UNIVERSITY OF CALIFORNIA
RIVERSIDE

Chemical Proteomic Approaches Toward Understanding Cell Signaling and Arsenic Toxicity

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

Doctor of Philosophy

in

Chemistry

by

Xuejiao Dong

September 2021

Dissertation Committee:

Dr. Yinsheng Wang, Chairperson
Dr. Wenwan Zhong
Dr. Min Xue

The Dissertation of Xuejiao Dong is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

Time flies, and now I am about to enter my founding year. I spent the best six years of my life at UC Riverside and this experience will be one that I will treasure forever. Along the way, there are too many people who I want to say thanks. I could never complete my Ph.D. study without the help and support from them.

First and foremost, I would like to express my deepest gratitude to my principle investigator, Professor Yinsheng Wang, for his careful guidance and continuous encouragement for me over the past six years. He gave me a platform to give full play to my strengths and increase abilities, taught me rigorous scientific research attitude and problem-solving thinking, led me into the palace of science. His profound professional knowledge, sincere and kind attitude, rigorous academic attitude and striving for excellence have set me up as a model of outstanding scientific research workers and a model for my study, gave me a lifetime of inexhaustible and inexhaustible treasure!

I would like to give my appreciation to Professor Wenwan Zhong and Professor Min Xue, for being my dissertation committee and Professor. Michael Pirrung, Professor Zhenbiao Yang for serving as my guidance committee, providing me their helpful comments during my Ph.D. study. Professor Jikui Song, for his kind assistance and advice on molecule direct binding. Professor Chia-en Chang, for all the wise advice she gave me and passive discussion on my research. I would also like to thank Professor Joseph Genereux for his assistance on LI-COR Imaging System, Professor Dan Borchardt for the training on NMR. Professor Ryan Julian, Professor Christopher Switzer for their help and

instructions of the curricula in graduate school. Professor Kevin Simpson, Professor Rena Hayashi for their patient training and instruction when I worked as a teaching assistant.

My life can't be fulfilled without all the current and former members in the Wang group. Great thanks go to Dr. Ming Huang, Dr. Rong Cai and Dr. Weili Miao for their generous help and patient guidance when I first started work on proteomic. I would give my gratitude to Dr. Xiaochuan Liu, Dr. Xiaoxia Dai, Dr. Lin Li, Dr. Xiaomei He and Dr. Yuxiang Sun who kindly guiding me on my molecular biological experiments; Dr. Yuxiang Cui, Dr. Quanqing Zhang, Jiekai Yin, Tianyu Qi, Yenyu Yang, Dr. Gwendolyn Gonzales and Dr. Jiabin Wu for their discussion on my research and also on instrument maintenance and trouble shooting. I also like to give my profound thanks to Dr. Nathan E. Price for all the assistance he offered, especially in presentation and writing. Special thanks go to our lab manager Shuli Zhai for her effort in organizing our lab environment. I wish to extend my sincerely thanks to all other former and current group members: Dr. Pengcheng Wang, Dr. Preston Williams, Dr. Nicole Williams, Dr. Changjun You, Dr. Ji Jiang, Dr. Yang Yu, Dr. Jun Wu, Dr. Lok Ming Tam, Dr. David Bade, Dr. Jiapeng Leng. Dr. Tianlu Wang, Ying Tan, Su Guo, Dr. Hua Du, Ross Furash, Jun Yuan, Zi Gao, Dr. Feng Tan, Yinan Wang, Dr. Feng Tang, Dr. Songbo Wei, Xingyuan Chen, Garrit Clabaugh, Xin Wang, Shiyuan Guo, Dr. Andrew Kellum and Dr. Kailin Yu. I would keep appreciate all the beautiful moments we have in this laboratory.

Another special thanks to other friends I have met in UC Riverside. I'd like to thank Jianan Sun, Dr. Linfeng Gao for the technical support and research discussion; Dr. Shuyue

Lan, Mengyao Shao, Dr. Hang Zhang, Dr. Tianyi Yu, Dr. Fan Yang, Dr. Zhe Song for our friendship, it has been so lucky to have you all in the journey of my life.

At the end, I would like to deeply thank my parents, Yanli Dai and Zhengjun Dong, your understanding and selfless help have enabled me to fulfill my dream of continuing my studies and successfully complete my doctoral studies. You have always been my strongest backing and will always be the driving force for me to move forward. In addition, I say thanks to my boyfriend Dr. Liang Zhou whose accompany, and love support me through every trough and depressed moments, thank you for always standing by my side.

COPYRIGHT ACKNOWLEDGEMENTS

The text of this dissertation, in full, is a reprint of the materials as they appear in the following publications:

Chapter 2: Xuejiao, D., Linfeng, G., Jikui, S., and Yinsheng, W. "Chemical Proteomic Profiling of Lysophosphatidic Acid-Binding Proteins". Anal. Chem. 2019, 91, 24, 15365–15369

Chapter 4: Xuejiao, D., Jianan, S., Weili, M, Chia-En, C., Yinsheng, W. "Proteome-wide Characterizations of N6-Methyl-ATP- and N6-Furfuryl-ATP-Binding Capabilities of Kinases". Anal. Chem. 2021

DEDICATION

To my family,
the most precious part of my life.

I love you all.

ABSTRACT OF THE DISSERTATION

Chemical Proteomic Approaches Toward Understanding Cell Signaling and Arsenic Toxicity

by

Xuejiao Dong

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, September 2021
Dr. Yinsheng Wang, Chairperson

Recent advances in mass spectrometry (MS) instrumentation and methods have greatly facilitated protein identification and quantification in complex sample matrices. Chemical proteomic methods, which allow for selective labeling, enrichment, identification, and characterization of target proteins in the entire human proteome, have been widely employed to study functional subgroups of proteins and provided valuable resources for systematic study of biological pathways. Cell signaling pathways can be largely affected by biological small molecules through their binding to important target proteins; hence, the discovery and understanding of molecular interactions between small molecules and their protein targets are critical in drug discovery and toxicology. However, some of the observed effects of small molecule cannot be attributed to its interactions with their known binding proteins. Thus, there are likely other yet identified ligand-binding proteins. With this in mind, we focused on the discovery of novel binding proteins for several important

small molecules. In this dissertation, we developed and employed chemoproteomics methods to study the binding of small molecules and the mechanism of their related cell signaling pathways. We have utilized such techniques, to analyze the LPA-binding proteins, As(III)-binding proteins and N^6-modifed ATP binding proteins comprehensively, which represented the identification of a class of crucial signaling related binding proteins and characterized the regulation of their activity or expression from the entire human proteome.

In the first study, we described a strategy, relying on the use of a desthiobiotin-lysophosphatidic acid (LPA) probe in conjunction with LC-MS/MS analysis, to enrich and identify putative LPA-binding proteins in the human proteome. By combining this strategy with SILAC, we were able to discover 86 proteins exhibiting highly selective interactions with LPA at the entire proteome level. We further validated that, among these identified proteins, ANXA5 and PGK1 can bind directly with LPA.

In the second project, we employed SILAC-based metabolic labeling, biotin-As probe pull-down and LC-MS/MS analysis to achieve proteome-wide identification of arsenic-binding proteins in human cells and to characterize quantitatively arsenic-protein interactions. Our work revealed HSPA4 as a potential new molecular target contributing to the proteotoxic effects of arsenic exposure.

Lastly, we employed isotope-coded ATP acyl-phosphate probes, together with a multiple-reaction monitoring (MRM)-based targeted proteomic workflow, for proteome-wide identification of endogenous kinases that can bind to two N^6-modified ATP derivatives. Additionally, *in-vitro* biochemical assay showed that GSK3β could employ N^6-Me-ATP as the phosphate group donor to phosphorylate its substrate peptide.

Each study shows that MS-based chemical proteomic approaches can serve as a robust and efficient tool for the discovery of ligand-binding proteins at the global proteome level, which provide important resources for understanding LPA and kinase-mediated cell signaling as well as toxic mechanisms of action of arsenic species.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 General Overview

Substantial advances have been made in mass spectrometry (MS) in the past two decades. It becomes a vital tool in a wide range of fields to generate qualitative and quantitative information on the isotopic, elemental and molecular compositions of organic and inorganic species in the gas, liquid, or solid state.[1] Challenges encountered when MS techniques are used alone such as sample complexity and quantitation can be overcome, to some extent, by combining MS with powerful separation techniques.[2, 3] With the development of liquid chromatography-tandem mass spectrometry (LC-MS/MS), the application of MS to the identification of chemical compositions in complex biological mixtures, especially for the large-scale study of proteins has been sharply expanded due to its high specificity, accuracy and high throughput , which was referred to as MS-based proteomics.[4]

Proteins are found throughout the body and play many critical roles in function and regulation of biological pathways and influence a variety of cellular processes. Thus, the information on the identity of a protein, the amount of the protein that is present, and the modifications the protein contains can provide a better understanding of molecular mechanisms underlying life span and diseases. [5-7]

With MS-based proteomics, protein identification, protein quantification, proteins with post-translational modifications can be assessed.[8, 9] Among many different technologies used for proteomic analysis, bottom-up mass spectrometry proteomics a.k.a. shotgun

proteomics, has become the most widely used approach for characterization and quantification of proteins present in a biological samples or systems such as biological fluids and disease tissues.[9, 10] The most commonly used MS method to identify proteins is based on sequence-specific proteolytic enzyme digestion, the resulting unique amino acid sequences and their specific fragment ions are measured via tandem MS.[11, 12] (Figure 1.1) Proteomic study provides information about not only the identities of proteins that are expressed but also their expression levels in different systems.[13] Quantitative analysis of protein expression using MS can be achieved by metabolic labeling with a stable heavy isotope, such as ^{15}N, ^{13}C,[14, 15] chemical labeling such as isotope-coded affinity-tag(ICAT)-based protein profiling method or by using isotope-labeled internal standards.[16]

Post-translational modifications (PTMs) of proteins can partially determine the function and activity of a protein. Over 200 distinct types of covalent modifications have been reported, and they include phosphorylation, ubiquitination and glycosylation.[17, 18] A variety of MS approaches have been developed to study PTMs.[19-21] By comparing the mass increases of a peptide with the predicted value can indicate covalent modification[17], and MS/MS can further reveal the exact position of the modification.[21] Approaches involving the treatment with an enzyme to remove the modification (with subsequent mass spectra comparison), neutral loss scanning, or precursor ion scanning have also been employed for PTMs study.[20]

In this chapter, I will first introduce the commonly used data acquisition methods, including data-dependent acquisition (DDA) and data-independent acquisition (DIA) in discovery proteomics and multiple-reaction monitoring (MRM; also referred to as selected-

reaction monitoring, SRM) and parallel-reaction monitoring (PRM) in targeted proteomics. Then, I will discuss the MS-based labeling strategies to study quantitative proteomics, including label-free quantification (LFQ), metabolic labeling quantification such as stable isotope labeling by amino acids in cell culture (SILAC), chemical labeling quantification such as ICAT and absolute quantification of proteins (AQUA) using stable isotope-labeled (SIL) internal standards. Lastly, I will focus on the development and applications of novel discovery or targeted proteomic methods in the identifications and quantification of LPA-binding proteins, arsenic-binding proteins and ATP analog-binding proteins analysis as well as the characterization of their biological functions.

1.2 Detection Strategies in Bottom-up Proteomics

1.2.1 Discovery Proteomics

Bottom-up proteomics provides an indirect measurement of proteins through peptides derived from proteolytic digestion of intact proteins.[9] In a typical discovery proteomics experiment, it does not require any prior knowledge about the composition of the proteomic sample, the peptide mixture is fractionated and subjected to LC-MS/MS analysis, and further identified by comparing the tandem mass spectra derived acquired experimentally with theoretical tandem mass spectra derived from a protein database.[9-11] (Figure 1.2)

There are currently two broad approaches toward generating bottom-up MS proteomic data: data-dependent acquisition (DDA) and data-independent acquisition (DIA).[9, 22] Here, I will compare the DDA and DIA approaches in proteomic analysis to gain a useful overview of their features and applications.

1.2.1.1 Data-dependent Acquisition (DDA)

In tandem MS, the DDA approach selects a certain number of the most abundant peptide ions generated during the first stage of MS (MS1) for further fragmentation in the second stage of MS (MS/MS).[9] (Figure 1.3a) Typically, the precursors of highest abundance (called the "top N" precursors) are selected for further analysis (in most cases, N=10-20).[23] These selected peptides are chosen within a narrow range of mass-to-charge (m/z) signal intensity and the MS/MS data acquisition occurs sequentially for each peptide.[24] As with data acquisition, data analysis can be performed using database search, which compares measured spectra with those in established databases.

DDA performs better if the target peptides are in an existing database, as it is simple to set up and analyze, has low demand on computational resources and the database-dependent algorithms used for DDA analysis are generally faster than DIA.[25, 26] It also allows for relative quantification of peptides between samples when cooperated with various chemical labeling approaches (e.g., SILAC or iTRAQ)[27-29]

The MS instrument selects the top N precursors and then fragments them one after another can introduce a level of bias. Proteins with relatively low abundance compared to the entire proteome may be under-represented. As a result, the least complex sample can give the highest proteome coverage. For complex samples, to overcome this problem and to obtain the highest protein coverage, protein or peptide fractionation techniques prior to LC-MS/MS analysis are essential. Many prefractionation schemes have been developed, such as strong cation exchange (SCX), two-dimensional sodium dodecyl sulfate

polyacrylamide gel electrophoresis (2D-SDS-PAGE) and 2D-LC separation as implemented in multi-dimensional protein identification technology (MudPIT).[2, 30-32]

Due to the complexity of the entire proteome, shotgun proteomics cannot guarantee that every precursor ion will be selected and fragmented in every replicate. As a result, DDA datasets can contain "gaps" where peptides, especially those of low abundance, have been detected in some samples only.[33] Therefore, the reproducibility of DDA proteomic analysis among replicate runs is not very high.

1.2.1.2 Data-independent Acquisition (DIA)

In order to achieve an unbiased and deeper look at the proteome of samples, particularly for samples from little-studied organism or cell type, data-independent acquisition (DIA) was developed and widely employed as sequential window acquisition of all theoretical mass spectra (SWATH-MS) on TripleTOF platforms in 2012 with quantitative consistency and accuracy.[34, 35]

In the DIA approach, all peptides generated during the first MS stage can be fragmented and analyzed during the second stage of tandem MS. The MS/MS spectra are acquired either by fragmenting all ions that enter the mass spectrometer at a given time (called broadband DIA) or by sequentially focusing on a narrow *m/z* window of precursors and fragmenting all precursors detected within that window.[35, 36] (Figure 1.3b) It is less biased as all peptides are included in the analysis and allows for greater temporal resolution. Peptide quantification in complex mixtures can be extracted over a large dynamic range and offers higher precision and better reproducibility.[37, 38] However, it has a high demand

on computational resources and data analysis is challenging because of the multiplexed fragment ion spectra derived from multiple precursor ions. In terms of quantification, DIA has lower sensitivity than DDA as the complete spectrum must be scanned, reducing the acquisition time per data point.[33]

1.2.2 Targeted Proteomics

Discovery proteomics has emerged over the past decade as the most effective method for the qualitative study of complex proteomes; however, the sample complexity of the digested proteomes, the variation in automated top-N peptide selection and protein abundances limitation result in the compromise of reproducibility and the sensitivity for protein quantification of this stochastic strategy.[39] Thus, it is critical to develop an alternative MS approaches aiming at the systematic quantification of a predefined set of proteins involved in distinct gene families or signaling pathways in systems biology for biomarker discovery and clinical studies.

The MS-based targeted proteomics, which relies on multiple-reaction monitoring (MRM)[40, 41] or parallel-reaction monitoring (PRM)[42], aims to deliver highly reproducible and sensitive measurement. In this approach, only specific proteins of interest are selectively measured, thus, it requires information about predefined m/z ranges of the analytes.[43] However, the number of events that can be monitored per cycle time are limited; hence, it has restriction in the total number of peptides that can be monitored in each run.[44] This limitation can be diminished by scheduling the acquisition if the retention time (RT) windows of each peptide is known. To achieve this high-throughput quantification,

normalized RT (iRT), which is an empirically determined dimensionless peptide-specific retention time that allows for the highly accurate prediction of RT of targeted peptides has been introduced by Escher et al.[45] The iRT of a peptide is a fixed number relative to a standard set of reference iRT-peptides that can be transferred across different chromatographic systems and laboratories.

Ten chromatographically well distributed abundant peptides from tryptic digested bovine serum albumin (BSA) were selected as the reference peptides, as they have relatively high abundance and not interference with natural protein sequences. The iRT of reference peptides are assigned from 0 to 100 based on their elution times, and with the generation of linear regression obtained from the measured retention time versus iRT of the reference peptides, the iRT of targeted peptides can be calculated by the retention time of each peptide from the full-MS scan. After establishing the correlation between RT and iRT on the new instrument for the BSA reference peptides, and having the information of iRT for all targeted peptides, we can predict retention times of all targeted peptides from their iRT scores using the linear regression obtained from the reference peptides. (Figure 1.4)

1.2.2.1 Multiple-Reaction Monitoring (MRM)

Multiple reaction monitoring (MRM), also termed selected-reaction monitoring (SRM)), is a classic targeted quantification technique for peptides. It has been the foundation of high-quality peptide quantitation strategies when used with isotope-labeled peptides or proteins as internal standards. MRM analysis is often performed on a triple-quadrupole

(QqQ) mass spectrometer.[46, 47] The first and the third quadrupoles act as mass filters to specifically select series of transitions of peptides with predefined m/z values, and the second quadrupole serves as the collision cell.(Figure 1.5a) MRM requires previously acquired tandem mass spectra of target peptides to define transitions for MRM analysis. In most cases, at least three transition pairs are required to accurately quantify a single peptide. These transitions can be determined either from synthetic peptides[48], or from high-quality MS/MS spectra generated from shotgun proteomic analysis.[41] With the calculation and adjustment of the scheduled RT windows, the analytical robustness and throughput capacity of MRM can be dramatically enhanced.[49, 50]

The MRM-based targeted MS analysis is superior in that it provides unparalleled ability to characterize and quantify specific peptides ions with much higher sensitivity and specificity when compared with shotgun proteomics.[46] Moreover, the MRM-MS approach has better reproducibility since no full-scan MS/MS are recorded in QQQ-based SRM analysis, which results in a linear response over a wide dynamic range.[46] Therefore, MRM-MS contributes to the routine quantitative pipeline for sensitive and reproducible large-scale biomarker discovery and quantitative metabolomics analysis.[51]

Although the LC-MRM-MS strategy has been used widely in large-scale quantification analysis, the key challenge lies in is the low resolution of the triple quadrupole instruments used for MRM assays, as well as the filter of concomitantly eluted interfering background species, especially for complicated biological samples.[52]

1.2.2.2 Parallel-Reaction Monitoring (PRM)

An alternative approach has been developed to expand the options for large-scale targeted proteomics study is parallel reaction monitoring (PRM).[42, 53] Typically, PRM performed on high-resolution instruments such as hybrid quadrupole-Orbitrap (Q-OT) and time-of-flight (TOF) which provides high mass accuracy and minimizing the background signal.[53, 54]

In brief, a peptide precursor ion of interest is first isolated by the quadrupole and then fragmented in the high-energy collisional dissociation (HCD) cell. The resulting fragment ions are then analyzed with a high-resolution mass analyzer.[53] As it permits parallel detection of all detectable target product ions, there is no need for prior selection of target peptide transitions. (Figure 1.5b) Moreover, PRM offers higher specificity for peptide identification and better accuracy for quantification than MRM on QqQ instruments since the effect of interfering ions is largely reduced in high-resolution MS/MS.[55]

Sensitivity of PRM is also increased by the trapping capabilities of the C-trap, through the use of longer fill times in the process of fragmentation/accumulation of fragments, the signal-to-noise ratio will increase and help detecting of low-abundance species in complex background.[56, 57]

Study in model isotope-labeled peptides and tryptic digests of yeast showed that PRM facilitated quantitative detection in a wider dynamic range than MRM, and displayed comparable linearity, and reproducibility.[55] Collectively, the PRM-based targeted

proteomics represents a specific and accurate new technique for protein identification and quantification.

1.3 Labeling Strategies in Quantitative Proteomics

MS-based quantitative proteomics usually include the comparison of samples from different conditions, measure the changes in abundance of proteins which are functionally related to biological processes affected by the applied treatments.[58] Several quantitative proteomics approaches have been developed, including label-free quantitation by comparing spectra of the same peptide from two separate MS analyses.[59] More advanced quantitative proteomics is usually performed by incorporating stable isotopes to analytes of interest.[29, 60] Here I am going to introduce the metabolic labeling, such as stable isotope labeling with amino acids in cell culture (SILAC); chemical labeling such as isotope-coded affinity tagging (ICAT), tandem mass tags (TMT) and isobaric tagging for relative and absolute quantification (iTRAQ); and absolute quantitation (AQUA) using isotope-labeled internal standards.

1.3.1 Label-free Quantification

Label-free quantification (LFQ) is a highly economical and rapid MS analysis approach which has been widely employed in shotgun or targeted proteomic studies that determine the relative amounts of proteins in two or more biological samples.[61, 62] It allows for simultaneous identification and quantification of peptides or proteins and obviates the need of costly process of introducing stable isotope labeling.[59, 61]

In the LFQ-based proteomics approach, samples are typically digested separately and then peptide signals are detected by liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS), and their isotopic patterns are then tracked across the retention time dimension and used to reconstruct a chromatographic elution profile of the monoisotopic peptide mass.[59] The LFQ can be achieved by the counting the number of identified MS/MS spectra of each peptide/protein;[62] integrating the total product ion intensity;[63, 64] and utilizing chromatographic peak area extracted from the selected peptide precursor in different samples[65].

This approach has reasonable accuracy with careful sample preparation; however, LFQ presents quantitative issues when different peptides with a similar *m*/*z* ratio and the close retention time, or spectra contain the overlap with other interferences.[66] Thus, it is necessary to run an additional MS/MS scan which is time-consuming and reduces the resolution of the experiment. Moreover, LFQ is more prone to batch-to-batch variation, every sample must be measured separately in a label-free experiment.[66] Peptides that are expected to have the same expression levels in different biological samples but the differential processing of samples can still compromise results, such as the technical differences in sample preparation, cell culture conditions, digestion efficiency and instrument conditions.[67, 68] Thus, label-free approaches have limitations on quantification accuracy and reproducibility.[69]

1.3.2 Metabolic Labeling

Metabolic labeling approaches have also been developed to quantify protein expression.[70] In metabolic labeling, stable heavy isotopes (usually in the form of labeled amino acids) are incorporated biosynthetically into proteins, by culturing cells in controlled media containing radiolabeled amino acids, ^{15}N-enriched nitrogen source (^{15}N labeling)[71] or isotopically labeled essential amino acids (stable isotope labeling by amino acids in cell culture or SILAC[60]) via the process of translation. Two cellular states can then be compared by pre-mixing the "light" and "heavy" tagged cell/ lysate/protein/peptide at 1:1 ratio before analysis.

Radioisotopes has been utilized for a long time, it labels proteins using radioactive amino acids.[72] [^{35}S] methionine is the most commonly used radioisotope as it is sufficiently radioactive (~40 TBq/mmol) and it allows for the accumulation measurement of radio labeled material in a relatively short labeling time and with modest incorporation.[73, 74] The shortcoming of radio-labeling is the noise floor of common mass spectrometer which might obscure the degree of labeling when a precursor stable was isotope-labeled at a low incorporation.[75]

To overcome this limitation, a series of stable isotope labeling strategies have been developed for accurate quantification in comparative proteomics, including SILAC. SILAC was first introduced by Ong *et al.* in 2002, and it enables a pair-wise comparison of one components in the global proteome.[60] In this approach, it is assumed that the isotope-labeled "heavy" precursor amino acid is not differentially metabolized in cells and the proteins will be fully labeled with "light" or "heavy" isotope-labeled amino acids

after multiple times cell subculture in SILAC medium, such that the heavy/light ratio can be taken to directly represent the relative amounts of the analyte in the two cell populations.[60]

Typically, heavy labeled essential amino acids lysine and arginine are most widely used, since it can combine with trypsin digestion to ensure labeling of every peptide except for the C-terminal peptide of the protein.[71] Cells were first cultured in medium containing "light" or "heavy" lysine and arginine, respectively, and after cells are fully isotope-labeled, "light" and "heavy" cell lysates are generated and equal-mass mixed. Followed by tryptic digestion, the resulting peptide mixtures are purified or fractionated prior to LC-MS/MS analysis. Isotope-labeled peptide usually appears as a "doublet" in the mass spectra with amino acid-related m/z shifts.[60] By comparing the relative signal intensities of the paired isotope-derived peptide peaks, the differences in protein abundance between these two cellular states can be directly calculated. (Figure 1.6) Multiple SILAC-derived methodology have been developed for the comparison among three or more cell populations, such as "triple-SILAC", namely SILAC with three isotope labeling states, expanded the applicability of stable isotope labeling in quantitative proteomics.[76]

Labeling protein *in vivo* through metabolic incorporation of label has many advantages, where it enables earlier introduction of labeling and combination of samples grown in different states at the cell/cell lysate level. Therefore, it eliminates systematic variation and errors from downstream sample preparation and MS measurement, thereby providing better quantitative accuracy and reproducibility compared with LFQ approaches.[77]

Metabolic labeling is particularly suited to cultured cells whereas many biological systems are not amenable to efficient metabolic labeling, such as tissue samples, natural microbial communities and biological fluids.[77, 78] It should also be pointed out that many other complicating factors must be considered to conduct meaningful quantitative measurements, including incomplete labeling, metabolic lability of the amino acid precursor, and the role of protein turnover in labeling kinetics.[78]

1.3.3 Chemical Labeling

To further improve quantification performance in biological systems not amenable to metabolic labeling, a more general term whereby label is incorporated after cell lysis or preparation of the protein mixture, called chemical or enzymatic labeling methods have been developed,[79] such as isotope-coded affinity tagging (ICAT),[32] tandem mass tags (TMT)[80] and isobaric tagging for relative and absolute quantification (iTRAQ)[81]. After cell lysis or protein digestion, relevant peptides can be labeled chemically with a small functional group, which retains similar chemical and physical properties yet distinguishable m/z values, compared to their natural counterparts. In proteomics measurements, the ratios of isotopically labeled peptides in comparison to the unlabeled standard are determined using their signal intensities in full parent ion scans of the LC−MS/MS analysis.[79]

ICAT was the first chemical tagging strategy in quantitative proteomics developed in 1999 by Gygi *et al*.[32] Typically, the ICAT reagent consists of three moieties: an affinity tag, a light or heavy isotope-labeled linker containing either hydrogens or deuterons, and a

reactive moiety which reacts specifically with cysteine or other amino acid residues. By targeting different amino acid residues for active-site labeling, ICAT represents a flexible labeling strategy for investigation of potentially undiscovered analytes and proteome-wide quantification differences between samples, especially when coupled with subsequent affinity-based enrichment (Figure 1.7). Abundance ratios of affinity captured peptides are then used to infer abundance ratios of their parent proteins.[32, 78]

In addition to ICAT-based quantitation, another two isobaric chemical labeling method iTRAQ and TMT, have become increasingly used in quantitative proteomics. After proteolysis, samples are labeled separately with different isotopic variants of iTRAQ or TMT and are then combined for LC−MS/MS analysis.

iTRAQ was brought up in 2004 by Applied Biosystems (Ross *et al*.), in a 4-plex fashion.[82] In addition, tandem mass tags (TMT) 6-plex isobaric mass tagging kit was first released by Thermo Fisher Scientific.[80] 6-Plex TMT and 4-plex iTRAQ allow for simultaneous quantification of up to 6 and 4 samples in a single LC−MS/MS analysis, respectively. In 2012, McAlister *et al*. expanded the multiplexing capacity of the TMT reagents to at least 10-plex.[83] Recently, the TMTpro 16-plex reagent was developed to further expand the multiplexity.[84] The design of the reagents for both iTRAQ and TMT consists of a reporter group, a mass normalization group and an reactive moiety.[29] For example, an amine-reactive group can covalently react with N-terminal amine groups and the lysine residues specifically. The mass normalization groups balance the mass difference among the reporter ion groups such that different isotopic variants of the tag have the same mass. They generate isobaric peptides with the same molecular mass and elution behavior

in liquid chromatography separation, which cannot be differentiated in MS survey spectra (MS1), which prevents increasing the full-scan complexity after mixing multiple samples. After collision-induced dissociation, reporter ions of different masses are dissociated from isobaric peptide precursors, and the relative intensity of reporter ions can be used for relative parent protein abundance quantification (Figure 1.8). In addition, due to the multiplexed quantification capability, iTRAQ and TMT techniques are particularly useful for monitoring protein changes in biological systems upon multiple treatments or over multiple time points.[78]

High multiplexing and precise quantitative analyses are unique attributes of iTRAQ- or TMT-based approaches in comparison to the other labeling techniques. However, quantification of complex proteomic samples could be biased due to the "ratio compression" issue originated from co-fragmentation of peptides.[85] This is usually linked with reporter ion pattern distortion caused by inevitable co-isolated precursor interference during the MS/MS-based quantitation.[85] To circumvent this issue, the SPS-MS3 technique, namely isolation waveforms with multiple frequency notches (i.e., synchronous precursor selection, SPS) coupled to MS/MS/MS, can effectively enhance quantitative accuracy and sensitivity.[86, 87] That said, MS2-based TMT outperforms MS3-based TMT since its higher precision and larger identification numbers enable detection of a greater number of significantly altered proteins.[88] Lastly, the introduction of heavy labels at a late stage of sample preparation is more susceptible to experimental errors, compared to metabolic labeling.[89]

1.3.4 Absolute QUAntitation (AQUA) using Isotope-labeled Internal Standards

Chemical labeling is usually performed with the peptide and proteins of interest. In contrast, the use of stable isotope dilution of synthetic peptides standards in absolute quantitation of proteins (AQUA) can act as internal standards in mass spectrometry analysis for a specific subset of peptides.[90] In AQUA strategy, a known quantity of AQUA peptides, which are simply synthesized with incorporated stable isotopes, are added to a biological peptide sample, they are chemically and physically indistinguishable from their normal peptide of interest with same retention time, ionization efficiency and fragmentation mechanism. Therefore, they can serve as an ideal internal standard by mimicking native peptides formed from proteolysis. The mixture is then digested and analyzed by LC-MS/MS.(Figure 1.9) The AQUA peptides are chemically identical to their native counterparts derived from protein mixture, but the labeled and native forms of these peptide pairs can be easily differentiated by molecular weight in MS analysis.[90, 91] The absolute quantity of the native peptide as well as the corresponding protein can be highly accurately and selectively determined by integrating the MS peak areas of the native peptide and their known amount synthetic isotope-labeled AQUA peptide.[90, 91] It also reduces the sample amount necessary for peptide detection, and thus allows high sensitive identification and quantification of low abundance targets.

