UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mass Spectrometry-Based Chemical and Quantitative Proteomics

A Dissertation submitted in partial satisfaction of
the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Haibo Qiu

March 2009

Dissertation Committee:

Dr. Yinsheng Wang, Chairperson
Dr. Quan Jason Cheng
Dr. Ryan Julian

The Dissertation of Haibo Qiu is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

There are a number of individuals to whom I would like to express my sincere gratitude. Completing this Ph.D. would have been impossible without the assistance and support I received throughout this dissertation work.

First and foremost, I wish to thank my academic advisor, Dr. Yinsheng Wang, for the opportunity to work in his laboratory, and for his great guidance and support. His guidance and support gave me the passion and freedom in completing this work. It has always been and will always be my great honor to have him as my mentor.

I wish to express thanks to other members of my thesis committee, Dr. Quan Jason Cheng and Dr. Ryan Julian for their help, advice and evaluation of this work. Their critical assessment of the content of my dissertation is truly appreciated. I would also like to thank Dr. Cynthia Larive and Dr. Dallas Rabenstein for their help during the past few years.

I want to thank the staff members in our department, Center of Plant Cell Biology and Core Instrumentation Facility for their assistance in my work over the past years. Special thanks go to Dr. Songqin Pan for his help with QSTAR MALDI mass spectrometer. Thanks go to Dr. David Carter for his training and assistance in confocal microscopy, and to Ms. Barbara Walter for her help with flow cytometry. I also want to thank Dr. Wei Sun from the Division of Biomedical Sciences for his suggestion and discussion in immunocytochemistry and Western blot.

In the laboratory, I express my sincere thanks to Drs. Huachuan Cao, Qibin Zhang and Yuesong Wang for their help with organic synthesis. I would also like to give thanks to Drs.Yan Zou, Haizheng Hong, Yuan Gao and Qingchun Zhang for their help with cell culture and LC-MS/MS when I started the work. Special thanks go to Hongxia Wang for her valuable discussion and help in my research. Thanks to Yong, Lei, Jianshuang, Mario, Renee and Nisana for their assistance in the laboratory; thanks to Drs. Liyan Ping and Bifeng Yuan for their help and discussion in molecular biology. I am very happy to be associated with such a nice group and good friends.

Finally, I am grateful to my parents and brother for their love and continuous support that I will never be able to pay back. I am deeply indebted to my wife, who has been through all the difficulties and joys with me for the past seven years. I love you and our baby more than words can say.

COPYRIGHT ACKNOWLEDGEMENTS

The text and figures in Chapter two, in part or full, are a reprint of the material as it appears in *Anal. Chem.*, **2007**, 79, 5547-5556. The coauthor (Dr. Yinsheng Wang) listed in that publication directed and supervised the research that forms the basis of this chapter.

The text and figures in Chapter three, in part or full, are a reprint of the material as it appears in *J. Proteome Res.* **2009,** 8, in press. The coauthor (Dr. Yinsheng Wang) listed in that publication directed and supervised the research that forms the basis of this chapter.

The text and figures in Chapter four, in part or full, are a reprint of the material as it appears in *J. Proteome Res.* **2008,** 7, 1904-1915. The coauthor (Dr. Yinsheng Wang) listed in that publication directed and supervised the research that forms the basis of this chapter.

ABSTRACT OF THE DISSERTATION

Mass Spectrometry-Based Chemical and Quantitative Proteomics

by

Haibo Qiu

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, March 2009
Dr. Yinsheng Wang, Chairperson

As one of the most powerful techniques in analytical sciences, mass spectrometry (MS), coupled with the techniques for sample preparation and separation, plays an essential role on the rise of proteomics. In this dissertation, we focused on the development and application of MS and chemical tools for the identification and quantification of several important classes of proteins, including nucleotide-binding proteins, nucleic acid-binding proteins and plasma membrane proteins.

In Chapter two, we described the synthesis and application of a biotin-conjugated chemical probe in probing nucleotide-binding proteins. This activity-based chemical probe can react specifically with the lysine residue at the nucleotide-binding site. The strategy developed for proteomic application is relatively simple, fast, and straightforward. This method is useful in the activity-based functional study and in profiling nucleotide-binding proteins, e.g., kinases, in complex biological samples.

In Chapter three, a strategy for the comprehensive investigation of DNA-binding proteins with *in vivo* chemical cross-linking and LC-MS/MS was developed. DNA-

binding proteins were isolated via the isolation of DNA-protein complexes and released from the complexes by reversing the cross-linking. By using this method, we were able to identify more than one hundred DNA-binding proteins, including proteins involved in transcription, gene regulation, DNA replication and repair.

In Chapter four, we described a strategy, including SILAC, cell-surface biotinylation, affinity purification and LC-MS/MS, for the identification and quantification of cell-surface membrane proteins. Integrins, cell adhesive molecules, CD antigens and receptors, which are essential in tumor development, were quantified in this work. The identification of aberrantly expressed membrane proteins of melanoma cells sets the stage for the future investigation of these proteins in the progression of human melanoma. The method held a great potential in the comprehensive identification of tumor progression makers, as well as in the discovery of protein-based new therapeutic targets.

In Chapter five, a comparative study of protein expression in human leukemia cells upon 6-thioguanine treatment was performed by using a MS-based proteomic method together with SILAC. The biological implications of the changed expression of some proteins were discussed. This study may offer new insights into the molecular mechanisms of action of this drug.

TABLE OF CONTENTS

Acknowledgements……iv

Copyright Acknowledgements……vi

Abstract of the Dissertation……vii

Table of Contents……ix

List of Figures……...xvi

List of Schemes……..xxiii

List of Tables……..xxiv

List of Supporting Information……xxv

Chapter 1. General Overview

1.1 Introduction ……...1

1.2 MS-based chemical proteomics……2

1.2.1 The general structure of a chemical probe……..3

1.2.2 Application of chemical probes in MS-based proteomics……..5

1.2.2.1 Activity-based protein profiling ……..6

1.2.2.2 Membrane protein biotinylation……..8

1.2.2.3 Chemical cross-linking……..11

1.3 MS-based quantitative proteomics……...13

1.3.1 Chemical tagging strategies for quantitative proteomics……..14

1.3.1.1 Isotope-coded affinity tag (ICAT)…….14

1.3.1.2 Isobaric tag for relative and absolute quantitation (iTRAQ)…….17

1.3.1.3 Other chemical labeling strategies…….18

1.3.2 Metabolic labeling strategies……20

1.3.2.1 Stable isotope labeling with amino acids in cell culture (SILAC)……. ……..20

1.3.2.2 ^{15}N labeling for bacteria and plants……...……23

1.3.3 Other quantitative strategies for MS-based quantitative proteomics……24

1.4 Scope of this dissertation…….24

References…….27

Chapter 2. Probing Adenosine Nucleotide-binding Proteins with an Affinity Labeled-nucleotide Probe and Mass Spectrometry

Introduction…….. …….40

Materials and Methods

Materials……...43

Preparation of the biotinylated ATP probe……...43

Reaction of the biotin-ATP probe with lysine……..46

Labeling of RecA and YADH-I with the biotin-ATP probe…….47

Cell lysate preparation and labeling with the biotin-ATP probe…….. ……...47

Enzymatic digestion and affinity purification……...48

LC-MS/MS……..…….50

Data processing…….50

Results and Discussion

Reactivity of the biotin-ATP probe with lysine……51

Labeling of adenosine nucleotide-binding proteins……..53

Affinity purification and enrichment of the labeled peptides……...58

Identification of biotin-ATP probe labeled proteins in whole cell lysates……...59

Conclusions……66

References……..68

Supporting Information……..73

Chapter 3. Exploring DNA-binding Proteins with In Vivo Chemical Cross-linking and Mass Spectrometry

Introduction……79

Materials and Methods

Cell culture……82

In vivo formaldehyde cross-linking……..82

Isolation of nuclei…….83

Isolation of DNA-protein complexes……....83

Cross-linking reversal and DNA removal…….84

SDS-PAGE separation and enzymatic digestion……..85

Western blotting……85

Extraction and enzymatic digestion of nucleic acids……86

HPLC separation……...87

Nanoflow LC-MS/MS analysis…….87

Data processing…….88

Results and Discussion

Strategy for the identification of DNA-binding Proteins…….89

In Vivo DNA-protein Cross-linking…….92

Isolation of DNA-binding Proteins……...93

The selectivity of the method toward the isolation of DNA-binding proteins…….95

Identification and characterization of proteins with In vivo DNA-protein cross linking……...98

Conclusions……...…108

References……109

Supporting Information……115

Chapter 4. Quantitative Analysis of Surface Plasma Membrane Proteins of Primary and Metastatic Melanoma Cells

Introduction……..117

Materials and Methods

Cell culture……..120

Biotinylation of cell-surface membrane proteins……121

Preparation of cell-surface membrane proteins …….....121

Enzymatic digestion……122

Affinity purification ……...123

LC-MS/MS…….124

Quantitative analysis and data processing……..125

Immunocytochemistry and confocal microscopy …….126

Flow cytometric analysis ……..127

Results and Discussion

General strategy for the comparative studies of cell surface proteins……128

Proteomic analysis of membrane proteins of primary and metastatic melanoma Cells……135

Immunocytochemistry for the validation of surface localization and quantification……..140

Characterization of the differential expression of cell surface proteins…….148

Conclusions……..151

References……....153

Supporting Information……166

Chapter 5. Quantitative Proteomic Analysis of HL-60 Human Leukemia Cells upon Treatment with 6-Thioguanine

Introduction……..174

Materials and Methods

Cell culture ……176

6-TG treatment and sample preparation ……177

SDS-PAGE separation and in-gel digestion……..177

Western blotting……178

LC-MS/MS for protein identification and quantification……..179

Data processing……..180

Results and Discussion

Protein identification and quantification……182

Western blot analysis……...186

Characterization of differentially expressed proteins……..186

Conclusions……..191

References……192

Chapter 6. Summary and Future Directions……...201

LIST OF FIGURES

Figure 1.1……4

The general structure of a chemical probe and some common tag, linker and reactive groups.

Figure 1.2 ……...7

General procedures for activity-based proteomics.

Figure 1.3 ……10

Amine reactive biotinylation reagents from Pierce.

Figure 1.4 ……16

(A) The structure of original ICAT reagent (B) General procedures of ICAT experiment for quantitative proteomics.

Figure 1.5 ……19

(A) The structure of iTRAQ reagents (B) General procedures of iTRAQ experiment for quantitative proteomics.

Figure 1.6 ……22

General procedures of SILAC experiment for quantitative proteomics.

Figure 2.1 …….52

(A) Negative-ion ESI-MS of 4-h reaction mixture of biotin-ATP probe and lysine after dilution. The biotin-ATP conjugate of lysine, the product of this reaction, has an m/z of 442 in negative-ion ESI-MS. (B) Product-ion spectrum of the ion of m/z 442 observed in (A).

Figure 2.2 …….55

The product-ion spectra of the probe-labeled peptides. (A) The identified peptide with the sequence of IVEIYGPESSGK*TTLTLQVIAAAQR (m/z 981.5, triply charged) from RecA with site-directed labeling at K72; (B) the identified peptide with the sequence of VLGIDGGEGK*EELFR (m/z 958.4, doubly charged) from YADH-I with site-directed labeling at K206.

Figure 2.3 …….57

X-ray crystal structures of RecA and YADH-I. The bound ADP and NAD molecules and all lysine residues are shown in stick mode, whereas the lysine residues at the probe-labeling sites are shown in space-filling mode. (B) Structure of YADH-I complexed with NAD (PDB entry, 2HCY; only subunit A is shown).

Figure 2.4 …….60

Selected-ion chromatograms (SICs) of ions, generated from LC-MS/MS analyses, corresponding to the biotinylated peptide, VLGIDGGEGK*EELFR (m/z 958.4), and two internal standard peptides, SIGGEVFIDFTK (m/z 656.8) and

DIVGAVLK (m/z 814.5). The internal standard peptides were chosen from the unlabeled peptides identified in LC-MS/MS analysis of the tryptic digestion mixture of YADH-I without (A) and with (B) magnetic beads purification.

Figure 2.5 …….62

Biotin-ATP probe-labeled proteins in HeLa-S3 and WM-266-4 cells. (A) The identified proteins were classified according their nucleotide-binding ability. AxP: adenosine nucleotide; GxP: guanosine nucleotides; NAD(P): nicotinamide adenine dinucleotide (phosphate); RNA/DNA and others. (B) The Venn diagram shows the numbers of proteins and the overlap (intersection) of proteins identified from both cell lines.

Figure 2.6 …….63

Product ion spectra of the probe-labeled peptides from adenylate kinase 2, ALESPERPFLAILGGAK*VADK (m/z 944.1, doubly charged) with labeling site of K28 (A) and phosphoglycerate kinase 1, ALESPERPFLAILGGAK*VADK (m/z 827.2, triply charged) with labeling site of K216 (B).

Figure 2.7 …….65

The structures of adenylate kinase 2 complexed with bis(adenosine)-5'-tetraphosphate (PDB entry, 2C9Y) and yeast phosphoglycerate kinase 1 complexed with ATP (PDB entry, 3PGK) are shown in (A) and (B), respectively.

In (B), K213 at the ATP binding site of yeast PGK1 corresponds to K216 in the human counterpart.

Figure 3.1 …….91

Strategy for the identification of DNA-binding proteins with chemical cross-linking and LC-MS/MS.

Figure 3.2 …….94

Western blotting of MCM2 was performed in whole cell lysate (whole), nuclear fraction (nuclear), and the DNA-binding protein fraction (DBP) of HL-60 cells. Actin was used as the internal standard for quantitative comparison.

Figure 3.3 …….97

HPLC traces for the separation of nucleoside mixtures arising from the enzymatic digestion of nucleic acids that were isolated from the in-vivo chemically cross-linked DNA-protein complexes by the standard phenol extraction method (A) and the DNAzol method (B).

Figure 3.4 …….100

The identified proteins were classified using GO annotations. (A) Distribution of proteins in major cellular localizations; (B) Classification of proteins according to their function.

Figure 4.1 ……...134

Evaluation of the quality of quantification based on SILAC and biotinylated peptides. Selected-ion chromatograms (SICs) of two pairs of light (L) and heavy (H) biotinylated peptides of MCAM from WM-266-4 cells.

Figure 4.2 ……...……136

A typical isotopic pair of biotinylated peptides from WM-115 (light, “L”) and WM-266-4 cells (heavy, “H”) used for quantification by ESI-MS. The identities of peptides were confirmed by MS/MS acquired in data-dependant scan mode.

Figure 4.3 ……...138

Pie chart of subcellular classification of the 109 proteins identified from WM-115 and WM-266-4 cells, based on GO database annotations and previous publications.

Figure 4.4 ……...144

A summary of the quantified proteins, which were up-(black bar) and down-(gray bar) regulated in WM-266-4 with respect to WM-115 cells.

Figure 4.5 ……...145

Immunocytochemical detection of cell surface proteins. The left and right panels represent the staining of WM-115 and WM-266-4 cells, respectively. Cells were stained with primary mAbs to (A) H2B, (B) MCAM, (C) ALCAM and (D)

ITAG3 followed by staining with Alexa Fluor 488-labeled secondary antibody (green).

Figure 5.1 ……...183

A schematic illustration of comparative analysis of protein expression in HL-60 cells upon 6-TG treatment.

Figure 5.2 ……...185

A summary of the quantization results. The $\log_{1.5}$ ratios were plotted against protein numbers.

Figure 5.3 ……..188

Western blot analysis of histones H2A, H2B, H3, H4 and ferritin light polypeptide with lysates of 6-TG-treated “6-TG” and untreated “Con” HL-60 cells. Actin was used as the loading control. The black arrows indicate the histone bands confirmed with purified histones.

LIST OF SCHEMES

Scheme 2.1……... 44

Synthesis of the biotin-ATP probe. Reagents: (a) (i) isobutylchloroformate, tri-*n*-butylamine, DMF; (ii) $H_2NCH_2CH_2COOH$. (b) (i) ethylchloroformate, tri-*n*-butylamine, CH_2Cl_2/DMF; (ii) ATP, CH_2Cl_2/DMF.

Scheme 4.1 …….129

Strategy for the relative quantification of surface membrane proteins in primary and metastatic human melanoma cells.

Scheme 4.2 …….130

Cell surface protein biotinylation with sulfo-NHS-LC-biotin. The chemical reaction between the sulfo-NHS-LC-biotin and amino groups in proteins is also shown.

LIST OF TABLES

Table 3.1…….102

A list of transcription factors identified in the present study.

Table 3.2 ……105

A list of identified proteins involved in DNA repair.

Table 4.1 …….133

A list of biotinylated proteins and their peptides identified from WM-266-4 cells used to evaluate the quantification strategy based on cell surface biotinylation and affinity-enriched peptides.

Table 4.2 …….141

A list of quantified proteins and their corresponding ratios of expression in WM-266-4 cells with respect to WM-115 cells.

Table 4.3 …….147

Comparison of quantitative results from immunofluorescence staining and MS-based proteomics.

Table 5.1 …….187

Proteins quantified with changes > 1.5 or < 1.5.

LIST OF SUPPORTING INFORMATION

Table S2.1 ……73

Nucleotide-binding proteins identified with labeling from HeLa-S3 cells.

Table S2.2.....76

Nulceotide-binding proteins identified with labeling from WM-266-4 cells.

Table S4.1 ……..168

Control list of proteins from light and heavy lysates (1:1 ratio) of WM-266-4 cells.

Table S4.2 ……..171

A list of quantified proteins from WM-115 and WM-266-4 cells.

Figure S3.1 …….116

Western blotting of MCM2 in whole cell lysates (whole), nuclear fractions (nuclear), and DNA fractions of HL-60 cells. Actin was used as the internal standard for the quantitative comparison.

Figure S4.1 …….166

Control images of WM-115 and WM-266-4 cells stained with Alexa Fluor 488-conjugated rabbit secondary antibody before fixation. Cells were also stained with DAPI (blue) to identify nuclei. The left and right panels represent the staining of WM-115 and WM-266-4 cells, respectively. (A) Fluorescent image (B)

transmitted light image. Images of WM-115 and WM-266-4 cells were captured Leica TCS SP2/UV confocal microscope with identical parameters to compare the staining of the two cell lines.

Figure S4.2 …….167

Flow cytometic analysis of cell surface biotinylation. (A) Histograms of WM-115 (left panel) and WM-266-4 (right panel) cells with 1, 3, 5, 10, 20 and 30 minutes of biotinylation reaction. After reaction, the cells were stained with SAPE. The cells without labeling reaction were stained with SAPE as control (Con). (B) The fluorescent intensities of SAPE were normalized as relative intensities by using the highest signal as “1” within a serial labeling reaction of same cells.

CHAPTER 1

General Overview

1.1 Introduction

The dream of having human genome completely sequenced has become a reality.[1] However, understanding the functions of large number of proteins and the complexity of biological systems still has a long way to go. Proteomics is the qualitative and quantitative comparison of proteomes under different conditions to further unravel biological processes. Mass spectrometry (MS), as one of the most powerful techniques in analytical sciences, coupled with sample preparation and separation techniques, plays an essential role on the rise of proteomics.[2]

Although MS has served as a formidable tool for the current proteomics to identify and quantify proteins in an unprecedented scale, the proteomic samples, e.g., whole cell proteomes, tissues, body fluids, etc., are often complicated, and the complexity of these samples is far beyond the power of the state-of-the-art mass spectrometers. Many protein-based drug targets and biomarkers have very low abundance, and cannot usually be covered by the general MS-based proteomic methods. For instance, plasma, serum and cerebrospinal fluids provide a wealth of biomolecules for proteomics studies; however, many of these interesting proteins are in extremely low concentrations in these fluids.[3] It was estimated that there is a presence of $> 10^6$ different protein molecules spanning a concentration range of 11 orders of magnitude in human plasma, where 20 of the most abundant proteins constitute approximately 97-99% of the total protein mass.[4, 5]

In addition to the wide dynamic range of proteins in complex biological samples, another challenge for MS-based proteomics is the determination of the localization, interactions, and, ultimately, the function of a protein. Conventional two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and LC-MS/MS are useful for the separation, identification and relative quantification of proteins at a large scale, but these techniques have substantial limitations in the detection of proteins in low abundance due to the masking effects of abundant proteins and ion suppression during electrospray ionization.[6, 7] Hydrophobic membrane proteins are also known to be difficult to analyze by 2D gel electrophoresis.[8]

To date, various methods have been developed to meet these challenges. In this chapter, we review the strategies that implement chemical tools and stable isotope labeling techniques for MS-based proteomics.

1.2 MS-based Chemical Proteomics

Chemical proteomics is the use of chemical probes to study protein function and mechanism in a large scale.[9, 10] The interactions between proteins and small molecules are also of great interest in drug development.[11-13] So far, different chemical probes have been developed to target specific families of proteins, mainly based on their functions and/or activities.[14] Therefore, a subset of the chemical proteomics using activity-based chemical probes is known as activity-based proteomics (ABP), or activity-based protein profiling (ABPP).[12]

1.2.1 The general structure of a chemical probe

A good chemical probe is critical for the success of chemical proteomics. A typical chemical probe usually consists of three basic elements: (1) a reactive functional group that covalently conjugates with the protein active site, (2) a linker region that provides spacing and directs binding to the target, and (3) a chemical tag or reporter used for the identification, purification, or direct detection of the probe-labeled proteins.[11, 15] Figure 1.1 depicts the general structure of a chemical probe for chemical proteomics.

The reactive functional group is perhaps the most crucial component of a chemical probe. This specific chemical moiety has to assume the ability to recognize a conserved catalytic or functional motif in a family of proteins and to bind covalently to the active site of the targets. The reactive group is often derived from the electrophilic group of an irreversible enzyme inhibitor, such as an epoxide[16], fluorophosphonate[17] or sulfonate ester.[18] The linker region incorporated in a chemical probe connects the reactive group to the tag used for the identification and/or purification. Its primary function is to supply the space between the reactive group and the tag to avoid steric hindrance. The common linkers are long-chain alkyl groups or poly(ethylene glycol). In an activity-based chemical probe, the linker moiety can also influence the specificity of the probe, or provide the specificity by using a peptide or peptide-like structure.[19] In addition, the linkers of many chemical probes can be used to incorporate stable isotopes for MS-based quantitative proteomics, such as ICAT (isotope-coded affinity tag) [20], which will be discussed later in this chapter. Finally, the tag of a chemical probe facilitates the

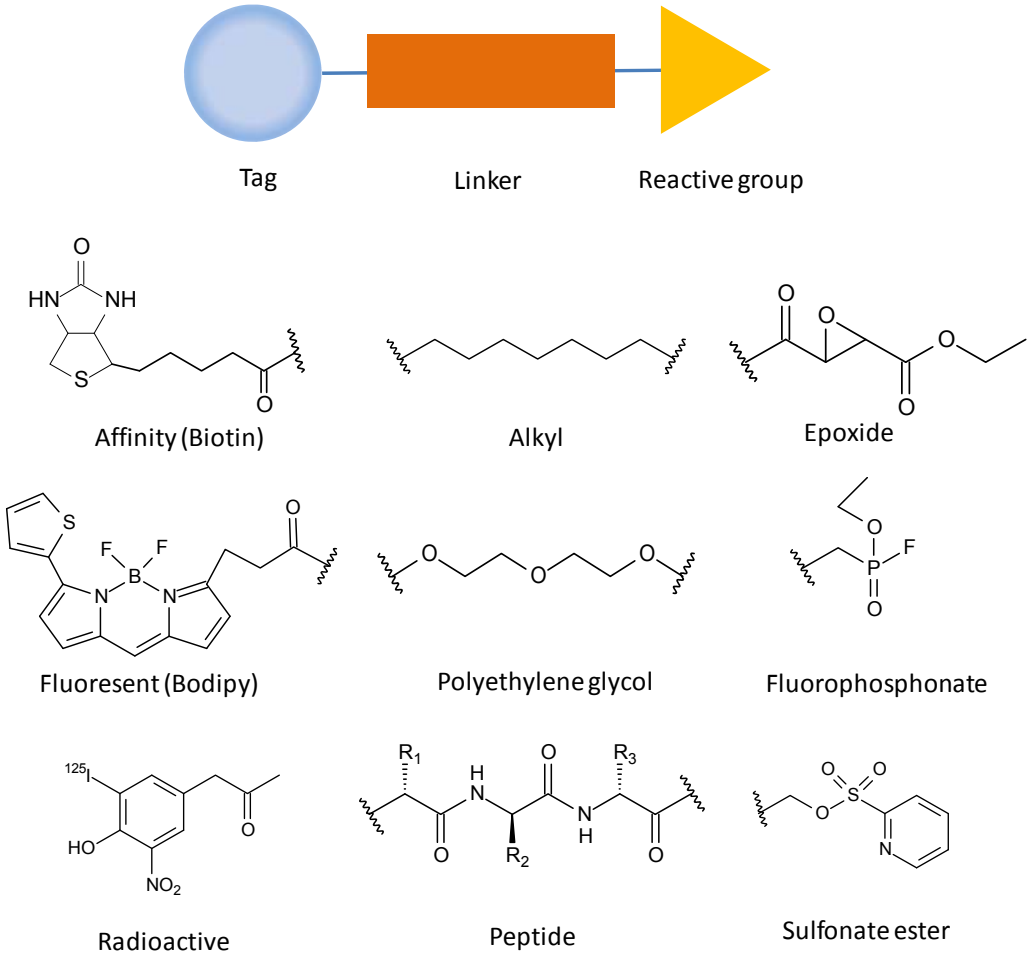


Figure 1.1 The general structure of a chemical probe and some common tag, linker and reactive groups.

purification and identification of probe-labeled proteins. Different tags are available in chemical proteomics, such as biotin, fluorescent and radioactive tags, as shown in Figure 1.1. In general, fluorescent and radioactive tags are used in gel-based screening and imaging assays,[21] in which the labeled-proteins can be visualized by scanning the gels with fluorescent scanners or phosphorimagers. These fluorescent and radioactive imaging approaches are usually simple and sensitive, which require less time to handle the samples than that of the biotin-labeled chemical probes.[22, 23] However, biotin is the most commonly used tags in chemical probes because of the availability of well-established techniques for both affinity purification and detection of biotinylated proteins, based on the strong and specific interactions between biotin and avidin/streptavidin.[23] More importantly, a biotin tag is fully compatible with MS-based proteomics, and allows the affinity isolation, purification and enrichment of biotinylated proteins or peptides.[20, 24-27]

1.2.2 Application of chemical probes in MS-based proteomics

Chemical probes have been used to label specific protein targets, facilitate the isolation, purification and identification of low-abundance proteins, and address their activities, interactions, localizations, and functions by the combination with MS-based proteomic techniques.[9, 11, 12, 15, 28]

1.2.2.1 Activity-based protein profiling

One application of chemical probes is the activity-based proteomics, or ABP.[14] Figure 1.2 depicts the general procedures of ABP. In ABP, the activity-based chemical probes have been used to identify classes of proteins based on their affinity toward the specifically designed chemical probes. For example, serine hydrolases are a family of large and diverse enzymes that include proteases, peptidases, lipases, esterases, and amidases.[29] By the regulation of growth factor activation and the metabolism of small-molecule signals, these enzymes may contribute to the malignant behavior of aggressive cancer cells. To facilitate the global analysis of serine hydrolase activities in complex proteomes, a biotinylated fluorophosphonate was synthesized by Cravatt and co-workers[17, 30] to target this family of enzymes. These studies demonstrated that this chemical probe can greatly facilitate the functional characterization and molecular identification of active enzymes in complex proteomes. Using a similar method, Jessani and colleagues[31] profiled globally serine hydrolase activities in human breast carcinoma and melanoma cell lines with fluorescent fluorophosphonate probes. Following the gel-based quantitative comparison of enzyme activities, the researchers employed biotinylated fluorophosphonate probes to enrich and identify these enzymes with avidin chromatography and LC-MS/MS, respectively.

To improve the standard ABP method, which utilizes 1D or 2D gels to separate and detect probe-labeled proteins, Adam and colleagues[32] described a general gel-free strategy utilizing activity-based chemical probes and LC-MS/MS for profiling

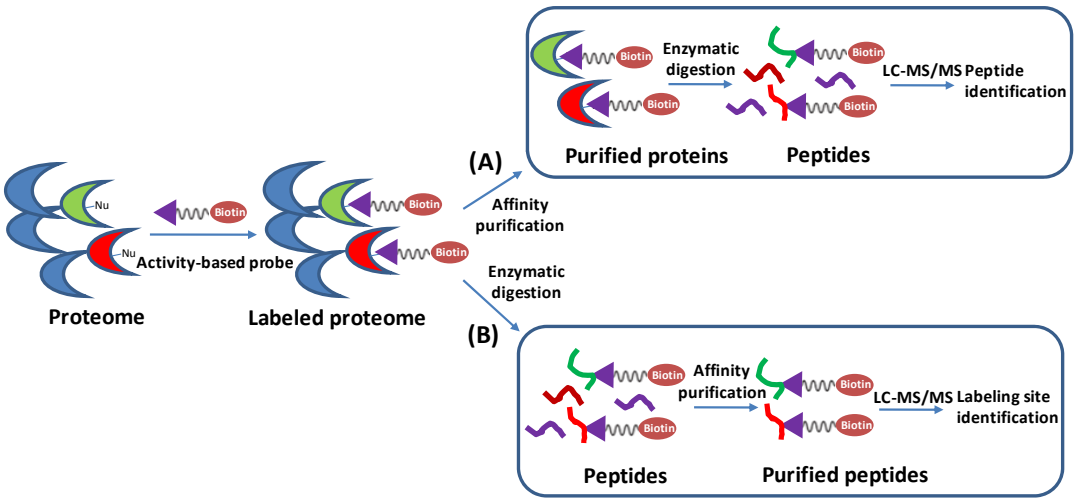


Figure 1.2 General procedures for activity-based proteomics. A proteome is first reacted with an activity-based chemical probe. After labeling reaction, the labeled proteome can be treated in two ways. (A) The labeled proteins are isolated and enriched by affinity purification. The purified proteins are subsequently digested with trypsin, and the resulting peptide mixtures are analyzed by LC-MS/MS for target identification. (B) The labeled proteins are digested with trypsin to peptides. The labeled peptides are isolated and enriched by affinity purification. After affinity purification, the resulting labeled peptide mixtures are analyzed by LC-MS/MS for target identification.

enzyme active sites in whole proteomes. In this method, the sulfonate ester probe-labeled proteomes were digested in solution with trypsin and incubated with agarose beads conjugated with avidin or anti-rhodamine antibodies to isolate probe-labeled peptides. The labeled peptides were then eluted from the affinity matrix and analyzed by LC-MS/MS to identify the protein target of the ABP probes and their labeling site. Wither and co-workers [22] also developed MS-based strategy for mechanism-based profiling of β-endoglycosidases in complex proteomes. Using a biotinylated, cleavable 2-deoxy-2-fluoroxylobioside probe, the authors isolated and identified the active-site peptides of target retaining β-1,4-glycanases.

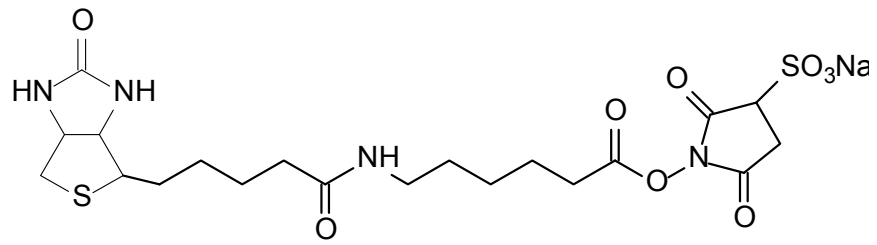
1.2.2.2 Membrane protein biotinylation

In addition to ABP, chemical probes have been used for labeling and enrichment of low-abundance proteins. For example, plasma membrane proteins play critical biological functions in cell-to-cell recognition, signal transduction and material transport.[33] Despite their important biological functions, membrane proteins are very challenging to study because of their high hydrophobicity and low abundances.[34-36] Conventional 2D gel electrophoresis is widely used for proteomic analysis, but this technique does not resolve membrane proteins very well because of the poor solubility of membrane proteins in the buffers used for the separation in the first dimension.[8] To study the membrane proteins, a group of labeling reagents for membrane protein have been developed and commercialized.[37] These biotin-conjugated chemical probes usually contain a reactive

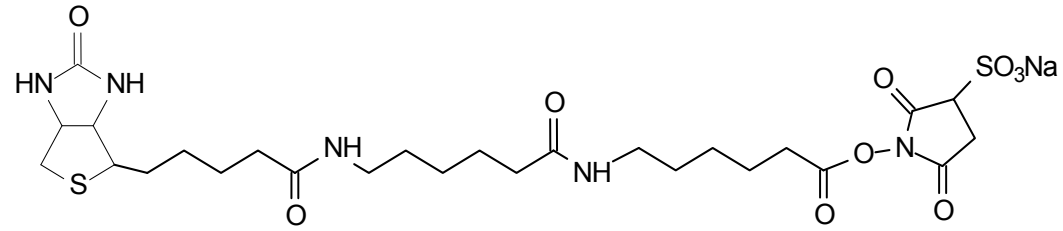
group of *N*-hydroxysuccinimide (NHS) ester (Figure 1.3), which can react with primary amines in proteins. A sulfonate-substituted NHS (i.e., *N*-hydroxysulfosuccinimide, or sulfo-NHS) is membrane-impermeable and has better water solubility than the underivatized counterpart. Sulfo-NHS-SS-biotin, containing an internal disulfide bond, is a more hydrophilic reagent than sulfo-NHS-LC-biotin.[34] The disulfide bond allows for the cleavage of labeled proteins or peptides from avidin resin upon treatment with a reducing agent. Following the labeling reaction, affinity enrichment can be performed at either the peptide or protein level; enrichment of proteins allows for broader sequence coverage, whereas enrichment of peptides usually facilitates the identification of labeled amino acid residues and peptides.

Cell-surface biotinylation has been widely used for membrane protein study[34, 35, 38] Recently, this method has been coupled with 1D or 2D gel electrophoresis and LC-MS/MS for membrane proteomics.[25-27, 39-48] Sostaric and co-workers[46] globally profiled the plasma membrane proteome of oviducal epithelial cells using sulfo-NHS-LC-biotin for membrane biotinylation. After labeling reaction, cultured cells were lysed and the labeled proteins were affinity purified. The purified proteins were then analyzed by 1D and 2D gel electrophoresis and LC-MS/MS. Among the identified proteins, 56% of which were integral membrane proteins. Zhao and colleagues[48] employed cell-surface biotinylation and affinity purification of labeled membrane proteins, along with the use of high-concentration salt or high-pH buffers, to minimize nonspecific bindings during

(A) Sulfo-NHS-LC-biotin



(B) Sulfo-NHS-LC-LC-biotin



(C) Sulfo-NHS-S-S-biotin

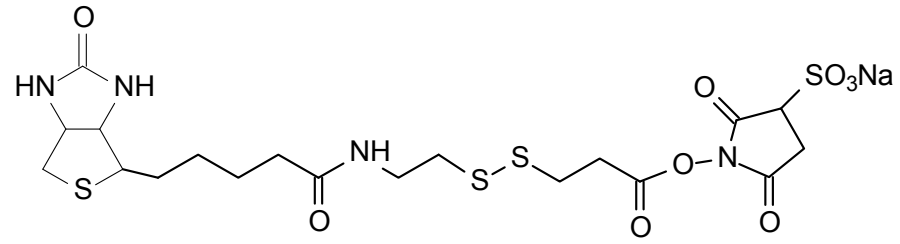


Figure 1.3 Amine-reactive biotinylation reagents from Pierce.[37] (A) EZ-Link™ Sulfo-NHS-Biotin (Sulfosuccinimidobiotin) (B) EZ-Link™ Sulfo-NHS-LC-Biotin [Sulfosuccinimidyl-6-(biotinamido)hexanoate] (C) EZ-Link™ Sulfo-NHS-SS-Biotin [Sulfosuccinimidyl 2-(biotinamido)-ethyl-1,3-dithiopropionate].

affinity purification, and they were able to identify more than 800 proteins, 87% of which were plasma membrane or membrane-associated proteins.

By perfusion or injection of biotinylation reagents, Yu and co-workers[49] labeled the apical and basolateral plasma membranes of rat kidney inner medullary collecting ducts. After biotinylation, the plasma membranes were isolated by differential centrifugation. Labeled proteins were purified by using avidin-immobilized agarose beads, and separated with 1D SDS-PAGE. After in-gel tryptic digestion, proteins were identified by LC-MS/MS analysis. A total of more than 200 plasma membrane proteins were identified in the basolateral and apical membranes. Among the identified proteins, 18% were integral or glycosylphosphatidylinositol-linked membrane proteins.

1.2.2.3 Chemical cross-linking

MS-based chemical proteomics is useful in the interrogation of interactions between biomolecules, such as protein-protein or protein-nucleic acid interactions at large scale.[50, 51] The identification of the interaction partners of a protein in a cellular system is important to understand the protein's function.[50, 52, 53] Different chemical cross-linking approaches combined with MS have been developed for the study of protein-protein interactions.[51, 54] The simplest chemical cross-linker is formaldehyde, which is highly reactive and widely used to generate protein-protein and protein-DNA cross-linking.[55, 56] For instance, Huang and colleagues[57, 58]developed an *in vivo* chemical cross-linking strategy by using formaldehyde for the quantitative study of protein-protein interactions,

including the identification of weak and/or transient interacting partners. This strategy, termed quantitative analysis of tandem-affinity purified cross-linked protein complexes or QTAX, was used to assess comprehensively the protein interaction network of the yeast 26S proteasome.[57]

The formaldehyde cross-linking reaction can occur in living cells under physiological conditions, thereby preventing subsequent dissociation and redistribution of proteins during sample preparation.[55, 58-60] Amino and imino groups of amino acids (lysine, arginine and histidine) and DNA (primarily adenine and cytosine) can react with formaldehyde and form cross-links.[56, 61] An important feature of formaldehyde cross-linking is that the cross-link is fully reversible when exposed to high temperature (> 67 ºC) at low pH in aqueous solution, mainly due to the protonation of imino groups. The disadvantage of formaldehyde cross-linking is the lack of specificity and short cross-linking distance (2 Å).

To date, different chemical cross-linking reagents with long arms or specificity have been developed for mapping protein-protein interactions.[51, 54] For example, a biotin-conjugated cross-linker was synthesized by Chu and colleagues.[62] This cross-linking reagent is homobifunctional and amine-reactive, containing two groups of NHS esters. A thiol-cleavable site was also introduced to facilitate the recovery of cross-linked peptides after an affinity purification step. The eluted cross-linked peptides can then be analyzed by MS. Other cross-linking reagents, such as heterobifunctional and trifunctional cross-linkers, have also been used for the study of protein-protein interactions.[51]

Another important field of application of chemical probes is quantitative proteomics, which is discussed in the next section.

1.3 MS-based Quantitative Proteomics

MS-based proteomics has shown to be extremely powerful in the large-scale and high-throughput identification of proteins in complex samples.[2, 63] However, quantitative analysis is highly demanded in the study of perturbation-induced changes in biological systems and in the identification of biomarkers of different diseases.[64, 65] Prior to the development of quantitative proteomic techniques, the studies of cellular changes were mainly limited to the investigation of either gross morphological changes or small numbers of proteins. Quantitative studies were heavily dependent on antibody-based immunostaining for Western blotting or fluorescent analysis.[66, 67] These methods are usually semiquantitative due to the variation in the binding affinities of antibodies. Although 2D-PAGE has been widely used to separate and compare the relative abundances of proteins in different samples, this technique is labor-intensive and often not compatible with specific classes of proteins, such as hydrophobic membrane proteins, extremely acidic or basic proteins, and very small or large proteins.[8, 68]

In the past decade, different quantitative strategies, especially stable isotope labeling techniques together with MS, have been developed for assessing quantitatively and comprehensively the differential expressions of proteomes.[64, 69] With these chemical and metabolic techniques for stable isotope labeling, MS-based quantitative proteomics is

now routinely used for the relative quantification of proteins. Recently, methods have been introduced to enable the absolute quantification of proteomes, in particular, with the development of strategies with isotope-labeled standard peptides.[70, 71] On the other hand, label-free quantification methods based on the numbers of identified peptides from different proteomes have achieved remarkable progress.[72] In this section, we summarize general and widely used approaches for MS-based quantitative proteomics.

1.3.1 Chemical tagging strategies for quantitative proteomics

These chemical approaches are based on the use of stable isotope-coded chemical reagents to differentially label proteins or peptides and to compare the relative abundances of the proteins or peptides in different samples.[9] Once proteins or peptides are differentially labeled with stable isotopes, they can be distinguished in MS or MS/MS by characteristic mass shifts. The relative abundance of labeled peptide pairs can be determined by comparing the signal intensities of isotope-labeled peptides in mass spectrum.

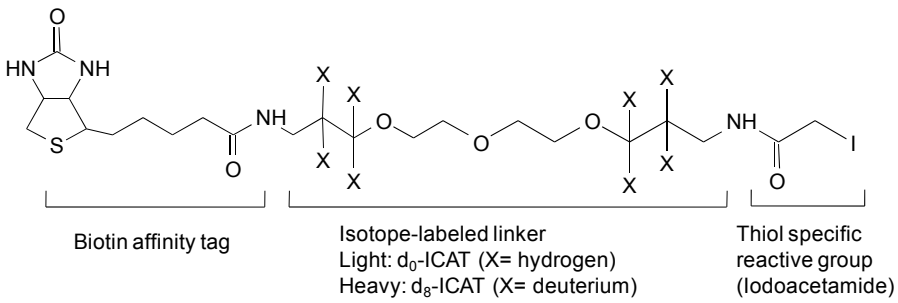
1.3.1.1 Isotope-coded affinity tag (ICAT)

The isotope-coded affinity tag (ICAT), developed by Aebersold and co-workers[20], is among the most successful chemical labeling strategies for quantitative proteomic analysis. The ICAT reagent consists of three elements: an affinity tag (biotin), which is

used to isolate and enrich ICAT-labeled peptides; a poly(ethylene glycol) linker containing stable isotope signatures, where the light form contains eight hydrogens and the heavy form carries eight deuteriums; and a reactive group with specificity toward thiol groups. The structure of ICAT is shown in Figure 1.4. Among the 20 natural amino acids, cysteine is often the target for chemical labeling owing to the available chemical reactions specific for thiol groups. For ICAT, the iodoacetyl group can react with the thiol group of cysteine side chain in a physiological pH range. In addition to stable isotope labeling, ICAT also allows the isolation of labeled peptides through the biotin-avidin affinity chromatography. The affinity purification of labeled peptides significantly reduces the sample complexity because only a relatively small number of peptides contain cysteine.

