Revealing Differential Proteomic Events by Mass Spectrometry

by

Chenxi Yang

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2015

Date of final oral examination: 6/18/2015

The dissertation is approved by the following members of the Final Oral Committee:
Lingjun Li, Professor, Chemistry and Pharmacy
Lloyd M. Smith, Professor, Chemistry
John C. Wright, Professor, Chemistry
Emery Bresnick, Professor, Cell and Regenerative Biology
Xinyu Zhao, Professor, Neuroscience

Acknowledgements

I would like to firstly express my sincere gratitude to my advisor, Professor Lingjun Li. Without her support and guidance, I cannot finish most of the projects presented here and pursue my Ph.D. degree. Her passion in research motivated me to be a better scientist than what I expected.

I would like to thank my committee members: Dr. Lingjun Li, Dr. Lloyd M. Smith, Dr. John C. Wright, Dr. Xinyu Zhao and Dr. Emery Bresnick. I appreciate that they take time out of their busy schedules to provide valuable input during my Ph.D. work.

I would like to thank all members from the Li lab for all the help they provided throughout my graduate school. My special thanks go to Dr. Xuefei Zhong, Chuanzi Ouyang and Bingming Chen for their patience of listening to my ideas and providing scientific input. I especially want to thank a former Li lab member, Dr. Di Ma, who selflessly devoted a lot of time to mentoring me when I joined the group. I also thank Amanda Buchberger for editing my thesis.

I would also thank Dr. Yu Gao, Professor Xinyu Zhao, Dr. Koichi R Katsumura, Professor Emery Bresnick, Dr. Xudong Shi, Dr. Di Ma, Professor K. Craig Kent, Arick C. Park, Nicholas A. Davis, Dr. Jason D. Russell, Professor Greenspan and Professor Joshua Coon for their support and help in the collaborative projects. I also appreciate the instrument support from Dr. Cameron Scarlett at School of Pharmacy.

In particular, I would like to thank my parents. Without their constant love and support, it would not have been possible to complete this journey.

Table of Contents

Table of Figures

Table of Tables

List of Abbreviations and Acronyms

ACN	Acetonitrile
Ad	Adenoviral
AdGFP	Adenoviral vectors expressing green fluorescent protein
AdSmad3	Adenoviral vectors expressing Smad3
AGC	Automatic gain control
AGM	Aorta Gonad Mesonephros
AP	Affinity purification
aNSCs	Adult neural stem/progenitor cells
BGE	Background electrolytes
CE	Capillary electrophoresis
ChIP	Chromatin Immunoprecipitation
CID	Collision-induced dissociation
CNBr	Cyanogen bromide
Col(V)	Collagen type V
DHB	2,5-dihydroxybenzoic acid
DMEM	Dulbecco's modified eagle's medium
DTT	Dithiothreitol
ECD	Electron-capture dissociation
ECM	Extracellular matrix

ERLIC	Electrostatic repulsion-hydrophilic interaction chromatography
ESI	Electrospray ionization
ETD	Electron-transfer dissociation
FA	Formic acid
FDR	False discovery rate
FT	Fourier transform
Gal	Galactosyl
GFP	Green fluorescent protein
Glc	Glucosyl
GST	Glutathione-S-transferase
HA	Hemaglutinin
HCD	Beam-type CID
HILIC	Hydrophilic interaction chromatography
HSPS	High sensitivity porous sprayer
Hyl	Hydroxylysine
Hyp	Hydroxyproline
ICR	Ion cyclotron resonance
IGFBP	Insulin-like growth factor binding protein
IMAC	Immobilized metal affinity chromatography
IP	Immunoprecipitation
LC	Liquid chromatography

LC-ESI MS/MS	Liquid chromatography electrospray ionization tandem mass spectrometry
LDL	Low-density lipoprotein
MAE	Mouse Aortic Endothelial
MALDI	Matrix-assisted laser desorption ionization
MAPK	Mitogen-activated protein kinase
MBD1	Methyl-CpG binding protein 1
MOAC	Metal oxide affinity chromatography
MS	Mass spectrometry
MS^n	Tandem mass spectrometry
m/z	mass-to-charge ratio
N	Asparagine
NSAF	Normalized spectral abundance factor
NTA	Nitrilotriacetate
P3H	Prolyl 3-hydroxylase
P3H2	Prolyl 3-hydroxylase 2
P4H	Prolyl 4-hydroxylase
PD	Proteome discoverer
PDGF-BB	Platelet derived growth factor-BB
PDI-A3	Protein disulfide-isomerase A3
PolyMAC	Polymer-based metal ion affinity capture
Pro	Proline

PSMs	Peptide spectrum matches
PTMs	Post-translational modifications
Q	Glutamine
RPLC	Reversed Phase Liquid Chromatography
SAX	Strong anion exchange chromatography
SCX	Strong cation exchange chromatography
SILAC	Stable isotope labeling by amino acids in cell culture
Smad3	SMAD family member 3
SMCs	Smooth muscle cells
TGF-β	Transforming growth factor β
t-ITP	Transient isotachophoresis
TOF	Time-of-flight
VSMCs	Vascular smooth muscle cells
WAX	Weak anion exchange chromatography

Revealing Differential Proteomic Events by Mass Spectrometry

Chenxi Yang

Under the supervision of Professor Lingjun Li

At the University of Wisconsin-Madison

Abstract

The significance of proteins in biological processes has driven increasing numbers of studies focusing on proteomics in various biological systems during the past decade. These studies include: profiling proteins in simple or complex samples, investigating protein-protein interactions, characterizing post-translational modifications of proteins, and analyzing expression level of proteins in different biological or disease states. Such studies can provide valuable information to decipher the roles of proteins in numerous of biological processes and advance our understanding of mechanisms of various diseases. Mass spectrometry (MS), with its ability of parallel analysis of thousands of proteins, has become the standard technology to investigate proteome in recent years. In my dissertation research, I have employed MS-based tools to investigate differential proteomic events in various biological systems. The applications include: revealing the interacting proteins of MBD1 (chapter 3), characterizing phosphorylation of GATA-2 (chapter 4) and hydroxylation of α1 (V) collagen chain (chapter 6), and profiling secretome of vascular smooth muscle cells (VSMCs) in response to TGF-β/Smad3 stimulation

(chapter 5). Results presented here have provided insights into: 1) the function of MBD1 in neural development; 2) the mechanism of GATA-2 in regulating hematopoiesis; 3) the role of Smad3-dependent TGF-β signaling in the atherosclerotic process; and 4) the unique biological properties of α1 (V) collagen chain to induce the autoimmunity in lung transplantation. Collectively, this thesis research has developed improved proteomics methodology, showcased the power of MS-based tools, and enabled biological discovery in several important processes of biomedical and clinical relevance.

Chapter 1

Introduction: Brief Background and Research Summary

Partially adapted from: "Recent advances in enrichment and separation strategies for mass spectrometry-based phosphoproteomics." **Yang, C.**, Zhong, X., Li, L. *Electrophoresis* 2014, 35, (24), 3418-29

Mass Spectrometry

Over the last 15-20 years, MS-based proteomics has become a mainstream approach for protein identification due to its ability to globally profile the proteins within a complex biological system in a relatively high-throughput fashion.[1] In a mass spectrometer, the mass-to-charge ratio (*m/z*) of an ion/charged molecule in the gas phase is measured under vacuum. In high resolution mass spectrometers, the charge state of the measured molecules can be determined according to the distribution of their isotopes. The mass of the molecules therefore can be calculated based on the measured *m/z* value and charge state. To analyze the molecules of interest in a mass spectrometer, the molecules are firstly converted into ions in the ionization source. Once transferred into the vacuum region, the ions are separated according to their *m/z* by the mass analyzer. Finally, the ions are detected by an ion detector which enlarges electrical signals and these signals are transformed to mass spectra by a computer. Due to the general analysis process of molecules in MS, three basic instrumental components are standard: an ionization source, a mass analyzer and an ion detector.

An ionization source is utilized to produce ions from neutral molecules. Although various ionization approaches have been applied in MS analysis, electrospray ionization (ESI) was employed in all projects presented here. In addition to its high ionization efficiency, ESI can be directly coupled with liquid chromatography (LC) separation. The mobile phase eluted from an LC column is sprayed under a high electric field, and charged droplets are formed in the ion source.

As the solvent droplet evaporates, the radii of the charged droplet decrease. The shrinkage of the droplet volume results in the increasing charge density on the droplet surface. When the mutual Coulombic repulsion between charges on the surface of each droplet exceeds the surface tension, the droplet breaks down into a set of daughter droplets. This solvent evaporation-Columbic fission process happens repeatedly and individual gas-phase ions are generated.[2]

Ions generated in the ion source are then directed into the mass analyzer. The commonly used mass analyzers include time-of-flight (TOF), ion trap, Orbitrap, magnetic sector, quadrupole, and ion cyclotron resonance (ICR). The projects presented here were mainly performed on Orbitrap-type MS instrument platform because of their high resolution and high mass accuracy. In the Orbitrap-type of mass spectrometer, a spectrum of ions with wide *m/z* range are collected in a "C" trap and injected simultaneously to the Orbitrap analyzer.[3] Ions with different *m/z* oscillate in the Orbitrap at different frequencies, generating image currents. Then Fourier transform (FT) is utilized to deconvolute the synthesized waveform resulted by the motion of ions at different frequencies. The signals are then transformed into a mass spectrum through a computer.

To map the sequence of a peptide, a full MS scan is performed to measure the intact mass of the peptide followed by isolation of the peptide of interest's ions for MS/MS analysis. In MS/MS analysis, the isolated peptide ions are fragmented into smaller pieces through collisions with neutral molecules (collision-induced dissociation) (CID)) or ion-ion reactions (electron-transfer dissociation (ETD)[4],

electron-capture dissociation (ECD)[5]). Sequence information can be extracted from MS/MS spectra by matching the *m/z* of fragment ions to the theoretical *m/z* of the ions with certain chemical structure (Figure 1.1).

Characterization of Post-translational Modifications (PTMs) by Mass Spectrometry

About 5% of genomes in eukaryotes are coded for enzymes that are dedicated to PTMs of proteome. Understanding of dynamic PTMs of proteins in various biological systems can facilitate deciphering the signaling transduction in biological processes and discovering the therapeutics for various diseases. Characterization of PTMs by MS is focused on two aspects: 1) localization of the modified sites (e.g., phosphorylation, acetylation, glycosylation, and hydroxylation) in a single protein or large-scale manner and 2) characterization of the structure of PTMs themselves, including uncovering the structure of a glycan carried by the proteins of interests.[6] Projects presented here are mainly focusing on the localization of PTMs in proteins. Therefore, I will take phosphorylation as an example to illustrate how MS is utilized for localization of PTMs in proteins.

In a general "bottom-up" phosphoproteomic workflow (Figure 1.2), protein mixtures are extracted from cells or tissues and digested by proteases. Trypsin is the most widely used enzyme. However, Glu-C, Lys-C, Asp-N, or other enzymes are also used alone or with trypsin to achieve in-depth phosphoproteomic analysis according to the need of individual projects.[7,8] In some of studies, an additional off-line separation is employed to reduce the complexity of samples. For example, the generated peptide mixture is separated by off-line strong cation exchange

chromatography (SCX). The eluted samples are collected and combined into tens of fractions according to their elution profiling. An enrichment technique, like immobilized metal affinity chromatography (IMAC), is then utilized to isolate phosphopeptides from the mixtures and facilitates the detection of phosphopeptides in subsequent LC-MS/MS analysis. Details on enrichment and separation strategies applied in MS-based phosphoproteomics are described in Chapter 2. Finally, phosphopeptides are identified by a database search, and phosphorylation sites are confidently assigned by a site localization algorithm.

During MS analysis, ions will be fragmented into small pieces in order to obtain sequence information. In MS, a cycle will be conducted in which a full MS survey scan is performed to measure the intact mass of peptides and the most intense ions are selected for MS/MS analysis to map the sequence of peptides and localization of phosphorylation sites. CID is the most widely used fragmentation methods for peptide sequencing, but phosphopeptides are often preferentially fragmented at the phosphate group in CID. This gives rise to nonsequence neutral and charged losses from the precursor and sequence-specific product ions,[9] limiting the identification of phosphopeptides and phosphorylation site localization. In comparison to CID, high-energy collision dissociation was shown to identify more phosphopeptides and phosphorylation sites, as it produces less abundant neutral loss and more interpretable sequence informative product ions.[10,11] ETD and ECD, unlike CID or high-energy collision dissociation, fragment peptide ions independently of the sequence of peptides. As a result, labile PTMs, such as phosphate groups, will stay

intact using these two approaches,[4] however, these electron based fragmentation techniques require the presence of multiply charged peptide ions.[12] Furthermore, ECD requires the use of FT-ICR instrumentation. These limitations prevent their broader usage in large-scale phosphoproteomics.

With thousands of tandem MS spectra identified by database searches, phosphorylation sites on each identified peptide need to be verified. Unfortunately, it is impossible to validate the assignments of phosphorylation sites by manual inspection of thousands of MS/MS spectra. To address this challenge, site localization algorithms, such as A score,[13] PhosphoRS,[14] and MASCOT delta score,[15] have been introduced in the field of phosphoproteomics to facilitate the validation process.

Analysis of Protein-Protein Interactions by Mass Spectrometry

Biological processes in an organism are usually organized and coordinated through the dynamic signaling networks of interacting proteins.[16] Thus, the simply identification of proteins in the organism are not sufficient to describe biological processes and understand protein functions, which have made characterization of protein-protein interactions become critical in biological research. Recently, affinity purification (AP) coupled with MS has been widely used to advance our understanding of protein-protein interactions due to their ability to isolate protein complexes under near physiological conditions and perform the analysis in a simple and fast manner with high selectivity and sensitivity.[17] In a basic workflow of AP-MS, the target protein (bait) and its interacting proteins (prey) are isolated from the lysates of cells or tissues. During the isolation, the target protein is recognized and captured

by a capture molecule immobilized on a solid support (resin). Because of the interactions with target protein, the interacting proteins are simultaneously captured. The isolated protein complexes are then eluted and analyzed by MS after few washing steps to reduce nonspecific interactions (Figure 1.3). Although isolation of endogenous proteins can be performed, such isolation is limited by the number of commercial available antibodies to recognize the proteins of interest. An alternative is to isolate proteins through an affinity epitope tag. Proteins of interest are expressed with an epitope tag. It thus enables the purification of target protein along with its interacting proteins through an antibody specific to the tag. The common tags used in AP-MS include FLAG, six histidines, c-Myc, hemaglutinin (HA), green fluorescent protein (GFP), glutathione-S-transferase (GST), and protein A.[18,19] Because the usage of tags is universally applicable for proteins of interest and achieves parallel sample preparation without additional optimization of the protocol for individual protein complex of interest, it is suitable for large-scale and high-throughput experiments.[20,21]

In theory, proteins that have specific interactions with the target protein can only be co-precipitated during the affinity purification. However, contaminants always present. They are co-precipitated through the nonspecific interactions with the resin (e.g., magnentic beads, agarose), immunoglobulin molecules (e.g., heavy or light chains of antibodies), and tags (e.g., FLAG, GFP). Even worse, some of them can bind to the isolated proteins of interest through nonspecific interactions. To distinguish the interacting proteins from contaminants, control experiments are required for all AP-MS experiments. Further quantitative MS analysis can provide

valuable information to decide the specificity of interactions by revealing the relatively high abundant proteins in the samples compared to a control. Once the interacting proteins for the target protein are identified, validation experiments must be performed.

Mass Spectrometry based Quantification Approaches

To date, the growing demand for information on differential expression of proteins in various biological systems under different physiological or pathological conditions has driven the field of proteomics from qualitative analysis to quantitative analysis.[1] Quantitative proteomics can be classified into two major approaches: label-based and label-free approaches. Labeling approaches, which incorporate isotopes into proteins or peptides and perform quantification analyses in MS (stable isotope labeling by amino acids in cell culture (SILAC)) or MS/MS (isobaric tags) scans, are often considered to be more accurate in relative protein quantification. However, the required expensive isotope labels and the limited number of samples analyzed in a single experiment make the label-free approach as a more cost-effective method for quantitative proteomics. Label-free quantitation is generally based on two categories of measurements: 1) ion intensity such as peptide peak areas or peak heights in chromatography, which are proportional to the concentration of measured peptide, and 2) spectral count based on the observation that an increase in protein abundance is associated with an increased number of proteolytic peptides and their frequencies of being selected to produce MS/MS spectra. Because of the different features of

these quantification approaches, they have been utilized in various studies dependent on the need of individual study.[22,23]

Overview of Projects

My thesis projects presented here are focused on utilization of MS to investigate the different proteomic events in various biological systems. In chapter 2, a review on recent advances for enrichment and separation techniques in MS-based phosphoprotemics is given. In chapter 3, I applied MS to reveal the interacting proteins of MBD1 in mouse brain because of the importance of MBD1 in regulation of neutral development. A MBD1 mouse model was generated. A series of AP experiments were then conducted to isolate the MBD1 complexes from lysate of mice brain. Control groups and quantification analysis were performed to uncover the potential interacting proteins of MBD1 from contaminants. Quantification analysis via spectral counting found five proteins as significant potential interacting of MBD1.

The projects described in chapter 4 and 6 relied on MS to map PTMs of proteins. In chapter 4, the phosphorylation of GATA-2 was analyzed by MS. Results from western blots indicated that the wild-type GATA-2 is phosphorylated, while GATA-2 (T354M) is hyperphosphorylated. Therefore, the wild-type and mutant of GATA-2 were isolated separately through immunoprecipitation, and further purifications were performed by SDS-PAGE separation. The phosphorylated GATA-2 proteins (wild-type and mutant) were digested in-gel. Peptides generated were extracted and subjected to LC-MS/MS analysis. Here, we revealed the S192-dependent phosphorylation of GATA-2. In chapter 6, a comprehensively mapping

hydroxylation of α1(V) collagen chain was demonstrated. α1(V) collagen chain in human and bovine were digested by multiple proteases to achieve high sequence coverage in MS analysis. Beam-type CID (HCD) was selected as the main fragmentation approach because of its ability to generate immonium ions and "proline-effect" in ETD fragmentation. In the MS analysis, about 89% and 95% sequence coverage were achieved separately in human pro-α1 (V) chain and bovine α 1 (V) chain. In total, 120 hydroxyproline residues, 8 hydroxylysines, and 23 glycosylated hydroxyprolines were identified in human pro-α1 (V) chain. Meanwhile, 124 hydroxyprolines, 3 hydroxylysines, and 31 glycosylated hydroxyprolines were found in bovine α 1 (V) chain. The comprehensively map of hydroxylation of α1(V) collagen chain in human and bovine has provided valuable information to perform downstream functional analysis for hydroxylation of α1(V) collagen chain. Eventually, such studies may explain the role of hydroxylation carried by α1(V) collagen chain in autoimmune response against collagen type V in lung transplant.

Chapter 5 presents the comparative secretome analysis of vascular muscle cells in response to Smad3-dependent TGF-β signaling. The conditioned media from control and treatment were collected. Secreted proteins in the media were isolated and concentrated by molecular weight cut-off filter. After tryptic digestion and desalting, peptide mixtures were analyzed by LC-MS/MS. A substantially larger list of putative secreted proteins is identified. Further quantification analysis by spectral counting revealed 38 secreted proteins that were significantly up or down-regulated in response to TGF-β. A conclusion chapter is then included in the end of my thesis.

REFERENCES

1. Boja, E. S.; Rodriguez, H., Mass spectrometry-based targeted quantitative proteomics: achieving sensitive and reproducible detection of proteins. *Proteomics* **2012,** *12* (8), 1093-110.

2. Hoffmann, E. d.; Stroobant, V., *Mass spectrometry : principles and applications*. 2nd ed.; Wiley: Chichester ; New York, 2002; p xii, 407 p.

3. Perry, R. H.; Cooks, R. G.; Noll, R. J., Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass spectrometry reviews* **2008,** *27* (6), 661-99.

4. Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (26), 9528-33.

5. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc* **1998,** *120* (13), 3265-3266.

6. Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angewandte Chemie* **2005,** *44* (45), 7342-72.

7. Wisniewski, J. R.; Mann, M., Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Analytical chemistry* **2012,** *84* (6), 2631-7.

8. Bian, Y.; Ye, M.; Song, C.; Cheng, K.; Wang, C.; Wei, X.; Zhu, J.; Chen, R.; Wang, F.; Zou, H., Improve the coverage for the analysis of phosphoproteome of HeLa cells by a tandem digestion approach. *Journal of proteome research* **2012,** *11* (5), 2828-37.

9. Palumbo, A. M.; Smith, S. A.; Kalcic, C. L.; Dantus, M.; Stemmer, P. M.; Reid, G. E., Tandem mass spectrometry strategies for phosphoproteome analysis. *Mass spectrometry reviews* **2011,** *30* (4), 600-25.

10. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods* **2007,** *4* (9), 709-12.

11. Nagaraj, N.; D'Souza, R. C.; Cox, J.; Olsen, J. V.; Mann, M., Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *Journal of proteome research* **2010,** *9* (12), 6786-94.

12. Swaney, D. L.; McAlister, G. C.; Coon, J. J., Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nature methods* **2008,** *5* (11), 959-64.

13. Beausoleil, S. A.; Villen, J.; Gerber, S. A.; Rush, J.; Gygi, S. P., A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nature biotechnology* **2006,** *24* (10), 1285-92.

14. Taus, T.; Kocher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K., Universal and confident phosphorylation site localization using phosphoRS. *Journal of proteome research* **2011,** *10* (12), 5354-62.

15. Savitski, M. M.; Lemeer, S.; Boesche, M.; Lang, M.; Mathieson, T.; Bantscheff, M.; Kuster, B., Confident phosphorylation site localization using the Mascot Delta Score. *Molecular & cellular proteomics : MCP* **2011,** *10* (2), M110 003830.

16. Pawson, T.; Scott, J. D., Signaling through scaffold, anchoring, and adaptor proteins. *Science* **1997,** *278* (5346), 2075-80.

17. Miteva, Y. V.; Budayeva, H. G.; Cristea, I. M., Proteomics-based methods for discovery, quantification, and validation of protein-protein interactions. *Analytical chemistry* **2013,** *85* (2), 749-68.

18. Brizzard, B., Epitope tagging. *BioTechniques* **2008,** *44* (5), 693-5.

19. Trinkle-Mulcahy, L., Resolving protein interactions and complexes by affinity purification followed by label-based quantitative mass spectrometry. *Proteomics* **2012,** *12* (10), 1623-38.

20. Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sorensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W.; Figeys, D.; Tyers, M., Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **2002,** *415* (6868), 180-3.

21. Terpe, K., Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology* **2003,** *60* (5), 523-33.
22. Ning, Z.; Zhou, H.; Wang, F.; Abu-Farha, M.; Figeys, D., Analytical aspects of proteomics: 2009-2010. *Analytical chemistry* **2011,** *83* (12), 4407-26.

23. Shen, X.; Young, R.; Canty, J. M.; Qu, J., Quantitative proteomics in cardiovascular research: global and targeted strategies. *Proteomics. Clinical applications* **2014,** *8* (7-8), 488-505.

24. Yang, C.; Zhong, X.; Li, L., Recent advances in enrichment and separation strategies for mass spectrometry-based phosphoproteomics. *Electrophoresis* **2014,** *35* (24), 3418-29.

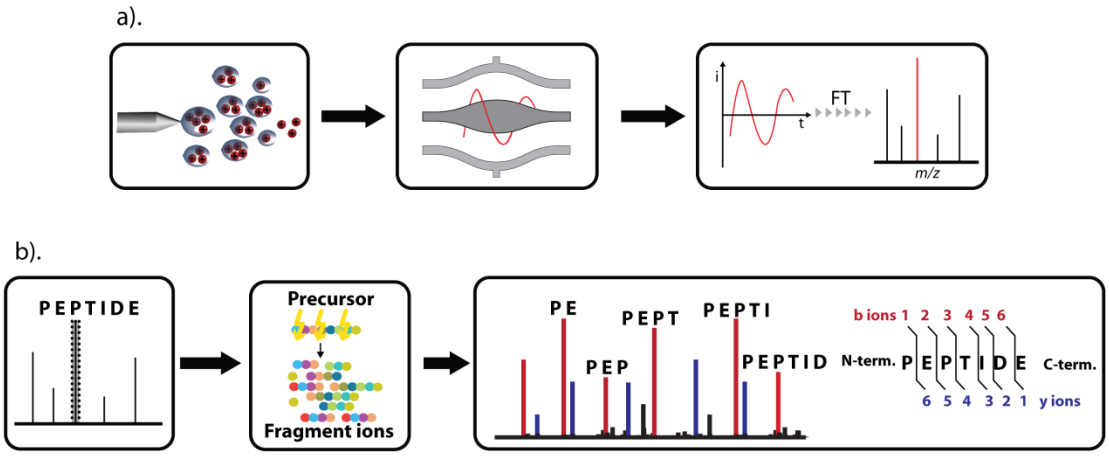


Figure 1.1 A general schematic of an Orbitrap and MS, MS/MS analysis. a). The general schematic of an Orbitrap mass spectrometer. Ions are generated in the ions source through ESI and transferred into the Orbitrap. Once in Orbitrap, ions with various m/z oscillate, resulting in the generation of image currents. A spectrum is produced in a computer through Fourier transform. b). Visualization MS and MS/MS analysis. The intact mass of peptide ions is measured in a full MS scan. The intact peptide ions with specific m/z are selected and fragmented to generate MS/MS spectrum. In MS/MS spectrum, the sequence of the selected peptides can be deciphered by matching the m/z value in the spectrum to the theoretical m/z value of a single or multiple amino acids.

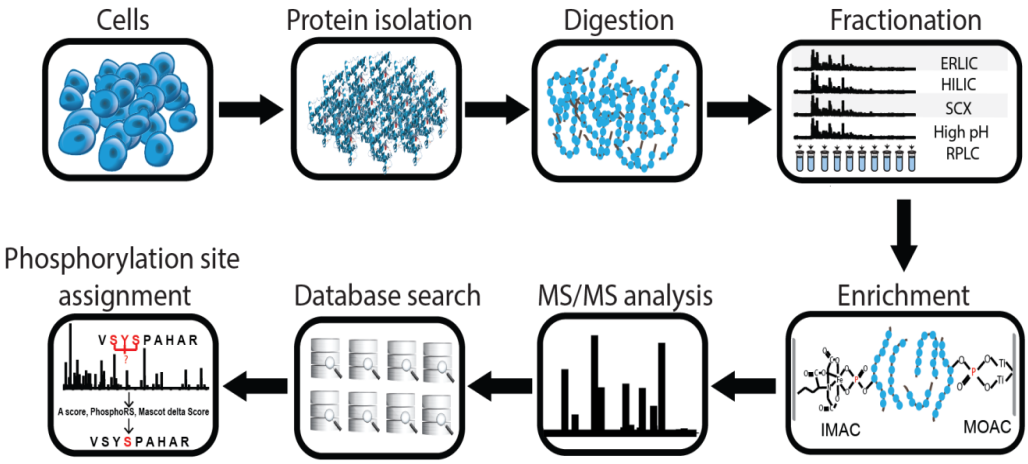


Figure 1.2 A general "bottom-up" phosphoproteomic workflow depicting the major steps, which comprise protein extraction from cells, enzymatic digestion, and fractionation of the resulting peptide mixtures. Phosphopeptides in each fraction are then enriched and subjected to LC-MS/MS analysis. Finally, phosphopeptides are identified by a database search and phosphorylation sites are confidently assigned by a site localization algorithm.[24] (ERLIC: electrostatic repulsion-hydrophilic interaction chromatography; HILIC: hydrophilic interaction chromatography; MOAC: metal oxide affinity chromatography)

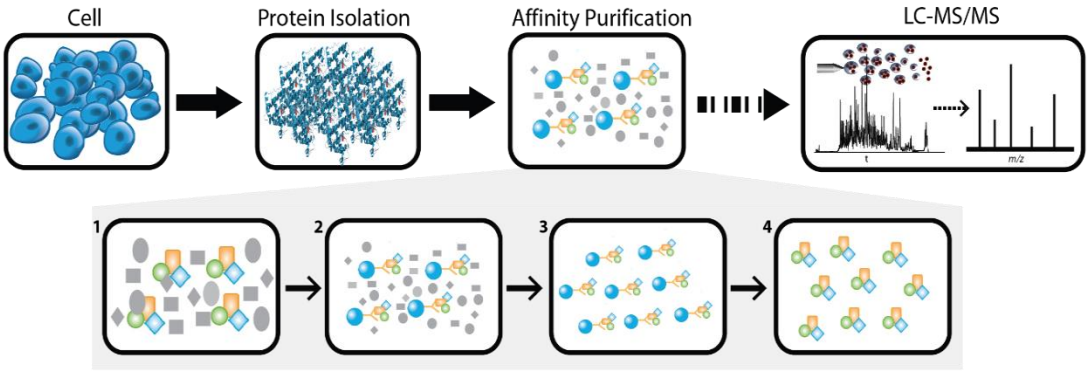


Figure 1.3 A basic workflow of AP-MS. Initially, proteins are isolated from cells. The proteins of interest and their interacting proteins are then precipitated by AP. After a series of sample preparation steps, samples are analyzed by LC-MS/MS. The further quantification and statistical analysis reveal the potential interacting proteins from contaminants. 1. Lsate from cells or tissues; 2. Incubation of capture molecules (antibodies here) with lysate; 3. Washing steps to remove nonspecific binding proteins; 4. Elution of protein complexes of interest.

Chapter 2

Recent Advances in Enrichment and Separation Strategies for Mass Spectrometry-based Phosphoproteomics

Adapted from: "Recent advances in enrichment and separation strategies for mass spectrometry-based phosphoproteomics." **Yang, C.**, Zhong, X., Li, L. *Electrophoresis* 2014, 35, (24), 3418-29

ABSTRACT

Due to the significance of protein phosphorylation in various biological processes and signaling events, new analytical techniques for enhanced phosphoproteomics have been rapidly introduced in recent years. The combinatorial use of the phospho-specific enrichment techniques and prefractionation methods prior to MS analysis enables comprehensive profiling of phosphoproteome and facilitates deciphering the critical roles that phosphorylation plays in signaling pathways in various biological systems. This review places special emphasis on the recent five-year (2009-2013) advances for enrichment and separation techniques that have been utilized for phosphopeptides prior to MS analysis.

INTRODUCTION

Protein phosphorylation is one of the most common post-translational modifications (PTMs) in various organisms. One-third of proteins in eukaryotic cells are estimated to be phosphorylated at any time.[1] The reversible phosphorylation of proteins plays critical roles in the regulation of intracellular biological processes, such as signal transduction, transcription and translation regulation, and metabolism.[2] Aberrant phosphorylation can result in diseases including cancer and cardiovascular disease.[3,4] Due to the significance of protein phosphorylation in biological processes, tremendous efforts have been made to investigate protein phosphorylation for decades.[5] Mass spectrometry (MS) as a promising tool to analyze protein phosphorylation has gained great deal of attention due to its ability to profile thousands of proteins in a single analysis.[1, 6] However, analysis of protein phosphorylation is not straightforward. The low stoichiometry, wide dynamic range and various isoforms of phosphorylated proteins present in the biological systems pose significant challenges to current analytical techniques. Fortunately, these challenges can be alleviated to some extent by enriching phosphoproteins or phosphopeptides prior to MS analysis.[7] Enrichments also increase the sensitivity of MS to detect phosphorylated proteins or peptides by reducing the ionization competition with non-phosphorylated proteins or peptides. [8]

To increase the number of identified proteins from a biological sample, multidimensional separation strategies are commonly utilized in current MS analysis in order to reduce the complexity of samples.[9] In a typical MS-based

phosphoproteomic analysis, a prefractionation step and enrichment step are often employed before MS analysis.[10] Although some prefractionation methods can partially enrich phosphopeptides or phosphoproteins from biological samples, such as ion-exchange based chromatography, we categorize those techniques as separation strategies for phosphoproteins or phosphopeptides. Herein, we will focus on the advances in enrichment and separation techniques for MS-based phosphoproteomics that have been published in recent years during 2009-2013. The citations are not comprehensive due to vast amount of literature and explosive rate of growth in the field. Rather, we attempt to highlight some representative examples from recent work and advances.

Advances in Enrichment Techniques

A variety of enrichment techniques, including immunoprecipitation (IP), chemical modification, immobilized metal affinity chromatography (IMAC), and metal oxide affinity chromatography (MOAC), have been developed and applied to study the phosphoproteome in different biological samples.[3, 11] Among these techniques, chemical modifications, such as β-elimination coupled with a Michael addition, often suffer from incomplete reactions resulting in significant sample loss and increased sample complexity. Despite an effective enrichment strategy, IP is mainly used for enrichment of phosphotyrosine-containing proteins.[12,13] Thus, we will focus our discussion primarily on IMAC and MOAC in the following sections.

Immobilized Metal Affinity Chromatography

Currently, immobilized metal affinity chromatography (IMAC) is one of the most widely used enrichment techniques for phosphopeptides prior to MS analysis. It enriches phosphopeptides by the affinity of positive charges of metal ions for negative charges of phosphate. Fe^{3+}, Al^{3+}, Ga^{3+}, Zr^{4+} and Ti^{4+} are the commonly used metal ions in IMAC.[14] In IMAC, metal ions are chelated to a support, such as magnetic beads, stationary phase in a chromatographic column or MALDI plate, via a chelating group. However, IMAC often suffers from high level of non-specific binding for non-phosphorylated peptides that contain acidic residues (glutamic and aspartic acid). Low pH loading and washing buffers are usually used in IMAC enrichment to improve the specificity of enrichment. The pH of loading and washing buffer is often between the pK_a values of acidic amino acids and pK_{a1} value of phosphoric acids and results in the protonation of acidic residues and deprotonation of phosphate groups. Thus, negative charge of phosphopeptides will be retained in IMAC resin while neutral acidic non-phosphorylated peptides will be washed away from IMAC resin during loading and washing steps.[13, 15] But complete protonation of all acidic residues can only be achieved in highly acidic pH (pH=1 to 1.5) and such low pH condition could result in metal ions leaching from supports and contaminating phosphopeptides.[16] Another way to overcome this weakness is esterification of phosphopeptides, which converts carboxyl groups of amino acid residues into methyl esters.[17] But sample losses and increased sample complexity, resulted by incomplete reactions and side reactions, prevent the wide spread usage of esterification.[18]

Significant efforts have been made to address these problems by optimizing the IMAC materials or enrichment protocols. Herein, we will highlight recent advances of IMAC from these two aspects.

Advances in Materials for Immobilized Metal Affinity Chromatography

The improvements on IMAC materials include selection of different metal ions, chelating group and supports. Stationary phases of LC columns are the common supports for IMAC. Recently, various metal ions were immobilized onto monolithic columns.[19] Because of the porous structure, monolith can provide larger surface area with high specificity to bind more metal ions and enable rapid mass transport for phosphopeptides. A Ti^{4+}-monolithic based IMAC was used to perform phosphoproteomic study of rat liver mitochondrion.[19a] The sol-gel method with tetraethoxysilane and 3-aminopropyltriethoxysilane as precursors was used to prepare the monolithic support in a 250 μm i.d. capillary. Ti^{4+} ions were chelated onto the monolithic support by further functionalizing amine groups. A total of 224 phosphopeptides were identified in comparison with 28 and 72 phosphopeptides which were identified by commercial Fe^{3+}-IMAC and TiO_2, respectively.

However, enriching phosphopeptides in an LC column usually requires multiple sample-handling steps and an LC system to achieve sample loading, washing and eluting. These limitations motivated researchers to design new supports to simplify the procedure of phosphopeptide enrichment. Magnetic beads have attracted a great deal of interest due to its simplicity for isolating phosphopeptides from a sample solution by an external magnetic field and provide a high throughput platform

for automated enrichment processes.[20] Although magnetic non-porous beads immobilizing with different metal ions have been successfully used for various biological samples and are commercially available, magnetic mesoporous beads with high density pores and large surface areas have gained increasing attention in the past five years. [21,22] Zirconium-containing magnetic mesoporous beads were synthesized and applied for enriching phosphopeptides.[22b] The magnetic mesoporous beads were prepared by directly coating mesoporous silica on to Fe_3O_4 magnetic microspheres following addition of phosphate to chelate with Zr^{4+} ions. This Zr^{4+}-functionalized materials facilitated identification of 218 phosphopeptides in 100 μg rat brain which showed excellent potential for selective enrichment of phosphopeptides.

Directly enriching phosphopeptides on a modified MALDI plate has been developed and successfully applied to enable rapid enriching of a small amount of samples.[14c, 23] Phosphopeptides in 20 fmol purified β-casein digests were successfully enriched by a functionalized MALDI plate with a phosphonate self-as-sembled monolayer immobilizing to Zr^{4+} ions.[23c] The sensitivity is significantly higher than TiO_2 or Fe-IMAC microcolumns or magnetic beads due to reduced sample loss caused by additional sample transfer steps. Recently, Si wafers with a micro-array of functionalized microspots was fabricated as a MALDI plate. [23a] In each microspots, Fe^{3+} ions were immobilized onto Si wafers via nitrilotriacetate (NTA) which provides a potential platform for high throughput parallel enrichment of phosphopeptides.

Further, to immobilizing metal ions onto a solid support, polymer-based metal ion affinity capture (PolyMAC), which immobilizes metal ions to soluble

nanopolymers, was recently introduced.[24] This technique allows for enriching limited phosphopeptides in a homogeneous, aqueous solution which overcomes the inconsistent enrichment resulting from the heterogeneity of solid phase extraction-based isolation methods. Phosphopeptides are typically bound to PolyMAC-Ti agents and isolated from solution through coupling to hydrazide-agarose gels. The entire enrichment process, including capture of phosphopeptides in solution, recovery of PolyMAC-phosphopeptide complexes on agarose beads, wash off non-phosphorylated peptides, and elution of phosphopeptides, requires less than 30 min and provides higher selectivity, reproducibility and greater yields than commonly used TiO_2 method.