Considering the availability and costs for isotope-labeled synthetic peptides for absolute quantification, and the limitation of the number of proteins that can be measured in a single LC-MS run, AQUA cannot be applied in assays with a great number of proteins. Thus, AQUA is specifically attractive when combined with targeted proteomics strategy, such

as MRM-MS.[92-94] This combined approach has been reported to detect proteins with concentrations less than 50 copies per cell in unfractionated lysates, and been widely used for analysis and validation of potential biomarkers in clinical samples with high reproducibility and sensitivity.[91]

1.4 Biologically Important Small Molecules

1.4.1 Lysophosphatidic acid (LPA) and LPA receptors

One of the bioactive phospholipids[95], known as lysophosphatidic acid or LPA[95-97], has garnered interest for its multiple biological activities especially for its extracellular signaling properties. LPA is a small glycerophospholipid that consists of a glycerol backbone, a free phosphate group and one saturated or unsaturated fatty acid chain. It is produced during the synthesis of cell membranes and is described as a robust extracellular signaling molecule present in all eukaryotic tissues and blood plasma.[98] LPA can derive from the precursor membrane phospholipids such as phosphatidylcholine, phosphatidylserine, and phosphatidylethanolamine, by first producing respective lysophospholipids through the action of phospholipase A1 (PLA1) or PLA2. These lysophospholipids are then converted into LPA by autotoxin (ATX), which is mainly responsible for LPA maintenance at a physiological concentration in plasma after birth and also during vascular and embryonic development.[98, 99] It also can be synthesized intracellularly, phosphatidic acid generated from phospholipids by phospholipase D (PLD) can be cleaved by PLA1 and PLA2 to yield LPA.[98,99]

Because of the heterogeneity of LPA-binding protein subtypes, expression patterns, and effector pathways, LPA affects various cellular processes, such as smooth muscle contraction, platelet aggregation, cell proliferation, and cytoskeletal reorganization.[100] It is also reported that aberrant LPA-signaling is associated with numerous cancers.[98, 101] It has been confirmed that the activation of seven-transmembrane domain G protein-coupled receptors (GPCRs) were linked to the cellular responses triggered by LPA.[102] The first result correlating LPA signaling through these GPCRs showed that the upregulation of Edg2 in murine neuronal cultures could lead to morphological changes and adenylyl cyclase inhibition in response to LPA.[98] Even before the year 2000, another eight genes encoding Edg receptors in the human genome were identified. Among them, three encoding LPA receptors, namely LPAR1/Edg2,[103] LPAR2/Edg4[104] and LPAR3/Edg7,[105, 106] which, in turn, are part of the rhodopsin GPCR family alpha subclass.[107] They can mediated LPA signaling effect via the Rho/Rho-associated protein kinase (ROCK) signaling pathway, activating phospholipase C (PLC), Ras, and phosphatidylinositol 3-kinase pathways or inhibiting the production of cAMP. In 2003, beyond the classical Edg family LPA receptors, the first non-Edg LPA receptor, which was further called LPAR4, was identified during experiments to find a P2Y9/GPR23 ligand.[108, 109] This discovery opened up the possibility of investigating the existence of other LPA receptors and revealed the existence of two more LPA receptors, the LPAR5 (GPR92/93) and LPAR6 (P2Y5)[110]. (Figure 1.10)

Although several cell-surface LPA receptors have been discovered, the mechanism of LPA regulated cell signaling pathways cannot be fully explained; thus, there are likely

many yet to be identified LPA-binding proteins. Identification and quantification of LPA-binding proteins at the proteome-wide scale is important for investigating biological events underlying the complicated signaling network comprised of LPA receptors and down-stream effector proteins.

1.4.2 As (III) and As(III)-binding Proteins

Arsenic is a widespread environmental contaminant found in the earth's crust at an average concentration of ~5 μg/g (ppm) and it is a component of 245 minerals, associated most frequently with other metals such as copper, gold, lead, and zinc in sulfidic ores.[111] When disturbed by natural processes, such as weathering, biological activity, and volcanic eruption, arsenic may be released into the environment. Long-term exposure to arsenic from groundwater is known to be associated with the development of human diseases. More than 100 million people are at risk of cancer and other arsenic-related diseases, such as skin, bladder cancers,[112-115] neurological diseases,[116, 117] diabetes mellitus,[118] cardiovascular and peripheral vascular disease[119, 120] and hematological disorders.[121] On the other hand, arsenic trioxide is used as a chemotherapeutic agent for treating acute promyelocytic leukemia patient carrying the promyelocytic leukemia/retinoic acid receptor alpha (PML-RARα) oncogenic fusion.[122]

It is known that trivalent arsenic has a tendency to bind to sulfhydryl (SH) groups in vicinal cysteine residues in proteins.[111, 123, 124] Moreover, the binding of arsenic with cysteine (Cys) residues can result in conformational changes in the target protein, which may result in altering the conformation and function of the protein as well as aberrant

protein–protein or protein-DNA interactions.[125-127] For example, in the model of As(III) binding to the *E.coli* repressor protein ArsR.[128] The two cysteine residues within an α-helix must unravel from one end in order to bind As(III). Melting of the helix produces the conformational change that dissociates the repressor from the DNA, inducing gene expression. (Figure 1.11)

Because arsenic is a cancer inducer and at the same time a therapeutic agent, many studies have been conducted to examine the biological consequences of arsenic exposure. However, the mechanisms through which it elicits its toxic effects are not fully understood. Therefore, identification of arsenic-binding proteins is essential for understanding the mechanisms through which arsenic confers its biological effects. Many arsenic-binding proteins have been identified in mammalian cells, including hemoglobin,[129] tubulin[130], actin[131], arsR[111], nicotinic receptor[132], galectin-1[133] and thioredoxin peroxidase II.[134] However, some of the observed toxic effects of arsenic cannot be attributed to its interactions with these proteins.[111] Thus, there are likely other yet identified arsenic-binding proteins.

Several proteomic approaches were employed to identify arsenic-binding proteins. In this vein, a global profiling using a human proteome microarray led to the identification of 360 arsenic-binding proteins, including hexokinase 2 (HK2), whose enzymatic activity can be inhibited by As(III).[135] MS has great advantages in the identification and quantification of proteins in complex sample matrices.[136] Le and colleagues[137] employed an approach combining-LC-MS/MS with click chemistry using an azide-labeled arsenical to selectively capture and identify 48 arsenic-binding proteins in A549 cells.

An amide linked conjugate of *p*-aminophenylarsine oxide and biotin is conveniently prepared in a one-pot procedure by the reaction of biotinyl chloride and *p*-aminophenyldichloroarsine,[132, 138](Figure 1.12) which combines the characteristics of biotin and an arsenic reagent and therefore is bifunctional for thiols and avidin (or streptavidin), has been used in the study for the identification of arsenic-binding proteins in MCF-7 human breast cancer cells and 50 arsenic-binding proteins were identified using MALDI-TOF MS.[138]

1.4.3 ATP binding protein and ATP analog

Adenosine triphosphate (ATP) is an abundant nucleoside triphosphate that often functions as a form of intracellular energy transfer. ATP binds to numerous proteins called "ATP-binding proteins" that play pivotal roles in many cellular processes like metabolism, synthesis, active transport and cell signaling. There are almost 1500 proteins assigned as ATP-binding proteins in the UniProt database; however, the whole picture of the ATP-binding proteome remains unclear. Biotin-conjugated acyl-ATP probes have been developed and widely employed for large-scale identification of ATP-binding protein.[139, 140] In mammalian cells, 349 ATP-binding proteins were identified in HL-60 cells using a desthiobiotin-conjugated ATP-affinity probe.[141]

Kinases are one of the most important enzyme families in eukaryotic cells, and they mediate cellular protein and lipid phosphorylation to regulate downstream signaling cascades(Figure 1.13)[142]. Aberrant expression and/or activation of kinases are closely associated with disease development and resistance toward cancer therapy.[143] Therefore,

kinases have become one of the most intensively pursued enzyme superfamilies as drug targets for cancer chemotherapy and more than 30 small-molecule kinase inhibitors have been approved by the Food and Drug Administration (FDA) for cancer chemotherapy.[144] Furthermore, kinase inhibitors have been extensively developed for more than 400 kinases, covering the majority of the human kinome.[145-147]

Based on the cosubstrate role that ATP plays in the kinase mechanism, modified ATP analogs have been useful tools to probe kinase mechanism, structure, activity, interactions, and function. These analogs differ structurally from natural ATP through alteration of atoms or groups on the base, sugar, and/or triphosphate regions of the molecule.[148]

Shokat et al.[149] introduced a creative approach, termed analog substrate-kinases, to define kinase-substrate relationship. The approach harnesses the power of chemical syntheses of N^6-alkylated ATP analogs or purine analogs and genetic engineering of the kinase of interest to accommodate these analogs for substrate phosphorylation. An assumption underlying the aforementioned approach that the ATP analogs are not tolerated by natural kinases. This assumption, however, has been rejected when N^6-furfuryl-ATP (a.k.a. kinetin triphosphate, KTP) was found to bind and enhance the activity of wild-type PINK1 kinase and restore the catalytic activity of a Parkinson's disease-related mutant (i.e. G309D) of PINK1 to near-wild-type levels.[150] Hence, it is important to assess, at the proteome-wide scale, whether these analog substrates can also be recognized by endogenous kinases. If so, the results obtained from such studies may provide an opportunity to discover substrates for native kinases in cells.

1.5 Scope of the Dissertation

Small molecules such as LPA, As(III) and ATP and their binding proteins play pivotal roles in regulation of many cellular processes. However, the proteomics study of their binding protein and the mechanism behind their regulatory pathways remains challenging. In this dissertation, I combined discovery and targeted proteomics, with the using of affinity-labeled chemical probe pull-down strategy to identify, and quantify several small molecule- and modified nucleotide- binding proteins in the entire human proteome.

In Chapter 2, I described a strategy that used the desthiobiotin- LPA probe to enrich and identify putative LPA-binding proteins in the human proteome. We successfully synthesized the desthiobiotin-tagged LPA-affinity probe and employed it, together with LC-MS/MS analysis, to enable the identification of 939 putative LPA-binding proteins from two different cell lines. By combining this strategy with SILAC, we were able to discover 86 proteins exhibiting highly selective interactions with LPA at the entire proteome level. Moreover, we validated that ANXA5 and PGK1 can bind directly with LPA.

In Chapter 3, by employing a similar chemoproteomic approach, which relies on SILAC-based metabolic labeling, biotin-As probe pull-down and LC-MS/MS analysis, we achieved the proteome-wide identification of arsenic-binding proteins in human cells. We were able to identify a total of 409 candidate arsenic-binding proteins from HEK293T cells and quantitatively characterized the arsenic-protein interactions, which results in the identification of 51 specific As(III)-binding proteins. Among them, HSPA4's interaction with As(III) may contribute to the proteotoxic effects of arsenic exposure.

In Chapter 4, we employed isotope-coded ATP acyl-phosphate probes, in conjunction with a multiple-reaction monitoring (MRM)-based targeted proteomic method, for proteome-wide identifications of endogenous kinases that can bind to two N^6-modified ATP derivatives. We found that, among the ~300 quantified kinases, 28 and 18 are candidate kinases that can bind to N^6-methyl-ATP and N^6-furfuryl-ATP, respectively. Additionally, *in-vitro* biochemical assay showed that GSK3β, one of these kinases, could employ N^6-Me-ATP as the phosphate group donor to phosphorylate its substrate peptide.

Figure 1. 1 General procedures of experiment for bottom-up proteomics. Scheme created using Biorender.com

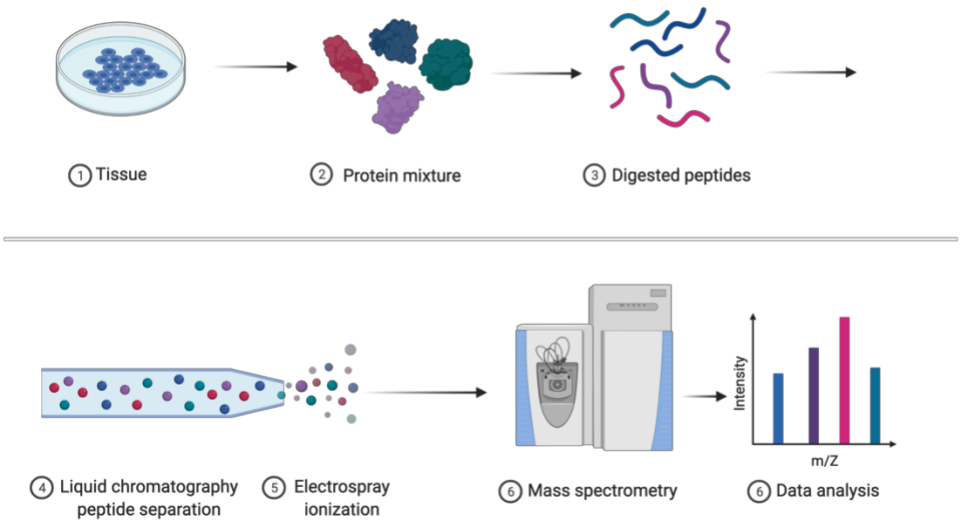


Figure 1. 2 Schematic representation of the bottom-up proteomics and top-down proteomics spectrometry that are used to identify and characterize. Scheme created using Biorender.com

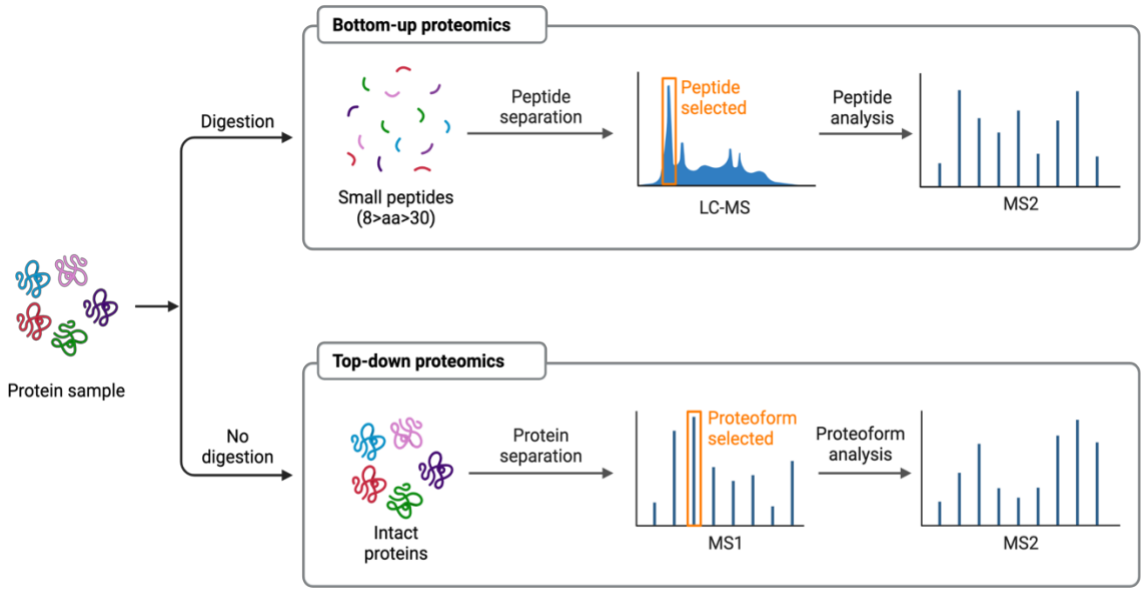


Figure 1. 3 Representative scanning modes used in MS-based bottom-up discovery proteomics. Shown is a schematic diagram representing common ion detection methods in (a)DDA and (b)DIA. Scheme created using Biorender.com

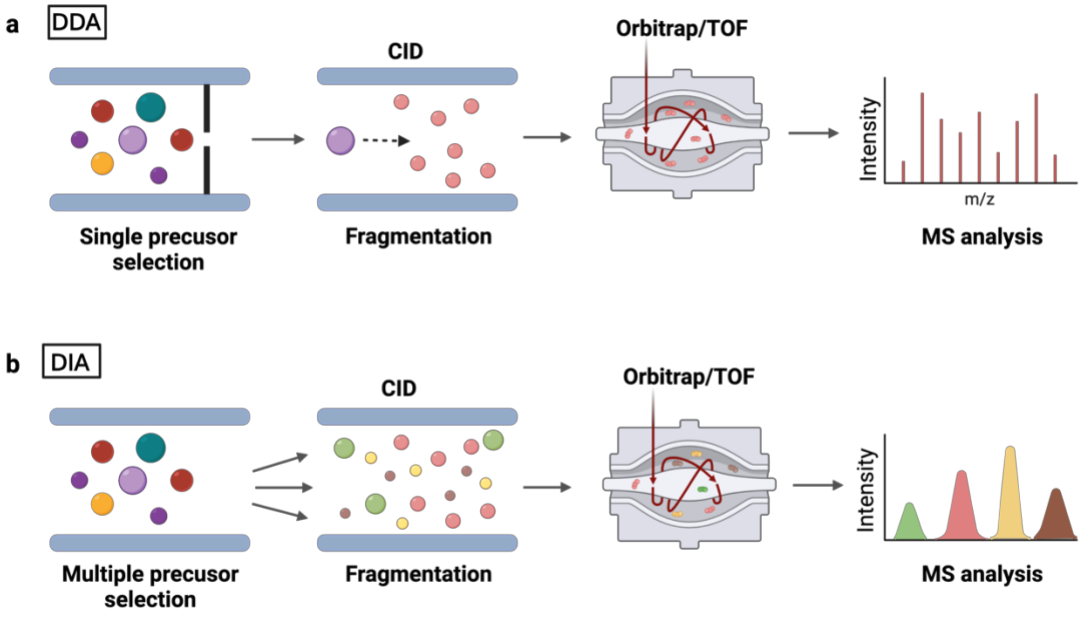


Figure 1. 4 A schematic workflow that outlines the empirical determination of iRT scale, conversion of RTs for targeted peptides into iRTs, and RT scheduling for MRM. To simplify, 10 reference peptides (iRT-peptides) used for iRT transformation based upon linear regression. The retention time of the target peptide (RT_x) is can be empirically transformed into iRT using the established linear regression of the iRT-peptides. iRT values can then be transferred to a different chromatographic setup by using an RT calibration of the iRT-peptides. Scheme created using Biorender.com

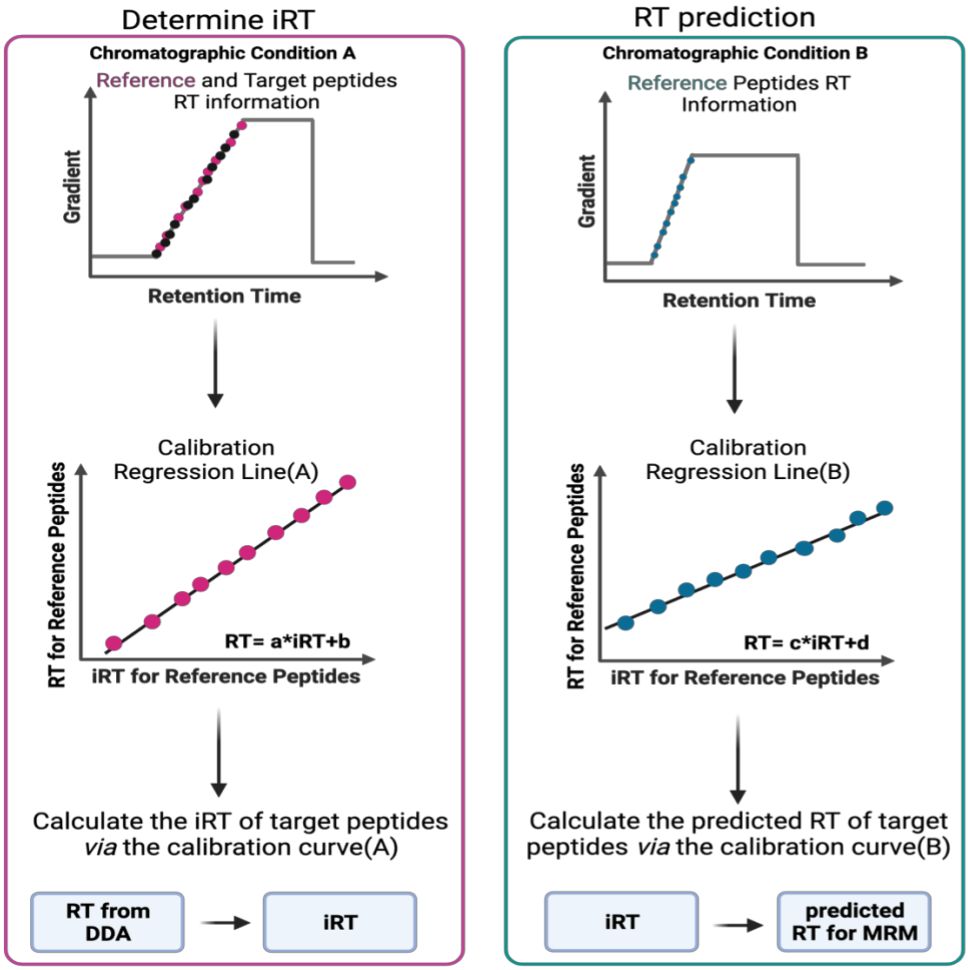


Figure 1. 5 Representative scanning modes used in MS-based bottom-up targeted proteomics. Shown is a schematic diagram representing common ion detection methods in (a)MRM and (b)PRM. Scheme created using Biorender.com

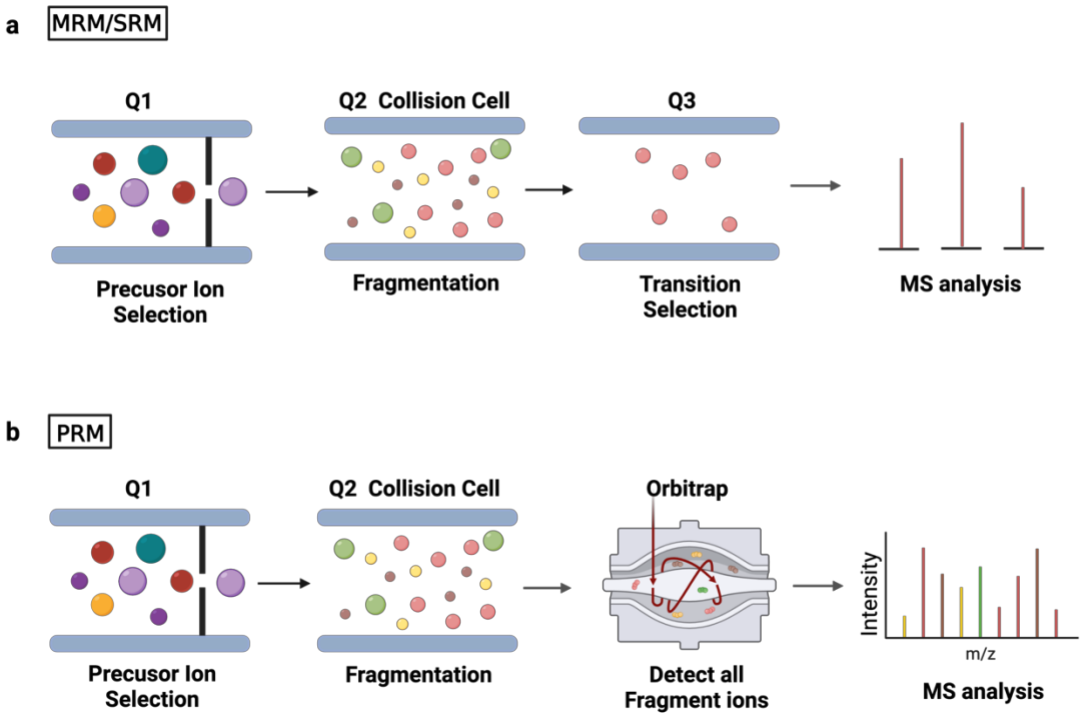


Figure 1. 6 General procedures of SILAC experiment for quantitative proteomics. Scheme created using Biorender.com

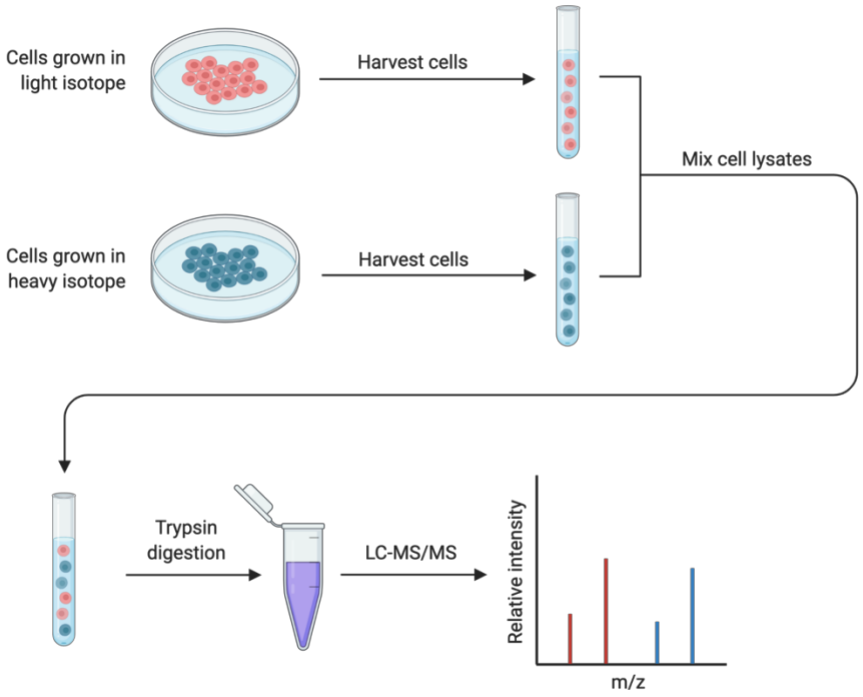


Figure 1. 7 General procedures of ICAT experiment for quantitative proteomics.

Scheme created using Biorender.com

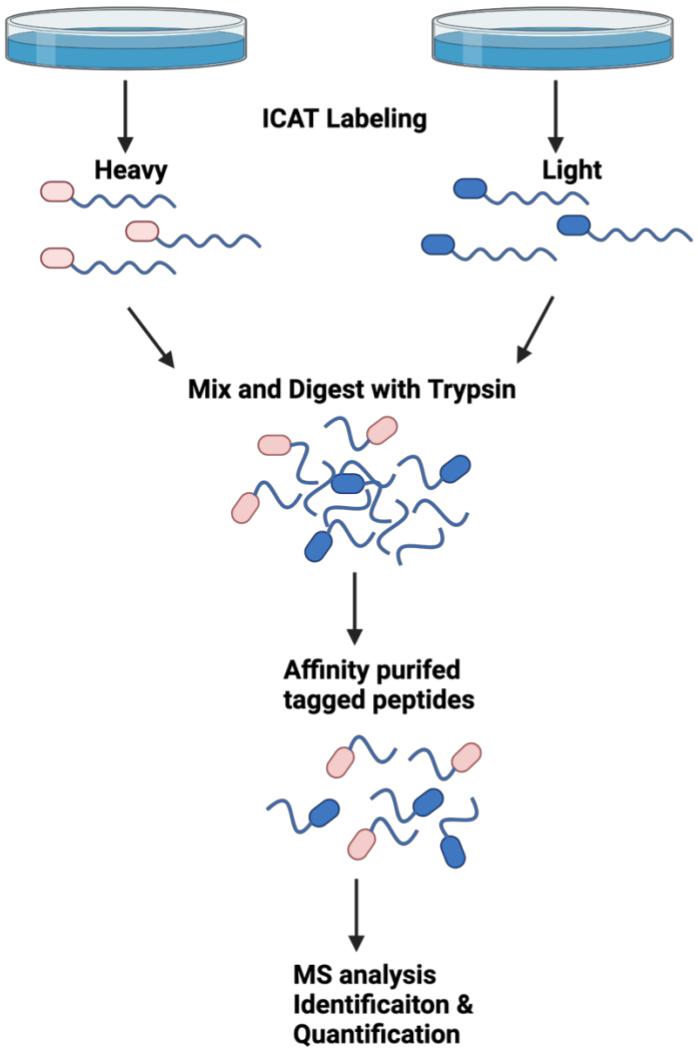


Figure 1. 8 General procedures of TMT experiment for quantitative proteomics.

Scheme created using Biorender.com

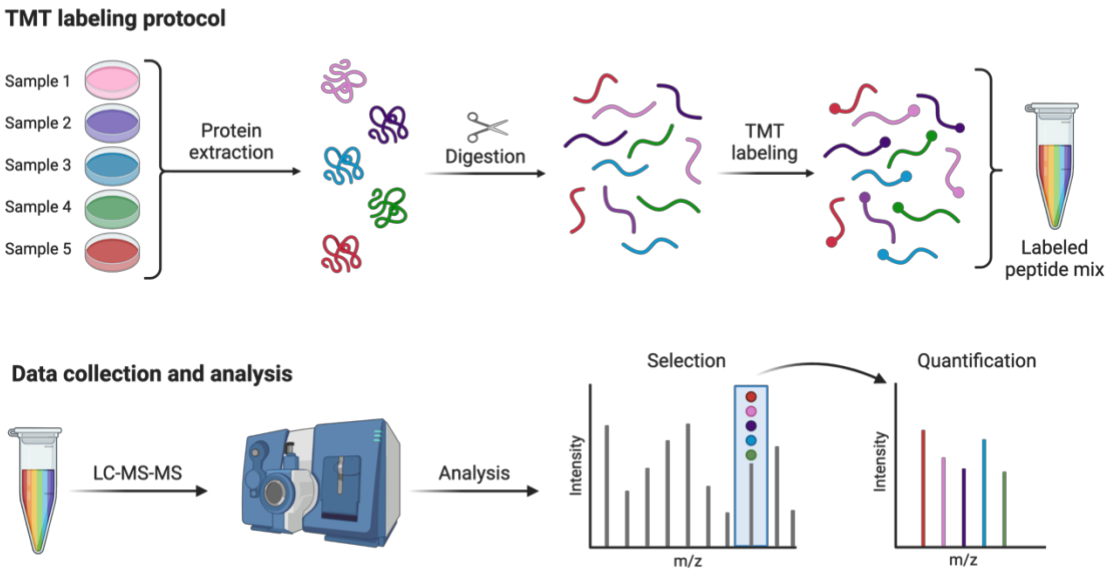


Figure 1. 9 General procedures of AQUA experiment for quantitative proteomics.

