass spectrum of a peptide mixture from the three synthetic peptides corresponding to the sequences identified in Panel A, showing similar ion intensity.

Figure 5-3

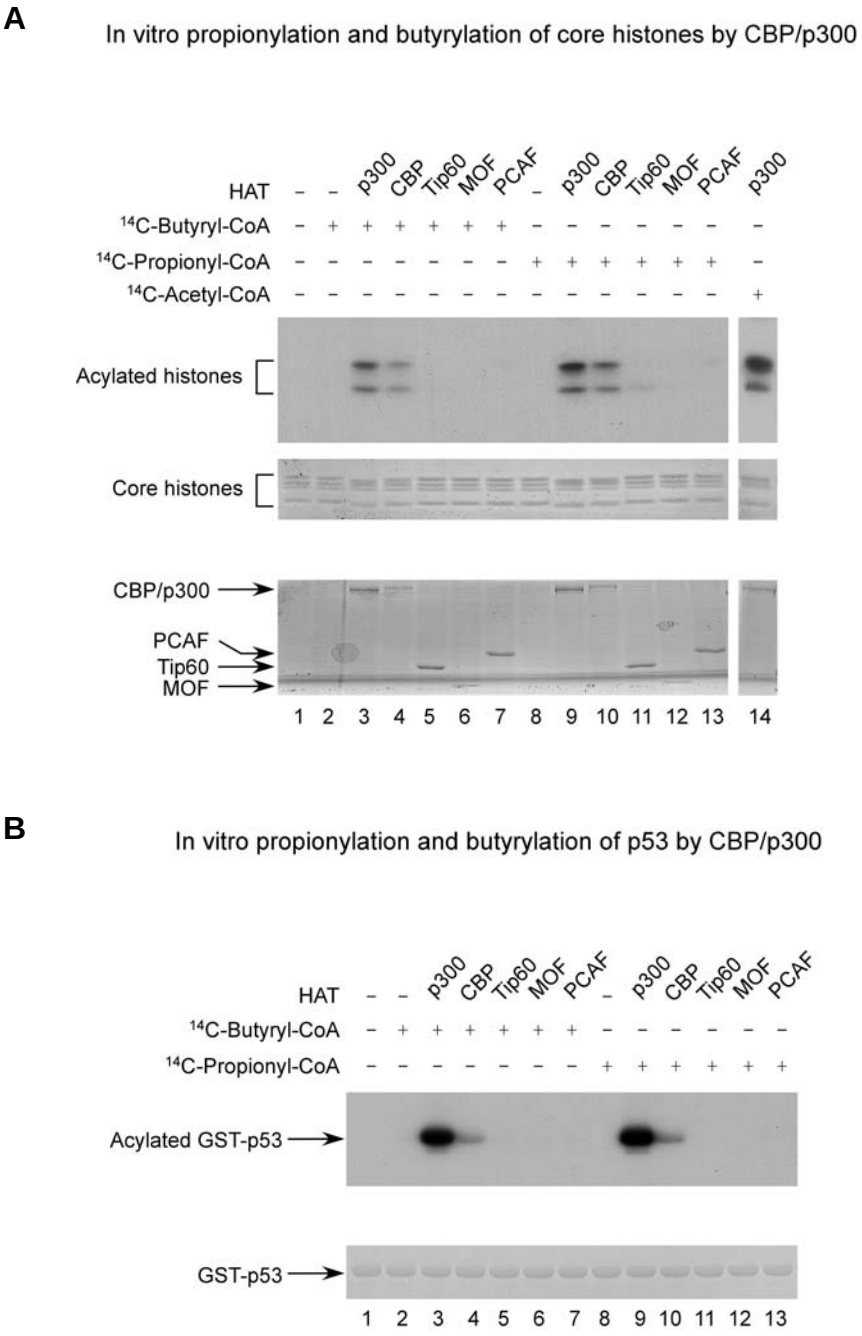
A
In vitro propionylation and butyrylation of core histones by CBP/p300
HAT
p300
CBP
Tip60
MOF
PCAF
14C-Butyryl-CoA
14C-Propionyl-CoA
14C-Acetyl-CoA
Acylated histones
Core histones
CBP/p300
PCAF
Tip60
MOF
1 2 3 4 5 6 7 8 9 10 11 12 13 14
B
In vitro propionylation and butyrylation of p53 by CBP/p300
HAT
p300
CBP
Tip60
MOF
PCAF
14C-Butyryl-CoA
14C-Propionyl-CoA
Acylated GST-p53
GST-p53
1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 5-3. ***In vitro*** **propionylation and butyrylation of core histones and p53 by acetyltransferases**. (A) *In vitro* propionylation and butyrylation of core histones. The core histone proteins were incubated with the purified acetyltransferase in the presence of either (^{14}C)-propionyl-CoA or (^{14}C)-butyryl-CoA, as indicated. The reaction products were resolved by SDS-PAGE and visualized by autoradiography (top panel). The amounts of the core histone substrates and the acetyltransferases are shown in the middle and bottom panels, respectively, by Coomassie Blue staining. (B) *In vitro* propionylation and butyrylation of p53. The GST-p53 recombinant protein was incubated with the purified acetyltransferase in the presence of (^{14}C)-acyl-CoA, as indicated. The reaction mixtures were resolved by SDS-PAGE and visualized by autoradiography (upper panel). The amount of the GST-p53 substrate is shown in the lower panel stained by Coomassie Blue. Experiments performed by the laboratory of Dr. Wei Gu.

Table 5-1

A

No.	Protein Name	gi#	Sequence	No. of Propionyl-Lys Site
1	Histone 4, H4	28173560	GK^GGK*GLGK^GGAK*R	2
2	Histone 4, H4	28173560	GGK^GLGK*GGAK*R	1
3	Histone 4, H4	28173560	GGK*GLGK^GGAK*R	1
4	Histone 4, H4	28173560	GK^GGK*GLGK*GGAK*R	1
5	Histone 4, H4	28173560	GK*GGK^GLGK*GGAK*R	1
6	Histone 4, H4	28173560	GK*GGK*GLGK^GGAK*R	1
7	Histone 4, H4	28173560	GK^GGK*GLGK*GGAK	1
8	Histone 4, H4	28173560	GK*GGK^GLGK*GGAK	1
9	Histone 4, H4	28173560	GK*GGK*GLGK^GGAK	1
10	Histone 4, H4	28173560	GGK*GLGK^GGAK	1
11	Histone 4, H4	28173560	GGK^GLGK*GGAK	1

B

No.	Protein Name	gi#	Sequence	No. of Butyryl-Lys Site
12	Histone 4, H4	28173560	GK"GGK*GLGK*GGAK*R	1
13	Histone 4, H4	28173560	GK*GGK*GLGK"GGAK*R	1

C

In vitro analysis	p53	p300	Histone H4	CBP
Propionyl-Lys	2	21	9	12
Butyryl-Lys	2	11	9	7

Table 5-1. List of lysine-propionylated and butyrylated peptides identified *in vivo* and the number of modification sites identified in four proteins *in vitro*. Protein name, gene index number, modified peptide sequence and the number of modification sites are specified in two lists. (A) A list of lysine propionylated peptides and (B) a list of lysine butyrylated peptides identified *in vivo*. Modified lysine residues are marked as: * - acetylated lysine, ^ - propionylated lysine, " - butyrylated lysine. (C) The total number of lysine-propionylation and butyrylation sites identified in p53, histone H4, p300 and CBP *in vitro*.

CHAPTER SIX

EVIDENCE FOR ABUNDANT ASPARTIC AND GULTAMIC ACID METHYLATION IN EUKARYOTES

SUMMARY

Post-translational methylation has been well studied in the context of reversible lysine and arginine modification and is appreciated for its role in regulating gene expression. Yet, relatively little is known about the extent and consequences of methylation among glutamic and aspartic acid residues, especially in mammalian cells. The millimolar concentration in cells of the source for these methyl groups, S-adenosylmethionine, as well as the high number of methyltransferases, suggests that methylation among aspartic and glutamic acid may be a widespread protein modification. Given the significant alteration in charge and hydrophobicity that result upon methylation of acidic residues, the modification has the potential to be a dynamic regulatory mechanism with unexplored potential. To address this, we applied nano-HPLC/Orbitrap mass spectrometry to conduct a systematic proteomic screening for methylated residues among diverse organisms. From our preliminary data, it is estimated that

methylation at aspartic and glutamic acid in HeLa cells is likely to occur at an abundance greater than tyrosine phosphorylation.