In addition to the improvements on IMAC supports, an increasing number of studies have focused on the improvements of chelating groups. [19a, 22b, 23c, 25] The phosphate group is recently employed in IMAC as a promising metal ion chelator. Unlike in NTA-IMAC or imidodiacetic acid-IMAC method that each metal ion only coordinates to one ligand, each metal ion binds to more than one phosphate group producing strong binding between metal ions and ligands.[26] Ti or Zr ions immobilized on the phosphate materials were reported to be coated onto various supports, including MALDI plates, magnetic beads, celluloses and monolithic columns.[19a, 22b, 23c, 25a] Moreover, many new types of complexes have also been utilized in IMAC, such as dinuclear zinc (II) complex Phos-tag and polydopamine. [25b-d]

Advances in Protocol for Immobilized Metal Affinity Chromatography

A recent study discovered that IMAC not only can enrich phosphopeptides, but also can bind to non-phosphorylated peptides with deamidation or carbamylation.[27] Therefore, it is recommended to avoid using urea, a commonly used denaturant, in phosphoprotemic studies due to its tendency to induce these modifications (deamidation and carbamylation) on peptides. Additionally, because nucleic acids contain abundant phosphate groups that have strong affinities for IMAC beads causing interference for phosphopeptides enrichment, acetonitrile (ACN) precipitation has been employed for protein extraction to efficiently remove nucleic acids before IMAC enrichment step.[28]

To comprehensively map the heterogeneous types of phosphopeptides, different strategies have been developed for IMAC enrichment.[29] A tandem IMAC-IMAC strategy (Figure 2.1) was utilized to address the problem of relatively low number of identified monophosphorylated peptides in IMAC enrichment due to its weak interaction with monophosphorylated peptides.[29a] To recover the monophosphorylated peptides in the washing fractions from the first IMAC enrichment, a second IMAC enrichment step was performed to capture these unbound phosphopeptides. This tandem IMAC strategy yielded about 60% increase in the number of phosphopeptides identified in comparison to one-step IMAC enrichment. Furthermore, more than 90% of the identified phosphopeptides in the second IMAC fractions were monophosphorylated peptides. The combined use of different metal ions for enriching phosphopeptides in a single experiment has also been demonstrated

[29b, 29c]. The Fe^{3+}-IMAC and Ti^{4+}-IMAC methods could complementarily enrich phosphopeptides in Raji cell lysate with only 10% overlapping.[29b] Sequential enrichment of phosphopeptides from Raji cell lysate by Ga^{3+}-Fe^{3+}IMAC also resulted in more than 1.5-fold greater coverage of phosphoproteome compared to that of single IMAC (Fe^{3+}, Ti^{4+}, Ga^{3+} and Al^{3+}).[29c]

Metal Oxide Affinity Chromatography

In comparison with IMAC, MOAC has a better tolerance with low pH loading and washing buffers which efficiently protonate carboxyl group while keeping the negative charges on phosphorylated residues.[16] This technique is based on the affinity of oxygen in the phosphate groups for metal atoms in the MOAC resin. Peptide mixtures are usually loaded onto the column under acidic conditions and the phosphopeptides are eluted by addition of basic solutions. Various metal oxides including TiO_2, ZrO_2, Fe_3O_4, SnO_2, HfO_2 and CeO_2 have been used for phosphopeptide enrichment and exhibit different affinity and specificity for phosphopeptides.[23a, 30] Turecek group and Ge group independently demonstrated that ZrO_2 outperforms TiO_2.[31] However, the variations in phosphopeptide enrichment capabilities are highly dependent on the preparation techniques of metal oxide as the morphological properties of the solid coating, in addition to the differences of metal centers, can also significantly contribute to the enrichment capability of metal oxide. The performance of MOAC for enriching phosphopeptides is also influenced by the physical characteristics of the MOAC resin. Nanoparticles generate superior enrichment compared with microparticles due to larger surface area with high

specificity to provide more ligand binding sites.[32] Herein, we will also cover recent advances of the MOAC in materials and enrichment protocols.

Advances in Materials for Metal Oxide Affinity Chromatography

In the recent five years, substrates with various structures and different compositions were synthesized to improve the enrichment performance of MOAC.[33] TiO_2 and iron oxide were coated onto a monolithic capillary column to enrich phosphopeptides from casein or human serum.[34] Mesoporous TiO_2 and ZrO_2 aerogel were recently designed for selective enrichment of phosphopeptides in rat liver mitochondrion.[35] The binding capacity of mesoporous TiO_2 aerogel was six times higher than that of conventional TiO_2 microparticles for a single phosphopeptide standard. A total of 216 phosphoprotein groups was identified by mesoporous TiO_2 aerogel from rat liver mitochondrion with comparison to less than 150 phosphoprotein groups identified by TiO_2 nanoparticles or microparticles.[35b] Magnetic particles decorated with metal oxide have also been widely applied for enriching phosphopeptides.[36] Magnetic carbon-encapsulated iron nanoparticles were functionalized with poly (acrylic acid) to produce carboxyl-rich polymer surface which greatly facilitated the immobilization of TiO_2.[36c] The magnetic TiO_2-coated carbon-encapsulated iron nanoparticles enabled identification of 1415 unique phosphopeptides and 1093 phosphorylation sites from 200 μg of HeLa cell lysates. This performance was significantly superior to that of commercial magnetic TiO_2 beads.

Although metallic MALDI plates are the main supports used for on-plate enrichment of phosphopeptides, the high cost of these metallic plates makes alternative substrates attractive such as glass slides. A mesoporous TiO_2 stripe coated microscope glass slide not only provided a platform for on-target enrichment, but also enabled separation of phosphopeptides based on the number of phosphate groups on the peptides.[37] Ammonium dihydrogen phosphate solution was used to elute phosphopeptides. In a similar way to thin layer chromatography, the stripe was placed in a chamber containing elution buffer and mono- and multiphosphorylated peptides of β-casein were clearly separated. TiO_2 arrays were also fabricated on a gold-covered glass slide via photocatalysis and provided the capability for high-throughput processing.[38] Furthermore, single-use TiO_2-coated aluminum foil was introduced for on-plate enrichment of phosphopeptides as an economical disposable layer attached to MALDI plates.[39] This platform avoided the time-consuming polishing and washing steps to regenerate MALDI plates and was used for matrix-free laser desorption/ionization of peptides with sample amounts down to several hundred femtomoles.

Advances in Protocol for Metal Oxide Affinity Chromatography

To improve the enrichment performance of MOAC, considerable efforts have focused on the optimization of sample loading and elution buffers. In addition to conventional additives including 2,5-dihydroxybenzoic acid (DHB), glutamic acid and lactic acid, which contain carboxyl groups, glycerol was demonstrated to improve the phosphopeptide selectivity of TiO_2[40,41,42, 43]. It is speculated that glycerol adding

into loading and washing buffers could reduce acidic peptides binding to TiO_2 by disruption of carboxy group-based bidentate chelation. It is also demonstrated that NH_4OH, a common elution solution, tended to elute shorter phosphopeptides (1000-1500 Da), while bis-Tris propane was suitable for elution of longer phosphopeptides (1000-4000 Da), in some cases, which contained more hydrophilic and/or more acidic residues. Therefore, a best optimized method which combined the glycerol additive and two-step elution (NH_4OH and bis-Tris propane) was applied to enrich phosphopeptides in PC2 cell lysate. In comparison to the conventional method, a 1.4-fold increase in the number of identified phosphopeptides was obtained (Figure 2.2). Further to these results, glycerol additive enabled recovery of more monophosphorylated peptides than lactic acid.[43] Recently, asparagine (N) or glutamine (Q)-rich peptides, which did not show a propensity to binding to IMAC beads, were proven to exhibit non-specific binding to TiO_2 beads.[44] Furthermore, phosphopeptide enrichment was particularly low for yeast cells compared to flies and human. The low specificity in yeast cells was caused by the relative higher abundance of N/Q-rich proteins in yeast (2.7%) than that in flies (1.5%) and human (1.3%). A mixture of asparagine and glutamine amino acids as decoy amino acids were added into washing buffers to reduce the extent of non-specific peptide binding and improve the recovery and detection of low abundance phosphopeptides. A 30% increase in the enrichment of phosphopeptides was observed compared to that of washing buffers without decoy amino acids.

Peptide-to-metal oxide beads ratio is also a significant factor for phosphopeptide enrichment (Figure 2.3).[45] Insufficient or excessive TiO2 beads could decrease the selectivity. While excessive TiO_2 beads could increase non-specific binding, insufficient TiO_2 beads favored enrichment of multiphosphorylated peptides due to their higher affinity compared to monophosphorylated peptides. Therefore, pre-experiments are recommended to determine an optimum peptide-to-beads ratio when handling different biological samples. It is also suggested that incubation of samples in insufficient beads could be utilized as a strategy to separate multiphosphorylated peptides from monophosphorylated peptides. Separate analysis of mono- and multiphosphorylated peptides can enhance the detection of multiphosphorylated peptides as monophosphorylated peptides can suppress the signal of multiphosphorylated peptides. A pre-separation of mono- and multiphosphorylated peptides could also be achieved by eluting phosphopeptides at different high pH conditions.[46] Moreover, citric acid also could be used for stepwise separation of phosphopeptide since concentration of citric acid can greatly affect the binding of mono- and multiphosphorylated peptides with TiO_2.[47] 1 mg of HeLa cell digests was loaded onto TiO_2 beads in 1M citric acid containing loading buffer. The flow-through fraction was diluted to a final concentration of 50 mM citric acid and enriched by another aliquot of TiO_2 beads. 69% of phosphopeptides in the first enrichment were multiphosphorylated and 92% of them in the second enrichment were monophosphorylated. In comparison to the commonly used DHB/TiO_2 enrichment strategy, a 37% increase for the total number of identified

phosphopeptides and 2.6-fold increase for the number of identified multiphosphorylated peptides were observed.

Advances in Separation Techniques

Although IMAC or MOAC can effectively enrich thousands of phosphopeptides from tissue extract or cell lysate samples, the single step affinity purification is often not effective enough for a successful in-depth analysis of the phosphoproteome. Thus, pre-fractionation with different LC methods before routine RPLC-MS/MS analysis to reduce sample complexity is essential. In these pre-fractionation methods, strong cation exchange chromatography (SCX) is the most frequently used strategy prior to affinity purification. Hydrophilic interaction chromatography (HILIC), electrostatic repulsion-hydrophilic interaction chromatography (ERLIC), and high pH RPLC have also been used for fractionation of phosphopeptides before routine RPLC-MS/MS. Each method has its own merits to provide optimal fractionation for different physicochemical properties of phosphopeptides. For example, SCX yields the highest number of identified phosphopeptides, while ERLIC is optimal for the identification of multiphosphorylated peptides among SCX-TiO_2, HILIC-TiO_2, and ERLIC-TiO_2. [48] Furthermore, capillary electrophoresis (CE) is also applied for separation of phosphopeptides or phosphoproteins prior to MS analysis. Herein, we will highlight the five-year advances in separation methods for phosphoproteomics including ion exchange based chromatography, HILIC, ERLIC, RP and CE.

Ion Exchange Based Chromatography

SCX separates phosphopeptides from non-phosphorylated peptides based on the net positive charge in acidic condition. In acidic condition (pH=3), phosphopeptides usually possess a lower net charge compared to non-phosphopeptides because of the protonation of acidic residues and deprotonation of phosphate groups. Thus, phosphopeptides have poor retention on the SCX column leading to multiphosphorylated peptides eluting first, followed by N-acetylated peptides, monophosphorylated peptides and finally regular peptides with a differing number of basic residues. However, phosphopeptides with multiple basic residues will coelute with non-phosphorylated peptides because of the same net positive charge. A subsequent SCX separation at more acidic condition (pH=1) was used to address this problem by protonation of phosphate groups.[49] The neutralization of phosphate groups result in increasing net charge for phosphopeptides, whereas the net charge of regular peptides remains unchanged. Consequently, the unaffected non-phosphopeptides will elute first followed by the "basic" phosphopeptides. This tandem SCX at different pH approach enabled identification of over 10,000 unique "basic" phosphopeptides in HeLa cells. However, this approach requires LC equipment and SCX columns to operate at pH=1 which is not compatible with the majority of chromatographic systems and columns.

Unlike poor retention in SCX, phosphopeptides are eluted later compared to non-phosphorylated peptides during strong anion exchange chromatography (SAX) separation. In a typical neutral-to-alkaline SAX buffer system, basic residues tend to

be uncharged, whereas the C-terminal carboxyl group, acidic residues and phosphate groups carry negative charge. It has been demonstrated that SAX-TiO_2 enabled identification of more acidic and multiphosphorylated peptides in comparison with SCX-TiO_2.[50] This difference in performance can be attributed to the fact that phosphopeptides with or without multiple acidic residues elute simultaneously from the SCX column due to their same net charge at low pH condition. Thus, the acidic phosphopeptides are not effectively separated by SCX. However, acidic phosphopeptides can be separated according to the number of acidic amino acid residues they have by SAX due to the deprotonation of acidic residues under high pH condition. In contrast, phosphopeptides with multiple basic residues, which can be separated by SCX, cannot be distinguished from phosphopeptides without multiple basic residues in SAX. Due to the different properties of SCX and SAX , Dong et al. further separated the basic phosphopeptides from the early eluted fractions of SAX by SCX to improve the coverage for detection of basic phosphopeptides.[51] Based on similar idea, weak anion exchange (WAX) chromatography, using a low concentration of salt to elute phosphopeptides compared to SAX, was applied to further separate a SCX fraction which is dominated by phosphopeptides with one basic amino acid and a free N-terminus.[52]

Hydrophilic Interaction Chromatography

HILIC is chosen as a fractionation method for phosphopeptides due to its orthogonality to the subsequent routine low pH RPLC-MS/MS analysis.[53] In HILIC, samples are loaded at high organic solvent concentration and eluted by increasing the

polarity of the mobile phase, which is opposite to the operation mode for RPLC.[54] Due to the hydrophilic separation of HILIC, multiphosphorylated peptides and peptides with multiple acidic residues are strongly retained in HILIC and are more likely to be fractionated into the same fractions. An increased selectivity for subsequent IMAC enrichment is obtained since multiphosphorylated peptides will compete more effectively for IMAC binding sites compared to monophosphorylated peptides in the presence of high abundance acidic peptides.[55] In RPLC, phosphopeptides and their unmodified counterparts or isomers typically elute in a narrow window since the polar phosphate group only weakly interacts with the hydrophobic stationary phase resulting in little contribution to the retention of the peptide. In contrast, HILIC effectively separates phosphopeptides from their isomers or unmodified counterparts and elute phosphopeptides according to the number of phosphate groups they carry. Peptides (τ226-240) from human τ protein with different number of phosphate groups were successfully separated on a HILIC column with aminopropyl stationary phases. τ226-240, τ226-240 with one phosphate group on Ser238, τ226-240 with two phosphate groups on Thr231 and Ser235, and τ226-240 with two phosphate groups on Thr231 and Ser238 were sequentially eluted at 27.9, 31.8, 35.6 and 36.2 min over a 30 min linear gradient from 90 to 50% aqueous ACN in 5 mmol/L ammonium formate. [56]

Electrostatic Repulsion-hydrophilic Interaction Chromatography

Recently, ERLIC, which has the hydrophilic and weak anion-exchange characteristics, has been introduced as a novel separation method for

phosphopeptides.[57] In the low pH and high organic content of the mobile phase, the majority of peptides with acidic residues and carboxyl groups at C terminus are largely protonated, thus poorly retained on an ERLIC column, whereas phosphopeptides and highly hydrophilic peptides will be retained due to deprotonation of phosphate groups and hydrophilic interaction. Because of the characteristics of ERLIC, multiphosphorylated peptides possessing more negative charges are eluted later and separated with a relatively high resolution. However, mono- and non-phosphorylated peptides have very little retention on the ERLIC column and are eluted in a narrow window resulting in complex fractions. As mentioned above, SCX usually generates more complex fractions for multiphosphorylated peptides but produces better resolution for monophosphorylated peptides. Therefore, additional complementary separation for those complex fractions generated by SCX or ERLIC is necessary for in-depth phosphoproteomics. Combinatorial use of SCX and ERLIC to further separate their respective complex fractions has been investigated.[58] In this study, multiphosphorylated peptides were fractionated by ERLIC, while monophosphorylated peptides enriched ERLIC fractions were further separated by SCX. The number of identified phosphopeptides for HeLa cell was increased by about 48% with the use of combinatorial strategies (SCX-ERLIC and ERLIC-SCX) as compared to single step fractionation by SCX. A significant increase in the number of multiphosphorylated peptides was also noted when using combinatorial strategies that overcome the weakness of single SCX fractionation methods (Figure 2.4). The combinatorial strategies have been further

simplified by performing solid-phase extraction for both SCX and ERLIC.[59] In this simplified workflow, a total of 9952 unique phosphopeptides was identified for HeLa cells from totally 13 fractions. 2137 phosphopeptides of 9952 unique phosphopeptides were multiphosphorylated peptides.

Reversed Phase Liquid Chromatography

RPLC is also widely used for fractionation of phosphopeptides due to their high resolution and capability to provide orthogonal separation in different pH conditions. A high pH RPLC is typically utilized as the first dimension separation followed by routine on-line low pH RPLC in a phosphoproteomic analysis.[53, 60] Complex peptide mixtures are usually fractionated into tens of fractions. To reduce the instrument analysis time, the conventional way is to pool adjacent fractions together, thus reducing the total number of fractions required to be analyzed. Recently, a new pooling strategy was applied to 2D RPLC in phosphoproteomic analysis resulting in significantly increased number of identified phosphopeptides.[61] The fractions for enriched phosphopeptides were collected every minute over a 90 min gradient of the high pH RPLC. The fractions eluted in the first 45 min were labeled as early group and the last 45 min eluted fractions were called later eluted group. Then every two fractions from each group with equal time interval were mixed for the subsequent on-line low pH RPLC separation. The new pooling strategy yielded a 33% increase in the number of unique identified phosphopeptides for mouse liver lysate compared with that obtained by the conventional approach. This increase was likely due to a more random distribution of phosphopeptides in one pooled fraction. It

reduced the probability of coelution of phosphopeptides in the second separation dimension as similar physiochemical properties of phosphopeptides were pooled together in conventional adjacent pooling approach. A combinatorial use of multistep IMAC and high pH reversed phase fractionation from solid-phase extraction was introduced to simplify the RP-RP workflow for phosphoproteomics.[62] Phosphopeptides from MCF-10A cell lysate were firstly enriched by IMAC, and then fractionated by a hydrophilic−lipophilic-balanced reversed-phase cartridge using a serial mixed ratio of ACN/NH_4HCO_3 buffers. 3 mg of starting materials was used compared to 15 mg in a typical SCX-IMAC workflow and enabled the identification of 8969 unique phosphopeptides. [62-63]

Capillary Electrophoresis

As the phosphate group introduces negative charges and modifies the pIs of peptides or proteins, capillary electrophoresis, which resolves analytes based on their different charge-to-size ratio, offers distinct advantages for separation of phosphorylated peptides or proteins. Both ESI and MALDI MS have been frequently employed in CE-MS coupling for proteomic analysis [64]. Heemskerk et al. have explored the potential of sheathless ultra-low flow CE-ESI-MS in phosphoproteomics.[65] By using the high sensitivity porous sprayer (HSPS) for direct infusion of multiphosphorylated peptides, the authors demonstrated that a near equimolar ESI response could be approached when the flow rate was reduced below 15 nL/min, which might overcome the ionization bias against the phosphorylated peptides at conventional flow rates (> 50 nL/min). Although achieving this ultra-low

flow by conventional nanoLC is difficult, it is possible to operate sheathless CE-ESI-MS at this low flow rate with the HSPS and a neutrally coated capillary that EOF. Combined with an in-line preconcerntration technique, transient isotachophoresis (t-ITP), improved sensitivity of phosphopeptides from milk digest by CE-ESI-MS compared to that by RP nanoLC-MS under conditions of equal sample loading and equal sample concentrations was demonstrated. This sheathless nanospray CE-ESI-MS setup was also implemented for the characterization of post-translationally modified H1 histones isolated from rat testis by Sarg et al..[66] With IMAC enrichment prior to MS analysis, a total of 55 phosphopeptides were identified by combining the results of RP nanoLC-MS and CE-MS, 22 of which were uniquely identified by CE-MS and 19 were uniquely identified by RPLC-MS, showing the complementarity of the two techniques. A major drawback of CE analysis is the small sample volume that could be injected onto the column. In addition to the t-ITP mentioned above, pH-mediated stacking strategies were also adopted to increase the loading amount of phosphopeptides.[67] Dong et al. applied this method in CE-ESI-MS analysis of phosphopeptide isomers which have the same amino acid sequence but with phosphate group on different residues. [67b]

For MALDI-MS detection, the selection of background electrolytes (BGE) for CE analysis is more versatile compared to CE-ESI-MS. Bachmann et al. employed a latex-coated capillary and a BGE containing 80 mM of phosphoric acid, 40 mM of triethylamine and 20% ACN for CE-MALDI-TOF analysis of tryptic digests of phosphoproteins. [67a] With pH-mediated stacking and pressure assisted sample

deposition, even tetra- and penta-phosphorylated peptides (pI<2) from tryptic digests of α-casein and β-casein could be detected by MALDI-TOF MS.

Selected Applications

Recently, Mohammed group demonstrated that extensive fractionation in phosphoproteomics can boost the number of identified phosphopeptides from 3700 in a regular 1D LC-MS to 22,000 in a 3D LC-MS in human cancer cells.[68] The increasing depth in the coverage of phosphoproteomes has led to advances in biology.[3, 11, 69] Herein, we highlight a few recent examples of phosphoproteomic applications with improved enrichment and separation strategies that enabled better understanding of the biological processes. A recent study employed SCX and TiO_2 to fractionate and enrich phosphopeptides from AT_1R (angiotensin II type 1 receptor) stable transfected HEK293 (AT_1R-HEK293) cells as AT_1R is an important drug target in cardiovascular diseases.[40] A comprehensive understanding of the AT_1R signaling pathways is critical for the development of effective treatment for cardiovascular diseases. A total of 10,965 unique phosphopeptides were identified and led to the discovery of protein kinase D as a critical mediator in the AT_1R signaling pathways by comparing differential phosphoproteomes in two different agonists treated AT_1R-HEK293 cells. Another work used ERLIC to enrich and separate phosphopeptides in partner of PIX 2 (POPX2) over-expressing cells in order to elucidate the regulatory mechanism of cancer cell motility and invasiveness. POPX2 is a serine/threonine phosphatase and is found in many cancer types, regulating cancer cell motility and invasiveness. After ERLIC enrichment and LC-

MS/MS analysis, 3700 phosphopeptides were identified. With this approach, POPX2 was found to exert its cellular function through the regulation of the activity of glycogen synthase kinase-3. IMAC or MOAC combined with HILIC have also been implemented in a phosphoproteomic workflow to elucidate molecular pathways involved in biological processes.[70] TiO_2 and HILIC were utilized to facilitate monitoring cell signaling changes during mouse brain development.[70a] A total of 7682 unique phosphopeptides and 3246 unique formerly sialylated glycopeptides were identified. IMAC-HILIC was also used in conjunction to study 17β-Estradiol (E2)-modulated phosphorylation in order to investigate transcriptional activity regulated by E2.[70b] A collection of 2857 unique phosphorylation sites were quantified and E2 was found to modulate gene transcription through a HSP90 phosphorylation-mediated chaperoning process.

Concluding Remarks and Future Prospects

As we presented in this review, new materials and improved sample processing protocols have been rapidly developed for IMAC and MOAC strategies that enable phosphopeptide enrichment with high sensitivity and specificity. Furthermore, different separation methods with high resolution and efficiency have been utilized for enhanced separation of different types of phosphopeptides. Integrated use of enrichment techniques and multiple separation methods provides a powerful platform, enabling in-depth coverage of the phosphoproteome. Although MS-based phosphoproteomics have gained significant successes in revealing the roles of phosphorylation in various biological systems, several limitations still hinder the

broader applications of MS-based phosphoproteomics.[3, 11] Non-specific binding and preferential enrichment of certain types of phosphopeptides still challenge the current enrichment techniques. Furthermore, although incorporating multidimensional separation strategies into phosphoproteomic workflows increases the phosphoproteome coverage, the workflow with increased complexity results in more sample-handling steps and greater sample losses. Therefore, the continuing development of simplified and integrated phosphoproteomic workflow that offers high selectivity for phosphopeptides and provides improved-coverage and sensitivity would be in great demand. To better decipher biological processes under diseased and healthy conditions, quantitative approaches with multiplexing capability and reduced experimental cost are also highly attractive with great promise for systems phosphoproteomics.

REFERENCES

1. Mann, M.; Ong, S. E.; Gronborg, M.; Steen, H.; Jensen, O. N.; Pandey, A., Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends in biotechnology* **2002,** *20* (6), 261-8.

2. Hunter, T., Signaling--2000 and beyond. *Cell* **2000,** *100* (1), 113-27.

3. Harsha, H. C.; Pandey, A., Phosphoproteomics in cancer. *Molecular oncology* **2010,** *4* (6), 482-95.

4. Sun, Z.; Hamilton, K. L.; Reardon, K. F., Phosphoproteomics and molecular cardiology: techniques, applications and challenges. *J Mol Cell Cardiol* **2012,** *53* (3), 354-68.

5. Pawson, T.; Scott, J. D., Protein phosphorylation in signaling--50 years and counting. *Trends in biochemical sciences* **2005,** *30* (6), 286-90.

6. Piggee, C., Phosphoproteomics: miles to go before it's routine. *Analytical chemistry* **2009,** *81* (7), 2418-20.

7. (a) McLachlin, D. T.; Chait, B. T., Analysis of phosphorylated proteins and peptides by mass spectrometry. *Current opinion in chemical biology* **2001,** *5* (5), 591-602; (b) Mann, M.; Jensen, O. N., Proteomic analysis of post-translational modifications. *Nature biotechnology* **2003,** *21* (3), 255-61.

8. (a) Craig, A. G.; Hoeger, C. A.; Miller, C. L.; Goedken, T.; Rivier, J. E.; Fischer, W. H., Monitoring protein kinase and phosphatase reactions with matrix-assisted laser desorption/ionization mass spectrometry and capillary zone electrophoresis: comparison of the detection efficiency of peptide-phosphopeptide mixtures. *Biological mass spectrometry* **1994,** *23* (8), 519-28; (b) Liao, P. C.; Leykam, J.; Andrews, P. C.; Gage, D. A.; Allison, J., An approach to locate phosphorylation sites in a phosphoprotein: mass mapping by combining specific enzymatic degradation with matrix-assisted laser desorption/ionization mass spectrometry. *Analytical biochemistry* **1994,** *219* (1), 9-20.

9. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology* **2001,** *19* (3), 242-7.

10. Thingholm, T. E.; Jensen, O. N.; Larsen, M. R., Analytical strategies for phosphoproteomics. *Proteomics* **2009,** *9* (6), 1451-68.

11. Nakagami, H.; Sugiyama, N.; Ishihama, Y.; Shirasu, K., Shotguns in the front line: phosphoproteomics in plants. *Plant & cell physiology* **2012,** *53* (1), 118-24.

12. Nika, H.; Lee, J.; Willis, I. M.; Angeletti, R. H.; Hawke, D. H., Phosphopeptide characterization by mass spectrometry using reversed-phase supports for solid-phase beta-elimination/Michael addition. *Journal of biomolecular techniques : JBT* **2012,** *23* (2), 51-68.

13. Fila, J.; Honys, D., Enrichment techniques employed in phosphoproteomics. *Amino acids* **2012,** *43* (3), 1025-47.

14. (a) Zou, X. J.; Liu, D.; Zhong, L. J.; Yang, B.; Lou, Y. X.; Hu, B. H.; Yin, Y. X., Highly specific capture and direct MALDI-MS analysis of phosphorylated peptides using novel multifunctional chitosan-GMA-IDA-Fe (III) nanosphere. *Analytical and bioanalytical chemistry* **2011,** *401* (4), 1251-1261; (b) Hu, L.; Zhou, H.; Li, Y.; Sun, S.; Guo, L.; Ye, M.; Tian, X.; Gu, J.; Yang, S.; Zou, H., Profiling of endogenous serum phosphorylated peptides by titanium (IV) immobilized mesoporous silica particles enrichment and MALDI-TOFMS detection. *Analytical chemistry* **2009,** *81* (1), 94-104; (c) Hu, Y.; Peng, Y.; Lin, K.; Shen, H.; Brousseau, L. C., 3rd; Sakamoto, J.; Sun, T.; Ferrari, M., Surface engineering on mesoporous silica chips for enriching low molecular weight phosphorylated proteins. *Nanoscale* **2011,** *3* (2), 421-8; (d) Qian, K.; Wan, J.; Liu, F.; Girault, H. H.; Liu, B.; Yu, C., A phospho-directed macroporous alumina-silica nanoreactor with multi-functions. *Acs Nano* **2009,** *3* (11), 3656-62.

15. Dunn, J. D.; Reid, G. E.; Bruening, M. L., Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. *Mass Spectrom Rev* **2010,** *29* (1), 29-54.

16. Negroni, L.; Claverol, S.; Rosenbaum, J.; Chevet, E.; Bonneu, M.; Schmitter, J. M., Comparison of IMAC and MOAC for phosphopeptide enrichment by column chromatography. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2012,** *891-892*, 109-12.

17. Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M., Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nature biotechnology* **2002,** *20* (3), 301-5.

18. Stewart, II; Thomson, T.; Figeys, D., 18O labeling: a tool for proteomics. *Rapid communications in mass spectrometry : RCM* **2001,** *15* (24), 2456-65.

19. (a) Hou, C.; Ma, J.; Tao, D.; Shan, Y.; Liang, Z.; Zhang, L.; Zhang, Y., Organic-inorganic hybrid silica monolith based immobilized titanium ion affinity chromatography column for analysis of mitochondrial phosphoproteome. *Journal of proteome research* **2010,** *9* (8), 4093-101; (b) Krenkova, J.; Lacher, N. A.; Svec, F., Control of selectivity via nanochemistry: monolithic capillary column containing hydroxyapatite nanoparticles for separation of proteins and enrichment of phosphopeptides. *Analytical chemistry* **2010,** *82* (19), 8335-41; (c) Wang, H.; Duan, J.; Xu, H.; Zhao, L.; Liang, Y.; Shan, Y.; Zhang, L.; Liang, Z.; Zhang, Y., Monoliths with immobilized zirconium ions for selective enrichment of phosphopeptides. *Journal of separation science* **2011**.

20. Ficarro, S. B.; Adelmant, G.; Tomar, M. N.; Zhang, Y.; Cheng, V. J.; Marto, J. A., Magnetic bead processor for rapid evaluation and optimization of parameters for phosphopeptide enrichment. *Analytical chemistry* **2009,** *81* (11), 4566-75.

21. Batalha, I. L.; Lowe, C. R.; Roque, A. C., Platforms for enrichment of phosphorylated proteins and peptides in proteomics. *Trends in biotechnology* **2012,** *30* (2), 100-10.

22. (a) Li, X. S.; Su, X.; Zhu, G. T.; Zhao, Y.; Yuan, B. F.; Guo, L.; Feng, Y. Q., Titanium-containing magnetic mesoporous silica spheres: effective enrichment of peptides and simultaneous separation of nonphosphopeptides and phosphopeptides. *Journal of separation science* **2012,** *35* (12), 1506-13; (b) Lu, J.; Li, Y.; Deng, C., Facile synthesis of zirconium phosphonate-functionalized magnetic mesoporous silica microspheres designed for highly selective enrichment of phosphopeptides. *Nanoscale* **2011,** *3* (3), 1225-33.

23. (a) Wang, W. H.; Bruening, M. L., Phosphopeptide enrichment on functionalized polymer microspots for MALDI-MS analysis. *The Analyst* **2009,** *134* (3), 512-8; (b) Lu, J.; Liu, S.; Deng, C., Facile synthesis of alumina hollow spheres for on-plate-selective enrichment of phosphopeptides. *Chemical communications (Cambridge, England)* **2011,** *47* (18), 5334-6; (c) Hoang, T.; Roth, U.; Kowalewski, K.; Belisle, C.; Steinert, K.; Karas, M., Highly specific capture and direct MALDI MS analysis of phosphopeptides by zirconium phosphonate on self-assembled monolayers. *Analytical chemistry* **2010,** *82* (1), 219-28.

24. Iliuk, A. B.; Martin, V. A.; Alicie, B. M.; Geahlen, R. L.; Tao, W. A., In-depth analyses of kinase-dependent tyrosine phosphoproteomes based on metal ion-functionalized soluble nanopolymers. *Molecular & cellular proteomics : MCP* **2010,** *9* (10), 2162-72.

25. (a) Shen, F.; Hu, Y.; Guan, P.; Ren, X., Ti(4+)-phosphate functionalized cellulose for phosphopeptides enrichment and its application in rice phosphoproteome

analysis. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2012,** *902*, 108-15; (b) Yan, Y.; Zheng, Z.; Deng, C.; Li, Y.; Zhang, X.; Yang, P., Hydrophilic polydopamine-coated graphene for metal ion immobilization as a novel immobilized metal ion affinity chromatography platform for phosphoproteome analysis. *Analytical chemistry* **2013,** *85* (18), 8483-7; (c) Nabetani, T.; Kim, Y. J.; Watanabe, M.; Ohashi, Y.; Kamiguchi, H.; Hirabayashi, Y., Improved method of phosphopeptides enrichment using biphasic phosphate-binding tag/C18 tip for versatile analysis of phosphorylation dynamics. *Proteomics* **2009,** *9* (24), 5525-33; (d) Kinoshita, E.; Kinoshita-Kikuta, E.; Sugiyama, Y.; Fukada, Y.; Ozeki, T.; Koike, T., Highly sensitive detection of protein phosphorylation by using improved Phos-tag Biotin. *Proteomics* **2012,** *12* (7), 932-7.

26. Han, G.; Ye, M.; Zou, H., Development of phosphopeptide enrichment techniques for phosphoproteome analysis. *The Analyst* **2008,** *133* (9), 1128-38.

27. Worthington, J.; Cutillas, P. R.; Timms, J. F., IMAC/TiO(2) enrich for peptide modifications other than phosphorylation: implications for chromatographic choice and database searching in phosphoproteomics. *Proteomics* **2011,** *11* (23), 4583-7.

28. Li, Y.; Luo, Y.; Wu, S.; Gao, Y.; Liu, Y.; Zheng, D., Nucleic acids in protein samples interfere with phosphopeptide identification by immobilized-metal-ion affinity chromatography and mass spectrometry. *Molecular biotechnology* **2009,** *43* (1), 59-66.

29. (a) Ye, J.; Zhang, X.; Young, C.; Zhao, X.; Hao, Q.; Cheng, L.; Jensen, O. N., Optimized IMAC-IMAC protocol for phosphopeptide recovery from complex biological samples. *Journal of proteome research* **2010,** *9* (7), 3561-73; (b) Lai, A. C.; Tsai, C. F.; Hsu, C. C.; Sun, Y. N.; Chen, Y. J., Complementary Fe(3+)- and Ti(4+)-immobilized metal ion affinity chromatography for purification of acidic and basic phosphopeptides. *Rapid communications in mass spectrometry : RCM* **2012,** *26* (18), 2186-94; (c) Tsai, C. F.; Hsu, C. C.; Hung, J. N.; Wang, Y. T.; Choong, W. K.; Zeng, M. Y.; Lin, P. Y.; Hong, R. W.; Sung, T. Y.; Chen, Y. J., Sequential Phosphoproteomic Enrichment through Complementary Metal-directed Immobilized Metal Ion Affinity Chromatography. *Analytical chemistry* **2013**.

30. (a) Sun, S.; Ma, H.; Han, G.; Wu, R.; Zou, H.; Liu, Y., Efficient enrichment and identification of phosphopeptides by cerium oxide using on-plate matrix-assisted laser desorption/ionization time-of-flight mass spectrometric analysis. *Rapid communications in mass spectrometry : RCM* **2011,** *25* (13), 1862-8; (b) Lu, Z.; Duan, J.; He, L.; Hu, Y.; Yin, Y., Mesoporous TiO(2) nanocrystal clusters for selective enrichment of phosphopeptides. *Analytical chemistry* **2010,** *82* (17), 7249-58; (c) Wan, H.; Yan, J.; Yu, L.; Zhang, X.; Xue, X.; Li, X.; Liang, X., Zirconia layer coated mesoporous silica microspheres used for highly specific phosphopeptide enrichment. *Talanta* **2010,** *82* (5), 1701-7; (d) Chen, S. Y.; Juang, Y. M.; Chien, M. W.; Li, K. I.;

Yu, C. S.; Lai, C. C., Magnetic iron oxide nanoparticle enrichment of phosphopeptides on a radiate microstructure MALDI chip. *The Analyst* **2011,** *136* (21), 4454-9; (e) Lu, J.; Qi, D.; Deng, C.; Zhang, X.; Yang, P., Hydrothermal synthesis of alpha-Fe(2)O(3)@SnO(2) core-shell nanotubes for highly selective enrichment of phosphopeptides for mass spectrometry analysis. *Nanoscale* **2010,** *2* (10), 1892-900; (f) Cheng, G.; Zhang, J. L.; Liu, Y. L.; Sun, D. H.; Ni, J. Z., Synthesis of novel Fe3O4@SiO2@CeO2 microspheres with mesoporous shell for phosphopeptide capturing and labeling. *Chemical communications (Cambridge, England)* **2011,** *47* (20), 5732-4.

31. (a) Blacken, G. R.; Volny, M.; Diener, M.; Jackson, K. E.; Ranjitkar, P.; Maly, D. J.; Turecek, F., Reactive landing of gas-phase ions as a tool for the fabrication of metal oxide surfaces for in situ phosphopeptide enrichment. *Journal of the American Society for Mass Spectrometry* **2009,** *20* (6), 915-26; (b) Nelson, C. A.; Szczech, J. R.; Dooley, C. J.; Xu, Q.; Lawrence, M. J.; Zhu, H.; Jin, S.; Ge, Y., Effective enrichment and mass spectrometry analysis of phosphopeptides using mesoporous metal oxide nanomaterials. *Analytical chemistry* **2010,** *82* (17), 7193-201.

32. Vilasi, A.; Fiume, I.; Pace, P.; Rossi, M.; Pocsfalvi, G., Enrichment specificity of micro and nano-sized titanium and zirconium dioxides particles in phosphopeptide mapping. *Journal of mass spectrometry : JMS* **2013,** *48* (11), 1188-98.

33. (a) Tang, L. A.; Wang, J.; Lim, T. K.; Bi, X.; Lee, W. C.; Lin, Q.; Chang, Y. T.; Lim, C. T.; Loh, K. P., High-performance graphene-titania platform for detection of phosphopeptides in cancer cells. *Analytical chemistry* **2012,** *84* (15), 6693-700; (b) Tan, Y. J.; Sui, D.; Wang, W. H.; Kuo, M. H.; Reid, G. E.; Bruening, M. L., Phosphopeptide enrichment with TiO2-modified membranes and investigation of tau protein phosphorylation. *Analytical chemistry* **2013,** *85* (12), 5699-706.

34. (a) Wang, S. T.; Wang, M. Y.; Su, X.; Yuan, B. F.; Feng, Y. Q., Facile preparation of SiO2/TiO2 composite monolithic capillary column and its application in enrichment of phosphopeptides. *Analytical chemistry* **2012,** *84* (18), 7763-70; (b) Krenkova, J.; Foret, F., Iron oxide nanoparticle coating of organic polymer-based monolithic columns for phosphopeptide enrichment. *Journal of separation science* **2011,** *34* (16-17), 2106-2112.