The ICAT strategy for quantifying differential protein expression is shown in Figure 1.4. Two protein mixtures representing two different states of a proteome under investigation are labeled with light and heavy ICAT reagents. Theoretically, the ICAT reagents can react with all cysteine residues in every protein. The two labeled protein mixtures are combined and enzymatically digested to peptides. The ICAT-labeled peptides are then purified using the biotin tag present in the ICAT reagents by avidin affinity chromatography.

(A)



(B)

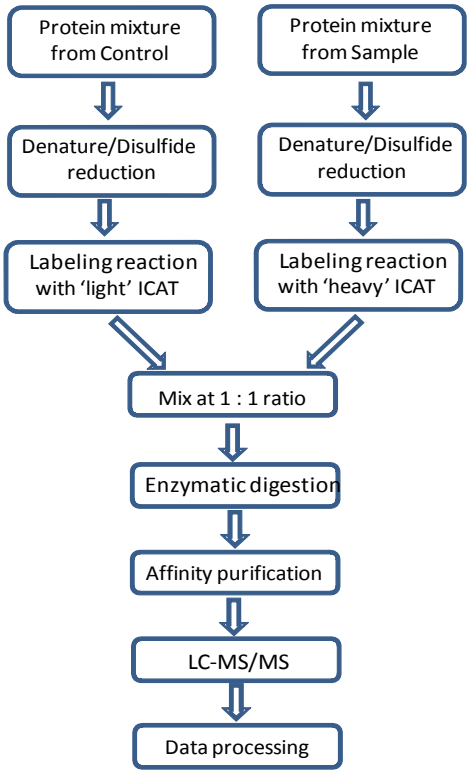


Figure 1.4 (A) The structure of original ICAT reagent, (B) general procedures of ICAT experiment for quantitative proteomics.

Since its introduction, ICAT has been widely used for the MS-based quantitative analysis.[73] However, Regnier and co-workers first found that the light (d_0) and heavy (d_8) forms of ICAT-labeled peptides can be partially resolved during reversed-phase chromatography, which could affect severely the accuracy of quantitation. Another concern is that the large moiety of biotin tag could affect the fragmentation of the labeled peptides. These problems have been resolved by using ^{13}C-coded reagent in place of the original d_8 linker, and by introducing an acid-cleavable site between biotin and the thiol-reactive group (named cICAT) that allows the removal of biotin prior to MS analysis.[24, 74] In addition, a solid-phase ICAT was also developed to improve the original ICAT by using an isotope-coded tag containing a photocleavable linker immobilized on glass beads.[75]

1.3.1.2 Isobaric tag for relative and absolute quantitation (iTRAQ)

The isobaric tag for relative and absolute quantitation (iTRAQ™) strategy was developed by Pappin and colleagues at Applied Biosystems.[76] iTRAQ has caught great attention of researchers, as it allows up to four different biological samples to be analyzed simultaneously.[77] In this approach, digested peptides from different samples to be compared are labeled with amine-reactive isobaric tags. All peptides with the same sequence but from different samples carry different versions of the tag, and are indistinguishable during LC separation and MS. The iTRAQ quantification relies on the MS/MS level rather than on the MS level by comparing the peak area of the reporter

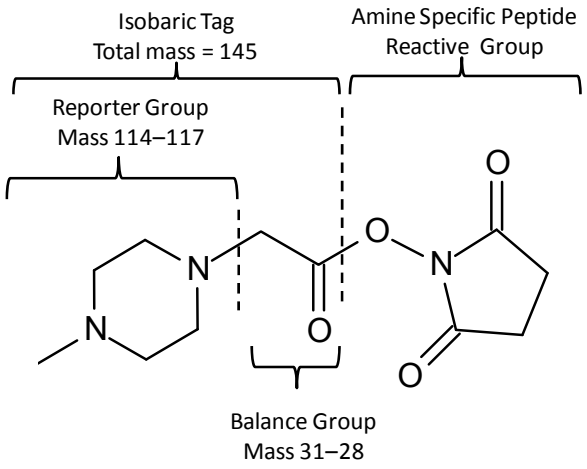
fragment ions found in the tandem mass spectrum. For this purpose, a group of amine-reactive tagging reagents were designed that are available in four isotope-coded variants with identical overall molar mass, as shown in Figure 1.5. The multiplex isobaric tags generate abundant MS/MS signature ions at *m/z* 114.1, 115.1, 116.1, and 117.1, and the relative areas of these peaks are proportional to the abundances of the labeled peptides from different samples.

There are several advantages of using iTRAQ over other isotope-labeling techniques. For instance, the isobaric tags allow quantifying multiple samples at the same time without increasing MS complexity at the peptide level, indicating that more MS/MS can be acquired for the identification and quantification. Recently, a new version of iTRAQ 8-plex labeling reagents was introduced by Applied Biosystems and can be used for the simultaneous quantification of eight samples. Besides relative quantification, iTRAQ can also be used for absolute quantification by using tagged standard peptides in the multiplex mixtures.[76]

1.3.1.3 Other chemical labeling strategies

In addition to ICAT and iTRAQ, there are many other chemical labeling techniques to introduce stable isotopes into proteins or peptides. For example, N-terminal isotope labeling with global internal standard technology (GIST)[78] and C-terminal isotope

(A)



(B)

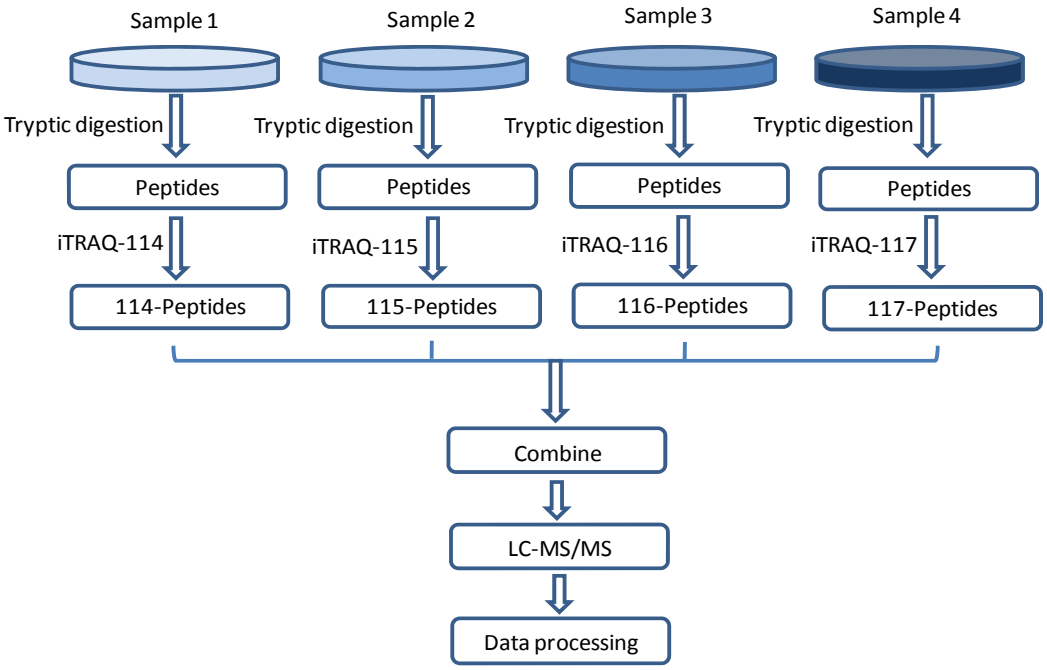


Figure 1.5 (A) The structure of iTRAQ reagents, (B) general procedures of iTRAQ experiment for quantitative proteomics.

labeling with esterification reaction[79] are also important and useful techniques for MS-based quantitative proteomic analysis.

1.3.2 Metabolic labeling strategies

Metabolic incorporation of stable isotopes has been widely used for the MS-based quantitative analysis.[64, 80] In this approach, stable isotope-labeled nutrients are added in the growth media of living cells. Through cell growth, stable isotope-labeled precursors are incorporated into all cellular proteins. A major advantage of metabolic labeling over chemical labeling is that the isotopes are present in live cells. Different cells labeled with light and heavy stable isotopes can be mixed prior to various steps of sample manipulation, such as lysis, protein extraction, fractionation and digestion, thereby minimizing differential sample loss and facilitating improved accuracy in protein quantification. The major limitation of metabolic labeling lies in that the current metabolic labeling strategies cannot be readily extended to clinical studies.

1.3.2.1 Stable isotope labeling with amino acids in cell culture (SILAC)

Stable isotope labeling with amino acids in cell culture (SILAC) is a simple metabolic or *in vivo* labeling strategy for MS-based quantitative proteomics.[81] Using SILAC, cells under two biological conditions can be grown in culture media that are identical except that one of them contains light, and the other heavy, form of particular

amino acids (e.g. ^{12}C and ^{13}C labeled L-lysine and L-arginine). The strategy of SILAC is shown in Figure 1.6. After a number of cell doublings, each instance of the particular amino acids will be replaced by their isotope-labeled analogs. If there is no chemical difference between the labeled amino acids and the natural amino acids, the cells should behave exactly like a control cell population grown in the presence of normal amino acids. Cells from both samples are then harvested, and because the isotopes are encoded directly into the amino acid sequence of each protein, the extracts can be mixed directly at certain ratios, usually in equal amounts. Protein mixtures can be further purified or fractionated, and enzymatically digested to peptides. The purified proteins and peptides will preserve the exact mixing ratio of the labeled to unlabeled proteins. The mixtures of light and heavy peptides can then be analyzed as they are analyzed in non-quantitative proteomics. Quantitation usually takes place at the level of the mass spectrum or tandem mass spectrum of peptides, similar to ICAT.[20]

SILAC is an efficient and reproducible quantitative technique, as almost complete incorporation of the isotope labels can be readily achieved. Therefore, SILAC is very suitable for the comparative study of protein expression in different cell populations, and accurate experimental results could be obtained with minimal bias, allowing relative quantification of small changes in protein abundance. A ratio of greater than 1.3-2.0 has been widely accepted as cut-offs for a change that is of statistical and biological significance.[82]

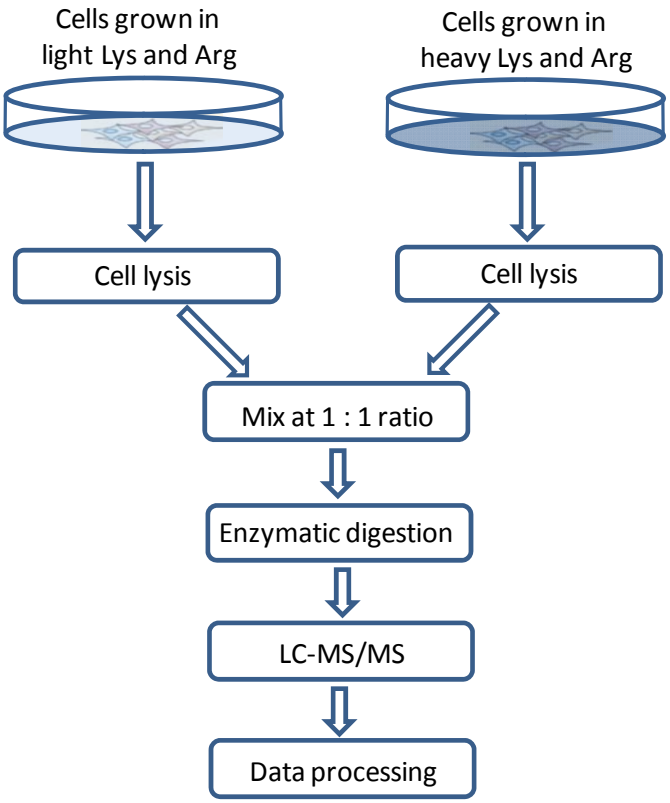


Figure 1.6 General procedures of SILAC experiment for quantitative proteomics.

In addition to the global proteome quantitation, SILAC can also be used for the target protein quantification, as well as for the quantitation in functional and subcellular proteomics.[64, 82] For example, SILAC is suitable for quantitative phosphoproteomics, which is usually difficult to study, to distinguish the differences of phosphorylation events.[83, 84] Another valuable utility of SILAC is the membrane protein quantitattion. It has been demonstrated that SILAC is suitable for the quantitative analysis of cell membrane proteins,[85] where ICAT has been rarely employed, probably due to the poor compatibility of ICAT with the high concentration of detergent required for dissolving membrane proteins.[86]

1.3.2.2 ^{15}N labeling for bacteria and plants

In fact, the approach utilizing ^{14}N and ^{15}N minimal media for bacteria and plants has been well established before SILAC method.[87-89] The procedures developed by Oda and colleagues [88] in 1999 can be generally applied to any cell systems including mammalian cell lines. In the case of ^{15}N labeling, the heavy isotope is not associated with a specific amino acid but distributed over the whole protein, resulting in an undefined mass difference between the labeled and unlabeled proteins. The mass difference varies with the nitrogen content of an individual protein or peptide. ^{15}N-labeling has the similar advantage of SILAC to label the samples at an early stage; however, the major disadvantage is the variable mass difference, which makes the data interpretation complicated.

1.3.3 Other quantitative strategies for MS-based quantitative proteomics

Aside from the techniques discussed above, there are many approaches that are used for MS-based quantitative proteomics. These methods include ^{18}O-labeling,[90] mass-coded abundance tagging (MCAT),[91] label-free [92] and absolute [93] quantitative strategies.

1.4 Scope of This Dissertation

As the preceding discussion has shown, many innovative and promising techniques have been developed for the MS-based chemical and quantitative proteomics. These chemical and quantitative strategies have greatly facilitated the analysis of global protein expression, as well as the isolation, purification, detection and quantitative analysis of low-abundance proteins in complex biological samples. However, these techniques also have their own limitations, and are far from enough to address all challenges in the complicated biological and clinical questions. More efforts and advancements still need to be made to improve the current techniques in MS-based proteomics.

In this dissertation, we focus on the development of chemical and quantitative strategies for MS-based proteomics to identify and quantify several specific classes of proteins, including nucleotide-binding proteins, nucleic acid-binding proteins and plasma membrane proteins, as well as for the proteomic study of the response of human cells toward anticancer drug treatment.

In Chapter two, we describe the application of a chemical probe, biotin-conjugated acyl nucleotide, for probing adenosine nucleotide-binding proteins. We demonstrate that this chemical probe reacts specifically with the lysine residue at the nucleotide-binding site of two purified adenosine nucleotide-binding proteins. This method allowed us to enrich and identify nucleotide-binding proteins from complex samples, e.g., whole cell lysates. The strategy, involving labeling reaction, enzymatic digestion, affinity purification, and LC-MS/MS analysis, is relatively simple, fast, and straightforward.

In addition to nucleotide-binding proteins, we also investigated nucleic acid-binding proteins. In Chapter three, we describe an approach for the comprehensive investigation of DNA-binding proteins with *in vivo* chemical cross-linking and LC-MS/MS. DNA-binding proteins could be purified via the isolation of DNA-protein complexes and released from the complexes by reversing the cross-linking. By using this method, we were able to identify more than one hundred DNA-binding proteins, such as proteins involved in transcription, gene regulation, DNA replication and repair, and a large number of proteins that are potentially associated with DNA and DNA-binding proteins.

In Chapter four, a strategy, including SILAC, cell surface biotinylation, affinity peptide purification and LC-MS/MS, for the identification and quantification of cell-surface membrane proteins is described. With this strategy, we identified over 100 membrane and its associated proteins from a pair of melanoma cell lines derived from the primary and metastatic tumor sites of the same individual. Moreover, we quantified the relative expression of 62 identified proteins in these two cell lines.

In Chapter five, a comparative study of protein expression in HL-60 human leukemia cells upon 6-thioguanine treatment was performed by using a MS-based proteomic method together with SILAC. Our results revealed that the expression levels of fifteen proteins, such as ferritin, and major histones H1, H2A, H2B, H3, and H4, were significantly altered upon treatment with 6-thioguanine. The biological implications of the changed expression of some proteins were discussed.

References

1. Collins, F. S.; Lander, E. S.; Rogers, J.; Waterston, R. H.; Conso, I. H. G. S., Finishing the euchromatic sequence of the human genome. *Nature* **2004,** 431, 931-945.
2. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** 422, 198-207.
3. Hanash, S. M.; Pitteri, S. J.; Faca, V. M., Mining the plasma proteome for cancer biomarkers. *Nature* **2008,** 452, 571-579.
4. Anderson, N. L.; Anderson, N. G., The human plasma proteome - History, character, and diagnostic prospects. *Mol. Cell. Proteomics* **2002,** 1, 845-867.
5. Anderson, L., Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J. Physiol.* **2005,** 563, 23-60.
6. Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C., The dynamic range of protein expression: A challenge for proteomic research. *Electrophoresis* **2000,** 21, 1104-1115.
7. Harry, J. L.; Wilkins, M. R.; Herbert, B. R.; Packer, N. H.; Gooley, A. A.; Williams, K. L., Proteomics: Capacity versus utility. *Electrophoresis* **2000,** 21, 1071-1081.
8. Santoni, V.; Molloy, M.; Rabilloud, T., Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* **2000,** 21, 1054-1070.
9. Leitner, A.; Lindner, W., Chemistry meets proteomics: The use of chemical tagging reactions for MS-based proteomics. *Proteomics* **2006,** 6, 5418-5434.

10. Smith, J. C.; Lambert, J. P.; Elisma, F.; Figeys, D., Proteomics in 2005/2006: Developments, applications and challenges. *Anal. Chem.* **2007,** 79, 4325-4343.

11. Jeffery, D. A.; Bogyo, M., Chemical proteomics and its application to drug discovery. *Curr.Opin. Biotechnol.* **2003,** 14, 87-95.

12. Speers, A. E.; Cravatt, B. F., Chemical strategies for activity-based proteomics. *Chembiochem* **2004,** 5, 41-47.

13. Verhelst, S. H. L.; Bogyo, M., Chemical proteomics applied to target identification and drug discovery. *Biotechniques* **2005,** 38, 175-177.

14. Cravatt, B. E.; Wright, A. T.; Kozarich, J. W., Activity-based protein profiling: From enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* **2008,** 77, 383-414.

15. Paulick, M. G.; Bogyo, M., Application of activity-based probes to the study of enzymes involved in cancer progression. *Curr. Opin. Genet. Dev.* **2008,** 18, 97-106.

16. Greenbaum, D.; Medzihradszky, K. F.; Burlingame, A.; Bogyo, M., Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **2000,** 7, 569-581.

17. Liu, Y. S.; Patricelli, M. P.; Cravatt, B. F., Activity-based protein profiling: The serine hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** 96, 14694-14699.

18. Adam, G. C.; Sorensen, E. J.; Cravatt, B. F., Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype. *Nat. Biotechnol.* **2002,** 20, 805-809.

19. Borodovsky, A.; Ovaa, H.; Kolli, N.; Gan-Erdene, T.; Wilkinson, K. D.; Ploegh, H. L.; Kessler, B. M., Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme. *Chem. Biol.* **2002,** 9, 1149-1159.

20. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999,** 17, 994-999.

21. Patricelli, M. P.; Giang, D. K.; Stamp, L. M.; Burbaum, J. J., Direct visualization of serine hydrolase activities in complex proteomes using fluorescent active site-directed probes. *Proteomics* **2001,** 1, 1067-1071.

22. Hekmat, O.; Kim, Y. W.; Williams, S. J.; He, S. M.; Withers, S. G., Active-site peptide "fingerprinting" of glycosidases in complex mixtures by mass spectrometry - Discovery of a novel retaining beta-1,4-glycanase in Cellulomonas fimi. *J. Biol. Chem.* **2005,** 280, 35126-35135.

23. Bayer, E. A.; Wilchek, M., Application of avidin biotin technology to affinity-based separations. *J. Chromatogr.* **1990,** 510, 3-11.

24. Hansen, K. C.; Schmitt-Ulms, G.; Chalkley, R. J.; Hirsch, J.; Baldwin, M. A.; Burlingame, A. L., Mass spectrometric analysis of protein mixtures at low levels using cleavable C-13-isotope-coded affinity tag and multidimensional chromatography. *Mol. Cell. Proteomics* **2003,** 2, 299-314.

25. Chen, W. N. U.; Yu, L. R.; Strittmatter, E. F.; Thrall, B. D.; Camp, D. G.; Smith, R. D., Detection of in situ labeled cell surface proteins by mass spectrometry:

Application to the membrane subproteome of human mammary epithelial cells. *Proteomics* **2003,** 3, 1647-1651.

26. Goshe, M. B.; Blonder, J.; Smith, R. D., Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. *J. Proteome Res.* **2003,** 2, 153-161.
27. Zhang, W.; Zhou, G.; Zhao, Y. X.; White, M. A.; Zhao, Y. M., Affinity enrichment of plasma membrane for proteomics analysis. *Electrophoresis* **2003,** 24, 2855-2863.
28. Greenbaum, D.; Baruch, A.; Hayrapetian, L.; Darula, Z.; Burlingame, A.; Medzihradszky, K. F.; Bogyo, M., Chemical approaches for functionally probing the proteome. *Mol. Cell. Proteomics* **2002,** 1, 60-68.
29. Rubin, G. M.; Yandell, M. D., Comparative genomics of the eukaryotes. *Science* **2000,** 287, 2204-2215.
30. Kidd, D.; Liu, Y. S.; Cravatt, B. F., Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **2001,** 40, 4005-4015.
31. Jessani, N.; Liu, Y. S.; Humphrey, M.; Cravatt, B. F., Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** 99, 10335-10340.
32. Adam, G. C.; Burbaum, J.; Kozarich, J. W.; Patricelli, M. P.; Cravatt, B. F., Mapping enzyme active sites in complex proteomes. *J. Am. Chem. Soc.* **2004,** 126, 1363-1368.

33. Wallin, E.; von Heijne, G., Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **1998,** 7, 1029-1038.

34. Speers, A. E.; Wu, C. C., Proteomics of integral membrane proteins-theory and application. *Chem. Rev.* **2007,** 107, 3687-3714.

35. Josic, D.; Clifton, J. G., Mammalian plasma membrane proteomics. *Proteomics* **2007,** 7, 3010-3029.

36. Tan, S.; Tan, H. T.; Chung, M. C. M., Membrane proteins and membrane proteomics. *Proteomics* **2008,** 8, 3924-3932.

37. Thermo Fisher Scientific. http://www.piercenet.com/Products/Browse.cfm?fldID=8D38BA83-EFDC-421A-853F-E96EBA380612

38. DelaFuente, E. K.; Dawson, C. A.; Nelin, L. D.; Bongard, R. D.; McAuliffe, T. L.; Merker, M. P., Biotinylation of membrane proteins accessible via the pulmonary circulation in normal and hyperoxic rats. *Am. J. Physiol.* **1997,** 16, L461-L470.

39. Hastie, C.; Saxton, M.; Akpan, A.; Cramer, R.; Masters, J. R.; Naaby-Hansen, S., Combined affinity labelling and mass spectrometry analysis of differential cell surface protein expression in normal and prostate cancer cells. *Oncogene* **2005,** 24, 5905-5913.

40. Jang, J. H.; Hanash, S., Profiling of the cell surface proteome. *Proteomics* **2003,** 3, 1947-1954.

41. Lee, S. J.; Kim, K. H.; Park, J. S.; Jung, J. W.; Kim, Y. H.; Kim, S. K.; Kim, W. S.; Goh, H. G.; Kim, S. H.; Yoo, J. S.; Kim, D. W.; Kim, K. P., Comparative analysis of cell surface proteins in chronic and acute leukemia cell lines. *Biochem. Biophys. Res. Commun.* **2007,** 357, 620-626.

42. Mojica, W.; Sun, J. L., Plasma membrane protein identification by cell surface biotinylation and tandem mass spectrometry of clinical specimens. *Mod. Pathol.* **2006,** 19, 162-162.

43. Nunomura, K.; Nagano, K.; Itagaki, C.; Taoka, M.; Okamura, N.; Yamauchi, Y.; Sugano, S.; Takahashi, N.; Izumi, T.; Isobe, T., Cell surface labeling and mass spectrometry reveal diversity of cell surface markers and signaling molecules expressed in undifferentiated mouse embryonic stem cells. *Mol. Cell. Proteomics* **2005,** 4, 1968-1976.

44. Scheurer, S. B.; Rybak, J. N.; Roesli, C.; Brunisholz, R. A.; Potthast, F.; Schlapbach, R.; Neri, D.; Elia, G., Identification and relative quantification of membrane proteins by surface biotinylation and two-dimensional peptide mapping. *Proteomics* **2005,** 5, 2718-2728.

45. Shin, B. K.; Wang, H.; Yim, A. M.; Le Naour, F.; Brichory, F.; Jang, J. H.; Zhao, R.; Puravs, E.; Tra, J.; Michael, C. W.; Misek, D. E.; Hanash, S. M., Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function. *J. Biol. Chem.* **2003,** 278, 7607-7616.

46. Sostaric, E.; Georgiou, A. S.; Wong, C. H.; Watson, P. F.; Holt, W. V.; Fazeli, A., Global profiling of surface plasma membrane proteome of oviductal epithelial cells. *J. Proteome Res.* **2006,** 5, 3029-3037.

47. Tang, X. T.; Yi, W.; Munske, G. R.; Adhikari, D. P.; Zakharova, N. L.; Bruce, J. E., Profiling the membrane proteome of Shewanella oneidensis MR-1 with new affinity labeling probes. *J. Proteome Res.* **2007,** 6, 724-734.

48. Zhao, Y. X.; Zhang, W.; Kho, Y. J.; Zhao, Y. M., Proteomic analysis of integral plasma membrane proteins. *Anal. Chem.* **2004,** 76, 1817-1823.

49. Yu, M. J.; Pisitkun, T.; Wang, G. H.; Shen, R. F.; Knepper, M. A., LC-MS/MS analysis of apical and basolateral plasma membranes of rat renal collecting duct cells. *Mol. Cell. Proteomics* **2006,** 5, 2131-2145.

50. Monti, M.; Orru, S.; Pagnozzi, D.; Pucci, P., Interaction proteomics. *Biosci. Rep.* **2005,** 25, 45-56.

51. Sinz, A., Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* **2006,** 25, 663-682.

52. Burckstummer, T.; Bennett, K. L.; Preradovic, A.; Schutze, G.; Hantschel, O.; Superti-Furga, G.; Bauch, A., An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat. Methods* **2006,** 3, 1013-1019.

53. Moll, D.; Prinz, A.; Gesellchen, F.; Drewianka, S.; Zimmermann, B.; Herberg, F. W., Biomolecular interaction analysis in functional proteomics. *J. Neural. Transm.* **2006,** 113, 1015-1032.

54. Trakselis, M. A.; Alley, S. C.; Ishmael, F. T., Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics. *Bioconjugate Chem.* **2005,** 16, 741-750.

55. Vasilescu, J.; Guo, X. C.; Kast, J., Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics* **2004,** 4, 3845-3854.

56. Solomon, M. J.; Varshavsky, A., Formaldehyde-mediated DNA protein crosslinking - a probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. U.S.A.* **1985,** 82, 6470-6474.

57. Guerrero, C.; Milenkovic, T.; Przulj, N.; Kaiser, P.; Huang, L., Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, 13333-13338.

58. Guerrero, C.; Tagwerker, C.; Kaiser, P.; Huang, L., An integrated mass spectrometry-based proteomic approach - Quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. *Mol. Cell. Proteomics* **2006,** 5, 366-378.

59. Jackson, V., Formaldehyde cross-linking for studying nucleosomal dynamics. *Methods* **1999,** 17, 125-139.

60. Orlando, V.; Strutt, H.; Paro, R., Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods* **1997,** 11, 205-214.

61. Orlando, V., Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* **2000,** 25, 99-104.

62. Chu, F. X.; Mahrus, S.; Craik, C. S.; Burlingame, A. L., Isotope-coded and affinity-tagged cross-linking (ICATXL): An efficient strategy to probe protein interaction surfaces. *J. Am. Chem. Soc.* **2006,** 128, 10362-10363.

63. Cravatt, B. F.; Simon, G. M.; Yates, J. R., The biological impact of mass-spectrometry-based proteomics. *Nature* **2007,** 450, 991-1000.

64. Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **2005,** 1, 252-262.

65. Ahn, N. G.; Shabb, J. B.; Old, W. M.; Resing, K. A., Achieving in-depth proteomics profiling by mass spectrometry. *ACS Chem. Biol.* **2007,** 2, 39-52.

66. Towbin, H.; Staehelin, T.; Gordon, J., Electrophoretic Transfer of Proteins from Polyacrylamide Gels to Nitrocellulose Sheets - Procedure and Some Applications. *Proc. Natl. Acad. Sci. U. S. A.* **1979,** 76, 4350-4354.

67. Ramos-Vara, J. A., Technical aspects of immunohistochemistry. *Vet. Pathol.* **2005,** 42, 405-426.

68. Garrels, J. I.; McLaughlin, C. S.; Warner, J. R.; Futcher, B.; Latter, G. I.; Kobayashi, R.; Schwender, B.; Volpe, T.; Anderson, D. S.; MesquitaFuentes, R.; Payne, W. E., Proteome studies of Saccharomyces cerevisiae: Identification and characterization of abundant proteins. *Electrophoresis* **1997,** 18, 1347-1360.

69. Julka, S.; Regnier, F., Quantification in proteomics through stable isotope coding: A review. *J. Proteome Res.* **2004,** 3, 350-363.

70. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** 100, 6940-6945.

71. Kirkpatrick, D. S.; Gerber, S. A.; Gygi, S. P., The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* **2005,** 35, 265-273.

72. Roddy, T. P.; Horvath, C. R.; Stout, S. J.; Kenney, K. L.; Ho, P. I.; Zhang, J. H.; Vickers, C.; Kaushik, V.; Hubbard, B.; Wang, Y. K., Mass spectrometric techniques for label-free high-throughput screening in drug discovery. *Anal. Chem.* **2007,** 79, 8207-8213.

73. Leitner, A.; Lindner, W., Current chemical tagging strategies for proteome analysis by mass spectrometry. *J. Chromatogr. B* **2004,** 813, 1-26.

74. Li, J. X.; Steen, H.; Gygi, S. P., Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents - The yeast salinity stress response. *Mol. Cell. Proteomics* **2003,** 2, 1198-1204.

75. Zhou, H. L.; Ranish, J. A.; Watts, J. D.; Aebersold, R., Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat. Biotechnol.* **2002,** 20, 512-515.

76. Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed

protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004,** 3, 1154-1169.

77. Wu, W. W.; Wang, G. H.; Baek, S. J.; Shen, R. F., Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *J. Proteome Res.* **2006,** 5, 651-658.

78. Chakraborty, A.; Regnier, F. E., Global internal standard technology for comparative proteomics. *J. Chromatogr. A* **2002,** 949, 173-184.

79. Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J. K.; von Haller, P.; Aebersold, R.; Kolker, E., Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid. Commun. Mass Spectrom.* **2001,** 15, 1214-1221.

80. Beynon, R. J.; Pratt, J. M., Metabolic labeling of proteins for proteomics. *Mol. Cell. Proteomics* **2005,** 4, 857-872.

81. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002,** 1, 376-386.

82. Mann, M., Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **2006,** 7, 952-958.

83. Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N., Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* **2005,** 4, 310-327.

84. Zhang, G. A.; Spellman, D. S.; Skolnik, E. Y.; Neubert, T. A., Quantitative phosphotyrosine proteomics of EphB2 signaling by stable isotope labeling with amino acids in cell culture (SILAC). *J. Proteome Res.* **2006,** 5, 581-588.

85. Liang, X. Q.; Zhao, J.; Hajivandi, M.; Wu, R.; Tao, J.; Amshey, J. W.; Pope, R. M., Quantification of membrane and membrane-bound proteins in normal and malignant breast cancer cells isolated from the same patient with primary breast carcinoma. *J. Proteome Res.* **2006,** 5, 2632-2641.

86. Smolka, M. B.; Zhou, H. L.; Purkayastha, S.; Aebersold, R., Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal. Biochem.* **2001,** 297, 25-31.

87. Kohno, T.; Kusunoki, H.; Sato, K.; Wakamatsu, K., A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: An application to the production of mastoparan-X uniformly enriched with N-15 and N-15/C-13. *J. Biomol. NMR* **1998,** 12, 109-121.

88. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** 96, 6591-6596.

89. Putz, S.; Reinders, J.; Reinders, Y.; Sickmann, A., Mass spectrometry-based peptide quantification: applications and limitations. *Expert Rev. Proteomics* **2005,** 2, 381-392.

90. Yao, X. D.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C., Proteolytic O-18 labeling for comparative proteomics: Model studies with two serotypes of adenovirus. *Anal. Chem.* **2001,** 73, 2836-2842.

91. Cagney, G.; Emili, A., De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nat. Biotechnol.* **2002,** 20, 163-170.

92. Ono, M.; Shitashige, M.; Honda, K.; Isobe, T.; Kuwabara, H.; Matsuzuki, H.; Hirohashi, S.; Yamada, T., Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. *Mol. Cell. Proteomics* **2006,** 5, 1338-1347.

93. Bronstrup, M., Absolute quantification strategies in proteomics based on mass spectrometry. *Expert Rev. Proteomics* **2004,** 1, 503-512.

CHAPTER 2

Probing Adenosine Nucleotide-binding Proteins with an Affinity Labeled-nucleotide Probe and Mass Spectrometry

Introduction

Mass spectrometry (MS) has progressed extremely rapidly during the past two decades. The application of MS to the identification of chemical compounds in a mixture, including determining the structural composition of large biomolecules, becomes increasingly popular.[1] When the analysis is directed toward complex biological mixtures or protein functional investigations, a few challenges, such as sample complexity and quantitation, are encountered when MS is used alone. Fortunately, this can be overcome, to some extent, by combining MS with powerful separation techniques, such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), in which proteins are separated based on their isoelectric points and molecular masses, or LC-based strategies, e.g., the multidimensional protein identification technology (MudPIT).[2, 3]

Aside from these technologies, chemical tagging methods that involve the modification of functional groups of amino acid residues in proteins and peptides have been described.[4] These chemical tagging or labeling reagents target specific amino acid residues or post-translational modifications (PTMs), which facilitate the enrichment of subfractions of interest via affinity purification. When stable isotope-labeled tags are

employed, relative quantitation of protein expression can be readily achieved. In this context, the isotope-coded affinity tag (ICAT) has become widely used.[5] Only those peptides containing certain amino acids (in this case, cysteine) can be targeted; an affinity tag, usually containing a biotin moiety, is attached to the functional group of interest via covalent linkage, which allows for reducing the sample complexity by affinity purification. However, these chemical tags, by measuring protein abundance, lack specificity for functional study of proteins, especially for various enzymes. To address this limitation, Cravatt and co-workers[6, 7] developed a series of activity-based chemical tagging approaches, known as activity-based protein profiling, for functional proteomic studies. For instance, they reported an LC-MS strategy to identify the sites of labeling on several enzymes targeted by sulfonate ester probes.[8] In this approach, proteomes were treated with a rhodamine-tagged phenyl sulfonate ester, followed by denaturation, thiol reduction, alkylation, and trypsin digestion. The peptide mixture was then incubated with an affinity capture matrix to isolate specifically the probe-labeled peptides for the subsequent LC-MS/MS analysis.

In addition to the sulfonate ester probes, a variety of nucleotide analogues, which are usually fluorescent, photoactive, or affinity-labeled, have been developed for different applications.[9-11] Among these nucleotide analogues, ATP derivatives are the most widely used because ATP is essential for almost all living organisms and it is a substrate for numerous enzymes and ATP-binding proteins. For example, 5'-*p*-fluorosulfonylbenzoyladenosine (FSBA), a reactive ATP analogue, has been used as an activity-based probe to target nucleotide-binding proteins.[12] In addition, a class of acyl

nucleotide probes, as well as the methods for their synthesis and use in proteomic analysis, has been described.[13] These probes, termed tagged acyl phosphate probes (TAPPs) or more specifically nucleotide-binding protein-directed affinity probes (NBAPs), were claimed to provide enhanced simplicity and accuracy in identifying changes in the presence, amount, or activity of target proteins in a complex protein mixture, preferably nucleotide-binding proteins including protein kinases. Recently, several hundred protein kinases were identified by using these acyl phosphate probes in an analysis of more than 100 proteomes including human, mouse, rat, and dog, which has demonstrated the ability of the acyl phosphate probes to broadly profile protein kinases in native proteomes.[14]

In this work, we chose two adenosine nucleotide-binding proteins, *Escherichia coli* recombinase A (RecA), an ATP/ADP-binding protein, and *Saccharomyces cerevisiae* alcohol dehydrogenase-I (YADH-I), a nicotinamide adenine dinucleotide (NAD)-binding protein, to demonstrate the utility of the affinity-labeled acyl phosphate probe with MS in elucidating protein structure and probing nucleotide-binding sites. We also applied the probe to profile the nucleotide-binding proteins in cell lysates. The method shows the potential application of this probe in the purification, enrichment and identification of nucleotide-binding proteins from whole-cell lysates.

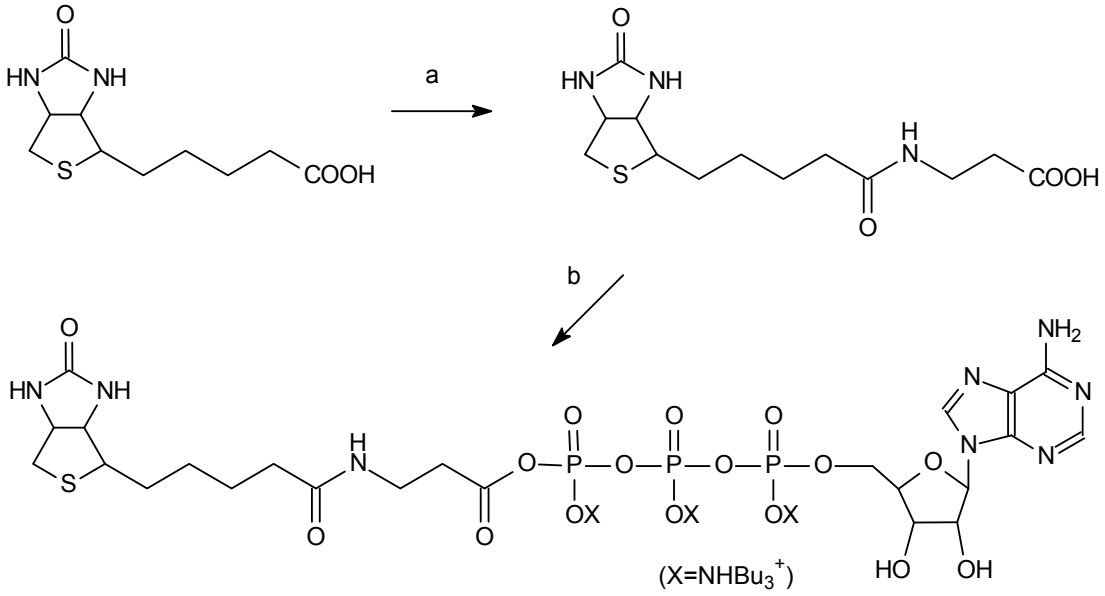
Materials and Methods

Materials

ATP, in disodium salt form, was from MP Biochemicals (Solon, OH). *D*-Biotin was purchased from AnaSpec Inc. (San Jose, CA), and β-alanine was from TCI America (Portland, OR). *E. coli* RecA protein was obtained from Epicentre Biotechnologies (Madison, WI). YADH-I and bovine serum albumin (BSA) were purchased from Sigma-Aldrich (St. Louis, MO) and BioRad (Hercules, CA), respectively. These proteins were used without further purification. Streptavidin-conjugated magnetic particles and sequencing-grade modified trypsin were obtained from Roche Applied Science (Indianapolis, IN). Common reagents for synthesis were obtained from VWR. Other reagents, unless otherwise specified, were from Sigma-Aldrich.

Preparation of the Biotinylated ATP Probe

The general methods for the preparation of acyl nucleotide probes were described previously.[13] Briefly, the biotinylated ATP probe was synthesized via a two-step procedure as outlined in Scheme 2.1. In this regard, we first prepared 3-*N*-d-biotinylaminopropanoic acid following previously published procedures.[15, 16] Isobutyl chloroformate (0.32 mL, 2.47 mmol) was added to a solution containing biotin (500 mg, 2.05 mmol), DMF (40 mL), and tri-*n*-butylamine (0.64 mL, 2.69 mmol). After being incubated at room temperature for 10 min, the mixture was slowly added to a suspension of β-alanine (4.1 mmol) in DMF (40mL) at 5 °C. After stirring at 5 °C for 2 h, the solvent



Scheme 2.1 Synthesis of the biotin-ATP probe. Reagents: (a) (i) isobutylchloroformate, tri-*n*-butylamine, DMF; (ii) $H_2NCH_2CH_2COOH$. (b) (i) ethylchloroformate, tri-*n*-butylamine, CH_2Cl_2/DMF; (ii) ATP, CH_2Cl_2/DMF.

was removed under reduced pressure and the crude product was dissolved in a 36-mL water and ethanol (1:1, v/v) solution at 40 °C. The solution pH was adjusted to 2 with 2.0 N HCl and the mixture was kept at 0 °C for 12 h. The desired 3-*N*-d-biotinylaminopropanoic acid was precipitated out of solution. The precipitate was filtered, washed with water, and dried under vacuum.