INTRODUCTION

Protein methylation catalyzed by S-adenosyl-L-methionine (SAM)-dependent methyltransferase represents a major group of post-translational modifications (Schubert et al., 2003). Methylation of lysine and arginine at histones is known to play an important role in the regulation of epigenetics and diseases (Strahl and Allis, 2000). However, methylation in other ribosomally coded amino acid residues has not yet been carefully examined. In addition to arginine and lysine, methylation has been found at aspartate (D) and glutamate (E) residues. Glutamate methylation has been shown to modulate chemotactic responses in *E. coli* (Shapiro and Koshland, 1994). Nevertheless, definitive evidence of E-methylation in eukaryotes is lacking and no eukaryotic E-specific methyltransferase has been reported (Clarke, 1993). D-methylation has been reported in both eukaryotic and prokaryotic organisms, where it has been implicated as a protein repair mechanism (Clarke, 1993). However, the sites and substrates of D-methylation *in vivo* remain largely unknown. Given the importance of aspartate and glutamate in protein folding and functions, it is expected that neutralization of the amino acids' negative charge, an event similar to dephosphorylation, would have a significant impact on the proteins in a similar fashion as protein phosphorylation and lysine acetylation.

The work presented here represents a proteomics screening of D- and E-methylation by exhaustive identification of aspartate- and glutamate-methylated (D- and E-methylated) peptides and mapping their corresponding modification sites by nano-HPLC/MS/MS analysis (Figure 6-1). This study identified 2 E-methylation sites from 1 protein in *S. cerevisiae*, and 3 D-methylation substrates in HeLa cells. The methylated peptides were confirmed by either MS/MS of synthetic peptides and *in vivo* labeling. While E-methylation is known among prokaryotes, potential substrates in eukaryotic cells have not been conclusively established. The results suggest that D/E-methylation are abundant post-translational modifications, potentially occurring in more than 1% of proteins in eukaryotic cells. Given the important roles of D/E residues in a protein structure and function, as well as significant structural change induced by methylation, it is expected that D/E methylation is likely to have higher impact on functions of its substrate proteins than lysine and arginine methylation. Therefore, our study reveals previously unappreciated post-translational modifications with functional consequences.

RESULTS

Proteomics Screening of D/E-methylation

D-methylation was initially detected during routine proteomic analyses for mapping post-translational modification sites in individual proteins. It was observed the spectra could often be better explained when allowing for methyl-esterification of D and E residues. To test if the D/E-methylated proteins exist *in vivo*, we carried out a proteomics screening to identify D/E-methylated proteins, and to evaluate the abundance and substrate diversity of the modifications.

Identification of D/E-methylated Proteins

To identify D/E-methylated substrate proteins, a proteomics screening was carried out to identify D/E-methylated peptides from *S. cerevisiae* and HeLa cells, representing two biological species: yeast and mammalian cells.

Identification of D/E-methylated proteins was accomplished by exhaustive HPLC/MS/MS analysis of tryptic digests and protein sequence database search (Figure 6-1). Briefly, forty-μg protein whole-cell lysate from each cell type was resolved by SDS-PAGE. Twenty protein bands of equal size were excised from each sample lane. Each protein band was in-gel digested with trypsin; the resulting tryptic peptides were analyzed by nano-HPLC/Orbitrap mass spectrometer for exhaustive peptide identification. The experimental protocol was designed in such a way that methanol was excluded in the procedure, thereby reducing the possibility for *in vitro* D/E-methylation. The MASCOT search algorithm used the resulting MS/MS spectra to identify peptide candidates from a

protein sequence database. Two variable modifications, D-methylation and E-methylation, were included in the protein sequence database search.

Validation Protocol for Peptide Identification

All the positive identifications were manually inspected to ensure the quality of analysis (Chen et al., 2005). Since methylation could be present in a variety of amino acid residues, including arginine, lysine, histidine and asparagine (Clarke, 1993; Walsh et al., 2005), special care was taken during the manual verification to make sure the methylated residue was located in either aspartate or glutamate residues instead of other amino acid residues. Those methylated peptides with ambiguous methylation sites were excluded from our list. The putative methylated peptides were further confirmed by either MS/MS of synthetic peptides or *in vivo* labeling.

D/E-methylated Peptides Identified

This study on D/E-methylation led to the identification of 1 E-methylated protein from yeast and 3 D-methylated proteins from HeLa lysates (Table 6-1).

Verification of D/E-methylated Peptides

To verify the quality of the peptide identification, we carried out MS/MS of synthetic peptides corresponding to the identified E- or D-methylated peptides,

the standard for verification of peptide identification using mass spectrometry. The complete overlap between MS/MS spectra of the synthetic peptide and the original raw spectrum verifies the peptide identification (Figure 6-2A and B).

D-methylated residues have a molecular weight equivalent to glutamate; therefore, either the MS/MS of a D-methylated peptide or a D-to-E-mutation can explain the same MS/MS spectrum. We took two steps to address this problem. We first examined polymorphisms of the modified D residue. BLAST searching was carried out substituting E for our putative methyl-D identifications to ensure that our identifications did not arise from known sequence variants. Also, since many post-translationally modified peptides are present in low stoichiometry, both modified and unmodified peptides will be present in a sample and their corresponding MS/MS spectra may be identified in the same HPLC-MS/MS run. Therefore the identified D-methylated peptides may be paired with MS/MS spectra of their corresponding non-methylated peptides. Identification of pairs of D-methylated peptides and non-D-methylated peptides for each of the three identifications suggests that these proteins are indeed D-methylated rather than polymorphisms resulting from D-to-E mutations (Figure 6-2D and E).

DISCUSSION

Prevalence of D/E-methylation

The data obtained from our unbiased screen for methylated proteins prompted us to estimate the abundance of methylation based on the ratio of the methylated peptides and the total peptides identified. From the HeLa samples, 7052 unique peptides were positively identified, of which 3 were confirmed through manual validation to be D-methylated. Thus, the abundance of D-methylation is estimated to be around 4 in 10000 peptides (0.043% of peptides). If it is assumed that the average protein size is about 60 KD and will generate about 40 possible tryptic peptides, it is estimated that ~2% of human proteins may be subject to D-methylation. Likewise, the abundance of E-methylation among S. *cerevisiae* proteins is estimated as 0.6%.

It is likely that these estimates are lower than the true modification abundance because those methylated peptides with ambiguous modification residues were excluded from the list of identifications. It should also be noted that this estimate only reflects the abundance of the modifications among the most abundant proteins, as the samples were not processed to enrich for the modification. Nevertheless, this analysis suggests that D- and E-methylation are likely to occur at and abundance greater than tyrosine phosphorylation.

E-methylation in S. cerevisiae

TDH3, a yeast glyceraldehyde-3-phosphate dehydrogenase ortholog, was found to be E-methylated protein. The protein is a functional homotetramer

involved in glycolysis and gluconeogenesis that is found in the cell wall and cytoplasm. Our analysis identified three differentially modified forms of a peptide unambiguously derived from TDH3, including an unmodified, a singly E-methylated and a doubly E-methylated form. One of the methylated glutamate residues, E170, is well conserved among species from bacteria to humans, suggesting a possible regulatory influence for the modification. It is interesting to note that previous two-dimensional SDS-PAGE analysis of the yeast TDH isozymes resulted in resolution of two forms of TDH3 that differed only in charge (McAlister and Holland, 1985). The difference in charge was not due to phosphorylation, since the enzyme could not be labeled in extracts from yeast grown in the presence of ^{32}P (McAlister and Holland, 1985). It is possible that methylation of glutamate residues may explain the observed charge variation.

TDH3 has been identified in diverse protein complexes by tandem affinity purification coupled with mass spectrometry (Gavin et al., 2002). In addition to energy metabolism complexes, TDH3 is associated with protein complexes involved in RNA metabolism, signaling and chromatin structural maintenance. Although the functional significance of the observed E-methylation is unclear, the modification may influence the composition of the complexes in which it is found. Methylation has previously been shown to be able to modulate the assembly of complexes, since c-terminal carboxymethylation of protein

phosphatase 2A has been shown to be important for the binding of regulatory B subunits.