35. (a) Zhang, L.; Xu, J.; Sun, L.; Ma, J.; Yang, K.; Liang, Z.; Zhang, L.; Zhang, Y., Zirconium oxide aerogel for effective enrichment of phosphopeptides with high binding capacity. *Analytical and bioanalytical chemistry* **2011,** *399* (10), 3399-405; (b) Zhang, L.; Liang, Z.; Yang, K.; Xia, S.; Wu, Q.; Zhang, L.; Zhang, Y., Mesoporous TiO2 aerogel for selective enrichment of phosphopeptides in rat liver mitochondria. *Analytica chimica acta* **2012,** *729*, 26-35.

36. (a) Ji, L.; Wu, J. H.; Luo, Q.; Li, X.; Zheng, W.; Zhai, G.; Wang, F.; Lu, S.; Feng, Y. Q.; Liu, J.; Xiong, S., Quantitative mass spectrometry combined with separation and enrichment of phosphopeptides by titania coated magnetic mesoporous silica microspheres for screening of protein kinase inhibitors. *Analytical chemistry* **2012,** *84* (5), 2284-91; (b) Lu, J.; Deng, C.; Zhang, X.; Yang, P., Synthesis of Fe3O4/graphene/TiO2 composites for the highly selective enrichment of phosphopeptides from biological samples. *ACS applied materials & interfaces* **2013,** *5* (15), 7330-4; (c) Zeng, Y. Y.; Chen, H. J.; Shiau, K. J.; Hung, S. U.; Wang, Y. S.; Wu, C. C., Efficient enrichment of phosphopeptides by magnetic TiO(2)-coated carbon-encapsulated iron nanoparticles. *Proteomics* **2012,** *12* (3), 380-90.

37. Eriksson, A.; Bergquist, J.; Edwards, K.; Hagfeldt, A.; Malmstrom, D.; Hernandez, V. A., Mesoporous TiO2-based experimental layout for on-target enrichment and separation of multi- and monophosphorylated peptides prior to analysis with matrix-assisted laser desorption-ionization mass spectrometry. *Analytical chemistry* **2011,** *83* (3), 761-6.

38. Wang, H.; Duan, J.; Cheng, Q., Photocatalytically patterned TiO2 arrays for on-plate selective enrichment of phosphopeptides and direct MALDI MS analysis. *Analytical chemistry* **2011,** *83* (5), 1624-31.

39. Bi, H.; Qiao, L.; Busnel, J. M.; Devaud, V.; Liu, B.; Girault, H. H., TiO(2) printed aluminum foil: single-use film for a laser desorption/ionization target plate. *Analytical chemistry* **2009,** *81* (3), 1177-83.

40. Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular & cellular proteomics : MCP* **2005,** *4* (7), 873-86.

41. Wu, J.; Shakey, Q.; Liu, W.; Schuller, A.; Follettie, M. T., Global profiling of phosphopeptides by titania affinity enrichment. *Journal of proteome research* **2007,** *6* (12), 4684-9.

42. Sugiyama, N.; Masuda, T.; Shinoda, K.; Nakamura, A.; Tomita, M.; Ishihama, Y., Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Molecular & cellular proteomics : MCP* **2007,** *6* (6), 1103-9.

43. Fukuda, I.; Hirabayashi-Ishioka, Y.; Sakikawa, I.; Ota, T.; Yokoyama, M.; Uchiumi, T.; Morita, A., Optimization of Enrichment Conditions on TiO2 Chromatography Using Glycerol As an Additive Reagent for Effective Phosphoproteomic Analysis. *Journal of proteome research* **2013,** *12* (12), 5587-97.

44. Kanshin, E.; Michnick, S. W.; Thibault, P., Displacement of N/Q-rich peptides on TiO2 beads enhances the depth and coverage of yeast phosphoproteome analyses. *Journal of proteome research* **2013,** *12* (6), 2905-13.

45. Li, Q. R.; Ning, Z. B.; Tang, J. S.; Nie, S.; Zeng, R., Effect of peptide-to-TiO2 beads ratio on phosphopeptide enrichment selectivity. *Journal of proteome research* **2009,** *8* (11), 5375-81.

46. Park, S. S.; Maudsley, S., Discontinuous pH gradient-mediated separation of TiO2-enriched phosphopeptides. *Analytical biochemistry* **2011,** *409* (1), 81-8.

47. Zhao, X.; Wang, Q.; Wang, S.; Zou, X.; An, M.; Zhang, X.; Ji, J., Citric acid-assisted two-step enrichment with TiO2 enhances the separation of multi- and monophosphorylated peptides and increases phosphoprotein profiling. *Journal of proteome research* **2013,** *12* (6), 2467-76.

48. Zarei, M.; Sprenger, A.; Metzger, F.; Gretzmeier, C.; Dengjel, J., Comparison of ERLIC-TiO2, HILIC-TiO2, and SCX-TiO2 for global phosphoproteomics approaches. *Journal of proteome research* **2011,** *10* (8), 3474-83.

49. Hennrich, M. L.; van den Toorn, H. W.; Groenewold, V.; Heck, A. J.; Mohammed, S., Ultra acidic strong cation exchange enabling the efficient enrichment of basic phosphopeptides. *Analytical chemistry* **2012,** *84* (4), 1804-8.

50. Dai, J.; Wang, L. S.; Wu, Y. B.; Sheng, Q. H.; Wu, J. R.; Shieh, C. H.; Zeng, R., Fully automatic separation and identification of phosphopeptides by continuous pH-gradient anion exchange online coupled with reversed-phase liquid chromatography mass spectrometry. *Journal of proteome research* **2009,** *8* (1), 133-41.

51. Dong, M.; Ye, M.; Cheng, K.; Song, C.; Pan, Y.; Wang, C.; Bian, Y.; Zou, H., Depletion of acidic phosphopeptides by SAX to improve the coverage for the detection of basophilic kinase substrates. *Journal of proteome research* **2012,** *11* (9), 4673-81.

52. Hennrich, M. L.; Groenewold, V.; Kops, G. J.; Heck, A. J.; Mohammed, S., Improving depth in phosphoproteomics by using a strong cation exchange-weak anion exchange-reversed phase multidimensional separation approach. *Analytical chemistry* **2011,** *83* (18), 7137-43.

53. Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C., Orthogonality of separation in two-dimensional liquid chromatography. *Analytical chemistry* **2005,** *77* (19), 6426-34.

54. Alpert, A. J., Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of chromatography* **1990,** *499*, 177-96.

55. McNulty, D. E.; Annan, R. S., Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Molecular & cellular proteomics : MCP* **2008,** *7* (5), 971-80.

56. Singer, D.; Kuhlmann, J.; Muschket, M.; Hoffmann, R., Separation of multiphosphorylated peptide isomers by hydrophilic interaction chromatography on an aminopropyl phase. *Analytical chemistry* **2010,** *82* (15), 6409-14.

57. Alpert, A. J., Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Analytical chemistry* **2008,** *80* (1), 62-76.

58. Zarei, M.; Sprenger, A.; Gretzmeier, C.; Dengjel, J., Combinatorial use of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) and strong cation exchange (SCX) chromatography for in-depth phosphoproteome analysis. *Journal of proteome research* **2012,** *11* (8), 4269-76.

59. Zarei, M.; Sprenger, A.; Gretzmeier, C.; Dengjel, J., Rapid Combinatorial ERLIC-SCX Solid-Phase Extraction for In-Depth Phosphoproteome Analysis. *Journal of proteome research* **2013,** *12* (12), 5989-95.

60. Delmotte, N.; Lasaosa, M.; Tholey, A.; Heinzle, E.; Huber, C. G., Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis. *Journal of proteome research* **2007,** *6* (11), 4363-73.

61. Song, C.; Ye, M.; Han, G.; Jiang, X.; Wang, F.; Yu, Z.; Chen, R.; Zou, H., Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. *Analytical chemistry* **2010,** *82* (1), 53-6.

62. Yue, X. S.; Hummon, A. B., Combination of multistep IMAC enrichment with high-pH reverse phase separation for in-depth phosphoproteomic profiling. *Journal of proteome research* **2013,** *12* (9), 4176-86.

63. Villen, J.; Gygi, S. P., The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature protocols* **2008,** *3* (10), 1630-8.

64. Zhong, X.; Zhang, Z.; Jiang, S.; Li, L., Recent advances in coupling capillary electrophoresis-based separation techniques to ESI and MALDI-MS. *Electrophoresis* **2013**.

65. Heemskerk, A. A.; Busnel, J. M.; Schoenmaker, B.; Derks, R. J.; Klychnikov, O.; Hensbergen, P. J.; Deelder, A. M.; Mayboroda, O. A., Ultra-low flow electrospray ionization-mass spectrometry for improved ionization efficiency in phosphoproteomics. *Analytical chemistry* **2012,** *84* (10), 4552-9.

66. Sarg, B.; Faserl, K.; Kremser, L.; Halfinger, B.; Sebastiano, R.; Lindner, H. H., Comparing and combining capillary electrophoresis electrospray ionization mass spectrometry and nano-liquid chromatography electrospray ionization mass spectrometry for the characterization of post-translationally modified histones. *Molecular & cellular proteomics : MCP* **2013,** *12* (9), 2640-56.

67. (a) Bachmann, S.; Bakry, R.; Huck, C. W.; Polato, F.; Corradini, D.; Bonn, G. K., Peptide mapping using capillary electrophoresis offline coupled to matrix-assisted laser desorption ionization time of flight mass spectrometry. *Electrophoresis* **2011,** *32* (20), 2830-2839; (b) Dong, Y. M.; Chien, K. Y.; Chen, J. T.; Lin, S. J.; Wang, T. C.; Yu, J. S., Site-specific separation and detection of phosphopeptide isomers with pH-mediated stacking capillary electrophoresis-electrospray ionization-tandem mass spectrometry. *Journal of separation science* **2013,** *36* (9-10), 1582-9.

68. Zhou, H.; Di Palma, S.; Preisinger, C.; Peng, M.; Polat, A. N.; Heck, A. J.; Mohammed, S., Toward a comprehensive characterization of a human cancer cell phosphoproteome. *Journal of proteome research* **2013,** *12* (1), 260-71.

69. Tobe, B. T.; Hou, J.; Crain, A. M.; Singec, I.; Snyder, E. Y.; Brill, L. M., Phosphoproteomic analysis: an emerging role in deciphering cellular signaling in human embryonic stem cells and their differentiated derivatives. *Stem cell reviews* **2012,** *8* (1), 16-31.

70. (a) Palmisano, G.; Parker, B. L.; Engholm-Keller, K.; Lendal, S. E.; Kulej, K.; Schulz, M.; Schwammle, V.; Graham, M. E.; Saxtorph, H.; Cordwell, S. J.; Larsen, M. R., A novel method for the simultaneous enrichment, identification, and quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain development. *Molecular & cellular proteomics : MCP* **2012,** *11* (11), 1191-202; (b) Wu, C. J.; Chen, Y. W.; Tai, J. H.; Chen, S. H., Quantitative phosphoproteomics studies using stable isotope dimethyl labeling coupled with IMAC-HILIC-nanoLC-MS/MS for estrogen-induced transcriptional regulation. *Journal of proteome research* **2011,** *10* (3), 1088-97.

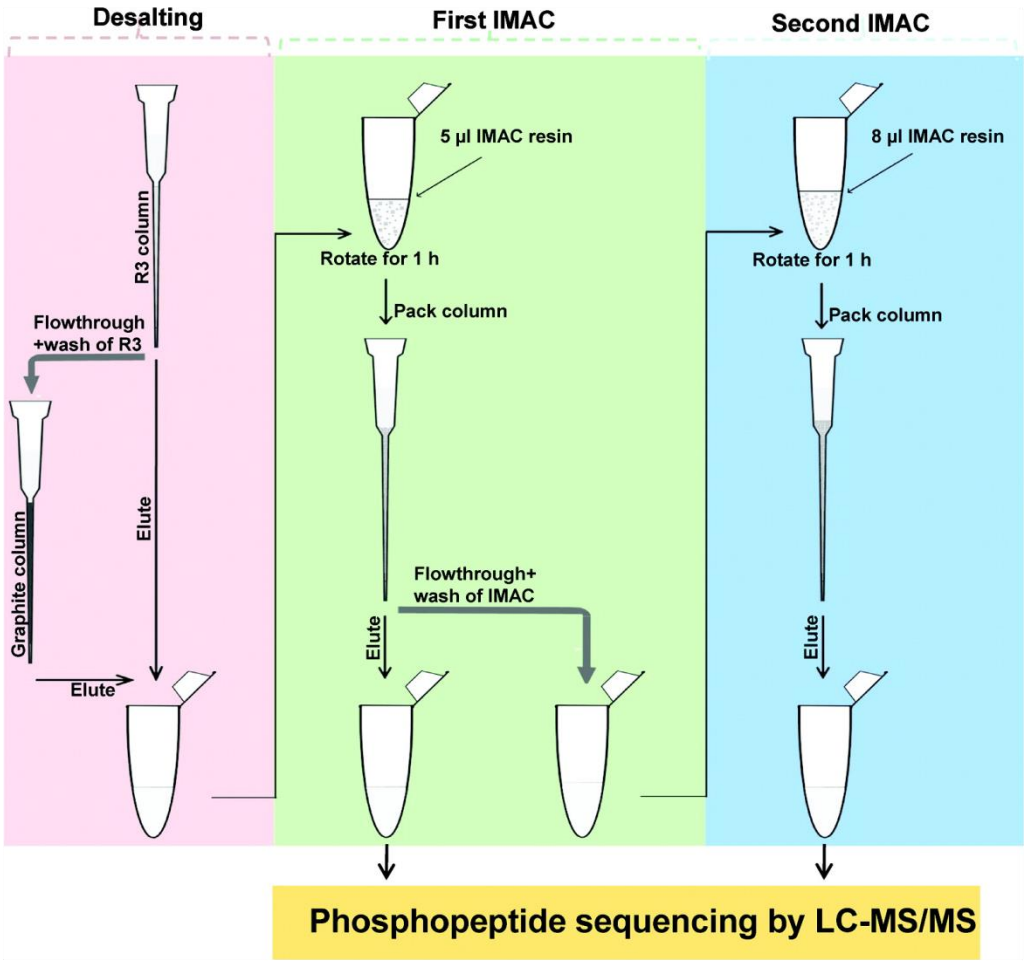


Figure 2.1 IMAC-IMAC strategy for enrichment of phosphopeptides from complex biological samples. Reprinted with permission from Ref [29a]. Copyright American Chemical Society 2010.

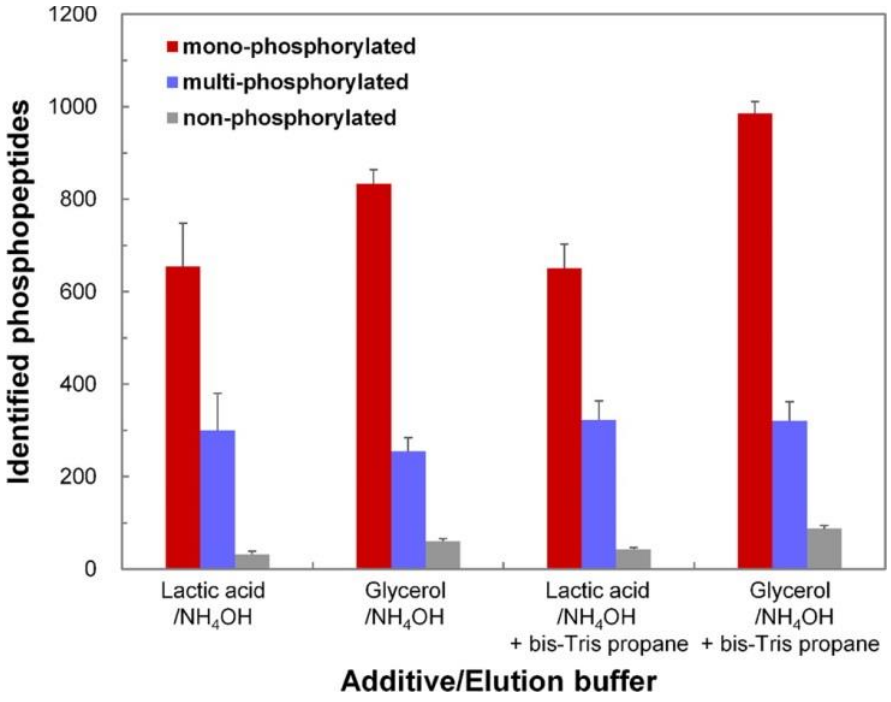


Figure 2.2 Comparison of the number of identified phosphopeptides obtained from PC3 cell lysate digest using different elution conditions for the TiO_2 enrichment. LC-MS/MS measurement was performed in duplicate. Reprinted with permission from Ref [43]. Copyright American Chemical Society 2013.

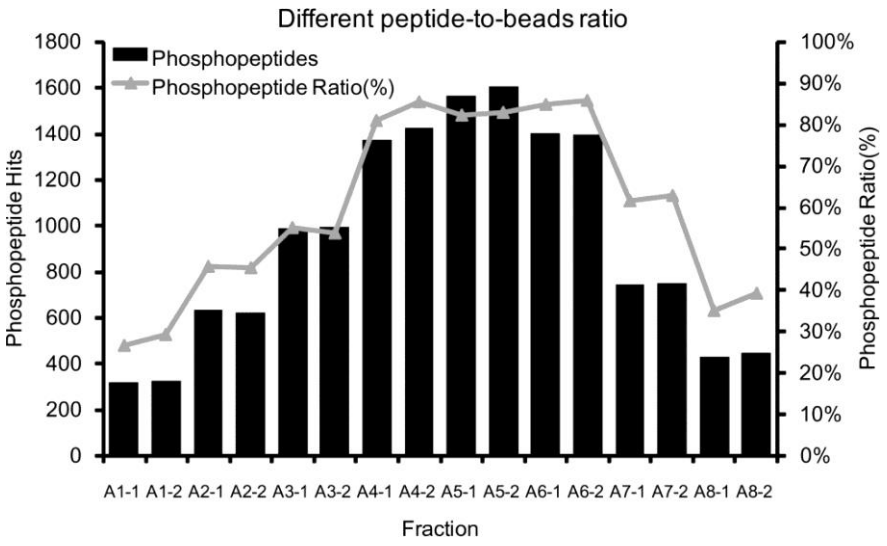


Figure 2.3 Profiles of identified phosphopeptides from 500 μg of HeLa cell lysate digest when different peptide-to-TiO_2 beads ratio used during enrichment. Phosphopeptide ratio for each fraction is calculated by the number of identified phosphopeptides divided by the total identified peptides. A1 to A8 represent 125, 250, 500, 1000, 2000, 4000, 10 000, and 20 000 μg TiO_2 beads respectively, and "−1" and "−2" correspond to duplicates of each experiment. Reprinted with permission from Ref [45]. Copyright American Chemical Society 2009.

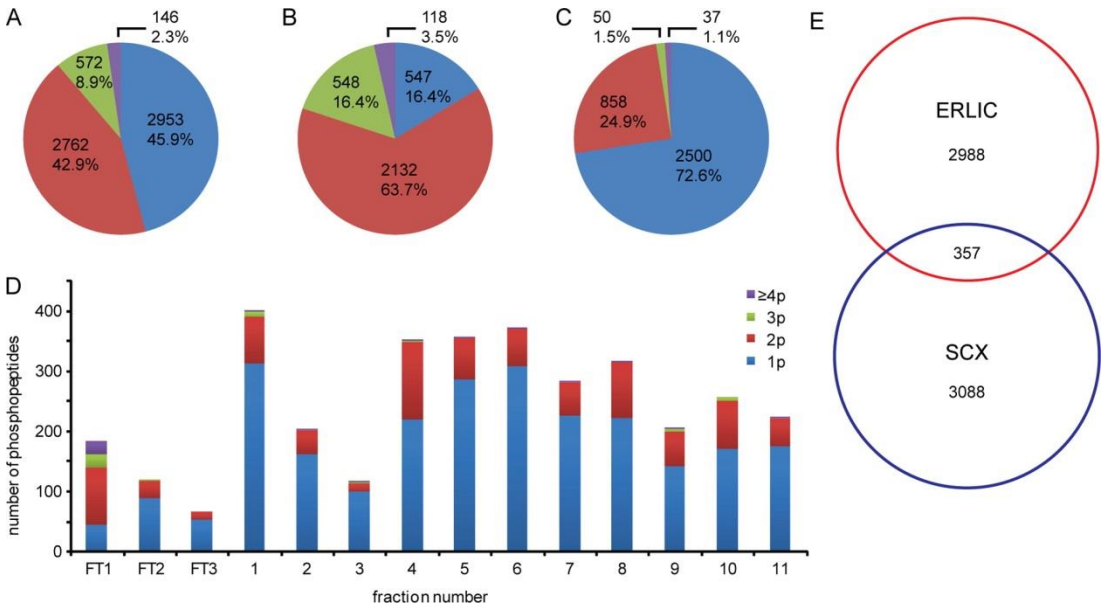


Figure 2.4 Distributions of identified phosphopeptides from HeLa cell lysate by combinatorial ERLIC–SCX fractionation. The number and percentage of nonredundant identified mono- and multiphosphorylated peptides by ERLIC-SCX (A). (B) The number and percentage of nonredundant identified mono- and multiphosphorylated peptides by ERLIC. (C) The number and percentage of nonredundant identified mono- and multiphosphorylated peptides from further separation of flow-through (FT) fractions of ERLIC by SCX. (D) Distribution of identified mono- and multiphosphorylated peptides in each SCX fraction. (E) Venn diagram for nonredundant identified phosphopeptides in both methods. Different colors represent the number of phosphate groups (1p-4,5p) carried by each phosphopeptide. Reprinted with permission from Ref [58]. Copyright American Chemical Society 2012.

Chapter 3

Revealing the Interacting Proteins of MBD1 in Mouse Brain by Mass Spectrometry

Adapted from: "Revealing the interacting proteins of MBD1 in Mouse brain by mass spectrometry." **Yang, C.**, Guo, Y., Zhao, X., Li, L. In preparation.

ABSTRACT

Methyl-CpG binding protein 1 (MBD1) is the member of the MBD family, which can translate the regulatory information inherent in DNA methylation into functional states. Previous studies demonstrated that MBD1 deficiency resulted in reduced adult neural stem/progenitor cells (aNSCs) differentiation and brain-associated phenotypes in mice, suggesting a role of MBD1 in regulation of neural development. Although considerable progress has been made to decipher the regulatory mechanism of MBD1 in neural development via interaction with DNA and RNA, proteins that mediate MBD1 in regulation of neural development remain to be mystery. Here, we utilized affinity purification (AP) and mass spectrometry (MS) to characterize the protein interacting networks of MBD1 in the mouse brain. A spectral counting based quantification approach revealed five proteins with significant changes as potential binders for MBD1 from hundreds of proteins present in the mouse brain extract. Our study here has provided a list of candidates for future validations and led to an insight into the function of MBD1 in neural development.

INTRODUCTION

DNA methylation is a common epigenetic modification in the genomes of diverse organisms.[1] Mutation in the DNA methyltransferase gene can result in developmental defects and embryonic lethality, proving the essential nature of DNA methylation in mammals.[2,3] Methyl-CpG-binding domain proteins (MBDs), including MBD1, MBD2, MBD3, MBD4 and MeCp2, regulate gene expression by recognizing and binding to methylated DNA through their MBD domain. Unlike other proteins in MBD family, MBD1 can not only bind to methylated DNA but also interact with unmethylated DNA through the CXXC3 domain, which makes it a unique member of the MBD family.[4] Recent studies have reported that MBD1 deficiency in mice resulted in reduced differentiation of adult neural stem/progenitor cell (aNSCs) and impacted adult neurogenesis and spatial learning.[5] Furthermore, mice lacking functional MBD1 exhibited autism-like behavior, including reduced social interaction, depression and abnormal brain serotonin activity.[6] The MBD1-deficiency-induced cognitive deficits were also studied in human.[7,8] All of these facts suggest that MBD1 plays a critical role in the neural development.

To decipher the mechanism of MBD1 in regulation of neural development, efforts have been made on investigation of the interacting network of MBD1.[9,10,11,6] Through binding to the *Fgf-2* promoter, as a transcriptional repressor of *Fgf-2*, MBD1 is able to maintain the normal neural differentiation in mice.[11] MBD1 also interacts with the promotor of *Htr2c* to regulate the serotonin activity in brain.[6] Additionally, by forming the regulatory networks with microRNA-184 and microRNA-195, MBD1 controls the balance between proliferation and differentiation of aNSCs.[9,11] Although these findings help elucidate how MBD1 regulates

neural development through DNA and RNA, little is known about how MBD1 interacts with proteins to regulate the neural development.

Biological processes are usually regulated by dynamic signaling networks of interacting proteins, therefore, it is essential to probe the protein-protein interaction networks to better understand the function of individual proteins.[12] The versatile combination of affinity purification and mass spectrometry has become the method of choice to advance our understanding of these networks due to the improvements in affinity tags and powerful capabilities of a mass spectrometer.[13] Here, we generated mice which expressed the flag tag containing MBD1 proteins, thus enabling affinity purification. After isolation of MBD1 complexes from the extraction of mice brains, a high resolution and high mass accuracy mass spectrometer was utilized to analyze the isolated MBD1 complexes. Further label-free quantification analysis revealed that twenty-two proteins might be the potential interacting proteins of MBD1. Five of them were identified as significant potential interacting partners of MBD1. Our analysis here provided a list of potential interacting proteins of MBD1 in neural system, which facilitates the future biological validation and functional analysis for MBD1.

EXPERIMENTAL PROCEDURES

Immunoprecipitation (IP) of MBD1 Interacting Proteins

The MBD1 mice were generated according to the previous studies.[5,9] IP experiments were performed as described elsewhere.[14] Brains from six mice were extracted and pooled together for one IP experiment. In total, eighteen MBD1 mice and eighteen control mice were used.

In-solution Digestion

After co-precipitated proteins were sequentially eluted by 8 M Urea and 0.1% TFA, samples were combined and the pH was adjusted to 8 by adding 1 M Tris-HCl (pH=8) buffer. The eluted proteins were reduced with dithiothreitol (DTT) at a final concentration of 5 mM for 45 min at 37 ^{0}C and alkylated with iodoacetamide at a final concentration of 15 mM for 30 min in darkness. Alkylation was quenched by incubation in 5 mM DTT for 15 min. After dilution with 50 mM Tris-HCl (pH=8) to a final concentration of 1 M for urea, 2 μg trypsin (Promega) was added into each sample, and incubation was performed overnight at 37 ^{0}C.The generated peptide mixtures were desalted with 100 μl OMIX C_{18} pipette tips (Agilent Technologies) and dried down under vacuum. Each sample was suspended in 20 μl of 0.1% formic acid (FA) for liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI MS/MS) analysis.

MS Analysis

Each sample (2 μl) was loaded onto a 5 μm, 180 μm × 20 mm C_{18} (Waters Symmetry, Waters Corp.) trap column and separated on a 1.7 μm, 75 μm × 100 mm C_{18} (Waters BEH 130, Waters Corp.) analytical column using a nanoACQUITY UPLC and autosampler system (Waters Corp.). Peptides were separated in a 60 min linear gradient of 5-35% mobile phase B at 300 nl/min (mobile phase A: 0.1% FA in water; mobile phase B: 0.1% FA in acetonitrile (ACN)) and analyzed by a Q Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with the manufacturer's nanospray ion source. The mass spectrometer was operated in a data-dependent mode with the twenty most intense peptide ions ($z \geq 2$) selected for

MS/MS. Full scans were acquired in 350-1800 *m/z* range at a resolution of 70,000 (at *m/z* 400) in the Orbitrap with an automatic gain control (AGC) target value of 1×10^6 charges and maximum fill time of 250 ms. MS/MS spectra were collected at a resolution of 17,500 (at *m/z* 400) with an AGC target value of 2×10^5 charges and maximum fill time of 120 ms. The underfill ratio is 0.1%. A precursor ion isolation width of 2 *m/z* and 30 s of dynamic exclusion were used.

MS Data Analysis

Raw data files were processed by Proteome Discoverer (PD) (v.1.4.0.288, Thermo Fisher Scientific). Peak lists were generated and searched with Sequest HT against an in-house built mouse protein database containing the sequence of flag tagged MBD1 protein. Precursor mass and fragment mass were searched with mass tolerance of 10 ppm and 0.2 Da respectively. Carbamidomethylcysteine was set as fixed modification and oxidized methionine was searched as a variable modification. Up to two missed cleavage sites for trypsin were allowed. The false discovery rate (FDR) was set to 1% for peptide and protein identifications. To perform quantification analysis, a spectral counting approach was used. Spectral counts for each identified protein were determined by the number of peptide spectrum matches (PSMs) in Sequest HT searches. In total, three biological replicates were analyzed and each biological sample had two technical replicates. Number of PSMs in technical replicates was summed for each protein and subject to normalized spectral abundance factor (NSAF) analysis. Details for this method can be found elsewhere.[15] In brief, the NSAF for each protein was calculated after the spectral counts of the

protein normalized to its length and divided by the sum of SpC/L for all the proteins in the experiment. To carry out statistical analysis, the natural log of each NSAF values was calculated followed by *t*-test comparison between the ln(NSAF) from the three biological replicates of tagged MBD1 mice samples and the ln(NSAF) from the three biological replicates of control mice samples.

RESULTS AND DISCUSSION

Optimization of Workflow

To reveal the interacting proteins of MBD1, a proteomic workflow based on affinity purification and label-free quantification has been applied (Figure 3.1). Here, a group of MBD1 mice and a group of control mice were generated to perform quantification analysis. In MBD1 mice, a sequence of flag tag (MDYKDHDGDYKDHDIDYKDDDDK) was imbedded into the sequence of MBD1 in the mouse. The flag tags can enable the affinity purification as the antibody for MBD1 did not successfully capture MBD1 in our western blot results (data was not shown here). Further biological experiments suggested that such extra sequence in MBD1 did not affect the behavior of mouse. The wild-type mice served as the control group, in which none of proteins have the specific interactions with the antibodies used here. Brain from each mouse was extracted. In our first experiments, proteins extracted from fifteen mice brains were pooled and incubated with antibodies in one IP experiment. After extensive washes, co-precipitated proteins were eluted with a large amount of excess flag tags. The samples were then subjected to MS analysis after digestion. About two hundred proteins were identified by MS either from MBD1

mice samples or control mice samples. In single MS analysis, more than 75% of peptide spectra matches (PSMs) were assigned to the sequence of flag tag, suggesting the presence of a large amount of excess flag tags in the samples. Such large amount of excess flag tags was introduced in the elution steps during IP experiments. To evaluate whether such large amount of excess flag tags affected the identification of other proteins and the sensitivity and accuracy of quantification method applied here, the relative abundance change of MBD1 between two groups was assessed. Because flag tagged MBD1 is known to bind to the antibodies, resulting in relatively high abundance in MBD1 mice samples comparing to control. Subsequent label-free quantification experiments revealed that the abundance of MBD1 in MBD1 mice samples was significantly higher than that in the control ($\log_2^{(\text{MBD1/control})} = 6.64$). It confirmed the capability of our strategy to reveal the potential interacting proteins of MBD1.

Six mice instead of fifteen were then tested in one IP experiment in order to obtain the best results with minimum number of mice in one experiment. MS investigation identified about fifty proteins either from control or MBD1 mice samples. Meanwhile, the percentage of PSMs of flag tag in one MS analysis increased to 90% in comparison with 75% in fifteen mice per IP experiment. Quantification analysis revealed that the abundance change of MBD1 between two groups was less than fivefold. These findings indicated that the large amount of excess flag tags significantly affected the detection of other peptides and the sensitivity of quantification when six mice were used in one IP experiment. To solve

this problem, an elution with reduced amount of flag tags and exclusion list enabled in data-dependent MS analysis were used. However, the number of proteins identified was still low when six mice were used in one IP experiment. A 10 kDa molecular weight cut-off filter was then applied before digestion to further remove the excess flag tags in the samples. But this additional step resulted in no identification for MBD1 during MS analysis due to the increased sample loss. Thus, alternative elution buffers were tested to replace the elution with the flag tags. In each IP experiment, samples were sequentially eluted with 8 M Urea and 0.1% TFA to avoid the usage of the excess flag tags but achieving completely elution. In comparison with the results from fifteen mice per IP experiment, the number of proteins identified in a single MS analysis almost doubled compared to those obtained from either control or MBD1 mice samples. This increase can be explained by the increased amount of contaminants due to non-specific elution of 8M urea and 0.1 % TFA. Another possible explanation might be that the excess flag tags obscure the signal of low abundance proteins in MS analysis. After the quantification analysis for three biological replicates (total of thirty-six mice), the abundance of MBD1 was found significantly higher in MBD1 mice samples than that in control ($\log_2^{(MBD1/control)} = 8.30$, $p=9.53^{-5}$), suggesting that the increased number of proteins did not affect the ability of our quantification strategy to discover the potential interacting proteins of MBD1. Finally, six mice and elution with 8M urea and 0.1 % TFA were selected for our study.

Identification and Label-free Quantification Analysis

Three biological replicates were analyzed in order to achieve statistical significance. For each biological replicate, six wild-type mice and six MBD1 mice were used. In total, thirty-six mice were scarified for the entire analysis. Each biological replicate was analyzed twice by MS. The generated MS data were searched by PD with 1% FDR, resulting in 5444 peptides and 795 proteins identified. To distinguish the potential interacting proteins from contaminants, a spectral counting based quantification approach was utilized.[15] A total of 568 proteins identified at least in two biological replicates with minimum one unique peptide were selected for quantification analysis. The number of PSMs for each identified protein was normalized to the total number of PSMs across all the replicates. Zero spectral count values were replaced by a fraction which was determined by itinerantly replacing 0 to 1 number until ln(NSAF) producing a normal distribution. Log_2 ratios of individual proteins were calculated by comparing the value of NSAF between two group samples. *p* values were determined by doing the *t*-test for the ln(NSAF) of individual proteins across all the biological samples.

Twenty-two proteins with $Log_2^{(MBD1/Control)} > 2$ were considered as significant increase in MBD1 mice samples (Table 2.1). It is not surprising that most of proteins found in Table I are involved in transcriptional process since MBD1 was reported as a transcriptional regulator through interacting with methylated and unmethylated promoters.[16,17] Furthermore, some of proteins, like Smarcb 1, have been previously reported to play a role in the neural development.[18] Among them, six proteins, including MBD1, Gatad2b, Smarcb1, Ruvbl2, Smarca2 and Rab10, had $p \leq 0.05$.

The proteins with $p \leq 0.05$ are identified as the significant potential interacting proteins of MBD1. Representative MS/MS spectra for those significant potential interacting proteins (Gatad2d, Smarcb1, Ruvbl2, Smarca2 and Rab10) are shown in Figure 3.2. Additional biological validations are required to confirm whether they are the interacting proteins of MBD1.

CONCLUSIONS

Here, 795 proteins were identified in AP-MS, while five of them were identified as significant potential interacting proteins of MBD1 by quantification analysis via spectral counting, which suggests the specificity of our method. Our studies have advanced the understanding of the function of MBD1 in neural development by revealing a list of candidates of interacting proteins of MBD1. Future biological validations are required to confirm our findings.

REFERENCES

1. Wilson, G. G.; Murray, N. E., Restriction and modification systems. *Annual review of genetics* **1991,** *25*, 585-627.

2. Okano, M.; Bell, D. W.; Haber, D. A.; Li, E., DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **1999,** *99* (3), 247-57.

3. Li, E.; Bestor, T. H.; Jaenisch, R., Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **1992,** *69* (6), 915-26.

4. Klose, R. J.; Bird, A. P., Genomic DNA methylation: the mark and its mediators. *Trends in biochemical sciences* **2006,** *31* (2), 89-97.

5. Zhao, X.; Ueba, T.; Christie, B. R.; Barkho, B.; McConnell, M. J.; Nakashima, K.; Lein, E. S.; Eadie, B. D.; Willhoite, A. R.; Muotri, A. R.; Summers, R. G.; Chun, J.; Lee, K. F.; Gage, F. H., Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** *100* (11), 6777-82.

6. Allan, A. M.; Liang, X.; Luo, Y.; Pak, C.; Li, X.; Szulwach, K. E.; Chen, D.; Jin,
P.; Zhao, X., The loss of methyl-CpG binding protein 1 leads to autism-like behavioral deficits. *Human molecular genetics* **2008,** *17* (13), 2047-57.

7. Cukier, H. N.; Rabionet, R.; Konidari, I.; Rayner-Evans, M. Y.; Baltos, M. L.; Wright, H. H.; Abramson, R. K.; Martin, E. R.; Cuccaro, M. L.; Pericak-Vance, M. A.; Gilbert, J. R., Novel variants identified in methyl-CpG-binding domain genes in autistic individuals. *Neurogenetics* **2010,** *11* (3), 291-303.

8. Li, H.; Yamagata, T.; Mori, M.; Yasuhara, A.; Momoi, M. Y., Mutation analysis of methyl-CpG binding protein family genes in autistic patients. *Brain & development* **2005,** *27* (5), 321-5.

9. Liu, C.; Teng, Z. Q.; Santistevan, N. J.; Szulwach, K. E.; Guo, W.; Jin, P.; Zhao, X., Epigenetic regulation of miR-184 by MBD1 governs neural stem cell proliferation and differentiation. *Cell stem cell* **2010,** *6* (5), 433-44.

10. Liu, C.; Teng, Z. Q.; McQuate, A. L.; Jobe, E. M.; Christ, C. C.; von Hoyningen-Huene, S. J.; Reyes, M. D.; Polich, E. D.; Xing, Y.; Li, Y.; Guo, W.;

Zhao, X., An epigenetic feedback regulatory loop involving microRNA-195 and MBD1 governs neural stem cell differentiation. *PloS one* **2013,** *8* (1), e51436.

11. Li, X.; Barkho, B. Z.; Luo, Y.; Smrt, R. D.; Santistevan, N. J.; Liu, C.; Kuwabara, T.; Gage, F. H.; Zhao, X., Epigenetic regulation of the stem cell mitogen Fgf-2 by Mbd1 in adult neural stem/progenitor cells. *The Journal of biological chemistry* **2008,** *283* (41), 27644-52.

12. Lu, L. J.; Sboner, A.; Huang, Y. J.; Lu, H. X.; Gianoulis, T. A.; Yip, K. Y.; Kim, P. M.; Montelione, G. T.; Gerstein, M. B., Comparing classical pathways and modern networks: towards the development of an edge ontology. *Trends in biochemical sciences* **2007,** *32* (7), 320-31.