To make ATP soluble in organic solvent, the commercially available disodium salt of ATP was converted to the tributylammonium form by passing it through a cation exchange column packed with Spectra/Gel IE 50×8 resin (40-75 μm), in which the H^+ was displaced by tributylammonium ion ($NHBu_3^+$). The exchange was performed by passing through the column twice at 4 °C, and the collected ATP-$NHBu_3^+$ fractions were lyophilized to form white solids.

The acyl adenosine nucleotide was synthesized following the procedures described by Kreimeyer et al.[17] with minor modifications. Briefly, 3-*N*-d-biotinylaminopropanoic acid (94.5 mg, 0.3 mmol) was dissolved in a 2.5-mL solvent mixture of dry CH_2Cl_2 and DMF (4:1, v/v). To the ice-cold solution were then added tri-*n*-butylamine (0.11 mL, 0.45 mmol) and ethyl chloroformate (0.03 mL, 0.3 mmol). After stirring at 0 °C for 5 min, the mixture was stirred at room temperature under argon atmosphere for another 60 min. ATP-$NHBu_3^+$ (0.1 mmol), dissolved in a 2.5-mL solution of CH_2Cl_2 and DMF (4:1, v/v), was then added to the above reaction mixture. The reaction was continued at room temperature and under argon atmosphere for 18 h, and to the reaction mixture was added toluene (5 mL), which produced white precipitates. The solid was collected by filtration

and dried under vacuum. To the filtrate was added water (2 mL), and the solution was extracted with toluene. The aqueous phase was collected, lyophilized, and combined with the precipitate.

The resulting solid products were purified by HPLC with a YMC ODS-AQ column (4.8×250 mm, 120 Å in pore size, 5 μm in particle size, Waters). The flow rate was set at 0.8 mL/min, and a 45-min linear gradient of 0—30% acetonitrile in 50 mM triethylammonium acetate (TEAA, pH 6.8) was used for the purification. A UV detector was set at 265 nm for monitoring the effluents. Appropriate HPLC fractions were collected, lyophilized, and stored at –20 ºC. The product was confirmed by phosphorus NMR and ESI-MS. ^{31}P NMR (121 MHz, D_2O, proton decoupled): δ (in ppm) −10.5 (d, 1P, $J_{P\text{-}P}$ = 20.1 Hz), −19.0 (d, 1P, $J_{P\text{-}P}$ = 16.4 Hz), −22.6 (d, 1P, $J_{P\text{-}P}$ = 21.0 Hz). ESI-MS: *m/z* 803.0 ($[M-H]^-$), 401.0 ($[M-2H]^{2-}$).

Reaction of the Biotin-ATP Probe with Lysine

The biotin-ATP probe solution was prepared by dissolving the lyophilized compound in water and was stored in a –20 ºC freezer. To investigate the labeling efficiency of the biotin-ATP probe, different concentrations of lysine, ranging from 0.1 μM to 1 mM, were incubated with the biotin-ATP probe at a concentration of 0.1 mM (reaction volume, 50 μL).

Labeling of RecA and YADH-I with the Biotin-ATP Probe

Different concentrations of the probe, which varied from 0.1 μM to 1 mM, were treated with 10 μg of the protein. For the labeling reaction with protein mixtures, the amount of BSA was varied from 10 μg to 1 mg while the amount of RecA was kept at 10 μg.

For the labeling reaction, the biotin-ATP probe solution was added to a solution of the protein in PBS buffer (pH 7.5) and mixed by quickly flicking the tube. The labeling reaction was continued with gentle shaking at room temperature for 1 h or at 4 ºC for 4 h. At the end of the labeling reaction, the excess probes were removed by using a Microcon YM-10 centrifugal filter (Millipore, Billerica, MA).

Cell Lysate Preparation and Labeling with the Biotin-ATP Probe

HeLa-S3 cells were purchased from National Cell Culture Center (Minneapolis, MN). The WM-266-4 cell line, derived from a metastatic site of a human malignant melanoma, was purchased from ATCC (Manassas, VA) and cultured in Eagle's Minimum Essential Medium (EMEM) supplied with 10% fetal bovine serum (Invitrogen, Carlsbad, CA), 100 IU/mL penicillin, 100 μg /mL streptomycin and 1.5 g/L sodium bicarbonate in 5% CO_2 at 37 ºC. Cells were harvested and washed three times with cold PBS. Approximately 6×10^6 cells were lysed in 500 μL of lysis buffer, which contained 0.7% CHAPS, 50 mM HEPES (pH 7.4), 0.5 mM EDTA, 100 mM NaCl and a protease inhibitor cocktail, on ice for 30 min. The cell lysates were centrifuged at 16,000g at 4 ºC

for 30 min, and the resulting supernatants were collected. To remove free endogenous nucleotides, particularly ATP, cell lysates were subjected to gel filtration separation by using NAP-25 columns (Amersham Biosciences). Cell lysates were eluted into 1.5 mL of elution buffer, which contained 50 mM HEPES (pH 7.4), 75 mM NaCl, and 5% glycerol. The gel-filtered cell lysates were further concentrated by using a Microcon YM-10 centrifugal filter. The proteins in cell lysates were quantified by using the Quick Start™ Bradford Protein Assay (BioRAD). Salts with divalent cations were added to the concentrated cell lysate until it contained 50 mM $MgCl_2$, 5 mM $MnCl_2$, and 5 mM $CaCl_2$. The biotin-ATP probe was then added until its final concentration reached 100 μM. Labeling reactions were carried out with gentle shaking at room temperature for 1 h. After reaction, cell lysates were cleaned and the buffer was exchanged to 100 mM NH_4HCO_3 solution (pH 8.5) by using Microcon centrifugal filter, followed by the addition of dithiothreitol (DTT) and iodoacetamide (IAA) to reduce and block cysteines.

Enzymatic Digestion and Affinity Purification

The labeled proteins were digested by using modified sequencing-grade trypsin at an enzyme/substrate ratio of 1:50 in 100 mM NH_4HCO_3 (pH 8.5) and at 37 °C for overnight, and the digestion was stopped by adding trifluoroacetic acid (TFA) to a final concentration of 5 mM. The peptide mixture was subsequently dried in a Speedvac.

Streptavidin-immobilized magnetic beads (MBs) were used to capture the biotin-labeled peptides from RecA and YADH-I digests. Prior to the binding, the MBs were

washed with buffer A, which contained 20 mM potassium phosphate and 0.15 M NaCl (pH 7.5), for at least 3 times and resuspended in 50 μL buffer A. The dried tryptic peptides were dissolved in 25 μL of buffer A followed by the addition of the MB suspension. Mixing was achieved by quickly flicking the tube. The mixture was then incubated at 25 °C for 30 min with gentle shaking.

After incubation, MBs were magnetically immobilized, the supernatant was removed, and the beads were washed with buffer A for three times and then with pure water for another three times. Following the washing procedures, MBs were resuspended in pure water (50 μL). The biotinylated peptides were eluted from the beads according to a modified protocol.[18] In this regard, the MB suspension was incubated at 70 °C for 5 min and the mixture was allowed to cool to room temperature. Once beads were magnetically captured, the supernatant, which contained eluted peptides, was transferred to a clean tube. This process was repeated once and the eluates were combined, dried in a Speedvac, and stored at –20 °C prior to LC-MS analysis.

Affinity columns packed with avidin-agarose resin (Sigma-Aldrich) were used to improve the capacity for capturing biotin-labeled peptides from HeLa-S3 and WM-266-4 cells. Briefly, the packed avidin-agarose resin was washed with 10 column volumes of buffer A before the tryptic digests were loaded. After a 30-min incubation, the unbound peptides were removed by washing the resin thoroughly with binding buffer and pure water. Following washing, labeled peptides were eluted with a mixture of CH_3CN and

H_2O (1:1, v/v) at 65 °C. The eluates were dried in a Speedvac and stored at –20 °C prior to mass spectrometric analysis.

LC-MS/MS

LC-MS/MS analysis of the affinity-purified peptides was performed in triplicate on an LTQ linear ion trap mass spectrometer (Thermo Electron, Waltham, MA) equipped with an electrospray ionization source. An Agilent 1100 capillary HPLC system (Agilent, Palo Alto, CA) with a Zorbax SB-C18 capillary column (5 μm in particle size, 0.5×250 mm, Agilent Technologies) was used for the separation, and the effluent from the column was coupled directly to the ionization source. Peptides were eluted from the column using a 63-min linear gradient of 2-65% CH_3CN in 0.6% aqueous solution of acetic acid, and the flow rate was 6 μL/min. The product ion spectra were acquired in data-dependent mode, where the eight most abundant ions observed in MS scan were chosen sequentially for collisional activation to obtain MS/MS.

Data Processing

The product ion spectra from RecA and YADH-I samples were searched against UniProt (Universal Protein Resource, http://www.ebi.ac.uk/trembl/index.html) using TurboSEQUEST in Bioworks 3.2 (Thermo Electron Corporation, San Jose, CA). MassAnalyzer 1.03, provided by Dr. Zhongqi Zhang at Amgen Inc., was used to further confirm the search results.[19, 20] The product-ion spectra from the analysis of samples of

HeLa-S3 and WM-266-4 cell lysates were searched against the NCBI human RefSeq protein database. The sequences and the labeling sites of all the identified biotin-labeled peptides were confirmed by analyzing manually the product-ion spectra of these peptides.

Results and Discussion

Reactivity of the Biotin-ATP Probe with Lysine

Many acyl phosphates are metabolic acylating reagents. A best-known example of these is acetyl coenzyme A, which is the acetyl group donor in many biological acetylation reactions.[21] In our work, the biotin-ATP probe was used as a chemical reagent for the acylation of ε-amino group of lysine residues in proteins (The structure of the biotin-ATP probe is shown in Scheme 2.1). The reactivity of this probe relies mainly on the mixed carboxylic phosphoric anhydride.[22] In this respect, acyl phosphates are known to react efficiently with amines to form amides, and the hydrolysis of acyl phosphates is very slow compared to the rate of amide bond formation when suitable concentrations of amines are used.[23]

To evaluate the reactivity of the biotin-ATP probe, different concentrations of lysine were used. ESI-MS and tandem MS (MS/MS, Figure 2.1) results revealed that, with high concentrations of lysine (1 mM) and long incubation time (4 h), an acylation product of lysine was formed. In this respect, the product ion spectrum of the $[M-H]^-$ ion of the acylation product showed the formation of a fragment ion of *m/z* 145, which is attributed to the deprotonated ion of lysine (Figure 2.1B). At low lysine concentrations

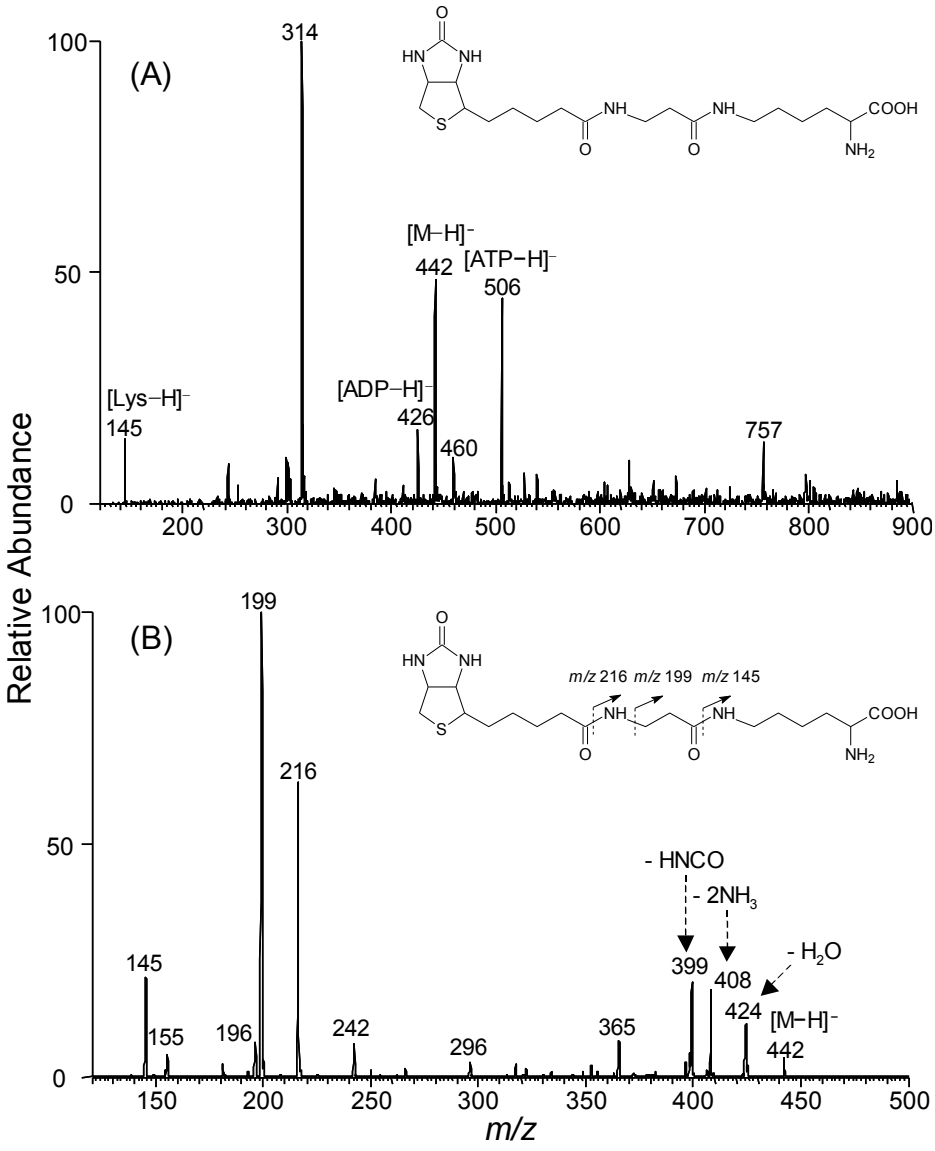


Figure 2.1 (A) Negative-ion ESI-MS of 4-h reaction mixture of biotin-ATP probe and lysine after dilution. The biotin-ATP conjugate of lysine, the product of this reaction, has an *m/z* of 442 in negative-ion ESI-MS. (B) Product-ion spectrum of the ion of *m/z* 442 observed in (A).

(< 10 μM), the formation of the acylation product was, however, not obvious under the otherwise identical reaction conditions.

Labeling of Adenosine Nucleotide-binding Proteins

To investigate the reaction of biotin-ATP probe with target proteins, we chose two adenosine nucleotide-binding proteins, RecA and YADH-I. The former protein is a multi-functional DNA-binding protein and a critical player in both homologous recombination and post-replicative DNA repair.[24] The X-ray crystal structures of *E. coli* RecA complexed with ADP or AMP-PNP, a nonhydrolyzable analogue of ATP, have been determined.[25-27] YADH-I is a member of zinc-containing alcohol dehydrogenases. It contains four identical subunits, and each subunit harbors an active site. The active site binds one NAD and a zinc atom, which is essential for the oxidation of alcohols to aldehydes.[28, 29] The X-ray crystal structure of YADH-I has been released very recently (PDB entry, 2HCY).

Similar to what we found for the reaction with lysine, the reaction of the acyl phosphate probe with proteins occurs at lysine residues as well. The difference lies in that the latter reaction is site-directed because the probe distinguishes many amino groups in the protein chain and reacts with a specific lysine residue in the two proteins.

We identify the probe labeling site(s) by digesting the RecA and YADH-I after the labeling reactions, treating the digestion mixtures with streptavidin-immobilized magnetic beads, eluting the bound peptides from the beads, and subjecting the peptide

mixtures to LC-MS/MS analysis. The LC-MS/MS results showed that only one probe-labeled peptide, IVEIYGPESSGK*TTLTLQVIAAAQR, was present in the tryptic digestion products of the reaction mixture of RecA (The asterisk symbol designates the labeled lysine residue, and MS/MS for the modified peptide is shown in Figure 2.2A. In this respect, the mass difference between the native and the probe-labeled peptide is 297.1 Da, which is consistent with the monoacylation of this peptide segment. In addition, the amide bond on the C-terminal side of the labeled lysine residue was no longer susceptible to trypsin cleavage. Therefore, this peptide with one internal lysine residue, which is K72, can be identified. This conclusion is further substantiated by the MS/MS results (Figure 2.2A).

In keeping with our observations here, studies on K72R mutant of RecA showed that K72 is critical for ATP hydrolysis.[30, 31] The X-ray crystal structure of the RecA-MgADP complex revealed that the phosphates of ADP form hydrogen bonds with the side chains of S70, K72, T73, and T74, and K72 is believed to stabilize the transition state of the ATP hydrolysis reaction mediated by RecA.[25-27] Moreover, the crystal structure of the RecA-MnAMP-PNP complex from the same report showed that K72 contacts directly with the γ phosphate of AMP-PNP, a nonhydrolyzable ATP analogue. Site-directed mutagenesis experiments and X-ray crystal structures of RecA also demonstrated that K72 is critical for ATP binding and the subsequent hydrolysis reaction. Different from these previous studies, our method facilitates the identification of the probe-labeled K72 residue, which might serve as direct evidence supporting that K72 is involved in ATP binding and hydrolysis.

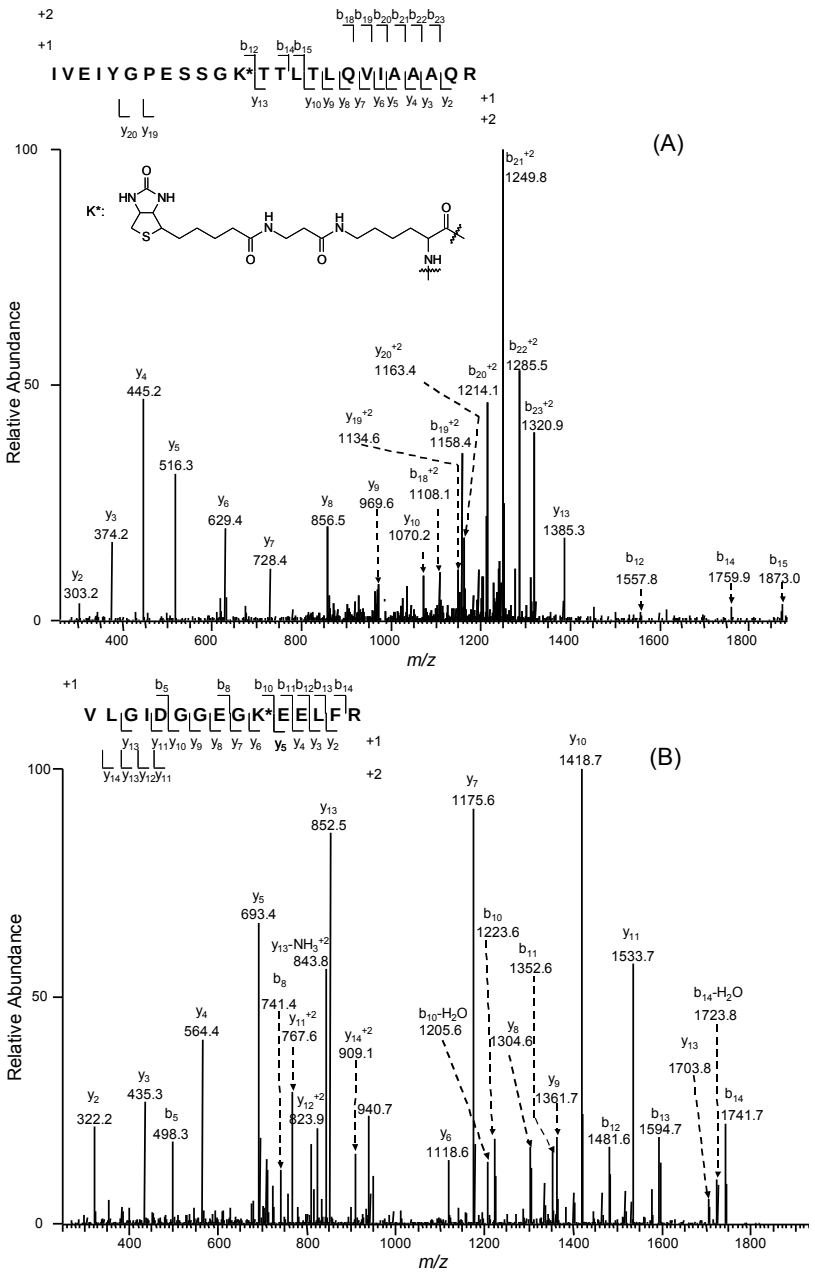


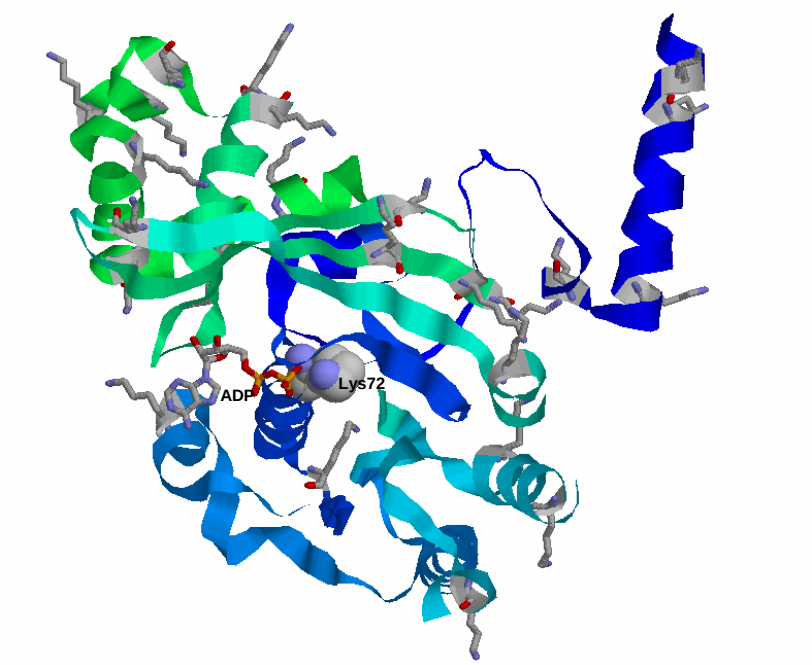
Figure 2.2 The product-ion spectra of the probe-labeled peptides. (A) The identified peptide with the sequence of IVEIYGPESSGK*TTLTLQVIAAAQR (*m/z* 981.5, triply charged) from RecA with site-directed labeling at K72; (B) the identified peptide with the sequence of VLGIDGGEGK*EELFR (*m/z* 958.4, doubly charged) from YADH-I with site-directed labeling at K206. The labeled lysine is designated with " * ", and its structure is shown in (A).

The structure of the RecA-ADP complex (Figure 2.3A) shows that, among all 26 lysine residues in this protein, only K72 is in proximity to the β- phosphate component of ADP (the distance between the ε-amino function of K72 and the β phosphate component of ADP is 3.6 Å). In this context, it is worth noting that the solvent accessible surface area of K72 is the smallest among all lysine residues (determined by using GETAREA,[32] http://www.scsb.utmb.edu/cgi-bin/get_a_form.tcl). Therefore, the specificity of the labeling reaction between the biotin-ATP probe and K72 is attributed to the binding of the acyl nucleotide probe to the ATP-binding site of the protein.

We next asked whether adenosine nucleotide-binding proteins other than ATP-binding proteins can also be labeled by the biotin-ATP probe. To this end, we employed YADH-I, an NAD-binding protein, to carry out the labeling reaction. The LC-MS/MS results revealed again the presence of a single, probe-labeled peptide with the sequence of VLGIDGGEGK*EELFR in the MB-enriched peptide mixture. In this regard, we found a mass difference of 425.2 Da, corresponding to the mass of the labeled lysine residue, between the y_5 and y_6 ions in the MS/MS of this peptide (Figure 2.2B), supporting that this lysine residue is modified. The labeled lysine was K206, which, among all the lysine residues, is the closest to NAD according to the X-ray crystal structure of YADH-I (Figure 2.3B; PDB entry, 2HCY).

It is worth noting that short reaction time and low concentration of biotin-ATP probe (10 μM probe and 1-h reaction time at room temperature) were employed for the above labeling experiments. When the labeling reaction was carried out with a higher concentration of the biotin-ATP probe and a longer incubation time, e.g., 1 mM probe and

(A)



(B)

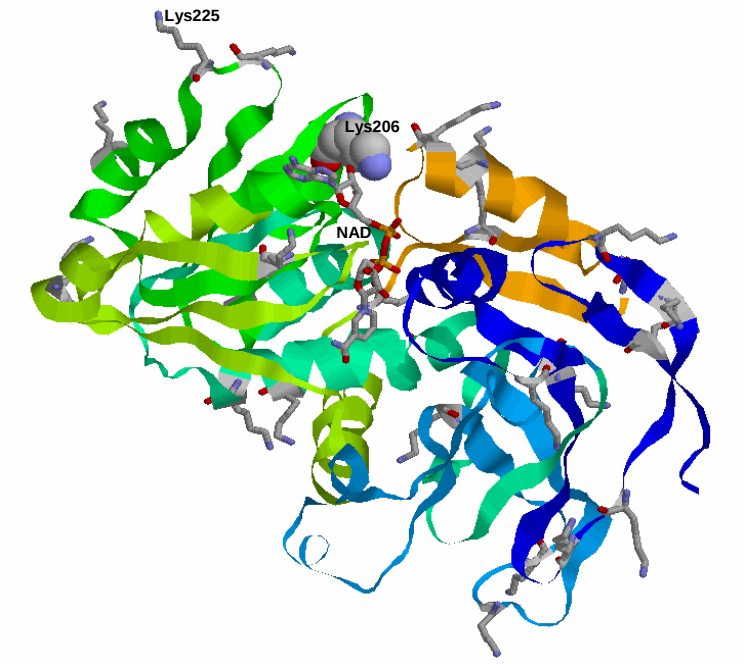


Figure 2.3 X-ray crystal structures of RecA and YADH-I. The bound ADP and NAD molecules and all lysine residues are shown in stick mode, whereas the lysine residues at the probe-labeling sites are shown in space-filling mode. (B) Structure of YADH-I complexed with NAD (PDB entry, 2HCY; only subunit A is shown). RasMol (Version 2.7.2.1.1, http://www.openrasmol.org/) was used to display the structures.

4-h incubation, however, the labeling may occur at K225 as well. We again calculated the total solvent-accessible areas of all lysine residues in YADH-I (subunit A) and it turned out that, among all the lysine residues, K225 assumes the largest surface area (209.8 $Å^2$). In contrast, K206 has only a small area exposed to solvent (15.3 $Å^2$). Nevertheless, at low probe concentrations, the lysine residue at the nucleotide-binding site of adenosine nucleotide-binding proteins can be labeled specifically. In this regard, NAD and ATP share the same adenosine diphosphate moiety, which accounts for the reactivity between the acyl nucleotide probe and the NAD-binding protein, YADH-I. This result showed that, to achieve labeling with high specificity, the concentration of the acyl nucleotide probe should be kept as low as feasible.

Affinity Purification and Enrichment of the Labeled Peptides

To demonstrate the potential application of the biotin-ATP probe for more complex samples, we examined the reaction of the probe with a protein mixture containing both RecA (10 μg) and BSA (from 10 μg to 1 mg). After the labeling reaction, we again digested the protein mixture with trypsin, subjected the peptide mixture to enrichment with magnetic beads as described above, and analyzed the resulting peptide mixture by LC-MS/MS. The LC-MS/MS results allowed us to identify, without ambiguity, the labeled K72-bearing peptide with the sequence of IVEIYGPESSGK*TTLTLQVIAAAQR in all samples containing up to 1 mg of BSA. We, however, did not identify any probe-labeled peptide from BSA, demonstrating the specificity of the reaction of the probe with nucleotide-binding proteins. This result also

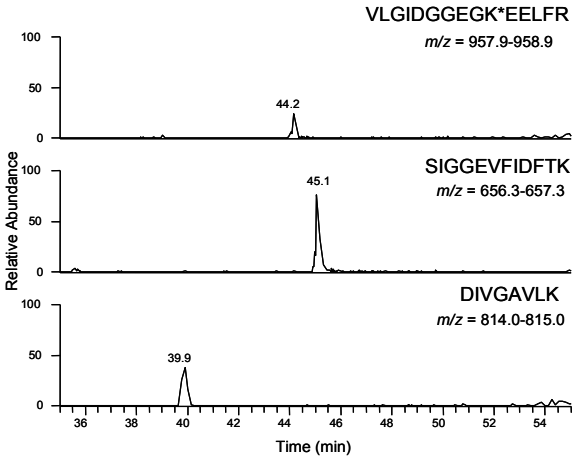
highlights that the affinity purification procedure allows for the enrichment of biotin-bearing peptide from a large quantity of peptide mixture.

To examine the effectiveness of the MBs in the enrichment of the probe-bearing peptides, we also analyzed directly, by LC-MS/MS, the tryptic digests of YADH-I after labeling reaction but without MB purification. We then evaluated the efficiency of the enrichment procedure by comparing the ion intensity for the probe-bearing peptide with respect to two other peptides which are not susceptible to reaction with the biotin-ATP probe, i.e., SIGGEVFIDFTK (*m/z* 656.8) and DIVGAVLK (*m/z* 814.5). Figure 2.4 depicts the selected-ion chromatograms (SICs) for these three peptides based on the LC-MS/MS analyses of the digestion mixture with (Figure 2.4B) or without (Figure 2.4A) MB enrichment. The data revealed that the enrichment procedure leads to significant decrease in the ion intensity for the unmodified peptides. In this respect, we estimated that the enrichment procedure allows for the selective enrichment for the probe-bearing peptide by at least 300-fold.

Identification of Biotin-ATP Probe Labeled Proteins in Whole Cell Lysates

From the above experiments with RecA and YADH-I, we showed that the biotin-ATP probe exhibits high specificity for adenosine nucleotide-binding proteins, and the affinity purification strategy can allow for the selective enrichment of the biotin-labeled peptides from a digestion mixture. To assess the performance of the biotin-ATP probe for real complex samples, we employed this strategy to the whole cell lysates of HeLa-S3

(A)



(B)

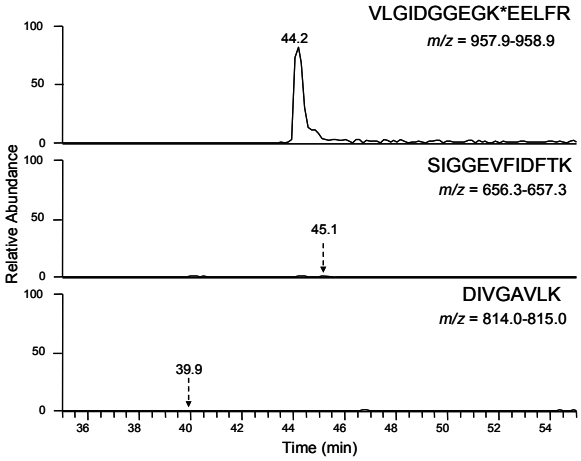
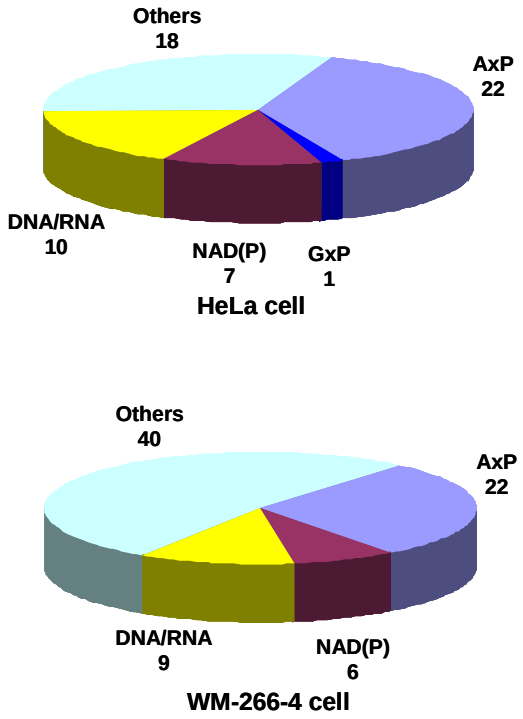


Figure 2.4 Selected-ion chromatograms (SICs) of ions, generated from LC-MS/MS analyses, corresponding to the biotinylated peptide, VLGIDGGEGK*EELFR (*m/z* 958.4), and two internal standard peptides, SIGGEVFIDFTK (*m/z* 656.8) and DIVGAVLK (*m/z* 814.5). The internal standard peptides were chosen from the unlabeled peptides identified in LC-MS/MS analysis of the tryptic digestion mixture of YADH-I without (A) and with (B) magnetic beads purification.

and WM-266-4 cells. The results are summarized in Figure 2.5. Overall, we identified probe-labeled peptides of 58 and 77 proteins from the whole-cell lysates of HeLa-S3 and WM-266-4 cells, respectively, and 15 common proteins were found in both cell lines. The identified peptides from these two cell lines, as well as the labeling site(s), are listed in the Supporting Information (Table S2.1 and S2.2).

Among these targets, 22 nucleotide-binding proteins were identified from the cell lysate of either HeLa-S3 or WM-266-4 cells (Figure 2.5A). Most of these identified nucleotide-binding proteins have clear adenosine nucleotide-binding ability. For example, adenylate kinase 2 is a ubiquitous enzyme that acts as a phosphotransferase catalyzing the production of ATP from ADP, a process involved in cellular energy homeostasis.[33] We identified one peptide, AVLLGPPGAGK*GTQAPR, with labeling by MS/MS (Figure 2.6A). The labeled lysine residue, K28, resides at the active site of this enzyme (Figure 2.7A). Phosphoglycerate kinase 1 (PGK1) is a transferase enzyme involved in glycolysis. It employs the energy released in ATP hydrolysis to catalyze the conversion of 1,3-bisphosphoglycerate to 3-phosphoglycerate. Recent work has revealed that PGK1 is overexpressed in pancreatic ductal adenocarcinoma.[34] The X-ray crystal structure of the yeast enzyme reveals that ATP can bind to the enzyme at the active site, where lys residues are present (Figure 2.7B). Our MS/MS result also confirmed that K216, within peptide ALESPERPFLAILGGAK*VADK, at the active site reacted with the biotin-ATP probe (Figure 2.6B).

(A)



(B)

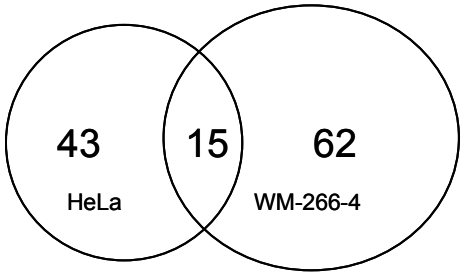


Figure 2.5 Biotin-ATP probe-labeled proteins in HeLa-S3 and WM-266-4 cells. (A) The identified proteins were classified according their nucleotide-binding ability. AxP: adenosine nucleotide; GxP: guanosine nucleotides; NAD(P): nicotinamide adenine dinucleotide (phosphate); RNA/DNA and others. (B) The Venn diagram shows the numbers of proteins and the overlap (intersection) of proteins identified from both cell lines.

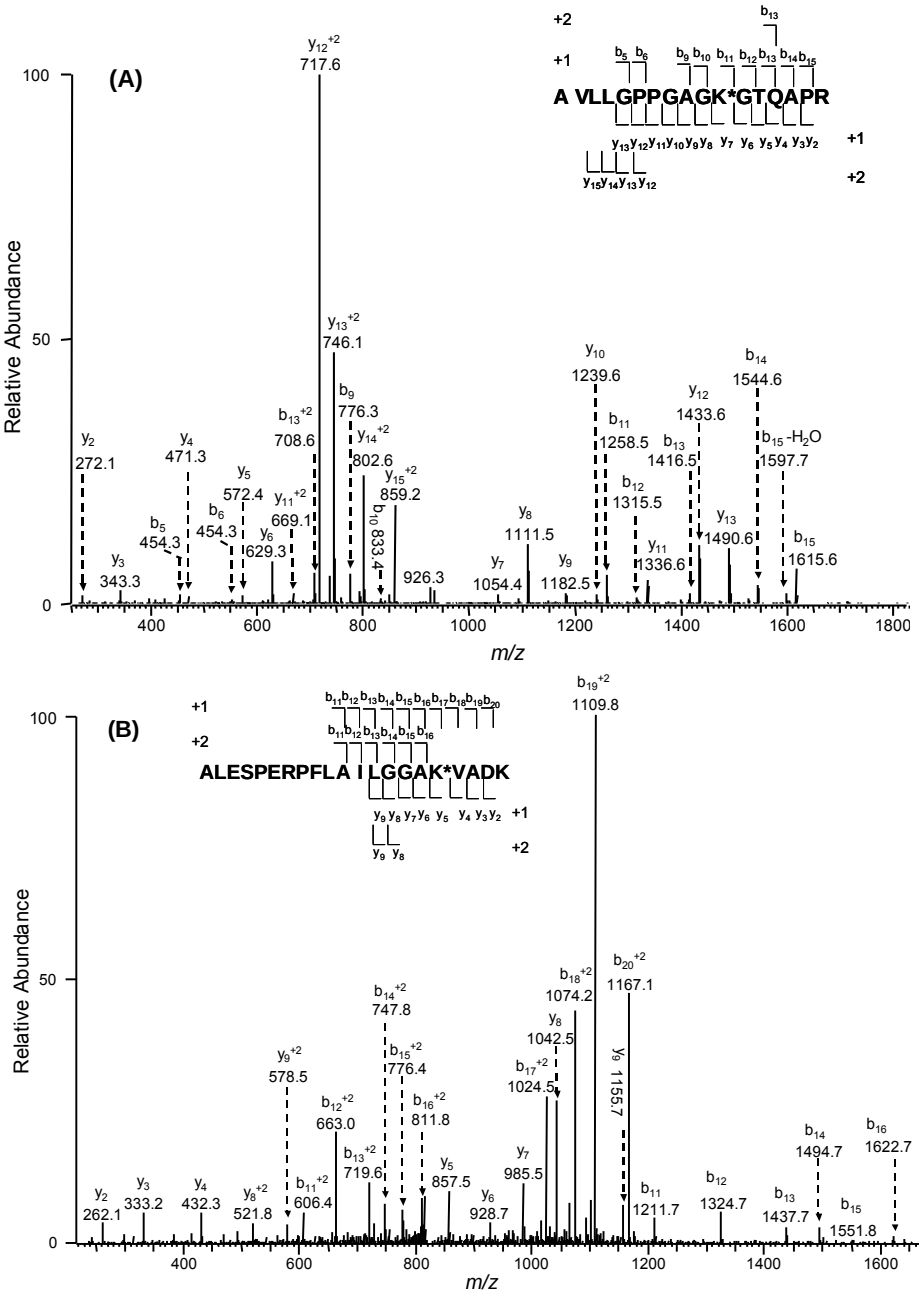


Figure 2.6 Product ion spectra of the probe-labeled peptides from adenylate kinase 2, ALESPERPFLAILGGAK*VADK (*m/z* 944.1, doubly charged) with labeling site of K28 (A) and phosphoglycerate kinase 1, ALESPERPFLAILGGAK*VADK (*m/z* 827.2, triply charged) with labeling site of K216 (B).

Actins and heat shock proteins are two families of abundant ATP-binding proteins. Some of these proteins have been identified with more than one labeling sites. The multiple labeling of these proteins may be attributed to the high abundances of these proteins, which result in the nonspecific labeling of the proteins. Alternatively, the labeling at multiple sites might also originate from the interactions of actins and heat shock proteins with other nucleotide-binding proteins. The latter might be particularly true for heat shock proteins, which are molecular chaperones and have many interaction partners.

Our results also revealed that, other than ATP-binding proteins, some nucleotide analogue-binding proteins can also couple with the probe. As shown above, YADH-I could react with the biotin-ATP probe specifically via lysine residue at the NAD-binding sites. With respect to cell lysates, seven and five NAD(P)-binding proteins were identified in HeLa-S3 and WM-266-4 cells, respectively (Figure 2.5A). Most of these proteins are dehydrogenases, which oxidize the substrate by transferring one or more protons and a pair of electrons to an acceptor, usually NAD/NADP or a flavin coenzyme such as flavin adenine dinucleotide (FAD). For example, lactate dehydrogenase A was identified in both cell lysates. The X-ray crystal structure revealed that it was labeled at one of the two conserved active site lys residues (PDB entry, 1I10). Similarly, lactate dehydrogenase B (PDB entry, 1T2F), glyceraldehyde-3-phosphate dehydrogenase (PDB entry, 1ZNQ) and catalase (PDB entry, 1DGB) were identified with specific labeling of the lys residues at the nucleotide analogue-binding sites, which is in keeping with the available X-ray crystal structures.