D-methylated Proteins in HeLa Cells

Analysis of proteomic data from HeLa cells revealed D/E-methylation among three proteins, cytokeratin 9, AHNAK, and Rab8. AHNAK was found to be methylated at a single position. This protein was originally identified in a screen for transcripts lost in neuroblastoma cells (Shtivelman et al., 1992). Its localization, and presumably function, varies depending on cell type and it has been shown to be targeted to the plasma membrane in response to its phosphorylation by either protein kinase B or protein kinase C (Hashimoto et al., 1995; Sussman et al., 2001). Although its function is currently unknown, it may serve as a platform for the integration of intracellular signals (Lee et al., 2004). Rab8, a GTPase, is known to be involved in cytoskeletal remodeling and in the regulation of intracellular trafficking (Lau and Mruk, 2003).

Mechanistic Impact of D/E-methylation

D/E methylation could be an important vehicle to modulate a protein's function, as it changes three major properties of the substrate amino acid: neutralization of negative charge, increasing size and hydrophobicity. Thus, D/E methylation is like a dephosphorylation event in terms of charge change.

The identification of diverse methylated structural proteins suggests the potential for the modification to be involved in the regulation of cytoskeletal structure. Methylation has been demonstrated to be important for motility in response to chemotactic stimuli in bacteria (Shapiro and Koshland, 1994), so it is enticing to consider the potential for methylation to modulate invasiveness and motility of mammalian cells mediated through cytoskeletal dynamics. Increased rates of protein carboxymethylation have been described in neutrophils treated with chemotactic stimuli (O'Dea et al., 1978).

CONCLUDING REMARKS

Proteomics screening identified 2 E-methylation sites from 1 substrate protein from *S. cerevisiae* as well as 3 D-methylation sites from 3 substrates from HeLa cells. The identified substrates include proteins with diverse functions, implicating the modifications as potential regulatory modifications involved in multiple pathways.

Identification of a number of D/E-methylated proteins was initially a surprise. Esterases are abundant proteins in cells. Pro-drugs take advantage of *in vivo*-labile ester bonds to improve the pharmacokinetic properties of drugs. In addition, methyl-esters are notoriously labile, particularly at higher pH. Nevertheless, several lines of evidence suggest that D/E-methylation observed

indeed occurred *in vivo* rather than *in vitro*. First, protein lysates from cells were prepared in denaturing buffer to prevent post-translational modifications from occurring *in vitro*. Second, methanol was avoided in each step of our experimental procedure. Finally, there is precedent for such modifications as E-methylation has been described in bacteria (Clarke, 2003) and methylation at C-terminal carboxylic acids is known to be present in cellular proteins such as prenylated proteins and PP2A.

D/E-methylation possesses a few unique features to serve as an ideal vehicle for regulatory functions in eukaryotic cells. First, a large group of methyltransferases, demethylases and esterases exist in cells, some of which might serve as enzymes for controlling the status of D/E methylation. Second, neutralization of D/E side chains provides a versatile approach to change a protein's structure, interactions and therefore its functions. Third, given the important roles of aspartate in enzymatic reactions, methylation of key aspartates in an active center is likely to alter the enzymatic activity of the substrate proteins. Fourth, D/E methylation is known to be reversible (Clarke, 1993). Fifth, acidic residues are present in many recognition motifs of post-translational modifications, such as the sumoylation recognition motif (hydrophobic K*X*E, where *X* is any amino acid) and phosphorylation consensus motifs (e.g., casein kinase motif). Neutralization of D/E side chains may provide an extra layer of regulation for those post-translational modifications in tandem cascades. Finally,

S-adenosyl-L-methionine is the most widely used enzyme substrate after ATP (Cantoni, 1975). It is distributed among virtually all body tissues and fluids.

Methylation of arginine and lysine residues has been extensively studied in the past decade, especially on histone proteins. Previous studies demonstrated that methylation of arginine and lysine residues is reversible and involved in the regulation of gene expression by altering chromatin structure or by creating a binding platform to recruit transcriptional regulators (Hake et al., 2004; Shi et al., 2004). In contrast to methylation of arginine and lysine, D/E-methylation leads to more significant changes in charge and hydrophobicity of the substrate residues, because methylation of arginine and lysine residues has relatively little effect on their side-chain's charge and hydrophobicity. Consequently, D/E methylation should elicit more dramatic alteration of the majority of substrate proteins than methylation of arginine and lysine.

Protein structure is influenced to a large extent by the charge state of ionizable groups on the side chains of several amino acids (Creighton, 1990). Copious literature documents how a decrease in an amino acid's net charge, induced by post-translational modifications such as acetylation and phosphorylation, alter diverse protein properties. However, similar alterations in net charge among amino acid residues with negative charge (D and E) have been neglected in the research community. Given the high prevalence and possible dynamic nature of D/E-methylation, and the resulting dramatic changes in

substrate residues' side chain properties, it is reasonable to assume that D/E-methylation is a gold mine that remains to be explored by the biomedical research community.

It is important to note that our proteomics analysis of D/E methylation did not involve prior enrichment for D/E methylated proteins. Given that protein modifications often occur sub-stoichiometrically, unmodified peptides would be expected to significantly outnumber modified peptides upon digestion, leading to suppression in the number of identified modifications. Our ability to identify E methylated peptides even without prior enrichment suggests a vast frontier the exploration of which awaits development of suitable enrichment strategies. Attempts have been made to identify D/E-methylated proteins in the past in the context of the modifications putative role as a protein repair mechanism (Clarke, 2003). For example, abundant D-methylation has been described in erythrocytes incubated with ^{3}H-methionine (O'Connor and Clarke, 1984). Validation was carried out by gel filtration of proteins and digestion of fractions with carboxypeptidase Y. HPLC analysis revealed radioactivity co-eluting with D-aspartic acid β-methyl ester standard. However, only 0.5% of initial radioactivity in the individual proteins was recovered as β-methyl ester, the rest eluting as methanol due to hydrolysis. Other work has identified tubulin, synapsin, histone H2B and αA-crystallin as substrates subject to methylation by protein-L-isoaspartyl methyltransferase 1 (PIMT1). In contrast to the preceding methods,

the study presented here represents the first step towards the generation of a catalog of D/E-methylation using a proteomics strategy. Furthermore, this is the first evidence of E-methylation occurring in eukaryotes.

Attempts were made to further validate the *in vivo* nature of the methylations using a stable isotope labeling method. For this experiment, yeast cells expressing a GST-tagged version of TDH3 were grown in the presence of methyl-[$^{13}C^{2}H_3$]-methionine. The methionine is taken up by the yeast and incorporated into S-adenosylmethionine during the normal course of metabolism. Any methylations resulting from the use of isotopically heavy methionine would be evident by a characteristic mass shift of 4 Daltons per methyl group appended. Such a method has previously been applied to the identification of arginine and lysine methylation sites in mammalian cells (Ong et al., 2004). However, we were unable to validate the methylation using this method. There are a couple of possible explanations for this phenomenon. Since yeast are capable of synthesizing all 20 amino acids (Dever and Hinnebusch, 2005), it could be that the yeast are synthesizing their own methionine in spite of the exogenous source. Also, it is possible that the turnover of the methyl esters depletes the pool of heavy methyl groups and the resulting S-adenosylhomocysteine is subsequently converted back to S-adenosylmethionine through a methyltransferase reaction involving the folate single carbon pool. Finally, reactions involving the single

carbon folate pool may be directly responsible for the modification and would not be subject to labeling via heavy-methyl methionine.

Nevertheless, the dataset of D/E-methylation demonstrates that these post-translational modifications are relatively abundant in proteins, reminding us of the potentially critical roles of the modifications in cellular regulation. The data raise many interesting questions for future consideration. What are the enzymes responsible for modulating the status of D/E-methylation? Which cellular pathways are regulated by D/E-methylation? Is this modification compartmentalized? How dynamic is the modification? Does D/E-methylation cross-talk with other post-translational modification pathways? How is the modification involved in the progression of disease? The D/E-methylation datasets generated in this study offer a steppingstone driving experimental efforts toward these directions.