13. Gingras, A. C.; Gstaiger, M.; Raught, B.; Aebersold, R., Analysis of protein complexes using mass spectrometry. *Nature reviews. Molecular cell biology* **2007,** *8* (8), 645-54.

14. Schulze, W. X.; Mann, M., A novel proteomic screen for peptide-protein interactions. *The Journal of biological chemistry* **2004,** *279* (11), 10756-64.

15. Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P., Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *Journal of proteome research* **2006,** *5* (9), 2339-47.

16. Jorgensen, H. F.; Ben-Porath, I.; Bird, A. P., Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Molecular and cellular biology* **2004,** *24* (8), 3387-95.

17. Nakao, M.; Matsui, S.; Yamamoto, S.; Okumura, K.; Shirakawa, M.; Fujita, N., Regulation of transcription and chromatin by methyl-CpG binding protein MBD1. *Brain & development* **2001,** *23 Suppl 1*, S174-6.

18. Lessard, J.; Wu, J. I.; Ranish, J. A.; Wan, M.; Winslow, M. M.; Staahl, B. T.; Wu, H.; Aebersold, R.; Graef, I. A.; Crabtree, G. R., An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **2007,** *55* (2), 201-15.

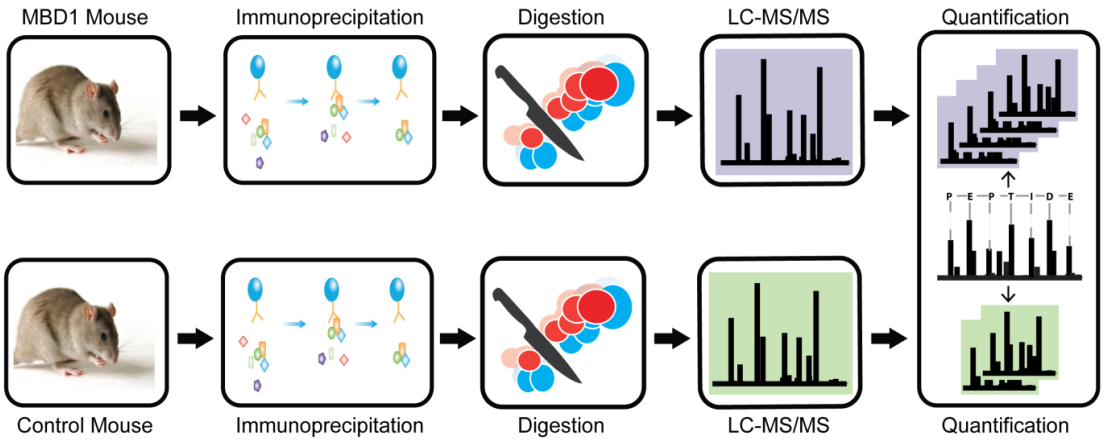


Figure 3.1 Workflow applied to reveal the interacting proteins of MBD1. Two groups of mice, MBD1 and control, were generated. Proteins were extracted from six mice brain for one IP experiment. The co-precipitated proteins were digested by trypsin and subjected to LC/MS analysis. A spectral counting based quantification approach was employed to distinguish the potential interacting proteins of MBD1 from contaminants.

Accession	Gene Symbol	ΣCoverage	avg.NSAF_MBD1(×10^{-6})	avg.NSAF_Control(×10^{-6})	$log_2^{(MBD1/Control)}$
***Q9Z2E2**	Mbd1	30.96	1897.32	6.00	8.3055
***Q9Z0H3-2**	Smarcb1	9.57	292.96	19.26	3.9268
***E9QAB8**	Smarca2	5.10	26.71	2.62	3.3513
Q923G2	Polr2h	23.33	262.79	26.35	3.3182
F8WHY8	Mta1	7.16	55.50	5.66	3.2931
***Q9WTM5**	Ruvbl2	18.79	142.37	15.64	3.1860
***Q8VHR5**	Gatad2b	11.95	135.44	15.33	3.1436
***P61027**	Rab10	11.00	313.25	36.22	3.1126
P14094	Atp1b1	8.88	202.87	23.83	3.0899
P61264	Stx1b	6.60	108.03	13.72	2.9769
A2A6A1	Gpatch8	3.99	34.62	4.81	2.8467
P63011	Rab3a	12.27	393.65	56.34	2.8047
O54946-2	Dnajb6	9.50	93.54	16.33	2.5180
Q9D198	Syf2	2.89	93.54	16.33	2.5180
Q9QZE7	Tsnax	3.10	76.93	13.63	2.4970
Q3UID0	Smarcc2	31.42	588.76	108.00	2.4466
Q9CXE2	Bcl7a	17.62	170.82	34.49	2.3082
Q9JME5	Ap3b2	12.38	184.23	39.06	2.2376
P60122	Ruvbl1	27.19	310.20	68.37	2.1818
P63024	Vamp3	23.30	394.54	88.39	2.1583
F7C528	Chd3	2.49	16.81	3.84	2.1296
P63271	Supt4h1a	27.35	263.48	61.91	2.0896

Table 3.1 Proteins identified as significant increase in MBD1 samples. Proteins presented here were identified at least in two biological replicates with minimum one unique peptide and $Log_2^{(MBD1/Control)} > 2$.

Note: accession is the Uniprot accession number for the identified protein. ΣCoverage is the sequence coverage of identified proteins. avg.NSAF_MBD1 (×10^{-6}) is the normalized spectral abundance factor representing proteins expression in MBD1 mice samples, average across three biological replicates. avg.NSAF_Control (×10^{-6}) is the normalized spectral abundance factor representing proteins expression in control samples, average across three biological replicates. * $p \leq 0.05$ in the *t-test*.

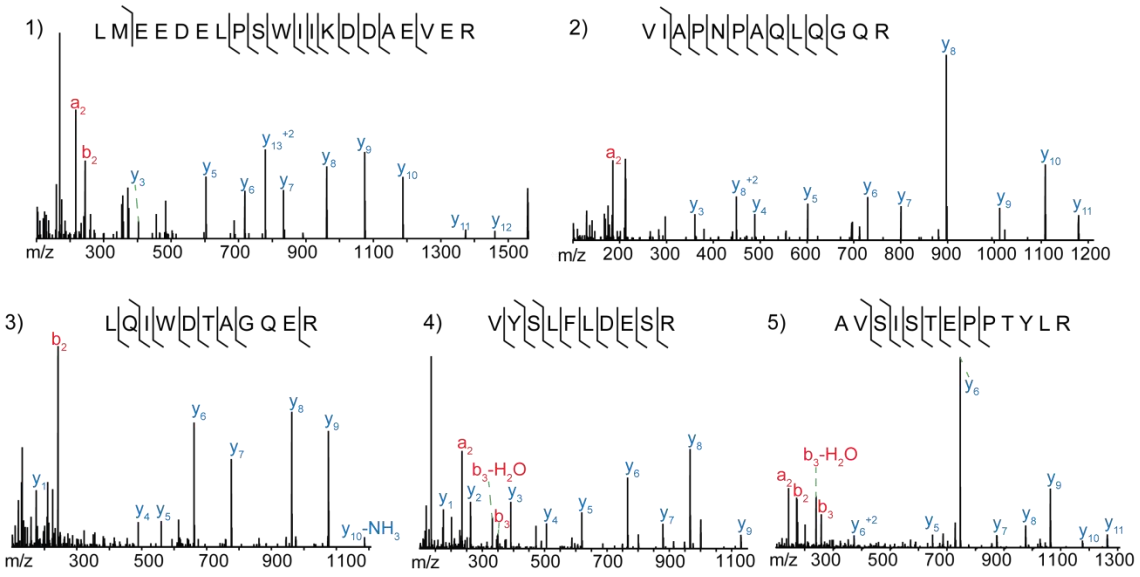


Figure 3.2 Representative MS/MS spectra for five potential interacting proteins of MBD1 (Gatad2d, Smarcb1, Ruvbl2, Smarca2 and Rab10). 1) a representative MS/MS spectrum for a tryptic peptide (precursor m/z: 806.3893, +2) from Smarca2; 2) a representative MS/MS spectrum for a tryptic peptide (precursor m/z: 696.3909, +2) from Gatad2b; 3) a representative MS/MS spectrum for a tryptic peptide (precursor m/z: 658.8333, +2) from Rab10; 4) a representative MS/MS spectrum for a tryptic peptide (precursor m/z: 688.2952, +3) from Ruvbl2; 5) a representative MS/MS spectrum for a tryptic peptide (precursor m/z: 717.3858, +2) from Smarcb1.

Chapter 4

Molecular Basis of Crosstalk between Oncogenic Ras and the Master Regulator of Hematopoiesis GATA-2

Adapted from: "Molecular basis of crosstalk between oncogenic Ras and the master regulator of hematopoiesis GATA-2." Katsumura, K.R., **Yang, C.**, Boyer, M.E., Li, L., Bresnick, E.H. *EMBO reports* 2014, 15, (9), 938-947

ABSTRACT

Disease mutations provide unique opportunities to decipher protein and cell function. Mutations in the master regulator of hematopoiesis GATA-2 underlie an immunodeficiency associated with myelodysplastic syndrome and leukemia. We discovered that a GATA-2 disease mutant (T354M) defective in chromatin binding was hyperphosphorylated by p38 Mitogen Activated Protein Kinase. p38 also induced multisite phosphorylation of wild type GATA-2, which required a single phosphorylated residue (S192). Phosphorylation of GATA-2, but not T354M, stimulated target gene expression. While crosstalk between oncogenic Ras and GATA-2 has been implicated as an important axis in cancer biology, its mechanistic underpinnings are unclear. Oncogenic Ras enhanced S192-dependent GATA-2 phosphorylation, nuclear foci localization, and transcriptional activation. These studies define a mechanism that controls a key regulator of hematopoiesis and a dual mode of impairing GATA-2-dependent genetic networks: mutational disruption of chromatin occupancy yielding insufficient GATA-2, and oncogenic Ras-mediated amplification of GATA-2 activity.

INTRODUCTION

The establishment and maintenance of genetic networks must be exquisitely orchestrated to ensure the fidelity of biological processes. While intense efforts are documenting genetic networks, many questions remain unanswered regarding how master regulator establish/maintain networks, how signaling mechanisms control network, and how alterations influence vulnerable nodes within a network. As GATA transcription factors control genetic networks in diverse contexts,[1] and pathological GATA factor mutants have been described,[2] it is instructive to investigate the consequences of such mutations.

Of the six mammalian GATA factors, GATA-1, GATA-2, and GATA-3 have important roles in regulating hematopoiesis. [1a, 3] GATA-1 regulates erythropoiesis, megakaryopoiesis, and the development of eosinophils and mast cells,[4] while GATA-2, controls primitive and definitive hematopoiesis.[5] GATA-2 induces HSC generation in the AGM and the HSPCs in the fetal liver.[6] As *Gata2*$^{+/-}$ HSCs are impaired,[7] and GATA-2 overexpression suppresses hematopoiesis,[8] increases and decreases in GATA-2 disrupt hematopoiesis.

Alterations in GATA-2 expression are implicated in myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML).[9] Elevated GATA-2 expression correlates with AML severity.[10] Heterozygous GATA-2 mutations cause an immunodeficiency (MonoMAC or Emberger syndrome) associated with MDS/AML.[6a,11,12] While these mutations, for example, T354M, often occur within the

GATA-2 DNA binding zinc finger, mutations of the +9.5 cis-element can also cause MonoMAC.[6a, 13]

GATA-2 (T354M) expressed in 293 cells was not competent for DNA binding.[9a] However, certain factors defective in DNA binding can function *in vivo,*[14] and chromatin occupancy can be regulated independent of DNA binding.[15] While dissecting how leukemogenic mutations alter GATA-2, we discovered that the functionally impaired T354M mutant was hyperphosphorylated, and a Ras-p38-dependent phosphorylation mechanism amplified wild-type GATA-2 activity.

EXPERIMENTAL PROCEDURES

Cell Culture

G1E and MAE cells were maintained as described.[6a, 12]

Plasmids

GATA-2 cDNA was cloned into pcDNA4TOFHA vector (from Dr. Danny Reinberg, NYU). H-Ras and H-Ras (G12V) expression vectors were from Dr. Jing Zhang (UW-Madison).

Quantitative Chromatin Immunoprecipitation (ChIP)

ChIP analysis in G1E cells was conducted as described.[16]

Protein Analysis

For analysis of phosphorylation, protein was prepared with RIPA buffer containing 50 mM β-glycerophosphate, 50 mM sodium fluoride, and 200 μM sodium vanadate. For phosphatase treatment, protein was incubated with 800 units of λ-

phosphatase (New England Biolabs) in 30 µl (293 cells) and 80 µl (G1E) at 30°C for 90 min.

Immunoprecipitation

Protein prepared by lysing 293 cells in RIPA buffer was immunoprecipitated with preimmune serum or rabbit anti-GATA-2 antibody.[12]

Statistical Analysis

Statistical significance was determined by Paired Student's *t*-test using GraphPad software (www.graphpad.com).

ChIP-Seq

ChIP-seq profiles for GATA-2 in HUVEC cells [17] and human CD34-positive cells [18] were generated using the UCSC Genome Browser (http://genome.ucsc.edu/). (GEO accession GSE29531).

Mass Spectrometry

GATA-2-containing bands were excised and cut into small pieces (1-5 mm^3 per piece). Each gel slice was destained with 200 μl of methanol:50 mM NH_4HCO_3 (1:1 v/v) for 1 min with vortexing. The supernatant was removed, and destaining step was repeated until the gel become colorless. The dehydration of each gel slice was performed by incubation in 200 μl of acetonitrile (ACN):50 mM NH_4HCO_3 (1:1 v/v) for 5 min followed by incubation in 200 μl of ACN for 30 seconds with vortex. After drying the gel with a Speed-Vac, proteins were alkylated with 100 μl of 25 mM DTT in 50 mM NH_4HCO_3 at 56°C for 20 min and reduced with 100 μl of 55 mM IAA in 50 mM NH_4HCO_3 in the dark for 20 min at room temperature. The supernatant was

removed, and each gel slice was washed with 200 μl of H_2O for 30 seconds with vortex. Gel slices were dehydrated and dried as mentioned before. Dried gel slices were covered with trypsin (Sequencing grade, Promega) (12 ng/μl in 0.01% ProteaseMAX^{TM} Surfactant (Promega): 50mM NH_4HCO_3) and incubated on ice for 30min to rehydrate gel slices. Each gel slice was overlayed with 50 μl of 0.01% ProteaseMAX^{TM} Surfactant:50 mM NH_4HCO_3 and incubated at 37°C for 2 hours. The supernatant containing extracted peptides was transferred into a new tube, and 10% TFA was added to inactivate trypsin with a final pH of ~2. To completely extract peptides from the gel, each gel slice was further extracted by the addition of 100 μl of 5% formic acid (FA): ACN (1:2 v/v) and incubated at 37°C for 15 min. The extraction was repeated twice, and extracts from all steps were combined. Extracts were vacuum-dried and resuspended in 20 μl of 0.1% FA for liquid chromatography electrospray tandem mass spectrometry (LC-ESI MS/MS) analysis.

3 μl of each sample was loaded onto a 5μm, 180 μm × 20 μm C18 (Waters Corp.) precolumn and separated on a 1.7μm, 100μm × 100μm C18 (Waters BEH 130, Waters Corp.) analytical column using a Waters nanoACQITY UPLC and autosampler (Waters Corp.). Peptides were separated and eluted from the column with a 60 min linear gradient of 5-35% mobile phase B at 300 nl/min (mobile phase A: 0.1% FA in water; mobile phase B: 0.1% FA in CAN).The eluate was introduced into the mass spectrometer via electrospray ionization (+2.1 kV), and peptide ions were subjected to tandem mass spectrometry using a Q Exactive mass spectrometer (Thermo Fisher Scientific, Bermen, Germany). The mass spectrometer was operated

in the data-dependent mode. Intact masses were measured in full scans with resolving power of 70,000 (at *m/z* 400) in the Orbitrap using an automated gain control (AGC) target value of 1×10^6 charges. The twenty most intense peptide ions ($z \geq 2$)) were fragmented in the HCD cell. MS/MS spectra were acquired at a resolution of 17,500 (at *m/z* 400) with an AGC target value of 1×10^5 charges and underfill ratio of 0.1%. A precursor ions isolation width of 2 *m/z* and 10s of dynamic exclusion were used.

Raw MS/MS spectra were processed with Proteome Discoverer (v.1.4.0.288, Thermo Fisher Scientific). Peak lists were generated and searched with Sequest HT against an in-house built human database containing the sequences of wild-type and mutated mouse GATA-2. A 10 ppm of precursor mass tolerance and 0.2 Da of fragment mass tolerance were used for Sequest HT. Carbamidomethylcysteine was set as fixed modification, and oxidized methionine, phosphorylation on serine, threonine, or tyrosine were searched as variable modifications. In addition, up to three missed cleavage sites for trypsin were allowed. The identified peptides were filtered by XCorr scores, and positions of phosphorylated residues were determined by site probabilities calculated by phosphoRS[19]. Peptides whose XCorr scores were lower than the default XCorr confidence thresholds in Sequest HT were discarded. Only phosphorylation sites assigned with a localization probability of at least 0.75 were considered and validated manually. Relative amounts of peptides were estimated by the peak area of the respective peptides in MS spectra. Peak areas were calculated by "Precursor Ions Area Detector" node in Proteome Discoverer. Peak area of each

peptide of interest was further normalized to the average protein amount in each MS analysis.[20]

RESULTS AND DISCUSSION

GATA-2 Disease Mutation Reduces Chromatin Occupancy and Target Gene Regulation

To determine how the T354M mutation (Figure 4.1A) influences GATA-2 function, we expressed GATA-2 (T354M) in G1E proerythroblasts, which express endogenous GATA-2.[21] Western blot analysis with anti-GATA-2 (Figure 4.1B, left) and anti-HA (Figure 4.1B, right) antibodies revealed GATA-2 (T354M) migration as two bands – with a mobility indistinguishable from wild-type GATA-2 or with a slower mobility. Wild type GATA-2 also migrated as two bands, although the upper band was less abundant versus GATA-2 (T354M).

Using the anti-HA antibody and a quantitative ChIP assay, we compared the capacities of expressed GATA-2 or GATA-2 (T354M) to occupy chromatin (Figure 4.1C).[12, 22] Whereas GATA-2 occupied the -3.9, -2.8, -1.8 and +9.5 GATA switch sites of the *Gata2* locus and the *Lyl1* locus, little to no GATA-2 (T354M) chromatin occupancy was detected (Figure 4.1C). GATA-2 (T354M) was expressed at least as high as GATA-2 (Figure 4.1B, right).

The reduced GATA-2 (T354M) chromatin occupancy suggested a defect in target gene regulation. A system does not exist to test whether exogenously expressed GATA-2 can regulate endogenous target genes. G1E cells express high-level endogenous GATA-2 and are not ideal for addressing this issue. We developed a

Mouse Aortic Endothelial (MAE) cell system in which exogenous GATA-2 regulates endogenous target genes. Immortalized MAE cells bear a normal endothelial phenotype and express endogenous GATA-2 considerably lower than G1E cells.[23]

Expression profiling in PECAM1^{+} cells and Aorta Gonad Mesonephros (AGM) from control and +9.5$^{-/-}$ embryos identified endogenous GATA-2 target genes.[6a, 24] We tested whether GATA-2 regulates these genes in MAE cells. Transiently expressed GATA-2 and GATA-2 (T354M) in MAE cells migrated as two bands, analogous to G1E cells, with the upper band more abundant with GATA-2 (T354M) versus GATA-2 (Figure 4.1D). GATA-2 activated *Hdc*, *IL13*, *Gfi1b*, and *Gfi1* expression 200, 10, 11, and 10-fold, respectively, whereas GATA-2 (T354M) had little activity (Figure 4.1E). The GATA-2 target genes *Gfi1b* and *Gfi1* function in hemogenic endothelium to control HSC generation.[22a, 24-25] Although *Hdc* and *IL13* are not established GATA-2 targets, ChIP-seq analysis in HUVECs and human CD34^{+} hematopoietic precursors revealed GATA-2 occupancy at *Hdc* and *IL13* loci (Figure 4.1F and Supplementary Figure S1A).[17, 26] Cotransfection of GATA-2 (T354M) did not influence GATA-2-mediated *Hdc* induction (Supplementary Figure S1B). Thus, T354 is critical for GATA-2 chromatin occupancy and target gene regulation.

p38 Mitogen-activated Protein Kinase-dependent GATA-2 Multisite Phosphorylation Governs GATA-2 Activity

To determine the basis of the T354M-enhanced mobility shift, GATA-2 or GATA-2 (T354M) was expressed in 293 cells and cell lysates were treated with λ-

phosphatase. λ-phosphatase abolished the GATA-2 (T354M) upper band and increased GATA-2 mobility equivalent to λ-phosphatase-treated GATA-2 (T354M) (Figure 4.2A). Screening signaling pathway inhibitors revealed that the p38 mitogen-activated protein kinase (MAPK) inhibitor SB203580, but not ERK (U0126) or JNK (SP600125) inhibitors, decreased the upper and increased the lower band (Figure 4.2B, Supplementary Figure S2A and B). In G1E lysates, λ-phosphatase (Figure 4.2C) and SB203580 (Figure 4.2E) decreased abundance of the GATA-2 (T354M) upper band and increased endogenous GATA-2 mobility (Figure 4.2D and F). The MDS mutant GATA-2 (D355T) and a C349A DNA binding-defective mutant were hyperphosphorylated (Supplementary Figure 2C).[9a] p38α knockdown reduced GATA-2 (T354M) hyperphosphorylation (Figure 4.2G) and wild-type GATA-2 regulation of target genes (Figure 4.2H). The protein phosphatase inhibitor okadaic acid induced GATA-2 hyperphosphorylation and GATA-2 target gene expression (*Hdc* and *Gfi1*) (Figure 4.2I and J). Thus, GATA-2 is phosphorylated in a p38α-dependent manner, and leukemogenic mutations that abrogate chromatin occupancy promote p38α-dependent phosphorylation and/or impair GATA-2 dephosphorylation (Figure 4.2). These results suggest that p38α phosphorylation of GATA-2 controls wild-type GATA-2 activity.

Mass spectrometry was used to identify GATA-2 and GATA-2 (T354M) phosphorylated residues (Figure 4.3A and 4.4). Analysis of proteins immunopurified from 293 cells revealed multi-site phosphorylation (Figure 4.3B) that was more abundant in GATA-2 (T354M) versus GATA-2 (Figure 4.3C). Unbiased deletion

mutants and point mutants that eliminate potential phosphorylatable residues were analyzed to delineate requirements for GATA-2 (T354M) hyperphosphorylation (Supplementary Table S1). S192A was the only point mutation that abrogated the shift (Figure 4.3D). While deleting amino acids 61–120 abrogated the shift, a smaller deletion including S73 and S119 did not (Supplementary Figure S3B–D and Supplementary Table S1). S192A mutation also increased GATA-2 mobility, resembling λ-phosphatase treatment (Figure 4.3D and E). Phosphonet (http://www.phosphonet.ca/) analysis suggested that S192 can be phosphorylated by p38α. Though GATA-2 S192 is evolutionarily conserved, other GATA factors lack a comparable serine. GATA-3 contains a potentially analogous threonine (Figure 4.3F). S192A reduced GATA-2 phosphorylation at S73, S119, and S290 (Figure 4.3G and H). Thus, Ser192 is phosphorylated, promotes multi-site phosphorylation, and is required for GATA-2(T354M) hyperphosphorylation.

In MAE cells, GATA-2 (S192A) (Figure 4.3I) had some capacity to activate *Hdc* and *IL13*, but its activity was significantly less than GATA-2 (Figure 4.3J). In G1E cells, HA-GATA-2 (S192A) chromatin occupancy was ~40% less than HA-GATA-2 (Figure 4.3K and L). As Ser192 is required for GATA-2 (T354M) hyperphosphorylation, and GATA-2 (T354M) is defective in chromatin binding and transcriptional activation, chromatin binding and transcriptional activation are not required for hyperphosphorylation. A S192A/T354M mutant was also inactive (Supplementary Figure S3E).

GATA-2 S192 Integrates Oncogenic Signals

GATA-2-Ras interactions are implicated in non-small cell lung and colon cancer.[27] Because Ras signaling activates p38α, p38α-dependent Ser192 phosphorylation may underlie Ras-GATA-2 crosstalk. Constitutively active Ras [Ras(G12V)] induced the upper GATA-2 band, which was abrogated by λ-phosphatase (Figure 4.5A), and was indistinguishable from the T354M-induced band; both required Ser192 and amino acids 61–120 (Figure 4.3D and 4.5B and Supplementary Figure S3C) and were reduced by SB203580 (Figure 4.2B and 4.5C). Though the function of amino acids 61–120 is not established, GATA factor N-termini can facilitate transactivation.[2a] Ras (G12V) expression in MAE cells enhanced GATA-2-mediated *Hdc* and *Gfi1* expression 10 and 5-fold, respectively, without affecting the weak GATA-2 (T354M) activity (Figure 4.5D and E). S192A attenuated GATA-2/Ras (G12V)-mediated *Hdc* induction (Figure 5F). Ras (G12V) expression in G1E cells enhanced GATA-2-mediated *Hdc*, *IL13*, and *Gfi1* expression 6, 3, and 2.5-fold, respectively (Figure 4.5G). Ras (G12V) increased GATA-2, but not S192A and Δ61–120, localization into nuclear foci [13.9–25.4%,− and + Ras (G12V), respectively] (Figure 4.6A and B). SB203580 reduced GATA-2 foci localization (Figure 4.6C and D). The foci partially colocalized with serine 2-phosphorylated Pol II, an active transcription marker (Figure 4.6E). Thus, p38α and Ras (G12V) regulate GATA-2 phosphorylation, subnuclear localization, and transcriptional activation (Figure 4.6F).

Mechanistic Considerations

Dissecting the deficits of a GATA-2 disease mutant uncovered a signal-dependent GATA factor pathway. While GATA-2 can be phosphorylated by cyclin-dependent kinases, Akt and MAPKs, the modified residues and mechanistic consequences were not known.[28] p38-dependent GATA-3 phosphorylation facilitates importin-α binding and cytoplasmic to nuclear translocation and regulates LT-HSC self-renewal.[29]

As p38α inhibition attenuated GATA-2 hyperphosphorylation and target gene activation (Figure 4.2 and 4.5F), this mechanism controls wild-type GATA-2 function. Besides regulating LT-HSC self-renewal via targeting GATA-3, p38 mediates reactive oxygen speciesdependent reductions in HSC lifespan and function of the HSC regulator thrombopoietin.[29b, 30] Inhibiting p38 restores hematopoiesis in defective MDS progenitors.[31] Elevated p38 activity may therefore target Ser192, increasing GATA-2 activity and deregulating the genetic network. By decreasing S192-dependent phosphorylation, p38 inhibition would reduce GATA-2 activity to a level compatible with normal hematopoiesis.

Ser192-dependent phosphorylation facilitates GATA-2 chromatin occupancy and target gene activation. Ser192 mediates GATA-2 and T354M phosphorylation. As Ser192 increased GATA-2 activity, Ser192-dependent hyperphosphorylated isoform is a minority of GATA-2, and phosphatase inhibition induced GATA-2 hyperphosphorylation, efficient GATA-2 dephosphorylation or dephosphosphorylation of a component in the activation mechanism restricts GATA-

2 activity. This mechanism would limit GATA-2 activity and HSPC generation/function. Since hyperphosphorylated T354M is defective in chromatin binding/target gene regulation, failure to dephosphorylate T354M would be inconsequential.

GATA-2–cancer links and oncogenic Ras targeting of Ser192, suggest the Ras-p38-GATA-2 axis may have vital roles in disease etiology.[2a, 32] It will be instructive to decipher how GATA-2 phosphorylation is opposed physiologically and how this balance is skewed in GATA-2-dependent pathologies not involving GATA-2 mutations that impair chromatin occupancy.

CONCLUSIONS

Here, we demonstrate that a disease mutation (T354M) in the GATA-2 DNA binding domain reduced GATA-2 chromatin occupancy and target gene expression and induced GATA-2 hyperphosphorylation. p38 Mitogen Activated Protein Kinase induced multi-site phosphorylation of wild type GATA-2 and the T354M mutant, which required a single phosphorylated residue (S192). Only phosphorylation of wild type GATA-2 stimulated target gene expression, as the T354M mutant was defective in chromatin binding. Oncogenic Ras enhanced S192-dependent GATA-2 phosphorylation, nuclear foci localization, and transcriptional regulation. These studies define a mechanism that controls a vital regulator of hematopoiesis and a dual mode of GATA-2 dysregulation in cancer: mutational disruption of chromatin occupancy, yielding insufficient GATA-2 to orchestrate hematopoiesis, and

oncogenic Ras-mediated amplification of GATA-2 activity, disrupting the GATA-2-dependent genetic network.

Supplemental Information Available Upon Request

REFERENCES

1. (a) Weiss, M. J.; Orkin, S. H., GATA transcription factors: key regulators of hematopoiesis. *Experimental hematology* **1995,** *23* (2), 99-107; (b) Charron, F.; Nemer, M., GATA transcription factors and cardiac development. *Semin. Cell. Dev. Biol.* **1999,** *10* (1), 85-91; (c) Craven, S. E.; Lim, K. C.; Ye, W.; Engel, J. D.; de Sauvage, F.; Rosenthal, A., Gata2 specifies serotonergic neurons downstream of sonic hedgehog. *Development* **2004,** *131* (5), 1165-1173; (d) Viger, R. S.; Guittot, S. M.; Anttonen, M.; Wilson, D. B.; Heikinheimo, M., Role of the GATA family of transcription factors in endocrine development, function, and disease. *Mol Endocrinol* **2008,** *22* (4), 781-98; (e) De Val, S.; Black, B. L., Transcriptional control of endothelial cell development. *Dev. Cell* **2009,** *16* (2), 180-195.

2. (a) Bresnick, E. H.; Katsumura, K. R.; Lee, H. Y.; Johnson, K. D.; Perkins, A. S., Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. *Nucleic acids research* **2012,** *40* (13), 5819-31; (b) Crispino, J. D., GATA1 in normal and malignant hematopoiesis. *Seminars in cell & developmental biology* **2005,** *16* (1), 137-47; (c) Rodrigues, N. P.; Tipping, A. J.; Wang, Z.; Enver, T., GATA-2 mediated regulation of normal hematopoietic stem/progenitor cell function, myelodysplasia and myeloid leukemia. *Int J Biochem Cell Biol* **2012,** *44* (3), 457-60.

3. Kim, S. I.; Bresnick, E. H., Transcriptional control of erythropoiesis: emerging mechanisms and principles. *Oncogene* **2007,** *26* (47), 6777-94.

4. (a) Tsai, S. F.; Martin, D. I.; Zon, L. I.; D'Andrea, A. D.; Wong, G. G.; Orkin, S. H., Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* **1989,** *339* (6224), 446-51; (b) Simon, M. C.; Pevny, L.; Wiles, M. V.; Keller, G.; Costantini, F.; Orkin, S. H., Rescue of erythroid development in gene targeted GATA-1- mouse embryonic stem cells. *Nat. Genet.* **1992,** *1* (2), 92-98; (c) Shivdasani, R. A.; Fujiwara, Y.; McDevitt, M. A.; Orkin, S. H., A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in mekagaryocyte growth and platelet development. *EMBO J.* **1997,** *16*, 3965-3973; (d) Migliaccio, A. R.; Rana, R. A.; Sanchez, M.; Lorenzini, R.; Centurione, L.; Bianchi, L.; Vannucchi, A. M.; Migliaccio, G.; Orkin, S. H., GATA-1 as a regulator of mast cell differentiation revealed by the phenotype of the GATA-1low mouse mutant. *J Exp Med* **2003,** *197* (3), 281-96; (e) Yu, C.; Cantor, A. B.; Yang, H.; Browne, C.; Wells, R. A.; Fukiwara, Y.; Orkin, S. H., Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. *J. Exp. Med.* **2002,** *195* (11), 1387-1395; (f) Fujiwara, Y.; Browne, C. P.; Cunniff, K.; Goff, S. C.; Orkin, S. H., Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93* (22), 12355-12358; (g) Pevny, L.; Simon, M.

C.; Robertson, E.; Klein, W. H.; Tsai, S. F.; D'Agati, V.; Orkin, S. H.; Costantini, F., Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **1991,** *349* (6306), 257-260.

5. (a) Tsai, F. Y.; Keller, G.; Kuo, F. C.; Weiss, M.; Chen, J.; Rosenblatt, M.; Alt, F. W.; Orkin, S. H., An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature* **1994,** *371* (6494), 221-6; (b) Tsai, F. Y.; Orkin, S. H., Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood* **1997,** *89* (10), 3636-43.

6. (a) Johnson, K. D.; Hsu, A. P.; Ryu, M. J.; Wang, J.; Gao, X.; Boyer, M. E.; Liu, Y.; Lee, Y.; Calvo, K. R.; Keles, S.; Zhang, J.; Holland, S. M.; Bresnick, E. H., Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. *The Journal of clinical investigation* **2012,** *122* (10), 3692-704; (b) Gao, X.; Johnson, K. D.; Chang, Y.-I.; Boyer, M. E.; Dewey, C. N.; Zhang, J.; Bresnick, E. H., Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. *J. Exp. Med.* **2013,** *In Press*.

7. (a) Ling, K. W.; Ottersbach, K.; van Hamburg, J. P.; Oziemlak, A.; Tsai, F. Y.; Orkin, S. H.; Ploemacher, R.; Hendriks, R. W.; Dzierzak, E., GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med* **2004,** *200* (7), 871-82; (b) Rodrigues, N. P.; Janzen, V.; Forkert, R.; Dombkowski, D. M.; Boyd, A. S.; Orkin, S. H.; Enver, T.; Vyas, P.; Scadden, D. T., Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood* **2005,** *106* (2), 477-84.

8. (a) Persons, D. A.; Allay, J. A.; Allay, E. R.; Ashmun, R. A.; Orlic, D.; Jane, S. M.; Cunningham, J. M.; Nienhuis, A. W., Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis. *Blood* **1999,** *93* (2), 488-99; (b) Tipping, A. J.; Pina, C.; Castor, A.; Hong, D.; Rodrigues, N. P.; Lazzari, L.; May, G. E.; Jacobsen, S. E.; Enver, T., High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. *Blood* **2009,** *113* (12), 2661-72.

9. (a) Hahn, C. H.; Chong, C.-E.; Carmichael, C. L.; Wilkins, E. J.; Brautigan, P. J.; Li, X.-C.; Babic, M.; Lin, M.; Carmagnac, A.; Lee, Y. K.; Kok, C. H.; Gagliardi, L.; Friend, K. L.; Ekert, P. G.; Butcher, C. M.; Brown, A. L.; Lewis, I. D.; To, L. B.; Timms, A. E.; Storek, J.; Moore, S.; Altree, M.; Escher, R.; Bardy, P. G.; Stutchers, G. K.; D'Andrea, R. J.; Horwitz, M. S.; Scott, H. S., Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat. Genet.* **2011,** *[Epub ahead of print]*; (b) Kazenwadel, J.; Secker, G. A.; Liu, Y. J.; Rosenfeld, J. A.; Wildin, R. S.; Cuellar-Rodriguez, J.; Hsu, A. P.; Dyack, S.; Fernandez, C. V.; Chong, C. E.; Babic, M.; Bardy, P. G.; Shimamura, A.; Zhang, M.

Y.; Walsh, T.; Holland, S. M.; Hickstein, D. D.; Horwitz, M. S.; Hahn, C. N.; Scott, H. S.; Harvey, N. L., Loss-of-function germline GATA2 mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature. *Blood* **2012,** *119* (5), 1283-91; (c) Hsu, A. P.; Sampaio, E. P.; Khan, J.; Calvo, K. R.; Lemieux, J. E.; Patel, S. Y.; Frucht, D. M.; Vinh, D. C.; Auth, R. D.; Freeman, A. F.; Olivier, K. N.; Uzel, G.; Zerbe, C. S.; Spalding, C.; Pittaluga, S.; Raffeld, M.; Kuhns, D. B.; Ding, L.; Paulson, M. L.; Marciano, B. E.; Gea-Banacloche, J. C.; Orange, J. S.; Cuellar-Rodriguez, J.; Hickstein, D. D.; Holland, S. M., Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood* **2011,** *118* (10), 2653-5; (d) Ostergaard, P.; Simpson, M. A.; Connell, F. C.; Steward, C. G.; Brice, G.; Woollard, W. J.; Dafou, D.; Kilo, T.; Smithson, S.; Lunt, P.; Murday, V. A.; Hodgson, S.; Keenan, R.; Pilz, D. T.; Martinez-Corral, I.; Makinen, T.; Mortimer, P. S.; Jeffery, S.; Trembath, R. C.; Mansour, S., Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat. Genet.* **2011,** *[Epub ahead of print]*; (e) Dickinson, R. E.; Griffin, H.; Bigley, V.; Reynard, L. N.; Hussain, R.; Haniffa, M.; Lakey, J. H.; Rahman, T.; Wang, X.-N.; McGovern, N.; Pagan, S.; Cookson, S.; McDonald, D.; Chua, I.; Wallis, J.; Cant, A.; Wright, M.; Keavney, B.; Chinnery, P. F.; Loughlin, J.; Hambleton, S.; Santibanez-Koref, M.; Collin, M., Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood* **2011,** *118* (10), 2656-2658.

10. (a) Vicente, C.; Vazquez, I.; Conchillo, A.; Garcia-Sanchez, M. A.; Marcotegui, N.; Fuster, O.; Gonzalez, M.; Calasanz, M. J.; Lahortiga, I.; Odero, M. D., Overexpression of GATA2 predicts an adverse prognosis for patients with acute myeloid leukemia and it is associated with distinct molecular abnormalities. *Leukemia* **2011,** *Sept. 9*, 1-5; (b) Luesink, M.; Hollink, I. H.; van der Velden, V. H.; Knops, R. H.; Boezeman, J. B.; de Haas, V.; Trka, J.; Baruchel, A.; Reinhardt, D.; van der Reijden, B. A.; van den Heuvel-Eibrink, M. M.; Zwaan, C. M.; Jansen, J. H., High GATA2 expression is a poor prognostic marker in pediatric acute myeloid leukemia. *Blood* **2012,** *120* (10), 2064-75.

11. Hsu, A. P.; Johnson, K. D.; Falcone, E. L.; Sanalkumar, R.; Sanchez, L.; Hickstein, D. D.; Cuellar-Rodriguez, J.; Lemieux, J. E.; Zerbe, C. S.; Bresnick, E. H.; Holland, S. M., GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* **2013**.

12. Grass, J. A.; Jing, H.; Kim, S. I.; Martowicz, M. L.; Pal, S.; Blobel, G. A.; Bresnick, E. H., Distinct functions of dispersed GATA factor complexes at an endogenous gene locus. *Molecular and cellular biology* **2006,** *26* (19), 7056-67.