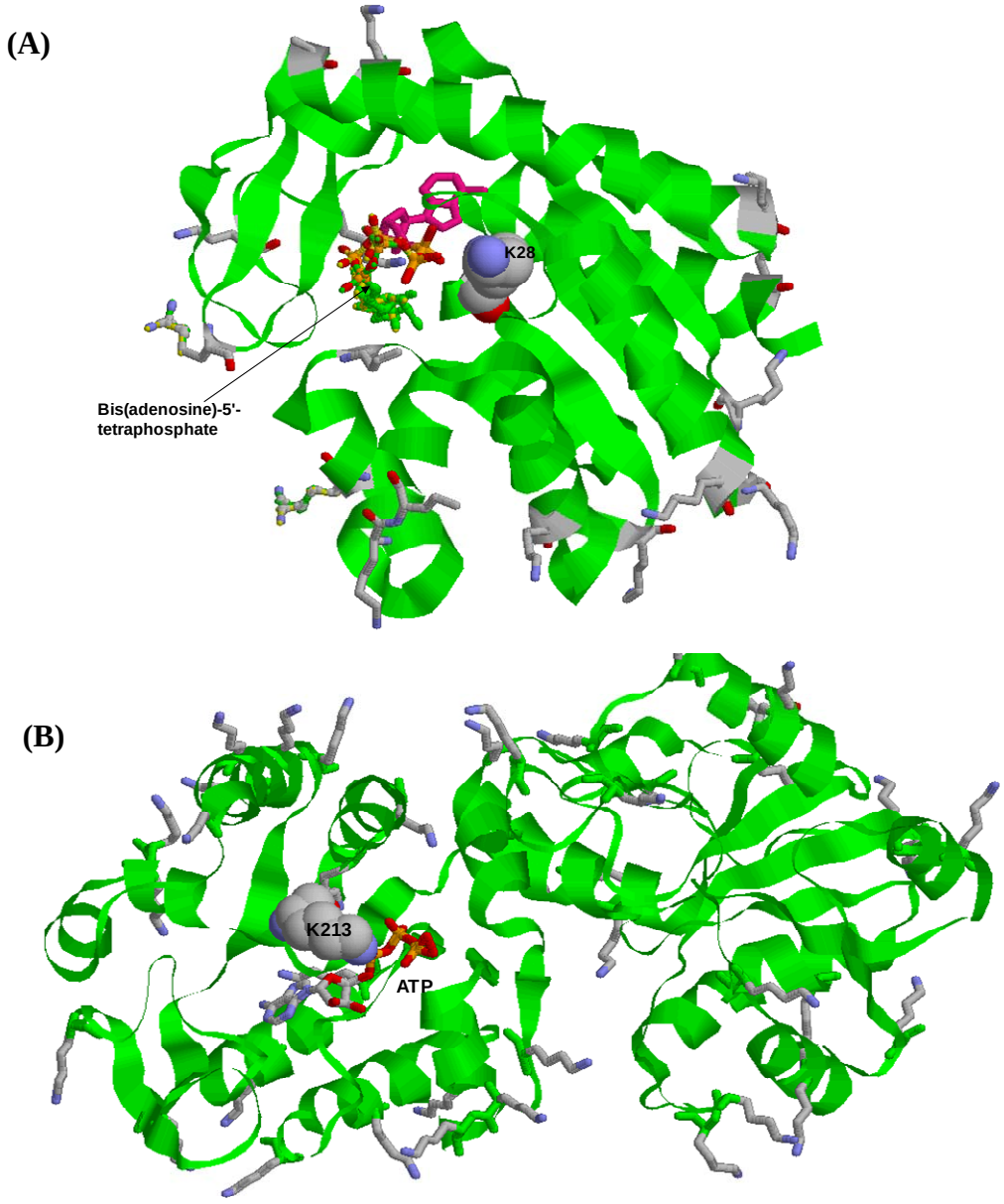


Figure 2.7 The structures of adenylate kinase 2 complexed with bis(adenosine)-5'-tetraphosphate (PDB entry, 2C9Y) and yeast phosphoglycerate kinase 1 complexed with ATP (PDB entry, 3PGK) are shown in (A) and (B), respectively. In (B), K213 at the ATP binding site of yeast PGK1 corresponds to K216 in the human counterpart.

In addition to the identified nucleotide-binding proteins, we also detected probe-labeled peptides from several RNA/DNA-binding proteins and proteins whose nucleotide-binding ability remains unknown. To the best of our knowledge, some of these proteins are actin binders, phosphate binders, kinase substrates, ribosomal proteins, and proteins without binding ability being described. Many of these proteins are abundant, and the surface lys residues in these proteins might be involved in the labeling reaction with the biotin-ATP probe. These results, therefore, indicate that the biotin-ATP probe may not be highly specific, especially when the sample is as complex as a whole proteome. Similar results were described in recent reports.[12, 14] In this context, it is important to note that some kinase substrates could also be labeled in a "crossover" manner.[12] Thus, some proteins that are capable of interacting with nucleotide-binding proteins can also be modified by the probe without binding directly to the probe. In most cases, however, we could not differentiate these proteins involved in the "crossover" labeling from those being labeled nonspecifically because of the lack of protein-protein interaction information.

Conclusions

In this work, we demonstrated the site-directed labeling of nucleotide-binding proteins using a biotin-ATP probe. Results from the study of two adenosine nucleotide-binding proteins, RecA and YADH-I, demonstrated that the probe can allow for the identification of the lysine residue residing at the nucleotide-binding sites of these proteins. In this respect, LC-MS/MS analysis enabled us to identify the nucleotide-

binding sites in proteins in a short period of time. K72, playing an important role in ATP hydrolysis for RecA , was labeled with the probe, providing direct evidence supporting that this residue is involved in ATP hydrolysis. K206 in YADH-I was also identified unambiguously by the labeling method. This strategy is relatively simple, fast, and straightforward, which involves labeling reaction, enzymatic digestion, affinity purification, and LC-MS/MS analysis. The method should be generally applicable for the identification of lysine residues at the binding sites of proteins that can bind to adenosine nucleotide and its analogues. The application of the biotin-ATP probe to cell lysates has shown a great potential in the activity-based functional study of other nucleotide-binding proteins, as well as in the profiling of nucleotide-binding proteins in complex samples. Quantitative analysis of the expression of nucleotide-binding proteins in cells with drug treatment or under different stress conditions may also be achieved by employing the acyl nucleotide probe and incorporating stable isotope labeling techniques.

References

1. de Hoffmann, E.; Stroobant, V., *Mass Spectrometry: Principles and Applications*. 2nd ed.; John Wiley & Sons: 2001.
2. Wolters, D. A.; Washburn, M. P.; Yates, J. R., An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001,** 73, 5683-5690.
3. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999,** 17, 676-682.
4. Leitner, A.; Lindner, W., Current chemical tagging strategies for proteome analysis by mass spectrometry. *J. Chromatogr. B, Anal. Technol. Biomed. Life Sci.* **2004,** 813, 1-26.
5. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999,** 17, 994-999.
6. Adam, G. C.; Sorensen, E. J.; Cravatt, B. F., Trifunctional chemical probes for the consolidated detection and identification of enzyme activities from complex proteomes. *Mol. Cell. Proteomics* **2002,** 1, 828-835.
7. Speers, A. E.; Cravatt, B. F., Chemical strategies for activity-based proteomics. *ChemBioChem* **2004,** 5, 41-47.

8. Adam, G. C.; Burbaum, J.; Kozarich, J. W.; Patricelli, M. P.; Cravatt, B. F., Mapping enzyme active sites in complex proteomes. *J. Am. Chem. Soc.* **2004,** 126, 1363-1368.

9. Yount, R. G., ATP Analogs. *Adv. Enzymol.* **1975,** 43, 1-56.

10. Cremo, C. R., Fluorescent nucleotides: Synthesis and characterization. *Methods Enzymol.* **2003,** 360, 128-177.

11. Bagshaw, C. R., ATP analogues at a glance. *J. Cell. Sci.* **2001,** 114, 459-460.

12. Hanoulle, X.; Van Damme, J.; Staes, A.; Martens, L.; Goethals, M.; Vandekerckhove, J.; Gevaert, K., A new functional, chemical proteomics technology to identify purine nucleotide binding sites in complex proteomes. *J. Proteome Res.* **2006,** 5, 3438-3445.

13. Campbell, D. A.; Liyanage, M.; Szardenings, A. K.; Wu, M. Acyl-nucleotide probes and methods of their synthesis and their use in protein labeling. WO 2004090154, **2004**.

14. Patricelli, M. P.; Szardenings, A. K.; Liyanage, M.; Nomanbhoy, T. K.; Wu, M.; Weissig, H.; Aban, A.; Chun, D.; Tanner, S.; Kozarich, J. W., Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **2007,** 46, 350-358.

15. Redeuilh, G.; Secco, C.; Baulieu, E. E., The use of the biotinyl estradiol-avidin system for the purification of nontransformed estrogen-receptor by biohormonal affinity-chromatography. *J. Biol. Chem.* **1985,** 260, 3996-4002.

16. Skander, M.; Humbert, N.; Collot, J.; Gradinaru, J.; Klein, G.; Loosli, A.; Sauser, J.; Zocchi, A.; Gilardoni, F.; Ward, T. R., Artificial metalloenzymes: (Strept)avidin as host for enantioselective hydrogenation by achiral biotinylated rhodium-diphosphine complexes. *J. Am. Chem. Soc.* **2004,** 126, 14411-14418.

17. Kreimeyer, A.; Ughetto-Monfrin, J.; Namane, A.; Huynh-Dinh, T., Synthesis of acylphosphates of purine ribonucleosides. *Tetrahedron Lett.* **1996,** 37, 8739-8742.

18. Holmberg, A.; Blomstergren, A.; Nord, O.; Lukacs, M.; Lundeberg, J.; Uhlen, M., The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* **2005,** 26, 501-510.

19. Zhang, Z. Q., De novo peptide sequencing based on a divide-and-conquer algorithm and peptide tandem spectrum simulation. *Anal. Chem.* **2004,** 76, 6374-6383.

20. Zhang, Z. Q.; McElvain, J. S., De novo peptide sequencing by two dimensional fragment correlation mass spectrometry. *Anal. Chem.* **2000,** 72, 2337-2350.

21. Lipmann, F., On chemistry and function of coenzyme-A. *Bacteriol. Rev.* **1953,** 17, 1-16.

22. Kluger, R., Acyl phosphate esters: Charge-directed acylation and artificial blood. *Synlett* **2000**, (12), 1708-1720.

23. Di Sabato, G.; Jencks, W. P., Mechanism and catalysis of reactions of acyl phosphates. 2. Hydrolysis. *J. Am. Chem. Soc.* **1961,** 83, 4400-4405.

24. Roca, A. I.; Cox, M. M., RecA protein: Structure, function, and role in recombinational DNA repair. *Prog. Nucleic Acids Res. Mol. Biol.* **1997,** 56, 129-223.

25. Story, R. M.; Steitz, T. A., Structure of the RecA protein-ADP Complex. *Nature* **1992,** 355, 374-376.

26. Xing, X.; Bell, C. E., Crystal structures of Escherichia coli RecA in a compressed helical filament. *J. Mol. Biol.* **2004,** 342, 1471-1485.

27. Xing, X.; Bell, C. E., Crystal structures of Escherichia coli RecA in complex with MgADP and MnAMP-PNP. *Biochemistry* **2004,** 43, 16142-16152.

28. Leskovac, V.; Trivic, S.; Latkovska, M., State and accessibility of zinc in yeast alcohol-dehydrogenase. *Biochem. J.* **1976,** 155, 155-161.

29. Trivic, S.; Leskovac, V., Structure and function of yeast alcohol dehydrogenase. *J. Serb. Chem. Soc.* **2000,** 65, 207-227.

30. Rehrauer, W. M.; Kowalczykowski, S. C., Alteration of the nucleoside triphosphate (NTP) catalytic domain within Escherichia-coli RecA protein attenuates NTP hydrolysis but not joint molecule formation. *J. Biol. Chem.* **1993,** 268, 1292-1297.

31. Cox, M. M., Why does RecA protein hydrolyze ATP. *Trends Biochem. Sci.* **1994,** 19, 217-222.

32. Fraczkiewicz, R.; Braun, W., Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **1998,** 19, 319-333.

33. Nobumoto, M.; Yamada, M.; Song, S. C.; Inouye, S.; Nakazawa, A., Mechanism of mitochondrial import of adenylate kinase isozymes. *J. Biochem.* **1998,** 123, 128-135.

34. Hwang, T. L.; Liang, Y.; Chien, K. Y.; Yu, J. S., Overexpression and elevated serum levels of phosphoglycerate kinase 1 in pancreatic ductal adenocarcinoma. *Proteomics* **2006,** 6, 2259-2272.

Supporting Information for Chapter 2

Table S2.1 Nucleotide-binding proteins identified with labeling from HeLa-S3 cells

Number	Protein Name	NCBI NP number	Labeling Peptide	Labling Residue	PDB ID	Binding
AxP-binding proteins						
1	adenylate kinase 2 isoform b	NP_037543	AVLLGPPGAGK*GTQAPR	K28	2C9Y	ATP
2	acyl-CoA synthetase long-chain family member 4 isoform 2	NP_075266	GDCTVLK*PTLMAAVPEIMDR	K367		ATP
3	aldolase A	NP_908930	GGVVGIK*VDK	K108	4ALD	ATP
4	alpha 2 actin	NP_001604	VAPEEHPTLLTEAPLNPK*ANR	K115	1RDW	ATP
			EITALAPSTMK*IK	K328		
			IK*IIAPPER	K330		
5	AMP-activated protein kinase beta 2 non-catalytic subunit	NP_005390	SEGAGGHAPGK*EHK	K31		ATP
6	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit	NP_001001937	VGLK*APGIIPR	K175		ATP
7	ATP synthase, H+ transporting, mitochondrial F1 complex, beta subunit	NP_001677	GQK*VLDSGAPIKIPVGPETLGR	K24		ATP
			GPIK*TK	K59		
8	ATP-binding cassette, sub-family A, member 6	NP_525023	LGLLGPNGAGK*SSSIR	K1326		ATP
9	ATP-binding cassette, sub-family A, member 9	NP_525022	QELGDLEEDFDPSVK*	K1615		ATP
10	chaperonin	NP_955472	TVIIEQSWGSPK*VTK	K72		ATP

			VTK*DGVTVAK	K75		
			GIIDPTK*VVR	K523		
11	fms-related tyrosine kinase 1	NP_002010	YVNAFK*FMSLER	K1218		ATP
12	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	NP_005338	K*TKPYIQVDIGGGQTK	K125		ATP
			QATK*DAGTIAGLNVMR	K185		
			VMEHFIK*LYK	K268		
			AK*FEELNMDLFR	K326		
			STMK*PVQK	K340		
			NK*ITITNDQNR	K522		
13	heat shock 70kDa protein 9B precursor	NP_004125	QATK*DAGQISGLNVLR	K206		ATP
14	inositol 1,4,5-trisphosphate 3-kinase C	NP_079470	DFEGPSIMDCK*MGSR	K486		ATP
15	phosphoglycerate kinase 1	NP_000282	ALESPERPFLAILGGAK*VADK	K216	3PGK	ATP
16	PI-3-kinase-related kinase SMG-1	NP_055907	QLK*EIER	K2763		ATP
17	polyribonucleotide nucleotidyltransferase 1	NP_149100	K*VLQSPATTVVR	K750		ATP
18	prolyl 4-hydroxylase, beta subunit	NP_000909	LK*AEGSEIR	K71		ADP
19	protein disulfide isomerase-associated 3 precursor	NP_005304	K*TFSHELSDFGLESTAGEIPVVAIR	K305		ATP
20	protein disulfide isomerase-associated 4	NP_004902	K*LAPEYEK	K211		ATP
21	protein kinase, DNA-activated, catalytic polypeptide	NP_008835	VCVNLMK*ALK	K1404		ATP/DNA
22	Ras-GTPase-activating protein SH3-domain-binding protein	NP_005745	LNVEEK*K	K413		ATP/DNA /RNA
GxP-binding proteins						
1	GDP dissociation inhibitor 2	NP_001485	EIIVQNGK*VIGVK	K264		GDP

NAD/NADP-binding proteins

1	alpha glucosidase II alpha subunit isoform 3	NP_938149	VVIIGAGK*PAAVVLQTK	K921		NAD
2	catalase	NP_001743	FHYK*TDQGIK	K237	1DGB	NADP
3	cytochrome b5 reductase isoform 2	NP_015565	FAIRPDK*K	K140	1UMK	NAD
4	glyceraldehyde-3-phosphate dehydrogenase	NP_002037	TVDGPSGK*LWR	K194	1ZNQ	NAD/NADP
			AITATQK*TVDGPSGK	K186		
5	lactate dehydrogenase A	NP_005557	DLADELALVDVIEDK*LK	K57	1I10	NAD
6	lactate dehydrogenase B	NP_002291	SLADELALVDVLEDK*LK	K58	1T2F	NAD
7	mitochondrial malate dehydrogenase precursor	NP_005909	IQEAGTEVVK*AK	K239	2DFD	NAD

Table S2.2 Nulceotide-binding proteins identified with labeling from WM-266-4 cells

Number	Protein Name	NCBI NP number	Labeling Peptide	Labling Residue	PDB ID	Binding
AxP-binding proteins						
1	adenylate kinase 2 isoform b	NP_037543	AVLLGPPGAGK*GTQAPR	K28	2C9Y	ATP
2	actin, gamma 1 propeptide	NP_001605	VAPEEHPVLLTEAPLNPK*ANR	K113		ATP
3	actinin, alpha 4	NP_004915	GVK*LVSIGAEEIVDGNAK	K125		ATP
4	aldolase A	NP_908930	GILAADESTGSIAK*R	K42	4ALD	ATP
			GGVVGIK*VDK	K108		
			ENLK*AAQEEYVK	K322		
5	alpha 2 actin	NP_001604	HQGVMVGMGQK*DSYVGDEAQSK	K52		ATP
			VAPEEHPTLLTEAPLNPK*ANR	K115		
			IK*IIAPPER	K330		
6	ATP-binding cassette, sub-family A member 4	NP_000341	ILK*NANSTFEELEHVR	K412		ATP
7	ATP-binding cassette, sub-family A, member 7 isoform b	NP_150651	ILK*QVFLIFPHFCLGR	K1547		ATP
8	chaperonin	NP_955472	TVIIEQSWGSPK*VTK	K72		ATP
			VTK*DGVTVAK	K75		
9	fructose-bisphosphate aldolase C	NP_005156	VDK*GVVPLAGTDGETTTQGLDGLSER	K111	1XFB	ATP

10	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	NP_005338	K*SQIFSTASDNQPTVTIK	K447		ATP
11	mitogen-activated protein kinase 4	NP_002738	K*LLPEVNSEAIDFLEK	K276		ATP
12	mitogen-activated protein kinase kinase kinase 3 isoform 2	NP_002392	K*YTRQILEGMSYLHSNMIVHR	K468		ATP
13	p21-activated kinase 1	NP_002567	IAKPLSSLTPLIAAAK*EATK	K525		ATP
14	p21-activated kinase 2	NP_002568	DPLSANHSLK*PLPSVPEEK	K38		ATP
15	PCTAIRE protein kinase 3 isoform a	NP_997668	LGEGTYATVFK*GR	K190		ATP
16	phosphoglycerate kinase 1	NP_000282	ALESPERPFLAILGGAK*VADK	K216	3PGK	ATP
17	prolyl 4-hydroxylase, beta subunit	NP_000909	LK*AEGSEIR	K71		ATP
			VLVGK*NFEDVAFDEK	K375		
18	protein disulfide isomerase-associated 3 precursor	NP_005304	EATNPPVIQEEK*PK	K494		ATP
19	protein disulfide isomerase-associated 4	NP_004902	K*LAPEYEK	K211		ATP
20	protein kinase, cGMP-dependent, type I	NP_006249	K*TWTFCGTPEYVAPEIILNK	K529		ATP/cGMP
21	serine/threonine kinase 25	NP_006365	K*TSFLTELIDR	K276		ATP
22	valosin-containing protein	NP_009057	IVSQLLTLMDGLK*QR	K336		ATP

NAD(P)-binding proteins

1	glyceraldehyde-3-phosphate dehydrogenase	NP_002037	TVDGPSGK*LWR	K194	1ZNQ	NAD/NADP
			AITATQK*TVDGPSGK	K186		
2	glutamate dehydrogenase 1	NP_005262	FGK*HGGTIPIVPTAEFQDR	K480	1HWZ	ADP/GTP/NAD/NADP
3	lactate dehydrogenase A	NP_005557	DLADELALVDVIEDK*LK	K57	1I10	NAD
4	lactate dehydrogenase B	NP_002291	SLADELALVDVLEDK*LK	K58	1T2F	NAD

5	mitochondrial malate dehydrogenase precursor	NP_005909	IQEAGTEVVK*AK	K239	2DFD	NAD
6	thioredoxin reductase 1	NP_003321	K*IGLETVGVK	K299	2CFY	NADP/FAD

CHAPTER 3

Exploring DNA-binding Proteins with *In Vivo* Chemical Cross-linking and Mass Spectrometry

Introduction

Although a number of proteins with specific or general affinity to nucleic acids have been identified, numerous proteins involved in gene regulation, DNA repair and oncogenesis are likely still unknown. Therefore, it has always been of interest to study proteins that interact with nucleic acids, with the motivation to understand fundamental biological processes such as chromatin organization, transcription, DNA replication, recombination and repair, which are often regulated by proteins that bind to nucleic acids.[1-3] Given the importance of nucleic acid-binding proteins and their interactions with DNA, RNA, or each other, it is necessary to develop a general analytical technique to identify comprehensively these proteins.

A classical technique used to detect nucleic acid-protein complexes is the electrophoretic mobility shift assay (EMSA), which is based on the principle that the electrophoretic mobility of a protein-nucleic acid complex is typically less than that of the free nucleic acid.[4] It is a core technology underlying a wide range of qualitative and quantitative assays for the characterization of protein-nucleic acid interactions.[4-6] Although EMSA is commonly used to detect nucleic acid-interacting factors, this assay is usually limited to evaluate *in vitro* interactions, by incubating a purified protein or a

protein mixture with a radioactively labeled DNA probe. Another widely used technique for characterizing DNA-binding proteins and their associated factors is chromatin immunoprecipitation (ChIP), which is employed to study the binding and interaction of modified histones or transcription factors with specific DNA sequences.[7, 8] ChIP assay allows the detection of *in vivo* interactions of specific proteins with particular genomic regions in living cells with formaldehyde cross-linking.[9, 10] However, both ChIP and EMSA are mainly restricted to known biological targets and have low throughput, making these two methods not suitable for identifying unknown nucleic acid-interacting factors or for studying the dynamics of gene regulation, a complex process requiring the interaction of numerous factors.

Recent advances in mass spectrometry (MS) have greatly facilitated protein identification and quantification.[11, 12] In the past several years, hundreds of previously unknown proteins have been identified as nuclear proteins that are potentially involved in the regulation of gene expression, DNA replication and repair.[13] The new field of nuclear proteomics has made some promising advances in elucidating the composition and dynamics of protein expression in nucleus and its subcompartments.[14-19] For example, nucleolar proteomic studies facilitated the identification of up to approximately 700 proteins from isolated nucleoli in HeLa cells.[20-24] However, most of these previous investigations were based on either crude nuclear pellets or purified subcellular compartments. Recently, the nuclear proteome from human Raji lymphoma cells was investigated by using 2D gel and MS.[25] In addition, DNA-binding proteins were isolated from the nuclear extracts by using agarose immobilized with calf thymus DNA, where

the *in vitro* interaction between the immobilized DNA and proteins constitutes the principle for the isolation. So far, very few studies have focused on the function-based comprehensive investigation of nucleic acid-binding proteins,[26, 27] probably owing to the difficulty in capturing the *in vivo* DNA-protein interactions at a large scale.

Formaldehyde is a highly reactive reagent, which can freeze the DNA-protein interactions occurring in living cells under physiological conditions *in situ*, thereby preventing subsequent dissociation and redistribution of proteins while working on sample preparation.[10, 27, 28] Formaldehyde cross-linking has been extensively used to study DNA-protein and protein-protein interactions.[28-31] Amino and imino groups of amino acids (lysine, arginine and histidine) and DNA (primarily adenine and cytosine) react readily with formaldehyde, leading to the formation of cross-links.[10, 32, 33] An attractive feature of formaldehyde cross-linking is that the cross-linking is fully reversible at high temperature (> 67 ºC) in aqueous solution.

In this work, we describe an approach for the comprehensive investigation of DNA-binding proteins with *in vivo* formaldehyde cross-linking. After the cross-linking reaction, cell nuclei were isolated, and the covalently bound DNA-protein complexes were subsequently purified. After purification, the DNA-protein cross-linking was reversed to release the DNA-binding proteins and the liberated DNA was removed by DNase digestion and centrifugal filtration. The purified DNA-binding proteins were resolved by SDS-PAGE, digested in-gel with trypsin, and the digestion mixtures were interrogated by LC-MS/MS. By using this method, we were able to identify more than one hundred

DNA-binding proteins according to the Gene Ontology (GO) annotations. In principle, this approach is not limited to the identification of DNA-binding proteins; it is also applicable to the investigation of proteins that bind to RNA.

Materials and Methods

Cell Culture

HL-60 Human acute promyelocytic leukemia cells (ATCC, Manassas, VA) were cultured in Iscove's modified minimal essential medium (IMEM) supplemented with 20% fetal bovine serum (FBS, Invitrogen, Carlsbad, CA), 100 IU/mL of penicillin and 100 µg/mL of streptomycin in 75 cm^2 culture flasks. Cells were maintained in a humidified atmosphere with 5% CO_2 at 37 ºC, with medium renewal of 2-3 times a week depending on cell density.

In Vivo Formaldehyde Cross-linking

HL-60 cells were collected by centrifugation at 300 g for 5 min at 4 ºC, and washed with ice-cold PBS to remove culture medium and FBS. *In vivo* cross-linking was achieved by adding 11% (w/v) formaldehyde to 5 mL of cell suspension in PBS to obtain a final concentration of 1% (w/v). After incubating at room temperature for 10 min, formaldehyde was quenched by the addition of 2.5 M glycine to a final concentration of 125 mM and incubated at room temperature for 5 min. The cross-linked cells were

collected by centrifugation (300 g at 4 ºC for 5 min) and the cell pellet was washed twice with cold PBS.

Isolation of Nuclei

Nuclei isolation was carried out using a protocol adapted from that reported by Henrich *et al*.[25] The cross-linked HL-60 cell pellet (~ 4×10^7 cells) was resuspended in 10 volumes of ice-cold hypotonic lysis buffer A containing 10 mM HEPES (pH 7.4), 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, 1 mM NaF, 1 mM Na_3VO_4, 1 mM PMSF, and a protease inhibitor cocktail. After incubation on ice for 30 min, NP-40 was added to the lysis buffer until its final concentration reached 0.5% (v/v), and the mixture was incubated on ice for 5-min. Cells were then gently lysed with a Dounce homogenizer with B type pestle (clearance ~ 0.7 mm) for 10 strokes on ice. The nuclear fraction was collected by centrifugation at 800 g at 4 ºC for 5 min, and the resulting crude nuclear pellet was resuspended in buffer B, which contained 250 mM sucrose, 10 mM $MgCl_2$, 20 mM Tris-HCl (pH 7.4) and 1mM DTT. The nuclei suspension was layered over a two-step sucrose gradient cushion [1.3 M sucrose, 6.25 mM $MgCl_2$, 20 mM Tris-HCl (pH 7.4), 0.5 mM DTT above 2.3 M sucrose in 2.5 mM $MgCl_2$ and 20 mM Tris-HCl (pH 7.4)], and centrifuged subsequently at 5000 g at 4 °C for 45 min. The isolated nuclei were washed with buffer A and collected by centrifugation at 1000 g.

Isolation of DNA-protein Complexes

DNA-binding proteins were isolated and copurified with genomic DNA as cross-

linked DNA-protein complexes. The purification of DNA-protein cross-links was carried out by using a method described by Baker *et al.* [34, 35] with modifications. The isolated nuclei were lysed in 500 μL DNAzol (Invitrogen) by repeated pipetting with a wide-bore pipette tip. DNA was precipitated by using a half volume of ice-cold 100% ethanol, and incubated at –20 °C for 1 h. The precipitates were pelletted by centrifugation at 5,000 g at 4 °C for 5 min. The pellet was washed with ice-cold 75% ethanol and resuspended in 50 mM Tris-HCl buffer (pH 7.4). Urea and SDS were added to the suspension until their final concentrations reached 8 M and 2% (w/v), respectively, to denature proteins and to dissociate the non-cross-linked proteins from DNA-protein complexes. The sample was incubated at 37 °C for 30 min with gentle shaking. To the sample solution, an equal volume of 5 M NaCl was added and the resulting mixture was incubated at 37 °C for 30 min. The DNA was precipitated again by the addition of 0.1 volume of 3 M sodium acetate and 3 volumes of ice-cold ethanol. Precipitated DNA was collected by centrifugation at 5,000 g at 4 °C for 5 min and washed thrice with ice-cold 75% ethanol to remove salts and detergents.

Cross-linking Reversal and DNA Removal

The purified DNA-protein complexes were resuspended in 0.5 M sodium acetate, and incubated at 68 °C overnight to reverse the DNA-protein cross-linking. After incubation, DNA was digested with 5 units of DNase I (Worthington Biochemical, Lakewood, NJ) and 5 units of S1 nuclease (Invitrogen) in a solution bearing 0.1 M sodium acetate (pH 5.5), 10 mM $MgCl_2$ and 10 mM $ZnCl_2$ at 37 °C for 1 h. The digested

nucleotides were removed by using a Microcon YM-10 centrifugal filter (Millipore, Billerica, MA). The purified DNA-binding proteins were quantified with Bradford Protein Assay kit (Bio-Rad, Hercules, CA).

SDS-PAGE Separation and Enzymatic Digestion

The purified DNA-binding proteins were separated by 1D SDS-PAGE using a 12% resolving gel with a 4% stacking gel, and stained with Coomassie blue. The gel was cut into 10 bands, in-gel reduced with dithiothreitol (DTT), alkylated with iodoacetamide (IAA) and digested with trypsin (Promega, Madison, WI). The digested peptides were collected, dried in a Speed-vac, and stored at –20 °C until further analysis.

Western Blotting

For Western blotting analysis, proteins were denatured and reduced by boiling in Laemmli loading buffer containing 80 mM DTT. After SDS-PAGE separation, proteins were transferred to a nitrocellulose membrane under standard conditions. The membrane was blocked with 2% non-fat milk ECL Advance blocking reagent (GE Healthcare, UK) and incubated with primary antibodies at 4 °C overnight. The rabbit polyclonal primary antibodies for MCM2 (Mini chromosome maintenance protein 2, a.k.a. DNA replication licensing factor MCM2) and actin were from Abcam (Cambridge, MA). The membrane was rinsed briefly with two changes of PBS-T washing buffer [PBS with 0.1% (v/v) Tween-20, pH 7.5] and washed with a large amount of washing buffer for 15 min, followed by 3 × 5 min wash with fresh changes of washing buffer at room temperature.

After washing, primary antibodies were recognized by incubating with horse radish peroxidase (HRP)-conjugated goat anti-rabbit IgG secondary antibody (Abcam) at room temperature for 1 h. The membrane was washed thoroughly with PBS-T (1× 15 min, then 3 × 5 min). The antibody binding was detected by using ECL Advance Western Blotting Detection Kit (GE Healthcare), and visualized with HyBlot CL Autoradiography Film (Denville Scientific Inc., Metuchen, NJ).

Extraction and Enzymatic Digestion of Nucleic Acids

After the above *in-vivo* chemical cross-linking, nucleic acids and nucleic acid-protein complexes were isolated from HL-60 cells using a standard phenol extraction protocol[36] or the above-described DNAzol method. For phenol extraction, the RNase digestion step was omitted so that both RNA and DNA could be isolated, and this sample was used as a control to estimate the relative amounts of RNA and DNA in the extract. Proteins present in the cross-linked complexes were removed by proteinase K digestion, and the remaining nucleic acids were precipitated by ethanol and digested to mononucleosides by using nuclease P1 (NP1, Sigma-Aldrich, St. Louis, MO) and calf intestinal phosphatase (CIP, Sigma-Aldrich). In this respect, 2 units of NP1 was added to a solution containing 30 μg of DNA, 50 mM sodium acetate and 1.0 mM $ZnCl_2$ (pH 5.5), and the digestion was continued at 37 °C for 12 h. The resulting sample was treated with 20 units of CIP in 50 mM Tris-HCl (pH 8.5) at 37 °C for 3 h. The digestion mixtures were passed through Microcon YM-10 centrifugal filter to remove enzymes and the resulting aliquots were subjected to HPLC analysis.

HPLC Separation

Off-line HPLC separation of nucleoside mixtures was performed on an Agilent 1100 HPLC pump with a 4.6 × 250 mm Grace Apollo C18 column (5 μm in particle size and 300 Å in pore size). A solution of 10 mM ammonium formate (solution A) and a 10 mM ammonium formate/acetonitrile mixture (70/30, v/v, solution B) were used as the mobile phases, and the flow rate was 0.8 mL/min. A gradient (0–5 min, 0–5% B; 5–45 min, 5–30% B, 45–50 min, 30–60% B) was used for the separation of the above nucleoside mixtures. The effluents were monitored by UV detection at 260 nm.

Nanoflow LC-MS/MS Analysis

Online LC-MS/MS analysis was performed on an Agilent 6510 Q-TOF system coupled with an Agilent HPLC-Chip Cube MS interface (Agilent Technologies, Santa Clara, CA). The sample enrichment, desalting, and HPLC separation were carried out automatically on the Agilent HPLC-Chip with an integrated trapping column (40 nL) and a separation column (Zorbax 300SB-C18, 75 μm × 150 mm, 5 μm in particle size). The peptide mixtures for LC-MS/MS analysis were first loaded onto the enrichment column and desalted with a solvent mixture of 0.1% formic acid in CH_3CN/H_2O (2:98, v/v) at a flow rate of 4 μL/min by using an Agilent 1200 capillary pump. After desalting, the peptide mixture was separated by using an Agilent 1200 nanoflow pump at a flow rate of 300 nL/min with the following gradient: 0-2min, 2% B; 2-10 min, 2-10% B; 10-90 min, 10-30% B; 90-120 min, 30-40% B; 120-130 min, 40-90% B. The gradient was held at 90% B for 5 min, and then changed to 2% B for equilibration for 15 min before the next

sample injection. The mobile phases were 0.1% formic acid in H_2O (A) and 0.1% formic acid in CH_3CN (B).

To maintain a stable nanospray during the whole analysis process, the capillary voltage (Vcap) applied to the HPLC-Chip capillary tip was 1900 V. For data collection, the Agilent Q-TOF was operated in an auto (data-dependent) MS/MS mode, where a full MS scan was followed by maximum of eight MS/MS scans (abundance-only precursor selection), with *m/z* ranges of 350-2000, and 60-2000 for MS and MS/MS scans, respectively. The active (dynamic) exclusion feature was enabled to discriminate against ions previously selected for MS/MS in two sequential scans. The acquisition rates were 6 and 3 spectra/s in MS and MS/MS modes, respectively. For collision-induced dissociation (CID), the collision energy was set at a slope of 3 V/100 Da and an offset of 2.5 V to fragment the selected precursor ions and give MS/MS.

Data Processing

Agilent MassHunter workstation software (Version B.01.03, Agilent Technologies) was used to extract the MS and MS/MS data from the LC-MS/MS results. The extracted LC-MS/MS data were converted to mzData files with MassHunter Qualitative Analysis. Mascot Server 2.2 with Mascot Daemon 2.2.2 (Matrix Science, London, UK) was used for protein and peptide identifications by searching LC-MS/MS data against UniProtKB/Swiss-Prot database (updated weekly). Carbamidomethylation of cysteine residues was used as a fixed modification. Methionine oxidation, serine, threonine and tyrosine phosphorylation were set as variable modifications. Stringent criteria were

employed for protein identification. The allowed maximum miscleavages per peptide was one, with a precursor tolerance of 20 ppm and a MS/MS tolerance of 0.6 Da (The average of absolute mass accuracy for all the identified peptides was calculated to be 7.9 ppm). Peptides identified with individual scores at or above the Mascot assigned homology score ($p < 0.01$ and individual peptide score > 40) were considered as specific peptide sequences. The false discovery rates (FDR, number of random matches divided by the total number of identified peptides) with homology or identity threshold, determined by using decoy database search, were less than 0.95%.

The cellular localization and function of identified proteins were assessed using Gene Ontology database (http://www.geneontology.org), Generic GO Term Mapper (http://go.princeton.edu) and GORetriever (http://www.agbase.msstate.edu).[37]

Results and Discussion

Strategy for the Identification of DNA-binding Proteins

DNA-binding proteins have a general or specific affinity for either single- or double-stranded DNA. Although a number of DNA-binding proteins have been identified, many proteins involved in gene regulation and DNA repair are likely still unknown because of their dynamic and/or weak interactions with DNA. DNA-protein cross-linking induced by ionizing radiation and chemotherapeutic agents such as aldehydes, cisplatin and other metal complexes has been studied as cytotoxic lesions.[38] Instead of studying the toxicity of cross-linking between DNA and proteins, we employed the *in vivo* DNA-protein cross-linking as a strategy to fix the DNA-protein interactions, and used a

standard DNA purification method to isolate DNA-binding proteins from complex biological samples. The idea of using immobilized oligodeoxyribonuleotides as probes has been demonstrated for purifying specific DNA-binding proteins, followed by protein identification with MS-based techniques.[39, 40] However, the previous methods, such as affinity-DNA probes and EMSA, usually depend on *in vitro* interactions with specific DNA sequences; therefore, they lack the ability to identify DNA-binding proteins at large scale and with high throughput.

To identify proteins with potential interactions with DNA, in this work, we described an approach for the comprehensive identification of DNA-binding proteins with formaldehyde cross-linking, which can fix DNA-protein interactions *in situ* and has been widely used in ChIP assays for probing *in vivo* chromatin structures and dynamics. As depicted in Figure 3.1, after the cross-linking reaction, cell nuclei were isolated, followed by the purification of cross-linked DNA-protein complexes. After DNA-binding proteins were copurified with genomic DNA, the DNA-protein cross-linking was reversed to release the DNA-binding proteins from DNA. The DNA was removed by using DNase digestion and centrifugation with Microcon centrifugal filters; the released DNA-binding proteins were further fractionated with SDS-PAGE and digested in-gel with trypsin. The extracted peptide mixtures from different gel bands were analyzed using LC-MS/MS, and the LC-MS/MS data were searched against a protein database for protein identification (Figure 3.1).

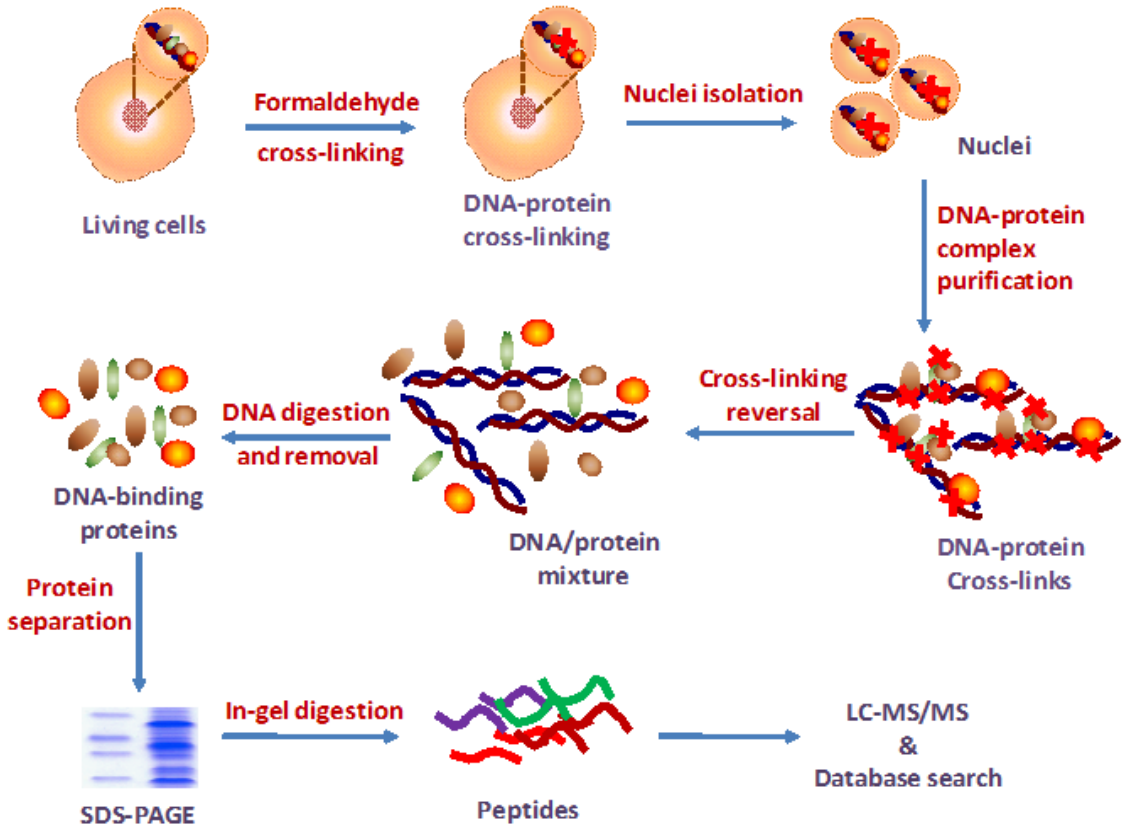


Figure 3.1 Strategy for the identification of DNA-binding proteins with chemical cross-linking and LC-MS/MS.

In Vivo DNA-protein Cross-linking

Formaldehyde-mediated DNA-protein cross-linking has been used for probing *in vivo* chromatin structures for more than two decades.[33] Formaldehyde is a tight and reactive reagent, which can lead to efficient cross-linking between nucleic acids and proteins within short distance *in vivo*. The formation of covalent DNA-protein complexes prevents the subsequent dissociation and redistribution of proteins due to changes in physiological condition and/or in the process of sample handling.[27] It has been found that formaldehyde is incapable of inducing protein-DNA cross-links *in vitro* even at extremely high concentrations.[33] This suggests that formaldehyde-mediated *in vivo* DNA-protein cross-linking is due to physiological DNA-protein interactions.

Aside from nucleic acid-protein cross-linking, protein-protein cross-linking can be generated by formaldehyde *in vivo*, especially upon long-term incubation and/or with high concentrations of cross-linking reagent. It is well-documented that nucleosomal proteins are normally analyzed following a cross-linking time within 10 min.[10] Longer exposure to formaldehyde favors the binding of nucleosome-associated proteins and protein-protein interactions, and most proteins are readily cross-linked following a 20-60 min cross-linking reaction with 1% formaldehyde, based on the study of specific targets of interest.[10, 31] However, in our experiment, it is unrealistic to optimize the cross-linking conditions for all DNA-interacting proteins based on specific DNA sequences and protein targets. We tested with different cross-linking reaction time with HL-60 cells and found that longer cross-linking time leads to low yield and poor solubility of DNA-protein

complexes. Therefore, we incubated HL-60 cells in 1% formaldehyde for 10 min to generate the DNA-protein cross-links.