METHODS

Preparation of cell lysates from HeLa cells and yeast. One dish (10 cm) of HeLa cells was grown to 80–90% confluence. The cells were washed with cold PBS buffer twice, and then 0.3 ml of cell lysis buffer was added (2% SDS, 62.5 mM Tris-Cl, pH 6.8, protease inhibitors). The cell lysate was harvested and sonicated three times for 5 s each with 20-s intervals between sonications. The

lysate was centrifuged at 4°C for 1 h at 21,000 x g. The supernatant was considered HeLa-cell whole-cell lysate.

S. cerevisiae 647 cells were grown in YPD medium at 30°C for 24 h to mid-log phase (1-5x10^7 cells/ml). Cells were harvested, washed with PBS, and resuspended in 0.2 ml of cell-breaking buffer (7 M urea, 2 M thiourea, 4% CHAPS, 50 mM Tris-Cl, pH 8.0, 10mM DTT and protease inhibitors), and lysed with acid-washed glass beads (0.45 to 0.55 mm). The lysate was centrifuged at 4°C for 1 h at 21,000 x g. The pellet was discarded, and the supernatant was considered S. *cerevisiae* whole-cell protein extract.

Forty micrograms of yeast and HeLa whole-cell lysates were resolved in a 12% SDS-PAGE gel. The gel was stained using a Colloidal Coomassie staining kit. The resolved proteins were cut out into 20 pieces of equal size for in-gel digestion.

In-gel digestion. Gel bands were sliced into small pieces (~1 mm^3) and destained with 25 mM ammonium bicarbonate solution (ethanol/water, 50:50 v/v). The destained gel pieces were washed in an acidic buffer (acetic acid/ethanol/water, 10:50:40, v/v/v) three times for one hour each time and in water two times for 20 min each time. The gel pieces were dehydrated in 100% acetonitrile and dried in a SpeedVac (ThermoFisher, Waltham, MA). One hundred fifty ng of porcine modified trypsin (Promega, Madison, WI) in 50 mM ammonium bicarbonate was added to the dried gels and incubated overnight at 37

°C. Tryptic peptides were sequentially extracted from the gel pieces with 50% acetonitrile buffer (acetonitrile/water/ trifluoroacetic acid (TFA), 50:45:5, v/v/v) and 75% acetonitrile buffer (acetonitrile/water/TFA, 75:24:1, v/v/v). The peptide extracts were pooled, dried in a SpeedVac, and desalted using a μ-C18 ZipTip (Millipore, Billerica, MA) prior to HPLC/MS/MS analysis.

HPLC-MS/MS analysis. Each sample was dissolved in 10 μl HPLC buffer A (0.1% formic acid in water (v/v)) and 2 μl were injected into Surveyor HPLC system (ThermoFinnigan, Waltham, MA) using an autosampler. Peptides were separated on a home-made capillary HPLC column (50 mm length X 75 μm ID, 4 μm particle size, 90 Å pore diameter) with Jupiter C12 resin (Phenomenex, St. Torrance, CA) and directly electrosprayed into mass spectrometer using nano-spray source. LTQ-Orbitrap was operated in the data-dependant mode acquiring fragmentation spectra of the 6 strongest ions. HPLC/LTQ mass spectrometer analysis was carried out as previously reported (Chen et al., 2005).

Protein sequence database search and manual verification. All MS/MS spectra were searched against NCBI-nr protein sequence database with the specification of D or E methylation using MASCOT database search engine. All putative D/E-methylated peptide identifications with MASCOT score > 35.0 were manually examined by rules previously described (Chen et al., 2005). All D/E

methylation sites were required to be identified by consecutive b- or y- ions so that the possibilities that methylation or +14 Da mass shift occurs on adjacent residues were eliminated.

Figure 6-1

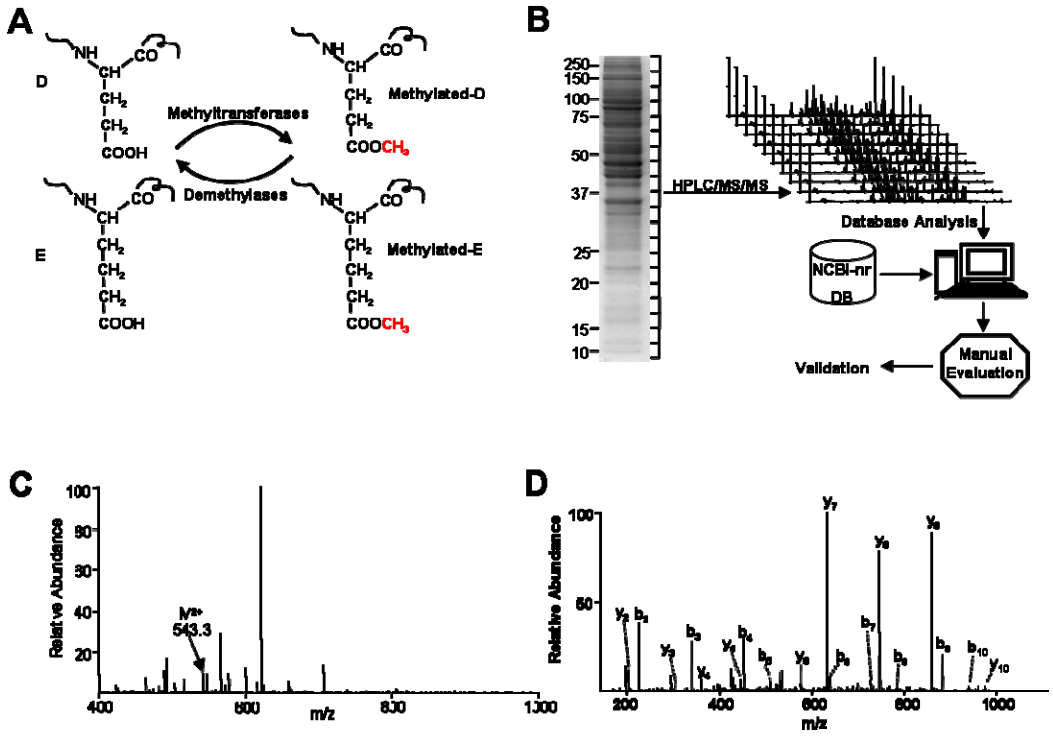

Figure 6-1. Scheme for proteomic analysis of D/E methylation. (A) Structural changes resulting from D/E-methylation; (B) The strategy for proteomic screening of D/E-methylation. The whole-cell lysate of interest was resolved by SDS-PAGE. 20 protein bands of equal size were excised and in-gel digested. The resulting peptides were analyzed by nano-HPLC/MS/MS. The mass spectrometric data was searched against the NCBI-nr database using the MASCOT search engine for peptide identification and localization of D- or E-methylation sites. All candidate modified peptides were manually verified and then further confirmed by MS/MS of synthetic peptides. (C) MS and (D) MS/MS of an ion at m/z 543.3 resulted in identification of a D-methylated peptide of sequence LLLIG**D***SGVGK. b and y ions designate N- and C-terminal fragment ions produced by collision-induced backbone fragmentation in the mass spectrometer.