13. Hsu, A. P.; Johnson, K. D.; Falcone, E. L.; Sanalkumar, R.; Sanchez, L.; Hickstein, D. D.; Cuellar-Rodriguez, J.; Lemieux, J. E.; Zerbe, C. S.; Bresnick, E. H.; Holland, S. M., GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* **2013,** *121* (19), 3830-7, S1-7.

14. (a) Porcher, C.; Liao, E. C.; Fujiwara, Y.; Zon, L. I.; Orkin, S. H., Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. *Development* **1999,** *126* (20), 4603-15; (b) Kassouf, M. T.; Chagraoui, H.; Vyas, P.; Porcher, C., Differential use of SCL/TAL-1 DNA-binding domain in developmental hematopoiesis. *Blood* **2008,** *112* (4), 1056-67; (c) Liu, N.; Barbosa, A. C.; Chapman, S. L.; Bezprozvannaya, S.; Qi, X.; Richardson, J. A.; Yanagisawa, H.; Olson, E. N., DNA binding-dependent and -independent functions of the Hand2 transcription factor during mouse embryogenesis. *Development* **2009,** *136* (6), 933-42; (d) Reichardt, H. M.; Kaestner, K. H.; Tuckermann, J.; Kretz, O.; Wessely, O.; Bock, R.; Gass, P.; Schmid, W.; Herrlich, P.; Angel, P.; Schutz, G., DNA binding of the glucocorticoid receptor is not essential for survival. *Cell* **1998,** *93* (4), 531-41.

15. (a) Pal, S.; Cantor, A. B.; Johnson, K. D.; Moran, T. B.; Boyer, M. E.; Orkin, S. H.; Bresnick, E. H., Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A* **2004,** *101* (4), 980-5; (b) Lamonica, J. M.; Vakoc, C. R.; Blobel, G. A., Acetylation of GATA-1 is required for chromatin occupancy. *Blood* **2006,** *108* (12), 3736-3738; (c) Letting, D. L.; Chen, Y. Y.; Rakowski, C.; Reedy, S.; Blobel, G. A., Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc Natl Acad Sci U S A* **2004,** *101* (2), 476-81.

16. Sun, M. K.; Alkon, D. L., Induced depressive behavior impairs learning and memory in rats. *Neuroscience* **2004,** *129* (1), 129-39.

17. Linnemann, A. K.; O'Geen, H.; Keles, S.; Farnham, P. J.; Bresnick, E. H., Genetic framework for GATA factor function in vascular biology. *Proceedings of the National Academy of Sciences of the United States of America* **2011,** *108* (33), 13641-6.

18. Beck, D.; Thoms, J. A.; Perera, D.; Schutte, J.; Unnikrishnan, A.; Knezevic, K.; Kinston, S. J.; Wilson, N. K.; O'Brien, T. A.; Gottgens, B.; Wong, J. W.; Pimanda, J. E., Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood* **2013,** *122* (14), e12-e22.

19. Taus, T.; Kocher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K., Universal and confident phosphorylation site localization using phosphoRS. *Journal of proteome research* **2011,** *10* (12), 5354-62.

20. Silva, J. C.; Gorenstein, M. V.; Li, G. Z.; Vissers, J. P.; Geromanos, S. J., Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Molecular & cellular proteomics : MCP* **2006,** *5* (1), 144-56.

21. Weiss, M. J.; Yu, C.; Orkin, S. H., Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol. Cell. Biol.* **1997,** *17* (3), 1642-1651.

22. (a) Fujiwara, T.; O'Geen, H.; Keles, S.; Blahnik, K.; Linnemann, A. K.; Kang, Y. A.; Choi, K.; Farnham, P. J.; Bresnick, E. H., Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Molecular cell* **2009,** *36* (4), 667-81; (b) Grass, J. A.; Boyer, M. E.; Pal, S.; Wu, J.; Weiss, M. J.; Bresnick, E. H., GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** *100* (15), 8811-6.

23. (a) Pal, S.; Wu, J.; Murray, J. K.; Gellman, S. H.; Wozniak, M. A.; Keely, P. J.; Boyer, M. E.; Gomez, T. M.; Hasso, S. M.; Fallon, J. F.; Bresnick, E. H., An antiangiogenic neurokinin-B/thromboxane A2 regulatory axis. *J Cell Biol* **2006,** *174* (7), 1047-58; (b) Bastaki, M.; Nelli, E. E.; Dell'Era, P.; Rusnati, M.; Molinari-Tosatti, M. P.; Parolini, S.; Auerbach, R.; Ruco, L. P.; Possati, L.; Presta, M., Basic fibroblast growth factor-induced angiogenic phenotype in mouse endothelium. A study of aortic and microvascular endothelial cell lines. *Arteriosclerosis, thrombosis, and vascular biology* **1997,** *17* (3), 454-64.

24. Gao, X.; Johnson, K. D.; Chang, Y. I.; Boyer, M. E.; Dewey, C. N.; Zhang, J.; Bresnick, E. H., Gata2 cis-element is required for hematopoietic stem cell generation in the mammalian embryo. *The Journal of experimental medicine* **2013,** *210* (13), 2833-42.

25. (a) Moignard, V.; Macaulay, I. C.; Swiers, G.; Buettner, F.; Schutte, J.; Calero-Nieto, F. J.; Kinston, S.; Joshi, A.; Hannah, R.; Theis, F. J.; Jacobsen, S. E.; de Bruijn, M. F.; Gottgens, B., Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nature cell biology* **2013,** *15* (4), 363-72; (b) Lancrin, C.; Mazan, M.; Stefanska, M.; Patel, R.; Lichtinger, M.; Costa, G.; Vargel, O.; Wilson, N. K.; Moroy, T.; Bonifer, C.; Gottgens, B.; Kouskoff, V.; Lacaud, G., GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood* **2012,** *120* (2), 314-22.

26. Beck, D.; Thoms, J. A.; Perera, D.; Schutte, J.; Unnikrishnan, A.; Knezevic, K.; Kinston, S. J.; Wilson, N. K.; O'Brien, T. A.; Gottgens, B.; Wong, J. W.; Pimanda, J. E., Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood* **2013,** *122* (14), e12-22.

27. (a) Kumar, M. S.; Hancock, D. C.; Molina-Arcas, M.; Steckel, M.; East, P.; Diefenbacher, M.; Armenteros-Monterroso, E.; Lassailly, F.; Matthews, N.; Nye, E.; Stamp, G.; Behrens, A.; Downward, J., The GATA2 transcriptional network is requisite for RAS oncogene-driven non-small cell lung cancer. *Cell* **2012,** *149* (3), 642-55; (b) Steckel, M.; Molina-Arcas, M.; Weigelt, B.; Marani, M.; Warne, P. H.; Kuznetsov, H.; Kelly, G.; Saunders, B.; Howell, M.; Downward, J.; Hancock, D. C., Determination of synthetic lethal interactions in KRAS oncogene-dependent cancer cells reveals novel therapeutic targeting strategies. *Cell research* **2012,** *22* (8), 1227-45.

28. (a) Koga, S.; Yamaguchi, N.; Abe, T.; Minegishi, M.; Tsuchiya, S.; Yamamoto, M.; Minegishi, N., Cell-cycle-dependent oscillation of GATA2 expression in hematopoietic cells. *Blood* **2007,** *109* (10), 4200-8; (b) Menghini, R.; Marchetti, V.; Cardellini, M.; Hribal, M. L.; Mauriello, A.; Lauro, D.; Sbraccia, P.; Lauro, R.; Federici, M., Phosphorylation of GATA2 by Akt increases adipose tissue differentiation and reduces adipose tissue-related inflammation: a novel pathway linking obesity to atherosclerosis. *Circulation* **2005,** *111* (15), 1946-53; (c) Towatari, M.; May, G. E.; Marais, R.; Perkins, G. R.; Marshall, C. J.; Cowley, S.; Enver, T., Regulation of GATA-2 phosphorylation by mitogen-activated protein kinase and interleukin-3. *The Journal of biological chemistry* **1995,** *270* (8), 4101-7.

29. (a) Maneechotesuwan, K.; Xin, Y.; Ito, K.; Jazrawi, E.; Lee, K. Y.; Usmani, O. S.; Barnes, P. J.; Adcock, I. M., Regulation of Th2 cytokine genes by p38 MAPK-mediated phosphorylation of GATA-3. *Journal of immunology* **2007,** *178* (4), 2491-8; (b) Frelin, C.; Herrington, R.; Janmohamed, S.; Barbara, M.; Tran, G.; Paige, C. J.; Benveniste, P.; Zuniga-Pflucker, J. C.; Souabni, A.; Busslinger, M.; Iscove, N. N., GATA-3 regulates the self-renewal of long-term hematopoietic stem cells. *Nature immunology* **2013,** *14* (10), 1037-44.

30. (a) Ito, K.; Hirao, A.; Arai, F.; Takubo, K.; Matsuoka, S.; Miyamoto, K.; Ohmura, M.; Naka, K.; Hosokawa, K.; Ikeda, Y.; Suda, T., Reactive oxygen species act through p38 MAPK to limit the lifespan of hematopoietic stem cells. *Nature medicine* **2006,** *12* (4), 446-51; (b) Qian, H.; Buza-Vidas, N.; Hyland, C. D.; Jensen, C. T.; Antonchuk, J.; Mansson, R.; Thoren, L. A.; Ekblom, M.; Alexander, W. S.; Jacobsen, S. E., Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells. *Cell stem cell* **2007,** *1* (6), 671-84.

31. Navas, T. A.; Mohindru, M.; Estes, M.; Ma, J. Y.; Sokol, L.; Pahanish, P.; Parmar, S.; Haghnazari, E.; Zhou, L.; Collins, R.; Kerr, I.; Nguyen, A. N.; Xu, Y.; Platanias, L. C.; List, A. A.; Higgins, L. S.; Verma, A., Inhibition of overactivated p38 MAPK can restore hematopoiesis in myelodysplastic syndrome progenitors. *Blood* **2006,** *108* (13), 4170-7.

32. (a) Yamazaki, H.; Suzuki, M.; Otsuki, A.; Shimizu, R.; Bresnick, E. H.; Engel, J. D.; Yamamoto, M., A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. *Cancer cell* **2014,** *25* (4), 415-27; (b) Groschel, S.; Sanders, M. A.; Hoogenboezem, R.; de Wit, E.; Bouwman, B. A.; Erpelinck, C.; van der Velden, V. H.; Havermans, M.; Avellino, R.; van Lom, K.; Rombouts, E. J.; van Duin, M.; Dohner, K.; Beverloo, H. B.; Bradner, J. E.; Dohner, H.; Lowenberg, B.; Valk, P. J.; Bindels, E. M.; de Laat, W.; Delwel, R., A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **2014,** *157* (2), 369-81.

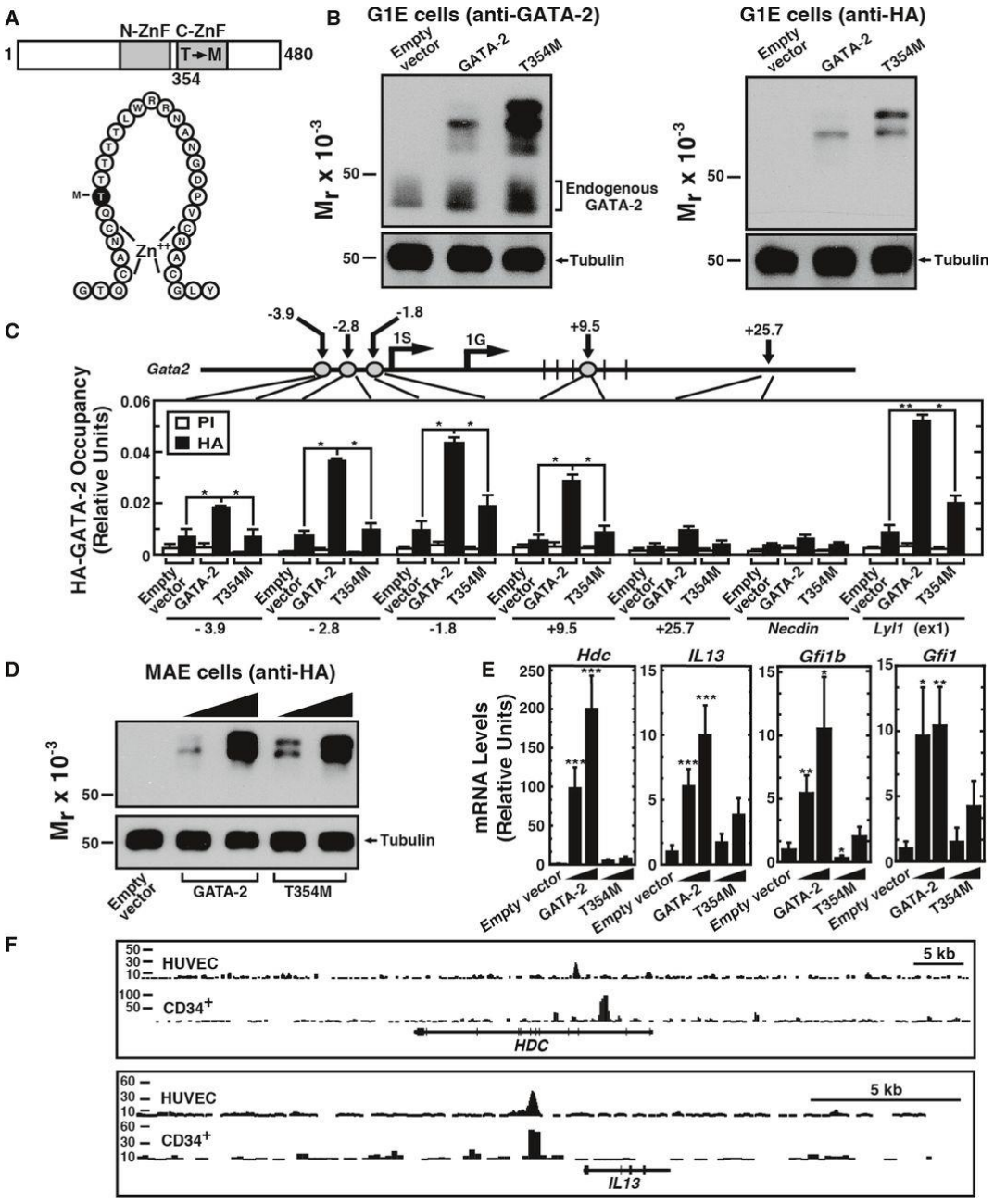


Figure 4.1 T354M mutation attenuates GATA-2 chromatin occupancy and endogenous target gene activation.

A. T354M mutant.

B. Western blot analysis with anti-GATA-2 or anti-HA antibodies of HA-GATA-2 and HA-GATA-2(T354M) transiently expressed in G1E cells.

C. Quantitative ChIP analysis of HA-GATA-2 and HA-GATA-2 (T354M) occupancy in G1E cells transiently expressing HA-GATA-2 or HA-GATA-2 (T354M) (n = 4, mean ± SE).

D. Western blot analysis of HA-GATA-2 and HA-GATA-2 (T354M) transiently expressed in MAE cells.

E. Quantitation of GATA-2 target gene expression. Real-time RT-PCR analysis in MAE cells expressing HA-GATA-2 or HA-GATA-2 (T354M) (n = 9, mean ± SE) *P < 0.05,

**P < 0.01, ***P < 0.001, relative to the empty vector.

F. ChIP-seq analysis of GATA-2 occupancy at HDC and IL13 loci in human CD34-positive hematopoietic cells and HUVECs.[17, 26]

Data information: Significance of the differences was estimated using Paired Student's t-test. *P < 0.05, **P < 0.01, ***P < 0.001.

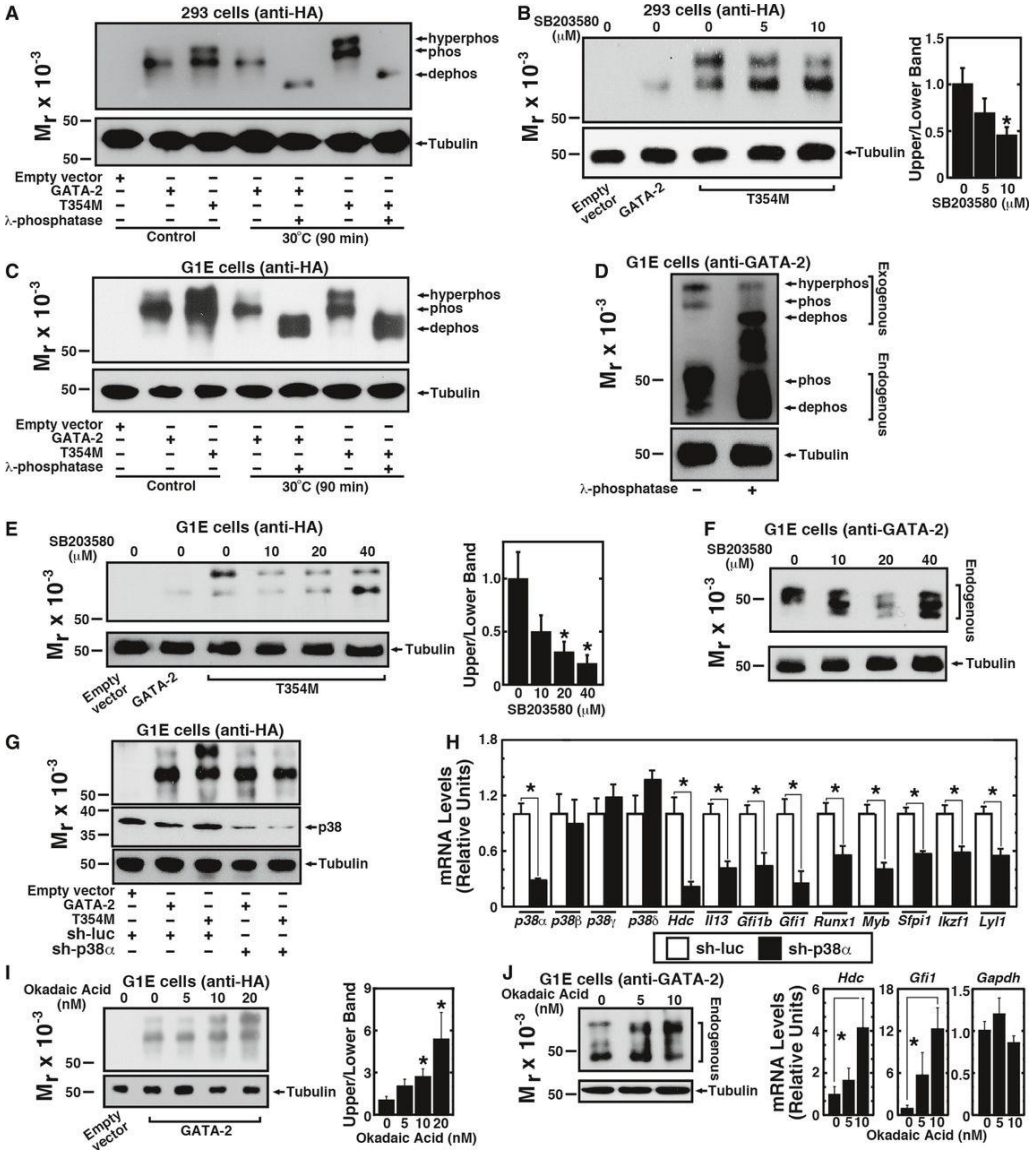


Figure 4.2 p38-mediated GATA-2 and GATA-2 (T354M) hyperphosphorylation and target gene regulation.

A. Total protein from 293 cells expressing HA-GATA-2, HA-GATA-2 (T354M), or empty vector was incubated with or without λ-phosphatase and analyzed by Western blotting with anti-HA antibody.

B Left, Western blot analysis of wild-type GATA-2 and HA-GATA-2 (T354M) from 293 cells treated with vehicle or p38 inhibitor SB203580. Right, densitometric analysis.

The ratio of intensities of the T354M upper to lower bands from control untreated cells was designated as 1 (n = 3, mean ± SE).

C. Total protein from G1E cells expressing HA-GATA-2, HA-GATA-2 (T354M), or empty vector was incubated with or without λ-phosphatase. Proteins were resolved by SDS–PAGE and analyzed by Western blotting with anti-HA antibody.

D. Total protein from G1E cells expressing HA-GATA-2 (T354M) was incubated with or without λ-phosphatase and analyzed by Western blotting with anti-GATA-2 antibody.

E Left, Western blot analysis of wild-type GATA-2 and HA-GATA-2 (T354M) from G1E cells treated with vehicle or p38 inhibitor SB203580. Right, densitometric analysis.

The ratio of the intensities of the T354M upper to lower bands from control untreated cells was designated as 1 (n = 3, mean ± SE).

F. Western blot analysis of GATA-2 in G1E cells treated with SB203580.

G. Western blot analysis of HA-GATA-2 (T354M) and p38 in G1E cells infected with sh-luc virus or sh-p38a virus.

H. Real-time RT-PCR analysis of p38 mRNA and transcripts of GATA-2 target genes in G1E cells infected with sh-luc virus or sh-p38a virus (n = 3, mean ± SE).

I Left, Western blot analysis of HA-GATA-2 in G1E cells treated with okadaic acid. Right, densitometric analysis. The upper to lower band ratio from control was designated as 1 (n = 4, mean ± SE). Cells were treated with okadaic acid for 24 h. The viability was more than 75%.

J Left, Western blot analysis of endogenous GATA-2 in G1E cells treated with okadaic acid (8% gel was used instead of 10% used in D and F). Right: Real-time RT-PCR analysis in okadaic acid-treated G1E cells (n = 4, mean ± SE).

Data information: Significance of the differences was estimated using Paired Student's t-test. *P < 0.05

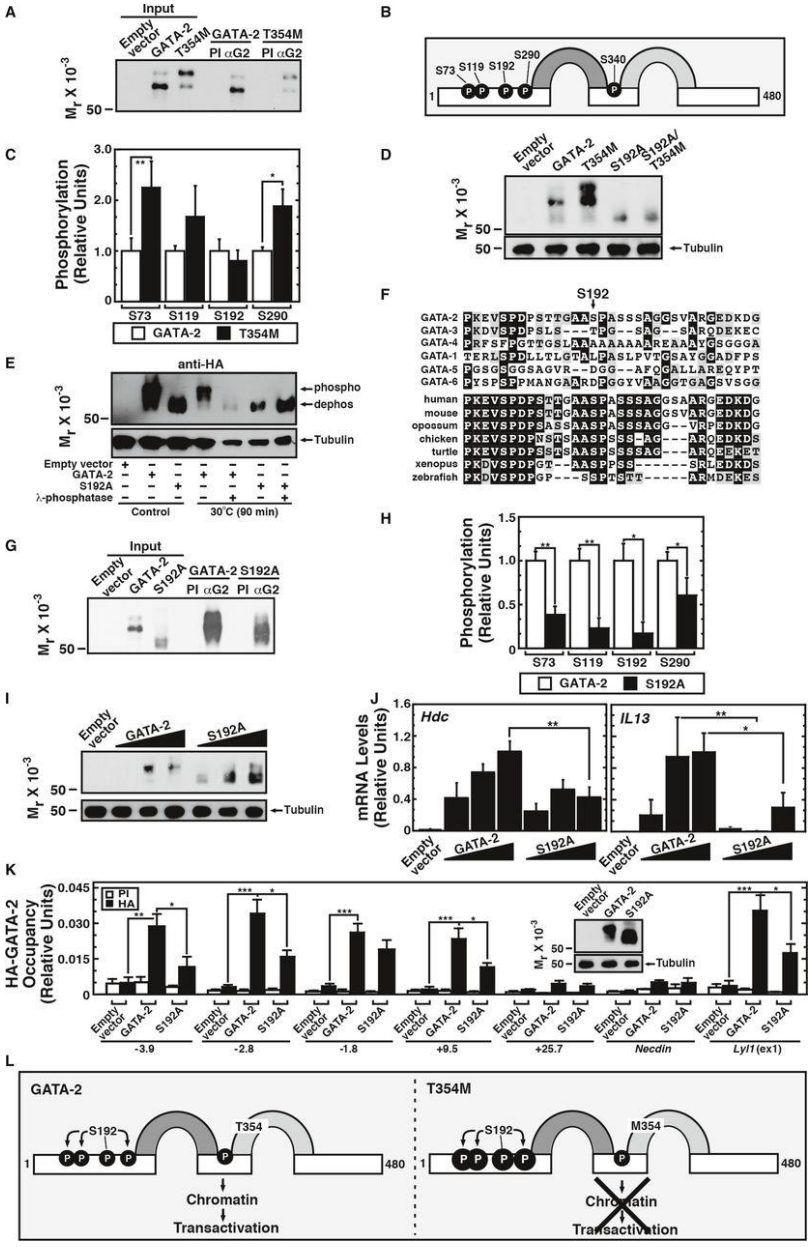


Figure 4.3 Ser192-mediated GATA-2 phosphorylation.

A. Immunoprecipitation of HA-GATA-2 and HA-GATA-2 (T354M) expressed in 293 cells. Proteins were detected by Western blotting with anti-HA antibody.

B. Phosphorylated residues detected by LC-ESI MS/MS.

C. Relative phosphorylation of S73, S119, S192, and S290 in GATA-2 and T354M (n = 3, mean ± SE).

D. Western blot analysis with anti-HA antibody of proteins expressed in G1E cells.

E. 293 cell proteins incubated with or without λ-phosphatase and analyzed by Western blotting with anti-HA antibody.

F. Murine GATA factor amino acid sequences (top) and GATA-2 sequences from multiple species (bottom).

G. Immunoprecipitation of HA-GATA-2 and HA-GATA-2 (S192A) expressed in 293 cells. Proteins were detected by Western blotting with anti-HA antibody.

H. Relative phosphorylation of S73, S119, S/A192, and S290 in GATA-2 and S192 mutant determined by mass spectrometry analysis (n = 3, mean ± SE).

I. Western blot analysis with anti-HA antibody of wild-type and mutant proteins expressed in MAE cells.

J. Real-time RT-PCR analysis of Hdc and IL13 mRNA levels in MAE cells expressing wild-type or mutant proteins (n = 3, mean ± SE).

K. Quantitative ChIP analysis of HA-GATA-2 and HA-GATA-2 (S192A) occupancy with anti-HA antibody in G1E cells expressing HA-GATA-2 or HA-GATA-2(S192A) (n = 5, Mean ± SE). The Western blot (anti-HA antibody) in the inset illustrates GATA-2 and S192A expression in representative samples used for ChIP.

L S192-mediated multi-site GATA-2 hyperphosphorylation.

Data information: Significance of the differences was estimated using Paired Student's t-test. *P < 0.05, **P < 0.01, ***P < 0.001.

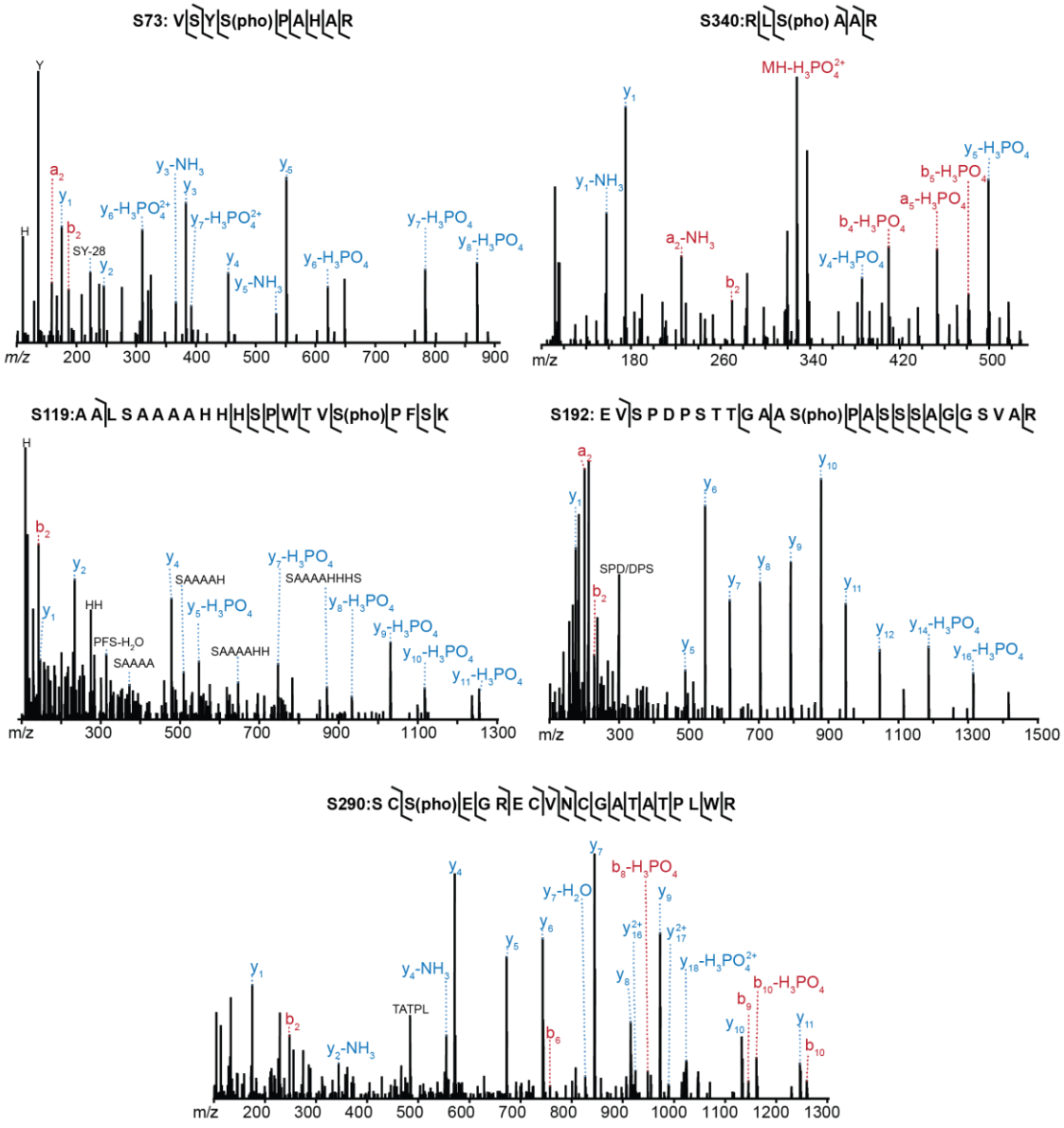


Figure 4.4 MS/MS spectra for phosphorylation sites identified in GATA-2 and GATA-2 (T354M). pho: phosphorylation.

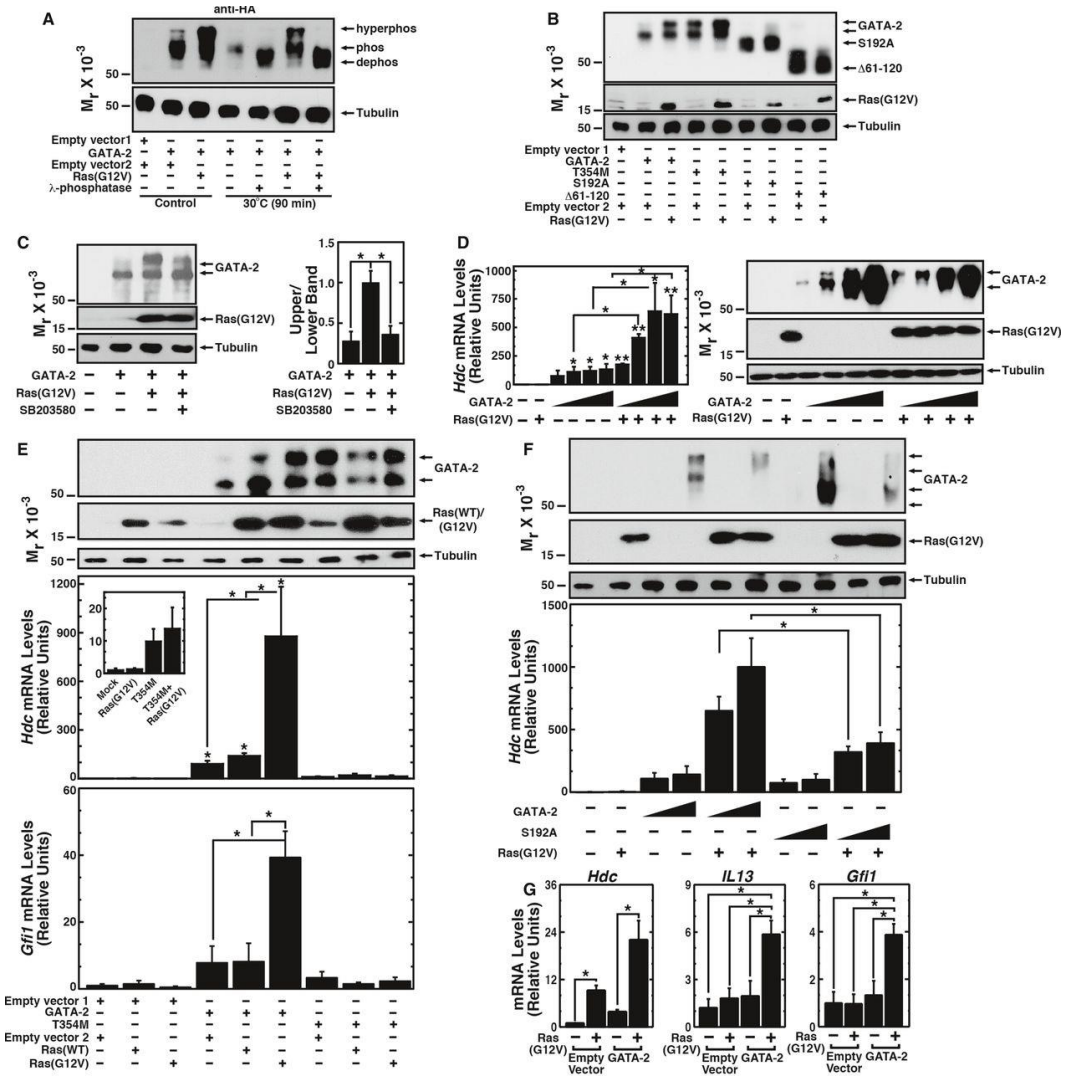


Figure 4.5 S192 requirement for oncogenic Ras-induced GATA-2 activity.

A. Total protein from 293 cells expressing HA-GATA-2, with or without H-Ras (G12V), was incubated with or without λ-phosphatase and analyzed by Western blotting with anti-HA antibody.

B. Western blot analysis with anti-HA antibody of proteins expressed in G1E cells with or without H-Ras (G12V).

C Left, Western blot analysis of G1E cells expressing HA-GATA-2 and H-Ras (G12V) with or without SB203580. Right, densitometric analysis. The ratio of intensities of the

upper to lower bands from Ras (G12V)-expressing cells was designated as 1 (n = 3, mean ± SE).

D. Influence of varying HA-GATA-2 expression vector concentration with or without H-Ras(G12V) on Hdc induction in MAE cells (n = 3, mean ± SE). Left, real-time RT-PCR analysis of Hdc mRNA level. Right, Western blot analysis with anti-HA antibody.

E Top, Western blot analysis of HA-GATA-2 and HA-GATA-2(T354M) proteins expressed in MAE cells with or without H-Ras (G12V). Bottom, real-time RT-PCR analysis of *Hdc* mRNA and *Gfi1* mRNA levels in MAE cells expressing HA-GATA-2 and HA-GATA-2 (T354M), with or without H-Ras (G12V) (n = 4, mean ± SE). *P < 0.05. Inset, expanded view of low-level *Hdc* expression.

F Top, Western blot analysis of HA-GATA-2 and HA-GATA-2(S192A) proteins expressed in MAE cells with or without H-Ras (G12V). Bottom, real-time RT-PCR analysis of *Hdc* mRNA levels in MAE cells expressing HA-GATA-2 and HA-GATA-2(S192A), with or without H-Ras (G12V) (n = 3, mean ± SE).

G. Real-time RT-PCR analysis of *Hdc*, *IL13*, and *Gfi1* mRNA levels in G1E cells expressing HA-GATA-2, with or without H-Ras (G12V) (n = 4, mean ± SE).

Data information: Significance of the differences was estimated using Paired Student's *t*-test. *P < 0.05, **P < 0.01.

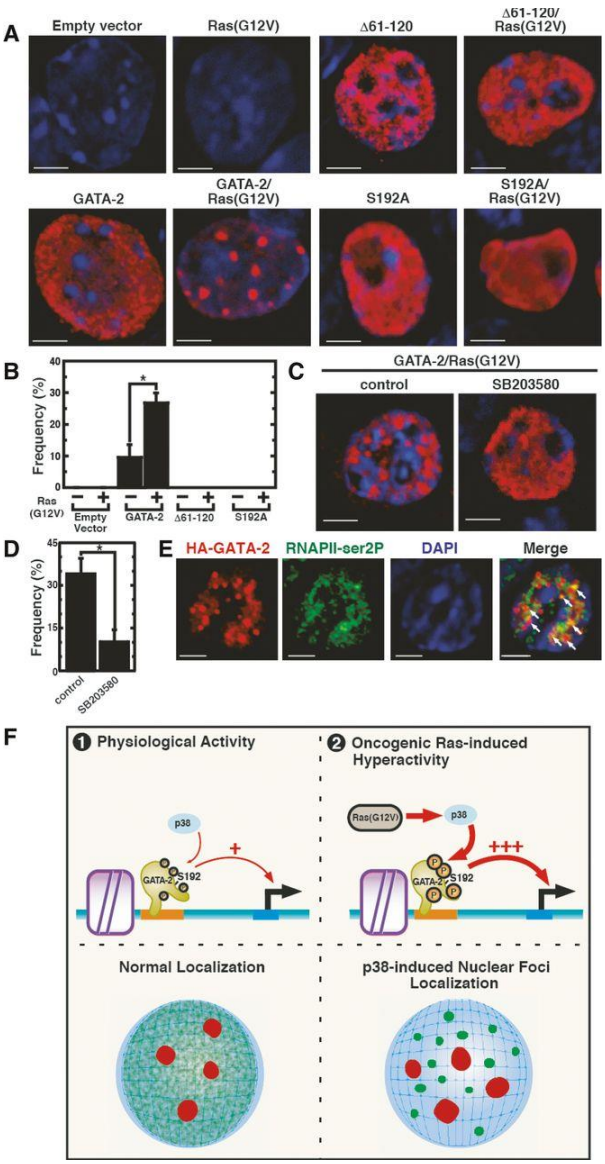


Figure 4.6 Ras-p38-regulated GATA-2 subnuclear localization.