Scheme created using Biorender.com

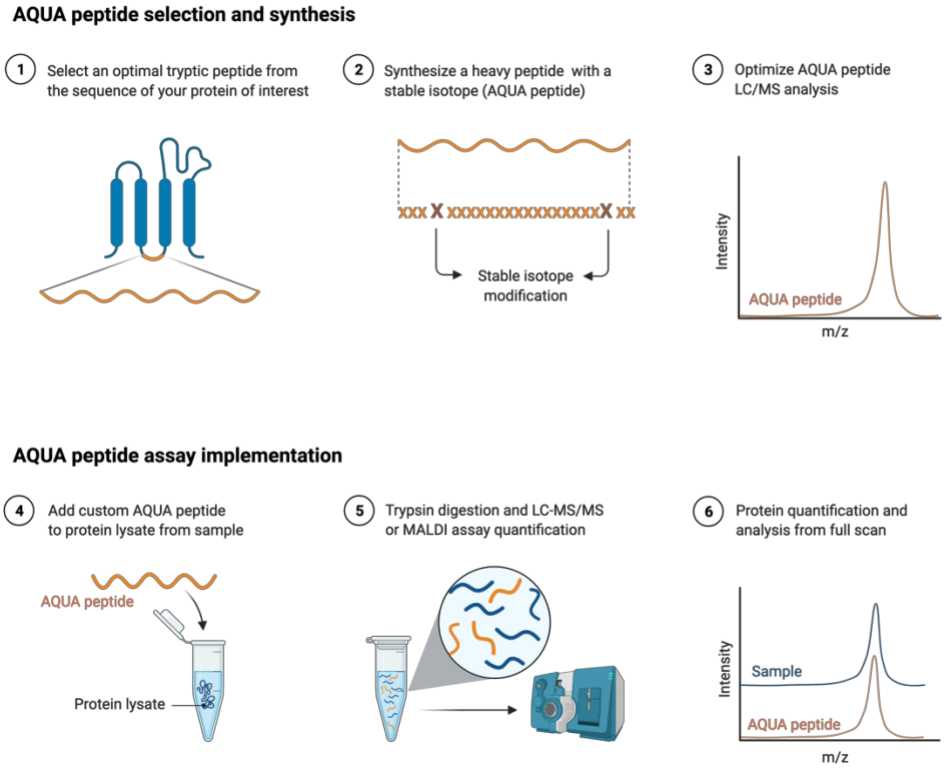


Figure 1. 10 LPA receptors and signaling intermediates. LPA signals through six different LPA transmembrane G protein-coupled receptors (GPCRs and LPAR1–6) and other GPCRs (such as P2Y10 and GRP87). These receptors activate different intracellular signaling mediators to elicit different/several cellular responses. (adopted from *Sig Transduct Target Ther*, 2021, 6, 45)

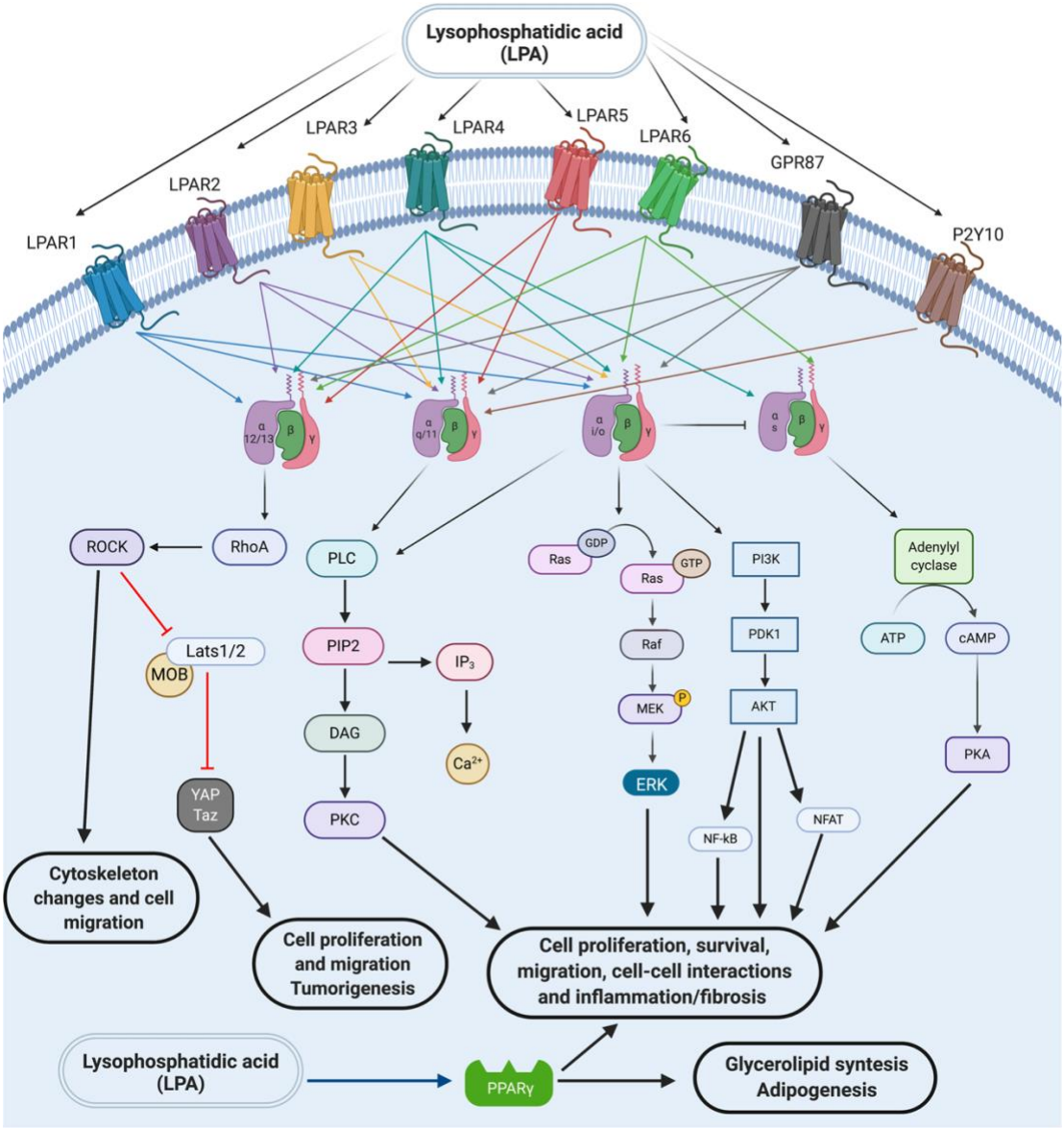


Figure 1. 11 Binding of iAsIII to the ArsR repressor from *E. coli* plasmid R773 (P15905) results in conformational change of the repressor. iAsIII binds to Cys32, Cys34, and Cys37 of the ArsR repressor. Unwinding the helix disrupts DNA binding, resulting in dissociation of the repressor from the operator site. Dissociation of the repressor induces gene expression. (Adapted from *J. Biol. Chem.* 2008, 283, 25706)

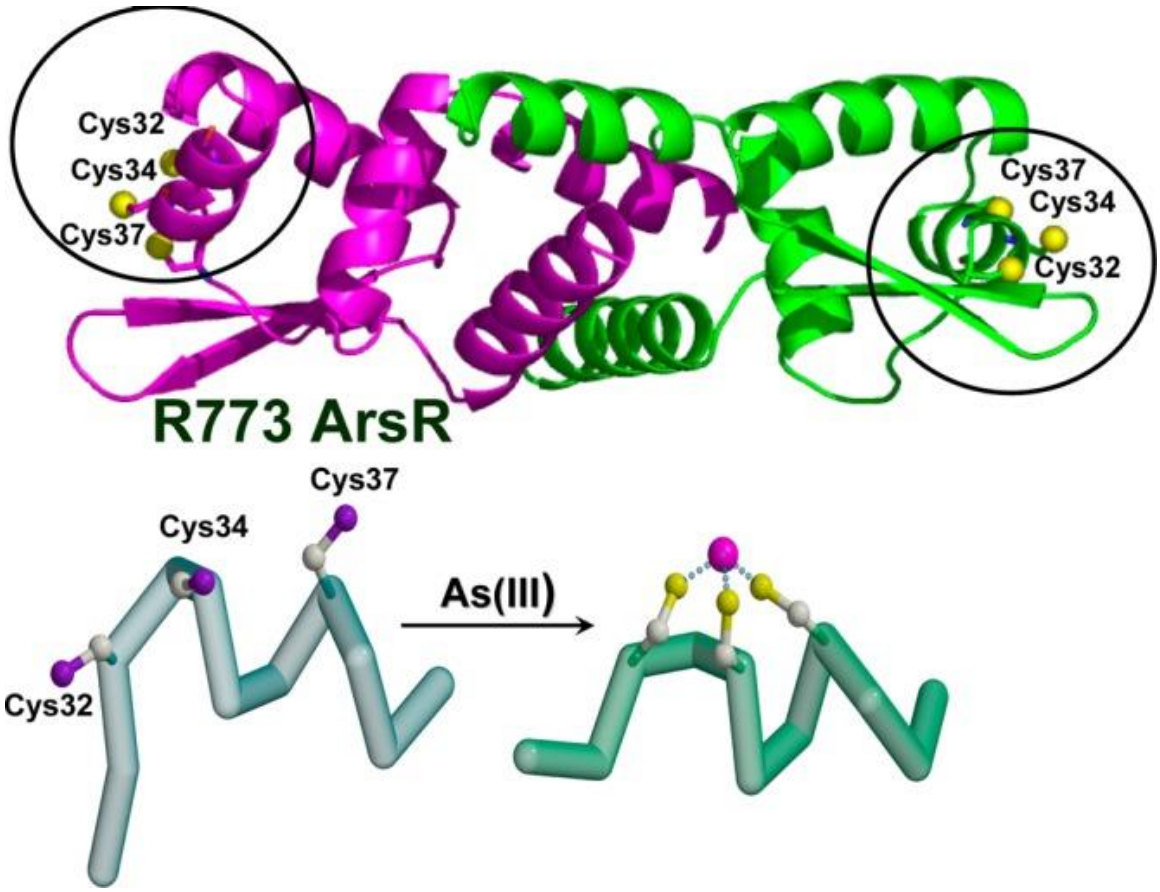


Figure 1. 12 Schematic representation of synthesis of arsenic-biotin conjugate. (Adopted from Cancer Lett. 2007 Sep 18; 255(1): 95–106)

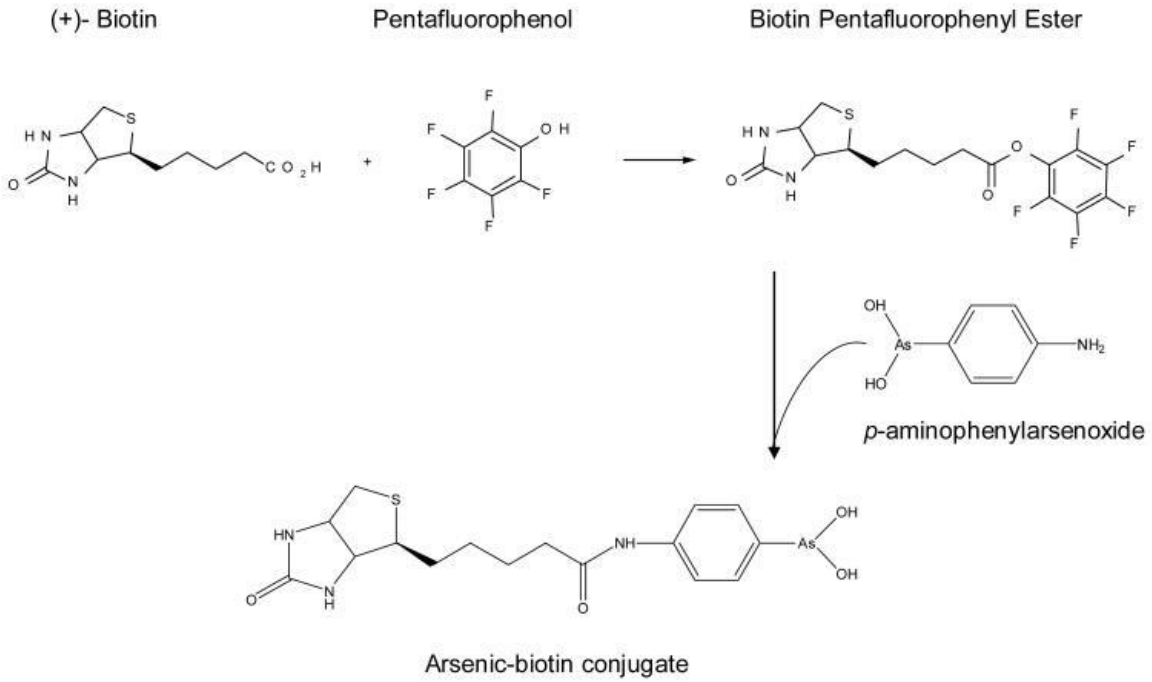
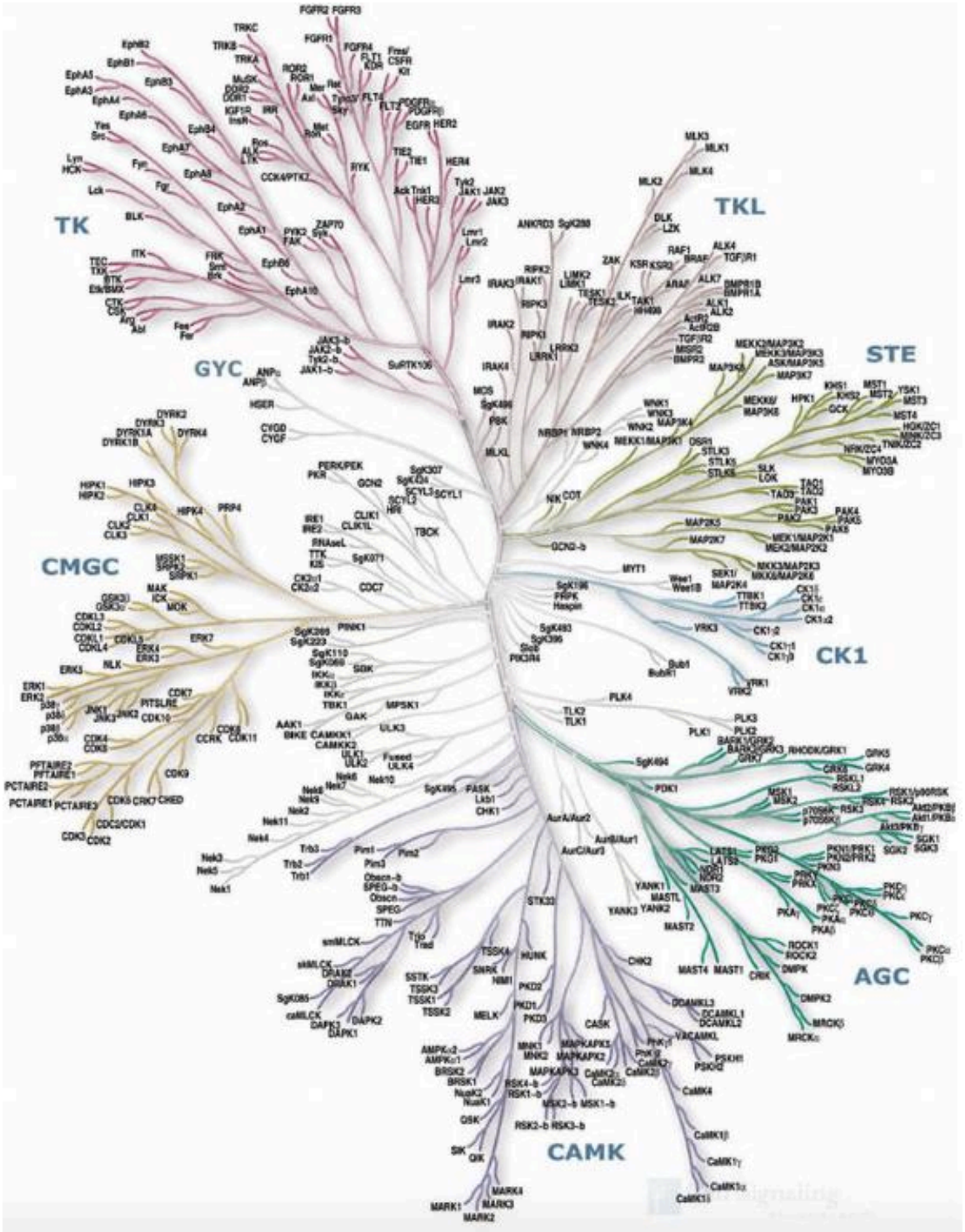


Figure 1. 13 The dendrogram of the human kinome. (adopted from Science. 2002, 298, 1912.)



References

1. Glish, G. L.; Vachet, R. W., The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery* **2003,** *2* (2), 140-150.

2. Wolters, D. A.; Washburn, M. P.; Yates, J. R., An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **2001,** *73* (23), 5683-5690.

3. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999,** *17* (7), 676-682.

4. Aebersold, R.; Goodlett, D. R., Mass Spectrometry in Proteomics. *Chem. Rev.* **2001,** *101* (2), 269-296.

5. Dyson, H. J.; Wright, P. E., Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* **2005,** *6* (3), 197-208.

6. Clevers, H.; Nusse, R., Wnt/β-Catenin Signaling and Disease. *Cell* **2012,** *149* (6), 1192-1205.

7. Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Muzio, L. L., The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int. J. Mol. Med.* **2017,** *40* (2), 271-280.

8. Dupree, E. J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B. A.; Darie, C. C., A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes* **2020,** *8* (3), 14.

9. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013,** *113* (4), 2343-2394.

10. Amunugama, R.; Jones, R.; Ford, M.; Allen, D., Bottom-Up Mass Spectrometry–Based Proteomics as an Investigative Analytical Tool for Discovery and Quantification of Proteins in Biological Samples. *Adv. Wound Care* **2013,** *2* (9), 549-557.

11. Han, X.; Aslanian, A.; Yates, J. R., Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **2008,** *12* (5), 483-490.

12. Karpievitch, Y. V.; Polpitiya, A. D.; Anderson, G. A.; Smith, R. D.; Dabney, A. R., Liquid chromatography mass spectrometry-based proteomics: Biological and technological aspects. *The Annals of Applied Statistics* **2010,** *4* (4), 1797-1823.

13. Graves, P. R.; Haystead, T. A. J., Molecular Biologist's Guide to Proteomics. *Microbiol. Mol. Biol. Rev.* **2002,** *66* (1), 39-63.

14. Hansen, K. C.; Schmitt-Ulms, G.; Chalkley, R. J.; Hirsch, J.; Baldwin, M. A.; Burlingame, A. L., Mass Spectrometric Analysis of Protein Mixtures at Low Levels Using Cleavable 13C-Isotope-coded Affinity Tag and Multidimensional Chromatography. *Mol. Cell. Proteomics* **2003,** *2* (5), 299-314.

15. Lahm, H.-W.; Langen, H., Mass spectrometry: A tool for the identification of proteins separated by gels. *Electrophoresis* **2000,** *21* (11), 2105-2114.

16. Gygi, S. P.; Rist, B.; Griffin, T. J.; Eng, J.; Aebersold, R., Proteome Analysis of Low-Abundance Proteins Using Multidimensional Chromatography and Isotope-Coded Affinity Tags. *J. Proteome Res.* **2002,** *1* (1), 47-54.

17. Schweppe, R. E.; Haydon, C. E.; Lewis, T. S.; Resing, K. A.; Ahn, N. G., The Characterization of Protein Post-Translational Modifications by Mass Spectrometry. *Acc. Chem. Res.* **2003,** *36* (6), 453-461.

18. Finehout, E. J.; Lee, K. H., An introduction to mass spectrometry applications in biological research. *Biochem. Mol. Biol. Educ.* **2004,** *32* (2), 93-100.

19. Larsen, M. R.; Trelle, M. B.; Thingholm, T. E.; Jensen, O. N., Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *BioTechniques* **2006,** *40* (6), 790-798.

20. Sickmann, A.; Mreyen, M.; Meyer, H. E., Identification of Modified Proteins by Mass Spectrometry. *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)* **2002,** *54* (2), 51-57.

21. Zhang, X.; Herring, C. J.; Romano, P. R.; Szczepanowska, J.; Brzeska, H.; Hinnebusch, A. G.; Qin, J., Identification of Phosphorylation Sites in Proteins Separated by Polyacrylamide Gel Electrophoresis. *Anal. Chem.* **1998,** *70* (10), 2050-2059.

22. Doerr, A., DIA mass spectrometry. *Nat. Methods* **2015,** *12* (1), 35-35.

23. Pino, L. K.; Rose, J.; O'Broin, A.; Shah, S.; Schilling, B., Emerging mass spectrometry-based proteomics methodologies for novel biomedical applications. *Biochem. Soc. Trans.* **2020,** *48* (5), 1953-1966.

24. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422* (6928), 198-207.

25. Barbier Saint Hilaire, P.; Rousseau, K.; Seyer, A.; Dechaumet, S.; Damont, A.; Junot, C.; Fenaille, F., Comparative Evaluation of Data Dependent and Data Independent

Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics. *Metabolites* **2020,** *10* (4), 158.

26. Guo, J.; Huan, T., Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography–Mass Spectrometry Based Untargeted Metabolomics. *Anal. Chem.* **2020,** *92* (12), 8072-8080.

27. Köcher, T.; Pichler, P.; Schutzbier, M.; Stingl, C.; Kaul, A.; Teucher, N.; Hasenfuss, G.; Penninger, J. M.; Mechtler, K., High Precision Quantitative Proteomics Using iTRAQ on an LTQ Orbitrap: A New Mass Spectrometric Method Combining the Benefits of All. *J. Proteome Res.* **2009,** *8* (10), 4743-4752.

28. Cox, J.; Matic, I.; Hilger, M.; Nagaraj, N.; Selbach, M.; Olsen, J. V.; Mann, M., A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nat. Protoc.* **2009,** *4* (5), 698-705.

29. Li, Z.; Adams, R. M.; Chourey, K.; Hurst, G. B.; Hettich, R. L.; Pan, C., Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos. *J. Proteome Res.* **2012,** *11* (3), 1582-1590.

30. Mcdonald, W. H.; Yates, J. R., Shotgun Proteomics and Biomarker Discovery. *Dis. Markers* **2002,** *18* (2), 99-105.

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

32. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R., Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999,** *17* (10), 994-999.

33. Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A.-J. L.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman, C., Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography−Tandem Mass Spectrometry. *J. Proteome Res.* **2010,** *9* (2), 761-776.

34. Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R., Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2018,** *14* (8), e8126.

35. Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R., Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012,** *11* (6), O111.016717.

36. Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R., OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotechnol.* **2014,** *32* (3), 219-223.

37. Kawashima, Y.; Watanabe, E.; Umeyama, T.; Nakajima, D.; Hattori, M.; Honda, K.; Ohara, O., Optimization of Data-Independent Acquisition Mass Spectrometry for Deep and Highly Sensitive Proteomic Analysis. *Int. J. Mol. Sci.* **2019,** *20* (23), 5932.

38. Venable, J. D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R., Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **2004,** *1* (1), 39-45.

39. Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M., Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proceedings of the National Academy of Sciences* **2007,** *104* (14), 5860-5865.

40. Marx, V., Targeted proteomics. *Nat Methods* **2013,** *10*, 19–22.

41. Prakash, A.; Tomazela, D. M.; Frewen, B.; Maclean, B.; Merrihew, G.; Peterman, S.; Maccoss, M. J., Expediting the Development of Targeted SRM Assays: Using Data from Shotgun Proteomics to Automate Method Development. *J. Proteome Res.* **2009,** *8* (6), 2733-2739.

42. Doerr, A., Targeting with PRM. *Nat. Methods* **2012,** *9* (10), 950-950.

43. Doerr, A., Mass spectrometry–based targeted proteomics. *Nat. Methods* **2013,** *10* (1), 23-23.

44. Kiyonami, R.; Schoen, A.; Prakash, A.; Peterman, S.; Zabrouskov, V.; Picotti, P.; Aebersold, R.; Huhmer, A.; Domon, B., Increased Selectivity, Analytical Precision, and Throughput in Targeted Proteomics. *Mol. Cell. Proteomics* **2011,** *10* (2), S1-S11.

45. Escher, C.; Reiter, L.; Maclean, B.; Ossola, R.; Herzog, F.; Chilton, J.; Maccoss, M. J.; Rinner, O., Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **2012,** *12* (8), 1111-1121.

46. Lange, V.; Picotti, P.; Domon, B.; Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2008,** *4* (1), 222.

47. Gallien, S.; Duriez, E.; Demeure, K.; Domon, B., Selectivity of LC-MS/MS analysis: Implication for proteomics experiments. *J. Proteomics* **2013,** *81*, 148-158.

48. Hüttenhain, R.; Surinova, S.; Ossola, R.; Sun, Z.; Campbell, D.; Cerciello, F.; Schiess, R.; Bausch-Fluck, D.; Rosenberger, G.; Chen, J.; Rinner, O.; Kusebauch, U.; Hajdúch, M.; Moritz, R. L.; Wollscheid, B.; Aebersold, R., N-Glycoprotein SRMAtlas. *Mol. Cell. Proteomics* **2013,** *12* (4), 1005-1016.

49. Krokhin, O. V., Sequence-Specific Retention Calculator. Algorithm for Peptide Retention Prediction in Ion-Pair RP-HPLC: Application to 300- and 100-Å Pore Size C18 Sorbents. *Anal. Chem.* **2006,** *78* (22), 7785-7795.

50. Gallien, S.; Peterman, S.; Kiyonami, R.; Souady, J.; Duriez, E.; Schoen, A.; Domon, B., Highly multiplexed targeted proteomics using precise control of peptide retention time. *Proteomics* **2012,** *12* (8), 1122-1133.

51. Whiteaker, J. R.; Lin, C.; Kennedy, J.; Hou, L.; Trute, M.; Sokal, I.; Yan, P.; Schoenherr, R. M.; Zhao, L.; Voytovich, U. J.; Kelly-Spratt, K. S.; Krasnoselsky, A.; Gafken, P. R.; Hogan, J. M.; Jones, L. A.; Wang, P.; Amon, L.; Chodosh, L. A.; Nelson, P. S.; Mcintosh, M. W.; Kemp, C. J.; Paulovich, A. G., A targeted proteomics–based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* **2011,** *29* (7), 625-634.

52. Reiter, L.; Rinner, O.; Picotti, P.; Hüttenhain, R.; Beck, M.; Brusniak, M.-Y.; Hengartner, M. O.; Aebersold, R., mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* **2011,** *8* (5), 430-435.

53. Rauniyar, N., Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *Int. J. Mol. Sci.* **2015,** *16* (12), 28566-28581.

54. Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J., Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics* **2012,** *11* (11), 1475-1488.

55. Ronsein, G. E.; Pamir, N.; Von Haller, P. D.; Kim, D. S.; Oda, M. N.; Jarvik, G. P.; Vaisar, T.; Heinecke, J. W., Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *J. Proteomics* **2015,** *113*, 388-399.

56. Hoffman, M. A.; Fang, B.; Haura, E. B.; Rix, U.; Koomen, J. M., Comparison of Quantitative Mass Spectrometry Platforms for Monitoring Kinase ATP Probe Uptake in Lung Cancer. *J. Proteome Res.* **2018,** *17* (1), 63-75.

57. Nakamura, K.; Hirayama-Kurogi, M.; Ito, S.; Kuno, T.; Yoneyama, T.; Obuchi, W.; Terasaki, T.; Ohtsuki, S., Large-scale multiplex absolute protein quantification of

drug-metabolizing enzymes and transporters in human intestine, liver, and kidney microsomes by SWATH-MS: Comparison with MRM/SRM and HR-MRM/PRM. *Proteomics* **2016,** *16* (15-16), 2106-2117.

58. Kaake, R. M.; Wang, X.; Huang, L., Profiling of Protein Interaction Networks of Protein Complexes Using Affinity Purification and Quantitative Mass Spectrometry. *Mol. Cell. Proteomics* **2010,** *9* (8), 1650-1665.

59. Zhu, W.; Smith, J. W.; Huang, C.-M., Mass Spectrometry-Based Label-Free Quantitative Proteomics. *J. Biomed. Biotechnol.* **2010,** *2010*, 1-6.

60. Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **2002,** *1* (5), 376-386.

61. Florens, L.; Carozza, M.; Swanson, S.; Fournier, M.; Coleman, M.; Workman, J.; Washburn, M., Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **2006,** *40* (4), 303-311.

62. Liu, H.; Sadygov, R. G.; Yates, J. R., A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.* **2004,** *76* (14), 4193-4201.

63. Egertson, J. D.; Kuehn, A.; Merrihew, G. E.; Bateman, N. W.; Maclean, B. X.; Ting, Y. S.; Canterbury, J. D.; Marsh, D. M.; Kellmann, M.; Zabrouskov, V.; Wu, C. C.; Maccoss, M. J., Multiplexed MS/MS for improved data-independent acquisition. *Nat. Methods* **2013,** *10* (8), 744-746.

64. Collins, B. C.; Gillet, L. C.; Rosenberger, G.; Röst, H. L.; Vichalkovski, A.; Gstaiger, M.; Aebersold, R., Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat. Methods* **2013,** *10* (12), 1246-1253.

65. Levin, Y.; Schwarz, E.; Wang, L.; Leweke, F. M.; Bahn, S., Label-free LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples. *J. Sep. Sci.* **2007,** *30* (14), 2198-2203.

66. Shalit, T.; Elinger, D.; Savidor, A.; Gabashvili, A.; Levin, Y., MS1-Based Label-Free Proteomics Using a Quadrupole Orbitrap Mass Spectrometer. *J. Proteome Res.* **2015,** *14* (4), 1979-1986.

67. Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M., Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* **2014,** *13* (9), 2513-2526.

68. Neilson, K. A.; Ali, N. A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; Van Sluyter, S. C.; Haynes, P. A., Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* **2011,** *11* (4), 535-553.

69. Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T., Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences* **1999,** *96* (12), 6591-6596.

70. Pan, C.; Oda, Y.; Lankford, P. K.; Zhang, B.; Samatova, N. F.; Pelletier, D. A.; Harwood, C. S.; Hettich, R. L., Characterization of Anaerobic Catabolism of p-Coumarate in Rhodopseudomonas palustris by Integrating Transcriptomics and Quantitative Proteomics. *Mol. Cell. Proteomics* **2008,** *7* (5), 938-948.

71. Belnap, C. P.; Pan, C.; Denef, V. J.; Samatova, N. F.; Hettich, R. L.; Banfield, J. F., Quantitative proteomic analyses of the response of acidophilic microbial communities to different pH conditions. *The ISME Journal* **2011,** *5* (7), 1152-1161.

72. Patton, W. F., Detection technologies in proteome analysis. *J. Chromatogr. B* **2002,** *771* (1-2), 3-31.

73. Willard, M.; Cowan, W. M.; Vagelos, P. R., The Polypeptide Composition of Intra-axonally Transported Proteins: Evidence for Four Transport Velocities. *Proceedings of the National Academy of Sciences* **1974,** *71* (6), 2183-2187.

74. Yeargin, J.; Haas, M., Elevated levels of wild-type p53 induced by radiolabeling of cells leads to apoptosis or sustained growth arrest. *Curr. Biol.* **1995,** *5* (4), 423-431.

75. Beynon, R. J.; Pratt, J. M., Metabolic Labeling of Proteins for Proteomics. *Mol. Cell. Proteomics* **2005,** *4* (7), 857-872.

76. Hilger, M.; Mann, M., Triple SILAC to Determine Stimulus Specific Interactions in the Wnt Pathway. *J. Proteome Res.* **2012,** *11* (2), 982-994.

77. Mann, M., Functional and quantitative proteomics using SILAC. *Nature Reviews Molecular Cell Biology* **2006,** *7* (12), 952-958.

78. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **2007,** *389* (4), 1017-1031.

79. Leitner, A.; Lindner, W., Current chemical tagging strategies for proteome analysis by mass spectrometry. *J. Chromatogr. B* **2004,** *813* (1-2), 1-26.

80. Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J.-C., Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. *Anal. Chem.* **2008,** *80* (8), 2921-2931.

81. Wiese, S.; Reidegeld, K. A.; Meyer, H. E.; Warscheid, B., Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics* **2007,** *7* (3), 340-350.

82. Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **2004,** *3* (12), 1154-1169.

83. Mcalister, G. C.; Huttlin, E. L.; Haas, W.; Ting, L.; Jedrychowski, M. P.; Rogers, J. C.; Kuhn, K.; Pike, I.; Grothe, R. A.; Blethrow, J. D.; Gygi, S. P., Increasing the Multiplexing Capacity of TMTs Using Reporter Ion Isotopologues with Isobaric Masses. *Anal. Chem.* **2012,** *84* (17), 7469-7478.

84. Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C.; Berfelde, A.; Martinez-Pinna, R.; Farztdinov, V.; Jung, S.; Kuhn, K.; Pike, I., TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Anal. Chem.* **2019,** *91* (24), 15941-15950.

85. Savitski, M. M.; Mathieson, T.; Zinn, N.; Sweetman, G.; Doce, C.; Becher, I.; Pachl, F.; Kuster, B.; Bantscheff, M., Measuring and Managing Ratio Compression for Accurate iTRAQ/TMT Quantification. *J. Proteome Res.* **2013,** *12* (8), 3586-3598.

86. Mcalister, G. C.; Nusinow, D. P.; Jedrychowski, M. P.; Wühr, M.; Huttlin, E. L.; Erickson, B. K.; Rad, R.; Haas, W.; Gygi, S. P., MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. *Anal. Chem.* **2014,** *86* (14), 7150-7158.

87. Ting, L.; Rad, R.; Gygi, S. P.; Haas, W., MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **2011,** *8* (11), 937-940.

88. Hogrebe, A.; Von Stechow, L.; Bekker-Jensen, D. B.; Weinert, B. T.; Kelstrup, C. D.; Olsen, J. V., Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nature Communications* **2018,** *9* (1).

89. Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* **2012,** *404* (4), 939-965.

90. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences* **2003,** *100* (12), 6940-6945.

91. Mayya, V.; Rezual, K.; Wu, L.; Fong, M. B.; Han, D. K., Absolute Quantification of Multisite Phosphorylation by Selective Reaction Monitoring Mass Spectrometry. *Mol. Cell. Proteomics* **2006,** *5* (6), 1146-1157.

92. Carvajal-Hausdorf, D. E.; Schalper, K. A.; Neumeister, V. M.; Rimm, D. L., Quantitative measurement of cancer tissue biomarkers in the lab and in the clinic. *Lab. Invest.* **2015,** *95* (4), 385-396.

93. Keshishian, H.; Addona, T.; Burgess, M.; Mani, D. R.; Shi, X.; Kuhn, E.; Sabatine, M. S.; Gerszten, R. E.; Carr, S. A., Quantification of Cardiovascular Biomarkers in Patient Plasma by Targeted Mass Spectrometry and Stable Isotope Dilution. *Mol. Cell. Proteomics* **2009,** *8* (10), 2339-2349.

94. Stahl-Zeng, J.; Lange, V.; Ossola, R.; Eckhardt, K.; Krek, W.; Aebersold, R.; Domon, B., High Sensitivity Detection of Plasma Proteins by Multiple Reaction Monitoring of N-Glycosites. *Mol. Cell. Proteomics* **2007,** *6* (10), 1809-1817.

95. Moolenaar, W. H., Lysophosphatidic Acid, a Multifunctional Phospholipid Messenger. *J. Biol. Chem.* **1995,** *270* (22), 12949-12952.

96. Tigyi, G.; Parrill, A. L., Molecular mechanisms of lysophosphatidic acid action. *Prog. Lipid Res.* **2003,** *42* (6), 498-526.

97. Mills, G. B.; Moolenaar, W. H., The emerging role of lysophosphatidic acid in cancer. *Nature Reviews Cancer* **2003,** *3* (8), 582-591.

98. Yung, Y. C.; Stoddard, N. C.; Chun, J., LPA receptor signaling: pharmacology, physiology, and pathophysiology. *J. Lipid Res.* **2014,** *55* (7), 1192-1214.

99. Tokumura, A., Metabolic pathways and physiological and pathological significances of lysolipid phosphate mediators. *J. Cell. Biochem.* **2004,** *92* (5), 869-881.

100. Aoki, J.; Taira, A.; Takanezawa, Y.; Kishi, Y.; Hama, K.; Kishimoto, T.; Mizuno, K.; Saku, K.; Taguchi, R.; Arai, H., Serum Lysophosphatidic Acid Is Produced through Diverse Phospholipase Pathways. *J. Biol. Chem.* **2002,** *277* (50), 48737-48744.

101. Choi;, J. W.; Herr;, D. R.; Noguchi;, K.; Yung;, Y. C.; Lee;, C.-W.; Mutoh;, T.; Lin;, M.-E.; Teo;, S. T.; Park;, K. E.; Mosley;, A. N.; Chun, J., LPA Receptors: Subtypes and Biological Actions. *Annu. Rev. Pharmacol. Toxicol* **2010,** *50*, 157–186.

102. Lin, M.-E.; Herr, D. R.; Chun, J., Lysophosphatidic acid (LPA) receptors: Signaling properties and disease relevance. *Prostaglandins Other Lipid Mediat.* **2010,** *91* (3-4), 130-138.

103. Hecht, J. H.; Weiner, J. A.; Post, S. R.; Chun, J., Ventricular zone gene-1 (vzg-1) encodes a lysophosphatidic acid receptor expressed in neurogenic regions of the developing cerebral cortex. *J. Cell Biol.* **1996,** *135* (4), 1071-1083.

104. An, S.; Bleu, T.; Hallmark, O. G.; Goetzl, E. J., Characterization of a Novel Subtype of Human G Protein-coupled Receptor for Lysophosphatidic Acid. *J. Biol. Chem.* **1998,** *273* (14), 7906-7910.

105. Im, D.-S.; Heise, C. E.; Ancellin, N.; O'Dowd, B. F.; Shei, G.-J.; Heavens, R. P.; Rigby, M. R.; Hla, T.; Mandala, S.; Mcallister, G.; George, S. R.; Lynch, K. R., Characterization of a Novel Sphingosine 1-Phosphate Receptor, Edg-8. *J. Biol. Chem.* **2000,** *275* (19), 14281-14286.

106. Bandoh, K.; Aoki, J.; Hosono, H.; Kobayashi, S.; Kobayashi, T.; Murakami-Murofushi, K.; Tsujimoto, M.; Arai, H.; Inoue, K., Molecular Cloning and Characterization of a Novel Human G-protein-coupled Receptor, EDG7, for Lysophosphatidic Acid. *J. Biol. Chem.* **1999,** *274* (39), 27776-27785.

107. Kihara, Y.; Maceyka, M.; Spiegel, S.; Chun, J., Lysophospholipid receptor nomenclature review: IUPHAR Review 8. *Br. J. Pharmacol.* **2014,** *171* (15), 3575-3594.

108. Noguchi, K.; Ishii, S.; Shimizu, T., Identification of p2y9/GPR23 as a Novel G Protein-coupled Receptor for Lysophosphatidic Acid, Structurally Distant from the Edg Family. *J. Biol. Chem.* **2003,** *278* (28), 25600-25606.

109. Lee, Z.; Cheng, C.-T.; Zhang, H.; Subler, M. A.; Wu, J.; Mukherjee, A.; Windle, J. J.; Chen, C.-K.; Fang, X., Role of LPA4/p2y9/GPR23 in Negative Regulation of Cell Motility. *Mol. Biol. Cell* **2008,** *19* (12), 5435-5445.

110. Yanagida, K.; Ishii, S., Non-Edg family LPA receptors: the cutting edge of LPA research. *J. Biochem.* **2011,** *150* (3), 223-232.

111. Shen, S.; Li, X.-F.; Cullen, W. R.; Weinfeld, M.; Le, X. C., Arsenic Binding to Proteins. *Chem. Rev.* **2013,** *113* (10), 7769-7792.

112. Saint-Jacques, N.; Parker, L.; Brown, P.; Dummer, T. J., Arsenic in drinking water and urinary tract cancers: a systematic review of 30 years of epidemiological evidence. *Environ. Health* **2014,** *13* (1), 44.

113. Yu, H.-S.; Liao, W.-T.; Chai, C.-Y., Arsenic Carcinogenesis in the Skin. *J. Biomed. Sci.* **2006,** *13* (5), 657-666.

114. Wild, P.; Bourgkard, E.; Paris, C., Lung Cancer and Exposure to Metals: The Epidemiological Evidence. In *Methods Mol. Biol.*, Humana Press: 2009; pp 139-167.

115. Jomova, K.; Jenisova, Z.; Feszterova, M.; Baros, S.; Liska, J.; Hudecova, D.; Rhodes, C. J.; Valko, M., Arsenic: toxicity, oxidative stress and human disease. *J. Appl. Toxicol.* **2011**, n/a-n/a.

116. Avram, S.; Udrea, A. M.; Negrea, A.; Ciopec, M.; Duteanu, N.; Postolache, C.; Duda-Seiman, C.; Duda-Seiman, D.; Shaposhnikov, S., Prevention of Deficit in Neuropsychiatric Disorders through Monitoring of Arsenic and Its Derivatives as Well as Through Bioinformatics and Cheminformatics. *Int. J. Mol. Sci.* **2019,** *20* (8), 1804.

117. Yoshinaga-Sakurai, K.; Shinde, R.; Rodriguez, M.; Rosen, B. P.; El-Hage, N., Comparative Cytotoxicity of Inorganic Arsenite and Methylarsenite in Human Brain Cells. *ACS Chem. Neurosci.* **2020,** *11* (5), 743-751.

118. Chen, C.-J.; Wang, S.-L.; Chiou, J.-M.; Tseng, C.-H.; Chiou, H.-Y.; Hsueh, Y.-M.; Chen, S.-Y.; Wu, M.-M.; Lai, M.-S., Arsenic and diabetes and hypertension in human populations: A review. *Toxicol. Appl. Pharmacol.* **2007,** *222* (3), 298-304.

119. States, J. C.; Srivastava, S.; Chen, Y.; Barchowsky, A., Arsenic and Cardiovascular Disease. *Toxicol. Sci.* **2009,** *107* (2), 312-323.

120. Navas-Acien, A.; Sharrett, A. R.; Silbergeld, E. K.; Schwartz, B. S.; Nachman, K. E.; Burke, T. A.; Guallar, E., Arsenic Exposure and Cardiovascular Disease: A Systematic Review of the Epidemiologic Evidence. *Am. J. Epidemiol.* **2005,** *162* (11), 1037-1049.

121. Ferzand, R.; Ali Gadahi, J.; Saleha, S.; Ali, Q., Histological and Haematological Disturbance Caused by Arsenic Toxicity in Mice Model. *Pak. J. Biol. Sci.* **2008,** *11* (11), 1405-1413.

122. Jeanne, M.; Lallemand-Breitenbach, V.; Ferhi, O.; Koken, M.; Le Bras, M.; Duffort, S.; Peres, L.; Berthier, C.; Soilihi, H.; Raught, B.; De Thé, H., PML/RARA Oxidation and Arsenic Binding Initiate the Antileukemia Response of As2O3. *Cancer Cell* **2010,** *18* (1), 88-98.

123. Huang, C.; Yin, Q.; Zhu, W.; Yang, Y.; Wang, X.; Qian, X.; Xu, Y., Highly Selective Fluorescent Probe for Vicinal-Dithiol-Containing Proteins and In Situ Imaging in Living Cells. *Angew. Chem. Int. Ed.* **2011,** *50* (33), 7551-7556.

124. Shi, W.; Dong, J.; Scott, R. A.; Ksenzenko, M. Y.; Rosen, B. P., The Role of Arsenic-Thiol Interactions in Metalloregulation of the ars Operon. *J. Biol. Chem.* **1996,** *271* (16), 9291-9297.

125. Hughes, M. F., Arsenic toxicity and potential mechanisms of action. *Toxicol. Lett.* **2002,** *133* (1), 1-16.

126. Kapahi, P.; Takahashi, T.; Natoli, G.; Adams, S. R.; Chen, Y.; Tsien, R. Y.; Karin, M., Inhibition of NF-κB Activation by Arsenite through Reaction with a Critical Cysteine in the Activation Loop of IκB Kinase. *J. Biol. Chem.* **2000,** *275* (46), 36062-36066.

127. Andrew, A. S.; Karagas, M. R.; Hamilton, J. W., Decreased DNA repair gene expression among individuals exposed to arsenic in United States drinking water. *Int. J. Cancer* **2003,** *104* (3), 263-268.

128. Ordóñez, E.; Thiyagarajan, S.; Cook, J. D.; Stemmler, T. L.; Gil, J. A.; Mateos, L. M.; Rosen, B. P., Evolution of Metal(loid) Binding Sites in Transcriptional Regulators. *J. Biol. Chem.* **2008,** *283* (37), 25706-25714.

129. Lu, M.; Wang, H.; Li, X.-F.; Lu, X.; Cullen, W. R.; Arnold, L. L.; Cohen, S. M.; Le, X. C., Evidence of Hemoglobin Binding to Arsenic as a Basis for the Accumulation of Arsenic in Rat Blood. *Chem. Res. Toxicol.* **2004,** *17* (12), 1733-1742.

130. Hoffman, R. D.; Lane, M. D., Iodophenylarsine oxide and arsenical affinity chromatography: new probes for dithiol proteins. Application to tubulins and to components of the insulin receptor-glucose transporter signal transduction pathway. *J. Biol. Chem.* **1992,** *267* (20), 14005-14011.

131. Menzel, D. B.; Hamadeh, H. K.; Lee, E.; Meacher, D. M.; Said, V.; Rasmussen, R. E.; Greene, H.; Roth, R. N., Arsenic binding proteins from human lymphoblastoid cells. *Toxicol. Lett.* **1999,** *105* (2), 89-101.

132. Moaddel, R.; Sharma, A.; Huseni, T.; Jones, G. S.; Hanson, R. N.; Loring, R. H., Novel Biotinylated Phenylarsonous Acids as Bifunctional Reagents for Spatially Close Thiols: Studies on Reduced Antibodies and the Agonist Binding Site of ReducedTorpedoNicotinic Receptors. *Bioconjugate Chem.* **1999,** *10* (4), 629-637.

133. Lin, C.-H.; Huang, C.-F.; Chen, W.-Y.; Chang, Y.-Y.; Ding, W.-H.; Lin, M.-S.; Wu, S.-H.; Huang, R.-N., Characterization of the Interaction of Galectin-1 with Sodium Arsenite. *Chem. Res. Toxicol.* **2006,** *19* (3), 469-474.

134. Chang, K. N.; Lee, T. C.; Tam, M. F.; Chen, Y. C.; Lee, L. W.; Lee, S. Y.; Lin, P. J.; Huang, R. N., Identification of galectin I and thioredoxin peroxidase II as two arsenic-binding proteins in Chinese hamster ovary cells. *Biochem. J* **2003,** *371* (2), 495-503.

135. Zhang, H.-N.; Yang, L.; Ling, J.-Y.; Czajkowsky, D. M.; Wang, J.-F.; Zhang, X.-W.; Zhou, Y.-M.; Ge, F.; Yang, M.-K.; Xiong, Q.; Guo, S.-J.; Le, H.-Y.; Wu, S.-F.; Yan, W.; Liu, B.; Zhu, H.; Chen, Z.; Tao, S.-C., Systematic identification of arsenic-

binding proteins reveals that hexokinase-2 is inhibited by arsenic. *Proceedings of the National Academy of Sciences* **2015,** *112* (49), 15084-15089.

136. Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M., Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Sci* **2009,** *325* (5942), 834-840.

137. Yan, X.; Li, J.; Liu, Q.; Peng, H.; Popowich, A.; Wang, Z.; Li, X.-F.; Le, X. C., p-Azidophenylarsenoxide: An Arsenical "Bait" for the In Situ Capture and Identification of Cellular Arsenic-Binding Proteins. *Angew. Chem. Int. Ed.* **2016,** *55* (45), 14051-14056.

138. Zhang, X.; Yang, F.; Shim, J.-Y.; Kirk, K. L.; Anderson, D. E.; Chen, X., Identification of arsenic-binding proteins in human breast cancer cells. *Cancer Lett.* **2007,** *255* (1), 95-106.

139. Patricelli, M. P.; Szardenings, A. K.; Liyanage, M.; Nomanbhoy, T. K.; Wu, M.; Weissig, H.; Aban, A.; Chun, D.; Tanner, S.; Kozarich, J. W., Functional Interrogation of the Kinome Using Nucleotide Acyl Phosphates. *Biochemistry* **2007,** *46* (2), 350-358.

140. Qiu, H.; Wang, Y., Probing Adenosine Nucleotide-Binding Proteins with an Affinity-Labeled Nucleotide Probe and Mass Spectrometry. *Anal. Chem.* **2007,** *79* (15), 5547-5556.

141. Xiao, Y.; Guo, L.; Jiang, X.; Wang, Y., Proteome-Wide Discovery and Characterizations of Nucleotide-Binding Proteins with Affinity-Labeled Chemical Probes. *Anal. Chem.* **2013,** *85* (6), 3198-3206.

142. G;, M.; DB;, W.; R;, M.; T;, H.; S., S., The protein kinase complement of the human genome. *Sci* **2002,** *298* (5600), 1912-34.

143. Blume-Jensen, P.; Hunter, T., Oncogenic kinase signalling. *Nature* **2001,** *411* (6835), 355-365.

144. Wu, P.; Nielsen, T. E.; Clausen, M. H., FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* **2015,** *36* (7), 422-439.

145. Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P., A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008,** *26* (1), 127-132.

146. Fabian, M. A.; Biggs, W. H.; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A.

G.; Lélias, J.-M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J., A small molecule–kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **2005,** *23* (3), 329-336.

147. Davis, M. I.; Hunt, J. P.; Herrgard, S.; Ciceri, P.; Wodicka, L. M.; Pallares, G.; Hocker, M.; Treiber, D. K.; Zarrinkar, P. P., Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2011,** *29* (11), 1046-1051.

148. Bagshaw, C. R., ATP analogues at a glance. *J. Cell Sci.* **2001,** *114* (3), 459-460.

149. Allen, J. J.; Li, M.; Brinkworth, C. S.; Paulson, J. L.; Wang, D.; Hübner, A.; Chou, W.-H.; Davis, R. J.; Burlingame, A. L.; Messing, R. O.; Katayama, C. D.; Hedrick, S. M.; Shokat, K. M., A semisynthetic epitope for kinase substrates. *Nat. Methods* **2007,** *4*, 511-6.

150. Hertz, N. T.; Berthet, A.; Sos, M. L.; Thorn, K. S.; Burlingame, A. L.; Nakamura, K.; Shokat, K. M., A neo-substrate that amplifies catalytic activity of parkinson's-disease-related kinase PINK1. *Cell* **2013,** *154*, 737-747.

Chapter 2 Chemical Proteomic Profiling of Lysophosphatic Acid-binding Proteins

2.1 Introduction

Lysophosphatidic acid (LPA) is a ubiquitous phospholipid that exists in all eukaryotic tissues and plays critical roles in various cellular processes. The impact on these cellular processes is due in part to the heterogeneity of LPA-binding protein subtypes, expression patterns, and effector pathways.[1,2] LPA can bind to and activate some G protein–coupled receptors (GPCRs) and thus act as an extracellular signaling molecule.[3] It was also reported that aberrant LPA signaling is associated with many types of cancer.[4–7] The ability of LPA signaling to impact such a wide array of important pathways renders LPA receptors an attractive targets for drug discovery.

To date, six cell-surface LPA receptors have been discovered, i.e. LPA1–LPA6.[8–13] Additional putative LPA receptors, i.e. GPR87 and P2Y10,[14–16] were proposed and await further validation. However, much remains to be learnt about the role of LPA signaling, where LPA's functions still remains largely unknown and some of its functions are not associated with known LPA receptors.[17,18] Thus, there are likely many yet-identified LPA-binding proteins. Identification and quantification of LPA-binding proteins at the entire proteome scale is important for investigating biological events underlying the complicated signaling network comprised of LPA receptors and down-stream effector proteins. Recent advances in mass spectrometry (MS) have greatly facilitated protein identification and

quantification in complex sample matrices;[19] however, proteomic studies of specific families of proteins by MS remains a challenge.

Here, we utilize a chemical proteomic method [20,21] – which has been widely used to study functional subgroups of proteins – to selectively label, identify, and characterize potential LPA-binding proteins in the entire human proteome. By employing this probe, in conjunction with an LC-MS/MS-based proteomic workflow, we were able to identify a number of previously unknown putative LPA-binding proteins. Along with stable isotope labeling by amino acids in cell culture (SILAC),[22] we further assessed the specificity of LPA−protein interactions at the entire proteome scale. We also confirmed the capabilities of some of these proteins in binding directly with LPA.

2.2 Experimental Procedures

2.2.1 Chemical Synthesis of LPA affinity Probe

On the basis of our previously published method for the synthesis of a desthiobiotin-ATP affinity probe,[23] we designed a desthiobiotin-conjugated LPA acylphosphate probe. Considering the generally higher hydrophobicity of LPA relative to ATP, we sought to increase the aqueous solubility of the probe by incorporating a polyethylene glycol (PEG) moiety.

The probes (Figure 2.1b) were prepared employing a two-step procedure (Figure 2.1a),[24] first by conjugating desthiobiotin with γ-aminobutyric acid or amino-PEG_4-acid, and then through reacting the resulting conjugate with LPA. Desthiobiotin (10 mg) and tri-*n*-butylamine (11 μL) were dissolved in a 1-mL solvent mixture of ice-cold CH_2Cl_2 and DMF

(4:1, v/v). After stirring at 0 °C for 5 min, ethyl chloroformate (6 μL) was added under an argon atmosphere, and the solution was maintained at 0 °C with stirring for 10 min, followed by stirring at room temperature for another 60 min. To the above reaction mixture was added LPA (30 mg) and the reaction was continued at room temperature under an argon atmosphere for 18 hr. Dichloromethane was removed by purging the reaction mixture with argon. Crude products were re-dissolved in water, loaded onto a Sep-pak C18 gravity column, and eluted with a gradient of 0-70% acetonitrile in 50 mM triethylammonium acetate (TEAA, pH 6.8). Appropriate elute fractions were pooled, lyophilized, and stored at -80°C. The structures of the products were confirmed by ESI-MS and MS/MS (Figure. 2.3 and Figure 2.4).

2.2.2 Cell Culture, Lysate Preparation and Probe Labeling

HEK293T, HeLa, MCF-7, GM04429, K562 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and penicillin (100 IU/mL). Approximately 2×10^7 cells were harvested, washed three times with cold PBS for three times and stored at -80 °C until use.

Cell lysis, endogenous small molecule removal and probe labeling were conducted in a similar way as described previously for nucleotide-binding proteins.[23] Briefly, the cells were lysed in a 1-mL lysis buffer, which contained 0.7% CHAPS, 50 mM HEPES (pH 7.4), 0.5 mM EDTA, and 100 mM NaCl with 1:100 protease inhibitor cocktail, at 0 °C for 30 min, followed by centrifugation at 16000g and 4 °C for 30 min. Gel filtration with a NAP-25 columns (Amersham Biosciences) was used to remove free endogenous

phospholipids and other small molecules from the resulting supernatants. Fractions were eluted into a buffer containing 50 mM HEPES (pH 7.4), 75 mM NaCl, and 5% glycerol. The recovered proteins were quantified using Quick Start Bradford Protein Assay (Bio-Rad, Hercules, CA) and stored at −80 °C.

$MgCl_2$, $CaCl_2$ and $CuCl_2$ were added to the above-mentioned protein mixture until their final concentrations reached 50, 5, and 0.5 mM, respectively. To the solution was added 100 μM synthesized desthiobiotin-LPA affinity probe at room temperature with gentle shaking for 2 hr. The remaining probe was removed by buffer exchange with a 50 mM NH_4HCO_3 solution using Amicon Ultra-4 filter (10000 NMWL, Millipore).

After probe labeling, the cell lysate was subjected to filter-aided sample preparation (FASP), where an 8 M urea buffer was used for protein denaturation, dithiothreitol (DTT) and iodoacetamide (IAA) for cysteine reduction and alkylation. The resulting proteins were digested at 37 °C overnight with trypsin at an enzyme/substrate ratio of 1: 100 and the desthiobiotin-tagged peptides were further enriched by affinity pull-down through incubation with avidin-agarose beads at 25 °C for 1 h with gentle shaking. The agarose resin was washed sequentially with 1×PBS buffer (6 mL), 10×PBS (6 mL), and water (6 mL). The desthiobiotin-labeled peptides were eluted with 1% TFA in CH_3CN/H_2O (7:3 v/v). The resulting peptide mixture was analyzed by LC-MS/MS on an LTQ Orbitrap Velos mass spectrometer in the data-dependent acquisition (DDA) mode. The raw data were searched using MaxQuant (version 1.3.0.3) against the IPI Human proteome database v3.86.

2.2.3 Purification of Recombinant Annexin A5 and PGK1

Recombinant human Annexin A5 (#19961)[25] and PGK1 (#38071)[26] plasmids were bought from addgene, we engineered them into a PET28a plasmid containing a polyhistidine tag for protein purification. Constructs were transfected into BL21 *Escherichia coli* cells. The bacteria were grown in LB medium containing 50 μg/mL of kanamycin. Bacteria were grown to OD_{600} 0.6-0.8 at 37 °C with shaking in LB medium. IPTG was added until its final concentration reached 1 mM, and the culture was grown for an additional 16 h at 16 °C.

The cells were harvested by centrifugation at 7000 rpm at 4 °C, resuspended in 10 mM Tris-HCl (pH 7.2), and lysed by sonication on ice. The supernatant was collected after centrifugation. The crude cell extract, containing the desired His-tagged proteins, was loaded onto a Ni^{2+}-NTA column and washed with a buffer containing 25 mM imidazole. The desired proteins were eluted with a buffer containing 100-300 mM imidazole. The purity of the His-tagged proteins was verified by SDS-PAGE and visualized by Coomassie staining and followed by dialysis.

2.2.4 Recombinant human LPAR1

Plasmids for expressing HA-tagged full-length human LPAR1 was obtained from the DNASU plasmid bank. HEK293T cells, maintained in complete media, were transfected with each plasmid individually using PolyFect transfection reagent (Qiagen). After incubating for 24 hr, the medium was replaced with fresh complete media and the cells were collected.

2.2.5 Isothermal Titration Calorimetry (ITC) Analysis

The ITC analyses of Annexin A5 and PGK1 proteins were performed on a MicroCal iTC200 (Malvern, UK). In particular, the macromolecule solution was placed in the sample cell. The reference cell with water and the sample cell with 0.1 mM protein of interest were set at 25 °C. LPA (1 mM) was loaded into a syringe. The direct observable parameter measured in an ITC experiment is the time-dependent input of power required to maintain an equal temperature in the sample and reference cells. In lipid titration experiments, both the protein and LPA solutions were adjusted to have the same Na^+ concentration of 30 mM in 50 mM Tris buffer (pH = 7.5) with 5% glycerol. A series of small aliquots of ligand were injected into the protein solution with a 180 s spacing between neighboring injections, beginning with an initial 0.5 μL injection followed by 16 × 2.48 μL injections, and the solution was stirred at 1000 rpm throughout the experiment. Heat changes were measured. After the titration, the data were analyzed using MicroCal ITC-ORIGIN analysis software and fit with the one-set-of-sites model to calculate the dissociation constant and the error of the fit. The data represent at least two separate experiments from two separate preparations of each protein.

2.3 Results and Discussion

2.3.1 Design of the LPA-Affinity Probe

Previous structural studies showed that many lipid-binding proteins frequently carry one or more lysine residues at their ligand binding sites. For example, Lys34 in sphingosine 1-

phosphate receptor 1 is very close to sphingolipid-binding pocket,[27] and Lys61 in phosphatidylinositol transfer protein α (PITPα) plays an essential role in binding with phosphatidylinositol.[28] This lysine can react with an acyl phosphate group via a nucleophilic substitution reaction, producing a covalently modified protein.[24] In light of these previous findings, we developed an LPA analog, bearing an acyl phosphate moiety to target the lysine residue in the LPA-binding site (Figure 2.1).

The LPA-binding probes comprise of LPA, which targets the LPA-binding pocket of cellular proteins; an enrichment moiety, i.e. desthiobiotin, which facilitates downstream purification; and a linker (Figure 2.1b). When the LPA affinity probe interacts with LPA-binding proteins, the carbonyl carbon in the acylphosphate reacts with the lysine residue at the phospholipid-binding pocket to yield a stable amide bond, thereby simultaneously modifying the proteins with a desthiobiotin tag (Figure 2.2). After the probe-labeling reaction, we digested the protein mixture with trypsin and enriched the resulting desthiobiotin-labeled peptides using streptavidin beads. We used desthiobiotin because it binds less tightly to avidin agarose than biotin,[29] thereby allowing for better recovery of the modified peptides or proteins during the enrichment. We analyzed the affinity-purified peptides with the desthiobiotin tag by LC−MS/MS.

During the synthesis and purification of the LPA probe, we noted its weak UV absorption in the 220-300 nm range (LPA Abs ~200-210 nm);[30] therefore, we separated the reaction mixture by using a C18 Sep-pak column, monitored the fractions by TLC, and identified the desired probe-containing fractions by ESI-MS analysis. Upon optimizing the elution conditions, we recovered the probe with an approximately 85% purity. We

subsequently conducted an initial labeling experiment with 10, 25, 100, 200 μM desthiobiotin-C3-affinity probe and 1 mg variant cell lysates; this experiment led to the identification of only a small number of LPA-binding proteins. We reasoned that this might be attributed to the poor aqueous solubility of the probe. Thus, we replaced the C3 linker with a PEG4 linker to improve the aqueous solubility of the probe in aqueous solution since PEG provides greater water solubility.[31] Synthesis of the PEG4 probe was carried out in the same way as previously described, where both the C3 and PEG4 linkers carry an NH_2 and a COOH functional groups on their termini. The new PEG4-probe could allow for the identification of about three times as many putative LPA-binding proteins compared to the C3-probe. (Figure 2.5)

We sought to further optimize the binding conditions by employing different concentrations of the probe. We incubated 1 mg cell lysate with 10, 25, 50, 100 or 200 μM of the desthiobiotin-PEG_4-LPA affinity probe under various binding conditions. We identified the largest number of putative LPA-binding proteins from the whole cell lysate when using 200 μM probe with the exist of Mg^{2+}, Ca^{2+}, Cu^{2+}. As we decreased the LPA probe concentration to 100 μM, the total number of identified LPA-binding proteins decreased to about one fifth of the number when compared with the use of 200 μM LPA probe; however, the relatively high probe concentration may lead to diminished specificity, which could result in the identification of proteins arising from non-specific reactions. Hence, we decided to employ 100 μM probe for the subsequent experiments.

2.3.2 Global Profiling of LPA-binding Proteins from the Whole Human Proteome

We first used the LPA probes to label the lysate from two different cell lines (HEK293T, HeLa), digested the proteins with trypsin, enriched the desthiobiotin-labeled peptides from the tryptic digestion mixtures using high capacity streptavidin agarose beads, and analyzed the resulting purified peptides on an LTQ Oribitrap Velos mass spectrometer in the DDA mode. This method led to the identification of 790 unique desthiobiotin-conjugated peptides from 318 distinct proteins with the desthiobiotin-C_3-LPA probe, and 3486 labeled peptides from 940 proteins with the desthiobiotin-PEG_4-LPA probe. Relative to the C_3-LPA probe results, the total number of probe-labeled proteins increased by approximately 3-fold with the PEG_4-LPA probe affinity pull-down under otherwise identical conditions, suggesting that the introduction of a PEG linker improves the water solubility of the probe and thus augments the labeling efficiency. However, during this initial analysis, none of the known LPA receptors were identified; this could be due to the relatively low expression levels of LPA receptors in these cell lines. Thus, we transfected an expression plasmid for LPA receptor 1 (LPAR1) into HEK293T cells. With the ectopic expression, we successfully identified the desthiobiotin-PEG4-modified peptide from LPAR1. (Figure 2.6)

Due to the relatively high reactivity of the acyl phosphate probe, lysine residues not at the LPA binding sites may also be modified through non-specific electrostatic interactions. To further investigate the specific LPA-binding proteins at the entire proteome scale, we devised an affinity profiling strategy combined with a SILAC-based workflow. In this vein, as described in previous proteome-wide discovery of reactive cysteine-containing proteins and ATP-binding proteins interaction study,[20,23] binding of a protein with a specific

reactive group of the probe greatly increases the rate for the coupling reaction between the acyl phosphate moiety and the lysine residue at the binding site. Thus, different probe concentrations will exhibit distinct labeling behaviors of lysine residues, which reflect the binding specificities between LPA and target proteins.