Figure 6-2

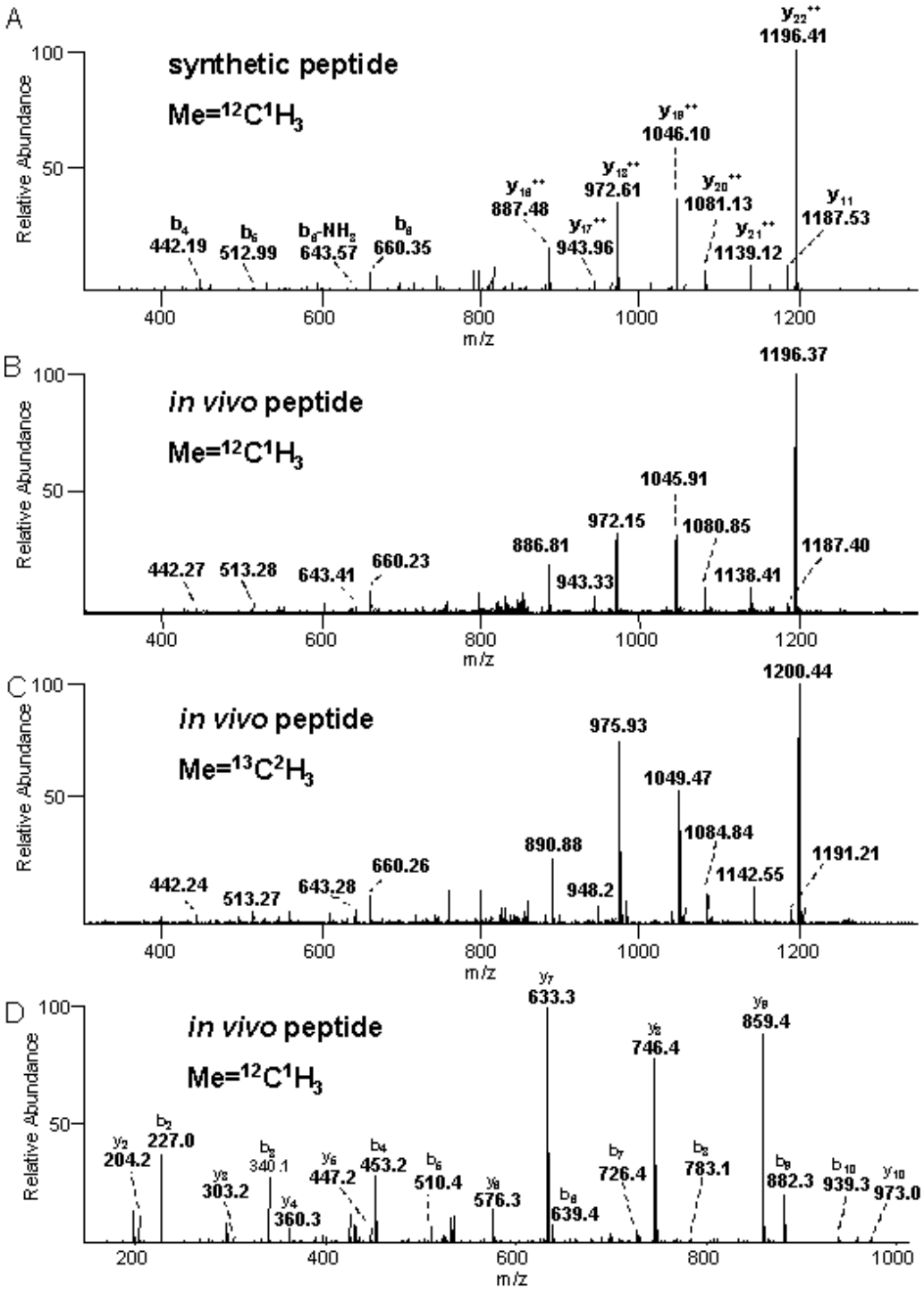
A
synthetic peptide
Me=12C1H3
Relative Abundance
m/z
y22++ 1196.41
y18++ 1046.10
y16++ 887.48
y12++ 972.61
y20++ 1081.13
y17++ 943.96
y21++ 1139.12
y11 1187.53
b4 442.19
b5 512.99
b6-NH3 643.57
b8 660.35
B
in vivo peptide
Me=12C1H3
1196.37
1045.91
972.15
1080.85
886.81
1187.40
442.27
513.28
643.41
660.23
943.33
1138.41
C
in vivo peptide
Me=13C2H3
1200.44
975.93
1049.47
1084.84
890.88
1191.21
442.24
513.27
643.28
660.26
948.2
1142.55
D
in vivo peptide
Me=12C1H3
y7 633.3
y8 859.4
y6 746.4
b2 227.0
y2 204.2
b3 340.1
y3 303.2
y6 447.2
b4 453.2
b5 510.4
y4 360.3
y8 576.3
b8 639.4
b7 726.4
b8 783.1
b9 882.3
b10 939.3
y10 973.0

Figure 6-2. Verification of methylated peptides. (A) MS/MS of an E-methylated synthetic peptide VINDAFGI**E*E***GLMTTVHSLTATQK (**E*** indicates methylated glutamic acid) and (B) peptide identified during proteomic screening of yeast; (C) MS/MS of a D-methylated peptide LLLIG**D***SGVGK and (D) its unmodified counterpart from a proteomic screening of HeLa lysate.

Table 6-1

Protein Name	gi#	Peptide Sequence	Functional Category
D-methylated proteins from HeLa cells			
cytokeratin 9	gi|435476	VQALEEANND*LENK	Structural
AHNAK nucleoprotein isoform 1	gi|61743954	ISMPDVD*LHLK	Structural
rab8 small GTP binding protein	gi|452318	LLLIGD*SGVGK	Signal Transduction
E-methylated proteins from *S. cerevisiae* cells			
TDH3	gi|1323341	VINDAFGIE*E*GLMTTVHSLTATQK	Glycolysis

Table 6-1. Proteins identified during proteomics screen for D- and E-methylation. Sites of methylation are indicated in boldface followed by an asterisk.

CHAPTER SEVEN

CONCLUSIONS AND RECOMMENDATIONS

It is clear from the preceding dissertation that the identification and characterization of proteins bearing post-translational modifications remains a challenging area of research. The vast chemical diversity afforded by the modifications allows for a dynamic system capable of changing in ways unpredictable using genomic methods alone and embodies a complex and intricate means of communication among the interconnected biochemical networks. Successful translation and understanding of this language will allow a more thorough understanding of disease processes and regulatory pathways. However, such understanding cannot arise out of a purely reductionist approach. A true systems-wide understanding of the interconnectedness of cellular processes will only come from an approach capable of integrating data on the functional state of the entire cellular content at once. As post-translational modifications are the most dynamic and adaptable means of modulating protein activity, the development of efficient enrichment strategies targeting specific modifications provides a window into cell-wide protein activity. In addition, such techniques facilitate a deeper analysis of post-translationally modified proteins, permitting detection of low-to-

medium abundance. These strategies bring the field of proteomics a step closer to its goal of comprehensively characterizing cellular networks in a single analysis.

Each post-translational modification poses unique challenges with respect to its isolation and functional characterization. The labeling and enrichment strategies described in this dissertation have made possible the proteomic dissection of the cellular targets of the O-GlcNAc modification and lysine acetylation. The datasets generated from these screens have provided new insight into the potential functional implications of the modifications by expanding our knowledge of their breadth of targets. In addition, the unambiguous identification of modification sites among proteins allows for the direct assessment of the functional consequences of its attachment. This feature of proteomic screens is illustrated in Chapter 3 where definitive localization of acetylation sites permitted mutational analysis for elucidation of the functional consequences of the modification. Moreover, the comprehensive analysis of peptides eluting during an LC separation allows for unbiased screening of the data in the search for novel protein modifications. This unbiased approach led to the identification of propionylation and butyrylation among histones and suggests potentially abundant methyl esterification of aspartic and glutamic acids.

While the field of proteomics possesses powerful tools uniquely suited to the system-wide characterization of proteins and their modifications, it remains fundamentally limited by the complexity of biological samples and the incomplete

nature of the databases on which the protein identifications rely. Mass spectrometers can only tell us the mass-to-charge ratio of a compound. It remains up to the investigator to interpret the results and determine the compound's identity. Although a typical LC separation and mass spectrometric analysis may last 2 hours and generate thousands of unique spectra, unambiguous protein identifications are possible for only about 5% of spectra. Some of this discrepancy can be attributed to co-elution of peaks from the chromatography column, yielding complicated spectra derived from multiple peptide species simultaneously. However, a large portion of the unassignable spectra is likely to arise from nucleotide polymorphisms at the DNA level, RNA splicing variability that remains to be annotated, protein processing events resulting in amino acid sequences that are not present in genome databases, and unknown post-translational modifications. Tools are currently available for the *de novo* sequencing of peptides, but will require additional refinement in order to tackle the complexity of even a modest microbial proteome. Such advanced search algorithms, when developed, will undoubtedly provide an unprecedented depth of view into the proteome and will likely challenge the current conception of regulatory complexity.

In conclusion, the work presented in this dissertation demonstrates the power of proteomic methods in the identification and characterization of post-translational modifications. Mass spectrometry alone possesses the sensitivity and

speed necessary to provide accurate identification of modified species and simultaneous pinpointing of the modification site, while the rapid analytical screening strategies afforded by the proteomics workflow permit the high-throughput, global characterization of post-translationally modified proteins. Thus, proteomics tools represent ideal methods for assessing the abundance and variety of cellular targets for specific modifications. True comprehensive characterization of a cellular proteome in a single analysis remains a daunting challenge, requiring significant advances in technology and methodology. In the meantime, the greatest impact of proteomics will stem from its strongest attributes: the potential for unbiased and system-wide screening approaches, unmatched sensitivity and the opportunity for relative and absolute quantitation of biomolecules in a high-throughput manner.