A. Immunofluorescence analysis with anti-HA antibody in G1E cells expressing HA-GATA-2 and mutant proteins with or without H-Ras (G12V). Scale bars: 5 lm.

B. The percentage of cells exhibiting foci, based on analysis of > 150 cells (n = 3, mean ± SE).

C. Immunofluorescence analysis with anti-HA antibody in G1E cells expressing HA-GATA-2 and H-Ras (G12V) with or without SB203580. Scale bars: 5 lm.

D. The percentage of cells exhibiting foci, based on examining > 150 cells (n = 3, mean ± SE).

E. Co-immunostaining with anti-HA and anti-serine 2-phosphorylated Pol II antibody in G1E cells expressing HA-GATA-2 and H-Ras (G12V). Arrows, colocalization signals. Bars: 5 lm.

F. Ras-p38 axis enhances GATA-2 phosphorylation, subnuclear localization, and target gene activation.

Data information: Significance of the differences was estimated using Paired Student's *t*-test. $P < 0.05$.

Chapter 5

Comparative Secretome Analysis of Vascular Smooth Muscle Cells in Response to Smad3-Dependent TGF-β Signaling

Adapted from: "Comparative Secretome Analysis of Vascular Smooth Muscle Cells in Response to Smad3-Dependent TGF-β Signaling". **Yang C**[*], Ma D[*], Shi X and Li L. In preparation.

ABSTRACT

Vascular smooth muscle cells (VSMCs) play a crucial role in cardiovascular disorders and have been used as a model system to study the pathogenesis of atherosclerosis. Transforming growth factor β (TGF-β) and its modulator, SMAD family member 3 (Smad3) play important roles in VSMC differentiation and function and are found to be critically involved in atherosclerotic process. To characterize the proteins secreted from cultured VSMCs and assess temporal changes in the secretome in response to TGF-β/Smad3 signaling, we performed LC ESI-MS/MS analysis followed by label-free quantification to identify proteins secreted into VSMC-conditioned medium before and after cells were treated with TGF-β. The use of one-dimensional LC-MS/MS resulted in identifications of 665 proteins from conditioned media. The quantitative analysis by spectral counting revealed 38 secreted proteins that were significantly up or down-regulated in response to TGF-β. The characterization of the secretome of VSMCs from this study provides new insights in vascular biology. Moreover, the TGF-β induced factors identified from the comparative analysis may provide novel targets for the investigations of the role of TGF-β in atherosclerotic process.

INTRODUCTION

According to the statistics from the World Health Organization and past research, atherosclerosis and atherosclerosis-associated cardiovascular diseases including myocardial infarction and stroke are the leading causes of death in developed countries including the United States of America where it is estimated that vascular diseases are to be the leading cause of global morbidity and mortality by 2020-2030.[1] High levels of plasma lipids, in particular low-density lipoprotein (LDL) cholesterol in blood is widely considered to increase the risk to develop cardiovascular diseases.[2] Although the emergence of atherogenesis is associated with accumulation of lipids within the arterial walls, the development of atherosclerosis is best described as a consequence of chronic inflammation consisting of a series of highly specific cellular and molecular responses with the involvement of both immune cells and vascular smooth muscle cells.[3]

Currently, a common clinical procedure to treat coronary and peripheral atherosclerosis is angioplasty due to its high success rate and minimally invasive nature. However, the efficacy of surgical procedure is often undermined by restenosis which occurs within 3–6 months post-angioplasty mainly due to intimal hyperplasia which mimics some aspects of the atherosclerotic process.[4] The proliferation of vascular smooth muscle cells (VSMCs) plays a key role in the development of intimal hyperplasia.[5] Intimal hyperplasia is a complex process associated with abnormal migration and proliferation of vascular smooth muscle cells (VSMCs). In response to injury, VSMCs change their quiescent-contractile

physiological phenotype and acquire an activated state endowed with proliferative and migratory properties, once subjected to external stimuli such as cytokines and growth factors.[6] As a consequence of their activation, VSMCs migrate to the subintimal space, proliferate, and secrete abundant amounts of extracellular matrix (ECM), which forms the bulk of the intimal hyperplastic lesion contributing to restenosis. Transforming growth factor-β (TGF-β) is believed to be a critical factor contributing to intimal hyperplasia as it was shown to play an important role in VSMC differentiation and function.[7] Recently, several studies reported that TGF-β enhances VSMC proliferation through Smad3. The up-regulation and activation of Smad3 at the time of arterial injury is mainly responsible for the stimulatory effect of TGF-β on VSMC proliferation and development of intimal plaque.[8] Suwanabol *et al*. reported that Smad3 acts as an intermediate between TGF-β and signaling pathways including ERK/MAPK[8a] and PI3K/Akt[8b] through which TGF-β stimulates VSMC proliferation. Although these findings provided evidence of the mechanism by which TGF-β enhances intimal hyperplasia, little is known about the Smad3-dependent TGF-β signaling in VSMCs. To this end, a global approach to proteomic profiling of VSMCs treated with TGF-β may serve to identify TGF-β/Smad3 induced factors that reveal previously unknown pathways and mechanisms that mediate intimal hyperplasia. Such knowledge provides potential novel targets that may be used to prevent restenosis.

Both profiling of intracellular proteome and extracellular secretome of primary culture of human arterial smooth muscle cells (SMCs) from patients

undergoing coronary artery bypass surgery has been previously reported.[9] The secretome typically includes the proteins of ECM as well as additional proteins shed from the cell surface.[10] The secreted proteins play important roles in both physiological and pathological processes such as cell-cell signaling, communication, and migration,[11] and reflect the state of the cells at various stages of disease progression. In secretome analysis, the proteins secreted by cells into conditioned media *in vitro* are studied to better understand the mechanisms *in vivo*. Although *in vitro* conditions may be different from the *in vivo* environment, the analysis of secretome in cultured cells provides a relatively easier and quicker manner to investigate possible biological pathways involved in these complicated cellular processes.

Previous studies demonstrated that VSMCs from injured rat aortas display elevated matrix production associated with activity of TGF-β and proposed that one of the mechanisms through which TGF-β enhances intimal hyperplasia is the production of ECM proteins including several collagens.[8c, 12] To elucidate the complex extracellular signaling in VSMCs in response to TGF-β that may be associated with pathogenesis of intimal hyperplasia, a comprehensive proteomic profiling of secreted proteins from VSMCs will serve an essential starting point for further investigation. Previous mass spectrometry (MS)-based secretome profiling in SMCs used 2-D gel electrophoresis coupled with MALDI-TOF for protein identification which limited the total protein identifications as well as the detection of lower abundance proteins.[9] Here, we employed shotgun proteomics with LC-MS/MS

for secretome profiling in VSMCs and our result yielded a substantially larger list of putative secreted proteins. Further quantitative analysis by spectral counting revealed 38 secreted proteins that were significantly up or down-regulated in response to TGF-β. Our analysis provides a basis for vascular biology investigations on VSMC protein changes that may lead to a better understanding of the mechanisms and factors that regulate the differentiation of VSMCs in response to TGF-β.

EXPERIMENTAL PROCEDURES

Smooth Muscle Cell Culture

Rat aortic vascular SMCs were isolated from the thoracoabdominal aorta of male Sprague-Dawley rats based on a protocol described by Clowes *et al.*[13] and maintained in DMEM containing 10% FBS at 37°C with 5% CO2. Adenoviral (Ad) vectors expressing Smad3 (AdSmad3) and green fluorescent protein (GFP) (AdGFP) were constructed as previously described.[14] AdGFP was used as a control. 5×10^5 VSMCs were infected with adenovirus in dulbecco's modified eagle's medium (DMEM) containing 2% FBS for 4 h at 37° C followed by recovery in 10% FBS overnight. Cells were then cultured in serum free DMEM for 12 h. After wash once with serum free media, 5 ng/ml TGF-β /or control (4 μl 0.2 N HCl in 4 ml media) was added to culture dishes (counted as time 0). After 6 hours, cell culture media was collected and concentrated with 10KDa MW cut off column (Millipore, Billerica, MA). The concentration of secreted proteins was determined by BCA assay.

Proteolysis

All protein samples were denatured with 8 M urea in 25 mM ammonium bicarbonate buffer, and reduced by incubating with 50 mM DTT at 37 °C for 1 h. The reduced proteins were alkylated for 1 h in darkness with 100 mM iodoacetamide. The alkylation reaction was quenched by adding DTT to a final concentration of 50 mM. The samples were diluted to a final concentration of 1 M urea. Trypsin was added to the sample at a 30:1 protein to trypsin mass ratio. The sample was incubated at 37 °C overnight. Digests were desalted by ZipTip® Pipette Tips (Millipore, Billerica, MA).

LC–ESI Ion Trap Mass Spectrometry and MS/MS Analysis

A Q-Exactive (Thermo Fisher Scientific, Bermen, Germany) mass spectrometer equipped with Waters nano Acquity UPLC (Waters Corp., Milford, MA) was utilized to analyze samples. For the chromatographic separation, solvent A consisted of 0.1% formic acid in water and solvent B consisted of 0.1% formic acid in ACN. 5µl of each sample was injected onto an Waters Symmetry C18 5µm 180µm x 20mm precolumn at a flow rate of 5 µl/min for 5 min at 95% A 5% B, followed by peptide separation performed on Waters BEH130 1.7 µm C18 100 µm x 100 mm analytical column using gradient from 0 to 45% solvent B at 300 nl/min over 120 min. The mass spectrometer was operated in a data-dependent mode. The twenty most intense ions ($z \geq 2$) in full scan were selected for MS/MS analysis. Full scans were acquired in 350-1600 *m/z* range at a resolution of 70,000 (at *m/z*=400) in the orbitrap with an automatic gain control (AGC) target value of 1×10^6 charges and maximum fill time of 250 ms. MS/MS spectra were collected at a resolution of 17,500 (at

m/z=400) with an AGC value of 1×10^5 charges and maximum fill time of 50 ms. The underfill ratio is 0.1%. A precursor isolation width of 2 *m/z* and 10s of dynamic exclusion were used.

Database Search

The generated MS data files were processed by Proteome Discoverer (v.1.4.0.288, Thermo Fisher Scientific). Peak lists were generated and searched with build-in Sequest HT against the uniprot rat database. Precursor mass and fragment mass were searched with mass tolerance of 10 ppm and 0.2 Da. Carbamidomethylcysteine was set as fixed modification and oxidized methionine was searched as variable modifications. Up to three missed cleavage sites for trypsin were allowed. The false discovery rate (FDR) was set to 1% for peptide and protein identifications. To perform quantification analysis, a spectral counting approach was used. Spectral counts for each identified protein were determined by the number of peptide spectrum matches (PSMs) in Sequest HT search. Number of PSMs in technical replicates was summed for each protein and subject to normalized spectral abundance factor (NSAF) analysis. Details for this method can be found elsewhere.[15] In brief, the NSAF for each protein was calculated after the spectral counts of the protein normalized to its length and divided by the sum of SpC/L for all the proteins in the experiment.

RESULTS AND DISCUSSION

VSMCs with Smad3 Overexpression

The array of effects of TGF-β on cultured VSMCs may be related to differential signaling through numerous downstream pathways. Abundant evidence from previous studies has shown that TGF-β has both stimulatory effects on VSMC fibronectin synthesis and inhibitory effects on VSMC proliferation and migration.[7b, 12a] Several recent studies reported that the up-regulation and activation of Smad3 at the time of arterial injury is mainly responsible for the stimulatory effect of TGF-β on VSMC proliferation and development of intimal plaque.[8] Since TGF-β enhances SMC proliferation only in the presence of elevated levels of Smad3, we infected cultured VSMCs with adenovirus-expressing Smad3 (AdSmad3) or control (AdGFP) followed by stimulation with or without TGF-β for 6 h. The overexpression of Smad3 enhanced the effect of Smad3-dependent TGF-β signaling, thus simplify the complex and contradictory effects of TGF-β VSMC function.

Given the biological uniformity of the cell cultures as opposed to individual organisms, secretome samples from three replicates of equal amounts of VSMC/AdGFP or VSMC/AdSmad3 cultured in supplement with or without TGF-β were pooled. To our knowledge, there was no evidence showing cell culture-to-cell culture variability from previous studies where immunological and biochemical approaches were utilized to characterize and quantify proteins in VSMCs.[8] Although the pooling strategy was not ideal for statistical purpose, it significantly reduces the time and complexity for MS and data analysis without compromising the accuracy of quantitative analysis.

Protein Identification and Data Analysis.

In total, 1534 non-redundant peptides and 665 non-redundant proteins were identified from control and TGF-β treated samples. 401 of 665 non-redundant proteins were identified from the conditioned media of control VSMC/AdGFP samples, while 576 non-redundant proteins were identified from that of TGF-β treated VSMC/AdSmad3 samples. Figure 5.1 shows the Venn Diagrams of the numbers of peptides and proteins identified from the conditioned media from VSMC/AdGFP cultures or VSMC/AdSmad3 cultures. A total of 312 non-redundant proteins and 609 non-redundant peptides were identified in both conditioned media where each sample was analyzed twice. Figure 5.2 shows a representative tandem MS/MS spectrum of a tryptic peptide (DLLTAYYDVDYEK) from protein disulfide-isomerase A3.

A spectral counting approach was employed for relative quantification where peptide spectral counts were added up for each protein and subjected to NSAF analysis. Details for this method can be found elsewhere.[15-16] Proteins with at least two unique peptides identified were considered for quantification analysis. A total of 255 proteins identified were selected for quantification by spectral counting, and 94 proteins were found with $\log_2^{(ratio)} \geq 2$ or ≤ -2. Given that protein identification in secretome analysis is often interfered by the presence of intracellular proteins in culture media mainly due to cell death. SignalP/SecretomeP was, therefore, used to distinguish the secreted proteins from intracellular proteins. Among 94 proteins with $\log_2^{(ratio)} >= 2$ or <=-2, 38 proteins are known to be secreted or predicted to be secreted proteins (Table 5.1).

Functional Annotation of TGF-β Induced Secreted Proteins

All TGF-β induced proteins were searched against the Swiss-Prot/TrEMBL annotated database using the online ExPASy interface (http://www.expasy.ch/) to confirm that they are extracellular proteins. Functional annotation analysis using public bioinformatics resources revealed that most TGF-β induced proteins identified in present work play important roles in cell proliferation, migration and differentiation. Based on the evidence from previous studies, some of the proteins may prove to be critical factors that are involved in the development of atherosclerotic process in response to TGF-β.

Insulin-like growth factor binding protein 7 (IGFBP-7) is a 30 kDa secretory glycoprotein which belongs to insulin-like growth factor binding protein (IGFBP) superfamily[17]. Previous *in vitro* analysis indicated that the expression of IGFBP-7 in human fibroblasts was enhanced by TGF-β and that IGFBP-7 significantly stimulated the proliferation and migration of fibroblasts[18]. However, no evidence so far has directly linked IGFBP-7 to TGF-β signaling in VSMCs. Our results showed that TGF-β stimulates the expression of IGFBP-7 in VSMCs thus further investigation of the role of IGFBP-7 in TGF-β stimulated VSMCs may lead to better understanding of TGF-β/Smad3 signaling pathways.

Protein disulfide-isomerase A3 (PDI-A3) belongs to the protein disulfide isomerase family, which encompasses several highly divergent proteins that participate in maturation of secretory proteins in the endoplasmic reticulum[19]. Functioning as chaperones, PDI-family proteins are involved in the proper folding and in the formation and reshuffling of the disulfide bridges of the proteins

synthesized in the rough ER. However, proteins of the PDI-family have also been found on the cell surface or as secreted proteins. Results from previous studies confirmed the presence of chaperones such as PDI-A3 and HSP-60 on cell surface suggesting that their phosphorylation might be responsible for reorientation of receptor complexes as well as key molecules in the process of cell activation[20]. Furthermore, it has been shown that the disulfide exchange function of PDI-A3 is required for cell mediated adhesion by integrins[21]. In a recent study, the phosphorylation of PDI-A3 was shown to be up-regulated in VSMCs activated by recombinant human platelet derived growth factor-BB (PDGF-BB), indicating PDI-A3 is a factor downstream of the receptor signaling cascade[22]. Given the similar effect that TGF-β may have in VSMCs, it is highly likely PDI family proteins are key factors involved in TGF-β dependent VSMC activation.

CONCLUSIONS

The secretome of VSMC was characterized by MS-based shotgun proteomics. In total, 665 non-redundant proteins were identified in the conditioned media of VSMC culture and TGF-β stimulated VSMC culture by a combination of LC-MS/MS techniques and bioinformatics analysis. A label-free quantification analysis via spectral counting revealed a list of 38 secreted proteins in that are up or down-regulated following Smad3-dependent TGF-β stimulation. Functional annotation of TGF-β/Smad3 induced proteins revealed several proteins with important functions related to TGF-β/Smad3 signaling that could potentially serve as the targets for future investigation of the mechanism through which TGF-β enhances intimal hyperplasia in

VSMCs. Altogether, our investigation sheds light on the intricate role of TGF-β signaling in the atherosclerotic process.

REFERENCES

1. (a) Mathers, C. D.; Loncar, D., Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* **2006,** *3* (11), e442; (b) Murray, C. J.; Lopez, A. D., Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet* **1997,** *349* (9064), 1498-504; (c) WHO publishes definitive atlas on global heart disease and stroke epidemic. *Indian J Med Sci* **2004,** *58* (9), 405-6.

2. Greenland, P.; Knoll, M. D.; Stamler, J.; Neaton, J. D.; Dyer, A. R.; Garside, D. B.; Wilson, P. W., Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. *JAMA* **2003,** *290* (7), 891-7.

3. (a) Perrins, C. J.; Bobryshev, Y. V., Current advances in understanding of immunopathology of atherosclerosis. *Virchows Arch* **2011,** *458* (2), 117-23; (b) Ross, R.; Glomset, J. A., Atherosclerosis and the arterial smooth muscle cell: Proliferation of smooth muscle is a key event in the genesis of the lesions of atherosclerosis. *Science* **1973,** *180* (4093), 1332-9.

4. Davies, M. G.; Hagen, P. O., Pathobiology of intimal hyperplasia. *Br J Surg* **1994,** *81* (9), 1254-69.

5. Rzucidlo, E. M.; Martin, K. A.; Powell, R. J., Regulation of vascular smooth muscle cell differentiation. *J Vasc Surg* **2007,** *45 Suppl A*, A25-32.

6. Newby, A. C.; Zaltsman, A. B., Molecular mechanisms in intimal hyperplasia. *J Pathol* **2000,** *190* (3), 300-9.

7. (a) McCaffrey, T. A.; Consigli, S.; Du, B.; Falcone, D. J.; Sanborn, T. A.; Spokojny, A. M.; Bush, H. L., Jr., Decreased type II/type I TGF-beta receptor ratio in cells derived from human atherosclerotic lesions. Conversion from an antiproliferative to profibrotic response to TGF-beta1. *J Clin Invest* **1995,** *96* (6), 2667-75; (b) Mii, S.; Ware, J. A.; Kent, K. C., Transforming growth factor-beta inhibits human vascular smooth muscle cell growth and migration. *Surgery* **1993,** *114* (2), 464-70.

8. (a) Suwanabol, P. A.; Seedial, S. M.; Shi, X.; Zhang, F.; Yamanouchi, D.; Roenneburg, D.; Liu, B.; Kent, K. C., Transforming growth factor-beta increases vascular smooth muscle cell proliferation through the Smad3 and extracellular signal-regulated kinase mitogen-activated protein kinases pathways. *J Vasc Surg* **2012,** *56* (2), 446-54; (b) Suwanabol, P. A.; Seedial, S. M.; Zhang, F.; Shi, X.; Si, Y.; Liu, B.; Kent, K. C., TGF-beta and Smad3 modulate PI3K/Akt signaling pathway in vascular smooth muscle cells. *Am J Physiol Heart Circ Physiol* **2012,** *302* (11), H2211-9; (c)

Tsai, S.; Hollenbeck, S. T.; Ryer, E. J.; Edlin, R.; Yamanouchi, D.; Kundi, R.; Wang, C.; Liu, B.; Kent, K. C., TGF-beta through Smad3 signaling stimulates vascular smooth muscle cell proliferation and neointimal formation. *Am J Physiol Heart Circ Physiol* **2009,** *297* (2), H540-9.

9. (a) Dupont, A.; Pinet, F., The proteome and secretome of human arterial smooth muscle cell. *Methods Mol Biol* **2007,** *357*, 225-33; (b) Dupont, A.; Corseaux, D.; Dekeyzer, O.; Drobecq, H.; Guihot, A. L.; Susen, S.; Vincentelli, A.; Amouyel, P.; Jude, B.; Pinet, F., The proteome and secretome of human arterial smooth muscle cells. *Proteomics* **2005,** *5* (2), 585-96.

10. Makridakis, M.; Vlahou, A., Secretome proteomics for discovery of cancer biomarkers. *J Proteomics* **2010,** *73* (12), 2291-305.

11. Doroudgar, S.; Glembotski, C. C., The cardiokine story unfolds: ischemic stress-induced protein secretion in the heart. *Trends Mol Med* **2011,** *17* (4), 207-14.

12. (a) Rasmussen, L. M.; Wolf, Y. G.; Ruoslahti, E., Vascular smooth muscle cells from injured rat aortas display elevated matrix production associated with transforming growth factor-beta activity. *Am J Pathol* **1995,** *147* (4), 1041-8; (b) Nabel, E. G.; Shum, L.; Pompili, V. J.; Yang, Z. Y.; San, H.; Shu, H. B.; Liptay, S.; Gold, L.; Gordon, D.; Derynck, R.; et al., Direct transfer of transforming growth factor beta 1 gene into arteries stimulates fibrocellular hyperplasia. *Proc Natl Acad Sci U S A* **1993,** *90* (22), 10759-63.

13. Clowes, M. M.; Lynch, C. M.; Miller, A. D.; Miller, D. G.; Osborne, W. R.; Clowes, A. W., Long-term biological response of injured rat carotid artery seeded with smooth muscle cells expressing retrovirally introduced human genes. *J Clin Invest* **1994,** *93* (2), 644-51.

14. Zhang, F.; Tsai, S.; Kato, K.; Yamanouchi, D.; Wang, C.; Rafii, S.; Liu, B.; Kent, K. C., Transforming growth factor-beta promotes recruitment of bone marrow cells and bone marrow-derived mesenchymal stem cells through stimulation of MCP-1 production in vascular smooth muscle cells. *J Biol Chem* **2009,** *284* (26), 17564-74.

15. Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P., Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *Journal of proteome research* **2006,** *5* (9), 2339-47.

16. Zhang, Y.; Wen, Z.; Washburn, M. P.; Florens, L., Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal Chem* **2010,** *82* (6), 2272-81.

17. Oh, Y.; Nagalla, S. R.; Yamanaka, Y.; Kim, H. S.; Wilson, E.; Rosenfeld, R. G., Synthesis and characterization of insulin-like growth factor-binding protein (IGFBP)-7. Recombinant human mac25 protein specifically binds IGF-I and -II. *J Biol Chem* **1996,** *271* (48), 30322-5.

18. Komiya, E.; Furuya, M.; Watanabe, N.; Miyagi, Y.; Higashi, S.; Miyazaki, K., Elevated expression of angiomodulin (AGM/IGFBP-rP1) in tumor stroma and its roles in fibroblast activation. *Cancer Sci* **2012,** *103* (4), 691-9.

19. Ferrari, D. M.; Soling, H. D., The protein disulphide-isomerase family: unravelling a string of folds. *Biochem J* **1999,** *339 (Pt 1)*, 1-10.

20. (a) Turano, C.; Coppari, S.; Altieri, F.; Ferraro, A., Proteins of the PDI family: unpredicted non-ER locations and functions. *J Cell Physiol* **2002,** *193* (2), 154-63; (b) Goplen, D.; Wang, J.; Enger, P. O.; Tysnes, B. B.; Terzis, A. J.; Laerum, O. D.; Bjerkvig, R., Protein disulfide isomerase expression is related to the invasive properties of malignant glioma. *Cancer Res* **2006,** *66* (20), 9895-902.

21. Lahav, J.; Wijnen, E. M.; Hess, O.; Hamaia, S. W.; Griffiths, D.; Makris, M.; Knight, C. G.; Essex, D. W.; Farndale, R. W., Enzymatically catalyzed disulfide exchange is required for platelet adhesion to collagen via integrin alpha2beta1. *Blood* **2003,** *102* (6), 2085-92.

22. Lande, C.; Boccardi, C.; Citti, L.; Mercatanti, A.; Rizzo, M.; Rocchiccioli, S.; Tedeschi, L.; Trivella, M. G.; Cecchettini, A., Ribozyme-mediated gene knock down strategy to dissect the consequences of PDGF stimulation in vascular smooth muscle cells. *BMC Res Notes* **2012,** *5*, 268.

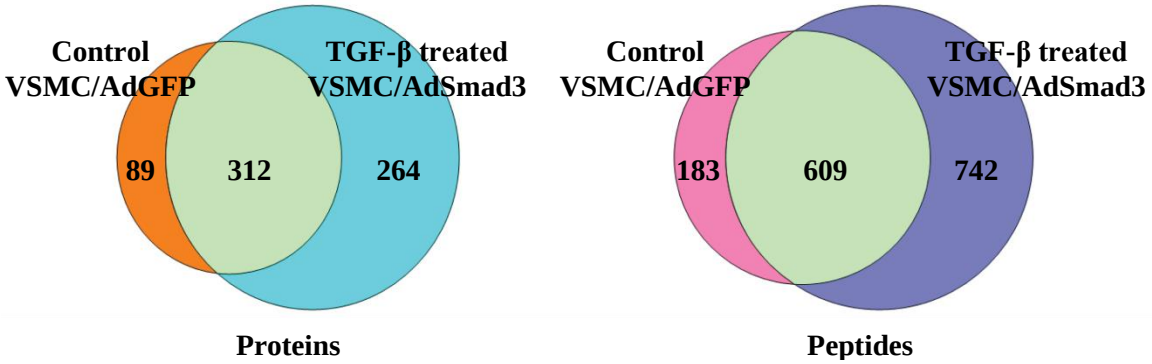
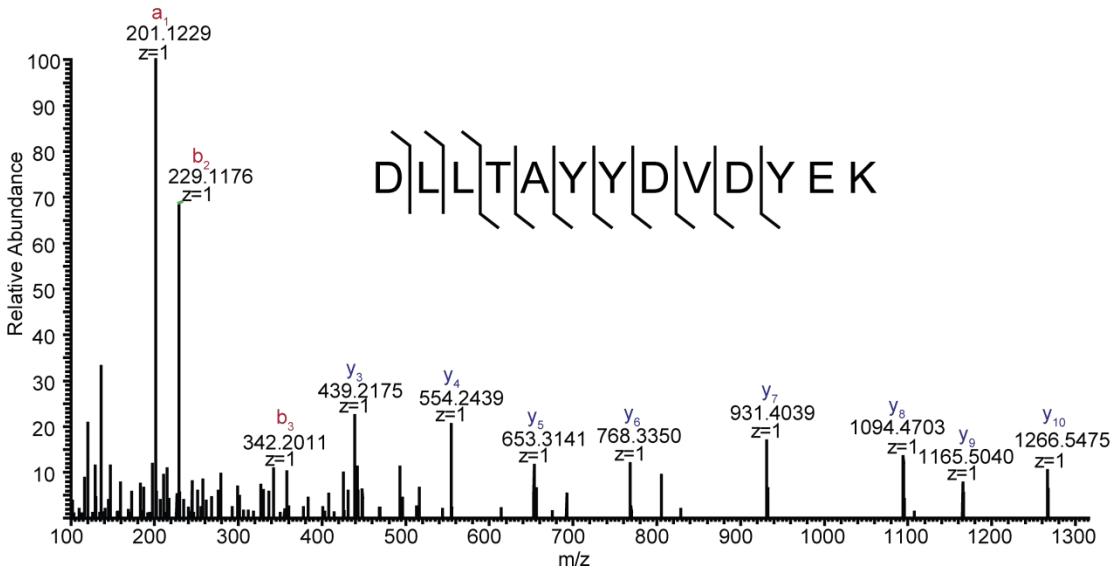


Figure 5.1 Venn diagram of the numbers of non-redundant proteins and peptides identified by LC-MS/MS from VSMC/AdGFP (control) conditioned media or TGF-β stimulated VSMC/AdSmad3 conditioned media. 312 non-redundant proteins and 609 non-redundant peptides were identified in the conditioned media of both cell cultures.



Protein Name: Protein disulfide-isomerase A3
Uniprot Accession number: P11598
Precursor *m/z*: 804.3754, 2+

Figure 5.2 A representative tandem MS/MS profile of the tryptic peptide DLLTAYYDVDYEK from protein disulfide-isomerase A3. MS/MS spectrum is from TGF-β treated sample with an elution at 77.06 minutes.

Accession	Gene Symbol	Σ# Unique Peptides	log2(TGF-β treated/control)	SecP	Signal peptide
G3V6X1	Fbln2	10	2.06	0.49	y
F1LTJ5	4	6	7.93	0.539	y
P11598	Pdia3	5	3.66	0.497	y
F1MAA7	Lamc1	4	7.12	0.266	y
D3ZGY4	Gapdh-ps2	4	6.50	0.513	-
P97615	Txn2	3	6.93	0.866	y
G3V818	Parva	3	6.70	0.802	-
F2Z3Q8	Kpnb1	3	5.44	0.592	-
Q62703	Rcn2	3	3.74	0.947	y
Q4FZT9	Psmd2	3	3.26	0.525	-
D4A3T3	Cbx1	3	2.58	0.872	-
P63281	Ube2i	3	2.38	0.749	-
D4ACB8	Cct8	3	2.38	0.539	-
Q9QZK5	Htra1	3	2.26	0.736	y
Q9JKL3	Bax	2	6.70	0.55	-
D4AE96	Ipo7	2	6.70	0.532	-
Q99MI5	Srm	2	6.12	0.533	-
F8WG91	Arl3	2	6.12	0.675	-
B1WC34	Prkcsh	2	6.12	0.419	y
Q07936	Anxa2	2	5.93	0.763	-
D4A0J7	Olfml2b	2	5.93	0.511	y
O54975	Xpnpep1	2	5.70	0.586	-
Q4FZU4	Adamtsl4	2	5.70	0.59	y
F1M335	4	2	5.70	0.484	y
Q00715	1	2	5.44	0.578	-
P83868	Ptges3	2	5.12	0.706	-
F8WFR8	LOC100359876	2	4.80	0.519	-
D4A259	Polr2d	2	4.70	0.918	-
Q71UF4	Rbbp7	2	4.60	0.645	-
P33436	Mmp2	2	4.12	0.586	y
P17246	Tgfb1	2	3.70	0.895	y
O08628	Pcolce	2	3.17	0.548	y
D3ZU54	LOC687565	2	2.38	0.8	-
F1M9B2	Igfbp7	2	2.23	0.573	y
F8WG89	Ptms	2	-5.11	0.741	-
P04785	P4hb	3	-2.20	0.738	y
Q6PEC4	Skp1	2	-4.69	0.649	-

Table 5.1 The list of significantly up- and down-regulated secreted proteins in VSMCs in response to Smad3-dependent TGF-β signaling.

*Note: accession is the Uniprot accession number for the identified protein. Proteins with $\log_2^{(\text{TGF-}\beta \text{ treated/control})} > 2$ were up-regulated, while those with $\log_2^{(\text{TGF-}\beta \text{ treated/control})} < -2$ were down-regulated. A number in the SecP column indicates the numerical score returned by SecretomeP (greater than 0.5 indicates high probability of

secretion). "y" in the signal peptide column indicates a predicted signal peptide from SignalP analysis.

Chapter 6

Comprehensive Mass Spectrometric Mapping of the Hydroxylated Amino Acid Residues of the α1(V) Collagen Chain

Adapted from: "Comprehensive Mass Spectrometric Mapping of the Hydroxylated Amino Acid residues of the α1(V) Collagen Chain." **Yang, C.**, Park, A. C., Davis, N. A., Russell, J.D., Kim, B., Brand, D., Lawrence, M. J., Ge, Y., Westphall, M. S., Coon, J.J., Greenspan, D. S. *J. Biol. Chem.* 2012, 287, (48), 40598-40610

ABSTRACT

Aberrant expression of the type V collagen α1(V) chain can underlie the connective tissue disorder classic Ehlers-Danlos syndrome, and autoimmune responses against the α1(V) chain are linked to lung transplant rejection and atherosclerosis. The α1(V) collagenous COL1 domain is thought to contain greater numbers of post-translational modifications (PTMs) than do similar domains of other fibrillar collagen chains, PTMs consisting of hydroxylated prolines and lysines, the latter of which can be glycosylated. These types of PTMs can contribute to epitopes that underlie immune responses against collagens and the high level of PTMs may contribute to the unique biological properties of the α1(V) chain. We here use high resolution mass spectrometry to map such PTMs in bovine placental α1(V) and human recombinant pro-α1(V) procollagen chains. Findings include the locations of PTMs that vary or are invariant between bovine tissue α1(V) and human recombinant pro-α1(V) chains. Notably, an unexpectedly large number of hydroxyproline residues were mapped to the X positions of Gly-X-Y triplets, contrary to expectations based on previous amino acid analyses of hydrolyzed α1(V) chains from various tissues. We attribute this difference to the ability of tandem mass spectrometry coupled to nanoflow chromatorgraphic separations to detect low-level PTM combinations with superior sensitivity and specificity. The data are consistent with the presence of a relatively large number of 3-hydroxyproline sites with less than 100 % occupancy, suggesting a previously unknown mechanism for the differential modification of α1(V) chain and type V collagen properties.

INTRODUCTION

Collagen type V (col(V)) is a low abundance fibrillar collagen that is widely distributed in vertebrate tissues as an $\alpha1(V)_2\alpha2(V)$ heterotrimer.[1] It is also found, with a more

limited tissue distribution, as an α1(V)α2(V)α3(V) heterotrimer, and as rare $\alpha1(V)_3$ homotrimers. [1-2] Thus, all forms of col(V) contain the α1(V) chain. Additionally, the closely related collagen type XI, first described as an α1(XI)α2(XI)α3(XI) heterotrimer in fetal cartilage,[3] incorporates α1(V) chains into heterotypic α1(XI)α1(V)α3(XI) trimers as cartilage matures,[4] implying additional roles for the α1(V) chain.

It is generally accepted that $\alpha1(V)_2\alpha2(V)$ heterotrimers are incorporated into growing fibrils of the more abundant collagen type I and are involved in regulating the geometry and properties of the resulting type I/V heterotypic fibrils.[5] Thus, mutations in the genes encoding either α1(V) or α2(V) chains result in the human connective tissue disorder classic Ehlers-Danlos syndrome, characterized by collagen I fibrils of abnormal shapes and diameters, and deficient tensile strength.[6] More recently, it has been demonstrated that anti-col(V) autoimmune responses can underlie chronic lung transplant rejection in both humans and animal models and that pre-transplant col(V)-specific autoimmunity is also a significant risk factor for primary graft dysfunction, the leading cause of early morbidity and mortality after lung transplantation.[7,8,9] Col(V) autoimmunity has also been identified as a consistent feature in both late stage human coronary artery disease and a mouse model of atherosclerosis.[10] In both human lung transplant rejection and coronary artery disease, immune responses have been shown to be specific to the α1(V) chain, with an absence of such responses to the α2(V) chain.[7, 10]

The α1(V) chain contains extensive post-translational modifications (PTMs), relative to other fibrillar collagen chains, consisting of 4-hydroxyproline (4-Hyp) and 3-hydroxyproline (3-Hyp) residues, and of hydroxylysine (Hyl) residues, most of which are decorated with di- and mono-saccharides.[11] Functionally, 4-Hyp residues have well

characterized effects in stabilizing triple helices,[12] while Hyl residues are important to the formation of stable covalent crosslinks between collagen chains,[13] and Hyl glycosylation may be important in assembly and secretion of at least some collagens.[14] The function of 3-Hyp residues is unclear, with conflicting reports on small stabilizing or destabilizing effects on triple helix stability.[15,16] However, loss of 3-Hyp residues in collagens can have catastrophic phenotypic consequences, implying important biological function(s).[17] In addition to functional roles, previous studies show various PTMs are capable of affecting immune responses against collagenous antigens.[18] For example, Pro hydroxylation can contribute to the antigenicity of collagen IV, while lysine hydroxylation and glycosylation are important to epitopes that affect the antigenic roles of collagen II in rheumatoid and experimentally induced arthritis.[18a,18b-h]

Recent studies investigating PTMs on collagenous proteins have utilized tandem mass spectrometry (MS^n) on low resolution, low mass accuracy ion trap mass spectrometers, or mass measurements of large peptides using matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) MS.[11a, 19] However, performing MS^n on instruments with low mass accuracy, or relying on MS measurements of peptides that have not been subjected to MS^n, makes confident peptide identification and unambiguous PTM site localization challenging. The analysis of collagenous proteins by mass spectrometry is complicated by the high degree of modification and the large number of proline residues. The former leads to generation of isobaric, but differentially modified, peptides that often co-elute, producing chimeric spectra difficult to interpret without the aid of high mass accuracy and abundant product ions from MS^n. The latter hampers MS^n sequencing efforts due the propensity of Pro/Hyp to preferentially cleave during collisional activation (the "proline effect").[20] This can limit the generation of a ladder of b- and y-type product ions most commonly used by

database search algorithms for peptide identification and PTM localization of spectra generated by collisional activation [20a, 21].

In the current study we use state-of-the-art proteomics workflow centered on the use of high resolution MS and nanoflow chromatographic separations to precisely map hydroxylation and glycosylation modifications in bovine placental α1(V) and human recombinant pro-α1(V) procollagen chains. Instead of relying on a single enzyme or chemical treatment to produce peptides amenable to MS, we used five different proteases (individually or sequentially) to maximize protein sequence coverage and facilitate PTM discovery. Peptides were subsequently separated by nanoflow liquid chromatography and introduced into an electron transfer dissociation (ETD)-enabled hybrid dual cell linear ion trap-Orbitrap mass spectrometer by electrospray ionization (ESI). The Orbitrap recorded the masses of eluting peptides with low ppm mass accuracy (< 5 ppm) affording confident peptide identification and localization of PTMs. The hybrid Orbitrap mass spectrometer permitted the use of several dissociation techniques; resonant excitation collision-activated dissociated (CAD), beam-type CAD (higher-energy collisional dissociation, HCD), and ETD.[22] The availability of multiple dissociation techniques and the use of MS^2 and MS^3 for peptide interrogation often allowed us to pinpoint the exact residue(s) carrying the individual PTMs. Such comprehensive, experiment-based PTM localization has not previously been achieved for collagenous proteins. Recent analyses of collagenous proteins by MS relied heavily on a priori biological knowledge to assess prolyl hydroxylation (i.e., PTM assignment based upon hydroxylation motifs), rather than PTM assignment based upon localizing fragments from tandem mass spectra.[19a, 19c] We report comprehensive mapping of all PTMs involving hydroxylated residues on bovine placenta α1(V) and human recombinant pro-α1(V) collagen chains and provide manually verified mass spectral evidence for each modified site. Our

analyses reveal PTMs that vary or are invariant between the bovine tissue α1(V) and human recombinant pro-α1(V) chains, and also reveal all hydroxylated residue PTMs in pro-α1(V) sequences NH_2-terminal to the COL1 major collagenous domain. We also identify an unexpectedly large number of Hyp residues in the X-position of Gly-X-Y triplets, attributed to our ability to identify modified peptides present in low stoichiometric amounts that go undetected by classic amino acid analyses or Edman sequencing. Another unexpected, and striking, finding was X position Hyp residues discovered in the unusual contexts of Gly-Hyp-Val and Gly-Hyp-Ala triplets in the bovine placental α1(V) COL1 domain. Findings presented herein may aid in characterizing and locating α1(V) autoimmune epitopes and may provide further insights into col(V) function.