We performed our SILAC experiment by allowing a low (10 μM) concentration of the LPA probe to react with light cell lysate and a high (100 μM) concentration of the LPA probe to react separately with the same amount of heavy cell lysate (forward experiment). We subsequently combined the two protein samples and digested the ensuing mixture with trypsin; the resulting desthiobiotin-labeled peptides were enriched with streptavidin beads. We also performed a reverse labeling experiment to minimize bias introduced by the labeling process. The affinity-purified peptides with the desthiobiotin tag were further analyzed by LC−MS/MS (Figure 2.7). We used peak intensity ratios to obtain accurate quantification results of light and heavy desthiobiotin-labeled peptides from the two experimental states to derive the LPA-binding affinity ratio, $R_{LPA10/1}$, which reflects the relative binding affinities of LPA toward specific lysine residues in individual proteins.

We assumed an $R_{LPA10/1}$ close to 1 will represent specific LPA-binding to lysine since even at a low probe concentration, the lysine residue at the LPA-binding site possesses hyper-reactivity and will be completely labeled. By contrast, lysines not involved with LPA binding will only be partially labeled because the limited amount of labeling reagent reacts preferentially with the lysine at the LPA-binding site, which will display a concentration-dependent increase in labeling efficiency, resulting in an $R_{LPA10/1} \gg 1$. We attempted to include all the putative LPA-binding proteins in our analysis with the use of lenient criteria:

$R_{LPA10/1} < 3$; with this criterion, we were able to quantify a total of 62 proteins, including 86 light or heavy desthiobiotin-modified lysine residues. Comparing our SILAC results with Gene Ontology analysis using DAVID, we observed many lipid-binding proteins exhibiting an $R_{LPA10/1} < 3$, while most ATP-binding proteins (e.g. creatine kinase B-type) had an $R_{LPA10/1} \gg 1$. This comparison indicates that specific labeling of lysine residues occurred mainly on lipid-binding proteins and non-specific labeling occurred more on ATP-binding proteins.

Together, our results revealed a large number of proteins displaying specific LPA-binding affinity (Figure 2.8) and demonstrated that our quantitative affinity profiling strategy could effectively eliminate false-positive targets. These features allow for application of our affinity reactive probes in protein-ligand binding studies.

2.3.3 Isothermal titration calorimetry analysis of LPA binding with Annexin A5 and PGK1

Based on the DAVID GO analysis, more than 50 targets predicted to be LPA-binding proteins from our quantitative affinity profiling results were not previously documented as LPA-binding proteins. To further explore these novel LPA-binding proteins, we studied several proteins that were identified as having related lipid-binding functions. A literature search revealed that some of these proteins may also be closely associated with LPA-binding.

Among them, annexin A5 (AnxA5) and Phosphoglycerate Kinase 1 (PGK1) presented high specificity with $R_{LPA10/1}$ values of 2.13 (take the average of its five quantified peptides) (Figure 2.9) and 1.71 (take the average of its four quantified peptides) (Figure 2.10),

respectively. These proteins also showed remarkable reproducibility in all the biological replicates. AnxA5 has been widely used to detect apoptotic cells owing to its high affinity for phosphatidylserine;[32] and it was also reported that AnxA5 can bind to lipopolysaccharide.[33] Previous studies suggest that AnxA5 has binding affinity for lipids with similar tail groups as LPA; therefore, we reason that LPA is likely a binding partner of AnxA5.

We next utilized isothermal titration calorimetry (ITC), which has been widely used for characterizing the interactions between proteins and small molecules,[34] to assess the direct interactions between LPA and recombinant AnxA5 and PGK1. To this end, we first purified the His-tagged AnxA5 and PGK1 and assessed its purity by SDS-PAGE analysis (Figure 2.11). Our results showed that PGK1 and AnxA5 bind to LPA with K_d values of 93.4 and 23 μM, respectively (Figure 2.12), supporting the direct binding of these two proteins to LPA. Hence, AnxA5 and PGK1 may play an important role in LPA signaling pathways.

2.4 Conclusions

Here, we report a strategy using the affinity chemical probe to enrich and identify putative LPA-binding proteins in the human proteome. We successfully synthesized the desthiobiotin-tagged LPA-affinity probe and used it to react with putative LPA-binding proteins, which, followed by tryptic digestion, affinity enrichment of the ensuing desthiobiotin-conjugated peptides, and LC-MS/MS analysis, allowed for the proteome-wide identification of LPA-binding proteins. A total of 1245 proteins with a desthiobiotin-

tag were identified and quantified from several different cell lines with the use of low milligram quantities of lysate. The method led to the identification of several previously reported lipid- or even LPA-interacting proteins. Our data show that LPA affinity probes can be employed as a useful reagent for the enrichment and subsequent detection of LPA-binding proteins from the human proteome, which may reveal new LPA signaling pathways.

This strategy can be combined with a quantitative method to characterize LPA−protein interactions and to identify specific LPA-binding proteins at the entire proteome level. The selectivities of LPA-binding proteins was systematically investigated by quantifying peak intensity ratios of desthiobiotin-labeled light and heavy tryptic peptides after reacting with our probes, demonstrating the potential of our affinity probes in quantitative proteomic analysis. Through ITC analysis, we validated that AnxA5 and PGK1 can bind directly with LPA, and measured their binding affinities. Moreover, the analytical method allows for identification of the specific lysine residues involved in binding with LPA, which may provide important knowledge for designing small molecule inhibitors for LPA-binding proteins and/or for designing mutant proteins for interrogating the biological functions of these proteins.

There are several advantages of this chemical proteomic approach. First, the method provides facile enrichment, identification and quantification of LPA-binding proteins. Second, the method allows for the identification of the specific lysine residues involved in binding with LPA, which may provide important knowledge for developing small molecule inhibitors for LPA-binding proteins and/or for designing mutant proteins for interrogating

the biological functions of these proteins. A limitation of the current probe resides in that metabolic labeling is required for the quantitative analysis of LPA-binding proteins. Nevertheless, this can be overcome by incorporating stable isotope-labeled linker into the probe, as what described previously for the ATP acyl phosphate probes.[1]

It can be envisaged that the method can be adapted for studying other lipid-binding proteins at the entire proteome scale. In particular, a similar acyl phosphate probe approach can be employed for the enrichment and subsequent identification of interaction proteins for those phosphate lipids carrying a terminal phosphate group.

References

(1) Choi, J. W.; Herr, D. R.; Noguchi, K.; Yung, Y. C.; Lee, C.-W.; Mutoh, T.; Lin, M.-E.; Teo, S. T.; Park, K. E.; Mosley, A. N.; et al. LPA Receptors: Subtypes and Biological Actions. *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50* (1), 157–186. https://doi.org/10.1146/annurev.pharmtox.010909.105753.

(2) Moolenaar, W. H. Lysophosphatidic Acid, a Multifunctional Phospholipid Messenger. *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc. 1995, pp 12949–12952. https://doi.org/10.1074/jbc.270.22.12949.

(3) Ishii, I.; Fukushima, N.; Ye, X.; Chun, J. Lysophospholipid Receptors: Signaling and Biology. *Annu. Rev. Biochem.* **2004**, *73* (1), 321–354. https://doi.org/10.1146/annurev.biochem.73.011303.073731.

(4) Lindholm, P. F.; Hwang, Y. S. LPA Increases Tumor Growth and Bone Destruction Through Enhancement of Osteoclastogenic Cytokines. *Anticancer Res.* **2016**, *36* (1), 61–70.

(5) Mills, G. B.; Moolenaar, W. H. The Emerging Role of Lysophosphatidic Acid in Cancer. *Nature Reviews Cancer*. August 2003, pp 582–591. https://doi.org/10.1038/nrc1143.

(6) Ha, J. H.; Radhakrishnan, R.; Jayaraman, M.; Yan, M.; Ward, J. D.; Fung, K. M.; Moxley, K.; Sood, A. K.; Isidoro, C.; Mukherjee, P.; et al. Lpa Induces Metabolic Reprogramming in Ovarian Cancer via a Pseudohypoxic Response. *Cancer Res.* **2018**, *78* (8), 1923–1934. https://doi.org/10.1158/0008-5472.CAN-17-1624.

(7) Wang, J.; Sun, Y.; Qu, J.; Yan, Y.; Yang, Y.; Cai, H. Roles of LPA Receptor Signaling in Breast Cancer. *Expert Rev. Mol. Diagn.* **2016**, *16* (10), 1103–1111. https://doi.org/10.1080/14737159.2016.1238763.

(8) Chun, J.; Hla, T.; Lynch, K. R.; Spiegel, S.; Moolenaar, W. H. International Union of Basic and Clinical Pharmacology. LXXVIII. Lysophospholipid Receptor Nomenclature. *Pharmacol. Rev.* **2010**, *62* (4), 579–587. https://doi.org/10.1124/pr.110.003111.

(9) Choi, J. W.; Chun, J. Lysophospholipids and Their Receptors in the Central Nervous System. *Biochim. Biophys. Acta* **2013**, *1831* (1), 20–32. https://doi.org/10.1016/j.bbalip.2012.07.015.

(10) Yanagida, K.; Kurikawa, Y.; Shimizu, T.; Ishii, S. Current Progress in Non-Edg Family LPA Receptor Research. *Biochim. Biophys. Acta* **2013**, *1831* (1), 33–41. https://doi.org/10.1016/j.bbalip.2012.08.003.

(11) Noguchi, K.; Ishii, S.; Shimizu, T. Identification of P2y9/GPR23 as a Novel G Protein-Coupled Receptor for Lysophosphatidic Acid, Structurally Distant from the Edg Family. *J. Biol. Chem.* **2003**, *278* (28), 25600–25606. https://doi.org/10.1074/jbc.M302648200.

(12) Lee, C.-W.; Rivera, R.; Gardell, S.; Dubin, A. E.; Chun, J. GPR92 as a New G12/13- and Gq-Coupled Lysophosphatidic Acid Receptor That Increases CAMP, LPA5. *J. Biol. Chem.* **2006**, *281* (33), 23589–23597. https://doi.org/10.1074/jbc.M603670200.

(13) Yanagida, K.; Masago, K.; Nakanishi, H.; Kihara, Y.; Hamano, F.; Tajima, Y.; Taguchi, R.; Shimizu, T.; Ishii, S. Identification and Characterization of a Novel Lysophosphatidic Acid Receptor, P2y5/LPA6. *J. Biol. Chem.* **2009**, *284* (26), 17731–17741. https://doi.org/10.1074/jbc.M808506200.

(14) Oka, S.; Ota, R.; Shima, M.; Yamashita, A.; Sugiura, T. GPR35 Is a Novel Lysophosphatidic Acid Receptor. *Biochem. Biophys. Res. Commun.* **2010**, *395* (2), 232–237. https://doi.org/10.1016/j.bbrc.2010.03.169.

(15) Tabata, K.; Baba, K.; Shiraishi, A.; Ito, M.; Fujita, N. The Orphan GPCR GPR87 Was Deorphanized and Shown to Be a Lysophosphatidic Acid Receptor. *Biochem. Biophys. Res. Commun.* **2007**, *363* (3), 861–866. https://doi.org/10.1016/j.bbrc.2007.09.063.

(16) Murakami, M.; Shiraishi, A.; Tabata, K.; Fujita, N. Identification of the Orphan GPCR, P2Y 10 Receptor as the Sphingosine-1-Phosphate and Lysophosphatidic Acid Receptor. *Biochem. Biophys. Res. Commun.* **2008**, *371* (4), 707–712. https://doi.org/10.1016/j.bbrc.2008.04.145.

(17) Yanagida, K.; Ishii, S. Non-Edg Family LPA Receptors: The Cutting Edge of LPA Research. *Journal of Biochemistry*. September 2011, pp 223–232. https://doi.org/10.1093/jb/mvr087.

(18) McIntyre, T. M.; Pontsler, A. V.; Silva, A. R.; St. Hilaire, A.; Xu, Y.; Hinshaw, J. C.; Zimmerman, G. A.; Hama, K.; Aoki, J.; Arai, H.; et al. Identification of an Intracellular Receptor for Lysophosphatidic Acid (LPA): LPA Is a Transcellular PPARγ Agonist. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (1), 131–136. https://doi.org/10.1073/pnas.0135855100.

(19) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science (80-.).* **2009**, *325* (5942), 834–840. https://doi.org/10.1126/science.1175371.

(20) Barglow, K. T.; Cravatt, B. F. Activity-Based Protein Profiling for the Functional

Annotation of Enzymes. *Nat. Methods* **2007**, *4* (10), 822–827. https://doi.org/10.1038/nmeth1092.

(21) Cravatt, B. F.; Wright, A. T.; Kozarich, J. W. Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry. *Annu. Rev. Biochem.* **2008**, *77* (1), 383–414. https://doi.org/10.1146/annurev.biochem.75.101304.124125.

(22) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **2002**, *1* (5), 376–386. https://doi.org/10.1074/mcp.M200025-MCP200.

(23) Xiao, Y.; Guo, L.; Wang, Y. Isotope-Coded ATP Probe for Quantitative Affinity Profiling of ATP-Binding Proteins. *Anal. Chem.* **2013**, *85* (15), 7478–7486. https://doi.org/10.1021/ac401415z.

(24) Qiu, H.; Wang, Y. Probing Adenosine Nucleotide-Binding Proteins with an Affinity-Labeled Nucleotide Probe and Mass Spectrometry. *Anal. Chem.* **2007**, *79* (15), 5547–5556. https://doi.org/10.1021/ac0622375.

(25) Wood, B. L.; Gibson, D. F.; Tait, J. F. Increased Erythrocyte Phosphatidylserine Exposure in Sickle Cell Disease: Flow-Cytometric Measurement and Clinical Associations. *Blood* **1996**, *88* (5), 1873–1880.

(26) Baltz, A. G.; Munschauer, M.; Schwanhäusser, B.; Vasile, A.; Murakawa, Y.; Schueler, M.; Youngs, N.; Penfold-Brown, D.; Drew, K.; Milek, M.; et al. The MRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol. Cell* **2012**, *46* (5), 674–690. https://doi.org/10.1016/j.molcel.2012.05.021.

(27) Hanson, M. A.; Roth, C. B.; Jo, E.; Griffith, M. T.; Scott, F. L.; Reinhart, G.; Desale, H.; Clemons, B.; Cahalan, S. M.; Schuerer, S. C.; et al. Crystal Structure of a Lipid G Protein-Coupled Receptor. *Science (80-.).* **2012**, *335* (6070), 851–855. https://doi.org/10.1126/science.1215904.

(28) Tilley, S. J.; Skippen, A.; Murray-Rust, J.; Swigart, P. M.; Stewart, A.; Morgan, C. P.; Cockcroft, S.; McDonald, N. Q. Structure-Function Analysis of Phosphatidylinositol Transfer Protein Alpha Bound to Human Phosphatidylinositol. *Structure* **2004**, *12* (2), 317–326. https://doi.org/10.1016/j.str.2004.01.013.

(29) Hirsch, J. D.; Eslamizar, L.; Filanoski, B. J.; Malekzadeh, N.; Haugland, R. P.; Beechem, J. M.; Haugland, R. P. Easily Reversible Desthiobiotin Binding to Streptavidin, Avidin, and Other Biotin-Binding Proteins: Uses for Protein Labeling, Detection, and Isolation. *Anal. Biochem.* **2002**, *308* (2), 343–357. https://doi.org/10.1016/s0003-2697(02)00201-4.

(30) Chen, Y. L.; Xu, Y. Determination of Lysophosphatidic Acids by Capillary Electrophoresis with Indirect Ultraviolet Detection. *J. Chromatogr. B Biomed. Sci. Appl.* **2001**, *753* (2), 355–363. https://doi.org/10.1016/S0378-4347(00)00582-X.

(31) Polyethylene Glycol (PEG) and Pegylation of Proteins - US.

(32) Reviakine, I.; Bergsma-Schutter, W.; Brisson, A. Growth of Protein 2-D Crystals on Supported Planar Lipid Bilayers Imaged in Situ by AFM. *J. Struct. Biol.* **1998**, *121* (3), 356–362. https://doi.org/10.1006/jsbi.1998.4003.

(33) Rand, J. H.; Wu, X. X.; Lin, E. Y.; Griffel, A.; Gialanella, P.; McKitrick, J. C. Annexin A5 Binds to Lipopolysaccharide and Reduces Its Endotoxin Activity. *MBio* **2012**, *3* (2). https://doi.org/10.1128/mBio.00292-11.

(34) Leavitt, S.; Freire, E. Direct Measurement of Protein Binding Energetics by Isothermal Titration Calorimetry. *Current Opinion in Structural Biology*. Elsevier Ltd September 1, 2001, pp 560–566. https://doi.org/10.1016/S0959-440X(00)00248-7.

Table 2. 1 Candidate LPA-binding proteins:ANXA5, PGK1, PPIA. Measured $R_{LPA10/1}$ ratio of peptides from 293T SILAC cell lysates with low (10 μM) and high (100 μM) concentrations of LPA affinity probe.

Protein Name	Peptide Sequence	Average Ratio (LPA 10/1)	S. D.
ANXA5	GDTSGDYK(0.999)K(0.001)	1.51	1.22
ANXA5	HALK(1)GAGTNEK	1.67	0.61
ANXA5	LYDAYELK(1)HALK	1.72	0.11
ANXA5	SELTGK(1)FEK	2.00	0.78
ANXA5	K(1)ALLLLCGEDD	2.74	1.64
PGK1	FHVEEEGK(1)GK	1.09	0.18
PGK1	GTK(1)ALMDEVVK(1)ATSR	1.22	1.58
PGK1	DLMSK(1)AEK	1.57	0.10
PGK1	K(1)YAEAVTR	1.86	1.26
PGK1	GTK(1)ALMDEVVK(1)ATSR	1.88	1.64
PGK1	IVKDLMSK	2.34	2.37
PGK1	SLLGK(1)DVLFLK	2.64	0.98
PPIA	GFGYK(1)GSCFHR	1.49	0.64
PPIA	ALSTGEK(1)GFGYK	1.70	1.21

Figure 2. 1 a) Synthetic scheme for the LPA probe. b) The chemical structures of the LPA probes.

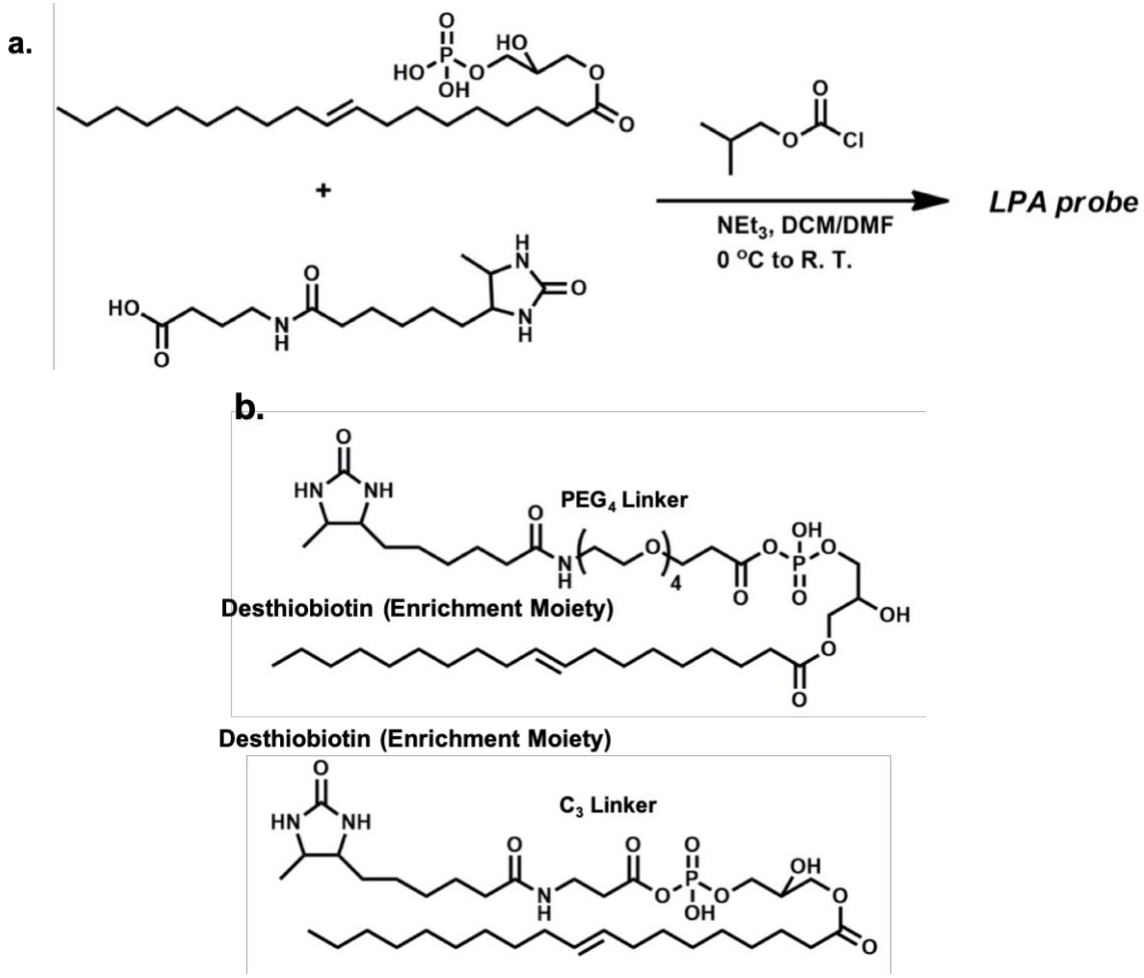


Figure 2. 2 A schematic diagram showing the reaction between the LPA affinity probe with an LPA-binding protein.

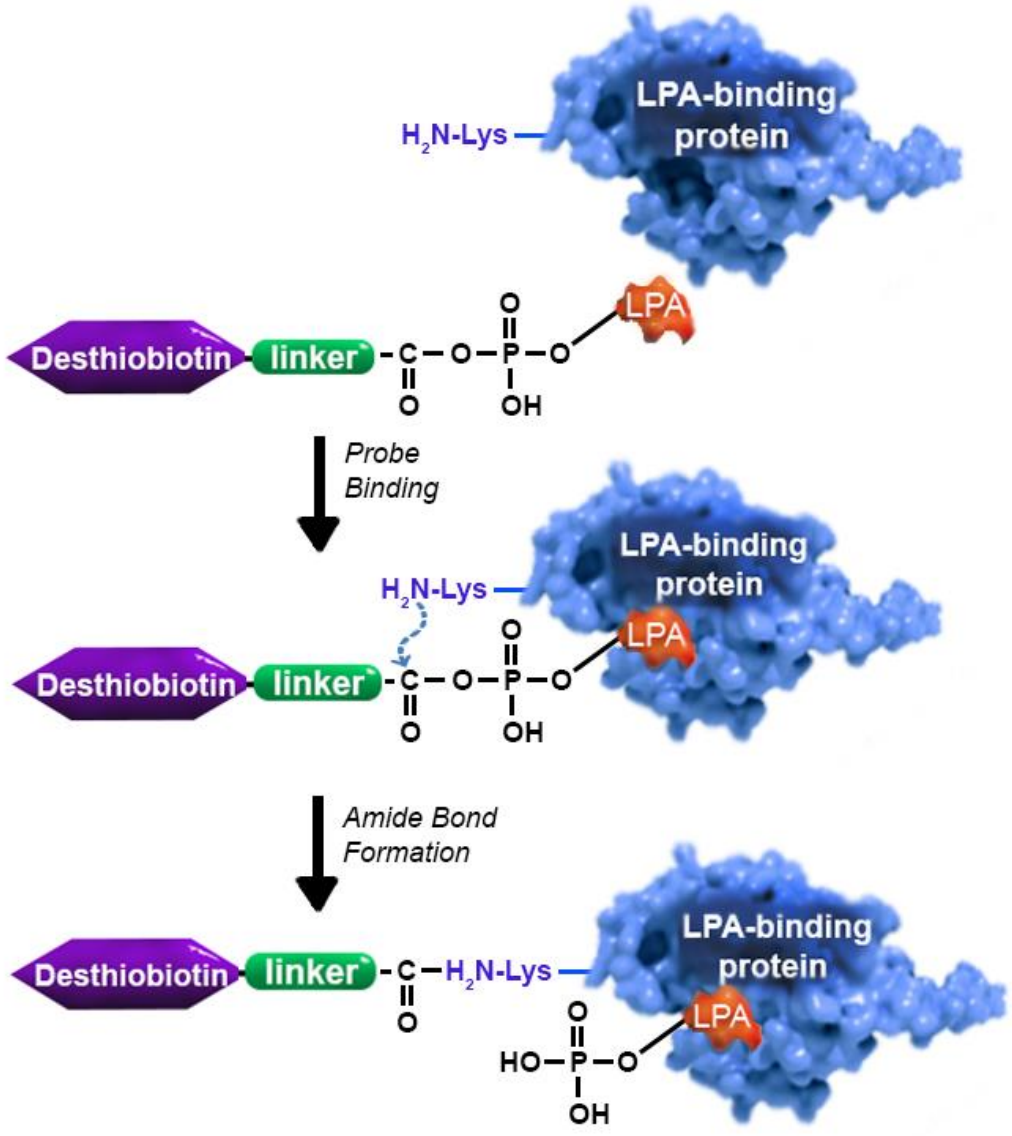


Figure 2. 3 Representative of (a)ESI-MS and (b)MSMS of the purified desthiobiotin-C3-LPA

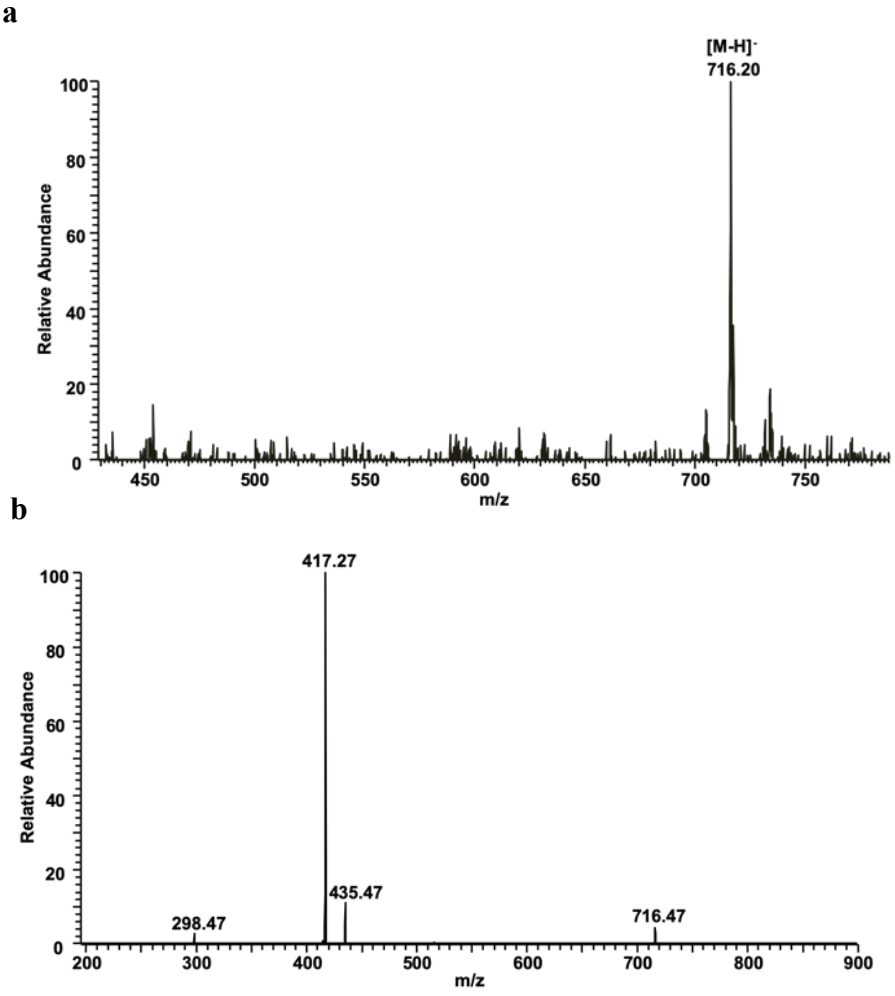


Figure 2. 4 Representative of (a)ESI-MS and (b)MSMS of the purified desthiobiotin-PEG$_4$-LPA probes

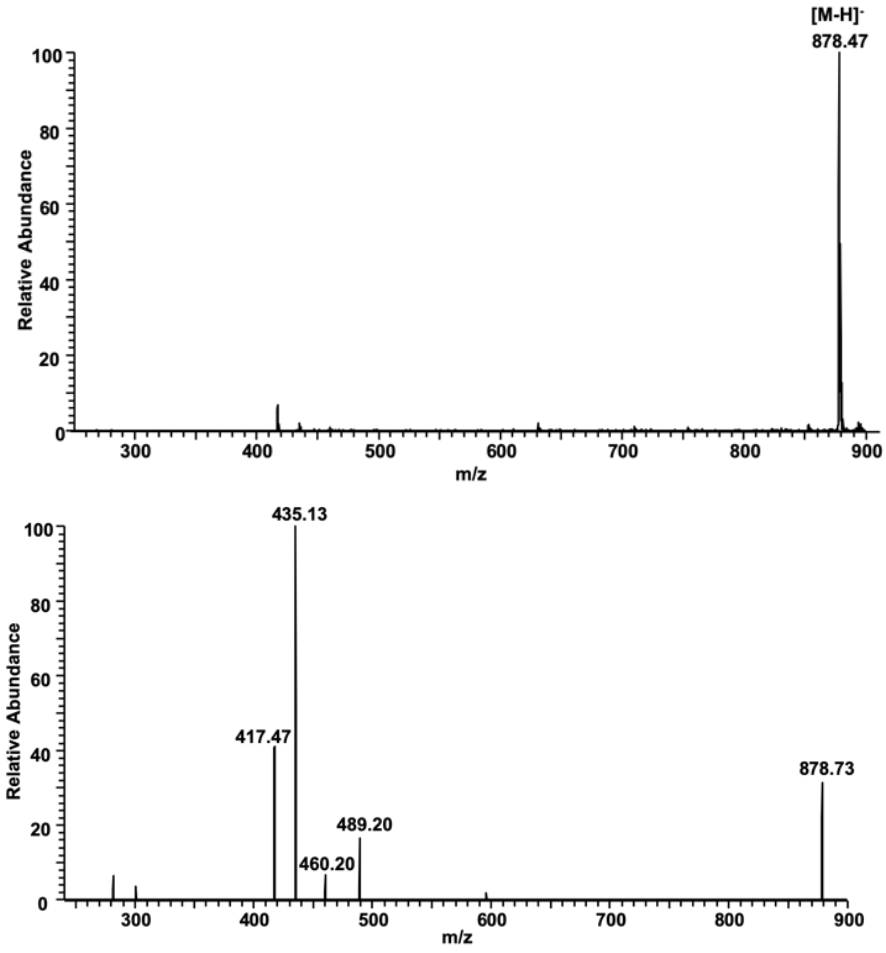


Figure 2. 5 The numbers of identified unique desthiobiotin-modified peptides (a) and proteins (b) from the use of the two desthiobiotin-LPA probes.