BIBLIOGRAPHY

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198-207.

Agalioti, T., Chen, G., and Thanos, D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. Cell *111*, 381-392.

Anastasiadis, P. Z., Moon, S. Y., Thoreson, M. A., Mariner, D. J., Crawford, H. C., Zheng, Y., and Reynolds, A. B. (2000). Inhibition of RhoA by p120 catenin. Nat Cell Biol *2*, 637-644.

Avantaggiati, M. L., Ogryzko, V., Gardner, K., Giordano, A., Levine, A. S., and Kelly, K. (1997). Recruitment of p300/CBP in p53-dependent signal pathways. Cell *89*, 1175-1184.

Barak, R., and Eisenbach, M. (2001). Acetylation of the response regulator, CheY, is involved in bacterial chemotaxis. Mol Microbiol *40*, 731-743.

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

Blander, G., and Guarente, L. (2004). The Sir2 family of protein deacetylases. Annu Rev Biochem *73*, 417-435.

Bordone, L., and Guarente, L. (2005). Calorie restriction, SIRT1 and metabolism: understanding longevity. Nat Rev Mol Cell Biol *6*, 298-305.

Borthwick, A. C., Holms, W. H., and Nimmo, H. G. (1984). Amino acid sequence round the site of phosphorylation in isocitrate dehydrogenase from Escherichia coli ML308. FEBS Lett *174*, 112-115.

Bouras, T., Fu, M., Sauve, A. A., Wang, F., Quong, A. A., Perkins, N. D., Hay, R. T., Gu, W., and Pestell, R. G. (2005). SIRT1 deacetylation and repression of p300 involves lysine residues 1020/1024 within the cell cycle regulatory domain 1. J Biol Chem *280*, 10264-10276.

Bradshaw, R. A., and Burlingame, A. L. (2005). From proteins to proteomics. IUBMB Life *57*, 267-272.

Browner, W. S., Kahn, A. J., Ziv, E., Reiner, A. P., Oshima, J., Cawthon, R. M., Hsueh, W. C., and Cummings, S. R. (2004). The genetics of human longevity. Am J Med *117*, 851-860.

Buhl, W. J. (1992). Annexins and phospholipase A2 inhibition. Eicosanoids *5 Suppl*, S26-28.

Cantoni, G. L. (1975). Biological methylation: selected aspects. Annu Rev Biochem *44*, 435-451.

Carrozza, M. J., Utley, R. T., Workman, J. L., and Cote, J. (2003). The diverse functions of histone acetyltransferase complexes. Trends Genet *19*, 321-329.

Chen, Y., Kwon, S. W., Kim, S. C., and Zhao, Y. (2005). Integrated approach for manual evaluation of peptides identified by searching protein sequence databases with tandem mass spectra. J Proteome Res *4*, 998-1005.

Chung, C. H., Levy, S., Chaurand, P., and Carbone, D. P. (2007). Genomics and proteomics: emerging technologies in clinical cancer research. Crit Rev Oncol Hematol *61*, 1-25.

Clarke, S. (1993). Protein methylation. Curr Opin Cell Biol *5*, 977-983.

Clarke, S. (2003). Aging as war between chemical and biochemical processes: protein methylation and the recognition of age-damaged proteins for repair. Ageing Res Rev *2*, 263-285.

Cohen, T., and Yao, T. P. (2004). AcK-knowledge reversible acetylation. Sci STKE *2004*, pe42.

Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A., and Hart, G. W. (2001). Characterization of a mouse monoclonal antibody specific for O-linked N-acetylglucosamine. Anal Biochem *293*, 169-177.

Creighton, T. E. (1990). Protein folding. Biochem J *270*, 1-16.

Dever, T. E., and Hinnebusch, A. G. (2005). GCN2 whets the appetite for amino acids. Mol Cell *18*, 141-142.

Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

Dinman, J. D. (2006). Programmed Ribosomal Frameshifting Goes Beyond Viruses: Organisms from all three kingdoms use frameshifting to regulate gene expression, perhaps signaling a paradigm shift. Microbe Wash DC *1*, 521-527.

Doyon, Y., Cayrou, C., Ullah, M., Landry, A. J., Cote, V., Selleck, W., Lane, W. S., Tan, S., Yang, X. J., and Cote, J. (2006). ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol Cell *21*, 51-64.

Driscoll, R., Hudson, A., and Jackson, S. P. (2007). Yeast Rtt109 promotes genome stability by acetylating histone H3 on lysine 56. Science *315*, 649-652.

Dutnall, R. N., Tafrov, S. T., Sternglanz, R., and Ramakrishnan, V. (1998). Structure of the yeast histone acetyltransferase Hat1: insights into substrate specificity and implications for the Gcn5-related N-acetyltransferase superfamily. Cold Spring Harb Symp Quant Biol *63*, 501-507.

Egger, G., Liang, G., Aparicio, A., and Jones, P. A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. Nature *429*, 457-463.

Etzkorn, F. A., Chang, Z. Y., Stolz, L. A., and Walsh, C. T. (1994). Cyclophilin residues that affect noncompetitive inhibition of the protein serine phosphatase activity of calcineurin by the cyclophilin.cyclosporin A complex. Biochemistry *33*, 2380-2388.

Farriol-Mathis, N., Garavelli, J. S., Boeckmann, B., Duvaud, S., Gasteiger, E., Gateau, A., Veuthey, A. L., and Bairoch, A. (2004). Annotation of post-translational modifications in the Swiss-Prot knowledge base. Proteomics *4*, 1537-1550.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science *246*, 64-71.

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

Fischle, W., Wang, Y., and Allis, C. D. (2003). Histone and chromatin cross-talk. Curr Opin Cell Biol *15*, 172-183.

Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M.*, et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Gray, M. W., Burger, G., and Lang, B. F. (1999). Mitochondrial evolution. Science *283*, 1476-1481.

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

Gu, W., and Roeder, R. G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell *90*, 595-606.

Gu, W., Shi, X. L., and Roeder, R. G. (1997). Synergistic activation of transcription by CBP and p53. Nature *387*, 819-823.

Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. Mol Cell Biol *19*, 1720-1730.

Hake, S. B., Xiao, A., and Allis, C. D. (2004). Linking the epigenetic 'language' of covalent histone modifications to cancer. Br J Cancer *90*, 761-769.

Hamilton, B., Dong, Y., Shindo, M., Liu, W., Odell, I., Ruvkun, G., and Lee, S. S. (2005). A systematic RNAi screen for longevity genes in C. elegans. Genes Dev *19*, 1544-1555.

Hashimoto, T., Gamou, S., Shimizu, N., Kitajima, Y., and Nishikawa, T. (1995). Regulation of translocation of the desmoyokin/AHNAK protein to the plasma membrane in keratinocytes by protein kinase C. Exp Cell Res *217*, 258-266.

Heine, A., Luz, J. G., Wong, C. H., and Wilson, I. A. (2004). Analysis of the class I aldolase binding site architecture based on the crystal structure of 2-deoxyribose-5-phosphate aldolase at 0.99A resolution. J Mol Biol *343*, 1019-1034.

Howitz, K. T., Bitterman, K. J., Cohen, H. Y., Lamming, D. W., Lavu, S., Wood, J. G., Zipkin, R. E., Chung, P., Kisielewski, A., Zhang, L. L.*, et al.* (2003). Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan. Nature *425*, 191-196.

Hubbard, S. J., Campbell, S. F., and Thornton, J. M. (1991). Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. J Mol Biol *220*, 507-530.

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., and Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res *32*, 1037-1049.

Inoki, K., Zhu, T., and Guan, K. L. (2003). TSC2 mediates cellular energy response to control cell growth and survival. Cell *115*, 577-590.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jones, D. T., and Ward, J. J. (2003). Prediction of disordered regions in proteins from position specific score matrices. Proteins *53 Suppl 6*, 573-578.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577-2637.

Kaiser, C., and James, S. R. (2004). Acetylation of insulin receptor substrate-1 is permissive for tyrosine phosphorylation. BMC Biol *2*, 23.