Isolation of DNA-binding Proteins

The enrichment of DNA-binding proteins was achieved through the isolation of DNA-protein complexes after *in vivo* chemical cross-linking. To evaluate the enrichment of DNA-binding proteins, we performed Western blotting experiments to assess the amounts of two identified proteins, MCM2 and actin, in whole cell lysates, nuclear fractions, and the enriched fraction of DNA-binding proteins.

DNA replication licensing factor MCM2 belongs to the MCM family proteins, which are DNA-dependent ATPases required for the initiation of eukaryotic DNA replication.[41] MCM2 is a component of the prereplicative complex; it is essential for eukaryotic DNA replication and is expressed only in proliferating cells.[42] Our Western blotting results revealed that MCM2, a DNA-binding protein, can be enriched by at least 17 fold after the isolation of DNA-protein complexes formed by *in vivo* chemical cross-linking compared to actin, which was used as an internal standard for this comparison (Figure 3.2). However, the nuclear fraction showed no significant enrichment of MCM2 with respect to actin. This could be attributed to the fact that a relatively high concentration of actin is present in the nuclear fraction.[43] The background signals of MCM2 in high mass ranges in DNA-binding protein fraction could emanate from the undissociated cross-linked nucleotides/or proteins to MCM2. The presence of residual DNA-protein cross-links may lead to the failure in identifying some DNA-binding

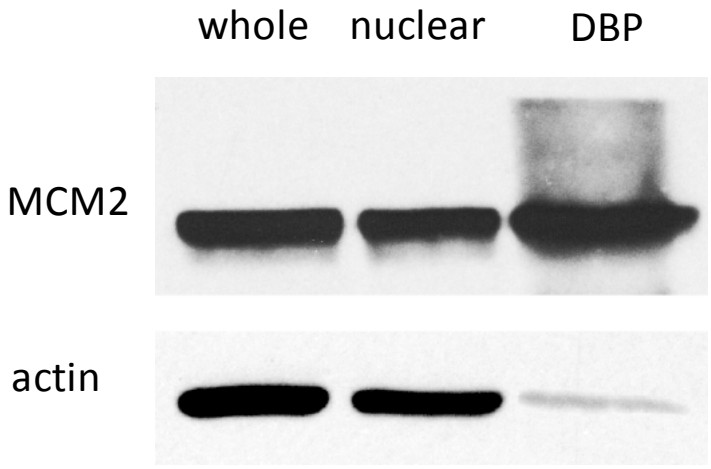


Figure 3.2 Western blotting of MCM2 was performed in whole cell lysate (whole), nuclear fraction (nuclear), and the DNA-binding protein fraction (DBP) of HL-60 cells. Actin was used as the internal standard for quantitative comparison.

proteins. In this context, because of the heterogeneity of the cross-linking, i.e., different nucleobases and amino acid residues may participate in the formaldehyde-mediated cross-linking,[10, 32, 33] it is difficult to incorporate the protein side-chain modifications, arising from the incomplete cross-linking reversal, into the database search for protein identification.

It is interesting to observe that actin can also be detected in the isolated DNA-binding protein fraction, which cannot be simply attributed to the contaminations from cytoplasmic or nuclear actin. In fact, it has been reported that actin could be associated with DNA and RNA during transcription.[43-45] An alternative explanation is that actin could cross-link to other proteins that can bind to DNA. In a separate experiment, we followed the identical protocol for DNA-protein complex purification and isolated the genomic DNA without chemical cross-linking reaction. The result showed that a relatively small amount of MCM2 can be detected in the DNA fraction (Supporting Information, Figure S3.1). By contrast, no obvious actin band was visible in the same DNA fraction. This result indicates that the DNA-bound actin might be lost during DNA isolation without cross-linking, or the amount of actin directly associated with DNA, if any, could be very small.

The Selectivity of the Method toward the Isolation of DNA-binding Proteins

Several precautions were exerted to improve the selectivity of the above-described method toward the isolation of DNA-binding proteins. First, we began with isolated nuclei rather than the whole lysate of the formaldehyde-treated cells, which minimizes

the contamination of cytosolic proteins. Second, we incubated the isolated protein-DNA complexes in a solution containing high concentrations of urea and SDS to dissociate and remove the non-cross-linked proteins from DNA-protein complexes. Third, we adopted a DNAzol-based protocol for the selective isolation of DNA and DNA-protein cross-links.[34] The basis of the DNAzol procedure lies in the use of a novel guanidine-detergent lysis solution that hydrolyzes RNA and allows the selective precipitation of DNA and DNA-protein cross-links from a cell lysate.

While it is difficult to evaluate directly the selectivity of this method because many proteins can bind both DNA and RNA, we chose to use an indirect method to assess how selective the method is toward the isolation of DNA- over RNA-binding proteins. In this respect, we isolated DNA and DNA-protein cross-links from HL-60 cells after formaldehyde-induced cross-linking reaction by using two different protocols, i.e., phenol extraction and the DNAzol method. We then removed the proteins by using proteinase K treatment, digested the remaining nucleic acids to nucleosides with two enzymes and analyzed the resulting nucleosides by HPLC analysis (See Materials and Methods). As depicted in Figure 3.3, the amount of ribonucleosides present in the nucleoside mixture emanating from the DNAzol method is much less than that of 2'-deoxynucleosides (<5%, Figure 3.3B), whereas ribonucleosides are present at a much higher level in the nucleoside mixture arising from the phenol extraction method (Figure 3.3A). This result, therefore, suggests that the DNAzol method is more selective toward the isolation of DNA and its associated proteins than RNA and its binding proteins.

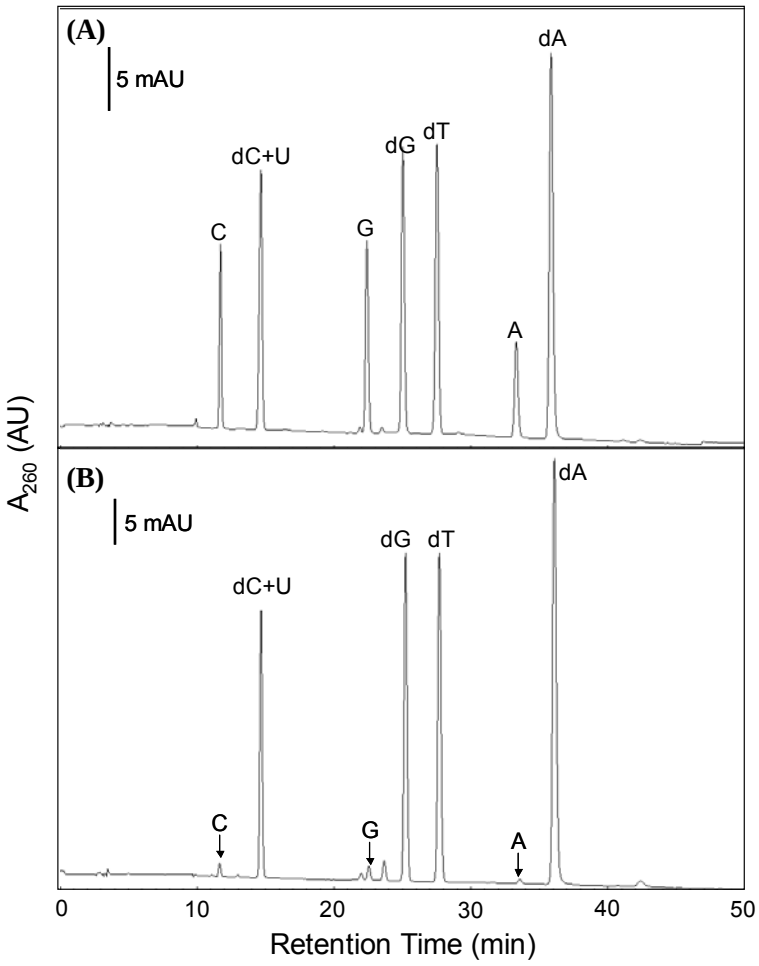


Figure 3.3 HPLC traces for the separation of nucleoside mixtures arising from the enzymatic digestion of nucleic acids that were isolated from the *in-vivo* chemically cross-linked DNA-protein complexes by the standard phenol extraction method (A) and the DNAzol method (B). The identities of nucleosides were confirmed by HPLC analysis of authentic compounds. “A”, “C”, “G”, “U” designate the four natural ribonucleosides, and “dA”, “dC”, “dG”, and “dT” represent the four natural 2’-deoxyribonucleosides.

Identification and Characterization of Proteins with In Vivo DNA-Protein Cross-linking

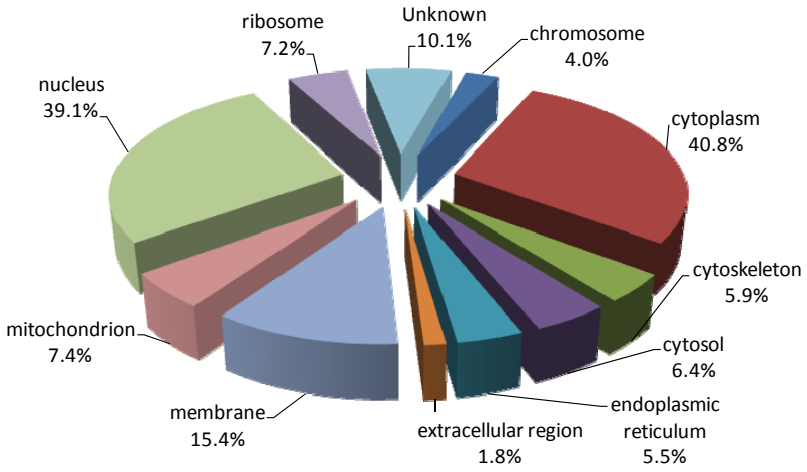
The current protocol with *in vivo* DNA-protein chemical cross-linking enabled us to enrich the DNA-binding proteins through the isolation of DNA-protein complexes. The cross-linking generated by formaldehyde stabilizes DNA-protein complexes and allows the capture of transient interactions between DNA and proteins. Based on this protocol, we were able to identify 780 proteins with Mascot database search.

To further understand the distribution and function of the identified proteins, we investigated the GO annotations of the identified proteins using Gene Ontology database (http://www.geneontology.org), Generic GO Term Mapper (http://go.princeton.edu) and GORetriever (http://www.agbase.msstate.edu).[37] Among the proteins with GO annotations, 305 unique proteins were classified as nuclear proteins (GO: 0005634), which represented 39.1% of the total proteins identified (Figure 3.4A). The presence of a relatively high percentage of nuclear proteins in the purified fraction could be attributed to the isolation of the DNA-protein complexes. In addition to the nuclear localization, 40.8% of the identified proteins were annotated as cytoplasmic proteins (GO: 0005737), followed by membrane (GO: 0016020, 15.4%), mitochondrial (GO: 0005739, 7.4%), ribosomal (GO: 0005840, 7.2%) proteins, etc. Because of the lack of annotation information, more than 10% of the identified proteins cannot be classified based on GO slim terms. On the other hand, many proteins can be grouped into multiple GO categories, which can cause overlaps. Therefore, the sum of percentages of GO categories is over 100%. It is also worth noting that many proteins were only annotated with their major cellular localizations and functions. For instance, actin is also localized in the nucleus and

is associated with gene regulation (Figure 3.2)[43-45]; however, the current GO databases classify actin only as a cytoplasmic protein.

Apart from the classification of the cellular distribution of the identified proteins, we also organized the identified proteins, according to their biological functions (Figure 3.4B). Among the identified proteins with GO molecular function annotation, 111 proteins (14.2% of the identified proteins) were clearly annotated as DNA-binding proteins (GO: 0003677), including double-stranded/single-stranded DNA binding, damaged DNA binding, sequence-specific/structure-specific DNA binding, transcription regulation, DNA helicase/ligase activity, etc. More functional categories were shown as nucleotide binding (GO: 0000166, 23.8%), RNA-binding (GO: 0003723, 20.0%), hydrolase activity (GO: 0016787, 9.0%), etc. The identification of proteins without known DNA-binding activity could arise from the cross-linking of these proteins with DNA-binding proteins through protein-protein interactions. Additionally, some of these proteins may bear DNA binding capabilities that have yet been characterized. Furthermore, formaldehyde may also give rise to non-specific protein-protein cross-links, thereby resulting in the isolation and identification of proteins that are not capable of binding to DNA or its associated proteins. Guisan et al.[46] showed that the addition of dextran could improve the specificity of aldehyde-induced cross-linking of purified proteins. In addition, it was found that weakly bound protein-protein complexes could be selectively adsorbed on lowly activated anionic exchange support, thereby preserving the weak protein-protein interactions.[47] These methods are, however, not applicable for the formaldehyde-induced *in-vivo* cross-linking discussed in the current paper.

(A)



(B)

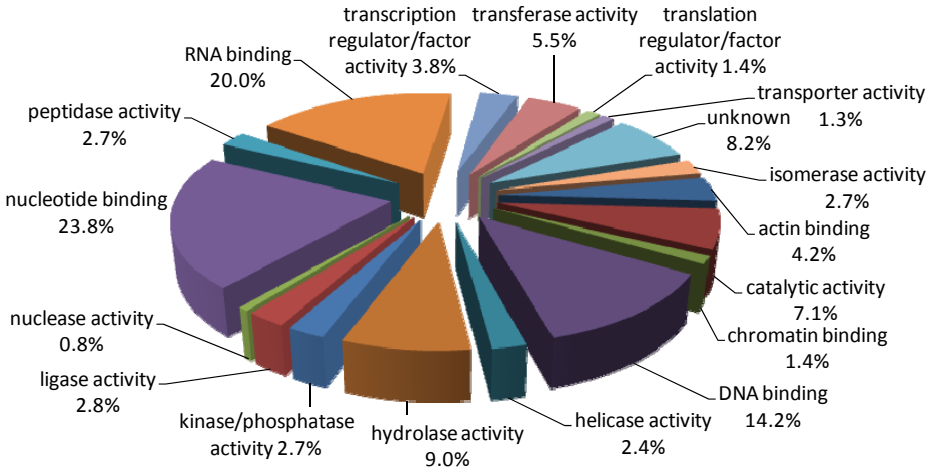


Figure 3.4 The identified proteins were classified using GO annotations. (A) Distribution of proteins in major cellular localizations; (B) Classification of proteins according to their function. Percentages indicate the identified proteins in each category relative to the total number of identified proteins.

To gain further knowledge about the identified DNA-binding proteins, we extracted the comprehensive GO annotations including the molecular function and biological process. According to the GO annotations, at least 67 identified proteins are involved in gene transcription, and 20 identified proteins are involved in DNA replication. For instance, 23 transcription factors were identified in our experiment (Table 3.1). In addition, almost a complete family of DNA replication licensing factors was identified in our experiments, which include all six subunits of the MCM protein complex, i.e., MCM2 through MCM7. These six subunits form a ring-shaped heterohexameric ATPase involved in DNA replication.[42, 48]

Besides the proteins involved in transcription and replication, we were able to identify at least 26 proteins involved in DNA damage and repair processes, such as base-excision repair, nucleotide-excision repair and double-strand break repair, as shown in Table 3.2. It is worth noting that some of identified proteins involved in DNA damage response and repair, however, are not annotated as DNA-binding proteins in the current GO databases. These proteins include MMS19 nucleotide excision repair protein homolog (MMS19), FACT complex subunit SPT16 (SUPT16H), transitional endoplasmic reticulum ATPase (VCP), etc. Because of the lack of proper GO annotations, these proteins were not counted as DNA-binding proteins depicted in Figure 3.4B.

Table 3.1 A list of transcription factors identified in the present study.

Uniprot ID	Protein Name	Gene Name	MW (Da)	Function
ENOA_HUMAN	Alpha-enolase	ENO1	47169	transcription factor activity, transcription repressor activity, transcription corepressor activity
ARNT_HUMAN	Aryl hydrocarbon receptor nuclear translocator	ARNT	86636	transcription factor activity, transcription coactivator activity, signal transduction
CREB1_HUMAN	cAMP response element-binding protein	CREB1	36688	transcription factor activity, sequence-specific DNA binding, signal transduction
CNBP_HUMAN	Cellular nucleic acid-binding protein	CNBP	19463	single-stranded DNA binding, single-stranded RNA binding, transcription activity
DBPA_HUMAN	DNA-binding protein A	CSDA	40090	double-stranded DNA binding, transcription factor activity, transcription corepressor activity
EDF1_HUMAN	Endothelial differentiation-related factor 1	EDF1	16369	regulation of transcription, transcription factor activity, transcription coactivator activity
ELF1_HUMAN	ETS-related transcription factor Elf-1	ELF1	67456	sequence-specific DNA binding, transcription activator activity, transcription factor activity, transcription repressor activity

FUBP1_HUMAN	Far upstream element-binding protein 1	FUBP1	67560	single-stranded DNA binding, transcription factor activity
HCLS1_HUMAN	Hematopoietic lineage cell-specific protein	HCLS1	53998	transcription factor activity, regulation of transcription
ROAA_HUMAN	Heterogeneous nuclear ribonucleoprotein A/B	HNRPAB	36225	transcription factor activity, positive regulation of gene-specific transcription
HMGB2_HUMAN	High mobility group protein B2	HMGB2	24034	DNA bending activity, base-excision repair, DNA replication, transcription factor activity
HMGA1_HUMAN	High mobility group protein HMG-I/HMG-Y	HMGA1	11676	AT DNA binding, transcription factor binding, transcription factor activity
HCFC1_HUMAN	Host cell factor	HCFC1	208732	transcription factor activity, transcription coactivator activity, regulation of transcription
MEF2D_HUMAN	Myocyte-specific enhancer factor 2D	MEF2D	55938	transcription factor activity, regulation of transcription
YBOX1_HUMAN	Nuclease-sensitive element-binding protein 1	YBX1	35924	transcription factor activity, regulation of transcription, transcription repressor activity
NDKB_HUMAN	Nucleoside diphosphate kinase B	NME2	17298	transcription factor activity, regulation of transcription

PA2G4_HUMAN	Proliferation-associated protein 2G4	PA2G4	43787	transcription factor activity, regulation of transcription
MAX_HUMAN	Protein max	MAX	18275	transcription regulator activity, transcription factor activity, transcription cofactor activity
STAT3_HUMAN	Signal transducer and activator of transcription 3	STAT3	88068	transcription factor activity, regulation of transcription, transcription factor binding
TADBP_HUMAN	TAR DNA-binding protein 43	TARDBP	44740	transcription factor activity, regulation of transcription
SPT6H_HUMAN	Transcription elongation factor SPT6	SUPT6H	199073	transcription factor activity, regulation of transcription
TIF1B_HUMAN	Transcription intermediary factor 1-beta	TRIM28	88550	transcription factor activity, regulation of transcription, transcription corepressor activity
ZN207_HUMAN	Zinc finger protein 207	ZNF207	50751	transcription factor activity, regulation of transcription

Table 3.2 A list of identified proteins involved in DNA repair.

Uniprot ID	Protein Name	Gene Name	MW (Da)	Function
KU70_HUMAN	ATP-dependent DNA helicase 2 subunit 1	XRCC6	69843	DNA ligation, double-strand break repair, DNA recombination
KU86_HUMAN	ATP-dependent DNA helicase 2 subunit 2	XRCC5	82705	DNA recombination, double-strand break repair
DDB1_HUMAN	DNA damage-binding protein 1	DDB1	126968	damaged DNA binding, nucleotide-excision repair, DNA damage removal
DNL1_HUMAN	DNA ligase 1	LIG1	101736	ligase activity, DNA replication, DNA repair
MCM7_HUMAN	DNA replication licensing factor MCM7	MCM7	81308	DNA replication, transcription, response to DNA damage stimulus
TOP2A_HUMAN	DNA topoisomerase 2-alpha	TOP2A	174385	DNA topoisomerase activity, DNA-dependent ATPase activity, DNA repair
APEX1_HUMAN	DNA-(apurinic or apyrimidinic site) lyase	APEX1	35555	lyase activity, base-excision repair, transcription coactivator activity
PRKDC_HUMAN	DNA-dependent protein kinase catalytic subunit	PRKDC	469089	DNA-dependent protein kinase activity, DNA recombination, DNA repair
SP16H_HUMAN	FACT complex subunit SPT16	SUPT16H	119914	DNA replication, DNA repair, nucleosome disassembly

HMG1X_HUMAN	High mobility group protein 1-like 10	HMG1L10	24218	protein binding, base-excision repair, regulation of transcription
HMGB1_HUMAN	High mobility group protein B1	HMGB1	24894	base-excision repair, signal transduction, transcription factor binding
HMGB2_HUMAN	High mobility group protein B2	HMGB2	24034	DNA bending activity, base-excision repair, DNA replication, transcription factor activity
MMS19_HUMAN	MMS19 nucleotide excision repair protein homolog	MMS19	113290	nucleotide-excision repair, transcription
NONO_HUMAN	Non-POU domain-containing octamer-binding protein	NONO	54232	RNA splicing, DNA replication, DNA repair
PARP1_HUMAN	Poly [ADP-ribose] polymerase 1	PARP1	113084	transferase activity, DNA repair, response to DNA damage stimulus
PCNA_HUMAN	Proliferating cell nuclear antigen	PCNA	28769	DNA binding, nucleotide-excision repair, DNA replication
RENT1_HUMAN	Regulator of nonsense transcripts 1	UPF1	124345	helicase activity, hydrolase activity, DNA replication, DNA repair
RBM14_HUMAN	RNA-binding protein 14	RBM14	69492	RNA binding, DNA recombination, DNA repair, transcription
RUVB2_HUMAN	RuvB-like 2	RUVBL2	51157	damaged DNA binding, DNA recombination, DNA repair

SFPQ_HUMAN	Splicing factor, proline- and glutamine-rich	SFPQ	76149	RNA splicing, DNA recombination, DNA repair, response to DNA damage stimulus
SMC1A_HUMAN	Structural maintenance of chromosomes protein 1A	SMC1A	143233	DNA repair, DNA damage response, signal transduction
SMC3_HUMAN	Structural maintenance of chromosomes protein 3	SMC3	141542	protein binding, ATPase activity, DNA repair, signal transduction
SODC_HUMAN	Superoxide dismutase [Cu-Zn]	SOD1	15936	double-strand break repair, DNA fragmentation during apoptosis
TERA_HUMAN	Transitional endoplasmic reticulum ATPase	VCP	89322	ATPase activity, double-strand break repair, response to DNA damage stimulus
TP53B_HUMAN	Tumor suppressor p53-binding protein 1	TP53BP1	213574	transcription activator activity, damaged DNA binding, DNA repair
RD23B_HUMAN	UV excision repair protein RAD23 homolog B	RAD23B	43171	damaged DNA binding, single-stranded DNA binding, nucleotide-excision repair

Conclusions

We described an approach for the comprehensive investigation of DNA-binding proteins with *in vivo* formaldehyde cross-linking. DNA-binding proteins can be purified via the isolation of DNA-protein complexes and released from the complexes by reversing the DNA-protein cross-linking. By using this method, we were able to identify more than one hundred DNA-binding proteins, including those involved in transcription, gene regulation, DNA replication and repair, and a large number of proteins which are potentially associated with DNA and DNA-binding proteins. This method should be generally applicable to the investigation of other nucleic acid-binding proteins, and hold great potential in the comprehensive study of gene regulation, DNA damage response and repair, as well as many other important biological processes at proteomic level. We believe that this function-oriented protein purification strategy may serve as a valuable tool for studying *in vivo* DNA-protein interaction and dynamic cellular responses to perturbations and stress, not only by identifying potential protein targets, transcription factors or DNA repair enzymes, but also by determining protein modifications and differential expression by incorporating other techniques, e.g., stable isotope labeling.

References

1. Carey, M.; Smale, S. T., *Transcriptional regulation in eukaryotes: concepts, strategies, and techniques*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2000.
2. Levine, M.; Tjian, R., Transcription regulation and animal diversity. *Nature* **2003,** 424, 147-151.
3. Naar, A. M.; Lemon, B. D.; Tjian, R., Transcriptional coactivator complexes. *Annu. Rev. Biochem.* **2001,** 70, 475-501.
4. Garner, M. M.; Revzin, A., The use of gel-electrophoresis to detect and study nucleic-acid protein interactions. *Trends Biochem. Sci.* **1986,** 11, 395-396.
5. Fried, M. G., Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis* **1989,** 10, 366-376.
6. Hellman, L. M.; Fried, M. G., Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2007,** 2, 1849-61.
7. Chen, H. W.; Lin, R. J.; Xie, W.; Wilpitz, D.; Evans, R. M., Regulation of hormone-induced histone hyperacetylation and gene activation via acetylation of an acetylase. *Cell* **1999,** 98, 675-686.
8. Spencer, V. A.; Sun, J. M.; Li, L.; Davie, J. R., Chromatin immunoprecipitation: a tool for studying histone acetylation and transcription factor binding. *Methods* **2003,** 31, 67-75.

9. Kuo, M. H.; Allis, C. D., In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* **1999,** 19, 425-433.

10. Orlando, V., Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* **2000,** 25, 99-104.

11. Ahn, N. G.; Shabb, J. B.; Old, W. M.; Resing, K. A., Achieving in-depth proteomics profiling by mass spectrometry. *ACS Chem. Biol.* **2007,** 2, 39-52.

12. Cravatt, B. F.; Simon, G. M.; Yates, J. R., The biological impact of mass-spectrometry-based proteomics. *Nature* **2007,** 450, 991-1000.

13. Nordhoff, E.; Lehrach, H., Identification and characterization of DNA-binding proteins by mass spectrometry. *Adv. Biochem. Eng. Biotechnol.* **2007,** 104, 111-195.

14. Barthelery, M.; Salli, U.; Vrana, K. E., Nuclear proteomics and directed differentiation of embryonic stem cells. *Stem Cells Dev.* **2007,** 16, 905-919.

15. Barthelery, M.; Salli, U.; Vrana, K. E., Enhanced nuclear proteomics. *Proteomics* **2008,** 8, 1832-1838.

16. Escobar, M. A.; Hoelz, D. J.; Sandoval, J. A.; Hickey, R. J.; Grosfeld, J. L.; Malkas, L. H., Profiling of nuclear extract proteins from human neuroblastoma cell lines: the search for fingerprints. *J. Pediatr. Surg.* **2005,** 40, 349-358.

17. Jung, E.; Hoogland, C.; Chiappe, D.; Sanchez, J. C.; Hochstrasser, D. F., The establishment of a human liver nuclei two-dimensional electrophoresis reference map. *Electrophoresis* **2000,** 21, 3483-3487.

18. Malmstrom, J.; Larsen, K.; Malmstrom, L.; Tufvesson, E.; Parker, K.; Marchese, J.; Williamson, B.; Patterson, D.; Martin, S.; Juhasz, P.; Westergren-Thorsson, G.; Marko-Varga, G., Nanocapillary liquid chromatography interfaced to tandem matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry: Mapping the nuclear proteome of human fibroblasts. *Electrophoresis* **2003,** 24, 3806-3814.

19. Tan, F.; Li, G. S.; Chitteti, B. R.; Peng, Z. H., Proteome and phosphoproteome analysis of chromatin associated proteins in rice (*Oryza sativa*). *Proteomics* **2007,** 7, 4511-4527.

20. Andersen, J. S.; Lam, Y. W.; Leung, A. K. L.; Ong, S. E.; Lyon, C. E.; Lamond, A. I.; Mann, M., Nucleolar proteome dynamics. *Nature* **2005,** 433, 77-83.

21. Andersen, J. S.; Lyon, C. E.; Fox, A. H.; Leung, A. K. L.; Lam, Y. W.; Steen, H.; Mann, M.; Lamond, A. I., Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **2002,** 12, 1-11.

22. Coute, Y.; Burgess, J. A.; Diaz, J. J.; Chichester, C.; Lisacek, F.; Greco, A.; Sanchez, J. C., Deciphering the human nucleolar proteome. *Mass Spectrom. Rev.* **2006,** 25, 215-234.

23. Hinsby, A. M.; Kiemer, L.; Karlberg, E. O.; Lage, K.; Fausboll, A.; Juncker, A. S.; Andersen, J. S.; Mann, M.; Brunak, S., A wiring of the human nucleolus. *Mol. Cell* **2006,** 22, 285-295.

24. Scherl, A.; Coute, Y.; Deon, C.; Calle, A.; Kindbeiter, K.; Sanchez, J. C.; Greco, A.; Hochstrasser, D.; Diaz, J. J., Functional proteomic analysis of human nucleolus. *Mol. Biol. Cell* **2002,** 13, 4100-4109.

25. Henrich, S.; Cordwell, S. J.; Crossett, B.; Baker, M. S.; Christopherson, R. I., The nuclear proteome and DNA-binding fraction of human Raji lymphoma cells. *Biochim. Biophys. Acta, Proteins Proteomics* **2007,** 1774, 413-432.

26. Afjehi-Sadat, L.; Engidawork, E.; Slavc, I.; Lubec, G., Comparative proteomic analysis of nucleic acid-binding proteins in ten human tumor cell lines. *Int. J. Oncol.* **2006,** 28, 173-190.

27. Vigneault, F.; Guerin, S. L., Regulation of gene expression: probing DNA-protein interactions in vivo and in vitro. *Expert Rev. Proteomics* **2005,** 2, 705-718.

28. Jackson, V., Formaldehyde cross-linking for studying nucleosomal dynamics. *Methods* **1999,** 17, 125-139.

29. Orlando, V.; Strutt, H.; Paro, R., Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods* **1997,** 11, 205-214.

30. Guerrero, C.; Tagwerker, C.; Kaiser, P.; Huang, L., An integrated mass spectrometry-based proteomic approach - Quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. *Mol. Cell. Proteomics* **2006,** 5, 366-378.

31. Vasilescu, J.; Guo, X. C.; Kast, J., Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics* **2004,** 4, 3845-3854.

32. Metz, B.; Kersten, G. F. A.; Hoogerhout, P.; Brugghe, H. F.; Timmermans, H. A. M.; de Jong, A.; Meiring, H.; ten Hove, J.; Hennink, W. E.; Crommelin, D. J. A.; Jiskoot, W., Identification of formaldehyde-induced modifications in proteins - Reactions with model peptides. *J. Biol. Chem.* **2004,** 279, 6235-6243.

33. Solomon, M. J.; Varshavsky, A., Formaldehyde-mediated DNA protein crosslinking - a probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. U.S.A.* **1985,** 82, 6470-6474.

34. Barker, S.; Murray, D.; Zheng, J.; Li, L.; Weinfeld, M., A method for the isolation of covalent DNA-protein crosslinks suitable for proteomics analysis. *Anal. Biochem.* **2005,** 344, 204-215.

35. Barker, S.; Weinfeld, M.; Zheng, J.; Li, L.; Murray, D., Identification of mammalian proteins cross-linked to DNA by ionizing radiation. *J. Biol. Chem.* **2005,** 280, 33826-33838.

36. Sambrook, J.; Russell, D. W., *Molecular Cloning: A Laboratory Manual*. 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2001.

37. McCarthy, F. M.; Bridges, S. M.; Wang, N.; Magee, G. B.; Williams, W. P.; Luthe, D. S.; Burgess, S. C., AgBase: a unified resource for functional analysis in agriculture. *Nucleic Acids Res.* **2007,** 35, D599-D603.

38. Reardon, J. T.; Cheng, Y.; Sancar, A., Repair of DNA-protein cross-links in mammalian cells. *Cell Cycle* **2006,** 5, 1366-1370.

39. Nordhoff, E.; Krogsdam, A. M.; Jorgensen, H. F.; Kallipolitis, B. H.; Clark, B. F. C.; Roepstorff, P.; Kristiansen, K., Rapid identification of DNA-binding proteins by mass spectrometry. *Nat. Biotechnol.* **1999,** 17, 884-888.

40. Yaneva, M.; Tempst, P., Affinity capture of specific DNA-binding proteins for mass spectrometric identification. *Anal. Chem.* **2003,** 75, 6437-6448.

41. Yan, H.; Gibson, S.; Tye, B. K., Mcm2 and Mcm3, 2 proteins important for Ars activity, are related in structure and function. *Genes Dev.* **1991,** 5, 944-957.

42. Ramnath, N.; Hernandez, F. J.; Tan, D. F.; Huberman, J. A.; Natarajan, N.; Beck, A. F.; Hyland, A.; Todorov, I. T.; Brooks, J. S. J.; Bepler, G., MCM2 is an independent predictor of survival in patients with non-small-cell lung cancer. *J. Clin. Oncol.* **2001,** 19, 4259-4266.

43. Valkov, N. I.; Ivanova, M. I.; Uscheva, A. A.; Krachmarov, C. P., Association of actin with DNA and nuclear matrix from guerin ascites tumor cells. *Mol. Cell. Biochem.* **1989,** 87, 47-56.

44. Nakayasu, H.; Yoshimura, Y.; Ueda, K., Association of actin-filaments with the nuclear matrix from bovine lymphocytes. *Cell Struct. Funct.* **1981,** 6, 424-424.

45. Verheijen, R.; Vanvenrooij, W.; Ramaekers, F., The nuclear matrix - structure and composition. *J. Cell Sci.* **1988,** 90, 11-36.

46. Fuentes, M.; Segura, R. L.; Abian, O.; Betancor, L.; Hidalgo, A.; Mateo, C.; Fernandez-Lafuente, R.; Guisan, J. M., Determination of protein-protein interactions through aldehyde-dextran intermolecular cross-linking. *Proteomics* **2004,** 4, 2602-2607.

47. Fuentes, M.; Mateo, C.; Pessela, B. C. C.; Guisan, J. M.; Fernandez-Lafuente, R., Purification, stabilization, and concentration of very weak protein-protein complexes: Shifting the association equilibrium via complex selective adsorption on lowly activated supports. *Proteomics* **2005,** 5, 4062-4069.

48. Chong, J. P. J.; Mahbubani, H. M.; Khoo, C. Y.; Blow, J. J., Purification of an Mcm-containing complexes a component of the DNA-replication licensing system. *Nature* **1995,** 375, 418-421.

Supporting Information for Chapter 3

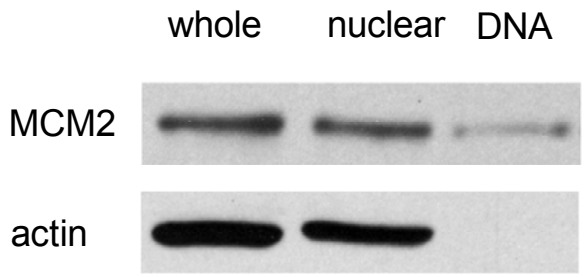


Figure S3.1 Western blotting of MCM2 in whole cell lysates (whole), nuclear fractions (nuclear), and DNA fractions of HL-60 cells. Actin was used as the internal standard for the quantitative comparison.

CHAPTER 4

Quantitative Analysis of Surface Plasma Membrane Proteins of Primary and Metastatic Melanoma Cells

Introduction

Membrane proteins represent about one-third of the proteins encoded by the human genome and assume a variety of key biological functions, such as cell-to-cell recognition, signal relay, and transport of ions and solutes.[1] Because of their accessibility, membrane proteins also constitute more than two-thirds of the current protein-based drug targets.[2] Despite their importance in biological processes and drug discovery, membrane proteins are usually challenging to study because of the difficulty in preparing clean plasma membrane (PM) fractions. In this context, PM proteins are hydrophobic and they have to be transferred into aqueous solution to be available for subsequent analysis. In addition, the relatively low abundances of most membrane proteins impose further challenges for determining their identities and functions.

Traditional two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is useful for the separation and relative quantification of proteins at large scale, but this technique does not resolve membrane proteins very well, mainly because of their poor solubility in the buffers used for protein separation in the first dimension.[3] Recently, advances in MS-based proteomics have facilitated the identification of a large number of

membrane proteins and their post-translational modifications by using multidimensional LC coupled with MS/MS, which could overcome the disadvantages associated with conventional 2D-PAGE.[4-9] These multidimensional LC-based techniques improved the separation power and could facilitate the identification of more proteins from complex mixtures. However, it is difficult to isolate and enrich PMs to reduce the sample complexity, and the majority of the proteins identified are not membrane-related. In addition, the PM protein fractions prepared by the typical ultracentrifugation method are usually contaminated with other cellular components.[10] Therefore, selective enrichment is frequently essential for the study of low-abundance membrane and membrane-associated proteins.

Biotinylation and affinity purification of cell-surface proteins, when coupled with MS, can improve the throughput and selectivity for the detection of membrane proteins in low abundance.[11-23] For example, relative enrichments of 1600- and 400-folds of PM proteins versus mitochondria and endoplasmic reticulum (ER), two major sources of contaminations in PM fractions prepared by ultracentrifugation methods, were achieved by using a commercial biotinylation reagent followed by 1D-SDS-PAGE separation and LC-MS/MS analysis.[11] In previous studies, the biotin-conjugated proteins were frequently purified and enriched after the biotinylation reaction and cell lysis. The affinity-enriched membrane fractions, however, are still a complex mixture and contain significant contamination of cytoplasmic proteins. Further separation of the enriched membrane proteins by using SDS-PAGE, 2-D gel or HPLC is often necessary to reduce sample complexity.[14, 19, 24]

Different MS-based strategies have been developed for the quantitative profiling of proteins.[25] However, quantification of membrane proteins at large scale seems to be more challenging than that of soluble proteins. For instance, isotope-coded affinity tag (ICAT),[26] a widely used chemical tagging reagent for protein quantification, has been rarely employed for interrogating membrane proteins, probably as a result of the poor compatibility of ICAT with the high concentration of detergent required for dissolving membrane proteins.[27] Recently, an optimized strategy for ICAT was published, and it was demonstrated that ICAT could be used for the quantitative analysis of hydrophobic proteins.[28] Nevertheless, a relatively small number of papers have been published for the quantitative assessment of membrane proteins.[19, 29-33] Among the various isotope-labeling approaches, stable isotope labeling by amino acids in cell culture (SILAC) is simple and accurate,[34] and SILAC has been demonstrated to be suitable for the quantitative analysis of cell membrane proteins using SDS-PAGE and LC-MS/MS.[29]

To improve the purification efficiency and to minimize contamination, in this Chapter, we describe an approach for the selective quantification of surface membrane proteins. In our strategy, intact cells cultured in SILAC light and heavy media were labeled *in situ* with a membrane impermeable biotin reagent, sulfo-NHS-LC-biotin, which can conjugate with the N-terminus or the side chain amino group of lysine residues in a protein to afford an amide linkage. The light and heavy cell lysates were mixed and digested with trypsin. The biotinylated peptides of membrane proteins were subsequently enriched via affinity purification and subjected to LC-MS/MS analysis.

We employed the above strategy and assessed the expression of cell surface membrane proteins in two human melanoma cell lines, WM-115 and WM-266-4. We were able to identify more than 100 membrane and membrane-associated proteins and quantified the relative expression of 62 identified proteins in these two cell lines. To the best of our knowledge, this is the first characterization of the differential expression of cell surface membrane proteins of primary and metastatic tumor cells using MS-based quantitative proteomic techniques.

Materials and Methods

Cell Culture

A pair of human malignant melanoma cell lines, WM-115 and WM-266-4, which were derived from the primary and metastatic tumor tissues of the same patient, were obtained from the American type Culture Collection (ATCC, Manassas, VA). The cells were cultured under ATCC-recommended conditions in Eagle's Minimum Essential Medium (EMEM) supplied with 10% fetal bovine serum (FBS, Invitrogen, Carlsbad, CA), 100 IU/mL of penicillin, 100 μg/mL of streptomycin and 1.5 g/L of sodium bicarbonate. For SILAC experiments, the EMEM medium in the absence of L-lysine was custom prepared by AthenaES (Baltimore, MD). [$^{13}C_6$,$^{15}N_2$]-L-lysine (heavy lysine) and unlabeled L-lysine (light lysine) were purchased from Cambridge Isotope Laboratories (Andover, MA) and Sigma (St. Louis, MO), respectively. Light and heavy EMEM media containing either light or heavy lysine were prepared according to the standard

formulation of EMEM except with the supplement of dialyzed FBS (Invitrogen). The WM-115 cells were grown in light medium, and the metastatic WM-266-4 cells were cultured in heavy medium for at least 6 days to allow the complete incorporation of the isotope-labeled amino acid.

Biotinylation of Cell-surface Membrane Proteins

Cells, which were attached to the culturing flasks, were washed gently with prewarmed PBS (pH 7.4) at 37 ºC three times to remove amine-containing culture media and FBS. A 1-mL solution of 0.5 mg/mL sulfo-NHS-LC-biotin (Pierce, Rockford, IL) in PBS was added to each flask and the cells were incubated at 37 ºC for 10 min. Following biotinylation, 1 mL of 100 mM Tris-HCl (pH 7.4) was added to each flask to quench the reaction by incubation at room temperature for 5 min, and the cells were washed three times with ice-cold PBS to remove excess reagent and byproducts.

Preparation of Cell-Surface Membrane Proteins

After a 10-min labeling reaction, biotinylated WM-115 and WM-266-4 cells from four 75-cm^2 culturing flasks (~1×10^7 cells in total) were harvested into ice-cold PBS by scraping and pelleted by centrifugation at 800 g at 4 ºC for 5 min. The WM-115 and WM-266-4 cells were lysed separately in lysis buffer on ice for 30 min with gentle shaking. The lysis buffer contained 50 mM HEPES (pH 7.4), 2% NP-40 (v/v), 0.5% SDS

(w/v), 100 mM NaCl, 0.5 mM EDTA, 1 mM NaF, 1 mM Na_3VO_4, 1.5 mM $MgCl_2$, 1 mM PMSF and a protease inhibitor cocktail. Cell lysates were centrifuged at 16,000 g and at 4 ºC for 30 min, and the resulting supernatants were collected. The cell lysates were diluted by at least 30-fold,and the protein concentrations of these diluted lysates were determined by using Quick Start™ Bradford Protein Assay (BioRad, Hercules, CA). The WM-115 and WM-266-4 cell lysates were then combined at a ratio of 1/1 (w/w).

To remove large amounts of salts and detergent, proteins were precipitated by adding trichloroacetic acid (TCA) to a final concentration of 20% and incubated on ice for at least 1 h. After incubation, the precipitates were pelleted by centrifugation at 10,000 g for 5 min and washed with an ice-cold solution of 10 mM HCl/90% acetone three times. The resultant protein pellets were dried under vacuum.