EXPERIMENTAL PROCEDURES

Preparation of Human Pro-α1(V) Procollagen and Bovine α1(V) Collagen Chains

Human recombinant pro-α1(V) chains were produced via expression from a modified pCEP-Pu vector in 293-EBNA human embryonic kidney cells (Invitrogen), followed by dialysis of conditioned media against 50 mM Tris-HCl, pH 8.6 buffer, containing 0.1 mM phenylmethylsulfonyl fluoride, 1 mM *N*-ethylmaleimide, 0.1 mM *p*-aminobenzoic acid, and 5 mM EDTA, for low salt precipitation of collagen chains, as previously described.[23] Bovine α1(V) chains were prepared essentially as previously described.[24] Briefly, minced and washed amnion, stripped off placenta from the area close to attachment of the umbilicus, was suspended in 0.5 M acetic acid, 0.2 M NaCl and digested with pepsin at 4 ^{o}C. Col(V), which is soluble in 0.7 M NaCl and precipitates in 1.2 M NaCl, was purified from supernatants via differential NaCl precipitation.[25] α1(V) chains were separated from α2(V) chains via chromatography on diethylaminoethyl cellulose.

Prolyl Hydroxylase Expression

Breast carcinoma cell line MB 436 (ATCC catalog no. HTB-130) and HEK293 T-REx cells were maintained in Dulbecco's modified Eagle's media (Cellgro) supplemented with 10% FBS and Pen/Strep in 5% CO_2. Total RNA was extracted from cell lines using TRIzol (Invitrogen). cDNA was synthesized from 1 μg RNA using SuperScript II Reverse Transcriptase kit (Invitrogen) with random hexamers. Subsequently, PCR was performed for 35 cycles with primer sets for human P3H1, P3H2, P3H3, CRTAP, and PPIB previously described by Frenandes et al. and for GAPDH previously described by Shah et al..[19b,26]

Pro-α1(V) Sequences for Comparison to MS Data

For human pro-α1(V) sequences, MS data were compared to sequences available in the databases. For bovine α1(V) sequences, a BLAST search of bovine genome databases, using human pro-α1(V) sequences identified mRNA sequence XM_002691720.1, predicted by automated analysis of genomic sequences, which encodes the pro-α1(V) C-propeptide, and ~38% of the α1(V) major triple helical (COL1) domain, interspersed with gaps and insertions of non-α1(V) sequences, probably due to annotative misidentification of intron/exon junctions.[27] Towards obtaining complete bovine α1(V) COL1 sequences, first strand cDNA was synthesized from bovine placenta total RNA (Zyagen) using oligo$(dT)_{20}$ primers. Primers 5'-AGGCCCCCCAGGCGAGGTC-3' (forward) and 5'-ATTCTGGCCCCTTCAGACTT-3' (reverse), based on XM_002691720.1 COL1 sequences, were then used to amplify a ~1.8 kb fragment of COL1 sequences, which was directly sequenced using the same forward primer. Another predicted mRNA sequence (XM_617549.3), identified via subsequent BLAST search, encodes partial pro-α1(V) NH_2-terminal globular sequences, and some inserted intronic sequences. To obtain remaining

bovine α1(V) COL1 sequences, forward primer 5'-GTGGCACAGAATTGCTCTCA-3', based on XM_617549.3 sequences, was used with primer 5'-GCCTTCACTGCCTTTCAGTC-3' (reverse primer A) based on XM_002691720.1 sequences, to amplify a ~2.7 kb fragment. The ~2.7 kb fragment could not be amplified from first strand cDNA synthesized using oligo$(dT)_{20}$ primers, as above, and was instead amplified from first strand cDNA obtained using reverse primer A. The ~2.7 kb PCR fragment was directly sequenced, sequentially, using reverse primers: 5'-GTGGCACAGAATTGCTCTCA-3', 5'-CTCCAAGTTTGCCCTTCTCC-3', and 5'-GCCCCTTTCTCCGTCTTC-3'. The full-length bovine α1(V) COL1 sequences have been submitted to the GenBank™/EMBL Data Bank with accession number JQ611730.

Digestion of Pro-α1(V) and α1(V) Samples for Mass Spectrometry

Pro-α1(V) and α1(V) samples were desalted using 50 mg, tC18 SepPak cartridges (Waters Corp., Milford, MA). Eluates were dried down and resuspended in digestion buffer (*vide infra*) optimized for each protease. Cysteine residues were reduced and alkylated by incubation in 5 mM dithiothreitol (DTT) for 45 min at 37 ^{0}C, followed by 30 min incubation at room temperature in 15 mM iodoacetamide, in the dark. Alkylation was quenched by 15-min incubation in 5 mM DTT at room temperature followed by addition of the protease.

A multiple protease approach was utilized to maximize protein sequence coverage.[28] For each protease, 50 µg protein aliquots of Pro-α1(V) or α1(V) were digested. To produce peptides cleaved N-terminal to Asp, protein was incubated overnight at 37 °C in 50 mM Tris, 2.5 mM Zinc sulfate, pH 8.0, with 5 µg of the protease AspN (New England Biolabs, Ipswich, MA). Peptides cleaved C-terminal to Glu were prepared by incubating protein with 5 µg of GluC (Roche Diagnostics, Indianapolis, IN) overnight in 50 mM Tris and 25 mM ammonium bicarbonate (pH ~ 7.8) at room temperature. Peptides predominately cleaved C-terminal to

Trp, Leu, Tyr, Phe (with slower kinetics for cleavage at Met, Asp, Glu, Ala) was achieved with 5 μg of chymotrypsin (Promega, Madison, WI) incubated overnight in 50 mM Tris, 1 mM $CaCl_2$ at room temperature. Protein digested overnight in 90 mM Tris, 8.5 mM $CaCl_2$, 5 mM DTT, 0.5 mM EDTA, pH 7.6 at 37 °C, with 5 μg of ArgC (Roche Diagnostics) produced peptides cleaved C-terminal to Arg. Tryptic peptides (cleavage C-terminal to Lys and Arg) were produced by digesting protein with 5 μg of trypsin (Promega) overnight in 50 mM Tris (pH 8.0), and 1 mM $CaCl_2$ at 37 °C. Sequential digestion using GluC and AspN was carried out by first digesting protein with GluC for 6 hrs at room temperature, followed by the addition of AspN and incubation overnight at 37 °C. Each digest was quenched by freezing at -80 ^{0}C, and desalted on a 50 mg, tC18 SepPak cartridge (Waters Corp.). Each eluate was then dried under vacuum and resuspended in 50 μl 0.2% formic acid resulting for mass spectrometry analysis (~ 1 μg/μl peptide concentration).

Liquid Chromatography Electrospray Tandem Mass Spectrometry (LC-MS/MS)

A Waters nanoACQUITY UPLC and autosampler (Waters Corp.) were used to load samples onto a fused-silica capillary precolumn (75 μm i.d. x 360 μm o.d.). Both vented-style configuration trapping and direct injection strategies were used (*vide infra*). Precolumns with cast chemical frits were slurry packed to 5 cm in length with a stationary phase consisting of 5 um diameter, 100 Å pore size, C18 particles (Magic C18AQ, Michrom Bioresources, Inc., Auburn CA).[29] Reversed-phase LC separation was achieved across a 13-cm long, 50 μm i.d. x 360 μm o.d. fused-silica analytical column packed with the same C18 stationary phase. An electrospray ionization emitter was integrated into the analytical column by use of a laser puller (Sutter Instrument Co., P-2000, Novato, CA). An estimated 1-4 μg of protein digest was loaded onto the precolumn and chromatographed over a 60-min linear gradient at 300 nl/min (2% B to 30% B, buffer A: 0.2% formic acid in water; buffer B: 0.2%

formic acid in acetonitrile). Eluate was introduced into the mass spectrometer via electrospray ionization (+2.0 kV) and peptide cations were subjected to tandem mass spectrometry using an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) enabled for electron-transfer dissociation (ETD).[22b, 30] Typical experiments consisted of MS^1 analysis in the Orbitrap mass analyzer using a resolving power of 60,000, followed by 10 data-dependent higher-energy collisional dissociation (HCD) MS^2 events. Product ion mass analysis was also conducted in the Orbitrap, using resolving powers of 7,500 – 15,000. Precursor and product ion mass error was typically < 10 ppm. Similar experiments were conducted using ETD, where charge-state dependent ETD activation times were utilized. To obtain additional sequence information of glycosylated peptides, an additional set of LC analyses were conducted, using an MS^3 mass spectrometry method. For MS^3 analyses, three data-dependent collision-induced dissociation (CID) or ETD MS^2 events were executed followed by data-dependent HCD MS^3 events, where the most intense peak in each MS^2 spectrum was dissociated. For all experiments, an automatic gain control (AGC) target value of 1,000,000 charges was used for MS^1. An AGC target of 50,000 charges was used for both MS^2 and MS^3. Precursors were dynamically excluded from data-dependent MS^2 for 30-60 s. A precursor isolation width of 2 -3 *m/z* was typically used.

Proteomic Data Analysis

Spectral reduction from raw data was performed by DTA Generator, a program available in the COMPASS software suite (freely available at http://www.chem.wisc.edu/~coon/software.php).[31] The ETD pre-processing option was used to remove known neutral loss peaks from ETD spectra.[32] The OMSSA (Open Mass Spectrometry Search Algorithm) search algorithm was used for database correlation.[33] Spectra were searched against a concatenated target-decoy version of human Pro-α1 (V) or

bovine α1 (V) sequence databases.[34] Several variable post-translational modifications were considered: hydroxylation of proline and lysine residues (+15.9949 Da, monoisotopic mass), glycosylation of lysine (monosaccharide, +178.0473 Da, disaccharide, +340.0995), and glycosylation-associated neutral losses upon activation. Finally, oxidation of methionine (+15.9949 Da) residues were specified as variable modification and carbamidomethylation of cysteine (+57.0215 Da) residues was searched as a fixed modification. A mass tolerance of ±5 Da from the average mass was used for precursors, while a mass tolerance of ± 0.01 Da from the monoisotopic mass was used for product ions.[22a]

Manual Validation of Peptide Modification Assignments

Peptides identified through database correlation with P-value scores larger than $1x10^{-9}$ were discarded and were not considered for manual validation. Although OMSSA usually identified the correct peptide sequence, the number, type, and specific location of modifications were occasionally incorrect. Manual verification of the peptide sequence and site(s) of modification was carried out for every modified peptide species discussed in this manuscript (supplemental spectra S1 and S2). Spectral validation was carried out as follows: the mass of the unmodified peptide sequence was compared to the modified peptide sequence identified with OMSSA. This mass difference represents the mass of the PTM(s) on the peptide (i.e. a mass difference of +15.995 Da relative to the unmodified peptide suggests hydroxylation or oxidation). The high mass accuracy of the Orbitrap limited the potential number of modifications that needed to be considered. Each peptide was iteratively evaluated considering all combinations of PTMs on residues that can accommodate the modifications. Theoretical product ion masses were produced using the MS-Product program available in the web version of Protein Prospector (http://prospector.ucsf.edu). Product ions were confidently mapped to peptide sequences if the production ions had a signal-to-noise

ratio > 3 and mass error < 10 ppm from the expected product ion monoisotopic mass. Modifications indicated as localized had product ion matches that unambiguously mapped the PTM to a specific residue in the peptide sequence. In situations where product ions did not support PTM assignment to a specific residue, the modification was considered unlocalized and is clearly indicated.

RESULTS AND DISCUSSION

Post-translational Modifications of Bovine Placental α1(V) Chains

Chromatographically purified α1(V) chains, extracted with acetic acid and pepsin from bovine placenta, were subjected to tandem mass spectrometry (MS^n) to map the positions of hydroxylated residues and of Galactosylhydroxylysyl (Gal-Hyl) and glucosylgalactosylhydroxylysyl (Glc-Gal-Hyl) residues. Treatment of fibrillar collagens with pepsin removes NH_2- and COOH-terminal globular sequences, leaving only the pepsin-resistant major triple helical COL1 domain. At the time this study began, bovine α1(V) cDNA sequences were unavailable and only a small portion (C-propeptide sequences and about the C-terminal third of COL1 sequences) of bovine α1(V) coding sequences were available from annotated bovine genome databases. Thus, full-length bovine α1(V) cDNA sequences were generated, as described in Experimental Procedures, for comparison to MS data. At the time of submission of this study, database annotated genomic sequences were mostly complete, but still had 29 amino acid differences compared to the cDNA sequences reported here (accession number JQ611730). The cDNA sequences reported here have been validated via comparison of MS analyses of bovine placental α1(V) protein.

MS^n of AspN, GluC, chymotrypsin, ArgC, and trypsin-generated peptides of bovine placental α1(V) chains produced 94% sequence coverage and identified the positions of 105

Hyp residues in the Y position of Gly-X-Y triplet repeats, and 21 Hyp residues in the X position of Gly-Hyp-Hyp triplet repeats (Figure 6.1). Y position Pro residues in Gly-X-Y triplets are hydroxylated to 4-Hyp residues by the enzyme prolyl 4-hydroxylase (P4H), for which the minimum substrate appears to be the tripeptide X-Pro-Gly.[35] Thus, the 105 Y position Hyp residues reported here are likely 4-Hyp. Gly-Pro-Hyp triplets are thought to constitute the substrate sequence for prolyl 3-hydroxylases (P3Hs), of which there are three.[36] We conclude the 21 Hyp residues in the X position of Gly-Hyp-Hyp triplet repeats in the COL1 domain of bovine placental α1(V) chain are likely to be 3-Hyp residues. Unexpectedly, Hyp residues were detected in the X positions of a Gly-X-Val triplet and a Gly-X-Ala triplet (residues 509 and 587, respectively) (Figs. 1 and 2). As P4H is thought to hydroxylate only Y position Pro residues, and as P3H is thought to hydroxylate only Pro residues within Gly-Pro-Gly triplets, it is not clear which enzyme has hydroxylated these sites, or whether these hydroxylated prolines are 3-Hyp or 4-Hyp residues. Our analysis also identified the positions of 3 Hyl residues, and 34 Glc-Gal-Hyl residues (Figure 6.1), Hyl residues linked to the dissacharide glucosylgalactose.

Hydroxylated Residues and Glycosylated Hyl Residues of Human Recombinant Pro-α1(V) Chains

A number of previous studies on the biology of pro-α1(V) collagen chains have been performed on human recombinant pro-α1(V) collagen chains produced in 293-HEK (human embryonic kidney) cells.[23, 37] As previous, partial characterizations of fibrillar collagen chain post-translational modifications have shown tissue- and cell type-specific variations in distributions of 3-Hyp residues, we mapped the hydroxylated residues and saccharide-bound Hyl residues of the experimentally important human recombinant full-length pro-α1(V) collagen chains produced in 293-HEK cells.[19] This study of procollagen chains also allowed

characterization of hydroxylated amino acid residues in sequences NH_2-terminal to the COL1 domain and in COOH-propeptide sequences, both of which are lost in pepsin-extraction of α1(V) chains from tissues.

MS^n of AspN, GluC, chymotrypsin, ArgC, and trypsin-generated peptides of human recombinant pro-α1(V) collagen chains produced 90 % sequence coverage of the entire protein, minus signal peptide sequences, and 96 % of the COL1 domain, and identified the positions of 98 Hyp residues in the Y position of Gly-X-Y triplet repeats, and 9 Hyp residues in the X position of Gly-Hyp-Hyp triplet repeats in the COL1 domain (Figure 6.3). As described above, Hyp residues in the Y position of Gly-X-Y repeats, or in the X position of Gly-Hyp-Hyp triplet repeats are likely 4- and 3-Hyp residues, respectively. Interestingly, within pro-α1(V) sequences NH_2-terminal to the COL1 domain, MS analysis mapped 6 Hyp residues in the Y position of Gly-X-Y triplets, 1 Hyp residue in the X position of a Gly-Hyp-Hyp triplet, one Hyl and two Glc-Gal-Hyl residues. Five of the Y position Hyp residues exist within a hypothetical short, interrupted collagenous (COL2) subdomain (Figure 6.3), consistent with the likelihood that this subdomain actually forms a triple helix, as the latter would be stabilized by these five 4-Hyp residues.[27] Possible functional significance of the sixth 4-Hyp, which lies upstream of this hypothetical collagenous domain, is unclear. All NH_2-terminal domain Glc-Gal-Hyl, Hyl and X position Hyp residues lie within the COL2 subdomain, suggestive of functional roles for these residues within this subdomain as well. Supportive of our findings, the single COL2 domain Hyl mapped here was previously identified by MS analysis as an "NH_2-telopeptide" Hyl residue involved in covalent cross-linking of α1(V) chains to α1(XI) chains in bovine cartilage.[4] MS^n analysis here of the human recombinant pro-α1(V) chain also localized the positions of 6 Hyl, 1 Gal-Hyl, and 22 Glc-Gal-Hyl residues, and one residue found as both Gal-Hyl and Glc-Gal-Hyl, within the

COL1 domain. It seems unlikely that species-specific differences in α1(V) sequences contributed much to differences in numbers and placement of Y position Hyp residues, as bovine and human α1(V) COL1 domain sequences are 99.3% identical, differing by only 7 primary amino acids, none of which is a Pro (Figs. 1 and 3, orange residues). Modifications could not be localized for Y position Pro residues 642, 714, 903, or 909 in either human recombinant pro-α1(V) or bovine placenta α1(V) chains.

Expression of Prolyl 3-hydroxylases in 293-HEK cells

Differential 3 hydroxylation of prolines at some, but not other, sites in clade A fibrillar collagen chain COL1 domains can be due to differential expression of the enzyme prolyl 3-hydroxylase 2 (P3H2) in different cell types and tissues.[19b] To test whether the absence of 3-Hyp residues at some sites in human recombinant pro-α1(V) chains that had been 3 hydroxylated in α1(V) chains from bovine placenta might be due to deficiency in levels of P3H2, we tested for P3H2 expression in 293-HEK cells. We also tested for expression of the other two proly 3-hydroxylases, P3H1 and P3H3, and for CRTAP (cartilage associated protein) and PPIB (peptidyl prolyl *cis-trans* isomerase B), which are components of the prolyl 3-hydroxylation complex.[38] As can be seen (Figure 6.4), all P3Hs are clearly expressed in 293-HEK cells, at levels similar to those of a cell line previously shown positive for expression of all three enzymes.[26] CRTAP and PPIB expression levels were readily detectable as well. These findings are thus consistent with the likelihood that the decreased numbers of 3-Hyp residues in human recombinant pro-α1(V) chains, compared to the tissue form of α1(V) chain obtained from bovine placenta, is unlikely due to decreased levels of P3Hs or other prolyl 3-hydroxylation complex components in 293-HEK cells. Rather, the reduction in 3-Hyp residues in the recombinant pro-α1(V) chains is likely due to some other

variable, such as the changed kinetics with which pro-α1(V) chains are incorporated into pro-α1(V)$_3$ homotrimers (see below), the form in which they were produced in HEK-293 cells for the current study, compared to the kinetics of incorporation of pro-α1(V) chains into pro-α1(V)2pro-α2(V) heterotrimers, in which they are more normally found in tissues.

A High Mass Accuracy, High Resolution Tandem Mass Spectrometric Approach for Analysis of Collagenous Proteins

Study of the PTMs of collagens has primarily involved traditional amino acid analysis, a methodology that determines amino acid compositions by hydrolysis of purified proteins/peptides followed by measurement of their individual amino acid constituents. Ordinarily, such analyses yield quantitative measurements of amino acid abundance, but not sequence information or the location of PTMs. More recently, MS has been employed to investigate PTMs in collagens.[11a, 19] Henkel and Dreisewerd utilized ultraviolet MALDI MS to analyze fetal calfskin collagens I, III and V, chemically cleaved by cyanogen bromide (CNBr) to generate a limited number of peptides with distinct masses.[11a] PTMs were deduced for each CNBr peptide by comparing the difference between experimental and theoretical masses – the mass difference representing the mass of the modification(s). However, MS^n was not performed to confirm sequence identity or to locate the positions of PTMs. This approach provided valuable information on the total PTM state, but without the use of MS^n, specific residues carrying the PTMs remained unknown. Eyre and colleagues have characterized collagen peptides by performing MS^2 using a low-resolution, low mass accuracy ion trap mass spectrometer.[19] They identified several novel sites of prolyl 3-hydroxylation and proposed a biological function for these PTMs in collagen. However, many of the modifications were not experimentally localized by MS^2. Since the acquired tandem mass spectra lacked the requisite information for localization assignment of Pro

hydroxylation was often inferred from known collagen motifs. Because of the highly modified nature of collagenous proteins, we contend that PTM mapping via MS should require spectral evidence in the form of localizing fragments to have the highest confidence in the validity of sequence-specific modifications.

Our strategy, founded on a high resolution MS platform, allowed characterization of the α1(V) chain with unparalleled sensitivity and PTM localization precision. The enabling technology was an ETD-enabled LTQ Orbitrap Velos hybrid mass spectrometer exhibiting low-ppm mass accuracy, high-resolution MS and MS^n capabilities, combined with multiple options for performing peptide dissociation.[30a] We began by using multiple proteases to generate a diverse pool of peptides amenable to MS^n.[28] The combination of mass accuracy and a capability to perform multistage activation using 3 different dissociation techniques often allowed us to pinpoint the exact residue(s) carrying individual PTMs. However, we were not always able to unambiguously define the site of localization. Thus, we classified PTM assignments into 1 of 3 categories: 1) localized – MS^n product ions support PTM assignment to a specific residue, 2) unlocalized – insufficient product ions to unambiguously assign a PTM to a specific residue, or 3) pseudolocalized - insufficient product ions to unambiguously assign a PTM to a specific residue, but localization can be inferred from known collagen modification motifs (supplemental Tables S1 and S2). Note that hydroxylation modification was only considered for Pro and Lys residues and that glycosylation (mono- and disaccharide) modifications were only considered for Hyl residues.[11b] The numbers of PTMs reported in the Results and Discussion sections are exclusively for localized sites.

α1(V) and pro-α1(V) chains were cleaved using five different individual proteases or by applying two proteases sequentially. Trypsin, the most common protease used for MS

experiments, has substantially reduced cleavage rates for Lys and Arg residues C-terminal to Pro, thus posing a problem for collagens, which have high Pro/Hyp content.[39] Such is also the case with chymotrypsin and Glu-C. However, Arg-C will cleave at Arg residues (also at Lys, but at lower rates) adjacent to Pro residues. Peptides were separated by on-line nanoflow reversed-phase liquid chromatography, where peptides with identical primary sequences differing only in degree of hydroxylation eluted across a wide elution window, while peptide isomers differing only in the position of hydroxylation typically co-eluted and were co-fragmented to produce chimeric MS^n spectra (Figure 6.5). Glycopeptides and short (< ~ 7 residues), highly modified peptides often eluted early during the chromatographic gradient, demonstrating poor and inconsistent retention. A direct injection style of sample loading was employed to enable analysis of highly polar and poorly retained peptides rather than the vented-style trapping normally used for rapid sample loading/concentrating.[40]

Due to high α1(V) Pro content and decreased efficacy of ETD-induced cleavage of Pro N-C_α bonds, ETD MS^2 analyses occasionally did not provide sufficient numbers of sequence-informative product ions to permit PTM localization.[41] To complement ETD data, data were also acquired utilizing HCD collected at different collision energies. Low HCD energies favor the generation of moderate mass fragment ions, whereas high HCD energies favor the production of low *m/z* fragments. Low mass fragments enabled screening of y_1-y_3, b_1-b_3, and immonium ions, improving the ability to localize PTMs residing on peptide termini. The high α1(V) Pro content also resulted in an unusually large number of internal fragments (multiple peptide backbone cleavages of a single ion) in spectra collected at moderate and high HCD energies.[21a] Many of these internal fragments were useful for PTM localization in the absence of localizing product ions produced by a single backbone cleavage event (supplemental Spectra 1 and 2). Despite our best efforts using multiple dissociation

techniques and MS^3, neutral loss of glycosylation was still common. In these cases, neutral loss product ions were used to aid in localization, as reported for phosphorylation localization (supplemental Spectra 1 and 2).[42]

Comparison of the Col1 PTMs of Bovine Placental α1(V) and Human Recombinant Pro-α1(V) Chains

Previous amino acid analyses of α1(V) chains from human skin or placenta, or produced in 293-HEK cells produced estimates of 4-Hyp content of from ~106 to ~111 per 1000 amino acids.[11b,43,37e] Here, we have localized 105 and 98 Y position Hyp residues by MS^n for the COL1 domains of bovine placenta α1(V) and human recombinant pro-α1(V) chains, respectively. Additional Y position Hyp residues were either not directly localized or are likely located on COL1 proteolytic fragments not recovered here by MS^n, as such fragments contain 5 additional Y position Pro residues for both the bovine and human. Eighty five Y position Hyp residues mapped to identical positions in the bovine and recombinant human COL1 domains (Figs. 1 and 2), suggesting that these Hyp residues may be relatively invariant in α1(V) chains from various sources. Of the 13 Y position Hyp residues that differ between the bovine α1(V) and human recombinant pro-α1(V) samples only 3 are found in the latter but not the bovine. Thus, Y position Hyp residues found in the human recombinant pro-α1(V) COL1 domain are, for the most part, a subset of those found in the bovine α1(V) chain. Differences in Y position Hyp residues between the two COL1 domains may be due to tissue-/cell type-specific differences in modifying enzymes, although the pro-α1(V) chains studied here were synthesized as pro-$\alpha1(V)_3$ homotrimers rather than as pro-$\alpha1(V)_2$pro-α2(V) heterotrimers more commonly found in tissues, which might also affect levels and placement of Hyp residues.

Previous estimates of the amino acid composition of α1(V) chains from human skin, suggested α1(V) to be more similar to nonfibrillar basement membrane collagen IV chains than to other fibrillar collagen chains in possessing a high content of Hyl residues, the majority of which are glycosylated.[11b] Specifically, human skin α1(V) chains were estimated to comprise ~39 hydroxylysines, 29 in the form of Glc-Gal-Hyl residues and 5 in the form of Gal-Hyl residues, suggesting Glc-Gal and Gal to be frequently and infrequently occurring PTMs, respectively.[11b] Henkel and Dreisewerd, employing MALDI MS analysis of fetal calf skin α1(V) chains, estimated that most, if not all, Hyl residues are likely glycosylated, with ~87% predicted to be Glc-Gal-Hyl.[11a] However, they lacked MS^n sequence data to support their assignments of PTMs, making it difficult to confidently determine the nature of saccharide species (e.g., two residues with monosaccharide modifications and one residue with a disaccharide modification share the same mass, such that these PTMs cannot be distinguished based solely on analysis of intact peptide masses). Rhodes and Miller employed amino acid analysis to estimate human placenta α1(V) chains to contain 35 Hyls, but did not analyze glycosylation.[43] Here, we've directly localized 37 hydroylysines in bovine placenta α1(V) chains, 34 of these as Glc-Gal-Hyl residues. Our estimates of Gal-Hyl modification of bovine placenta α1(V) may differ from those of previous reports because 1) We exclusively reported localized modifications, disregarding peptides with multiple potential glycosylation sites that could not be localized to a specific residue; and/or 2) Previous reports did not attempt to localize modifications and may have overestimated Gal-Hyl numbers.

We found the COL1 domain of human recombinant pro-α1(V) chains produced in 293-HEK cells to contain 30 hydroxylysines: 22 as Glc-Gal-Hyl residues, 1 as Gal-Hyl, and 1

as both Glc-Gal-Hyl and Gal-Hyl. These hydroxylated Lys residues are a subset of those found in bovine α1(V), except for Hyl, which is not hydroxylated in the bovine chain (Figs. 1 and 3).[150] Thus, 29 hydroxylated Lys residues found in α1(V) COL1 domains from two very different sources may be relatively invariant in α1(V) chains. The reduced numbers of Hyl and Glc-Gal-Hyl residues in the human recombinant protein suggests that 293-HEK cells, which produce little or no endogenous extracellular matrix proteins, , may possess reduced levels of the multifunctional enzyme lysyl hydroxylase 3, which has lysyl hydroxylase activity, and collagen galactosyltransferase and glycosyltransferase activities that allow it to glycosylate Hyl residues hydroxylated by itself and by other lysl hydroxylase isoforms.[37e,14] Interestingly, COL1 residue 84 was detected here as both a Glc-Gal-Hyl and a Gal-Hyl residue, showing glycosylation at this site to be dynamic, with partial occupancy by both Glc-Gal and Gal saccharides, and 100% occupancy by neither. Our direct localization of 30 COL1 Hyl residues within the human recombinant pro-α1(V) supersedes a previous amino acid analysis estimate of 6 COL1 Hyl residues for human recombinant pro-α1(V) chains produced in 293-HEK cells, and is more congruent with the observed efficient secretion of recombinant pro-$\alpha1(V)_3$ homotrimers from 293-HEK cells, as threshold levels of glycosylated Hyl residues are necessary for efficient secretion of at least some collagenous molecules.[37e, 23, 37e, 14]

Previous amino acid analysis of α1(V) chains from human placenta or human skin estimated 3-Hyp content at four or ten 3-Hyp residues, respectively, per 1000 amino acids, levels higher than those detected by similar analyses of the major fibrillar collagen chains.[43,11b,44] The 9 Hyp residues mapped here to the X positions of Gly-Hyp-Hyp repeats in the COL1 domain of human recombinant pro-α1(V) chains falls within this range.[11b, 43]

However, the 21 X position Hyp residues mapped here within the bovine placenta α1(V) chain lies beyond this range and is instead reminiscent of the range of ten to twenty 3-Hyp residues previously estimated by amino acid analysis to lie within the collagen IV chains of some tissues.[45]

Using low-resolution MS^2 and assignments based on known collagen motifs (see above), Eyre and colleagues recently mapped three X position Hyp residues in α1(V) chains from human bone, to positions 434 and 665, and to a site variously identified by them as 692 or 695 of the COL1 domain.[19c] They also predicted the tissue-specific occurrence of 3-Hyp residues within Gly-Pro-Pro repeats at the C-termini of the COL1 domains of α1(V) and other fibrillar collagen chains, based on mapping of such residues to Gly-Pro-Pro repeats at the COL1 COOH-termini of α1(I) and α2(I) chains from tendon, but not from skin or bone.[19a] Here, we confirm the presence of X position Hyp residues at sites 434, 665, and 695; and in all three COL1 COOH-terminal Gly-Pro-Pro repeats (sites 1004, 1007, and 1010), in bovine placenta α1(V) chains (supplemental spectra 1); whereas human recombinant pro-α1(V) chains contained an X position Hyp residue only at site 434. Thus, we have confirmed the positions previously identified or predicted for six α1(V) COL1 X position Hyp residues, and found one of these (residue 434) to be, thus far, invariable. In addition, we identified the positions of 15 additional X position Hyp residues in bovine placental α1(V) chains. The 9 human recombinant pro-α1(V) COL1 X position Hyp sites are also found in the bovine α1(V) chain and provide confirmatory evidence for Hyp occupancy of these 9 X position sites. Interestingly, COL1 X position Hyp sites in human recombinant pro-α1(V) chains are limited to the NH_2-terminal half of the domain. Although the reason for this is unknown, it may relate to the COOH- to NH_2-terminus direction of triple helix formation, and to ability of

prolyl 3-hydroxylases to modify unfolded, but not triple helical procollagen chains. Thus, since the recombinant pro-α1(V) chains are produced as pro-$\alpha1(V)_3$ homotrimers, which form triple helices with apparently increased kinetics than do pro-$\alpha1(V)_2$pro-α2(V) heterotrimers, prolyl 3-hydroxylases may be particularly limited in ability to hydroxylate residues in the COOH-terminal portions of these recombinant chains. [46] In fact, the NH_2-terminal COL1 distribution of X position Hyp residues in recombinant pro-$\alpha1(V)_3$ homotrimers is consistent with the suggestion that 3-Hyp content may be related to the speed of triple helix formation, but may be inconsistent with the suggestion that 3-hydroxylation of fibrillar collagens begins at the COOH-terminus and diminishes in the more NH_2-terminal portion of the COL1 domain. [36a,19b] Other PTMs were not limited to the NH_2-terminal portion of recombinant pro-α1(V) COL1 domains, perhaps reflecting differences in the kinetics of prolyl 3-, prolyl 4-, and lysyl hydroxylases.

We report here the unexpected finding of X position Hyp residues in the context of a Gly-X-Val and a Gly-X-Ala triplet in the bovine placenta α1(V) COL1 domain. It is unclear at this time whether these Hyp residues might be 3-Hyp or 4-Hyp or whether their appearance represents a lapse in fidelity of at least one of these enzymes, or whether such sites represent previously unknown substrates at which hydroxylation serves specific functions. Supportive of our findings, Pro hydroxylation in an Gly-Pro-Ala motif has previously been reported for bovine-derived collagen I.[47] It should be noted that although X position 4-Hyp residues are thought to destabilize triple helices when in the context of Gly-Hyp-Pro triplets, X position 4-Hyp residues do not necessarily destabilize the triple helix when in the context of triplets that lack Pro or Hyp in the Y position.[48, 49] Thus, the significance of the unexpected finding of X position Hyp in these non-canonical sites remains to be determined.

Numbers of Hyl, Glc-Gal-Hyl and Y position Hyp residues mapped here by MS^n in the bovine placenta α1(V) chain are consistent with numbers of 4-Hyp, Hyl, and Glc-Gal-Hyl residues in estimates of the amino acid compositions of hydrolyzed α1(V) chains from various tissues. In contrast, the number of X position Hyp, and likely 3-Hyp, residues detected here by tandem MS in Gly-Hyp-Hyp triplets in bovine placenta α1(V) chains is markedly higher than numbers of 3-Hyp residues previously estimated by amino acid analyses. Importantly, the numbers of PTMs predicted by early amino acid analyses were based on a presumption of 100 % occupancy at a fixed number of sites. Thus, although congruence in numbers of PTMs predicted by previous amino acid analyses and the MS^n localization results presented here suggests that this presumption holds true for most α1(V) PTMs, this does not appear to be the case for 3-Hyp residues. Rather, the large difference between the numbers of 3-Hyp predicted by amino acid analyses and the number of X position Hyp residues in Gly-Hyp-Hyp triplets mapped here by high resolution MS is best explained by the concept of a relatively large number of 3-Hyp sites that have less than 100 % occupancy in a population of α1(V) chains from a given tissue.

It is not surprising that with the improved sensitivity and specificity afforded by our mass spectrometry-driven workflow that we have identified numerous PTMs previously undiscovered, including those present at relatively low stoichiometric abundances. Interestingly, 3-Hyp residues point away from the triple helix, implying roles in protein-protein interactions as the bases of biological function for these residues. [50] The presence of numbers of sites that can be differentially modified by P3Hs in a given tissue suggests a previously unknown mechanism for dynamic modification of the functions and intermolecular interactions of α1(V) chains and col(V), and perhaps other collagen types as well.

CONCLUSIONS

The utilization of high resolution and high mass accuracy mass spectrometers have resulted in comprehensively mapping hydroxylated residues in bovine placental α1(V) and human recombinant pro-α1(V) procollagen chains. In total, 130 hydroxylated prolines and 37 hydroxylated lysines were found in bovine placental α1(V), while 110 hydroxylated prolines and 30 hydroxylated lysines were identified in human recombinant pro-α1(V) procollagen chains. The unexpectedly large number of X position hydroxyproline residues found here suggests a previously unknown mechanism for the differential modification of α1(V) chain and type V collagen properties.

Supplemental Information Available Upon Request

REFERENCES

1. Fichard, A.; Kleman, J. P.; Ruggiero, F., Another look at collagen V and XI molecules. *Matrix Biol* **1995,** *14* (7), 515-31.

2. Imamura, Y.; Scott, I. C.; Greenspan, D. S., The pro-alpha3(V) collagen chain. Complete primary structure, expression domains in adult and developing tissues, and comparison to the structures and expression domains of the other types V and XI procollagen chains. *J Biol Chem* **2000,** *275* (12), 8749-59.

3. Morris, N. P.; Bachinger, H. P., Type XI collagen is a heterotrimer with the composition (1 alpha, 2 alpha, 3 alpha) retaining non-triple-helical domains. *J Biol Chem* **1987,** *262* (23), 11345-50.

4. Wu, J. J.; Weis, M. A.; Kim, L. S.; Carter, B. G.; Eyre, D. R., Differences in chain usage and cross-linking specificities of cartilage type V/XI collagen isoforms with age and tissue. *J Biol Chem* **2009,** *284* (9), 5539-45.

5. (a) Birk, D. E.; Fitch, J. M.; Babiarz, J. P.; Linsenmayer, T. F., Collagen type I and type V are present in the same fibril in the avian corneal stroma. *J Cell Biol* **1988,** *106* (3), 999-1008; (b) Birk, D. E.; Fitch, J. M.; Babiarz, J. P.; Doane, K. J.; Linsenmayer, T. F., Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. *J Cell Sci* **1990,** *95 (Pt 4)*, 649-57.

6. (a) Toriello, H. V.; Glover, T. W.; Takahara, K.; Byers, P. H.; Miller, D. E.; Higgins, J. V.; Greenspan, D. S., A translocation interrupts the COL5A1 gene in a patient with Ehlers-Danlos syndrome and hypomelanosis of Ito. *Nat Genet* **1996,** *13* (3), 361-5; (b) Richards, A. J.; Martin, S.; Nicholls, A. C.; Harrison, J. B.; Pope, F. M.; Burrows, N. P., A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type II. *J Med Genet* **1998,** *35* (10), 846-8.