Karlberg, O., Canback, B., Kurland, C. G., and Andersson, S. G. (2000). The dual origin of the yeast mitochondrial proteome. Yeast *17*, 170-187.

Kelly, W. K., Richon, V. M., O'Connor, O., Curley, T., MacGregor-Curtelli, B., Tong, W., Klang, M., Schwartz, L., Richardson, S., Rosa, E.*, et al.* (2003). Phase I clinical trial of histone deacetylase inhibitor: suberoylanilide hydroxamic acid administered intravenously. Clin Cancer Res *9*, 3578-3588.

Khidekel, N., Ficarro, S. B., Peters, E. C., and Hsieh-Wilson, L. C. (2004). Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. Proc Natl Acad Sci U S A *101*, 13132-13137.

Kho, Y., Kim, S. C., Jiang, C., Barma, D., Kwon, S. W., Cheng, J., Jaunbergs, J., Weinbaum, C., Tamanoi, F., Falck, J., and Zhao, Y. (2004). A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. Proc Natl Acad Sci U S A *101*, 12479-12484.

Kim, S., Chin, K., Gray, J. W., and Bishop, J. M. (2004). A screen for genes that suppress loss of contact inhibition: identification of ING4 as a candidate tumor suppressor gene in human cancer. Proc Natl Acad Sci U S A *101*, 16251-16256.

Kim, S. C., Chen, Y., Mirza, S., Xu, Y., Lee, J., Liu, P., and Zhao, Y. (2006a). A clean, more efficient method for in-solution digestion of protein mixtures without detergent or urea. J Proteome Res *5*, 3446-3452.

Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L.*, et al.* (2006b). Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. Mol Cell *23*, 607-618.

King, M. T., and Reiss, P. D. (1985). Separation and measurement of short-chain coenzyme-A compounds in rat liver by reversed-phase high-performance liquid chromatography. Anal Biochem *146*, 173-179.

Kouzarides, T. (2000). Acetylation: a regulatory modification to rival phosphorylation? Embo J *19*, 1176-1179.

Kovacs, J. J., Murphy, P. J., Gaillard, S., Zhao, X., Wu, J. T., Nicchitta, C. V., Yoshida, M., Toft, D. O., Pratt, W. B., and Yao, T. P. (2005). HDAC6 regulates Hsp90 acetylation and chaperone-dependent activation of glucocorticoid receptor. Mol Cell *18*, 601-607.

Kwon, S. W., Kim, S. C., Jaunbergs, J., Falck, J. R., and Zhao, Y. (2003). Selective enrichment of thiophosphorylated polypeptides as a tool for the analysis of protein phosphorylation. Mol Cell Proteomics *2*, 242-247.

Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., and Jorgensen, T. J. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics *4*, 873-886.

Lau, A. S., and Mruk, D. D. (2003). Rab8B GTPase and junction dynamics in the testis. Endocrinology *144*, 1549-1563.

Lee, I. H., You, J. O., Ha, K. S., Bae, D. S., Suh, P. G., Rhee, S. G., and Bae, Y. S. (2004). AHNAK-mediated activation of phospholipase C-gamma1 through protein kinase C. J Biol Chem *279*, 26645-26653.

Lee, J., Xu, Y., Chen, Y., Sprung, R., Kim, S. C., Xie, S., and Zhao, Y. (2007). Mitochondrial phosphoproteome revealed by an improved IMAC method and MS/MS/MS. Mol Cell Proteomics *6*, 669-676.

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., 3rd (1999). Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol *17*, 676-682.

Luo, X., Budihardjo, I., Zou, H., Slaughter, C., and Wang, X. (1998). Bid, a Bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors. Cell *94*, 481-490.

Mann, M. (1990). Electrospray: Its Potential and Limitations as an Ionization Method for Biomolecules. Organic Mass Spectrometry *25*, 575-587.

Mann, M., and Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. Nat Biotechnol *21*, 255-261.

Margueron, R., Trojer, P., and Reinberg, D. (2005). The key to development: interpreting the histone code? Curr Opin Genet Dev *15*, 163-176.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838-849.

McAlister, L., and Holland, M. J. (1985). Differential expression of the three yeast glyceraldehyde-3-phosphate dehydrogenase genes. J Biol Chem *260*, 15019-15027.

McKinsey, T. A., and Olson, E. N. (2004). Cardiac histone acetylation--therapeutic opportunities abound. Trends Genet *20*, 206-213.

McLachlin, D. T., and Chait, B. T. (2003). Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. Anal Chem *75*, 6826-6836.

Michishita, E., Park, J. Y., Burneskis, J. M., Barrett, J. C., and Horikawa, I. (2005). Evolutionarily Conserved and Nonconserved Cellular Localizations and Functions of Human SIRT Proteins. Mol Biol Cell *16*, 4623-4635.

Minucci, S., and Pelicci, P. G. (2006). Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. Nat Rev Cancer *6*, 38-51.

Mira, J. P., Dubois, T., Oudinet, J. P., Lukowski, S., Russo-Marie, F., and Geny, B. (1997). Inhibition of cytosolic phospholipase A2 by annexin V in differentiated permeabilized HL-60 cells. Evidence of crucial importance of domain I type II Ca2+-binding site in the mechanism of inhibition. J Biol Chem *272*, 10474-10482.

Montreuil, J., and al., e. (1994). Glycoproteins, In Carbohydrate Analysis: A Practical Approach, M. F. Chaplin, and J. F. Kennedy, eds. (New York: IRL Press), pp. 181-293.

Mootha, V. K., Bunkenborg, J., Olsen, J. V., Hjerrild, M., Wisniewski, J. R., Stahl, E., Bolouri, M. S., Ray, H. N., Sihag, S., Kamal, M.*, et al.* (2003). Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. Cell *115*, 629-640.

Morris, A. J., and Tolan, D. R. (1994). Lysine-146 of rabbit muscle aldolase is essential for cleavage and condensation of the C3-C4 bond of fructose 1,6-bis(phosphate). Biochemistry *33*, 12291-12297.

Nandi, A., Sprung, R., Barma, D. K., Zhao, Y., Kim, S. C., Falck, J. R., and Zhao, Y. (2006). Global identification of O-GlcNAc-modified proteins. Anal Chem *78*, 452-458.

Neidhardt, F., ed. (1996). Escherichia coli and Salmonella Cellular and Molecular Biology (Washington D.C.).

Nelson, D. L., and Cox, M. M. (2005). Chapter 17 in Lehninger Principles of Biochemistry. (New York, NY: W.H. Freemand and Company).

O'Connor, C. M., and Clarke, S. (1984). Carboxyl methylation of cytosolic proteins in intact human erythrocytes. Identification of numerous methyl-accepting proteins including hemoglobin and carbonic anhydrase. J Biol Chem *259*, 2570-2578.

O'Dea, R. F., Viveros, O. H., Axelrod, J., Aswanikaumar, S., Schiffmann, E., and Corcoran, B. A. (1978). Raipid stimulation of protein carboxymethylation in leukocytes by a chemotatic peptide. Nature *272*, 462-464.

O'Donnell, N., Zachara, N. E., Hart, G. W., and Marth, J. D. (2004). Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. Mol Cell Biol *24*, 1680-1690.

Ong, S. E., Mittler, G., and Mann, M. (2004). Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. Nat Methods *1*, 119-126.

Pascal, M. C., Chippaux, M., Abou-Jaoude, A., Blaschkowski, H. P., and Knappe, J. (1981). Mutants of Escherichia coli K12 with defects in anaerobic pyruvate metabolism. J Gen Microbiol *124*, 35-42.

Patterson, S. D. (2000). Proteomics: the industrialization of protein chemistry. Curr Opin Biotechnol *11*, 413-418.

Payne, A. H., and Glish, G. L. (2005). Tandem mass spectrometry in quadrupole ion trap and ion cyclotron resonance mass spectrometers. Methods Enzymol *402*, 109-148.

Peterson, C. L., and Laniel, M. A. (2004). Histones and histone modifications. Curr Biol *14*, R546-551.

Polevoda, B., and Sherman, F. (2002). The diversity of acetylated proteins. Genome Biol *3*, reviews0006.

Posern, G., Miralles, F., Guettler, S., and Treisman, R. (2004). Mutant actins that stabilise F-actin use distinct mechanisms to activate the SRF coactivator MAL. Embo J *23*, 3973-3983.