Enzymatic Digestion

The TCA precipitates were resuspended in a 100-mM NH_4HCO_3 solution (pH 8.5) containing 0.2% SDS at room temperature for 30 min, and then heated at 95 ºC for 10 min. Urea was then added to the solution to a final concentration of 8 M to improve further the solubility of membrane proteins. These procedures were adapted from previously published procedures for enhancing the solubility of membrane proteins.[19, 35] It is worth noting that heating protein in the presence of urea may lead to the carbamylation of proteins.[36] Nevertheless, database search based on the LC-MS/MS

results revealed the absence of peptides with the N-terminus or the side chain of lysine residues being carbamylated.

The solution was cooled to room temperature, and the proteins were reduced with dithiothreitol (DTT) at a final concentration of 10 mM and alkylated in dark with iodoacetamide (IAA) at a final concentration of 50 mM for 30 min. The resulting protein solutions were diluted 4 times with 100 mM NH_4HCO_3 (pH 8.5) and digested with mass spectrometry-grade trypsin (Promega, Madison, WI) at an enzyme-to-substrate ratio of 1/100 (w/w) at 37 °C overnight. The peptide mixtures were dried in a SpeedVac.

Affinity Purification

Biotinylated peptides were purified and enriched by using an affinity column packed with avidin-agarose resin (Sigma-Aldrich) as previously described.[37] Briefly, the packed avidin-agarose resin was washed with 10 column volumes of binding buffer (20 mM potassium phosphate, 0.15 M NaCl, pH 7.5) prior to loading the tryptic digests. After a 30-min incubation, unbound peptides were removed by washing the resin thoroughly with the binding buffer and pure water. The bound peptides were eluted with an equi-volume mixture of CH_3CN and H_2O at 65 °C. The elution process was repeated, the eluates were combined, dried in a Speedvac, and stored at –20 °C for further analysis.

LC-MS/MS

The affinity-purified peptide mixtures were analyzed by LC-MS/MS in triplicate on an LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with an electrospray ionization (ESI) source. An Agilent 1100 capillary HPLC system with a Zorbax SB-C18 capillary column (5 μm in particle size, 0.5×250 mm, Agilent Technologies, Santa Clara, CA) was employed for peptide separation. The chromatographic separation was performed with a linear gradient of 2-15% CH_3CN in 0.6% aqueous solution of acetic acid in the first 13 min, followed by a 70-min gradient of 15-50% CH_3CN in 0.6% acetic acid; the flow rate was 6 μL/min.

The capillary temperature for the electrospray source was maintained at 275 °C and the source voltage was set at 4 kV. The maximum ion injection time was 50 ms with automatic gain control (AGC) values of 3×10^3, 1×10^3 and 1×10^3 ions for full-scan MS, MS in selected ion monitoring (SIM) mode, and MS/MS, respectively. The product-ion spectra were acquired in a data-dependent scan mode, where the eight most abundant ions observed in MS were chosen for collision-induced dissociation (CID). The normalized collision energy was 35% and the activation Q was 0.25. In addition, the dynamic exclusion feature was activated to maximize the detection of peptide ions of low abundance.

Quantitative Analysis and Data Processing

Raw data of product-ion spectra from the LC-MS/MS analysis were searched against NCBI human RefSeq protein database using TurboSEQUEST in Bioworks 3.2 (Thermo Electron, San Jose, CA). Variable modifications on lysine residues were allowed with a mass increase of 8.01 Da to recognize heavy lysine residues, and increases of 339.14 and 347.15 Da for biotinylated light and heavy lysine residues, respectively. The static carbamidomethylation (CAM) of cysteine residues was used with a mass increase of 57.01 Da. Peptide filters with cross-correlation (X_{corr}) via charge states (≥ 1.9, ≥ 2.2, ≥ 3.5 for peptide ions that are singly, doubly, and triply charged) and delta correlation ($\Delta C_n \geq 0.1$) were used to sort the search results. All the tandem mass spectra of the identified peptides with biotin labeling were manually inspected. The cellular localization and function of each identified protein were further investigated based on NCBI Gene Ontology (GO) annotation and Gene Ontology Annotation (GOA) database (http://www.ebi.ac.uk/GOA).

Protein quantification was based on the biotinylated peptides by comparing manually the relative ion abundances of unlabeled and isotope-labeled peptides with the average of at least three MS scans with a mass tolerance of ± 0.5 Da and an S/N ratio greater than 3. We also quantified some of the identified proteins by monitoring selectively the fragmentations of the precursor ions for the biotinylated peptides carrying either light- or heavy-lysine residue(s). The ratios of the unlabeled and isotope-labeled proteins were determined using the peak areas found in the corresponding selected-ion

chromatograms (SICs) of the monitored peptide pairs. For those ions of low relative abundance, SICs for monitoring the formation of the three most abundant fragment ions in MS/MS were employed for the quantification. All the quantification results were based on LC-MS/MS measurements of samples from three independent mixings of the cell lysates.

Immunocytochemistry and Confocal Microscopy

Four identified proteins, including histone H2B and three membrane proteins, MCAM (melanoma cell adhesion molecule, or CD146), ALCAM (activated leukocyte cell adhesion molecule, or CD166), and ITGA3 (integrin alpha 3, or CD49C), were assessed by immunofluorescence staining using confocal microscopy. To stain the three membrane proteins, WM-115 and WM-266-4 cells grown on coverslips were washed with PBS and fixed with 4% paraformaldehyde in PBS for 12 min. The fixed cells were treated with a blocking buffer (3% BSA in PBS) at room temperature for 30 min. After blocking, cells were stained with primary mouse monoclonal antibodies (mAbs) recognizing human MCAM, ALCAM, and ITGA3, respectively (Abcam, Cambridge, MA). After incubation in a humidified chamber at 37 °C for 1 h, cells were washed with PBS and blocking buffer, and then stained with an Alexa Fluor 488-conjugated rabbit secondary antibody to mouse IgG (Invitrogen) at 37 °C for 30 min. For histone H2B, cells were stained without fixation. In this case, all operations were performed on ice, and ice-cold PBS and blocking buffers were also used. Briefly, cells were first stained with

mouse monoclonal anti-human H2B antibody (MBL International, Woburn, MA) on ice and then with the same secondary antibody described above. Control experiments were also carried out for unfixed WM-115 and WM-266-4 cells stained with the secondary antibody alone. After staining, cells were fixed with 4% paraformaldehyde.

Following washing with PBS, coverslips with the antibody-stained WM-115 and WM-266-4 cells were mounted onto microscope slides with 10 μL of Vectashield mounting medium (Vector Laboratories, Burlingame, CA) containing DAPI (4',6-diamidino-2-phenylindole). Images were collected with a Leica TCS SP2/UV confocal microscope (Heidelberg, Germany) at a magnification factor of 63×. Images of WM-115 and WM-266-4 cells stained with the same antibody were captured with the same parameters to compare the staining of these two types of cells. Relative intensities of fluorescence staining were analyzed using ImageJ (http://rsb.info.nih.gov/ij/).

Flow Cytometric Analysis

After the biotinylation reaction, the WM-115 and WM-266-4 cells were detached from culturing flasks into trypsin-EDTA solution and collected by centrifugation at 800 g at 4 ºC for 5 min. The cells were washed with ice-cold PBS (Ca^{2+}/Mg^{2+}-free, 5 mM EDTA) and binding buffer (Ca^{2+}/Mg^{2+}-free PBS, 2% FBS, 1% sodium azide, 5 mM EDTA, 20 mM HEPES, pH 7.4). Before staining, the number of cells in each sample was determined by using a hemocytometer. An excess amount of streptavidin-R-phycoerythrin conjugate (SAPE, Invitrogen) was added to each cell sample (5×10^5 cells),

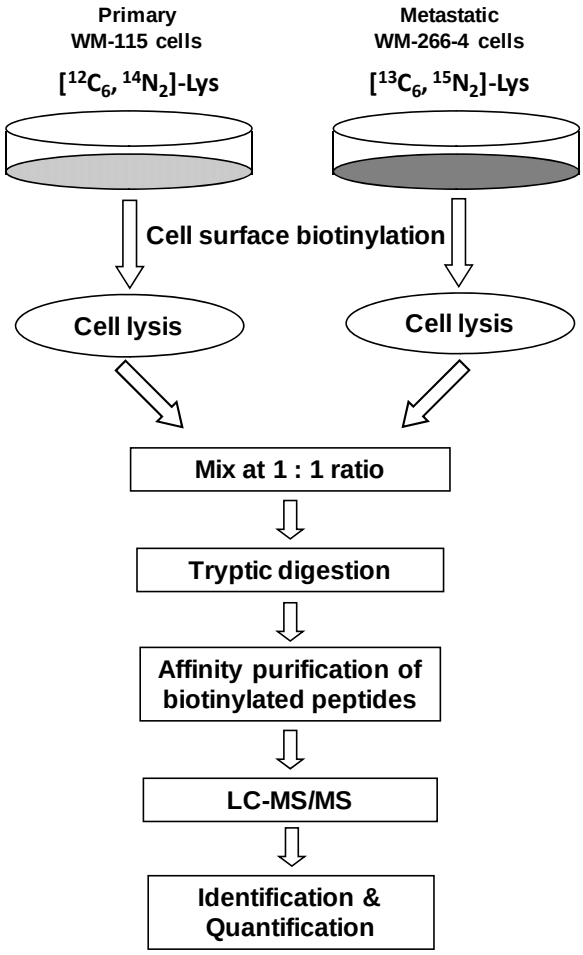
followed by incubation at 4 ºC in the dark for 30 min. The cells were subsequently washed twice with binding buffer and three times with PBS to remove the free SAPE. The cells were then fixed with 2% paraformaldehyde and kept at 4 ºC. Flow cytometric measurements were carried out on a BD FACSAria Cell Sorting System (BD Biosciences) and the data were analyzed using FlowJo (Tree Star, Inc.)

Results and Discussion

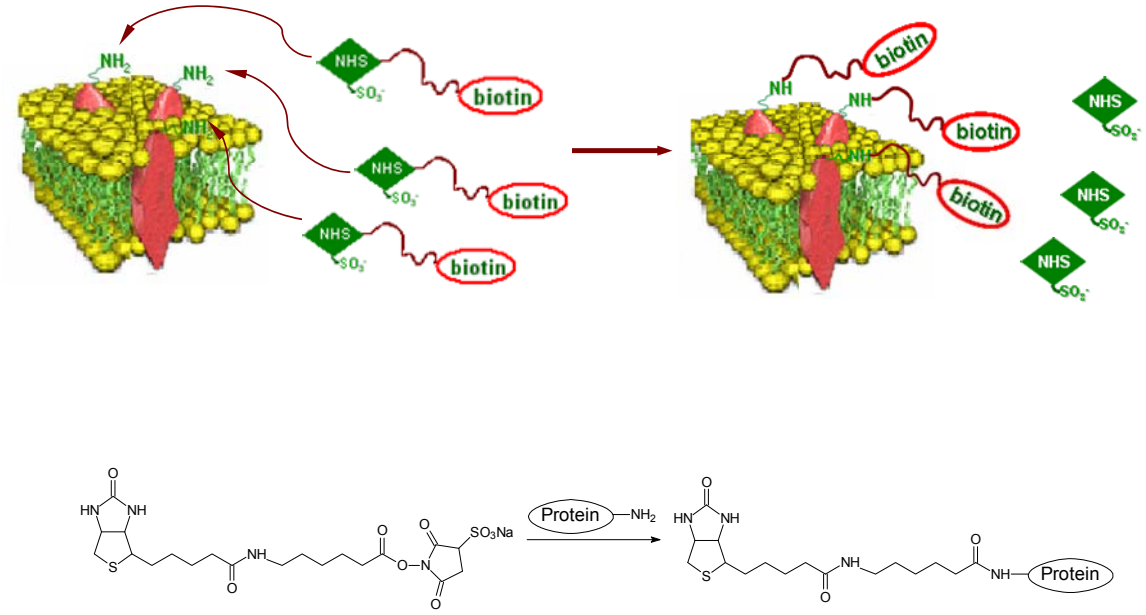
General Strategy for the Comparative Studies of Cell Surface Proteins

Our general strategy encompassed SILAC for the metabolic incorporation of isotope-labeled amino acids, *in situ* biotinylation of surface proteins of intact cells using the membrane-impermeable sulfo-NHS-LC-biotin, cell lysis, enzymatic digestion, affinity purification of biotinylated peptides, and LC-MS/MS for protein identification and quantification (Schemes 4.1 and 4.2). Since its first introduction, SILAC has been widely used for MS-based protein quantification and it has also been employed to quantify membrane and membrane-bound proteins.[29, 38, 39]

The strategy of labeling and selective enrichment of membrane proteins by biotinylation and subsequent affinity purification has been employed to improve the selectivity and throughput for profiling membrane proteins.[11, 12, 14-23] After affinity enrichment and purification, the resultant membrane fractions, however, are still contaminated with cytoplasmic proteins. The contamination could arise from the nonspecific binding of the cytoplasmic proteins to avidin resin, from noncovalent protein-



Scheme 4.1 Strategy for the relative quantification of surface membrane proteins in primary and metastatic human melanoma cells.



Scheme 4.2 Cell surface protein biotinylation with sulfo-NHS-LC-biotin. The chemical reaction between the sulfo-NHS-LC-biotin and amino groups in proteins is also shown.

protein interactions and/or from the biotinylation of proteins in nonviable cells.[18, 24] Washing with buffers at high-salt concentration and high pH during affinity purification could reduce contaminations to some extent. However, cytosolic proteins could still account for up to 15% of all the identified proteins.[11] In addition, the affinity-enriched membrane protein fractions from cell lysates are still a complex mixture. Further separation of the enriched membrane proteins by using 1-D SDS-PAGE, 2-D gel, or HPLC, is usually necessary to reduce the sample complexity prior to enzymatic digestion.[14, 19, 24]

To improve the efficiency of purification and to minimize contamination, we decided to digest the protein mixtures and enrich the biotinylated peptides using biotin/avidin-based affinity purification. After a single step of affinity purification, the biotinylated peptides of membrane proteins are subjected to LC-MS/MS analysis. Because biotinylation occurs only at lysine residues exposed on the cell surface and only the labeled peptides are selectively enriched, the sample complexity is markedly reduced.

A potential drawback for the purification of biotinylated peptides rather than biotinylated proteins is that the protein representation of the peptides is only limited to the biotinylated peptides.[13, 18, 21] However, our previous work,[37] together with the successful implementation of a similar approach for the selective enrichment of ICAT-conjugated cysteine-containing peptides,[26] revealed that protein identification and quantification based on the biotinylated peptides are reliable.

SILAC with [$^{13}C_6$,$^{15}N_2$]-L-lysine will render each lysine residue in the proteins from WM-266-4 cells being replaced with a heavy lysine, which results in a mass increase of 8.01 Da. After biotinylation, the lysine residues are no longer susceptible to trypsin cleavage. Therefore, these biotinylated peptides contain at least one internal lysine residue. As a result, the biotinylated light- and heavy-lysine residues exhibit mass increases of 339.14 and 347.15 Da, respectively. Relative quantification could be readily achieved by using the relative abundances of ions corresponding to the biotinylated light- and heavy-lysine-bearing peptide pairs.

To evaluate the accuracy of quantification based on SILAC and LC-MS/MS analysis of biotinylated peptides, we carried out control experiments by mixing the light and heavy cell lysates from WM-266-4 cells at ratios of 1/1, 3/1 and 10/1 (w/w). The SICs for monitoring the biotinylated peptides from five proteins identified in WM-266-4 cells were plotted, and the resulting ratios for these five proteins recovered from LC-MS analysis are shown in Table 4.1. These results demonstrated that the relative ratios of light and heavy peptides are consistent with the expected ratios (Table 4.1). The ratios of multiple isotopic peptide pairs from the same protein were also in good agreement with each other and with the expected ratios (Figure 4.1). Moreover, we mixed the light and heavy cell lysates at 1/1 (w/w) ratio and analyzed, for three times, the resulting tryptic digestion mixtures by using LC-MS/MS in data-dependent scan mode. The LC-MS/MS results allowed us to quantify 20 proteins, and the average ratio determined by LC-MS/MS was 1.01 with a relative standard deviation of 11% (Table S4.1).

Table 4.1 A list of biotinylated proteins and their peptides identified from WM-266-4 cells used to evaluate the quantification strategy based on cell surface biotinylation and affinity-enriched peptides.

NCBI Accession	Protein Name	Peptide Used for Quantification	Charge	*m/z*	Ratio (L/H) 1:1	Ratio (L/H) 3:1	Ratio (L/H) 10:1
NP_006491	melanoma cell adhesion molecule						
		R.IFLCQGK@R.P	2	684.9	1.12	3.14	10.35
		R.IFLCQGK#R.P	2	680.9			
		R.LPSGNHMK@ESR.E	3	534.9	0.99	3.18	9.91
		R.LPSGNHMK#ESR.E	3	532.3			
NP_001001390	CD44 antigen isoform 3 precursor						
		R.YVQK@GEYR.T	2	695.3	0.75	2.35	8.85
		R.YVQK#GEYR.T	2	691.3			
NP_002194	integrin alpha 2 precursor						
		R.K@YAYSAASGGR.R	2	739.3	0.98	2.95	10.02
		R.K#YAYSAASGGR.R	2	735.3			
NP_001670	Na+/K+ -ATPase beta 3 subunit						
		R.IIGLK@PEGVPR.I	2	763.4	1.04	3.07	10.11
		R.IIGLK#PEGVPR.I	2	759.4			
NP_005492	integrin alpha 3 isoform b precursor						
		R.TGAVYLCPLTAHK@DDCER.M	3	818.4	1.18	2.91	10.67
		R.TGAVYLCPLTAHK#DDCER.M	3	815.7			
			Mean		1.01	2.93	9.99
			SD		0.15	0.30	0.62
			RSD%		14.74	10.38	6.20

Note: # biotinylated light (L) lysine; @ biotinylated heavy (H) lysine; cysteine is CAM-modified.

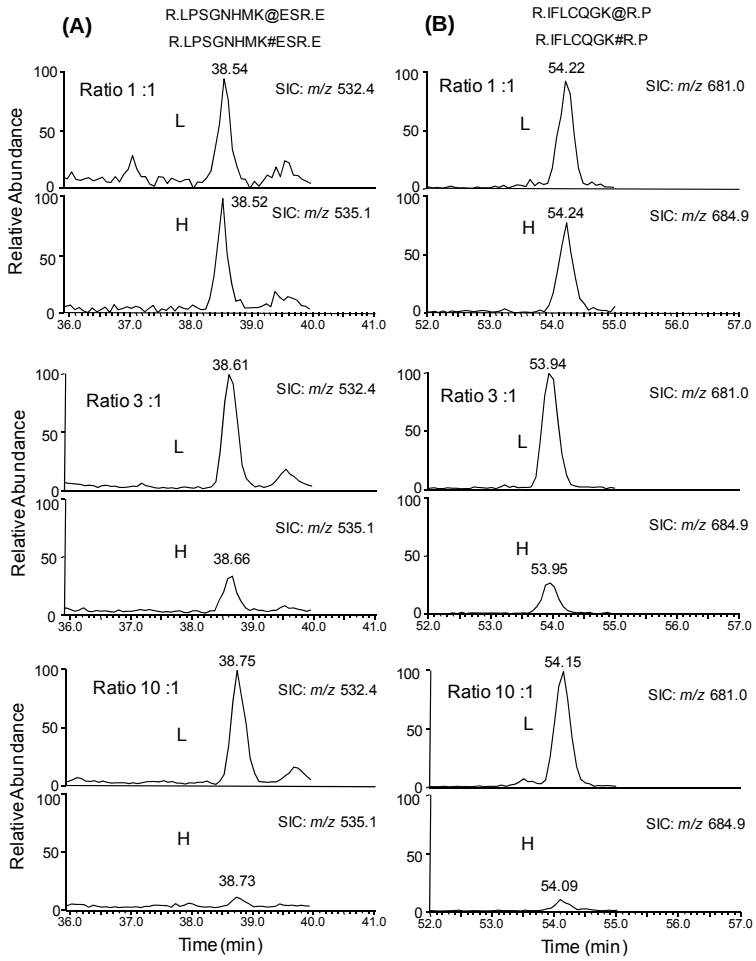


Figure 4.1 Evaluation of the quality of quantification based on SILAC and biotinylated peptides. Selected-ion chromatograms (SICs) of two pairs of light (L) and heavy (H) biotinylated peptides of MCAM from WM-266-4 cells. The expected ratios of peak area of light and heavy peptides are 1:1, 3:1 and 10:1 from top panel to bottom panel. The biotinylated light lysine ($[^{12}C_6,^{14}N_2]$-L-lysine) and heavy lysine ($[^{13}C_6,^{15}N_2]$-L-lysine) are labeled with "#" and "@", respectively. (A) SICs of the $[M+3H]^{3+}$ ions of LPSGNHMK#ESR (*m/z* 532.4) and LPSGNHMK@ESR (*m/z* 535.1); (B) SICs of the $[M+2H]^{2+}$ ions of IFLCQGK#R (*m/z* 681.0) and IFLCQGK@R (*m/z* 684.9), in which the cysteine residue is CAM-modified. The calculated ratios of peak area of these peptides are listed in Table 4.1.

A typical pair of peptides, which were used for quantification based on MS, as well as their product-ion spectra acquired in data-dependant scan mode, are depicted in Figure 4.2. The ratios of isotopic peptide pairs of different charge states (in this case, 2+ and 3+) were also examined, and they gave consistent results (data not shown). On the basis of the results from the above-mentioned control quantification experiments, we concluded that our strategy, which included SILAC, biotinylation reaction, one-step affinity purification of biotinylated peptides and LC-MS/MS analysis, can afford accurate and reliable quantification of protein expression.

It is worth noting that not all identified biotinylated peptides can be employed for quantifying the expression of their corresponding proteins. For instance, some peptides were identified based on their tandem mass spectra, but their molecular ions exhibit relatively low abundances or the backgrounds of the full-scan MS were relatively high. Under such circumstances, we cannot quantify these isotopic peptide pairs based on the intensities of peaks found in ESI-MS. Instead we acquired MS/MS for these peptide ions in the SIM mode, which offered improved selectivity, sensitivity and accuracy for the quantification. However, even with the SIM mode, a small portion of identified peptides, because of their low abundances, still could not be quantified (data not shown).

Proteomic Analysis of Membrane Proteins of Primary and Metastatic Melanoma Cells

To examine the differential expression of surface membrane proteins of melanoma cells at different tumor progression stages, we first labeled the cell-surface proteins of the

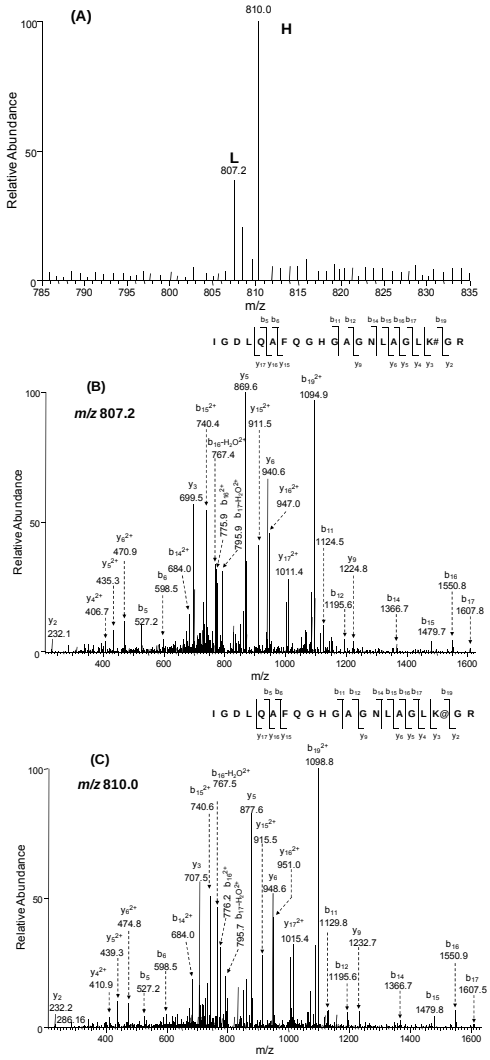


Figure 4.2 A typical isotopic pair of biotinylated peptides from WM-115 (light, "L") and WM-266-4 cells (heavy, "H") used for quantification by ESI-MS. The identities of peptides were confirmed by MS/MS acquired in data-dependant scan mode. (A) ESI mass spectra of an isotopic peptide pair, IGDLQAFQGHGAGNLAGLK#GR ($[M+3H]^{3+}$ ion, *m/z* 807.2) and IGDLQAFQGHGAGNLAGLK@GR ($[M+3H]^{3+}$ ion, *m/z* 810.0) of the solute carrier family 3 protein from WM-115 and WM-266-4 cells; (B) and (C) are the product-ion spectra of the biotinylated peptides carrying a light and heavy lysine, respectively.

WM-115 and WM-266-4 cells, which were cultured in media containing light and heavy lysines, respectively, with sulfo-NHS-LC-biotin *in situ*. After removing the excess labeling reagent, we lysed the cells, mixed the light and heavy cell lysates at a ratio of 1/1 (w/w), digested the proteins in the lysates with trypsin, and subjected the digestion mixture to LC-MS/MS analysis. From the LC-MS/MS results acquired in the data-dependant scan mode, we were able to identify 109 proteins (Figure 4.3). A majority (78%, Figure 4.3) of these proteins were localized on the cell surface and PM, as classified by GO database annotations. These include integral membrane, transmembrane and peripheral membrane proteins, such as a large set of integrins, cell adhesive molecules, CD antigens and receptors. All the identified proteins carry at least one peptide with internal lysine residue(s) tagged with a biotin moiety.

The rest (22%) of the identified proteins were not classified as cell surface and PM proteins by the GO database annotations. However, the membrane localization of the majority of these proteins was reported previously. These proteins contributed to 16.5% of all the identified proteins. For instance, several proteins with chaperone functions, including heat shock proteins (HSPs) and protein disulfide isomerases (PDIs), were identified.

HSPs and PDIs are generally considered to be cytoplasmic or ER proteins and the majority of them do not contain transmembrane domains. Several recent studies, however, revealed the cell membrane localization of these proteins.[15, 20, 21, 40, 41] In this respect, a previous comprehensive profiling of cell surface proteins revealed that HSPs have

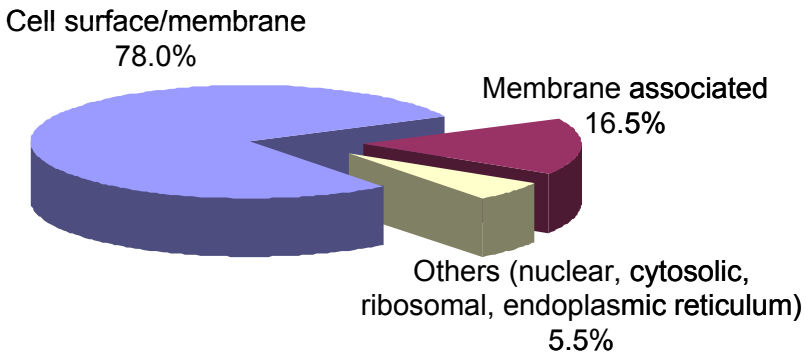


Figure 4.3 Pie chart of subcellular classification of the 109 proteins identified from WM-115 and WM-266-4 cells, based on GO database annotations and previous publications.

relatively high abundance, compared with other cell surface proteins.[15, 20, 21] Little is known about the cellular pathways through which the HSPs reach the cell surface, and it remains unclear what the functions are for these proteins to be present on the cell surface. In this context, it was suggested that the HSPs might be associated with misfolded proteins or peptide fragments and are transferred out of the cytosol via a nonclassic pathway.[20] In addition, the presence of these proteins on the cell surface may help maintain the structural integrity of various receptor complexes.[20]

Other than the chaperone proteins, histones H1, H2A, H2B, H3 and H4 were also found to be present on the surface of melanoma cells in our experiments. Histones are the major protein components of chromatin and play important roles in DNA packaging, chromosome stabilization and gene expression.[42, 43] However, histones were also shown to exist on the cell surface.[44-49]

As discussed above, a majority of the proteins identified were determined to be membrane and membrane-associated proteins by GO database annotations and recent publications. However, we also observed a few peptides arising from nonmembrane-related proteins, including those from cytosol, nucleus, ER and ribosome, which in total accounted for only 5.5% of all the proteins identified. It remains unexplored whether these proteins are also localized on the cell surface or membrane. Nevertheless, compared with the classical membrane enrichment techniques, such as ultracentrifugation and fractionation,[50] the proportion of nonmembrane-related proteins identified in our experiments is relatively small.

Because of the low abundances, some identified proteins were not quantified. In this regard, we were able to quantify the relative expression of 62 proteins in these two melanoma cell lines (Table 4.2). Among the quantified proteins, the majority of them exhibited only slight differences in expression between these two cell lines. Only about one-third of the proteins (22 out of 62, as diagramed in Figure 4.4) are down- or up-regulated in the metastatic melanoma cells, relative to the primary melanoma cells, if the ratio of 2 or above is considered as a significant change. With this criterion, the number of proteins that are up-regulated in WM-266-4 cells is comparable to the number of proteins that are down-regulated. For instance, several proteins, such as ALCAM, CD9, EphA2, erbB-2, histone H4, histone H2B and immunoglobulin superfamily member 4D isoform 1 (IGSF4), are down-regulated. On the other hand, butyrophilin, CD109, major histocompatibility complex class II DR alpha precursor (HLA-DRA), ITGB3, MCAM and several other proteins are up-regulated in WM-266-4 cells.

Immunocytochemistry for the Validation of Surface Localization and Quantification

A potential pitfall of the above quantification strategy lies in that the biotinylation step involves two different types of cells, which may result in different labeling efficiency with sulfo-NHS-LC-biotin, thereby giving rise to the observation of different levels of membrane proteins. Thus, it is important to assess the relative expression of some membrane proteins in these two cell lines by using an independent method.

Table 4.2 A list of quantified proteins and their corresponding ratios of expression in WM-266-4 cells with respect to WM-115 cells.

NCBI Accession	Protein Name	Surface/ Membrane GO[a]	Ratio WM-266-4/115 Mean ± SD (n=3)	WM266-4 Hits[b] (H)[c]	WM115 Hits (L)[d]
Up-regulated proteins					
NP_853509	butyrophilin, subfamily 3, member A1	Y	2.37 ± 0.59	3	3
NP_598000	CD109	Y	3.65 ± 0.04	4	1
NP_000102	down-regulated in adenoma protein	Y	5.97 ± 0.55	2	0
NP_000203	integrin beta chain, beta 3 precursor	Y	2.38 ± 0.07	19	9
NP_061984	major histocompatibility complex, class II, DR alpha precursor	Y	>5	2	0
NP_004986	matrix metalloproteinase 14 preproprotein	Y	2.18 ± 0.48	6	0
NP_006491	melanoma cell adhesion molecule	Y	3.13 ± 0.69	19	6
NP_005920	melanoma-associated antigen p97 isoform 1, precursor	Y	3.60 ± 1.05	2	0
XP_933500	PREDICTED: similar to immunoglobulin superfamily, member 3 isoform 1	Y	2.62 ± 0.10	2	0
NP_005496	scavenger receptor class B, member 1 isoform 1	Y	3.25 ± 0.35	6	0
NP_001143	solute carrier family 25, member 5	Y	>20	3	0
NP_001012679	solute carrier family 3	Y	2.53 ± 0.18	9	6
Down-regulated proteins					
NP_001618	activated leukocyte cell adhesion molecule	Y	0.07 ± 0.03	0	15

NP_001760	CD9 antigen	Y	0.31 ± 0.02	1	4
NP_004422	ephrin receptor EphA2	Y	0.36 ± 0.06	0	9
NP_001005862	erbB-2 isoform b	Y	0.29 ± 0.07	0	3
NP_003536	H4 histone family, member J	R	0.45 ± 0.26	19	20
NP_001019770	histone 2, H2bf	R	0.38 ± 0.19	1	9
NP_055148	immunoglobulin superfamily, member 4D isoform 1	Y	0.47 ± 0.11	3	4
NP_005388	podocalyxin-like precursor isoform 2	Y	<0.05	0	2
NP_543006	signal-regulatory protein beta 2 isoform 2	Y	0.29 ± 0.07	1	2
NP_003371	vimentin	R	0.43 ± 0.03	1	4

Note: [a] GO database annotations, Y, cell surface/membrane protein; R, membrane-associated protein based on previous publications (see Supporting Information for details). [b] Hits: total number of peptides (including same peptides with multiple identification and different *m/z*) identified with SEQUEST search based on LC-MS/MS data acquired in data-dependant scan mode. Although no peptides were identified for some of the listed proteins from LC-MS/MS data acquired in the data-dependent scan mode, biotinylated peptide(s) was (were) identified from LC-MS/MS results obtained in the SIM mode. [c] H, heavy. [d] L, light.

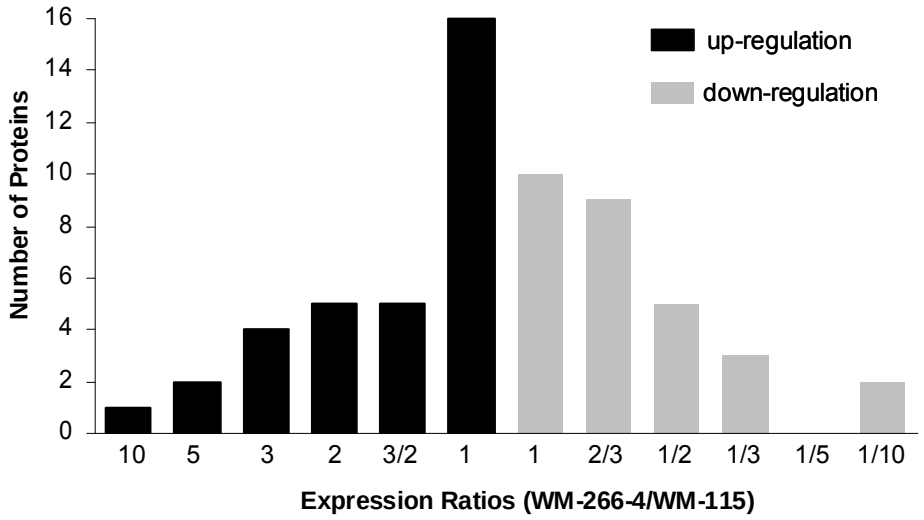


Figure 4.4 A summary of the quantified proteins, which were up-(black bar) and down-(gray bar) regulated in WM-266-4 with respect to WM-115 cells.

To evaluate the reliability of the above MS-based quantification results and to confirm the surface localization of the identified membrane proteins, we carried out immunocytochemistry experiments for four identified proteins, including H2B, MCAM, ALCAM, and ITGA3. As depicted in Figure 4.5, the confocal microscopic images confirmed the cell surface localization of these proteins (The unfixed WM-115 and WM-266-4 cells were also stained with the secondary antibody alone to serve as negative control, see Figure S4.1). Particularly, the expression of histone H2B, an abundant nuclear protein, was detected on the surface of most WM-115 cells with a punctuate distribution (Figure 4.5A, left panel). However, H2B showed a lower level of expression on the surface of most WM-266-4 cells (Figure 4.5A, right panel). It is worth noting that, to visualize the cell-surface distribution of histone H2B, we stained the WM-115 and WM-266-4 cells without fixation so that the integrity of PM can be maintained.

To date, the mechanism by which the histone proteins are expressed on PM has not been elucidated. A recent study showed that histone proteins are able to transverse directly PMs and mediate penetration of macromolecules covalently attached to them.[47] This process was believed to occur by direct translocation and not by typical endocytic pathways. It has also been shown that all major histone proteins, including H2A, H2B, H3, H4 and H1, could be located on the PM of activated T-cells.[45, 46] All these cell-surface histones are in the form of nucleosomes, which appear to be anchored to the cell surface through an interaction with sulfated polysaccharides; however, the function of the cell surface nucleosomes remains unclear. Considering the identification of almost all

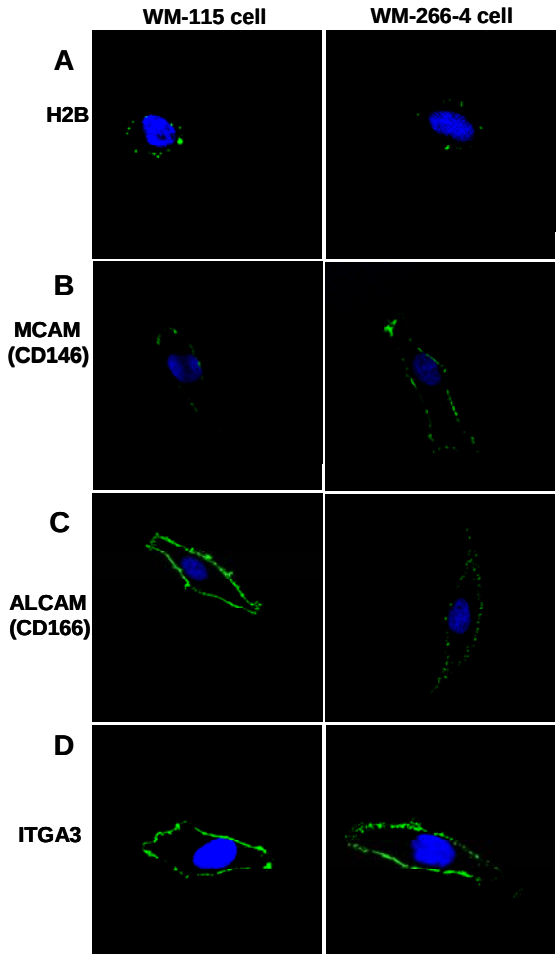


Figure 4.5 Immunocytochemical detection of cell surface proteins. The left and right panels represent the staining of WM-115 and WM-266-4 cells, respectively. Cells were stained with primary mAbs to (A) H2B, (B) MCAM, (C) ALCAM and (D) ITAG3 followed by staining with Alexa Fluor 488-labeled secondary antibody (green). Cells were also stained with DAPI (blue) to identify nuclei. Images of WM-115 and WM-266-4 cells stained with same primary mAbs were captured with identical parameters to compare the staining of the two cell lines.

major histones and their nonuniform distribution on the cell membrane (Figure 4.5A), our MS and immunofluorescence results are in line with a previous report.[45]

We next compared the relative intensities of immunofluorescence for these two cell lines and correlated these findings with the MS-based quantitative data (Table 4.3). The ratios calculated from the immunofluorescence measurements are generally in keeping with the quantification results obtained from the MS-based analysis. Therefore, wse may conclude that the WM-115 and WM-266-4 cells exhibit similar efficiency in the labeling of the amino group of surface proteins with the sulfo-NHS-LC-biotin.

To gain further insights into the labeling of the two types of cells, we stained the biotin-labeled cells with strepta-vidin-R-phycoerythrin and followed the labeling reaction by flow cytometric analysis. While it is difficult to calculate the labeling efficiency based on this assay, the results showed that the two cell lines exhibit similar kinetics in the labeling reaction (Figure S4.2). In this respect, we normalized the fluorescent intensities found at other time intervals against the highest fluorescent intensity observed. It is worth noting that the fluorescent intensity does not reach a plateau at 10 min; we, however, chose to label the cells for a relatively short period of time (i.e., 10 min) to minimize the potential effects of biotinylation on damaging the integrity of plasma membrane.

Table 4.3 Comparison of quantitative results from immunofluorescence staining and MS-based proteomics.

Protein Name	*Immunofluorescence* Ratio (WM266-4/115) Mean ± SD (n=5)	*MS* Ratio (WM266-4/115) Mean ± SD (n=3)
H2B	0.29 ± 0.06	0.38 ± 0.19
MCAM	5.08 ± 1.23	3.13 ± 0.69
ALCAM	0.04 ± 0.01	0.07 ± 0.03
ITGA3	1.19 ± 0.35	0.96 ± 0.11

Characterization of the Differential Expression of Cell Surface Proteins

The identification and characterization of the differential expression of cell surface proteins are potentially useful for screening candidates for biomarkers. Membrane proteins account for more than two-thirds of protein-based drug targets.[2] The above MS-based quantitative analysis facilitates the analysis of differential protein expression at large scale and holds great potential for screening purpose. In addition, understanding the difference of primary and metastatic cancers at proteomic level may potentially benefit cancer treatment. Next we will discuss the membrane and membrane-associated proteins that we quantified in this study.

CD (Cluster of Differentiation) molecules are protein or carbohydrate antigens recognized by specific antibodies, which are used to distinguish particular cell types. CD molecules often serve as receptors or ligands that are important to a cell. For example, MCAM (CD146), previously known as MUC18 or Mel-CAM, is an integral membrane glycoprotein functioning as Ca^{2+}-independent cell adhesive molecule involved in heterophilic cell-cell interactions.[51] Although an earlier study found no direct correlation between its expression and metastatic behavior of melanoma,[52] an increasing body of work has shed light on the functions of MCAM, particularly about its role in tumor growth and the progression of human malignant melanoma.[51, 53-55] Our MS-based quantitative proteomic data and immunostaining results (Figure 4.5B) consistently demonstrated the overexpression of MCAM in the metastatic over the primary melanoma

cells. These results support that MCAM may serve as a diagnostic marker for monitoring the progression of melanoma.

ALCAM (CD166) was found to be significantly down-regulated in WM-266-4 cells. ALCAM is a type-I transmembrane glycoprotein of the immunoglobin superfamily.[56, 57] It has been well-documented that ALCAM is the ligand for the lymphocyte cell-surface receptor CD6.[58] Several reports have claimed that ALCAM is associated with cancer development.[59-65] However, the expression pattern of ALCAM does not exhibit simple correlation with tumor development.[60, 62, 64] Although up-regulation of ALCAM was found in metastatic melanoma,[64] our MS data revealed a significant down-regulation of ALCAM in WM-266-4 cells, which was also confirmed by the immunofluorecence measurements (Figure 4.5C). Other CD molecules, such as lymphocyte function-associated antigen 3 (CD58), CD59 antigen p18-20 and CD9, were expressed at a slightly lower level in WM-266-4 cells (Table 4.2 and Table S4.2).