7. Burlingham, W. J.; Love, R. B.; Jankowska-Gan, E.; Haynes, L. D.; Xu, Q.; Bobadilla, J. L.; Meyer, K. C.; Hayney, M. S.; Braun, R. K.; Greenspan, D. S.; Gopalakrishnan, B.; Cai, J.; Brand, D. D.; Yoshida, S.; Cummings, O. W.; Wilkes, D. S., IL-17-dependent cellular immunity to collagen type V predisposes to obliterative bronchiolitis in human lung transplants. *J Clin Invest* **2007,** *117* (11), 3498-506.

8. Sumpter, T. L.; Wilkes, D. S., Role of autoimmunity in organ allograft rejection: a focus on immunity to type V collagen in the pathogenesis of lung transplant rejection. *Am J Physiol Lung Cell Mol Physiol* **2004,** *286* (6), L1129-39.

9. (a) Iwata, T.; Philipovskiy, A.; Fisher, A. J.; Presson, R. G., Jr.; Chiyo, M.; Lee, J.; Mickler, E.; Smith, G. N.; Petrache, I.; Brand, D. B.; Burlingham, W. J.; Gopalakrishnan, B.; Greenspan, D. S.; Christie, J. D.; Wilkes, D. S., Anti-type V collagen humoral immunity in lung transplant primary graft dysfunction. *J Immunol* **2008,** *181* (8), 5738-47; (b) Bobadilla, J. L.; Love, R. B.; Jankowska-Gan, E.; Xu, Q.; Haynes, L. D.; Braun, R. K.; Hayney, M. S.; Munoz del Rio, A.; Meyer, K.; Greenspan, D. S.; Torrealba, J.; Heidler, K. M.; Cummings, O.

W.; Iwata, T.; Brand, D.; Presson, R.; Burlingham, W. J.; Wilkes, D. S., Th-17, monokines, collagen type V, and primary graft dysfunction in lung transplantation. *Am J Respir Crit Care Med* **2008,** *177* (6), 660-8.

10. Dart, M. L.; Jankowska-Gan, E.; Huang, G.; Roenneburg, D. A.; Keller, M. R.; Torrealba, J. R.; Rhoads, A.; Kim, B.; Bobadilla, J. L.; Haynes, L. D.; Wilkes, D. S.; Burlingham, W. J.; Greenspan, D. S., Interleukin-17-dependent autoimmunity to collagen type V in atherosclerosis. *Circ Res* **2010,** *107* (9), 1106-16.

11. (a) Henkel, W.; Dreisewerd, K., Cyanogen bromide peptides of the fibrillar collagens I, III, and V and their mass spectrometric characterization: detection of linear peptides, peptide glycosylation, and cross-linking peptides involved in formation of homo- and heterotypic fibrils. *J Proteome Res* **2007,** *6* (11), 4269-89; (b) Chung, E.; Rhodes, K.; Miller, E. J., Isolation of three collagenous components of probable basement membrane origin from several tissues. *Biochem Biophys Res Commun* **1976,** *71* (4), 1167-74.

12. Shoulders, M. D.; Raines, R. T., Collagen structure and stability. *Annu Rev Biochem* **2009,** *78*, 929-58.

13. Kivirikko, K. I., Collagens and their abnormalities in a wide spectrum of diseases. *Ann Med* **1993,** *25* (2), 113-26.

14. Sipila, L.; Ruotsalainen, H.; Sormunen, R.; Baker, N. L.; Lamande, S. R.; Vapola, M.; Wang, C.; Sado, Y.; Aszodi, A.; Myllyla, R., Secretion and assembly of type IV and VI collagens depend on glycosylation of hydroxylysines. *J Biol Chem* **2007,** *282* (46), 33381-8.

15. Mizuno, K.; Peyton, D. H.; Hayashi, T.; Engel, J.; Bachinger, H. P., Effect of the -Gly-3(S)-hydroxyprolyl-4(R)-hydroxyprolyl- tripeptide unit on the stability of collagen model peptides. *FEBS J* **2008,** *275* (23), 5830-40.

16. Jenkins, C. L.; Bretscher, L. E.; Guzei, I. A.; Raines, R. T., Effect of 3-hydroxyproline residues on collagen stability. *J Am Chem Soc* **2003,** *125* (21), 6422-7.

17. Morello, R.; Bertin, T. K.; Chen, Y.; Hicks, J.; Tonachini, L.; Monticone, M.; Castagnola, P.; Rauch, F.; Glorieux, F. H.; Vranka, J.; Bachinger, H. P.; Pace, J. M.; Schwarze, U.; Byers, P. H.; Weis, M.; Fernandes, R. J.; Eyre, D. R.; Yao, Z.; Boyce, B. F.; Lee, B., CRTAP is required for prolyl 3- hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell* **2006,** *127* (2), 291-304.

18. (a) Arbogast, B. W.; Gunson, D. E.; Kefalides, N. A., The role of hydroxylation of proline in the antigenicity of basement membrane collagen. *J Immunol* **1976,** *117* (6), 2181-4; (b) Brand, D. D.; Myers, L. K.; Terato, K.; Whittington, K. B.; Stuart, J. M.; Kang, A. H.; Rosloniec, E. F., Characterization of the T cell determinants in the induction of autoimmune arthritis by bovine alpha 1(II)-CB11 in H-2q mice. *J Immunol* **1994,** *152* (6), 3088-97; (c) Corthay, A.; Backlund, J.; Broddefalk, J.; Michaelsson, E.; Goldschmidt, T. J.; Kihlberg, J.; Holmdahl, R., Epitope glycosylation plays a critical role for T cell recognition of type II collagen in collagen-induced arthritis. *Eur J Immunol* **1998,** *28* (8), 2580-90; (d) Myers, L. K.;

Miyahara, H.; Terato, K.; Seyer, J. M.; Stuart, J. M.; Kang, A. H., Collagen-induced arthritis in B10.RIII mice (H-2r): identification of an arthritogenic T-cell determinant. *Immunology* **1995,** *84* (4), 509-13; (e) Myers, L. K.; Myllyharju, J.; Nokelainen, M.; Brand, D. D.; Cremer, M. A.; Stuart, J. M.; Bodo, M.; Kivirikko, K. I.; Kang, A. H., Relevance of posttranslational modifications for the arthritogenicity of type II collagen. *J Immunol* **2004,** *172* (5), 2970-5; (f) Van den Steen, P. E.; Proost, P.; Brand, D. D.; Kang, A. H.; Van Damme, J.; Opdenakker, G., Generation of glycosylated remnant epitopes from human collagen type II by gelatinase B. *Biochemistry* **2004,** *43* (33), 10809-16; (g) Van den Steen, P. E.; Proost, P.; Grillet, B.; Brand, D. D.; Kang, A. H.; Van Damme, J.; Opdenakker, G., Cleavage of denatured natural collagen type II by neutrophil gelatinase B reveals enzyme specificity, post-translational modifications in the substrate, and the formation of remnant epitopes in rheumatoid arthritis. *FASEB J* **2002,** *16* (3), 379-89; (h) Michaelsson, E.; Malmstrom, V.; Reis, S.; Engstrom, A.; Burkhardt, H.; Holmdahl, R., T cell recognition of carbohydrates on type II collagen. *J Exp Med* **1994,** *180* (2), 745-9.

19. (a) Eyre, D. R.; Weis, M.; Hudson, D. M.; Wu, J. J.; Kim, L., A novel 3-hydroxyproline (3Hyp)-rich motif marks the triple-helical C terminus of tendon type I collagen. *J Biol Chem* **2011,** *286* (10), 7732-6; (b) Fernandes, R. J.; Farnand, A. W.; Traeger, G. R.; Weis, M. A.; Eyre, D. R., A role for prolyl 3-hydroxylase 2 in post-translational modification of fibril-forming collagens. *J Biol Chem* **2011,** *286* (35), 30662-9; (c) Weis, M. A.; Hudson, D. M.; Kim, L.; Scott, M.; Wu, J. J.; Eyre, D. R., Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly. *J Biol Chem* **2010,** *285* (4), 2580-90.

20. (a) Vaisar, T.; Urban, J., Probing the proline effect in CID of protonated peptides. *J Mass Spectrom* **1996,** *31* (10), 1185-7; (b) Huang, Y.; Triscari, J. M.; Pasa-Tolic, L.; Anderson, G. A.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H., Dissociation behavior of doubly-charged tryptic peptides: correlation of gas-phase cleavage abundance with ramachandran plots. *J Am Chem Soc* **2004,** *126* (10), 3034-5.

21. (a) Loo, J. A.; Edmonds, C. G.; Smith, R. D., Tandem mass spectrometry of very large molecules. 2. Dissociation of multiply charged proline-containing proteins from electrospray ionization. *Anal Chem* **1993,** *65* (4), 425-38; (b) Tang, X. J.; Thibault, P.; Boyd, R. K., Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. *Anal Chem* **1993,** *65* (20), 2824-34; (c) Breci, L. A.; Tabb, D. L.; Yates, J. R., 3rd; Wysocki, V. H., Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Anal Chem* **2003,** *75* (9), 1963-71.

22. (a) McAlister, G. C.; Phanstiel, D.; Wenger, C. D.; Lee, M. V.; Coon, J. J., Analysis of tandem mass spectra by FTMS for improved large-scale proteomics with superior protein quantification. *Anal Chem* **2010,** *82* (1), 316-22; (b) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **2004,** *101* (26), 9528-33.

23. Imamura, Y.; Steiglitz, B. M.; Greenspan, D. S., Bone morphogenetic protein-1 processes the NH2-terminal propeptide, and a furin-like proprotein convertase processes the COOH-terminal propeptide of pro-alpha1(V) collagen. *J Biol Chem* **1998,** *273* (42), 27511-7.

24. Mares, D. C.; Heidler, K. M.; Smith, G. N.; Cummings, O. W.; Harris, E. R.; Foresman, B.; Wilkes, D. S., Type V collagen modulates alloantigen-induced pathology and immunology in the lung. *Am J Respir Cell Mol Biol* **2000,** *23* (1), 62-70.

25. Seyer, J. M.; Kang, A. H., Covalent structure of collagen: amino acid sequence of three cyanogen bromide-derived peptides from human alpha 1(V) collagen chain. *Arch Biochem Biophys* **1989,** *271* (1), 120-9.

26. Shah, R.; Smith, P.; Purdie, C.; Quinlan, P.; Baker, L.; Aman, P.; Thompson, A. M.; Crook, T., The prolyl 3-hydroxylases P3H2 and P3H3 are novel targets for epigenetic silencing in breast cancer. *Br J Cancer* **2009,** *100* (10), 1687-96.

27. Greenspan, D. S.; Cheng, W.; Hoffman, G. G., The pro-alpha 1(V) collagen chain. Complete primary structure, distribution of expression, and comparison with the pro-alpha 1(XI) collagen chain. *J Biol Chem* **1991,** *266* (36), 24727-33.

28. Swaney, D. L.; Wenger, C. D.; Coon, J. J., Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* **2010,** *9* (3), 1323-9.

29. Ficarro, S. B.; Zhang, Y.; Lu, Y.; Moghimi, A. R.; Askenazi, M.; Hyatt, E.; Smith, E. D.; Boyer, L.; Schlaeger, T. M.; Luckey, C. J.; Marto, J. A., Improved electrospray ionization efficiency compensates for diminished chromatographic resolution and enables proteomics analysis of tyrosine signaling in embryonic stem cells. *Anal Chem* **2009,** *81* (9), 3440-7.

30. (a) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S., A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* **2009,** *8* (12), 2759-69; (b) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J., Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* **2007,** *6* (11), 1942-51.

31. Wenger, C. D.; Phanstiel, D. H.; Lee, M. V.; Bailey, D. J.; Coon, J. J., COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* **2011,** *11* (6), 1064-74.

32. (a) Good, D. M.; Wenger, C. D.; McAlister, G. C.; Bai, D. L.; Hunt, D. F.; Coon, J. J., Post-acquisition ETD spectral processing for increased peptide identifications. *J Am Soc Mass Spectrom* **2009,** *20* (8), 1435-40; (b) Good, D. M.; Wenger, C. D.; Coon, J. J., The effect of interfering ions on search algorithm performance for electron-transfer dissociation data. *Proteomics* **2010,** *10* (1), 164-7.

33. Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H., Open mass spectrometry search algorithm. *J Proteome Res* **2004,** *3* (5), 958-64.

34. Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **2007,** *4* (3), 207-14.

35. Gorres, K. L.; Raines, R. T., Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol* **2010,** *45* (2), 106-24.

36. (a) Risteli, J.; Tryggvason, K.; Kivirikko, K. I., Prolyl 3-hydroxylase: partial characterization of the enzyme from rat kidney cortex. *Eur J Biochem* **1977,** *73* (2), 485-92; (b) Tiainen, P.; Pasanen, A.; Sormunen, R.; Myllyharju, J., Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, an enzyme modifying the basement membrane collagen IV. *J Biol Chem* **2008,** *283* (28), 19432-9.

37. (a) Unsold, C.; Pappano, W. N.; Imamura, Y.; Steiglitz, B. M.; Greenspan, D. S., Biosynthetic processing of the pro-alpha 1(V)2pro-alpha 2(V) collagen heterotrimer by bone morphogenetic protein-1 and furin-like proprotein convertases. *J Biol Chem* **2002,** *277* (7), 5596-602; (b) Gopalakrishnan, B.; Wang, W. M.; Greenspan, D. S., Biosynthetic processing of the Pro-alpha1(V)Pro-alpha2(V)Pro-alpha3(V) procollagen heterotrimer. *J Biol Chem* **2004,** *279* (29), 30904-12; (c) Delacoux, F.; Fichard, A.; Geourjon, C.; Garrone, R.; Ruggiero, F., Molecular features of the collagen V heparin binding site. *J Biol Chem* **1998,** *273* (24), 15069-76; (d) Chanut-Delalande, H.; Fichard, A.; Bernocco, S.; Garrone, R.; Hulmes, D. J.; Ruggiero, F., Control of heterotypic fibril formation by collagen V is determined by chain stoichiometry. *J Biol Chem* **2001,** *276* (26), 24352-9; (e) Fichard, A.; Tillet, E.; Delacoux, F.; Garrone, R.; Ruggiero, F., Human recombinant alpha1(V) collagen chain. Homotrimeric assembly and subsequent processing. *J Biol Chem* **1997,** *272* (48), 30083-7; (f) Kessler, E.; Fichard, A.; Chanut-Delalande, H.; Brusel, M.; Ruggiero, F., Bone morphogenetic protein-1 (BMP-1) mediates C-terminal processing of procollagen V homotrimer. *J Biol Chem* **2001,** *276* (29), 27051-7.

38. Pyott, S. M.; Schwarze, U.; Christiansen, H. E.; Pepin, M. G.; Leistritz, D. F.; Dineen, R.; Harris, C.; Burton, B. K.; Angle, B.; Kim, K.; Sussman, M. D.; Weis, M.; Eyre, D. R.; Russell, D. W.; McCarthy, K. J.; Steiner, R. D.; Byers, P. H., Mutations in PPIB (cyclophilin B) delay type I procollagen chain association and result in perinatal lethal to moderate osteogenesis imperfecta phenotypes. *Hum Mol Genet* **2011,** *20* (8), 1595-609.

39. (a) Olsen, J. V.; Ong, S. E.; Mann, M., Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **2004,** *3* (6), 608-14; (b) Rodriguez, J.; Gupta, N.; Smith, R. D.; Pevzner, P. A., Does trypsin cut before proline? *J Proteome Res* **2008,** *7* (1), 300-5.

40. Licklider, L. J.; Thoreen, C. C.; Peng, J.; Gygi, S. P., Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal Chem* **2002,** *74* (13), 3076-83.

41. Wiesner, J.; Premsler, T.; Sickmann, A., Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* **2008,** *8* (21), 4466-83.

42. Savitski, M. M.; Lemeer, S.; Boesche, M.; Lang, M.; Mathieson, T.; Bantscheff, M.; Kuster, B., Confident phosphorylation site localization using the Mascot Delta Score. *Molecular & cellular proteomics : MCP* **2011,** *10* (2), M110 003830.

43. Rhodes, R. K.; Miller, E. J., The isolation and characterization of the cyanogen bromide peptides from the B chain of human collagen. *J Biol Chem* **1979,** *254* (23), 12084-7.

44. (a) Butler, W. T.; Piez, K. A.; Bornstein, P., Isolation and characterization of the cyanogen bromide peptides from the alpha-1 chain of rat skin collagen. *Biochemistry* **1967,** *6* (12), 3771-80; (b) Click, E. M.; Bornstein, P., Isolation and characterization of the cyanogen bromide peptides from the alpha 1 and alpha 2 chains of human skin collagen. *Biochemistry* **1970,** *9* (24), 4699-706; (c) Miller, E. J.; Lunde, L. G., Isolation and characterization of the cyanogen bromide peptides from the alpha 1(II) chain of bovine and human cartilage collagen. *Biochemistry* **1973,** *12* (17), 3153-9; (d) Seyer, J. M.; Kang, A. H., Covalent structure of collagen: amino acid sequence of alpha 1(III)-CB9 from type III collagen of human liver. *Biochemistry* **1981,** *20* (9), 2621-7.

45. (a) Kefalides, N. A., Basement membranes: structural and biosynthetic considerations. *J Invest Dermatol* **1975,** *65* (1), 85-92; (b) Gryder, R. M.; Lamon, M.; Adams, E., Sequence position of 3-hydroxyproline in basement membrane collagen. Isolation of glycyl-3-hydroxyprolyl-4-hydroxyproline from swine kidney. *J Biol Chem* **1975,** *250* (7), 2470-4.

46. Roulet, M.; Valkkila, M.; Chanut-Delalande, H.; Hamalainen, E. R.; Kessler, E.; Ala-Kokko, L.; Mannikko, M.; Bonod-Bidaud, C.; Ruggiero, F., The collagen V homotrimer [alpha1(V)](3) production is unexpectedly favored over the heterotrimer [alpha1(V)](2)alpha2(V) in recombinant expression systems. *J Biomed Biotechnol* **2010,** *2010*, 376927.

47. zhang, Y., Liu, T., Wang, Q., Chen, L., Lei, J., Luo, J., Ma, G., and Su, Z., Mass spectrometric detection of marker peptides in tryptic digests of gelatin: A new method to differentiate between bovine and porcine gelatin. *Food Hydrocolloids* **2009,** *23*, 2001-2007.

48. Inouye, K.; Kobayashi, Y.; Kyogoku, Y.; Kishida, Y.; Sakakibara, S.; Prockop, D. J., Synthesis and physical properties of (hydroxyproline-proline-glycine)10: hydroxyproline in the X-position decreases the melting temperature of the collagen triple helix. *Arch Biochem Biophys* **1982,** *219* (1), 198-203.

49. Bann, J. G.; Bachinger, H. P., Glycosylation/Hydroxylation-induced stabilization of the collagen triple helix. 4-trans-hydroxyproline in the Xaa position can stabilize the triple helix. *J Biol Chem* **2000,** *275* (32), 24466-9.

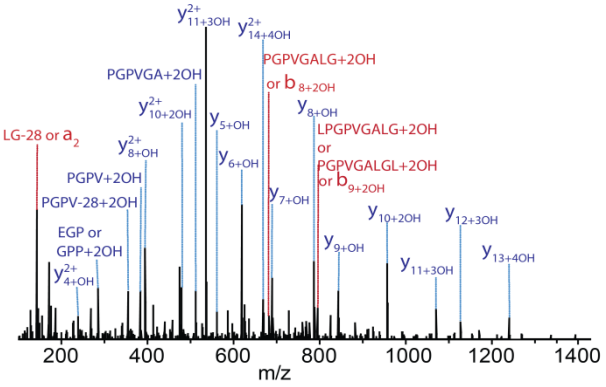
50. Schumacher, M. A.; Mizuno, K.; Bachinger, H. P., The crystal structure of a collagen-like polypeptide with 3(S)-hydroxyproline residues in the Xaa position forms a standard 7/2 collagen triple helix. *J Biol Chem* **2006,** *281* (37), 27566-74.

Figure 6.1 Hydroxylated amino acid residues and saccharide attachments of the bovine placenta α1(V) collagen chain. Sequence coverage was 94 %, with identification and mapping of 106 Y-position Hyp, 22 Gly-Hyp-Hyp X-position Hyp, 1 Gly-Hyp-Ala, 1 Gly-Hyp-Val, 3 Hyl, and 34 Glc-Gal-Hyl residues. Red, hydroxylated, non-glycosylated residues; green, Glc-Gal-Hyl residues; Underlined, hydroxylated/ glycosylated residues mapped in previous studies.[4, 19c] Sequences not identified in the course of mass spectrometry analysis are in light blue. The seven

COL1 amino acid residues that differ between human and bovine are orange. *, 13 Y-position Pro residues hydroxylated in human or bovine α1(V) COL1 domain, but not the other. ^, Lys residue hydroxylated in human recombinant pro-α1(V), but not in bovine placenta α1(V). #, Hyl residues identified by Wu et al. as being involved in interchain covalent cross-links.[4]

A)

7OH+Gal.Glc



	Theoretical m/z	Experimental m/z	Mass error (ppm)
$[M+4H]^{4+}$+7OH+Gal.Glc	775.3717	775.3749	4.1
LG-28 OR a2	143.1179	143.1181	1.4
y_4^{2+}+OH	237.619	237.6194	1.7
EGP or GPP+2OH	284.1241	284.1244	1.1
PGPV -28+2OH	355.1976	355.1976	0.0
PGPV+2OH	383.1925	383.1924	-0.3
y_8^{2+}+OH	393.6907	393.6912	1.3
y_{10}^{2+} +2OH	478.7252	478.7258	1.3
PGPVGA+2OH	511.2511	511.2491	-3.9
y_{11}^{2+} +3OH	535.2491	535.2492	0.2
y_5 +OH	561.2627	561.262	-1.2
y_6 +OH	618.2842	618.284	-0.3
y_{14}^{2+} +4OH	668.81	668.8109	1.3
b_8+2OH or PGPVGALG+2OH	681.3566	681.3574	1.1
y_7+OH	689.3213	689.3212	-0.1
y_8+OH	786.3741	786.3737	-0.5
b_9+2OH or LPGPVGALG+2OH or PGPVGALGL+2OH	794.4407	794.444	4.2
y_9 +OH	843.3955	843.3959	0.5
y_{10}+2OH	956.4432	956.4438	0.6
y_{11} +3OH	1069.491	1069.492	1.4
y_{12} +3OH	1126.512	1126.513	0.8
y_{13} +4OH	1239.56	1239.562	1.5

B)

3OH

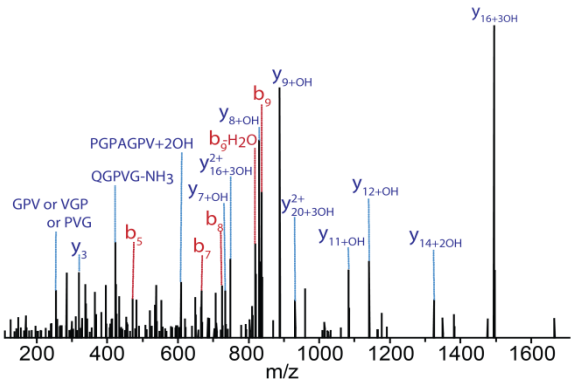


Figure 6.2 X position Hyp resdiues in Gly-X-Val and Gly-X-Ala triplets. Peptides were prepared by trypsin digestion of bovine placenta α1 (V) collagen chains. Mass error from the expected product ion monoisotopic mass is typically less than 5ppm. y_{14}+4OH fragment ion and internal fragment ions (PGPV-28+2OH, PGPV+2OH, and PGPVGA+2OH) confirm that one Hyp is in the Gly-X-Val triplet in the left hand spectrum. y_{11}+OH, y_{12}+OH and y_{14}+2OH fragment ions confirm that one Hyp is in the Gly-X-Ala triplet in the right hand spectrum. OH, one hydroxylation.

Figure 6.3 Hydroxylated amino acid residues and glycosylated Hyl residues of recombinant human pro-α1(V) collagen chains produced in 293-EBNA human

embryonic kidney cells. Sequence coverage was 90% of the entire pro-α1(V) chain and 96 % of the COL1 domain. Within the COL1 domain 98 Y-position Hyp, 9 Gly-Hyp-Hyp X-position Hyp, 1 Gly-Hyp-Val, 1 Gly-Hyp-Ala, 1 Gly-Hyp-Thr, 7-Hyl, and 23 Glc-Gal-Hyl residues were identified. Red, hydroxylated, non-glycosylated residues; green, Glc-Gal-Hyl residues; dark blue, Gal-Hyl residue; purple, residue found as both Glc-Gal-Hyl and Gal-Hyl. Underlined, hydroxylated/glycosylated residues mapped in previous studies.[4, 19c] Sequences not identified in the course of mass spectrometry analysis are in light blue. The seven COL1 amino acid residues that differ between human and bovine are orange. *, 13 Y-position Pro residues hydroxylated in human or bovine α1(V) COL1 domain, but not the other. #, Hyl residues identified by Wu et al. as being involved in interchain covalent cross-links. [4] ^, Lys residues hydroxylated in bovine placenta α1(V), but not in human recombinant pro-α1(V). A vertical arrow marks the site of proteolytic removal of the signal peptide, as previously determined by Edman degradation NH_2-terminal amino acid sequencing, and as confirmed by MS analysis in the present study.[23] Brackets indicate limits of the small COL2 hypothetical triple helical domain, in which interruptions of Gly-X-Y repeats are underlined.

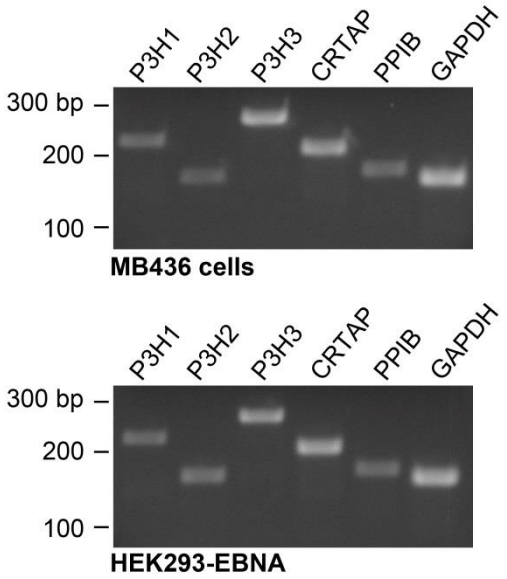


Figure 6.4 Expression of P3H1, P3H2, P3H3 and prolyl 3-hydroxylation complex components CRTAP and PPIB in 293-HEK cells. Levels of P3H1, P3H2, P2H3, CRTAP and GAPDH expression were ascertained by RT-PCR analysis of RNA from MB436 cells (top panel), a cell line previously shown to be positive for expression of all three P3Hs, and from 293-HEK cells (bottom panel). Positions of size markers, given in base pairs (bp), are shown to the left of each panel.[26]

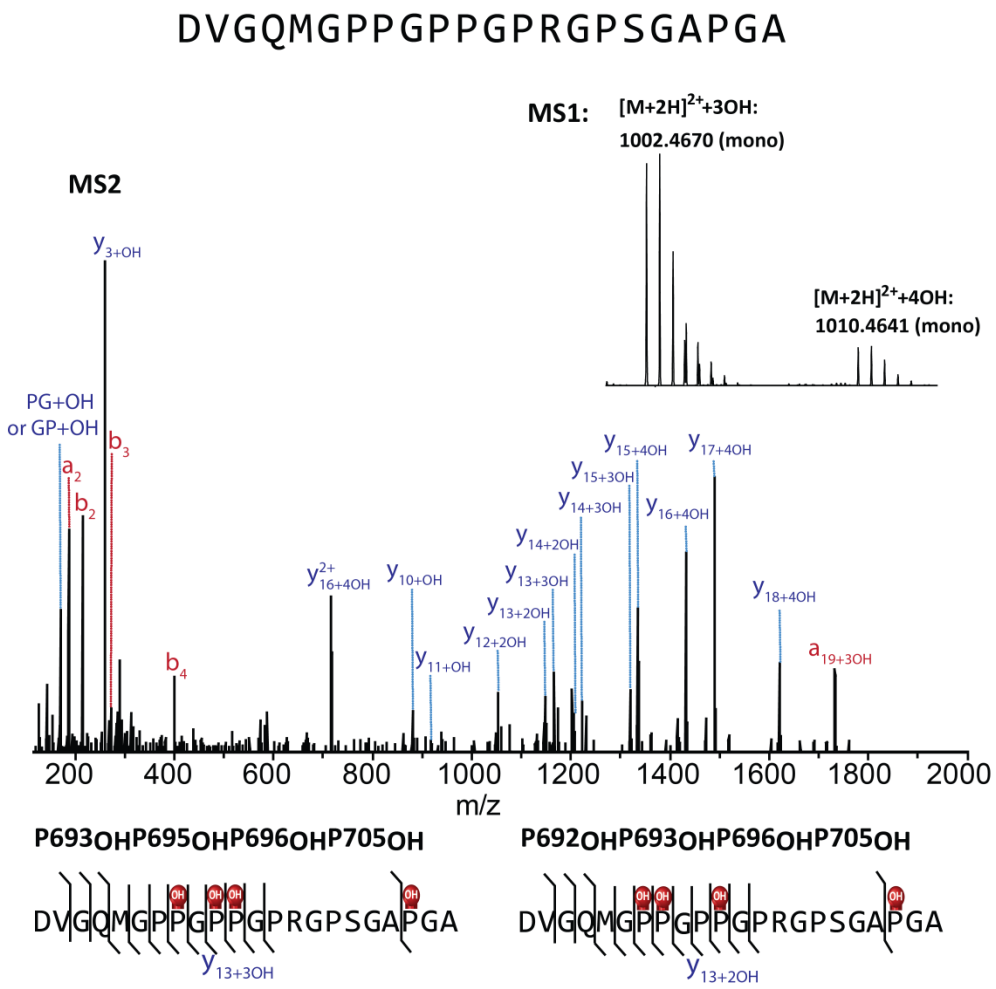


Figure 6.5 A chimeric MS2 spectrum revealing a pair of isomeric differently modified α1(V) chain peptides that co-eluted and co-fragmented. Peptides were generated by AspN cleavage of bovine placenta α1(V) chains. MS^1 in the insert was obtained by averaging MS^1 across the elution window. MS^2 revealed that the two separate cluster peaks in MS^1 belong to peptides with identical primary sequence, but differing degrees of hydroxylation. Peptides with a total of 3 hydroxylations contain Hyp^{693}, Hyp^{696}, and Hyp^{705} (MS^2 for peptides with a total of 3 hydroxylation in the

MS^1 of the insert is not shown here). MS^2 for peptides with a total of 4 hydroxylations (shown here) reveals that the extra Hyp can be either Hyp^{692} or Hyp^{695}.

Chapter 7

Conclusions

SUMMARY

In the past five years, I have devoted all my efforts to apply MS to investigate proteomic events in various biological systems. In each project, a series of tests was firstly conducted before analyzing real samples. These tests were used to optimize each experimental step, enabling the best results obtained from each analysis. Optimizations included: determining minimum amount of starting material needed in one analysis, improving sample preparation steps, selecting separation approaches, and optimizing instrumental parameters for MS investigation. Each optimized workflow was then applied for analyses of real samples. Results presented here have provided insights into: 1) the function of MBD1 in neural development; 2) the mechanism of GATA-2 in regulating hematopoiesis; 3) the role of Smad3-dependent TGF-β signaling in the atherosclerotic process; and 4) the unique biological properties of α1(V) collagen chain to induce the autoimmunity in lung transplantation.

Revealing the Interacting Proteins of MBD1 in Mouse Brain by Mass Spectrometry

Here, the MBD1 mice, which expressed flag tag containing MBD1 proteins, were generated. The flag tag containing MBD1 proteins and their interacting proteins were isolated through immunoprecipitation (IP) from lysate of mouse brains. Control groups and quantification analysis by MS were utilized to distinguish potential interacting proteins from contaminants. An optimization for the number of mice used in each replicate was firstly conducted. Finally, proteins from six mice brains were extracted and combined for one IP experiment. Analyses of three biological replicates

resulted in a total of 795 proteins identified. In each biological replicate, six MBD1 mice and six control mice were scarified. A spectral counting based quantification approach was employed here and revealed twenty-two proteins with higher abundance in MBD1 mice samples in comparison with those in control mice. However, only five of them were identified as significant interacting proteins after a *t*-test across three biological replicates. Our results have provided a list of candidates of interacting proteins for MBD1, including Gatad2d, Smarcb1, Ruvbl2, Smarca2 and Rab10. In the future, validation experiments will be conducted to prove our findings. Eventually, our findings will facilitate the understanding of function of MBD1 in neural development.

Molecular Basis of Crosstalk between Oncogenic Ras and the Master Regulator of Hematopoiesis GATA-2

GATA-2 (T354M) is the disease mutant of GATA-2, which impairs the ability of GATA-2 to bind to DNA. After performing western plot analyses, GATA-2 (T354M) migrated into two bands, which was significantly different than that found in wild-type GATA-2. Samples were then incubated with λ-phosphatase to test whether the band shift was caused by phosphorylation on GATA-2. λ-phosphatase has shown to decrease abundance of the GATA-2 (T354M) upper band and increase wild-type GATA-2 mobility. These results indicated that wild-type GATA-2 is phosphorylated, while GATA-2 (T354M) is hyperphosphorylated. MS was then employed to characterize the phosphorylation of wild-type GATA and GATA-2

(T354M). Wild-type GATA or GATA-2 (T354M) proteins were isolated from cell lysate through IP and subjected to SDS-PAGE separation. The purified wild-type GATA-2 or GATA-2 (T354M) proteins were in-gel digested by trypsin, extracted, and analyzed by LC-MS/MS. In total, eight phosphorylation sites were found by MS. A series of mutant tests was than conducted to reveal the functional role of each phosphorylation site in wild-type GATA and GATA-2 (T354M) proteins. Phosphorylation on S192 was demonstrated as the master regulator for phosphorylation on other residues in wild-type GATA and GATA-2 (T354M) proteins.

Comparative Secretome Analysis of Vascular Smooth Muscle Cells (VSMCs) in Response to Smad3-Dependent TGF-β Signaling

In order to identify the selected proteins involving in Smad3-dependent TGF-β signaling, rat VSMCs overexpressing Smad3 or green fluorescent protein (GFP) were treated with or without TGF-β. Proteins in the conditioned media under different treatments were then isolated and concentrated separately by 10 kDa molecular weight cut-off filters. The concentrated proteins were digested by trypsin and separated on-line in a C18 column before MS analysis. In total, 665 proteins were identified and subjected to SignalP and SecretomeP analyses. Further quantification revealed that 38 secreted proteins were significantly up or down-regulated in response to TGF-β.

Comprehensive Mass Spectrometric Mapping of the Hydroxylated Amino Acid Residues of α1(V) Collagen Chain

A comprehensively mapping of the hydroxylation of human and bovine α1(V) collagen chain was performed by MS. Human and bovine α1(V) collagen chain were isolated and subjected to digestion. Because of the high content of proline in the sequence of collagen, multiple proteases, including AspN, Glu-C, Chymotrypsin, ArgC and trypsin, were used alone or sequentially to achieve about 90% sequence coverage. Beam-type CID (HCD) and MS^3 scans were utilized to precisely localize the hydroxylation sites in α1(V) collagen chain. Manually inspections of all the modified sites indicated by database search were then conducted. Comprehensive analysis revealed 110 hydroxyproline residues, 7 hydroxylysines, and 23 glycosylated hydroxyprolines in human pro-α1 (V) chain with about 90% sequence coverage. I achieved about 94% sequence coverage in bovine α 1 (V) chain documenting 130 hydroxyprolines, 3 hydroxylysines, and 34 glycosylated hydroxyprolines. Additionally, an unexpectedly large number of hydroxyprolines were founded in X-positions of Gly-X-Y triplets, which is contrary to expectations based on previous amino acid analyses.

Appendix 1 List of Publications and Presentations

PUBLICATIONS

C. Yang, Y. Guo, X. Zhao, L. Li. "Revealing the interacting proteins of MBD1 in Mouse brain by mass spectrometry" In preparation.

C. Yang,* D. Ma,* X. Shi and Li L. "Comparative Secretome Analysis of Vascular Smooth Muscle Cells in Response to Smad3-Dependent TGF-β Signaling" In preparation.

KR. Katsumura, **C. Yang**, ME. Boyer, L. Li, EH. Bresnick. " Molecular Basis of Crosstalk between Oncogenic Ras and the Master Regulator of Hematopoiesis GATA-2", *EMBO Rep*., **2014**; 15: 938-947.

C. Yang, X. Zhong, L. Li. " Recent Advances in Enrichment and Separation Strategies for Mass Spectrometry-based Phosphoproteomics", *Electrophoresis*, **2014**; 35:3418-29.

C. Yang, A. Park, N. Davis, J. Russell, B. Kim, D. Brand, M. Lawrence, Y. Ge, M. Westphall, J. Coon, D. Greenspan. "Comprehensive Mass Spectrometric Mapping of the Hydroxylated Amino Acid Residues of the α1(V) Collagen Chain" *Journal of Biological Chemistry*, **2012**;287: 40598-610.

I. Agyekum, C. Nimley, **C. Yang**, P. Sun. "Combination of Scanning Electron Micorscopy in the Characterization of a Nanometer-Sized Electrode and Current Fluctuation Observed at a Nanometer-Sized Electrode" *Journal of Physical Chemistry C*, **2010**; 114: 14970-14974.

C. Yang, P. Sun. "Fabrication and Characterization of a Dual Submicrometer-Sized Electrode" *Analytical Chemistry*, **2009**; 81: 7496-7500.

CONFERENCE PRESENTATIONS

C. Yang, Russell, J., Ledvina, A., Westphall, M., Brumbugh, J., Coon, J. "Profiling Histone H3 Isoforms in Human Embryonic Stem Cells using a Velos Orbitrap Modified with a Segmented Cell" ,the 59th American Society for Mass Spectrometry Conference on Mass Spectrometry, 2011, Denver, CO.

C.Yang, Ma, D., Shi, X., Kent, KC., Li, L. " Charaterization of the Secretome of Vascular Smooth Muscle Cells in Response to TGF-β/Smad3 by Label-free MS^E Quantification", the 61st American Society for Mass Spectrometry Conference on Mass Spectrometry, 2013, Minneapolis, MN.

C.Yang, Ma, D., Shi, X., Kent, KC., Li, L. " Comparative Proteomic Analysis of Secretome in Vascular Smooth Muscle Cells by Label-free Quantitation via Data-Independent Acquisition Mass Spectrometry", Pittcon, 2014, Chicago, IL.

C. Yang, Katsumura, KR., Bresnick, EH., Li, L. " High Performance Mass Spectrometry Revealing the Role of Phosphorylation in Regulation of GATA-2 Function", the 62nd American Society for Mass Spectrometry Conference on Mass Spectrometry, 2014, Baltimore, MD.

**equal contribution*