Prescher, J. A., Dube, D. H., and Bertozzi, C. R. (2004). Chemical remodelling of cell surfaces in living animals. Nature *430*, 873-877.

Qian, W. J., Jacobs, J. M., Liu, T., Camp, D. G., 2nd, and Smith, R. D. (2006). Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. Mol Cell Proteomics *5*, 1727-1744.

Rojas, J. R., Trievel, R. C., Zhou, J., Mo, Y., Li, X., Berger, S. L., Allis, C. D., and Marmorstein, R. (1999). Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide. Nature *401*, 93-98.

Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X. M., Polakiewicz, R. D., and Comb, M. J. (2005). Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat Biotechnol *23*, 94-101.

Sakaguchi, K., Herrera, J. E., Saito, S., Miki, T., Bustin, M., Vassilev, A., Anderson, C. W., and Appella, E. (1998). DNA damage activates p53 through a phosphorylation-acetylation cascade. Genes Dev *12*, 2831-2841.

Saxon, E., and Bertozzi, C. R. (2000). Cell surface engineering by a modified Staudinger reaction. Science *287*, 2007-2010.

Schlaepfer, D. D., Jones, J., and Haigler, H. T. (1992). Inhibition of protein kinase C by annexin V. Biochemistry *31*, 1886-1891.

Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003). Many paths to methyltransfer: a chronicle of convergence. Trends Biochem Sci *28*, 329-335.

Shapiro, M. J., and Koshland, D. E., Jr. (1994). Mutagenic studies of the interaction between the aspartate receptor and methyltransferase from Escherichia coli. J Biol Chem *269*, 11054-11059.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shia, W. J., Pattenden, S. G., and Workman, J. L. (2006). Histone H4 lysine 16 acetylation breaks the genome's silence. Genome Biol *7*, 217.

Shtivelman, E., Cohen, F. E., and Bishop, J. M. (1992). A human gene (AHNAK) encoding an unusually large protein with a 1.2-microns polyionic rod structure. Proc Natl Acad Sci U S A *89*, 5472-5476.

Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J., and Udseth, H. R. (1990). New developments in biochemical mass spectrometry: electrospray ionization. Anal Chem *62*, 882-899.

Sprung, R., Nandi, A., Chen, Y., Kim, S. C., Barma, D., Falck, J. R., and Zhao, Y. (2005). Tagging-via-substrate strategy for probing O-GlcNAc modified proteins. J Proteome Res *4*, 950-957.

Starai, V. J., Celic, I., Cole, R. N., Boeke, J. D., and Escalante-Semerena, J. C. (2002). Sir2-dependent activation of acetyl-CoA synthetase by deacetylation of active lysine. Science *298*, 2390-2392.

Starai, V. J., and Escalante-Semerena, J. C. (2004). Identification of the protein acetyltransferase (Pat) enzyme that acetylates acetyl-CoA synthetase in Salmonella enterica. J Mol Biol *340*, 1005-1012.

Starai, V. J., Takahashi, H., Boeke, J. D., and Escalante-Semerena, J. C. (2003). Short-chain fatty acid activation by acyl-coenzyme A synthetases requires SIR2 protein function in Salmonella enterica and Saccharomyces cerevisiae. Genetics *163*, 545-555.

Stein, L. D. (2004). Human genome: end of the beginning. Nature *431*, 915-916.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Sussman, J., Stokoe, D., Ossina, N., and Shtivelman, E. (2001). Protein kinase B phosphorylates AHNAK and regulates its subcellular localization. J Cell Biol *154*, 1019-1030.

Tai, H. C., Khidekel, N., Ficarro, S. B., Peters, E. C., and Hsieh-Wilson, L. C. (2004). Parallel identification of O-GlcNAc-modified proteins from cell lysates. J Am Chem Soc *126*, 10500-10501.

Tang, Y., Luo, J., Zhang, W., and Gu, W. (2006). Tip60-Dependent Acetylation of p53 Modulates the Decision between Cell-Cycle Arrest and Apoptosis. Mol Cell *24*, 827-839.

Taylor, S. W., Fahy, E., Zhang, B., Glenn, G. M., Warnock, D. E., Wiley, S., Murphy, A. N., Gaucher, S. P., Capaldi, R. A., Gibson, B. W., and Ghosh, S. S. (2003). Characterization of the human heart mitochondrial proteome. Nat Biotechnol *21*, 281-286.

Thompson, P. R., Wang, D., Wang, L., Fulco, M., Pediconi, N., Zhang, D., An, W., Ge, Q., Roeder, R. G., Wong, J.*, et al.* (2004). Regulation of the p300 HAT domain via a novel activation loop. Nat Struct Mol Biol *11*, 308-315.

Thomson, J. J. (1897). Cathode Rays. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science *5th Series*, 293-316.

Thorsness, P. E., and Koshland, D. E., Jr. (1987). Inactivation of isocitrate dehydrogenase by phosphorylation is mediated by the negative charge of the phosphate. J Biol Chem *262*, 10422-10425.

Torres, C. R., and Hart, G. W. (1984). Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. J Biol Chem *259*, 3308-3317.

Vidali, G., Gershey, E. L., and Allfrey, V. G. (1968). Chemical studies of histone acetylation. The distribution of epsilon-N-acetyllysine in calf thymus histones. J Biol Chem *243*, 6361-6366.

Vosseller, K., Hansen, K. C., Chalkley, R. J., Trinidad, J. C., Wells, L., Hart, G. W., and Burlingame, A. L. (2005). Quantitative analysis of both protein expression and serine / threonine post-translational modifications through stable isotope labeling with dithiothreitol. Proteomics *5*, 388-398.

Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J., Jr. (2005). Protein posttranslational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed Engl *44*, 7342-7372.

Wang, J., Morris, A. J., Tolan, D. R., and Pagliaro, L. (1996). The molecular nature of the F-actin binding activity of aldolase revealed with site-directed mutants. J Biol Chem *271*, 6861-6865.

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

Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J., and Hart, G. W. (2002). Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. Mol Cell Proteomics *1*, 791-804.

Wells, L., Vosseller, K., and Hart, G. W. (2003). A role for N-acetylglucosamine as a nutrient sensor and mediator of insulin resistance. Cell Mol Life Sci *60*, 222-228.

Yang, X. J. (2004a). The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. Nucleic Acids Res *32*, 959-976.

Yang, X. J. (2004b). Lysine acetylation and the bromodomain: a new partnership for signaling. Bioessays *26*, 1076-1087.

Zachara, N. E., and Hart, G. W. (2002). The emerging significance of O-GlcNAc in cellular regulation. Chem Rev *102*, 431-438.

Zachara, N. E., and Hart, G. W. (2004a). O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. Biochim Biophys Acta *1673*, 13-28.

Zachara, N. E., and Hart, G. W. (2004b). O-GlcNAc modification: a nutritional sensor that modulates proteasome function. Trends Cell Biol *14*, 218-221.

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

Zhao, Y., Zhang, W., Kho, Y., and Zhao, Y. (2004). Proteomic analysis of integral plasma membrane proteins. Anal Chem *76*, 1817-1823.

Zhu, P., Huber, E., Kiefer, F., and Gottlicher, M. (2004). Specific and redundant functions of histone deacetylases in regulation of cell cycle and apoptosis. Cell Cycle *3*, 1240-1242.

VITAE

Robert William Sprung, Jr. was born to Robert, Sr. and Doris Jean Sprung on April 21, 1977 in St. Louis, Missouri. After graduation from St. Louis University High School, he began studying at Meramec Community College in St. Louis, Missouri in 1995. Transferring to St. Louis University in the fall of 1996, he received a Bachelor of Arts degree with a major in Chemistry in May, 2000. Undergraduate research in the laboratory of Dr. Dana M. Spence inspired Robert to pursue a Masters of Science (Research) degree in Analytical Chemistry, which he earned from St. Louis University in May, 2002. His persisting interest in biomedical research led him to continue graduate study at the University of Texas Southwestern Medical Center at Dallas. With an interest in mass spectrometry and the application of proteomic analysis techniques, Robert joined the laboratory of Dr. Yingming Zhao in the department of Biochemistry in the summer of 2003.

Permanent Address: 1432 McCausland Ave.
St. Louis, MO 63117