The CD9 antigen was described originally as a surface protein of non-T acute lymphoblastic leukemia cells and developing B-lymphocytes. [66] It plays a pivotal role in regulating cell adhesion, motility, and proliferation, and has been considered as an important inhibitory factor for metastasis of various human cancers. It has been reported that CD9 is expressed predominantly on primary rather than on metastatic tumors, and the down-regulation of CD9 in metastatic melanoma might have prognostic significance with respect to the potential for metastasis.[67] Recent studies showed that CD9 may reduce the metastatic spread of human small-cell lung carcinoma via the inhibition of cell

proliferation and motility,[68] and that it could also be a potential cell membrane marker in human breast cancer cell lines.[69]

In this study, several integrins, such as ITGAV (integrin alpha V or CD51) and ITGB3 (integrin beta chain beta 3, or CD61), were shown to be up-regulated in metastatic melanoma cells. In this context, the expression of these integrins (CD51/CD61) was found to be associated with tumor progression and metastasis formation in human malignant melanoma.[70, 71] They are well-recognized as molecular markers correlating with the change from radial to vertical growth in primary melanoma.[70] Changes in the expression or function of these cell adhesion molecules can therefore contribute to tumor progression by altering the adhesion status of cells and modulating cell signaling pathways.[72, 73]

We were also able to quantify several proteins that have potential roles in a variety of cell signaling pathways, i.e., erbB-2 (HER2), ephrin receptor (EPHA2), and signal-regulatory protein beta 2 (SIRPG). In this respect, erbB-2, a transmembrane tyrosine kinase receptor, is known to be associated with aggressive breast cancers, making this receptor an attractive therapeutic target.[74] However, conflicting results have been published regarding the expression pattern of erbB-2 in primary and metastatic melanoma.[75-78] Our proteomic data supported the down-regulation of this receptor in WM-266-4 cells. Other molecules, such as melanoma-associated antigen p97, scavenger receptor and several solute carrier family proteins, also displayed different expression

between these two cell lines. Further assessment of the roles of these proteins in melanoma progression is necessary.

Conclusions

We demonstrated that combining SILAC, cell surface biotinylation, affinity peptide purification and LC-MS/MS is very useful for the large-scale quantitative analysis of surface membrane proteins. Contaminations from cytoplasm and other nonmembrane related sources were greatly reduced by using *in situ* cell surface biotinylation and the subsequent affinity purification of biotinylated peptides; therefore, high selectivity and specificity were achieved for the identification of membrane proteins. Additionally, metabolic labeling with SILAC, together with LC-MS/MS, facilitates the reliable quantification of surface membrane proteins expressed on two human melanoma cell lines derived from the same individual at different stages of tumor development. Integrins, cell adhesive molecules, CD antigens and receptors, which are essential for tumor development, were quantified in our study. Other proteins, such as histones, were also detected on the surface of WM-115 and WM-266-4 cells, and the cell surface localization of histone H2B was further confirmed by immunocytochemistry. The identification of aberrantly expressed proteins on the surface of metastatic melanoma cells sets the stage for the future investigation on the implications of these proteins in the progression of human melanoma. Moreover, the application to quantitative analysis of membrane proteins of primary and metastatic melanoma cells using this strategy has demonstrated

its great potential in the comprehensive identification of tumor progression biomarkers in pathology and in the discovery of new protein-based drug targets.

References

1. Wallin, E.; von Heijne, G., Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **1998,** 7, 1029-1038.

2. Hopkins, A. L.; Groom, C. R., The druggable genome. *Nat. Rev. Drug Discov.* **2002,** 1, 727-730.

3. Santoni, V.; Molloy, M.; Rabilloud, T., Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* **2000,** 21, 1054-1070.

4. Wu, C. C.; Yates, J. R., The application of mass spectrometry to membrane proteomics. *Nat. Biotechnol.* **2003,** 21, 262-267.

5. Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* **2003,** 21, 532-538.

6. Xiang, R.; Shi, Y.; Dillon, D. A.; Negin, B.; Horvath, C.; Wilkins, J. A., 2D LC/MS analysis of membrane proteins from breast cancer cell lines MCF7 and BT474. *J. Proteome Res.* **2004,** 3, 1278-1283.

7. Fischer, F.; Wolters, D.; Rogner, M.; Poetsch, A., Toward the complete membrane proteome - High coverage of integral membrane proteins through transmembrane peptide detection. *Mol. Cell. Proteomics* **2006,** 5, 444-453.

8. Lohaus, C.; Nolte, A.; Bluggel, M.; Scheer, C.; Klose, J.; Gobom, J.; Schuller, A.; Wiebringhaus, T.; Meyer, H. E.; Marcus, K., Multidimensional chromatography: A powerful tool for the analysis of membrane proteins in mouse brain. *J. Proteome Res.* **2007,** 6, 105-113.

9. Speers, A. E.; Wu, C. C., Proteomics of integral membrane proteins-theory and application. *Chem. Rev.* **2007,** 107, 3687-3714.

10. Pasquali, C.; Fialka, I.; Huber, L. A., Subcellular fractionation, electromigration analysis and mapping of organelles. *J. Chromatogr. B.* **1999,** 722, 89-102.

11. Zhao, Y. X.; Zhang, W.; Kho, Y. J.; Zhao, Y. M., Proteomic analysis of integral plasma membrane proteins. *Anal. Chem.* **2004,** 76, 1817-1823.

12. Goshe, M. B.; Blonder, J.; Smith, R. D., Affinity labeling of highly hydrophobic integral membrane proteins for proteome-wide analysis. *J. Proteome Res.* **2003,** 2, 153-161.

13. Chen, W. N. U.; Yu, L. R.; Strittmatter, E. F.; Thrall, B. D.; Camp, D. G.; Smith, R. D., Detection of in situ labeled cell surface proteins by mass spectrometry: Application to the membrane subproteome of human mammary epithelial cells. *Proteomics* **2003,** 3, 1647-1651.

14. Hastie, C.; Saxton, M.; Akpan, A.; Cramer, R.; Masters, J. R.; Naaby-Hansen, S., Combined affinity labelling and mass spectrometry analysis of differential cell

surface protein expression in normal and prostate cancer cells. *Oncogene* **2005,** 24, 5905-5913.

15. Jang, J. H.; Hanash, S., Profiling of the cell surface proteome. *Proteomics* **2003,** 3, 1947-1954.

16. Lee, S. J.; Kim, K. H.; Park, J. S.; Jung, J. W.; Kim, Y. H.; Kim, S. K.; Kim, W. S.; Goh, H. G.; Kim, S. H.; Yoo, J. S.; Kim, D. W.; Kim, K. P., Comparative analysis of cell surface proteins in chronic and acute leukemia cell lines. *Biochem. Biophys. Res. Commun.* **2007,** 357, 620-626.

17. Mojica, W.; Sun, J. L., Plasma membrane protein identification by cell surface biotinylation and tandem mass spectrometry of clinical specimens. *Mod. Pathol.* **2006,** 19, 162-162.

18. Nunomura, K.; Nagano, K.; Itagaki, C.; Taoka, M.; Okamura, N.; Yamauchi, Y.; Sugano, S.; Takahashi, N.; Izumi, T.; Isobe, T., Cell surface labeling and mass spectrometry reveal diversity of cell surface markers and signaling molecules expressed in undifferentiated mouse embryonic stem cells. *Mol. Cell. Proteomics* **2005,** 4, 1968-1976.

19. Scheurer, S. B.; Rybak, J. N.; Roesli, C.; Brunisholz, R. A.; Potthast, F.; Schlapbach, R.; Neri, D.; Elia, G., Identification and relative quantification of membrane proteins by surface biotinylation and two-dimensional peptide mapping. *Proteomics* **2005,** 5, 2718-2728.

20. Shin, B. K.; Wang, H.; Yim, A. M.; Le Naour, F.; Brichory, F.; Jang, J. H.; Zhao, R.; Puravs, E.; Tra, J.; Michael, C. W.; Misek, D. E.; Hanash, S. M., Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function. *J. Biol. Chem.* **2003,** 278, 7607-7616.

21. Sostaric, E.; Georgiou, A. S.; Wong, C. H.; Watson, P. F.; Holt, W. V.; Fazeli, A., Global profiling of surface plasma membrane proteome of oviductal epithelial cells. *J. Proteome Res.* **2006,** 5, 3029-3037.

22. Tang, X. T.; Yi, W.; Munske, G. R.; Adhikari, D. P.; Zakharova, N. L.; Bruce, J. E., Profiling the membrane proteome of *Shewanella oneidensis* MR-1 with new affinity labeling probes. *J. Proteome Res.* **2007,** 6, 724-734.

23. Zhang, W.; Zhou, G.; Zhao, Y. X.; White, M. A.; Zhao, Y. M., Affinity enrichment of plasma membrane for proteomics analysis. *Electrophoresis* **2003,** 24, 2855-2863.

24. Zhao, Y. X.; Zhang, W.; White, M. A.; Zhao, Y. M., Capillary high-performance liquid chromatography/mass spectrometric analysis of proteins from affinity-purified plasma membrane. *Anal. Chem.* **2003,** 75, 3751-3757.

25. Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **2005,** 1, 252-262.

26. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999,** 17, 994-999.

27. Smolka, M. B.; Zhou, H. L.; Purkayastha, S.; Aebersold, R., Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal. Biochem.* **2001,** 297, 25-31.

28. Ramus, C.; de Peredo, A. G.; Dahout, C.; Gallagher, M.; Garin, J., An optimized strategy for ICAT quantification of membrane proteins. *Mol. Cell. Proteomics* **2006,** 5, 68-78.

29. Liang, X. Q.; Zhao, J.; Hajivandi, M.; Wu, R.; Tao, J.; Amshey, J. W.; Pope, R. M., Quantification of membrane and membrane-bound proteins in normal and malignant breast cancer cells isolated from the same patient with primary breast carcinoma. *J. Proteome Res.* **2006,** 5, 2632-2641.

30. Shi, Y.; Xiang, R.; Horvath, C.; Wilkins, J. A., Quantitative analysis of membrane proteins from breast cancer cell lines BT474 and MCF7 using multistep solid phase mass tagging and 2D LC/MS. *J. Proteome Res.* **2005,** 4, 1427-1433.

31. Foster, L. J.; Zeemann, P. A.; Li, C.; Mann, M.; Jensen, O. N.; Kassem, M., Differential expression profiling of membrane proteins by quantitative proteomics in a human mesenchymal stem cell line undergoing osteoblast differentiation. *Stem Cells* **2005,** 23, 1367-1377.

32. Stockwin, L. H.; Blonder, J.; Bumke, M. A.; Lucas, D. A.; Chan, K. C.; Conrads, T. P.; Issaq, H. J.; Veenstra, T. D.; Newton, D. L.; Rybak, S. M., Proteomic analysis of

plasma membrane from hypoxia-adapted malignant melanoma. *J. Proteome Res.* **2006,** 5, 2996-3007.

33. Bisle, B.; Schmidt, A.; Scheibe, B.; Klein, C.; Tebbe, A.; Kellermann, J.; Siedler, F.; Pfeiffer, F.; Lottspeich, F.; Oesterhelt, D., Quantitative profiling of the membrane proteome in a halophilic archaeon. *Mol. Cell. Proteomics* **2006,** 5, 1543-1558.

34. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002,** 1, 376-386.

35. Simpson, R. J., *Proteins and proteomics: a laboratory manual*. Cold Spring Harbor Laboratory Press: 2003.

36. McCarthy, J.; Hopwood, F.; Oxley, D.; Laver, M.; Castagna, A.; Righetti, P. G.; Williams, K.; Herbert, B., Carbamylation of proteins in 2-D electrophoresis - Myth or reality? *J. Proteome Res.* **2003,** 2, 239-242.

37. Qiu, H.; Wang, Y., Probing adenosine nucleotide-binding proteins with an affinity-labeled nucleotide probe and mass spectrometry. *Anal. Chem.* **2007,** 79, 5547-5556.

38. Loyet, K. M.; Ouyang, W. J.; Eaton, D. L.; Stults, J. T., Proteomic profiling of surface proteins on Th1 and Th2 cells. *J. Proteome Res.* **2005,** 4, 400-409.

39. Zhang, G.; Spellman, D. S.; Skolnik, E. Y.; Neubert, T. A., Quantitative phosphotyrosine proteomics of EphB2 signaling by stable isotope labeling with amino acids in cell culture (SILAC). *J. Proteome Res.* **2006,** 5, 581-588.

40. Multhoff, G.; Hightower, L. E., Cell surface expression of heat shock proteins and the immune response. *Cell Stress Chap.* **1996,** 1, 167-176.

41. Botzler, C.; Issels, R.; Multhoff, G., Heat-shock protein 72 cell-surface expression on human lung carcinoma cells in associated with an increased sensitivity to lysis mediated by adherent natural killer cells. *Cancer Immunol. Immunother.* **1996,** 43, 226-230.

42. Van Holde, K. E., *Chromatin*. New York: Springer-Verlag.: 1989.

43. Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **1997,** 389, 251-260.

44. Ojcius, D. M.; Muller, S.; Hasselkus-Light, C. S.; Young, J. D.; Jiang, S., Plasma membrane-associated proteins with the ability to partially inhibit perforin-mediated lysis. *Immunol. Lett.* **1991,** 28, 101-108.

45. Watson, K.; Gooderham, N. J.; Davies, D. S.; Edwards, R. J., Nucleosomes bind to cell surface proteoglycans. *J. Biol. Chem.* **1999,** 274, 21707-21713.

46. Watson, K.; Edwards, R. J.; Shaunak, S.; Parmelee, D. C.; Sarraf, C.; Gooderham, N. J.; Davies, D. S., Extra-nuclear location of histones in activated human

peripheral blood lymphocytes and cultured T-cells. *Biochem. Pharmacol.* **1995,** 50, 299-309.

47. Hariton-Gazal, E.; Rosenbluh, J.; Graessmann, A.; Gilon, C.; Loyter, A., Direct translocation of histone molecules across cell membranes. *J. Cell Sci.* **2003,** 116, 4577-4586.

48. Herren, T.; Burke, T. A.; Das, R.; Plow, E. F., Identification of histone H2B as a regulated plasminogen receptor. *Biochemistry* **2006,** 45, 9463-9474.

49. Parseghian, M. H.; Luhrs, K. A., Beyond the walls of the nucleus: the role of histones in cellular signaling and innate immunity. *Biochem. Cell Biol.* **2006,** 84, 589-604.

50. Ruth, M. C.; Old, W. M.; Emrick, M. A.; Meyer-Arendt, K.; Aveline-Wolf, L. D.; Pierce, K. G.; Mendoza, A. M.; Sevinsky, J. R.; Hamady, M.; Knight, R. D.; Resing, K. A.; Ahn, N. G., Analysis of membrane proteins from human chronic myelogenous leukemia cells: comparison of extraction methods for multidimensional LC-MS/MS. *J. Proteome Res.* **2006,** 5, 709-719.

51. Shih, I. M., The role of CD146 (Mel-CAM) in biology and pathology. *J. Pathol.* **1999,** 189, 4-11.

52. Denton, K. J.; Stretch, J. R.; Gatter, K. C.; Harris, A. L., A study of adhesion molecules as markers of progression in malignant melanoma. *J. Pathol.* **1992,** 167, 187-191.

53. Luca, M.; Hunt, B.; Bucana, C. D.; Johnson, J. P.; Fidler, I. J.; Bar-Eli, M., Direct correlation between MUC18 expression and metastatic potential of human melanoma cells. *Melanoma Res.* **1993,** 3, 35-41.

54. Xie, S.; Luca, M.; Huang, S.; Gutman, M.; Reich, R.; Johnson, J. P.; Bar-Eli, M., Expression of MCAM/MUC18 by human melanoma cells leads to increased tumor growth and metastasis. *Cancer Res.* **1997,** 57, 2295-2303.

55. Leslie, M. C.; Zhao, Y. J.; Lachman, L. B.; Hwu, P.; Wu, G. J.; Bar-Eli, M., Immunization against MUC18/MCAM, a novel antigen that drives melanoma invasion and metastasis. *Gene Ther.* **2007,** 14, 316-323.

56. Abidi, S. M.; Saifullah, M. K.; Zafiropulos, M. D.; Kaput, C.; Bowen, M. A.; Cotton, C.; Singer, N. G., CD166 expression, characterization, and localization in salivary epithelium: implications for function during sialoadenitis. *J. Clin. Immunol.* **2006,** 26, 12-21.

57. Bowen, M. A.; Bajorath, J.; D'Egidio, M.; Whitney, G. S.; Palmer, D.; Kobarg, J.; Starling, G. C.; Siadak, A. W.; Aruffo, A., Characterization of mouse ALCAM (CD166): the CD6-binding domain is conserved in different homologs and mediates cross-species binding. *Eur. J. Immunol.* **1997,** 27, 1469-1478.

58. Aruffo, A.; Bowen, M. A.; Patel, D. D.; Haynes, B. F.; Starling, G. C.; Gebe, J. A.; Bajorath, J., CD6-ligand interactions: a paradigm for SRCR domain function? *Immunol. Today* **1997,** 18, 498-504.

59. Burkhardt, M.; Mayordomo, E.; Winzer, K. J.; Fritzsche, F.; Gansukh, T.; Pahl, S.; Weichert, W.; Denkert, C.; Guski, H.; Dietel, M.; Kristiansen, G., Cytoplasmic overexpression of ALCAM is prognostic of disease progression in breast cancer. *J. Clin. Pathol.* **2006,** 59, 403-409.

60. Jezierska, A.; Matysiak, W.; Motyl, T., ALCAM/CD166 protects breast cancer cells against apoptosis and autophagy. *Med. Sci. Monit.* **2006,** 12, 263-273.

61. Klein, W. M.; Wu, B. P.; Zhao, S.; Wu, H.; Klein-Szanto, A. J.; Tahan, S. R., Increased expression of stem cell markers in malignant melanoma. *Mod. Pathol.* **2007,** 20, 102-107.

62. Kristiansen, G.; Pilarsky, C.; Wissmann, C.; Stephan, C.; Weissbach, L.; Loy, V.; Loening, S.; Dietel, M.; Rosenthal, A., ALCAM/CD166 is up-regulated in low-grade prostate cancer and progressively lost in high-grade lesions. *Prostate* **2003,** 54, 34-43.

63. Lunter, P. C.; van Kilsdonk, J. W.; van Beek, H.; Cornelissen, I. M.; Bergers, M.; Willems, P. H.; van Muijen, G. N.; Swart, G. W., Activated leukocyte cell adhesion molecule (ALCAM/CD166/MEMD), a novel actor in invasive growth, controls matrix metalloproteinase activity. *Cancer Res.* **2005,** 65, 8801-8808.

64. van Kempen, L. C.; van den Oord, J. J.; van Muijen, G. N.; Weidle, U. H.; Bloemers, H. P.; Swart, G. W., Activated leukocyte cell adhesion molecule/CD166,

a marker of tumor progression in primary malignant melanoma of the skin. *Am. J. Pathol.* **2000,** 156, 769-774.

65. Degen, W. G.; van Kempen, L. C.; Gijzen, E. G.; van Groningen, J. J.; van Kooyk, Y.; Bloemers, H. P.; Swart, G. W., MEMD, a new cell adhesion molecule in metastasizing human melanoma cell lines, is identical to ALCAM (activated leukocyte cell adhesion molecule). *Am. J. Pathol.* **1998,** 152, 805-813.

66. Kersey, J. H.; LeBien, T. W.; Abramson, C. S.; Newman, R.; Sutherland, R.; Greaves, M., P-24: a human leukemia-associated and lymphohemopoietic progenitor cell surface structure identified with monoclonal antibody. *J. Exp. Med.* **1981,** 153, 726-731.

67. Si, Z.; Hersey, P., Expression of the neuroglandular antigen and analogues in melanoma. CD9 expression appears inversely related to metastatic potential of melanoma. *Int. J. Cancer* **1993,** 54, 37-43.

68. Zheng, R.; Yano, S.; Zhang, H.; Nakataki, E.; Tachibana, I.; Kawase, I.; Hayashi, S.; Sone, S., CD9 overexpression suppressed the liver metastasis and malignant ascites via inhibition of proliferation and motility of small-cell lung cancer cells in NK cell-depleted SCID mice. *Oncol. Res.* **2005,** 15, 365-372.

69. Huang, H.; Groth, J.; Sossey-Alaoui, K.; Hawthorn, L.; Beall, S.; Geradts, J., Aberrant expression of novel and previously described cell membrane markers in human breast cancer cell lines and tumors. *Clin. Cancer Res.* **2005,** 11, 4357-4364.

70. Johnson, J. P., Cell adhesion molecules in the development and progression of malignant melanoma. *Cancer Metastasis Rev.* **1999,** 18, 345-357.

71. Saalbach, A.; Wetzel, A.; Haustein, U. F.; Sticherling, M.; Simon, J. C.; Anderegg, U., Interaction of human Thy-1 (CD 90) with the integrin alphavbeta3 (CD51/CD61): an important mechanism mediating melanoma cell adhesion to activated endothelium. *Oncogene* **2005,** 24, 4710-4720.

72. Cavallaro, U.; Christofori, G., Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nat. Rev. Cancer* **2004,** 4, 118-132.

73. Kuphal, S.; Bauer, R.; Bosserhoff, A. K., Integrin signaling in malignant melanoma. *Cancer Metastasis Rev.* **2005,** 24, 195-222.

74. Vogel, C. L.; Cobleigh, M. A.; Tripathy, D.; Gutheil, J. C.; Harris, L. N.; Fehrenbacher, L.; Slamon, D. J.; Murphy, M.; Novotny, W. F.; Burchmore, M.; Shak, S.; Stewart, S. J.; Press, M., Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J. Clin. Oncol.* **2002,** 20, 719-726.

75. Natali, P. G.; Nicotra, M. R.; Digiesi, G.; Cavaliere, R.; Bigotti, A.; Trizio, D.; Segatto, O., Expression of gp185HER-2 in human cutaneous melanoma: implications for experimental immunotherapeutics. *Int. J. Cancer.* **1994,** 56, 341-346.

76. Persons, D. L.; Arber, D. A.; Sosman, J. A.; Borelli, K. A.; Slovak, M. L., Amplification and overexpression of HER-2/neu are uncommon in advanced stage melanoma. *Anticancer Res.* **2000,** 20, 1965-1968.

77. Eliopoulos, P.; Mohammed, M. Q.; Henry, K.; Retsas, S., Overexpression of HER-2 in thick melanoma. *Melanoma Res.* **2002,** 12, 139-145.

78. Kluger, H. M.; DiVito, K.; Berger, A. J.; Halaban, R.; Ariyan, S.; Camp, R. L.; Rimm, D. L., Her2/neu is not a commonly expressed therapeutic target in melanoma -- a large cohort tissue microarray study. *Melanoma Res.* **2004,** 14, 207-210.

Supporting Information for Chapter 4

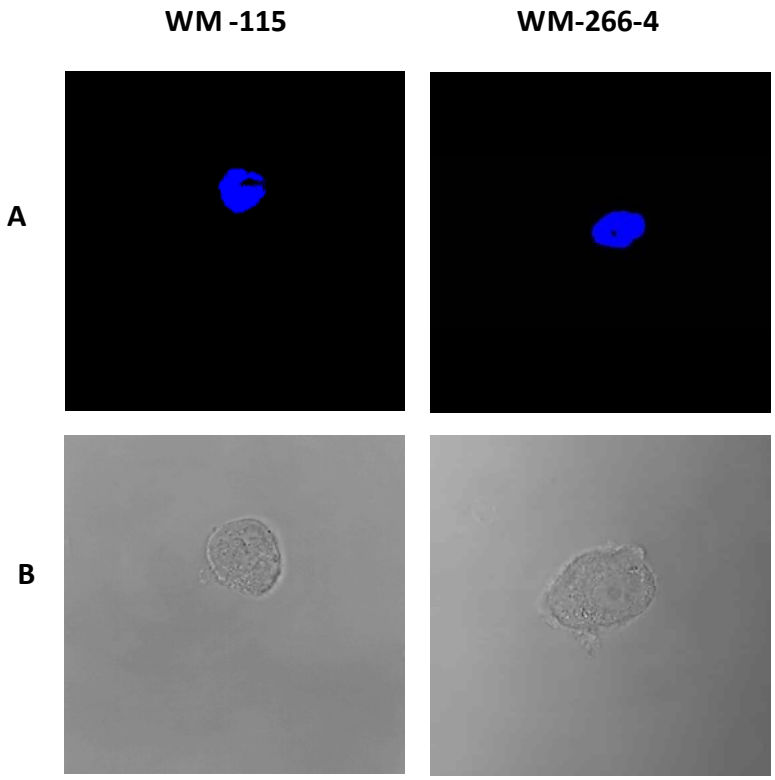


Figure S4.1 Control images of WM-115 and WM-266-4 cells stained with Alexa Fluor 488-conjugated rabbit secondary antibody before fixation. Cells were also stained with DAPI (blue) to identify nuclei. The left and right panels represent the staining of WM-115 and WM-266-4 cells, respectively. (A) Fluorescent image (B) transmitted light image. Images of WM-115 and WM-266-4 cells were captured Leica TCS SP2/UV confocal microscope with identical parameters to compare the staining of the two cell lines.

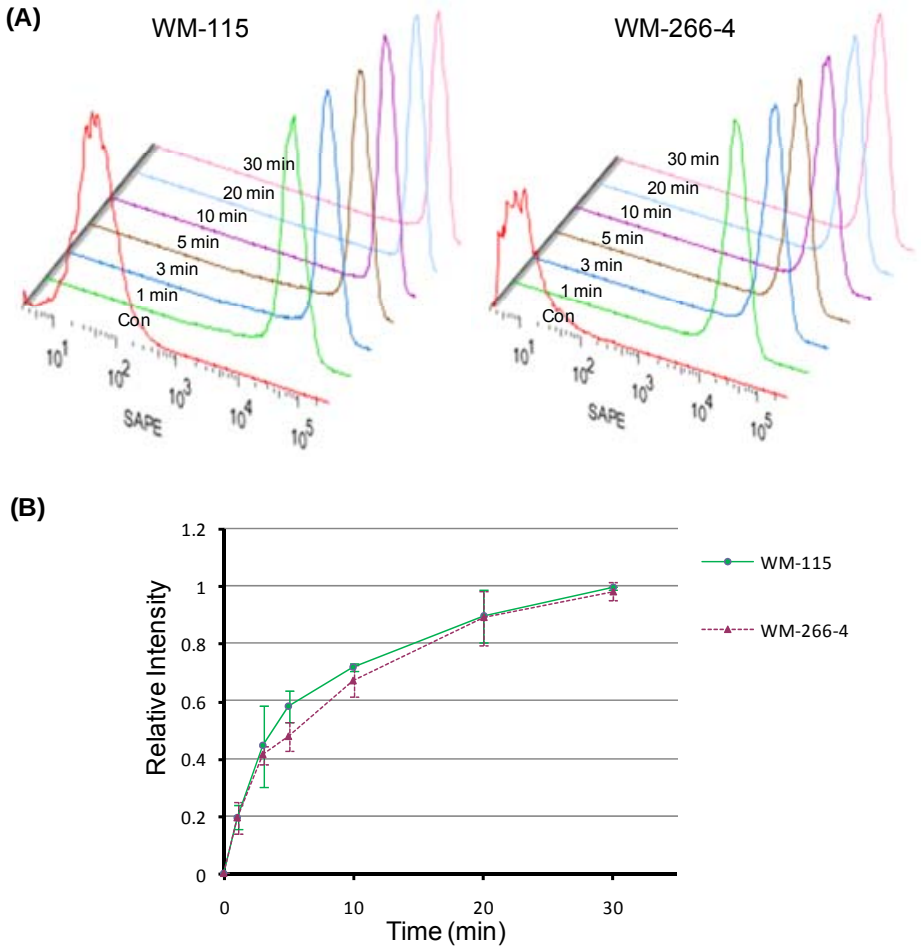


Figure S4.2 Flow cytometic analysis of cell surface biotinylation. (A) Histograms of WM-115 (left panel) and WM-266-4 (right panel) cells with 1, 3, 5, 10, 20 and 30 minutes of biotinylation reaction. After reaction, the cells were stained with SAPE. The cells without labeling reaction were stained with SAPE as control (Con). (B) The fluorescent intensities of SAPE were normalized as relative intensities by using the highest signal as "1" within a serial labeling reaction of same cells. The average relative intensities of three individual measurements were plotted with the different reaction time.

Table S4.1 Control list of proteins from light and heavy lysates (1:1 ratio) of WM-266-4 cells

NCBI accession	Protein name	Peptide used for quantification	charge state	m/z	L/H Ratio	SD (n=3)
NP_002517	5' nucleotidase, ecto				0.94	0.03
		R.LFTK#VQQIR.R	2	736.41		
		R.LFTK@VQQIR.R	2	740.42		
NP_853509	butyrophilin, subfamily 3, member A1				0.79	0.04
		R.DGITAGK#AALR.I	2	706.38		
		R.DGITAGK@AALR.I	2	710.38		
NP_598000	CD109				1.14	0.04
		R.K@YQPNIDVQESIHFLESEFSR.G	3	971.80		
		R.K#YQPNIDVQESIHFLESEFSR.G	3	969.13		
NP_001001390	CD44 antigen isoform 3 precursor				0.87	0.11
		R.YVQK#GEYR.T	2	691.34		
		R.YVQK@GEYR.T	2	695.34		
NP_976076	CD59 antigen p18-20				0.85	0.06
		K.AGLQVYNK#CWK.F	2	853.42		
		K.AGLQVYNK@CWK*.F	2	861.36		
NP_000328	delta-sarcoglycan isoform 1				1.12	0.10
		K.LDAAK#IR.L	2	563.31		
		K.LDAAK@IR.L	2	567.32		
NP_003527	H3 histone family, member K				1.05	0.14
		R.YQK#STELLIR.K	2	795.43		
		R.YQK@STELLIR.K	2	799.40		

NP_003536	H4 histone family, member J				1.02	0.10
		K.TVTAMDVVYALK#R.Q	3	602.65		
		K.TVTAMDVVYALK@R.Q	3	605.32		
NP_001019770	histone 2, H2bf				1.07	0.09
		R.LLLPGELAK#HAVSEGTK.A	2	1051.57		
		R.LLLPGELAK@HAVSEGTK*.A	2	1055.65		
NP_002194	integrin alpha 2 precursor				0.98	0.02
		R.K#YAYSAASGGR.R	2	735.35		
		R.K@YAYSAASGGR.R	2	739.32		
NP_005492	integrin alpha 3 isoform b, precursor				1.21	0.11
		R.TGAVYLCPLTAHK#DDCER.M	3	815.70		
		R.TGAVYLCPLTAHK@DDCER.M	3	818.35		
NP_000201	integrin alpha chain, alpha 6 isoform b precursor				0.92	0.04
		R.GLDSK#ASLILR.S	2	756.42		
		R.GLDSK@ASLILR.S	2	760.39		
NP_002201	integrin alpha-V precursor				1.12	0.13
		R.QVVCDLGNPMK#AGTQLLAGLR.F	3	860.77		
		R.QVVCDLGNPMK@AGTQLLAGLR.F	3	863.44		
NP_596867	integrin beta 1 isoform 1A precursor				0.98	0.05
		K.LK#PEDITQIQPQQLVLR.L	3	786.77		
		K.LK@PEDITQIQPQQLVLR.L	3	789.44		
NP_852478	integrin, alpha 1 precursor				0.90	0.11
		R.K#EYAQR.I	2	567.28		
		R.K@EYAQR.I	2	571.35		

NP_006491	melanoma cell adhesion molecule				1.01	0.10
		R.IFLCQGK#R.P	2	684.86		
		R.IFLCQGK@R.P	2	680.85		
NP_001670	Na+/K+ -ATPase beta 3 subunit				1.09	0.17
		R.IIGLK#PEGVPR.I	2	759.43		
		R.IIGLK@PEGVPR.I	2	763.44		
XP_950238	PREDICTED: similar to HLA class I histocompatibility antigen, A-29 alp				0.98	0.04
		R.YTCHVQHEGLPK#PLTLR.W	3	796.74		
		R.YTCHVQHEGLPK@PLTLR.W	3	799.41		
NP_065173	prostaglandin F2 receptor negative regulator				1.01	0.09
		R.K#GIVTTSR.R	2	600.83		
		R.K@GIVTTSR.R	2	604.83		
NP_001012679	solute carrier family 3 (activators of dibasic and neutral amino ac)				1.04	0.21
		R.IGDLQAFQGHGAGNLAGLK#GR.L	3	807.08		
		R.IGDLQAFQGHGAGNLAGLK@GR.L	3	809.75		

Note: @ biotinylated heavy lysine; # biotinylated light lysine; all cysteine residues are CAM-modifed with a mass increase of 57.01 Da.

Table S4.2 A list of quantified proteins from WM-115 and WM-266-4 cells

	NCBI Accession	Protein Name	cell surface/membrane GO	WM266-4/115 H/L Ratio	SD	WM266-4 Hits(H)	WM115 Hits(L)
1	NP_002517	5' nucleotidase, ecto	Y	0.52	0.03	12	18
2	NP_006661	5T4 oncofetal trophoblast glycoprotein	Y	0.72	0.06	0	2
3	NP_001618	activated leukocyte cell adhesion molecule	Y	0.07	0.03	0	15
4	NP_004039	beta-2-microglobulin precursor	Y	1.31	0.05	3	1
5	NP_853509	butyrophilin, subfamily 3, member A1	Y	2.37	0.59	3	3
6	NP_598000	CD109	Y	3.65	0.04	4	1
7	NP_001001390	CD44 antigen isoform 3 precursor	Y	1.84	0.09	3	3
8	NP_001770	CD58 antigen, (lymphocyte function-associated antigen 3)	Y	0.50	0.06	1	3
9	NP_976076	CD59 antigen p18-20	Y	0.52	0.10	14	20
10	NP_001760	CD9 antigen	Y	0.31	0.02	1	4
11	NP_015565	cytochrome b5 reductase isoform 2	Y	0.97	0.04	2	4
12	NP_000328	delta-sarcoglycan isoform 1	Y	1.18	0.04	2	1
13	NP_758447	delta-sarcoglycan isoform 2	Y	1.08	0.32	4	5
14	NP_006721	deoxyribonuclease I-like 1 precursor	N	1.12	0.08	1	1
15	NP_000102	down-regulated in adenoma protein	Y	5.97	0.55	2	0
16	NP_000109	endoglin precursor	Y	0.65	0.03	2	9
17	NP_004422	ephrin receptor EphA2	Y	0.36	0.06	0	9
18	NP_001005862	erbB-2 isoform b	Y	0.29	0.07	0	3
19	NP_001393	eukaryotic translation elongation factor 1 alpha 1	R	1.30	0.06	2	1
20	NP_066390	H2A histone family, member A	R	0.53	0.07	3	4
21	NP_003527	H3 histone family, member K	R	0.79	0.06	16	19
22	NP_003536	H4 histone family, member J	R	0.45	0.26	19	20
23	NP_005338	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	R	1.38	0.09	1	2

24	NP_066409	histone 1, H2ad	R	1.00	0.08	3	4
25	NP_001019770	histone 2, H2bf	R	0.38	0.19	1	9
26	NP_778235	histone H2A	R	0.55	0.01	3	5
27	NP_055148	immunoglobulin superfamily, member 4D isoform 1	Y	0.47	0.11	3	4
28	NP_002194	integrin alpha 2 precursor	Y	0.57	0.20	7	10
29	NP_005492	integrin alpha 3 isoform b, precursor	Y	0.96	0.11	16	16
30	NP_002196	integrin alpha 5 precursor	Y	1.11	0.32	1	3
31	NP_000201	integrin alpha chain, alpha 6 isoform b precursor	Y	0.99	0.11	16	15
32	NP_002201	integrin alpha-V precursor	Y	1.89	0.26	23	6
33	NP_596867	integrin beta 1 isoform 1A precursor	Y	1.04	0.07	8	5
34	NP_000203	integrin beta chain, beta 3 precursor	Y	2.38	0.07	19	9
35	NP_852478	integrin, alpha 1 precursor	Y	0.96	0.13	2	2
36	NP_002204	integrin, beta 5	Y	0.58	0.13	3	5
37	NP_085046	inter-alpha trypsin inhibitor heavy chain precursor 5 isoform 1	R	1.17	0.16	2	2
38	NP_061984	major histocompatibility complex, class II, DR alpha precursor	Y	>5	X	2	0
39	NP_004986	matrix metalloproteinase 14 preproprotein	Y	2.18	0.48	6	0
40	NP_006491	melanoma cell adhesion molecule	Y	3.13	0.69	19	6
41	NP_005920	melanoma-associated antigen p97 isoform 1, precursor	Y	3.60	1.05	2	0
42	NP_001888	melanoma-associated chondroitin sulfate proteoglycan 4	Y	1.46	0.18	7	2
43	NP_056344	monooxygenase, DBH-like 1 isoform 1	N	0.99	0.04	3	1
44	NP_001001586	Na+/K+ -ATPase alpha 1 subunit isoform b proprotein	Y	1.33	0.49	7	3
45	NP_001670	Na+/K+ -ATPase beta 3 subunit	Y	1.88	0.14	8	3
46	NP_055747	neuroligin 1	Y	1.08	0.06	1	1
47	NP_000933	peptidylprolyl isomerase B precursor	R	1.05	0.07	2	2
48	NP_036533	plexin B2	Y	0.84	0.03	1	3
49	NP_005388	podocalyxin-like precursor isoform 2	Y	<0.05	X		

50	XP_950238	PREDICTED: similar to HLA class I histocompatibility antigen, A-29 alp	Y	0.84	0.03	8	7
51	XP_933500	PREDICTED: similar to immunoglobulin superfamily, member 3 isoform 1	Y	2.62	0.10	2	0
52	NP_065173	prostaglandin F2 receptor negative regulator	Y	1.65	0.35	17	10
53	NP_005304	protein disulfide isomerase-associated 3 precursor	R	0.65	0.23	1	2
54	NP_002834	protein tyrosine phosphatase, receptor type, J isoform 1 precursor	Y	1.18	0.01	1	1
55	NP_005496	scavenger receptor class B, member 1 isoform 1	Y	3.25	0.35	6	0
56	NP_543006	signal-regulatory protein beta 2 isoform 2	Y	0.29	0.07	1	2
57	NP_001143	solute carrier family 25, member 5	Y	>20	X	3	0
58	NP_001012679	solute carrier family 3 (activators of dibasic and neutral amino ac)	Y	2.53	0.18	9	6
59	NP_065075	solute carrier family 39 (zinc transporter), member 10	Y	1.71	0.08	3	1
60	NP_055205	staphylococcal nuclease domain containing 1	N	1.23	0.12	2	1
61	NP_037522	transmembrane protein 2	Y	0.78	0.08	1	3
62	NP_003371	vimentin	R	0.43	0.03	1	4

Note.Y: Cell surface/membrane proteins based on GO annotation; R: Membrane-associated proteins based on previous publications; N: Non-membrane related membrane protein; Hits: Total number of peptides (including same peptides with multiple identification and different m/z) identified with SEQUEST search based on LC-MS/MS data acquired in data-dependant scan mode; H: heavy; L: light

CHAPTER 5

Quantitative Proteomic Analysis of HL-60 Human Leukemia Cells upon Treatment with 6-Thioguanine

Introduction

Thiopurine drugs, including azathioprine (Aza), 6-mercaptopurine (6-MP) and 6-thioguanine (6-TG), are widely used as cancer therapeutic and immunosuppressive agents.[1-4] For example, 6-MP and 6-TG are the most commonly prescribed drugs for acute lymphoblastic leukemia (ALL) and acute myelogenous leukemia (AML).[3, 5-7] The cytotoxicity of thiopurines involves changes in purine metabolism and the formation of thioguanine nucleotide and its subsequent incorporation into DNA.[1] For the latter process, the precise molecular mechanisms underlying the therapeutic activity remain unclear.[8, 9]

Microarray has been used for the quantitative evaluation of transcriptional responses in cancer cells that are treated with anticancer agents, and generates a large amount of mRNA expression data for different cancers.[10, 11] However, microarray analysis does not provide information for translational regulation of gene expression, which often exhibits a poor correlation with transcript levels due to the different kinetics of protein translation and turnover, and post-translational modifications.[12]

Recently, mass spectrometry (MS)-based proteomics and its related techniques have allowed the identification and quantification of a large number of proteins in complex

samples, which have shown a great potential in the identification of therapeutic target and in the development of anti-cancer therapies.[13-16] So far, numerous proteomic studies underlying the effect of treatments with different anti-cancer drugs, such as cisplatin,[17, 18] etoposide,[19, 20] ATRA,[21] have been performed by using MS-based proteomics and two-dimensional gel electrophoresis (2-DE).[13]

Other than 2-DE, several stable isotope labeling strategies, such as isotope-coded affinity tag (ICAT),[22] isobaric tags for relative and absolute quantitation (iTRAQ)[23] and stable isotope labeling with amino acids in cell culture (SILAC)[24], have been developed for the MS-based comparative analysis of differential protein expression between normal and cancer or drug-treated and untreated cancer cells. Among these isotope-labeling strategies, SILAC is a metabolic labeling method, which is very simple, efficient, and can facilitate almost complete heavy isotope incorporation. In addition, SILAC introduces heavy isotopes during cell culture; light and heavy samples are mixed prior to various steps of sample manipulation, thereby minimizing differential sample loss during protein extraction, fractionation and digestion. Therefore, SILAC is very suitable for the comparative study of protein expression in cells with and without drug treatments; accurate experimental results could be obtained with minimal bias, allowing relative quantitation of small changes in protein abundance.[24, 25]

In this work, we employed MS, together with metabolic labeling using SILAC, to examine the effect of 6-TG treatment on protein expression in HL-60 human leukemia cells. We were able to quantify a total of 299 proteins by using LC-MS/MS. Among these

quantified proteins, fifteen were differentially expressed upon 6-TG treatment. The MS-based proteomic results were further validated by using Western blot analysis of several proteins. This represents the first study of the proteomic response of human cells toward 6-TG treatment, which could facilitate the understanding of novel molecular pathways for the anti-cancer effects of the thiopurine drugs.