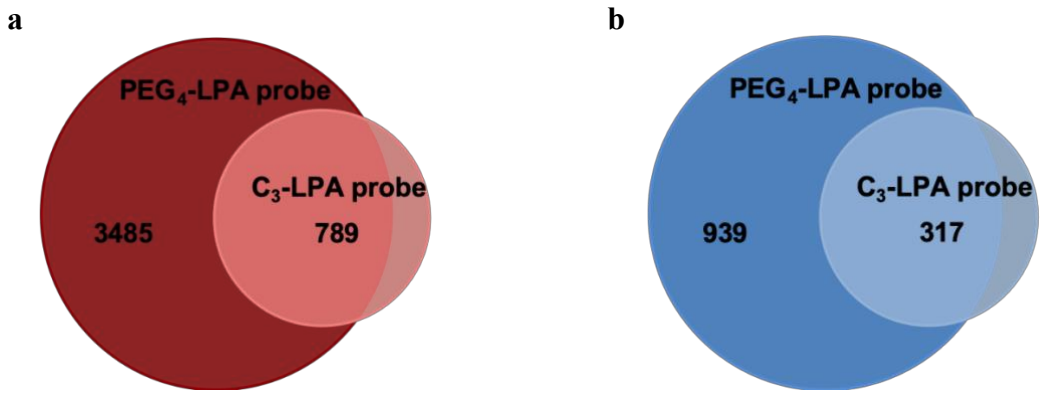


Figure 2. 6 MS/MS for the $[M + 2H]^{2+}$ ion of the desthiobiotin-labeled peptide of DKEMSATFR from LPAR1. This peptide was identified from 293T cells with ectopic expression of LPAR1.

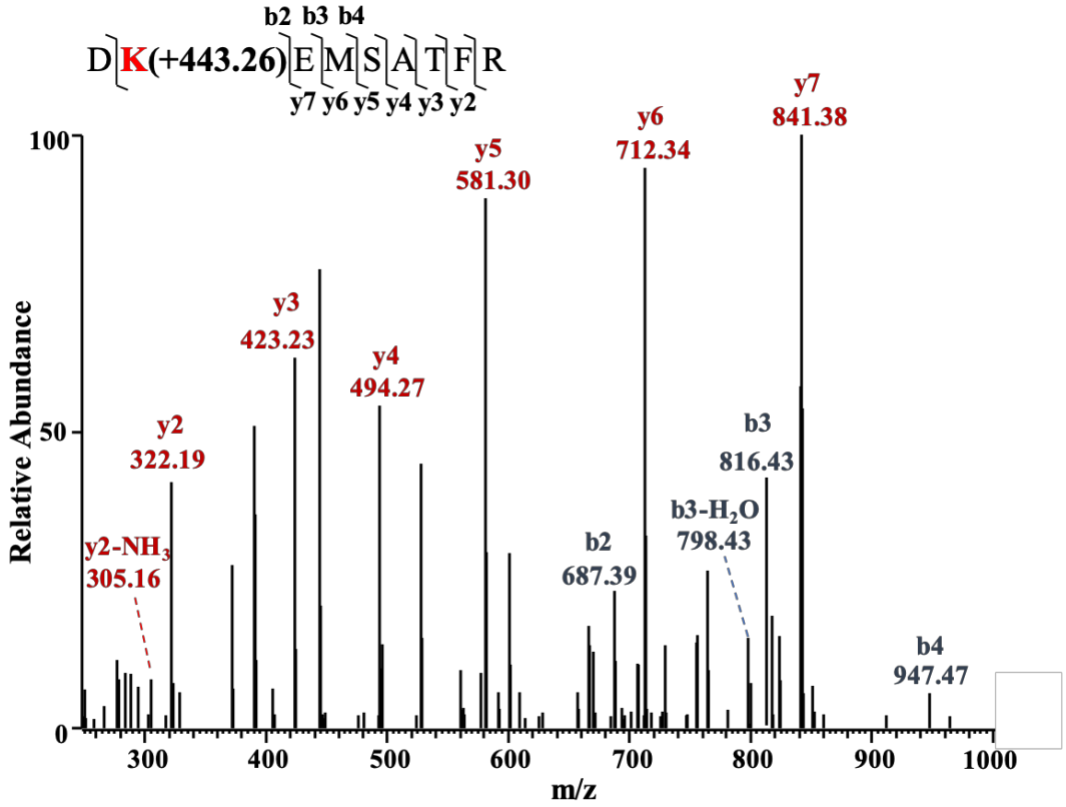


Figure 2. 7 A general SILAC-based strategy for quantitative LPA-affinity profiling of the global proteome

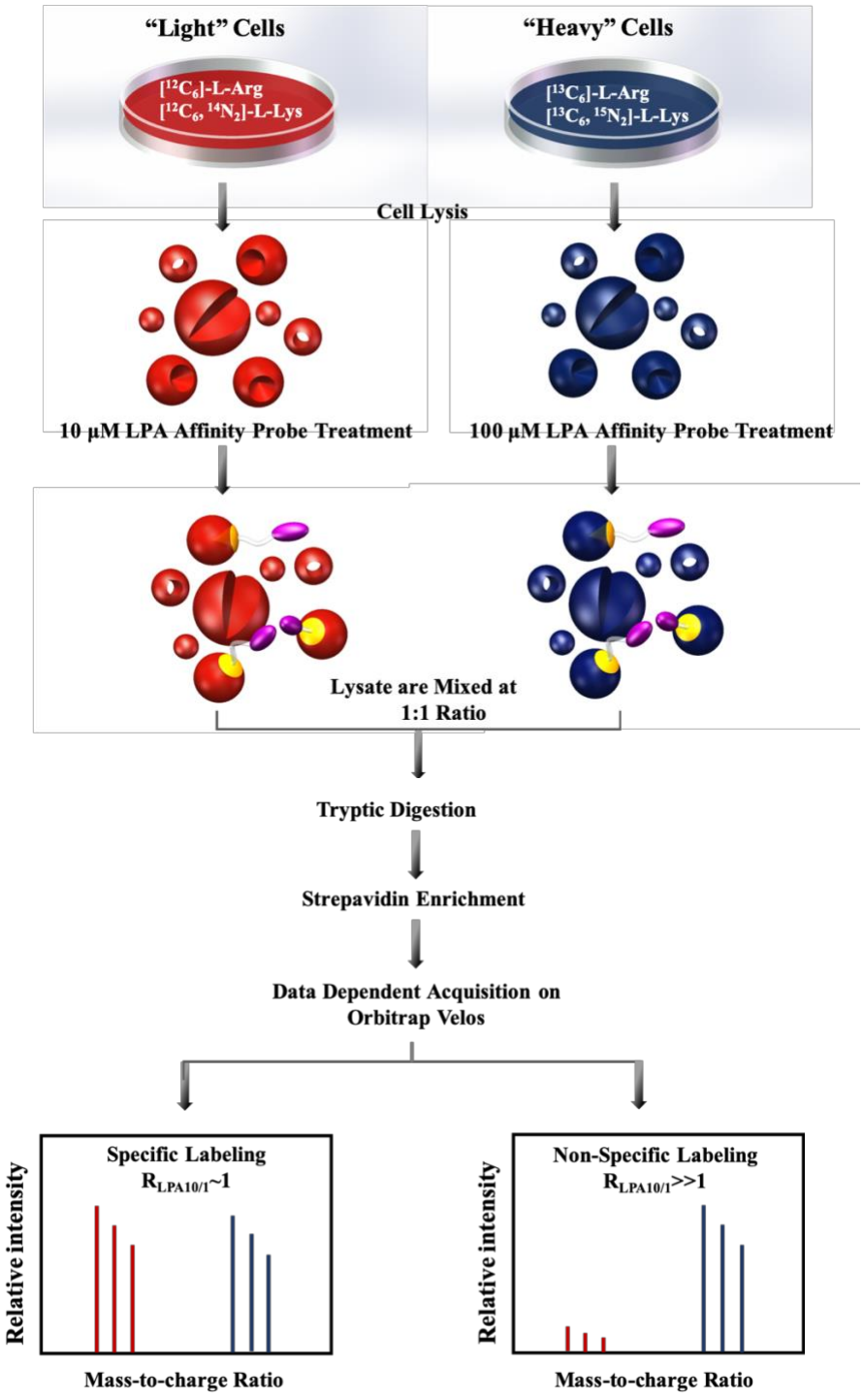


Figure 2. 8 a) Measured $R_{LPA10/1}$ ratio of peptides from 293T SILAC cell lysates with low (10 μM) and high (100 μM) concentrations of LPA affinity probe; b) Light- and heavy-labeled peptides from forward- and reverse-SILAC based affinity profiling experiments. (b, c) Peptide SELTGK#FEK with a low $R_{LPA10/1}$ ratio from Annexin A5; (d, e) Peptide AIEK# LAVEALSSLDGDLAGR with a high $R_{LPA10/1}$ ratio from creatine kinase B-type. "#" indicates the desthiobiotin-labeling site.

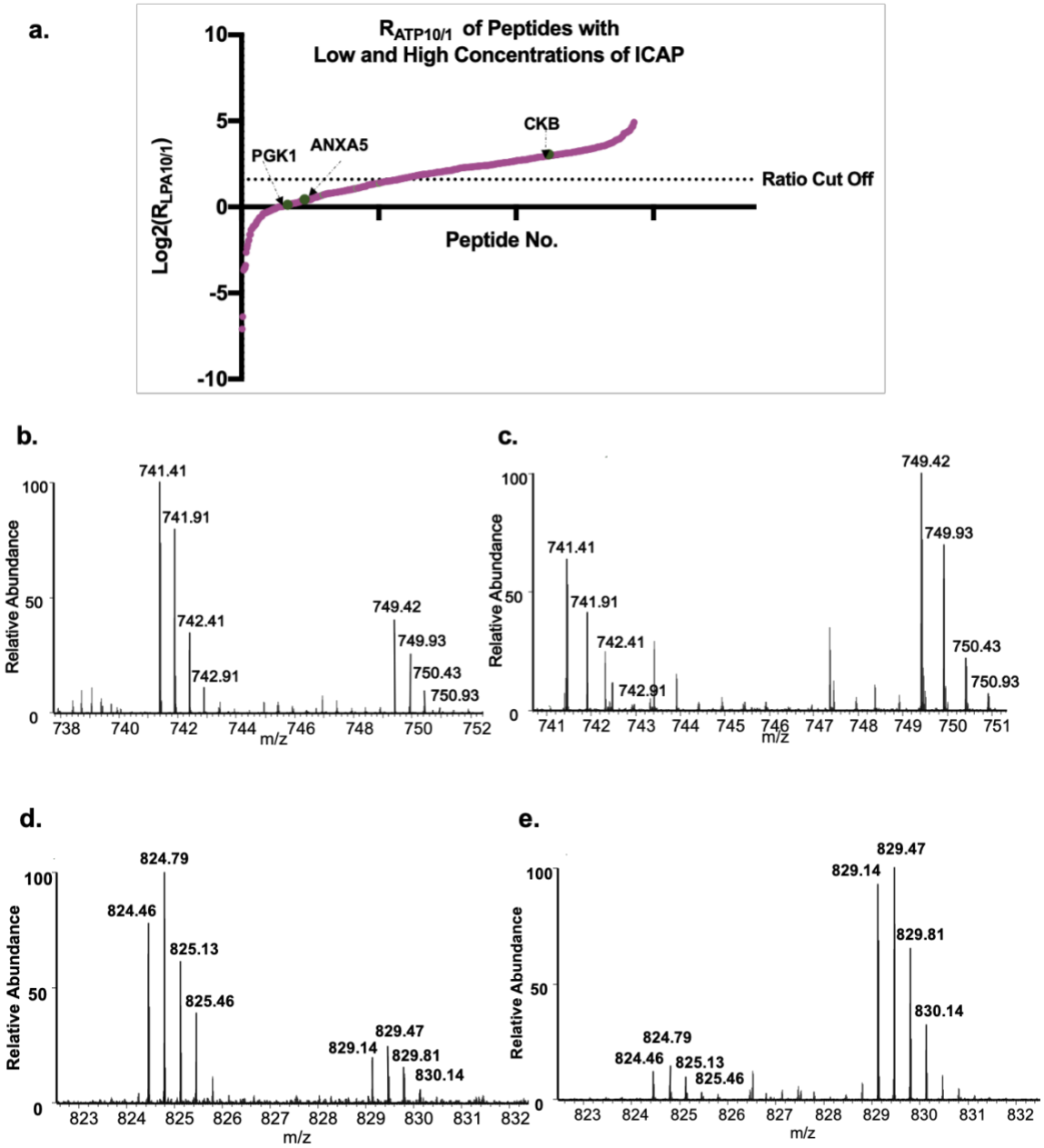
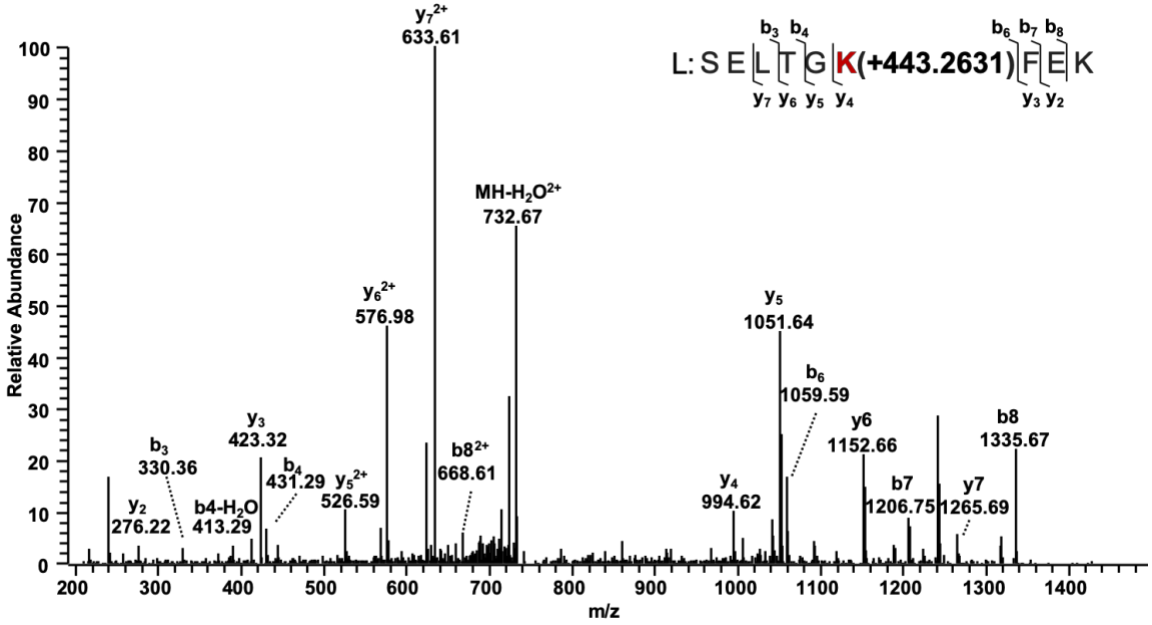


Figure 2. 9 MS/MS of the (a)light (L)- and (b)heavy (H)-labeled SELTGK#FEK from annexin A5 and AIEK#LAVEALSSLDGDLAGR (c, d) from casein kinase B.

a



b

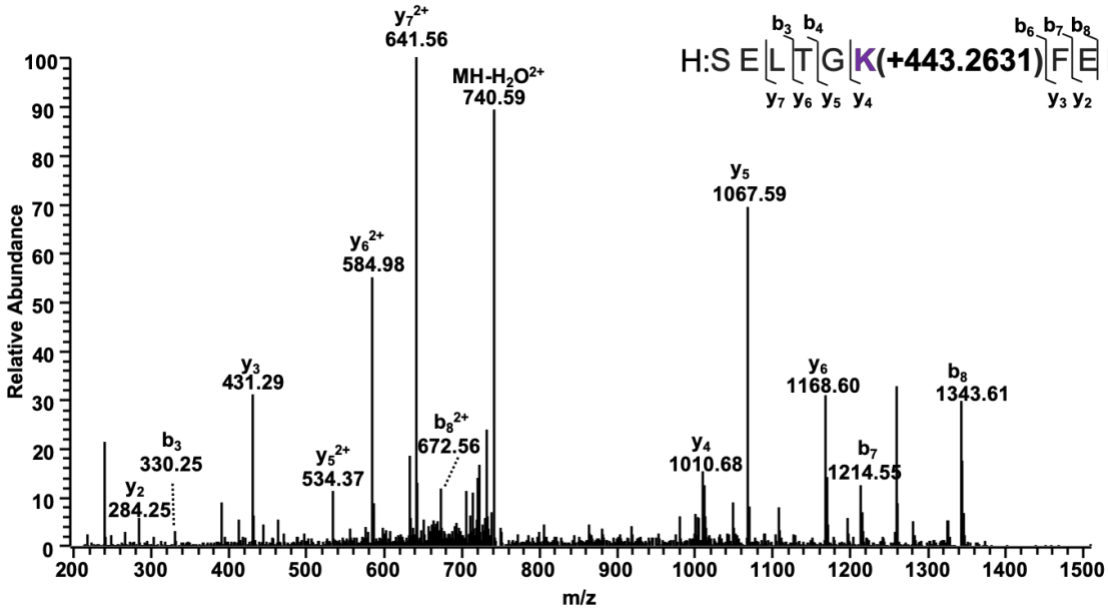
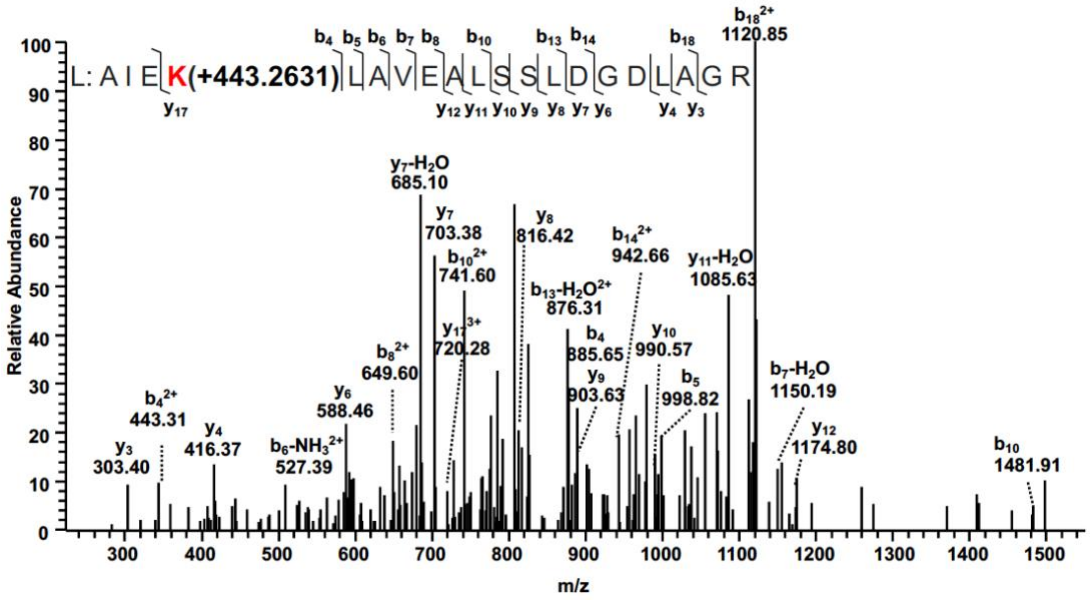


Figure 2. 10 MS/MS of the (a)light (L)- and (b)heavy (H)-labeled AIEK#LAVEALSSLDGDLAGR from casein kinase B

a



b

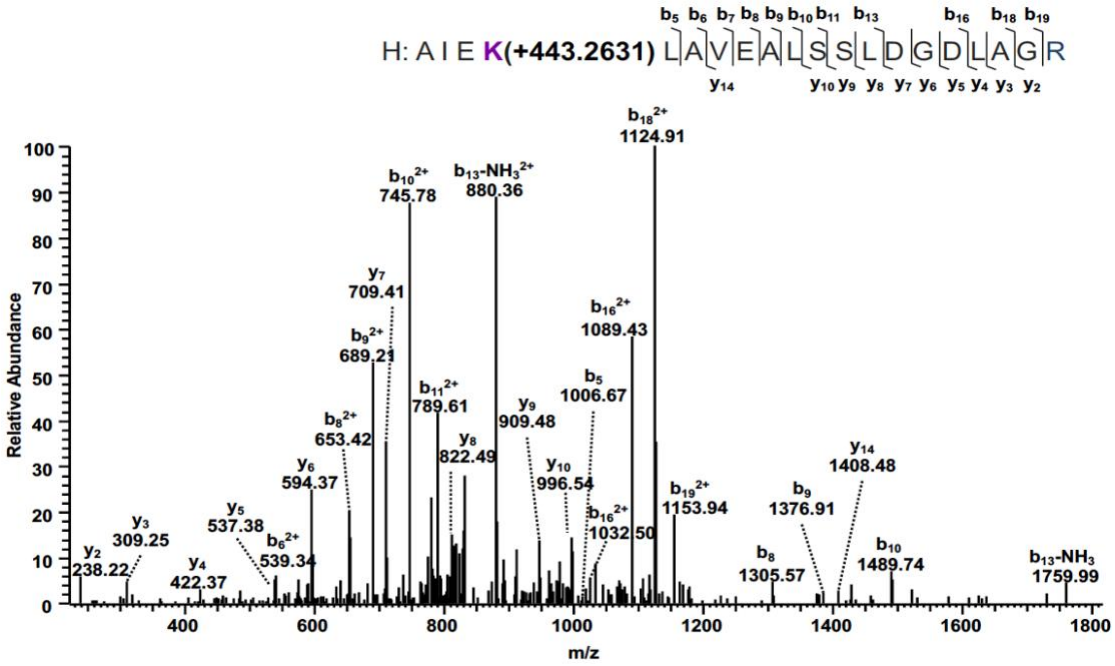


Figure 2. 11 Purification of recombinant His-tagged annexin A5 (a, b) and phosphoglycerate kinase 1 (b) proteins. The proteins were expressed in *E. coli* strain BL21. A lysate (L, lane 2 in a; lanes 2 and 8 in b) was made in lysis buffer and separated by centrifugation into pellet and supernatant fractions. The supernatant fraction (S, lanes 3 and 9 in b) was applied to a 1.0 mL HIS-Select HF Nickel Affinity Gel column and the flow through (F, lane 3 in a; lanes 4 and 10 in b), wash (W, lane 3 in a; lanes 5 and 11 in b), Wash with 25 mM imidazole (W-25 mM, lane 5 in a) and elution [labeled as 1-3 (lanes 5-7) in a; 1-2 (lanes 6-7 and lanes 12-13 in b)] fractions were collected. Samples of the fractions and markers (lane 1) were separated by SDS-PAGE and visualized by Coomassie staining.

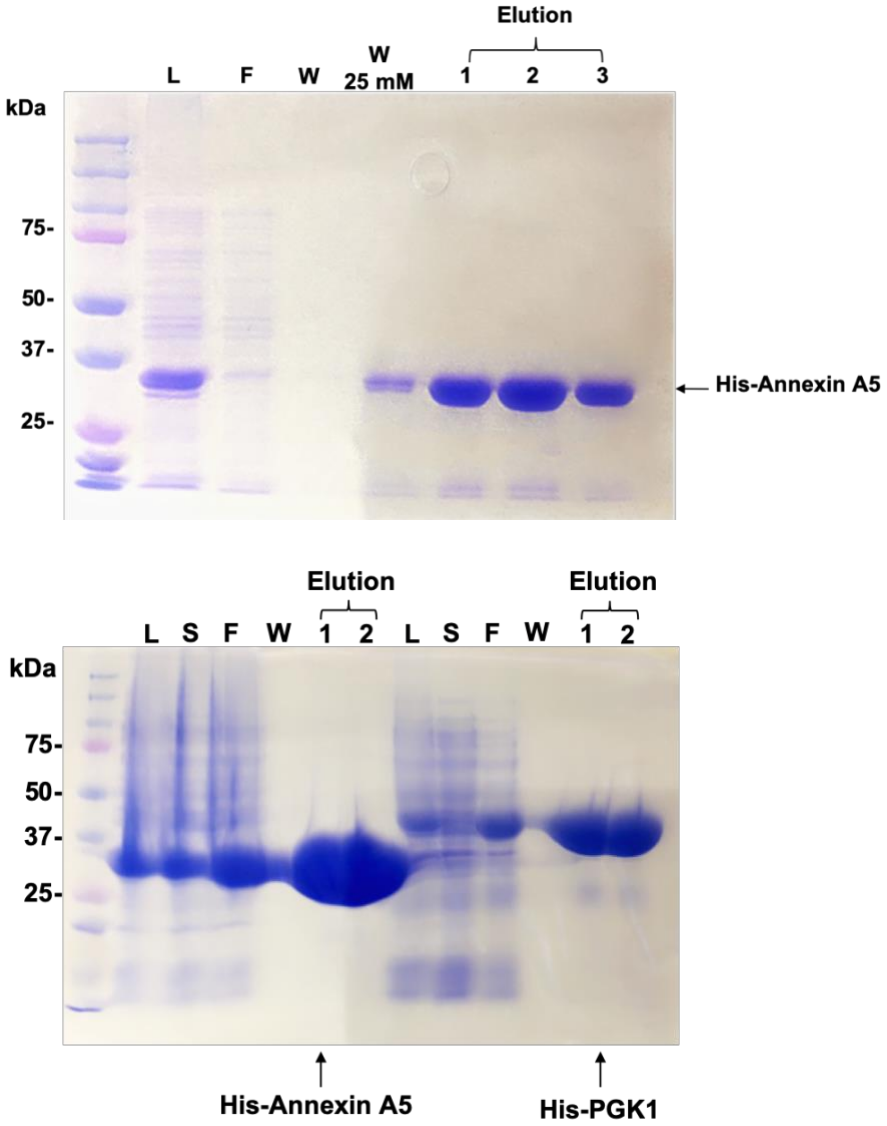
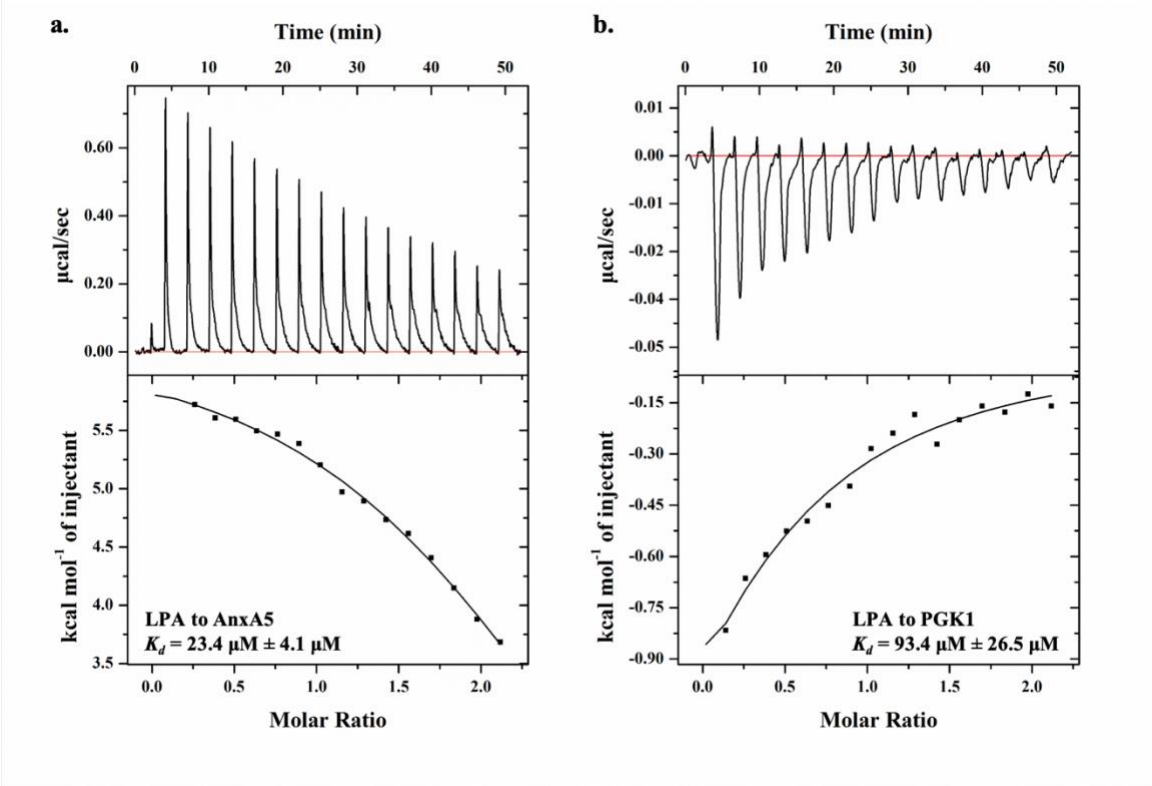


Figure 2. 12 ITC for assessing the binding affinity between LPA and proteins. a) Binding between LPA and AnxA5. 100 μM LPA and 1 μM AnxA5 in 30 mM Tris-HCl buffer (pH 7.5), 10 mM NaCl, 5% glycerol under 25 ^{o}C. b) Binding between LPA and PGK1. 100 μM LPA and 1 μM AnxA5 in 30 mM Tris-HCl buffer (pH 7.5), 10 mM NaCl, 5% gycerol under 25 ^{o}C



Chapter 3 A Chemoproteomic Approach for the Quantitative Identification of Arsenic-binding Proteins

3.1 Introduction

Arsenic is an element that is ubiquitously mobilized into the environment; it forms complexes with other elements, and exists in inorganic and organic forms.[1, 2] Long-term human exposure to arsenic species can result in a myriad of health disorders including cancer,[3-6] neurological diseases,[7, 8] diabetes mellitus,[9] cardiovascular and peripheral vascular disease[10, 11] and hematological disorders.[12]

Arsenic and its derivatives have also been used as chemotherapeutic agents for a variety of diseases.[13] For instance, Fowler's solution, which comprises of potassium arsenite, was used for treating various systemic illnesses in the 18-20th century.[14] In addition, arsenic therapy was reintroduced in the past few decades, where arsenic trioxide (ATO) was found to be very effective in treating acute promyelocytic leukemia (APL) patients carrying the PML-RARα oncogenic fusion.[15, 16]

Because arsenic is a cancer inducer and at the same time a therapeutic agent, many studies have been conducted to examine the biological consequences of arsenic exposure. However, the mechanisms through which it elicits its toxic effects are not fully understood. It is known that trivalent arsenic has a tendency to bind to sulfhydryl (SH) groups in vicinal cysteine residues in proteins.[17-19] Moreover, the binding of arsenic with cysteine (Cys) residues can result in conformational changes in the target protein, which may result in aberrant protein–protein or protein-DNA interactions.[20-22]

Identification of arsenic-binding proteins is essential for understanding the mechanisms through which arsenic confers its biological effects. Many arsenic-binding proteins have been identified in mammalian cells, including hemoglobin,[23] tubulin[24], actin[25], arsR[19], nicotinic receptor[26], galectin-1[27] and thioredoxin peroxidase II.[28] However, some of the observed toxic effects of arsenic cannot be attributed to its interactions with these proteins.[19] Thus, there are likely other yet identified arsenic-binding proteins.

Several proteomic approaches were employed to identify arsenic-binding proteins. In this vein, a global profiling using a human proteome microarray led to the identification of 360 arsenic-binding proteins, including hexokinase 2 (HK2), whose enzymatic activity can be inhibited by As(III).[29] Mass spectrometry (MS) has great advantages in the identification and quantification of proteins in complex sample matrices.[30] Le and colleagues[31] employed an approach combining-LC-MS/MS with click chemistry using an azide-labeled arsenical to selectively capture and identify 48 arsenic-binding proteins in A549 cells. Zhang et al.[32] employed a *p*-aminophenylarsenoxide (PAPAO)–biotin conjugate to treat with MCF-7 human breast cancer cells and 50 arsenic-binding proteins were identified using MALDI-TOF MS. While these methods allowed for the identification of arsenic-binding proteins, they did not incorporate quantification into the proteomic workflow. In particular, non-specific-binding of proteins to streptavidin beads may lead to false-positive identifications of As(III)-binding proteins.

Here, we utilized a chemoproteomic method, with the use of a biotin-conjugated As(III) probe, in combination with stable isotope labeling by amino acids in cell culture (SILAC),[33] to label, enrich and identify candidate inorganic and organic As(III)-binding

proteins from the entire human proteome in the absence or presence of excess amount of As(III) (i.e., PAPAO). By employing this strategy, we were able to identify a number of candidate arsenic-binding proteins. We also confirmed the binding of one of these proteins, HSPA4, with PAPAO using Western blotting and assessed the potential As(III)-binding sites in this protein.

3.2 Experimental Section

3.2.1 Cell Culture

HEK293T human embryonic kidney epithelial cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Corning) and 1% penicillin-streptomycin solution (PS, GE Healthcare). The cells were maintained in a humidified chamber with 5% CO_2 at 37°C and the culture medium was changed in every 2-3 days as needed.

3.2.2 Biotin-As Probe Labeling and PAPAO Competition Experiment

HEK293T cells were harvested at approximately 80% confluency. The resulting cell pellet was incubated on ice for 30 min with 1% protease inhibitor cocktail (Sigma) containing pre-chilled CelLytic™ M cell lysis reagent (Sigma Aldrich). The mixture was centrifuged at 16000*g* for 30 min at 4°C, and the supernatant, which contained cellular proteins, was loaded onto a NAP-10 column (Amersham Biosciences) for the removal of endogenous nucleotides, and proteins were eluted with a buffer containing 50 mM HEPES

(pH 7.4), 75 mM NaCl, and 5% glycerol. Protein concentrations were subsequently quantified by using Quick Start Bradford Protein Assay (Bio-Rad).

The biotin-As probe was synthesized following published procedures by coupling PAPAO with the activated pentafluorophenyl ester of biotin.[26, 34] In the forward SILAC experiment, heavy protein lysate (1.0 mg each) was first incubated with 80 μM PAPAO at 4°C for 30 min. The biotin-As probe was then added to the mixture, or the same amount of light lysate without preincubation with PAPAO, until the final concentration of the probe reached 20 μM. After incubation at room temperature with gentle shaking for 2 h, the biotin-As-conjugated proteins were enriched individually from the mixtures by incubating with high-capacity streptavidin beads (Sigma-Aldrich) at room temperature for 2 h. The reverse SILAC experiment was conducted in a similar way except that the light-labeled lysate was pre-incubated with 80 μM PAPAO. The enriched heavy- and light-biotin-labeled proteins were mixed and subjected to filter-aided sample preparation (FASP)[35], denatured with 8 M urea, reduced with dithiothreitol (20 mM), and alkylated with iodoacetamide (55 mM). The protein mixture was then subjected to buffer exchange with 50 mM ammonium bicarbonate and digested with trypsin at a 1:100 ratio of enzyme/protein. Prior to LC-MS and MS/MS analyses, the tryptic peptides were desalted with C18 ZipTip (Agilent), concentrated by Speed-Vac, and reconstituted in 20 μL of 0.1% formic acid.

3.2.3 LC-MS/MS Analysis and Data Processing

LC-MS/MS experiments were performed on a Q Exactive Plus quadrupole–Orbitrap mass spectrometer equipped with a Flex nanoelectrospray ion source (Thermo Fisher

Scientific) and coupled with an UltiMate 3000 RSLCnano system (Dionex, Thermo Fisher). Peptide mixtures were loaded onto a home-made 4 cm capillary trapping column (150 μm i.d.) packed with 5 μm C18 120 Å reversed-phase material (ReproSil-Pur 120 C18-AQ, Dr. Maisch GmbH HPLC) with buffer A (0.1% formic acid in water) at a flow rate of 3 μL/min. The peptides were eluted from the trapping column and separated on a 20-cm long fused silica analytical column (75 μm i.d.) packed in-house with C18 resin (3 μm in particle size and 120 Å in pore size, Dr. Maisch GmbH HPLC). A 200-min linear gradient of 3-45% buffer B (0.1% formic acid in 80% CH_3CN) was used and the flow rate was 300 nL/min. The mass spectrometer was operated in a data-dependent acquisition mode. Full-scan mass spectra were acquired using the Orbitrap mass analyzer with a resolution of 35,000 with lock mass option being enabled for the ion of *m/z* 445.120025. Up to 25 most abundant ions found in full-scan MS with charge state of ≥ 2 were sequentially isolated and fragmented with a normalized collisional energy of 28, an activation Q value of 0.25, and an activation time of 10 ms.

The raw data acquired were processed with the Maxquant search engine (version 1.5.2.8) against a human fasta database (Uniprot Human Proteome Fasta: UP000005640_9606) which contained 78,120 entries. The mass tolerances for precursor and fragment ions were set at 10 and 20 ppm, respectively. The maximum number of tryptic miss-cleavages was set as two per peptide. For SILAC-based quantification experiments, lysine (+8 Da) and arginine (+6 Da) mass shifts introduced by heavy stable isotope-labeling and cysteine carbamidomethylation were considered as fixed modifications. The reverse database

search option was enabled to filter the search results to satisfy a maximum false discovery rate of 1%.

3.2.4 Western Blot

The HEK293T cells were lysed following the above-described procedures, and 1.0 mg whole-cell protein lysate was used for each pull-down and Western blot experiment. High-capacity streptavidin beads (40 μL) were washed with PBS and incubated with 1 mL of 20 μM biotin-As at room temperature for 2 h. Excess free biotin-As was removed by washing sequentially with 10× PBS, 1× PBS, and water. The resulting biotin-As-saturated streptavidin beads were then added to 1.0 mg whole cell lysates pre-incubated with 0, 0.5, 1.0, 2.0 μM PAPAO. The mixtures were agitated at room temperature for 2 h. The resin was subsequently washed, and the bound proteins were eluted from the streptavidin beads by heating in a 25 μL of 2 × SDS loading buffer at 95°C for 10 min. The resulting mixtures were centrifuged, and the supernatants were subjected to Western blot analysis. HSPA4 antibody (Abclonal, #A19705, with a 1:1000 dilution) and Anti-Rabbit IgG (whole molecule)–Peroxidase antibody produced in goat (Sigma, #A0545, with a 1: 5000 dilution) were used as the primary and secondary antibodies, respectively.

3.3 Results and Discussion

3.3.1 Global Profiling of As(III)-Binding Proteins from the Whole Human Proteome

In this study, we aim to identify candidate As(III)-binding proteins from the lysates of HEK293T cells using LC-MS/MS. Toward this end, we devised an affinity profiling

strategy by using a biotin-conjugated As(III) probe, which can react with cysteine residues and install biotin tag(s) on the resulting proteins. To minimize false-positive identification of As(III)-binding proteins arising from non-specific adsorption of proteins to streptavidin beads, we employed an SILAC-based workflow by incorporating competition using PAPAO to reveal specific inorganic and organic As(III)-binding proteins.

We first incubated separately the 20 μM synthesized biotin-As(III) probe with the same amount of light and heavy HEK293T cell lysate, where the heavy cell lysate was pre-incubated with excess amount (80 μM) of PAPAO (forward experiment). To minimize bias introduced from the labeling, we also performed reverse labeling experiments, where excess amount of PAPAO was pre-incubated with the light cell lysate instead. In this vein, free arsenic-containing compounds (i.e., PAPAO) would also bind to proximal cysteine residues in proteins, thereby diminishing the labeling efficiency of the biotin-As probe. Thus, those proteins exhibiting attenuated labeling efficiencies in lysates pre-incubated with PAPAO are considered specific arsenic-binding proteins.

Biotin-labeled proteins from the reaction mixtures were enriched by using high-capacity streptavidin agarose beads, digested with trypsin, and the ensuing peptides analyzed on a Q Exactive Plus quadrupole-Orbitrap mass spectrometer (Figure 3.1). We used peak intensity ratios of the light- and heavy-labeled peptides from the two experimental states to derive the biotin-As labeling ratio. To classify a protein as an As(III)-binding protein, we employed a cut-off ratio of 0.5 for the peptide intensities derived from the experiments with over without competitor (i.e., PAPAO).

This method led to the quantification of a total of 409 proteins, where 51 putative arsenic-binding proteins were identified using the aforementioned criterion. The list contained several previously reported arsenic-binding proteins. For example, thioredoxin (TXN), which exhibits a high binding specificity to PAPAO with a ratio of 0.18 (Figure 3.2), was shown to have the highly conserved −Cys Gly Pro Cys– region and four trivalent arsenic species can bind to human TXN to form complexes.[28] Furthermore, exposure to ATO was shown to result in irreversible inactivation of TrxR arising from depletion of GSH, and GSH depletion by buthionine sulfoximine enhanced ATO-induced cell death, suggesting that the anticancer of ATO arises, in part, from TXN-mediated apoptosis.[36] Many enzymes involved in the glycolysis pathway, including transketolase (TKT), phosphoglycerate kinase 1 (PGK1) with ratio, and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with SILAC protein ratios of 0.51, 0.27 and 0.68, respectively, were shown to be impacted by arsenic exposure.[29, 37, 38] (Table 3.1) Together, our results revealed many proteins displaying specific arsenic-binding affinity and our quantitative affinity profiling strategy enabled proteome-wide assessment of protein-As(III) interactions.

3.3.2 HSPA4 can Bind to PAPAO

Among those identified proteins, 70 kDa heat shock protein 4 (HSPA4) and 110 kDa heat shock protein 1 (HSPH1) exhibited a relatively high specificity toward binding with PAPAO, with the quantified proteomic ratios being 0.37 and 0.23, respectively (Figure 3.3 a, Figure 3.5). In addition, the quantification results are highly reproducible among the

different replicate experiments. The representative MS and MS/MS for the light- and heavy-labeled peptides derived from HSPA4 are shown in Figure 3.3.

Heat-shock proteins (HSPs) represent the major components of molecular chaperones that facilitate the folding and assembly of newly synthesized proteins and the selection of unfolded proteins for degradation in different sub-cellular compartments such as cytosol, endoplasmic reticulum, and mitochondria.[37] HSPA4 is a member of the HSP70 family, and this family of HSPs and their co-chaperones participate in many important cellular processes, including protein folding, regulation of ubiquitination and other protein homeostasis network.[38-40]

We next examined further the interaction between HSPA4 and different arsenic-containing compounds by employing Biotin-As pull-down followed by Western blot analysis. To this end, we incubated the lysate of HEK293T cells with biotin-As-immobilized streptavidin beads in the presence of increasing concentrations of PAPAO (0, 0.5, 1 to 2 μM). Our Western blot results revealed that the amount of HSPA4 remained on the biotin-As-conjugated beads decreases with increasing concentration of added free PAPAO. These results substantiated the specific interaction between HSPA4 and As(III) (Figure 3.4).

Arsenic is known to affect protein functions through the formation of adducts with closely spaced cysteine sulfhydryl groups[41]. Thus, we analyzed the structure of HSPA4 to examine if As(III) can access specific Cys residues in HSPA4. While there is currently no high-resolution structure of human HSPA4, our sequence alignment result revealed a 66% amino acid identity between HSPA4 and HSP110 (Figure 3.7a). Thus, we analyzed the X-

ray crystal structure of HSP110 (PDB code: 6GFA), which revealed that the nucleotide binding domain of HSP110 contains several Cys residues (Figure 3.7b). As(III) in the biotin-As probe can bind to sulfhydryl groups of two closely placed cysteine residues in proteins[42]. To predict which Cys residues may be involved in arsenic binding, we further examined the distribution of Cys residues in HSPA4. In particular, we found that Cys139, Cys166, Cys375, Cys379are in close proximity (Figure 3.7b). Among them, Cys375 and Cys379 are located on the protein surface and within an α-helix which allows for the access and conjugation with arsenic.(Table 3.2) (Figure 3.6) We also observed that two cysteine residues (i.e., Cys13 and Cys34) are situated near the nucleotide-binding pocket, which may be responsible for the recognition and binding of ATP by creation of a proton bridge with N^6-amino group of ATP.[42] Thus, the binding with arsenic may elicit conformational change(s) to the protein and perturb its binding to ATP, thereby perturbing the chaperone activity of HSP70s during the ATP hydrolysis cycle (Figure 3.7c).[43]

3.4 Conclusions

Arsenic is a widespread environmental pollutant and ATO has been successfully employed in the treatment of acute promyelocytic leukemia. Numerous studies have been conducted to examine its mechanisms of action. Here, we report a chemoproteomic approach, which relies on SILAC-based metabolic labeling, biotin-As probe pull-down and LC-MS/MS analysis, to achieve proteome-wide identification of arsenic-binding proteins in human cells.

We were able to identify a total of 409 candidate arsenic-binding proteins from HEK293T cells. By combining a SILAC-based quantitative proteomic workflow and competitive labeling in the presence of PAPAO, the method facilitated the quantitative characterizations of arsenic-protein interactions and the identification of specific arsenic-binding proteins. The selectivity of arsenic-binding proteins was systematically investigated by quantifying peak intensity ratios of light- and heavy-labeled tryptic peptides. Using this method, we were able to identify 51 candidate As(III)-binding proteins. The results showed that the arsenic affinity pull-down can be employed as a useful method for the enrichment and subsequent detection of arsenic-binding proteins from the human proteome, which may reveal new protein targets that contribute to arsenic toxicity.

By analyzing the structure of HSP110, which exhibits a very high level of sequence identity as HSPA4, we found that the Cys residues located near the nucleotide-binding pocket of the protein (i.e., Cys12 and Cys33) [43] may constitute an As(III) binding site. This raises the possibility that As(III) binding may disrupt the ATPase activity of HSPA4, thereby perturbing the chaperone function of HSPA4. Along this line, arsenic exposure is known to elicit proteotoxic stress through multiple pathways, including the disruption of molecular chaperones that are crucial in maintaining the appropriate folding of the proteome.[44] Hence, our study revealed HSPA4 as a potential new molecular target underlying the proteotoxic effects of arsenic exposure. To our knowledge, this is the first identification of HSPA4 as a candidate arsenic-binding protein.

Together, we developed a chemoproteomic method for the quantitative identification of candidate arsenic-binding proteins, which provides mechanistic insights into mechanisms underlying the toxic effects of arsenic species.

Table 3. 1 Candidate PAPAO-binding proteins. Known As(III)-binding proteins are colored in red.

Gene ID	Gene names	Ave	S. D.
P10599	TXN[28]	0.176	0.041
Q92598	HSPH1	0.233	0.126
Q02790	FKBP4	0.254	0.183
P34932	HSPA4	0.255	0.098
P06733	ENO1[31]	0.262	0.131
P00558	PGK1[31]	0.273	0.138
Q9UBE0	SAE1	0.281	0.139
P31948	STIP1[31]	0.284	0.129
Q96JB1	DNAH8	0.293	0.081
Q9NR45	NANS	0.293	0.162
P40926	MDH2	0.309	0.250
P28072	PSMB6[29]	0.313	0.217
O95757	HSPA4L	0.313	NaN
Q8IX12	CCAR1	0.318	0.084
Q96G61	NUDT11	0.328	NaN
P49189	ALDH9A1	0.328	0.099
P13804	ETFA	0.340	NaN
P14174	MIF	0.340	0.198
Q9Y490	TLN1	0.355	NaN
Q96ER3	SAAL1	0.356	0.148
P24666	ACP1	0.360	0.219
P60900	PSMA6	0.363	0.234
Q9NXG2	THUMPD1	0.368	0.070
P49419	ALDH7A1	0.370	NaN
Q6P2E9	EDC4	0.376	NaN
O43175	PHGDH	0.389	0.192
P62917	RPL8	0.394	0.153
P62241	RPS8	0.397	0.157
P27694	RPA1	0.399	NaN
P14618	PKM	0.410	0.168
Q13126	MTAP	0.411	0.115
P84103	SRSF3	0.419	0.188
Q9BWD1	ACAT2	0.434	NaN

P41227	NAA10	0.437	NaN
P25398	RPS12	0.446	0.305
Q15185	PTGES3	0.447	0.092
Q13451	FKBP5[29]	0.450	0.186
Q9BY32	ITPA	0.451	0.235
P11310	ACADM	0.454	0.225
Q9UQ80	PA2G4	0.454	0.189
Q96GK7	FAHD2A	0.460	0.249
P07900	HSP90AA1[31]	0.467	0.151
P61978	HNRNPK	0.467	0.157
P26599	PTBP1	0.473	0.147
Q08211	DHX9	0.484	0.218
Q04760	GLO1	0.486	0.237
Q86VP6	CAND1	0.491	NaN
P13667	PDIA4	0.491	0.238
Q9NQW7	XPNPEP1	0.493	0.252
Q8NBS9	TXNDC5	0.493	0.270
Q00610	CLTC	0.499	0.185

Table 3. 2 (Continute) Candidate PAPAO-binding proteins. Known As(III)-binding proteins are colored in red.

Figure 3. 1 A general SILAC-based competition strategy for proteome-wide assessment of the binding capabilities of proteins toward arsenic-containing compounds.

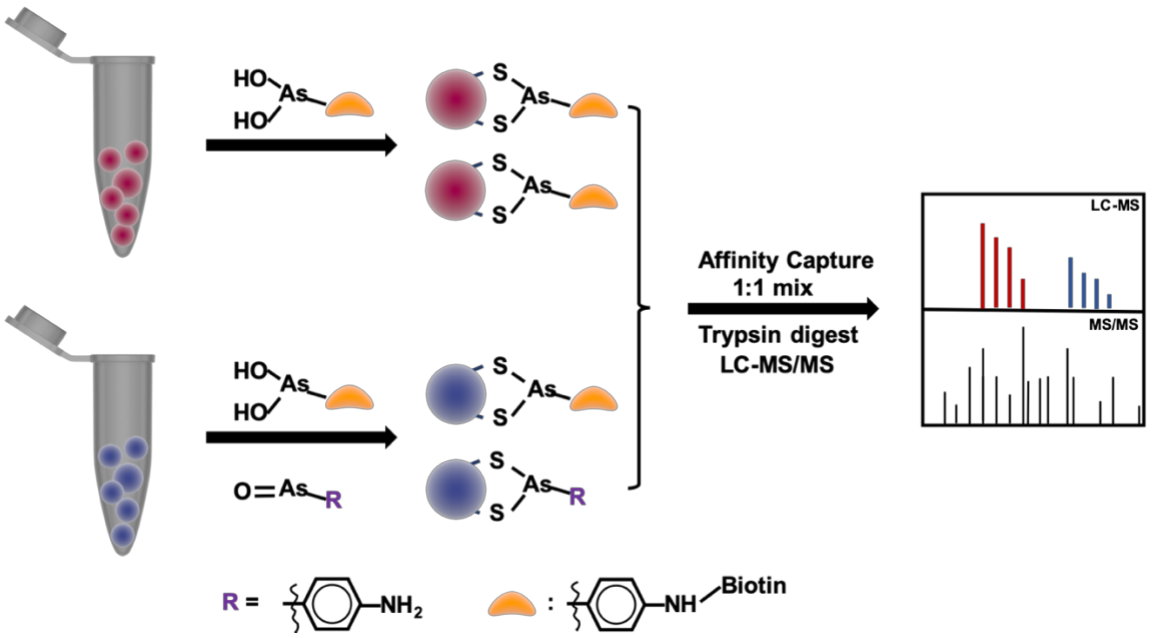


Figure 3. 2 LC-MS/MS revealed thioredoxin (TXN) as an As(III)-binding protein (a) A comparison of quantification results of TXN peptide VGEFSGANK obtained from forward and reverse competitive labeling experiment with 80 μM PAPAO; (b) MS/MS of light- and heavy-arginine-containing peptide, VGEFSGANR.

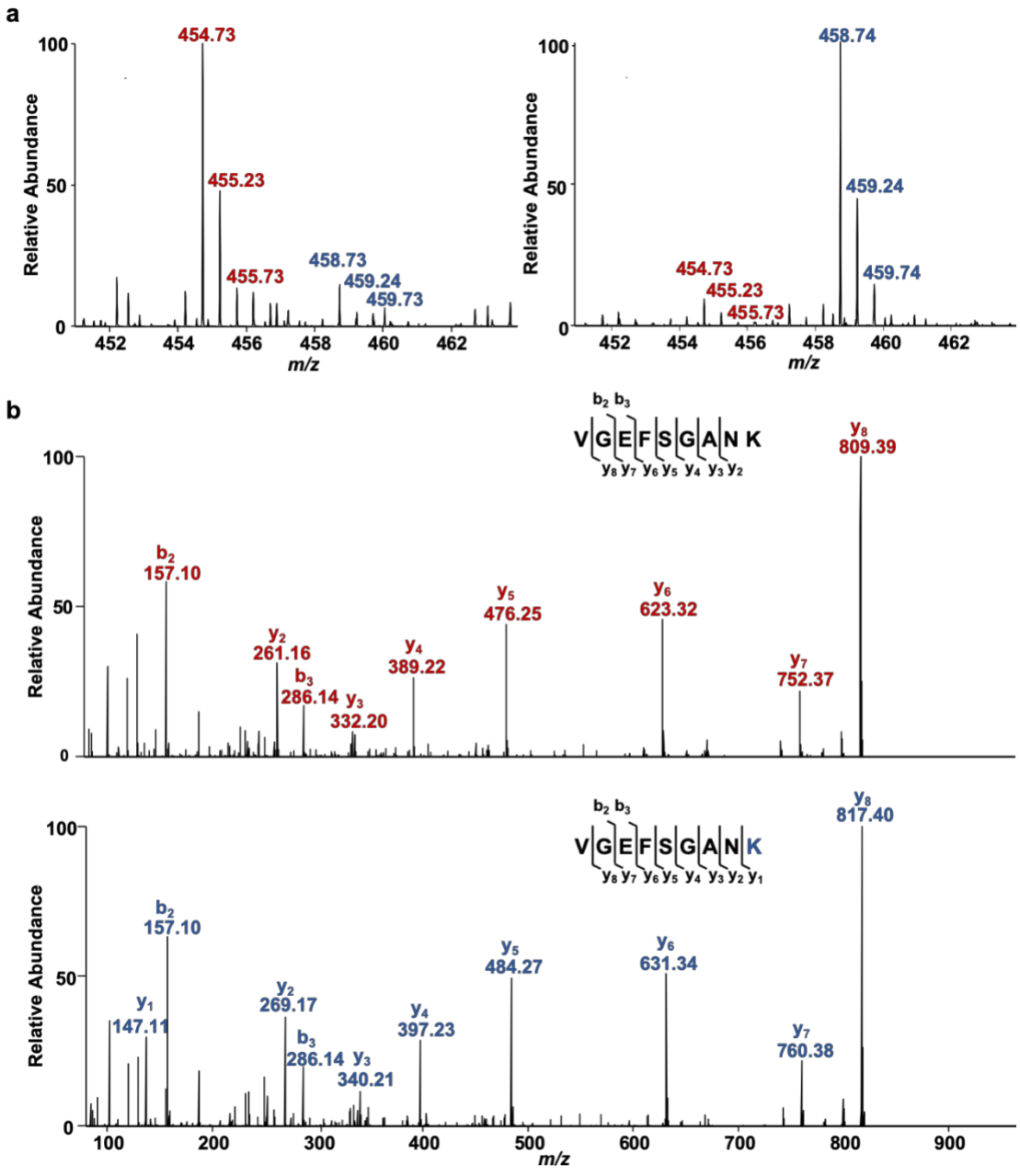


Figure 3. 3 LC-MS/MS revealed HSPA4 as an As(III)-binding protein (a) A comparison of quantification results for a tryptic peptide, TSTVDLPIENQLLWQIDR, of HSPA4 obtained from forward and reverse competitive labeling experiments with 80 μM PAPAO; (b) Representative MS/MS of the light- and heavy arginine-labeled peptide TSTVDLPIENQLLWQIDR

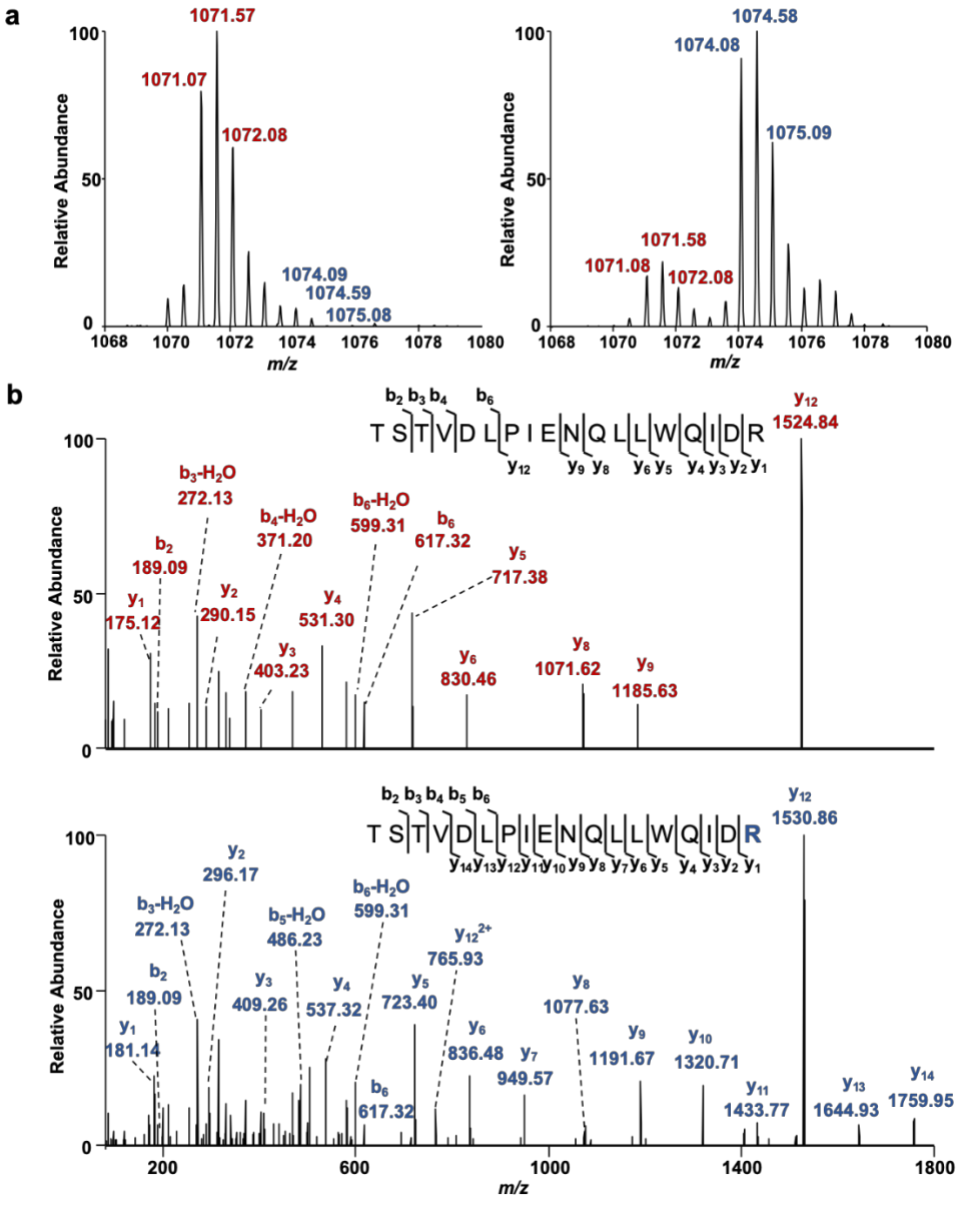


Figure 3. 4 Competitive binding of As(III) to HSPA4. Streptavidin affinity pull-down followed by Western blot analysis revealed the competitive interaction between HSPA4 and As(III) and Biotin-As probe. The interaction between Biotin-As and HSPA4 was diminished after co-incubated with 0.5, 1 and 2 μM PAPAO. Quantification results obtained from three separate experiments. The data represent mean ± S. D. The *p* values were calculated using two-tailed, paired *t* test, *** $p < 0.001$.