Materials and Methods

Cell Culture

Human promyelocytic leukemia (HL-60) cells (ATCC, Manassas, VA) were cultured in Iscove's modified minimal essential medium (IMEM) supplied with 20% fetal bovine serum (FBS, Invitrogen, Carlsbad, CA), 100 IU/mL of penicillin and 100 μg/mL of streptomycin in 75 cm^2 culture flasks. Cells were maintained in a humidified atmosphere with 5% CO_2 at 37 ºC, with medium renewal of 2-3 times a week depending on cell density. For SILAC experiments, the IMEM medium lacking L-lysine and L-arginine was prepared according to the ATCC formulation. Heavy lysine and arginine ($[^{13}C_6,^{15}N_2]$-L-lysine and $[^{13}C_6,^{15}N_4]$-L-arginine) were purchased from Cambridge Isotope Laboratories (Andover, MA), and light (unlabeled) lysine and arginine were obtained from Sigma (St. Louis, MO). Light and heavy IMEM media were prepared by the addition of either light or heavy lysine and arginine, with the supplement of dialyzed FBS (Invitrogen), to the above lysine, arginine-deficient medium. The HL-60 cells were

cultured in heavy IMEM medium for at least one week to achieve complete isotope incorporation.

6-TG Treatment and Sample Preparation

HL-60 cells, at a density of ~ 5×10^5 cells per mL in heavy IMEM medium, were treated with 3 μM 6-TG (Sigma) for 24 h. The mock-treated cells cultured in light IMEM medium were used as the internal standard. After treatment, the light and heavy cells were collected by centrifugation at 300 g at 4 ºC for 5 min, and washed three times with ice-cold PBS to remove culture medium and FBS. For cell lysis, CelLytic™ M lysis buffer (Sigma) was used with the supplement of 1 mM NaF, 1 mM Na_3VO_4, 1 mM PMSF and a protease inhibitor cocktail (Sigma). Cell lysates were centrifuged at 16 000 g at 4 °C for 30 min, and the resulting supernatants were collected. The cell lysates were diluted by at least 30-fold, and the protein concentrations of these diluted lysates were determined by using Quick Start Bradford Protein Assay (Bio-Rad, Hercules, CA). The light and heavy cell lysates were then combined at a ratio of 1/1 (w/w). Heavy HL-60 cells without treatment were used as an external mixing control, and the light and heavy lysates were prepared and mixed equally, following the identical protocol as the treated cells in parallel.

SDS-PAGE Separation and In-gel Digestion

The control and treated protein samples were denatured by boiling in Laemmli

loading buffer and separated by a 12% SDS-PAGE with 4% stacking gel. The gel was stained with Coomassie blue. After destaining, each gel lane was cut into 10 bands, in-gel reduced with dithiothreitol (DTT) and alkylated with iodoacetamide (IAA). The proteins were digested with trypsin (Promega, Madison, WI) for overnight. Following digestion, peptides were extracted into 0.1% trifluoroacetic acid in CH_3CN/H_2O (1:1, v/v) and dried in a Speed-vac. The resulting peptide mixtures were stored at –20 °C until further analysis.

Western Blotting

For Western blot analysis, 6-TG-treated and untreated HL-60 cells were harvested and the proteins were prepared following the same protocol as described above. After SDS-PAGE separation, proteins were transferred to a nitrocellulose membrane under standard conditions. A basic transfer buffer containing 10 mM $NaHCO_3$, and 3 mM Na_2CO_3 (pH 9.9) was used for immunoblotting histones. After protein transfer, the membranes were blocked with 2% non-fat milk ECL Advance blocking reagent (GE Healthcare, UK) and incubated with primary antibodies under optimized conditions. The membranes were rinsed briefly with two changes of PBS-T wash buffer [PBS solution containing 0.1% (v/v) Tween-20, pH 7.5] and washed with a large amount of wash buffer for 15 min, followed by 3 × 5-min wash with fresh changes of wash buffer at room temperature. After washing, primary antibodies were recognized by incubating with horse radish peroxidase (HRP)-conjugated secondary antibodies at room temperature for 1 h. The membranes were washed thoroughly with PBS-T (1× 15 min, then 3 × 5 min). The antibody binding was detected by using ECL Advance Western Blotting Detection Kit

(GE Healthcare), and visualized with HyBlot CL autoradiography film (Denville Scientific Inc., Metuchen, NJ). Standard histones, which were purified from untreated HL-60 cells with HPLC and confirmed by MALDI-MS analysis on a QSTAR oMALDI system (Applied Biosystems), were used as positive controls. The rabbit anti-human ferritin antibody was purchased from Bethyl Laboratories (Montgomery, TX). The rabbit anti-histone H2A, H3, actin and mouse anti-histone H4 antibodies were purchased from Abcam (Cambridge, MA). The mouse anti-histone H2B was from MBL International (Woburn, MA). HRP-conjugated goat anti-rabbit and anti-mouse IgG secondary antibodies were obtained from Abcam and Santa Cruz Biotechnology (Santa Cruz, CA), respectively.

LC-MS/MS for Protein Identification and Quantification

Online LC-MS/MS analysis for protein identification was performed on an LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with an electrospray ionization (ESI) source. An Agilent 1100 capillary HPLC system with a Zorbax SB-C18 capillary column (5 μm in particle size, 0.5 × 250 mm, Agilent Technologies, Santa Clara, CA) was employed for peptide separation. The chromatographic separation was performed with a linear gradient of 2-15% CH_3CN in 0.1% aqueous solution of formic acid in the first 13 min, followed by a 105-min gradient of 15-50% CH_3CN in 0.1% formic acid; the flow rate was 6 μL/min. The capillary temperature for the ESI source was maintained at 275 °C and the source voltage was set at 4 kV. The maximum ion injection time was 50 ms with automatic gain control (AGC)

values of 3×10^3 and 1×10^3 ions for full-scan MS and MS/MS, respectively. The product-ion spectra were acquired in a data-dependent scan mode, where the eight most abundant ions observed in MS were chosen for collision-induced dissociation (CID). The normalized collision energy was 35% and the activation *Q* was 0.25. In addition, the dynamic exclusion feature was activated to maximize the detection of low-abundance peptide ions.

For protein quantification, same samples analyzed on LTQ were analyzed in triplicate on an Agilent 6510 Q-TOF LC/MS system (Agilent Technologies) by using the same Zorbax SB-C18 capillary column and identical gradients. The Agilent Q-TOF was operated in an auto (data-dependent) MS/MS mode, where a full MS scan was followed by maximum of eight MS/MS scans (abundance-only precursor selection), with m/z ranges of 350-2000, and 60-2000 in MS and MS/MS modes, respectively. The active (dynamic) exclusion feature was enabled to discriminate against ions previously selected for MS/MS analyses in two sequential scans. The acquisition rates were 6 spectra/s and 3 spectra/s in MS and MS/MS modes, respectively. For CID, the collision energy was set to use a slope of 3 V per 100 m/z units with an offset of 2.5 V to fragment the selected precursor ions for MS/MS. The third analysis of the sample was performed in MS mode to improve the quality of protein quantitation.

Data Processing

The LC-MS/MS raw data acquired on LTQ were searched against NCBI human RefSeq protein database using TurboSEQUEST with Bioworks 3.2 for protein

identification. Variable modifications on lysine and arginine residues were allowed with mass increases of 8.01 Da and 10.0 to recognize heavy lysine and arginine residues, respectively. Methionine oxidation was set as a variable modification. The static carbamidomethylation (CAM) of cysteine residues was used with a mass increase of 57.01 Da. Peptide filters with cross-correlation (Xcorr $\geq 1.9, \geq 2.4, \geq 3.5$ for peptide ions that are singly, doubly, and triply charged) and delta correlation ($\Delta C_n \geq 0.1$) scores were used to sort the search results. In addition, the probability score was set as ≤ 0.1 to further filter the identified proteins.

Protein quantification was achieved by averaging manually the relative areas of peaks found in the extracted-ion chromatograms of precursor ions of light and heavy peptides based on Q-TOF data, and the data were processed using Agilent MassHunter workstation software (Version B.01.03). The retention times of peptides used for protein quantitation in Q-TOF data were checked to be consistent with the LTQ data. Some proteins that were identified by LTQ, but not by Q-TOF, were also quantified based on LC-MS results acquired on Q-TOF, where the accurate mass of peptide ions, retention time, and the numbers of isotope-labeled lysine and/or arginine in the peptide were employed to locate the light/heavy peptide pairs for the quantification.

Results and Discussion

6-MP and 6-TG have been widely used as anticancer drugs for the treatment of ALL and AML.[3, 5-7] Although the exact mechanisms for the cytotoxicity of these drugs remain unclear, it has been known that the conversions of these drugs to 6-TG

nucleotides and their subsequent incorporation into nucleic acids are important for the drugs to be effective.[1] It was estimated that during DNA replication approximately 0.01-0.1% of DNA guanines, corresponding to 10^5 to 10^6 per cell, may be substituted by the thionucleobase.[4, 26] The extensive substitution could be associated with chromosomal damage and cell death. It was also proposed that the covalent modifications of DNA 6-TG and the triggering of DNA mismatch repair (MMR) pathway may be implicated in the cytotoxicity of the thiopurine drugs.[2]

MS-based proteomics has already been applied for the identification of targets involved in cellular responses at the proteome level upon treatment with different drugs.[13, 17-21] Most of these studies identified the differential expression of proteins involved in the cellular stress response, metabolism, cell cycle, signal transduction, as well as many other biological processes. However, to the best of our knowledge, no study has been carried out for the quantitative evaluation of the effect of 6-TG treatment on protein expression in human leukemia cells. We hope that the current work could facilitate the understanding of novel molecular pathways for the anti-cancer effects of the thiopurine drugs.

Protein Identification and Quantification

We employed SILAC, SDS-PAGE and LC-MS/MS to assess the perturbation of protein expression in HL-60 human leukemia cells induced by 6-TG treatment. As depicted in Figure 5.1, the HL-60 cells cultured in light medium were used as internal

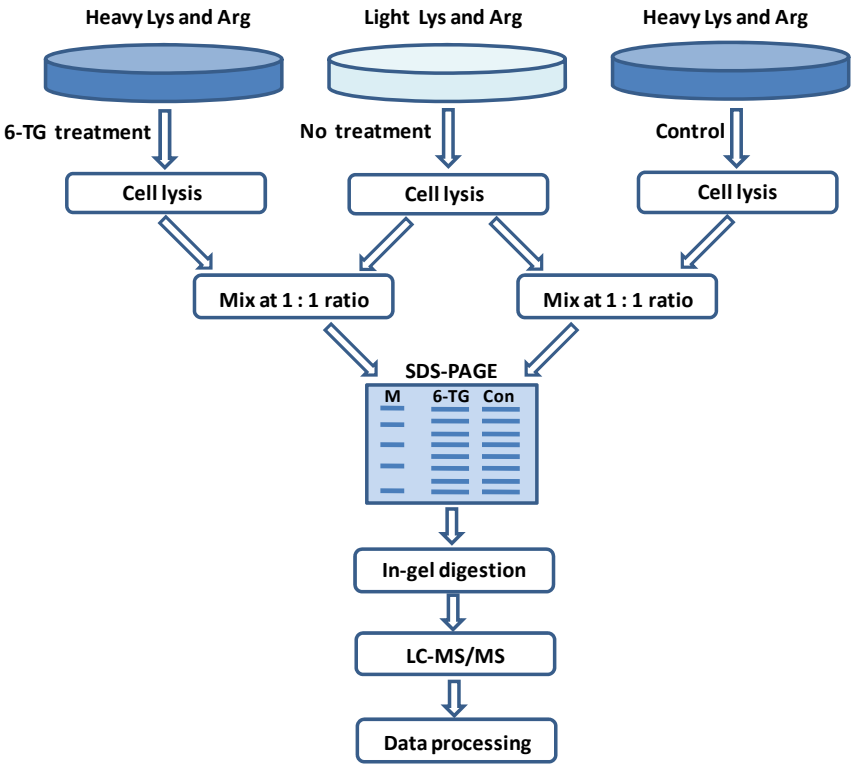


Figure 5.1 A schematic illustration of comparative analysis of protein expression in HL-60 cells upon 6-TG treatment.

standard, and cultured in parallel with HL-60 cells in heavy medium. After cell lysis, the lysate of light cells was mixed equally (by weight) with heavy lysate of untreated or 6-TG-treated cells. The use of both internal standard and control cells allows the minimization of systematic errors, which could emanate from Bradford assay for protein quantification and mixing steps. In addition, by using internal standard, the effects of potentially incomplete incorporation of heavy isotopes and the metabolic conversion of arginine to proline could be eliminated, though stable isotope-labeled arginine has been widely used for SILAC experiments.[27-30] Therefore, more accurate quantitative results could be obtained and subtle protein expression difference between the treated and untreated cells could be determined.

We identified a total of 635 proteins by using LC-MS/MS in data-dependant scan mode on an LTQ linear ion trap mass spectrometer. Among the identified proteins, ~ 300 were quantified based on peak areas found in the extracted-ion chromatograms of precursor ions of light/heavy peptide pairs detected by Q-TOF LC-MS. The other half of the identified proteins cannot be quantified, mainly because of the weak MS signal associated with the identified peptides, which made the manual inspection of retention time and quantitation difficult. The majority of the quantified proteins have similar expression levels in the treated and untreated cells. To make it easy to compare, the $\log_{1.5}$ values of relative abundance ratios were plotted against the number of proteins quantified. As depicted in Figure 5.2, the expression of a small number of proteins was significantly changed upon 6-TG treatment with fold changes >1.5 or <1.5,

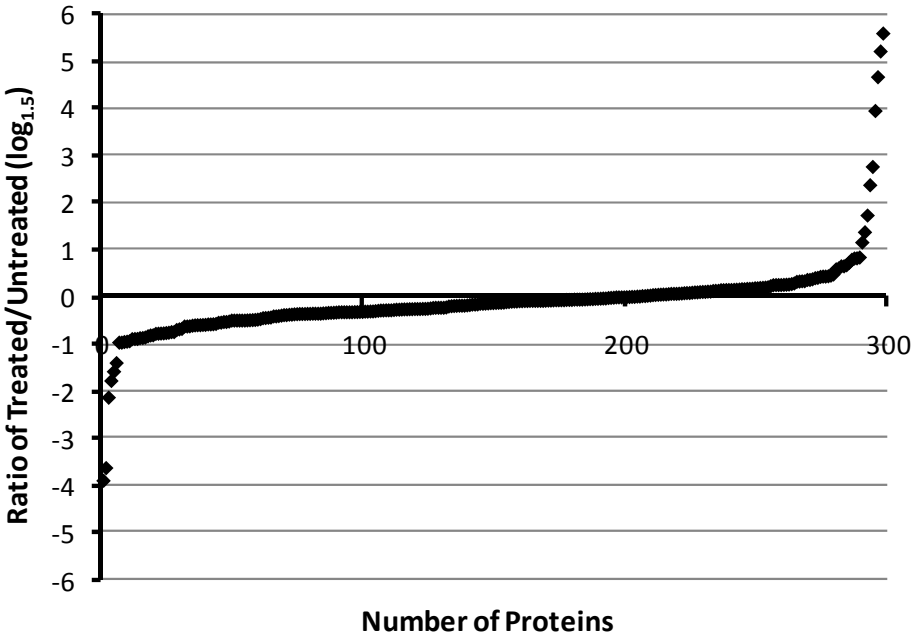


Figure 5.2 A summary of the quantitation results. The $log_{1.5}$ ratios were plotted against protein numbers.

corresponding to ± 1 at y-axis. Nine proteins were up-regulated and six were down-regulated in HL-60 cells treated with 6-TG. These differentially expressed proteins and the quantification results are summarized in Table 5.1.

Western Blot Analysis

The quantification results were further validated by using Western blot analysis of the whole lysates of untreated and treated cells, and we chose to assess the expression of ferritin light polypeptide and histone proteins including H2A, H2B, H3 and H4. Consistent with what we found from SILAC and LC-MS/MS analysis, Western blot analysis also revealed that histones and ferritin were expressed at higher levels in 6-TG-treated HL-60 cells than in untreated cells (Figure 5.3).

Characterization of Differentially Expressed Proteins

Among the differentially expressed proteins, almost all histone proteins, including H2A, H2B, H4 and linker histone H1, were up-regulated in 6-TG-treated cells. We were not able to quantify histone H3, but we identified three unique heavy isotope-labeled peptides from H3, whereas the corresponding light peptides were not identified, indicating that H3 was also expressed at a greater level in 6-TG-treated HL-60 cells. Histones are the major protein constituents of chromatin and play important roles in DNA packaging, chromosome stabilization and gene expression.[31, 32] All four core histones H2A, H2B, H3 and H4 share a similar structure with a fold domain and terminal

Table 5.1 Proteins quantified with changes > 1.5 or < 1.5.

NCBI accession	Protein Name	Treated/ Untreated	SD
	Up-regulated		
NP_001677	ATP synthase, H+ transporting, mitochondrial F1 complex, beta subunit	1.59	0.07
NP_001012679	solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2 isoform a	1.73	0.03
NP_002942	ribophorin II precursor	2.00	0.01
NP_002955	S100 calcium-binding protein A8	2.60	0.09
NP_000137	ferritin, light polypeptide	3.05	0.10
NP_005312	H1 histone family, member 4	4.94	0.78
NP_066390	H2A histone family, member A	6.61	0.16
NP_001019770	histone 2, H2Bf	8.25	0.90
NP_003536	histone H4	9.63	3.13
	Down-regulated		
NP_057667	palate, lung and nasal epithelium carcinoma associated protein precursor	0.20	0.02
NP_001087227	similar to ribosomal protein L23A	0.23	0.02
NP_780294	trifunctional purine biosynthetic protein adenosine-3	0.42	0.12
NP_004238	U5 snRNP-specific protein, 116 kD	0.48	0.14
NP_006392	acidic (leucine-rich) nuclear phosphoprotein 32 family, member B	0.52	0.03
NP_005817	heterogeneous nuclear ribonucleoprotein R	0.56	0.02

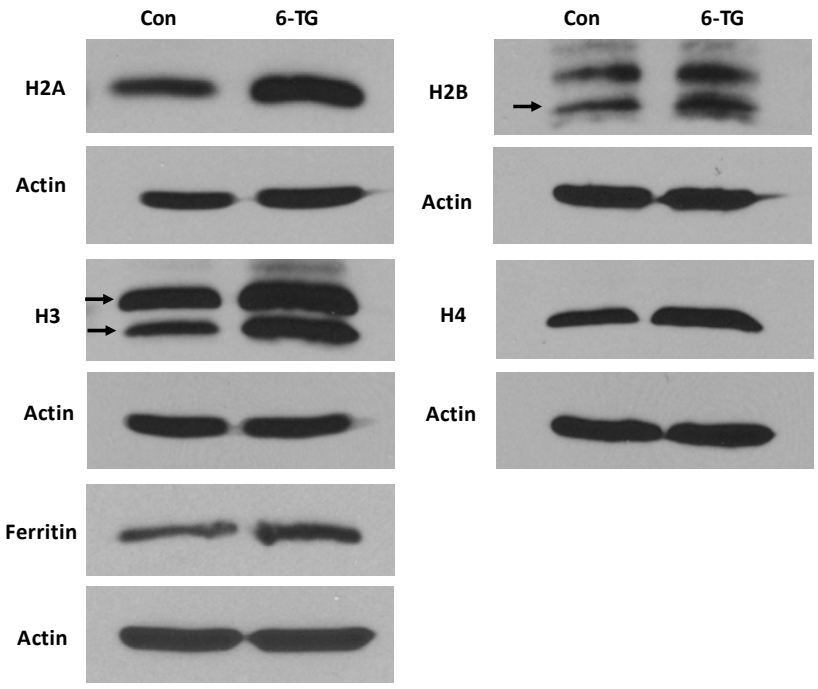


Figure 5.3 Western blot analysis of histones H2A, H2B, H3, H4 and ferritin light polypeptide with lysates of 6-TG-treated "6-TG" and untreated "Con" HL-60 cells. Actin was used as the loading control. The black arrows indicate the histone bands confirmed with purified histones.

tails.[31] Linker histone H1 binds nucleosomes together and maintains a condensed higher-order compaction of chromatin.[33] The major known functions of histones are associated with their nucleosomal localization. However, increasing lines of evidence has shown that histone proteins bear additional activities and novel functions, and are not confined within their normal nucleosomal context.[34-45] For instance, a recent study showed that a histone H2A variant, H2AX is overexpressed upon DNA replication stress and its overexpression is associated with increased apoptosis.[41] Another report from the same group showed that Imatinib, a selective small-molecule protein kinase inhibitor clinically used to treat chronic myelogenous leukemia (CML)[46], can trigger the apoptosis of gastrointestinal stromal tumor cells in part through the up-regulation of soluble histone H2AX.[42] Overexpression of histone H3 variant was also reported in a defect in kinetochore biorientation in *Saccharomyces cerevisiae*.[43] Besides these reports, our recent study demonstrated the differential expression of histones H2A, H2B, H3 and H4 on plasma membranes of human primary and metastatic melanoma cells.[47]

The aberrant expression of histone H1 and its variants has also been recognized in other studies.[39, 40, 44, 45, 48] It was found that high levels of non-chromatin-bound histones are cytotoxic and can lead to an increase in DNA damage sensitivity.[48] Because histone H1 has long been recognized as a major constituent of chromatin, it is difficult to believe that this protein could play roles other than packaging nucleosomes into chromatin. Nevertheless, a previous report has shown that non-nuclear histone H1 is upregulated in neurons and astrocytes in patients suffering from prion and Alzheimer's diseases but not in acute neurodegeneration, indicating that histone H1 may be involved in the

progression of neurodegenerative diseases.[44] In the present study, we found that the expression levels of major histones in human leukemia cells could be altered upon 6-TG treatment. Although the mechanisms through which the histone proteins are up-regulated upon 6-TG treatment remain unknown, a recent study showed that 6-thioguanine treatment can lead to a decrease in global DNA cytosine methylation in cultured human cells.[49] Therefore, we speculate that the treatment with 6-thioguanine may result in the alteration of global chromatin structure, and the increased expression of histone proteins might be reflective of the chromatin structure change.

In addition to histones, we found that ferritin was significantly upregulated upon 6-TG treatment. Ferritin is a ubiquitous and highly conserved intracellular iron storage protein, which contains 24 polypeptide subunits. In vertebrates, these subunits constitute both light and heavy forms.[50-52] It has been known that the expression of ferritin is regulated by iron, cytokines, hormones, oxidative stress and second messengers.[51] The altered ferritin levels have been implicated in the pathogenesis of many diseases including Parkinson's[53] and Alzheimer's diseases.[54] Previous studies revealed the accumulation of heavy form of ferritin protein and mRNA in HL-60 cells upon differentiation induced by phorbol ester PMA or dimethylsulfoxide.[55, 56] Another study showed that the induction of light ferritin subunit could be observed during monocytic differentiation.[57] However, the altered expression of ferritin in leukemia cells with 6-TG treatment has not been previously reported. Recently, a study showed that ferritin could also be upregulated in murine bone marrow upon exposure to ionizing radiation.[58] Although the exact mechanism underlying the cellular responses with altered ferritin

levels remains elusive, it was suggested that the elevated expression of ferritin could be mediated by different cytokines, which play a pivotal role in the cellular responses and communication.[51, 58]

Other proteins that were up- or down-regulated by 6-TG treatment included S100 calcium-binding protein A8 (S100A8),[59] and palate, lung and nasal epithelium carcinoma-associated protein precursor (PLUNC),[60, 61] etc. (Table 5.1).

Conclusions

Thiopurine drugs are among the most successful chemotherapeutic agents for treating leukemia and other human diseases. We reason that the assessment of the alteration of the whole proteome upon treatment with 6-thioguanine may offer new insights into the molecular mechanisms of action of these drugs. In the present study, we employed SILAC, together with LC-MS/MS, and assessed quantitatively the perturbations of protein expression in HL-60 leukemic cells upon treatment with 6-thioguanine. Our results revealed that the drug treatment led to elevated expression of major histones H2A, H2B, H3, H4 and H1. Other proteins, such as ferritin and S100A8, may also be involved in the cellular responses toward 6-thioguanine treatment. We believe that the outcome of this study may lead to a better understanding of the therapeutic activities of this drug. The pharmacoproteomic profiling could be a valuable tool for the identification of drug-response biomarkers and eventually lead to the discovery of new molecular pathways associated with the cytotoxic effects of the drug.

References

1. Elion, G. B., The purine path to chemotherapy. *Science* **1989,** 244, 41-47.
2. Karran, P., Thiopurines, DNA damage, DNA repair and therapy-related cancer. *Br. Med. Bull.* **2006,** 79-80, 153-170.
3. Pui, C. H.; Jeha, S., New therapeutic strategies for the treatment of acute lymphoblastic leukaemia. *Nat. Rev. Drug Discovery* **2007,** 6, 149-165.
4. Warren, D. J.; Andersen, A.; Slordal, L., Quantitation of 6-thioguanine residues in peripheral-blood leukocyte DNA obtained from patients receiving 6-mercaptopurine-based maintenance therapy. *Cancer Res.* **1995,** 55, 1670-1674.
5. Brox, L. W.; Birkett, L.; Belch, A., Clinical pharmacology of oral thioguanine in acute myelogenous leukemia. *Cancer Chemother. Pharmacol.* **1981,** 6, 35-38.
6. Vora, A.; Mitchell, C. D.; Lennard, L.; Eden, T. O. B.; Kinsey, S. E.; Lilleyman, J.; Richards, S. M.; Res, M. R. C. N. C., Toxicity and efficacy of 6-thioguanine versus 6-mercaptopurine in childhood lymphoblastic leukaemia: a randomised trial. *Lancet* **2006,** 368, 1339-1348.
7. Buchner, T.; Hiddemann, W.; Wormann, B.; Loffler, H.; Gassmann, W.; Haferlach, T.; Fonatsch, C.; Haase, D.; Schoch, C.; Hossfeld, D.; Lengfelder, E.; Aul, C.; Heyll, A.; Maschmeyer, G.; Ludwig, W. D.; Sauerland, M. C.; Heinecke, A., Double induction strategy for acute myeloid leukemia: the effect of high-dose cytarabine with mitoxantrone instead of standard-dose cytarabine with daunorubicin and 6-thioguanine: a randomized trial by the German AML Cooperative Group. *Blood* **1999,** 93, 4116-4124.

8. Duley, J. A.; Florin, T. H., Thiopurine therapies: problems, complexities, and progress with monitoring thioguanine nucleotides. *Ther. Drug Monit.* **2005,** 27, 647-654.

9. Gunnarsdottir, S.; Elfarra, A. A., Distinct tissue distribution of metabolites of the novel glutathione-activated thiopurine prodrugs cis-6-(2-acetylvinylthio)purine and trans-6-(2-acetylvinylthio)guanine and 6-thioguanine in the mouse. *Drug Metab. Dispos.* **2003,** 31, 718-726.

10. Cheok, M. H.; Yang, W.; Pui, C. H.; Downing, J. R.; Cheng, C.; Naeve, C. W.; Relling, M. V.; Evans, W. E., Treatment-specific changes in gene expression discriminate *in vivo* drug response in human leukemia cells. *Nat. Genet.* **2003,** 34, 85-90.

11. Ross, D. T.; Scherf, U.; Eisen, M. B.; Perou, C. M.; Rees, C.; Spellman, P.; Iyer, V.; Jeffrey, S. S.; Van de Rijn, M.; Waltham, M.; Pergamenschikov, A.; Lee, J. C.; Lashkari, D.; Shalon, D.; Myers, T. G.; Weinstein, J. N.; Botstein, D.; Brown, P. O., Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* **2000,** 24, 227-235.

12. Wilkins, M. R.; Sanchez, J. C.; Williams, K. L.; Hochstrasser, D. F., Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis* **1996,** 17, 830-838.

13. Kraljevic, S.; Sedic, M.; Scott, M.; Gehrig, P.; Schlapbach, R.; Pavelic, K., Casting light on molecular events underlying anti-cancer drug treatment: what can be seen from the proteomics point of view? *Cancer Treat. Rev.* **2006,** 32, 619-629.

14. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** 422, 198-207.

15. Ahn, N. G.; Shabb, J. B.; Old, W. M.; Resing, K. A., Achieving in-depth proteomics profiling by mass spectrometry. *ACS Chem. Biol.* **2007,** 2, 39-52.

16. Cravatt, B. F.; Simon, G. M.; Yates, J. R., The biological impact of mass-spectrometry-based proteomics. *Nature* **2007,** 450, 991-1000.

17. Castagna, A.; Antonioli, P.; Astner, H.; Hamdan, M.; Righetti, S. C.; Perego, P.; Zunino, F.; Righetti, P. G., A proteomic approach to cisplatin resistance in the cervix squamous cell carcinoma cell line A431. *Proteomics* **2004,** 4, 3246-3267.

18. Sinha, P.; Poland, J.; Kohl, S.; Schnolzer, M.; Helmbach, H.; Hutter, G.; Lage, H.; Schadendorf, D., Study of the development of chemoresistance in melanoma cell lines using proteome analysis. *Electrophoresis* **2003,** 24, 2386-2404.

19. Sinha, P.; Hutter, G.; Kottgen, E.; Dietel, M.; Schadendorf, D.; Lage, H., Increased expression of annexin I and thioredoxin detected by two-dimensional gel electrophoresis of drug resistant human stomach cancer cells. *J. Biochem. Biophys. Methods* **1998,** 37, 105-116.

20. Urbani, A.; Poland, J.; Bernardini, S.; Bellincampi, L.; Biroccio, A.; Schnolzer, M.; Sinha, P.; Federici, G., A proteomic investigation into etoposide chemo-resistance of neuroblastoma cell lines. *Proteomics* **2005,** 5, 796-804.

21. Bertagnolo, V.; Grassilli, S.; Bavelloni, A.; Brugnoli, F.; Piazzi, M.; Candiano, G.; Petretto, A.; Benedusi, M.; Capitani, S., Vav1 modulates protein expression during

ATRA-induced maturation of APL-derived promyelocytes: A proteomic-based analysis. *J. Proteome Res.* **2008,** 7, 3729-3736.

22. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999,** 17, 994-999.

23. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004,** 3, 1154-1169.

24. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **2002,** 1, 376-386.

25. Ong, S. E.; Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nature Chem. Biol.* **2005,** 1, 252-262.

26. Cuffari, C.; Li, D. Y.; Mahoney, J.; Barnes, Y.; Bayless, T. M., Peripheral blood mononuclear cell DNA 6-thioguanine metabolite levels correlate with decreased interferon-gamma production in patients with Crohn's disease on AZA therapy. *Digest. Disease. Sci.* **2004,** 49, 133-137.

27. Bendall, S. C.; Hughes, C.; Stewart, M. H.; Doble, B.; Bhatia, M.; Lajoie, G. A., Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. *Mol. Cell. Proteomics* **2008,** 7, 1587-1597.

28. Van Hoof, D.; Pinkse, M. W. H.; Oostwaard, D. W. V.; Mummery, C. L.; Heck, A. J. R.; Krijgsveld, J., An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat. Methods* **2007,** 4, 677-678.

29. Liang, X. Q.; Zhao, J.; Hajivandi, M.; Wu, R.; Tao, J.; Amshey, J. W.; Pope, R. M., Quantification of membrane and membrane-bound proteins in normal and malignant breast cancer cells isolated from the same patient with primary breast carcinoma. *J. Proteome Res.* **2006,** 5, 2632-2641.

30. Ong, S. E.; Kratchmarova, I.; Mann, M., Properties of C-13-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.* **2003,** 2, 173-181.

31. Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J., Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **1997,** 389, 251-260.

32. Van Holde, K. E., *Chromatin*. New York: Springer-Verlag.: 1989.

33. Th'ng, J. P. H.; Sung, R.; Ye, M.; Hendzel, M. J., H1 family histones in the nucleus - Control of binding and localization by the C-terminal domain. *J. Biol. Chem.* **2005,** 280, 27809-27814.

34. Hariton-Gazal, E.; Rosenbluh, J.; Graessmann, A.; Gilon, C.; Loyter, A., Direct translocation of histone molecules across cell membranes. *J. Cell Sci.* **2003,** 116, 4577-86.

35. Herren, T.; Burke, T. A.; Das, R.; Plow, E. F., Identification of histone H2B as a regulated plasminogen receptor. *Biochemistry* **2006,** 45, 9463-9474.

36. Parseghian, M. H.; Luhrs, K. A., Beyond the walls of the nucleus: the role of histones in cellular signaling and innate immunity. *Biochem. Cell Biol.* **2006,** 84, 589-604.

37. Watson, K.; Edwards, R. J.; Shaunak, S.; Parmelee, D. C.; Sarraf, C.; Gooderham, N. J.; Davies, D. S., Extra-nuclear location of histones in activated human peripheral blood lymphocytes and cultured T-cells. *Biochem. Pharmacol.* **1995,** 50, 299-309.

38. Watson, K.; Gooderham, N. J.; Davies, D. S.; Edwards, R. J., Nucleosomes bind to cell surface proteoglycans. *J. Biol. Chem.* **1999,** 274, 21707-21713.

39. Bianchi, M. E., Significant (re)location: how to use chromatin and/or abundant proteins as messages of life and death. *Trends Cell Biol.* **2004,** 14, 287-293.

40. Hadnagy, A.; Beaulieu, R.; Balicki, D., Histone tail modifications and noncanonical functions of histones: perspectives in cancer epigenetics. *Mol. Cancer Ther.* **2008,** 7, 740-748.

41. Liu, Y.; Parry, J. A.; Chin, A.; Duensing, S.; Duensing, A., Soluble histone H2AX is induced by DNA replication stress and sensitizes cells to undergo apoptosis. *Mol. Cancer* **2008,** 7, 61.

42. Liu, Y.; Tseng, M.; Perdreau, S. A.; Rossi, F.; Antonescu, C.; Besmer, P.; Fletcher, J. A.; Duensing, S.; Duensing, A., Histone H2AX is a mediator of gastrointestinal stromal tumor cell apoptosis following treatment with imatinib mesylate. *Cancer Res.* **2007,** 67, 2685-2692.

43. Collins, K. A.; Camahort, R.; Seidel, C.; Gerton, J. L.; Biggins, S., The overexpression of a Saccharomyces cerevisiae centromeric histone H3 variant mutant protein leads to a defect in kinetochore biorientation. *Genetics* **2007,** 175, 513-525.

44. Bolton, S. J.; Russelakis-Carneiro, M.; Betmouni, S.; Perry, V. H., Non-nuclear histone H1 is upregulated in neurones and astrocytes in prion and Alzheimer's diseases but not in acute neurodegeneration. *Neuropathol. Appl. Neurobiol.* **1999,** 25, 425-432.

45. PrymakowskaBosak, M.; Przewloka, M. R.; Iwkiewicz, J.; Egierszdorff, S.; Kuras, M.; Chaubet, N.; Gigot, C.; Spiker, S.; Jerzmanowski, A., Histone H1 overexpressed to high level in tobacco affects certain developmental programs but has limited effect on basal cellular functions. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** 93, 10250-10255.

46. Druker, B. J.; Talpaz, M.; Resta, D. J.; Peng, B.; Buchdunger, E.; Ford, J. M.; Lydon, N. B.; Kantarjian, H.; Capdeville, R.; Ohno-Jones, S.; Sawyers, C. L., Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* **2001,** 344, 1031-1037.

47. Qiu, H. B.; Wang, Y. S., Quantitative analysis of surface plasma membrane proteins of primary and metastatic melanoma cells. *J. Proteome Res.* **2008,** 7, 1904-1915.

48. Gunjan, A.; Verreault, A., A Rad53 kinase-dependent surveillsnce mechanism 3that regulates histone protein levels6 in S. cerevisiae. *Cell* **2003,** 115, 537-549.

49. Hogarth, L. A.; Redfern, C. P.; Teodoridis, J. M.; Hall, A. G.; Anderson, H.; Case, M. C.; Coulthard, S. A., The effect of thiopurine drugs on DNA methylation in relation to TPMT expression. *Biochem. Pharmacol.* **2008,** 76, (8), 1024-35.

50. Arosio, P.; Levi, S., Ferritin, iron homeostasis, and oxidative damage. *Free Radical Biol. Med.* **2002,** 33, 457-463.

51. Torti, F. M.; Torti, S. V., Regulation of ferritin genes and protein. *Blood* **2002,** 99, 3505-3516.

52. Zandman-Goddard, G.; Shoenfeld, Y., Ferritin in autoimmune diseases. *Autoimmun. Rev.* **2007,** 6, 457-463.

53. Linert, W.; Jameson, G. N. L., Redox reactions of neurotransmitters possibly involved in the progression of Parkinson's Disease. *J. Inorg. Biochem.* **2000,** 79, 319-326.

54. Kondo, T.; Shirasawa, T.; Itoyama, Y.; Mori, H., Embryonic genes expressed in Alzheimer's disease brains. *Neurosci. Lett.* **1996,** 209, 157-160.

55. Chou, C. C.; Gatti, R. A.; Fuller, M. L.; Concannon, P.; Wong, A.; Chada, S.; Davis, R. C.; Salser, W. A., Structure and expression of ferritin genes in a human promyelocytic cell-line that differentiates *in vitro*. *Mol. Cell. Biol.* **1986,** 6, 566-573.

56. Dorner, M. H.; Silverstone, A. E.; Desostoa, A.; Munn, G.; Desousa, M., Relative subunit composition of the ferritin synthesized by selected human lymphomyeloid cell-populations. *Exp. Hematol.* **1983,** 11, 866-872.

57. Cayre, Y.; Raynal, M. C.; Darzynkiewicz, Z.; Dorner, M. H., Model for intermediate steps in monocytic differentiation - C-Myc, C-Fms, and ferritin as markers. *Proc. Natl. Acad. Sci. U. S. A.* **1987,** 84, 7619-7623.

58. Chen, C. W.; Lorimore, S. A.; Evans, C. A.; Whetton, A. D.; Wright, E. G., A proteomic analysis of murine bone marrow and its response to ionizing radiation. *Proteomics* **2005,** 5, 4254-4263.

59. Hermani, A.; Hess, J.; De Servi, B.; Medunjanin, S.; Grobholz, R.; Trojan, L.; Angel, P.; Mayer, D., Calcium-binding proteins S100A8 and S100A9 as novel diagnostic markers in human prostate cancer. *Clin. Cancer Res.* **2005,** 11, 5146-5152.

60. Bartlett, J. A.; Hicks, B. J.; Schlomann, J. M.; Ramachandran, S.; Nauseef, W. M.; McCary, P. B., PLUNC is a secreted product of neutrophil granules. *J. Leukocyte Biol.* **2008,** 83, 1201-1206.

61. Sung, Y. K.; Moon, C.; Yoo, J. Y.; Moon, C.; Pearse, D.; Pevsner, J.; Ronnett, G. V., Plunc, a member of the secretory gland protein family, is up-regulated in nasal respiratory epithelium after olfactory bulbectomy. *J. Biol. Chem.* **2002,** 277, 12762-12769.

CHAPTER 6

Summary and Future Directions

In this dissertation, we developed several mass spectrometry-based chemical and quantitative proteomic methods for the identification and/or quantification of several specific classes of proteins, including nucleotide-binding proteins, nucleic acid-binding proteins and plasma membrane proteins, as well as for the proteomic analysis of the response of cells toward anticancer drug treatment.

In Chapter two, we described the synthesis and application of a biotin-conjugated acyl nucleotide probe for assessing adenosine nucleotide-binding proteins. This activity-based chemical probe can react specifically with the lysine residue at the nucleotide-binding site of purified adenosine nucleotide-binding proteins. The strategy, involving labeling reaction, enzymatic digestion, affinity purification and LC-MS/MS analysis, is relatively simple, fast, and straightforward. The method is useful for the enrichment and identification of nucleotide-binding proteins from complex biological samples. Quantitative analysis can be readily achieved by synthesizing stable isotope-coded chemical probes or by incorporating existing stable isotope labeling techniques, such as SILAC and iTRAQ.

In Chapter three, a strategy for the comprehensive investigation of DNA-binding proteins with *in vivo* chemical cross-linking and LC-MS/MS was developed. DNA-binding proteins can be isolated via the isolation of DNA-protein complexes and released from the complexes by reversing the cross-linking. By using this method, we were able to

identify more than one hundred DNA-binding proteins, including proteins involved in transcription, gene regulation, DNA replication and repair, and a large number of proteins potentially associated with DNA and DNA-binding proteins. This method should be generally applicable to the investigation of RNA-binding proteins, and hold great potential in the comprehensive study of gene regulation, DNA damage response and repair, as well as many other important biological processes at proteomic level.

In Chapter four, we described a strategy, including SILAC, cell-surface biotinylation, affinity peptide purification and LC-MS/MS, for the identification and quantification of cell-surface membrane proteins. Contaminations from cytoplasm and other nonmembrane related sources were greatly reduced by using *in situ* cell surface biotinylation and the subsequent affinity purification of biotinylated peptides. Metabolic labeling with SILAC, together with LC-MS/MS, facilitates the reliable quantification of surface membrane proteins expressed on two human melanoma cell lines derived from the same individual at different stages of tumor development. Integrins, cell adhesive molecules, CD antigens and receptors, which are essential for tumor development, were quantified in this work. Other proteins, such as histones, were also detected on the surface of WM-115 and WM-266-4 cells, and the cell surface localization of histone H2B and three other proteins was further confirmed by immunocytochemistry. The identification of aberrantly expressed proteins on the surface of metastatic melanoma cells sets the stage for the future investigation on the implications of these proteins in the progression of human melanoma.

In Chapter five, a comparative study of protein expression in human leukemia cells upon 6-thioguanine treatment was performed by using a MS-based proteomic method together with SILAC. The results revealed that the drug treatment led to elevated expression of major histones H2A, H2B, H3, H4 and H1. Other proteins, such as ferritin and S100A8, may also be involved in the cellular responses toward 6-thioguanine treatment. This study may offer new insights into the molecular mechanisms of action of this drug.