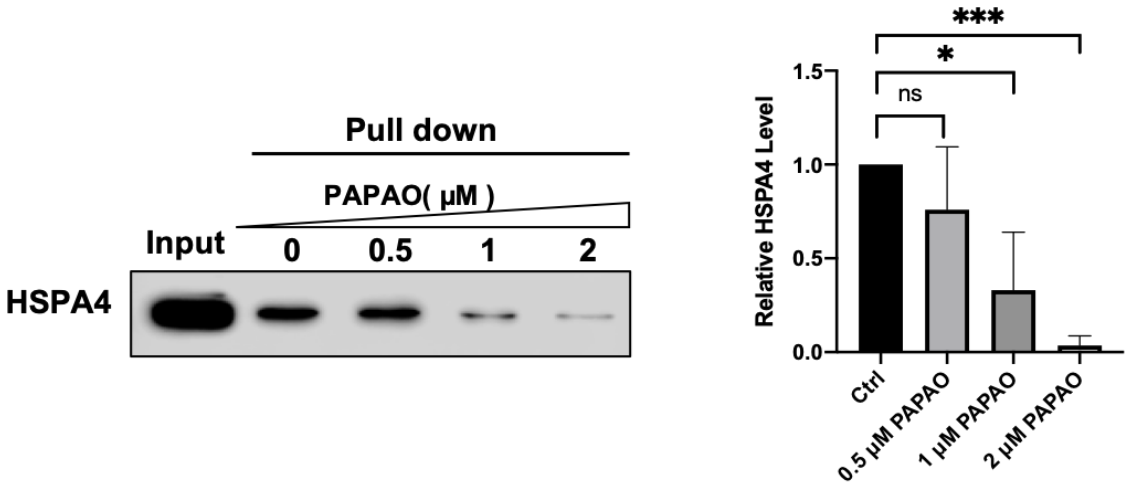


Figure 3. 5 Representative of LC-MS/MS of HSPH1. A quantitative comparison of (a) integrated areas and (b)MS for HSPH1(HSP110) peptide AGGIETIANEFSDR obtained from forward and reverse competitive labeling experiments with 80 μM PAPAO;(c)MS/MS of light- and heavy-arginine-containing peptide, AGGIETIANEFSDR

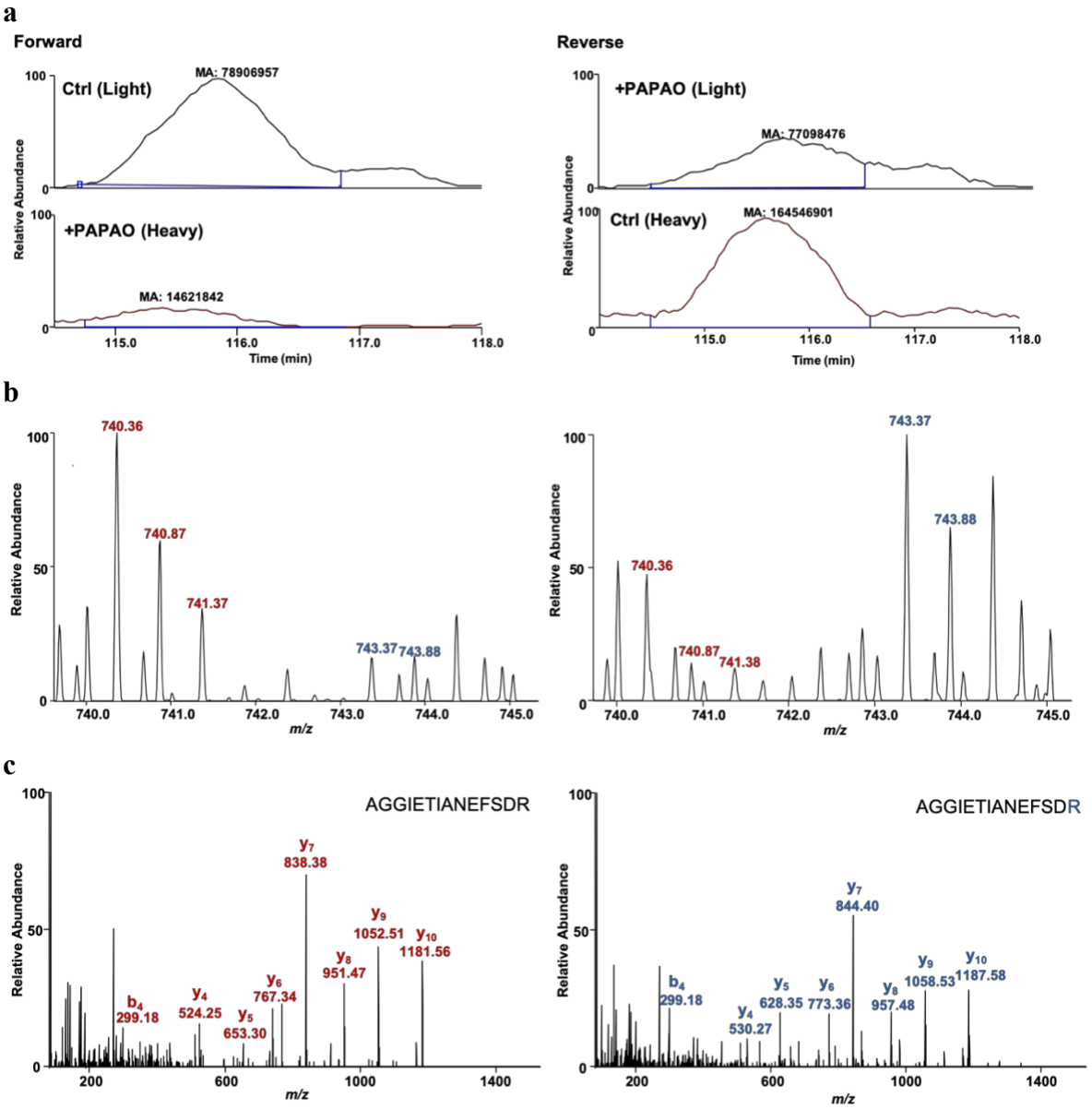


Table 3. 3 Accessible surface area and relative accessibility of HSP110 (PDB:6GFA) and corresponding cysteines position in HSPA4.

No.(HSP110)	No.(HSPA4)	aa	Area(Å2)	Relative (0.0 - 1.0)
12	12	Cys	**0.077**	**0.001**
33	33	Cys	**36.314**	**0.253**
139	139	Cys	**0.000**	**0.000**
166	166	Cys	**20.676**	**0.144**
212	212	Cys	**0.415**	**0.003**
244	244	Cys	**6.801**	**0.047**
269	269	Cys	**0.000**	**0.000**
289	289	Cys	**61.069**	**0.425**
309	309	Cys	**0.454**	**0.003**
375	375	Cys	**0.000**	**0.000**
379	379	Cys	**71.886**	**0.385**

Figure 3. 6 Solvent-accessible surface of HSP110. Sphere structure representation of HSP110 in two different orientations rotated 180° around z axis. The cysteine residues with solvent accessibility are colored by atoms and sulfur atoms are in yellow. ATP is labeled in red. The rest of protein is in gray.

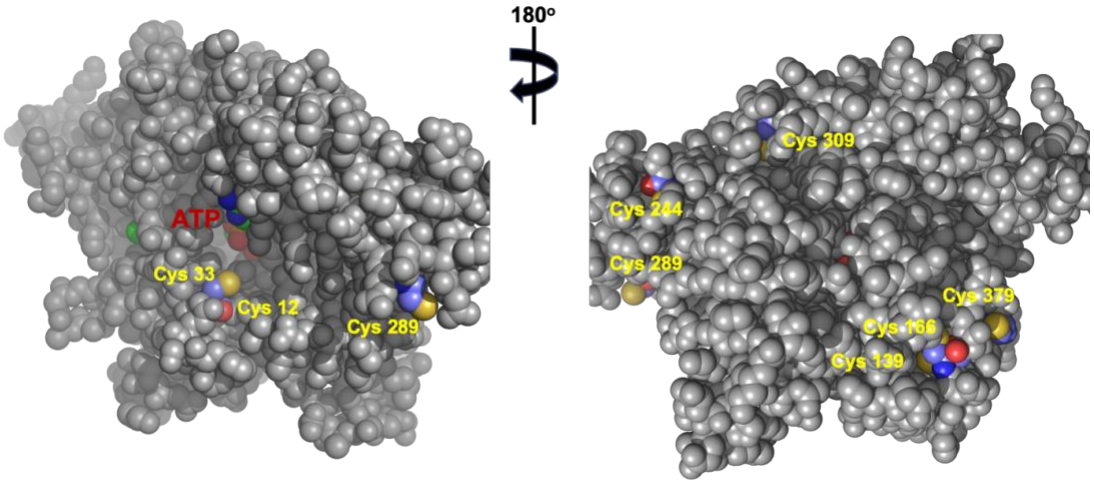
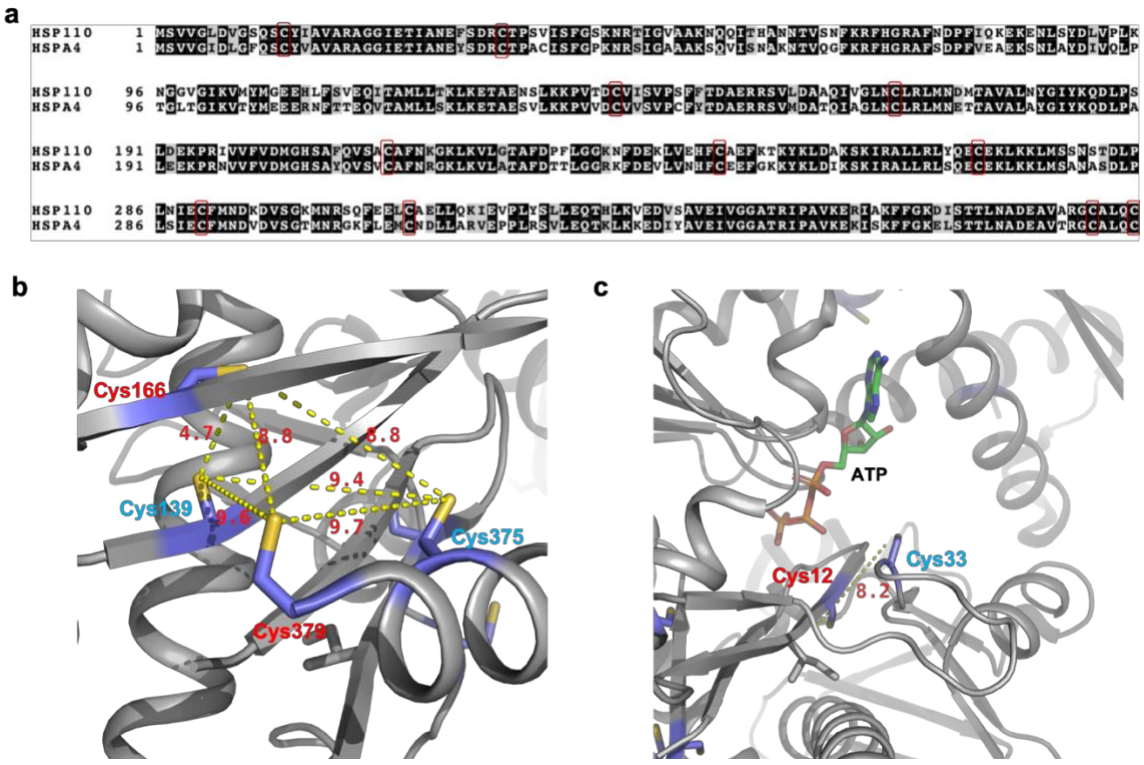


Figure 3. 7 Molecular modeling of HSPA4 and its interaction with arsenic. (a) A sequence comparation of HSP110 and HSPA4 revealed that cysteine residues in the two proteins are highly conserved. (b) Based on the X-ray crystal structure of HSP110 (PDB code: 6GFA), Cys139, Cys166, Cys375 and Cys379 are located in cysteine-rich region, where the distances between closely located Cys residues are labeled. (c) Cys12 and Cys33 are located near the ATP-binding site. For comparison, all Cys residues and ATP are represented in stick mode, where sulfur atom is shown in yellow. Cys residues with high solvent accessibility are labeled in red. Hydrogen atoms are omitted for better clarity.



References

1. Litzow, M. R., Arsenic trioxide. *Expert Opin. Pharmacother.* **2008,** *9* (10), 1773-1785.

2. Dilda, P. J.; Hogg, P. J., Arsenical-based cancer drugs. *Cancer Treat. Rev.* **2007,** *33* (6), 542-64.

3. Saint-Jacques, N.; Parker, L.; Brown, P.; Dummer, T. J., Arsenic in drinking water and urinary tract cancers: a systematic review of 30 years of epidemiological evidence. *Environ. Health* **2014,** *13* (1), 44.

4. Yu, H.-S.; Liao, W.-T.; Chai, C.-Y., Arsenic Carcinogenesis in the Skin. *J. Biomed. Sci.* **2006,** *13* (5), 657-666.

5. Wild, P.; Bourgkard, E.; Paris, C., Lung Cancer and Exposure to Metals: The Epidemiological Evidence. In *Methods Mol. Biol.*, Humana Press: 2009; pp 139-167.

6. Jomova, K.; Jenisova, Z.; Feszterova, M.; Baros, S.; Liska, J.; Hudecova, D.; Rhodes, C. J.; Valko, M., Arsenic: toxicity, oxidative stress and human disease. *J. Appl. Toxicol.* **2011**, n/a-n/a.

7. Avram, S.; Udrea, A. M.; Negrea, A.; Ciopec, M.; Duteanu, N.; Postolache, C.; Duda-Seiman, C.; Duda-Seiman, D.; Shaposhnikov, S., Prevention of Deficit in Neuropsychiatric Disorders through Monitoring of Arsenic and Its Derivatives as Well as Through Bioinformatics and Cheminformatics. *Int. J. Mol. Sci.* **2019,** *20* (8), 1804.

8. Yoshinaga-Sakurai, K.; Shinde, R.; Rodriguez, M.; Rosen, B. P.; El-Hage, N., Comparative Cytotoxicity of Inorganic Arsenite and Methylarsenite in Human Brain Cells. *ACS Chem. Neurosci.* **2020,** *11* (5), 743-751.

9. Chen, C.-J.; Wang, S.-L.; Chiou, J.-M.; Tseng, C.-H.; Chiou, H.-Y.; Hsueh, Y.-M.; Chen, S.-Y.; Wu, M.-M.; Lai, M.-S., Arsenic and diabetes and hypertension in human populations: A review. *Toxicol. Appl. Pharmacol.* **2007,** *222* (3), 298-304.

10. States, J. C.; Srivastava, S.; Chen, Y.; Barchowsky, A., Arsenic and Cardiovascular Disease. *Toxicol. Sci.* **2009,** *107* (2), 312-323.

11. Navas-Acien, A.; Sharrett, A. R.; Silbergeld, E. K.; Schwartz, B. S.; Nachman, K. E.; Burke, T. A.; Guallar, E., Arsenic Exposure and Cardiovascular Disease: A Systematic Review of the Epidemiologic Evidence. *Am. J. Epidemiol.* **2005,** *162* (11), 1037-1049.

12. Ferzand, R.; Ali Gadahi, J.; Saleha, S.; Ali, Q., Histological and Haematological Disturbance Caused by Arsenic Toxicity in Mice Model. *Pak. J. Biol. Sci.* **2008,** *11* (11), 1405-1413.

13. Antman, K. H., Introduction: The History of Arsenic Trioxide in Cancer Therapy. *The Oncologist* **2001,** *6* (S2), 1-2.

14. Doyle, D., Notoriety to respectability: a short history of arsenic prior to its present day use in haematology. *Br. J. Haematol.* **2009,** *145* (3), 309-317.

15. Wang, Z.-Y.; Chen, Z., Acute promyelocytic leukemia: from highly fatal to highly curable. *Blood* **2008,** *111* (5), 2505-2515.

16. Hayakawa, F.; Privalsky, M. L., Phosphorylation of PML by mitogen-activated protein kinases plays a key role in arsenic trioxide-mediated apoptosis. *Cancer Cell* **2004,** *5* (4), 389-401.

17. Huang, C.; Yin, Q.; Zhu, W.; Yang, Y.; Wang, X.; Qian, X.; Xu, Y., Highly Selective Fluorescent Probe for Vicinal-Dithiol-Containing Proteins and In Situ Imaging in Living Cells. *Angew. Chem. Int. Ed.* **2011,** *50* (33), 7551-7556.

18. Shi, W.; Dong, J.; Scott, R. A.; Ksenzenko, M. Y.; Rosen, B. P., The Role of Arsenic-Thiol Interactions in Metalloregulation of the ars Operon. *J. Biol. Chem.* **1996,** *271* (16), 9291-9297.

19. Shen, S.; Li, X.-F.; Cullen, W. R.; Weinfeld, M.; Le, X. C., Arsenic Binding to Proteins. *Chem. Rev.* **2013,** *113* (10), 7769-7792.

20. Hughes, M. F., Arsenic toxicity and potential mechanisms of action. *Toxicol. Lett.* **2002,** *133* (1), 1-16.

21. Kapahi, P.; Takahashi, T.; Natoli, G.; Adams, S. R.; Chen, Y.; Tsien, R. Y.; Karin, M., Inhibition of NF-κB Activation by Arsenite through Reaction with a Critical Cysteine in the Activation Loop of IκB Kinase. *J. Biol. Chem.* **2000,** *275* (46), 36062-36066.

22. Andrew, A. S.; Karagas, M. R.; Hamilton, J. W., Decreased DNA repair gene expression among individuals exposed to arsenic in United States drinking water. *Int. J. Cancer* **2003,** *104* (3), 263-268.

23. Lu, M.; Wang, H.; Li, X.-F.; Lu, X.; Cullen, W. R.; Arnold, L. L.; Cohen, S. M.; Le, X. C., Evidence of Hemoglobin Binding to Arsenic as a Basis for the Accumulation of Arsenic in Rat Blood. *Chem. Res. Toxicol.* **2004,** *17* (12), 1733-1742.

24. Hoffman, R. D.; Lane, M. D., Iodophenylarsine oxide and arsenical affinity chromatography: new probes for dithiol proteins. Application to tubulins and to components of the insulin receptor-glucose transporter signal transduction pathway. *J. Biol. Chem.* **1992,** *267* (20), 14005-14011.

25. Menzel, D. B.; Hamadeh, H. K.; Lee, E.; Meacher, D. M.; Said, V.; Rasmussen, R. E.; Greene, H.; Roth, R. N., Arsenic binding proteins from human lymphoblastoid cells. *Toxicol. Lett.* **1999,** *105* (2), 89-101.

26. Moaddel, R.; Sharma, A.; Huseni, T.; Jones, G. S.; Hanson, R. N.; Loring, R. H., Novel Biotinylated Phenylarsonous Acids as Bifunctional Reagents for Spatially Close Thiols: Studies on Reduced Antibodies and the Agonist Binding Site of ReducedTorpedoNicotinic Receptors. *Bioconjugate Chem.* **1999,** *10* (4), 629-637.

27. Lin, C.-H.; Huang, C.-F.; Chen, W.-Y.; Chang, Y.-Y.; Ding, W.-H.; Lin, M.-S.; Wu, S.-H.; Huang, R.-N., Characterization of the Interaction of Galectin-1 with Sodium Arsenite. *Chem. Res. Toxicol.* **2006,** *19* (3), 469-474.

28. Chang, K. N.; Lee, T. C.; Tam, M. F.; Chen, Y. C.; Lee, L. W.; Lee, S. Y.; Lin, P. J.; Huang, R. N., Identification of galectin I and thioredoxin peroxidase II as two arsenic-binding proteins in Chinese hamster ovary cells. *Biochem. J* **2003,** *371* (2), 495-503.

29. Zhang, H.-N.; Yang, L.; Ling, J.-Y.; Czajkowsky, D. M.; Wang, J.-F.; Zhang, X.-W.; Zhou, Y.-M.; Ge, F.; Yang, M.-K.; Xiong, Q.; Guo, S.-J.; Le, H.-Y.; Wu, S.-F.; Yan, W.; Liu, B.; Zhu, H.; Chen, Z.; Tao, S.-C., Systematic identification of arsenic-binding proteins reveals that hexokinase-2 is inhibited by arsenic. *Proceedings of the National Academy of Sciences* **2015,** *112* (49), 15084-15089.

30. Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M., Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Sci* **2009,** *325* (5942), 834-840.

31. Yan, X.; Li, J.; Liu, Q.; Peng, H.; Popowich, A.; Wang, Z.; Li, X.-F.; Le, X. C., p-Azidophenylarsenoxide: An Arsenical "Bait" for the In Situ Capture and Identification of Cellular Arsenic-Binding Proteins. *Angew. Chem. Int. Ed.* **2016,** *55* (45), 14051-14056.

32. Zhang, X.; Yang, F.; Shim, J.-Y.; Kirk, K. L.; Anderson, D. E.; Chen, X., Identification of arsenic-binding proteins in human breast cancer cells. *Cancer Lett.* **2007,** *255* (1), 95-106.

33. Ong, S.-E.; Mann, M., A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **2006,** *1* (6), 2650-2660.

34. Heredia-Moya, J.; Kirk, K. L., An improved synthesis of arsenic–biotin conjugates. *Biorg. Med. Chem.* **2008,** *16* (10), 5743-5746.

35. Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat. Methods* **2009,** *6* (5), 359-362.

36. Lu, J.; Chew, E.-H.; Holmgren, A., Targeting thioredoxin reductase is a basis for cancer therapy by arsenic trioxide. *Proceedings of the National Academy of Sciences* **2007,** *104* (30), 12288-12293.

37. Németi, B.; Gregus, Z., Reduction of Arsenate to Arsenite by Human Erythrocyte Lysate and Rat Liver Cytosol – Characterization of a Glutathione- and NAD-Dependent Arsenate Reduction Linked to Glycolysis. *Toxicol. Sci.* **2005,** *85* (2), 847-858.

38. Posey, T.; Weng, T.; Chen, Z.; Chintagari, N. R.; Wang, P.; Jin, N.; Stricker, H.; Liu, L., Arsenic-induced changes in the gene expression of lung epithelial L2 cells: implications in carcinogenesis. *BMC Genomics* **2008,** *9* (1), 115.

Chapter 4 Proteome-wide Characterizations of N^6-Methyl-ATP- and N^6-Furfuryl-ATP-Binding Capabilities of Kinases

4.1 Introduction

Protein kinases are encoded by more than 500 distinct genes, which constitute approximately 2% of all protein-coding human genes.[1-3] Studies in the last several decades established central roles of protein kinases in the regulation of crucial biological processes, including cell signaling, proliferation and metabolism.[4, 5] Amplifications, mutations, and abnormal expression/activation of kinase genes are associated with and contribute to human diseases.[1, 4] Therefore, kinase inhibitors and kinase-substrate interactions have been intensively investigated and exploited for the therapeutic interventions of various human diseases.[6]

Multiple kinases in different cell signaling pathways may converge on common phosphorylation sites in a substrate protein and sequential phosphorylations may occur in kinase-mediated signaling cascades, which render it very challenging to identify direct substrates for specific kinases.[7-9] To address this challenge, Shokat et al.[10, 11] introduced a creative approach, termed analog substrate-kinases, to define kinase-substrate relationships. The approach harnesses the power of chemical syntheses of N^6-modified ATP analogs or purine analogs and genetic engineering of the kinase of interest to accommodate these analogs for substrate phosphorylation. We reason that it is important to assess, at the proteome-wide scale, whether these analog substrates can also be recognized by endogenous kinases. Such characterizations of endogenous kinases that can recognize the ATP/purine analogs are important for rigorous interpretation of data obtained from analog

substrate kinases, where substrates phosphorylated by these off-target endogenous kinases may yield false-positive results. Additionally, the results obtained from these characterizations may also provide an opportunity to discover substrates for native kinases in cells. In this vein, N^6-furfuryl-ATP (a.k.a. kinetin triphosphate, KTP) was found to enhance the activity of wild-type PINK1 kinase and restore the catalytic activity of a Parkinson's disease-related mutant (i.e. G309D) of PINK1 to near-wild-type levels.[11]

Targeted proteomics, relying on multiple-reaction monitoring (MRM) on a triple-quadrupole mass spectrometer or parallel-reaction monitoring (PRM) on an Orbitrap mass spectrometer, have the attributes of providing enhanced sensitivity, reproducibility and accuracy relative to discovery proteomics.[12, 13] Recently a scheduled MRM-based targeted proteomic method, along with the use of isotope-coded desthiobiotin-ATP acyl phosphate probes, was employed for proteome-wide assessment of ATP-binding capabilities of kinases. The chemoproteomic method facilitates high-throughput interrogation of approximately 300 human kinases.[14]

In this study, we applied the chemoproteomic approach to examine the relative abilities of the human kinome in binding to N^6-methyl-ATP (N^6-Me-ATP) and KTP vs. native ATP (Figure 1a). We also revealed the ability of N^6-Me-ATP, but not KTP, as a phosphate group donor for GSK3β, one of the kinases that can bind to both nucleotide analogs, to phosphorylate its substrate peptide. Furthermore, we employed molecular dynamics (MD) simulations to understand the differences in interactions between GSK3β and N^6-Me-ATP and KTP.

4.2 Experimental Section

4.2.1 Chemical Synthesis of Isotope-Coded ATP Affinity Probe

The stable isotope-coded desthiobiotin-ATP acyl phosphate probes were synthesized in two steps procedure, first by conjugating desthiobiotin with γ-aminobutyric acid, and then through reacting the resulting conjugate with ATP. Desthiobiotin (10 mg) and tri-*n*-butylamine (11 μL) were dissolved in a 1-mL solvent mixture of ice-cold CH_2Cl_2 and DMF (4:1, v/v). After stirring at 0 °C for 5 min, ethyl chloroformate (6 μL) was added under an argon atmosphere, and the solution was maintained at 0 °C with stirring for 10 min, followed by stirring at room temperature for another 60 min. To the above reaction mixture was added ATP (50 mg) and the reaction was continued at room temperature under an argon atmosphere for 18 hr. Dichloromethane was removed by purging the reaction mixture with argon. Crude products were re-dissolved in water, loaded onto a SynergiTM 4 μm Fusion-RP 80 Å LC Column (10×150 mm, Phenomenex), and eluted using a gradient of 0-70% acetonitrile in 50 mM triethylammonium acetate (TEAA, pH 6.8). Appropriate eluted fractions were pooled, lyophilized, and stored at -80°C. The structures of the products were confirmed by ESI-MS analyses (Fig. S1).

4.2.2 Cell culture

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Corning), 100 U/ml penicillin and 100

μg/ml streptomycin (Cytiva). The cells were maintained in a humidified atmosphere with 5% CO_2 at 37°C and the culture medium was changed in every 2-3 days as needed.

Sample preparation and isotope-coded ATP affinity probe labeling

HEK293T cells were harvested at ~80% confluency, and washed for three times with cold PBS. The cells were then lysed by incubating, at 0 °C for 30 min, in a buffer containing 50 mM HEPES (pH 7.4), 100 mM NaCl, 0.5 mM EDTA, 0.7% CHAPS, and a protease inhibitor cocktail (1:100, v/v). The resulting mixture was centrifuged at 16000*g* and 4°C for 30 min. After removal of endogenous nucleotides using a NAP-10 column (Amersham Biosciences), proteins were eluted into a buffer containing 50 mM HEPES (pH 7.4), 75 mM NaCl, and 5% glycerol. Protein concentrations in the resultant lysates were subsequently quantified by Quick Start Bradford Protein Assay (Bio-Rad, Hercules, CA).

The isotope-labeled desthiobiotin-C3-ATP probes were synthesized previously.[15] $MgCl_2$, $MnCl_2$, and $CaCl_2$ were added to the concentrated protein lysate immediately before the labeling reaction, and their final concentrations were 50, 5 and 5 mM, respectively. The resulting protein lysates (1.5 mg each) were incubated with or without 100 μM N^6-Me-ATP or KTP at 4°C for 20 min. Heavy or light isotope-coded ATP affinity probe (100 μM) was subsequently added to the reaction mixture and incubated at room temperature with gentle shaking for 2 hr. The reactions were terminated by adding glycine until its final concentration was 50 mM. After incubation at room temperature for another 20 min, heavy and light ATP probe-labeled cell lysates were mixed at a 1:1 ratio (w/w).

For the concentration-dependent competition experiments, 1.5 mg protein lysate was pre-incubated separately with 10, 100, 200 μM N^6-Me-ATP or KTP, followed by incubation with 100 μM heavy ATP affinity probe, where the same amount of lysate was incubated with 100 μM light ATP affinity probe alone. The subsequent steps of sample preparation were the same as described above.

After ATP affinity probe labeling, the protein samples were subjected to filter-aided sample preparation (FASP)[16]. The desthiobiotin-conjugated peptides were enriched from the ensuing tryptic digestion mixture with high-capacity streptavidin beads (Sigma-Aldrich), and desalted using C18 pipet tips (Agilent), as described elsewhere.[15]

4.2.3 Scheduled LC-MRM analysis

Scheduled LC-MRM experiments were performed on a TSQ Altis triple-quadrupole mass spectrometer (Thermo Fisher, San Jose, CA) equipped with a nanoelectrospray ionization source, which was coupled to an UltiMate 3000 UPLC RSLCnano system (Dionex, Thermo Fisher), with the detailed conditions provided in the online Supporting Information. A 35-mm trapping column (150 μm i.d.) and a 250-mm analytical column (75 μm i.d., PicoTip Emitter, New Objective, Woburn, MA) were employed for the separation, and the two columns were packed in-house with reversed-phase ReproSil-Pur C18 material (120 Å in pore size, Dr. Maisch, Germany) with particle sizes of 5 and 3 μm, respectively.

Mobile phases A and B contained 0.1% formic acid in water (v/v) and 0.1% formic acid in 80.0% acetonitrile (v/v), respectively. The above-described desalted peptide samples were resolved using a 90-min linear gradient of 10–45% B at a flow rate of 300

nL/min. The spray voltage was 2200 V, and the ion transport tube temperature was 325°C. A resolution of 0.7 F.W.H.M. was set for both the first and third quadrupoles. Fragmentation of precursor ions in the collision cell (second quadrupole) was conducted with 1.5 mTorr argon, and the collision energy was derived from default settings in Skyline.[17] The cycle time was 3-5 s and varied with the total number of transitions per cycle. The LC-MS/MS data were processed using Skyline, version 19.1.0.193,[17] against the previously published kinome MRM library,[14] where we excluded those peptides with dotp[18] values for their fragment ions being < 0.7.

4.2.4 Purification of Recombinant GSK3β

The pGEX plasmid for expressing recombinant glutathione *S*-transferase (GST)-tagged GSK3β in *Escherichia coli* was purchased from Addgene (catalog number 15898), and the protein was purified using glutathione agarose beads following previously published procedures.[19] Briefly, the plasmid was transfected into BL21 *Escherichia coli* cells. The bacteria were grown in LB medium with 100 μg/mL of Ampicillin to OD_{600} 0.6-0.8 at 37 °C with shaking. IPTG was added until its final concentration reached 1 mM, and the culture was grown at 16 °C for an additional 16 h.

The cells were harvested by centrifugation at 7000 rpm at 4 °C, resuspended in 10 mM Tris-HCl (pH 7.2), and lysed by sonication on ice. The supernatant was collected after centrifugation. The crude cell extract, containing the desired GST-tagged proteins, was incubated with 800 μL glutathione agarose beads for 2 hr at 4 °C. After washing the beads to remove non-bound sample components, the desired proteins were then eluted with a

buffer containing 10 mM reduced glutathione. The purity of the GST-GSK3β was verified by SDS-PAGE analysis and visualized by Coomassie blue staining (Figure S2).

The elution fractions were concentrated by using Amicon Ultra-4 filter (10000 NMWL, Millipore), then purified by FPLC using a size-exclusion column (Superdex 200 Increase 10/300 GL) and a GE AKTA Purifier FPLC System, where phosphate-buffered saline (1×) was employed as the mobile phase. Fractions were monitored by UV absorbance at 280 nm, and those containing the desired protein were collected. The purity of protein was assessed with SDS-PAGE analysis (Figure S3).

4.2.5 *In vitro* Kinase Assay for GSK3β

Approximately 1 μM purified GST-GSK3β was incubated with 10 μM of a peptide substrate of GSK3β, YRRAAVPPSPSLSRHSSPHQ(pS)EDEEE (SignalChem Biotech Inc.), in a 50-μL reaction buffer (5 mM MOPS, pH 7.4, 0.1 mM EDTA, 10 mM $MgCl_2$, 1 mM sodium orthovanadate) containing 250 μM ATP, N^6-Me-ATP or KTP.[20, 21] The reaction was continued at 37°C for 2 h. A control reaction with heat-inactivated (boiled at 95°C for 5 min) GST-GSK3β was also included (Figure S6). The reaction mixtures were desalted with C18 Zip-Tip and injected for LC-MS and MS/MS analyses on an LTQ-XL liner ion-trap mass spectrometer. The mobile phases were 0.1 formic acid in water (Solution A) and 0.1% formic acid in acetonitrile (Solution B), and the gradient was 1-40% CH_3CN.

4.2.6 Molecular dynamics (MD) simulations

The initial structure for GSK3β was downloaded from the Protein Data Bank (ID: 1PYX).[22] The initial positions of the ligands ATP, KTP and N^6-Me-ATP in MD simulations were superimposed to that of AMP-PNP in 1PYX. Amber18 package with GPU implementation was used for unbiased all-atom MD simulations.[23] The Amber FF99SB and Li/Merz ion parameters were employed for GSK3β and Mg^{2+} ion, respectively.[24, 25] ATP force field parameters and partial charges were obtained from an existing publication using quantum mechanics at RHF/6-31+G* level of theory. Partial charges for adductions in KTP and N^6-Me-ATP were calculated using am1-bcc method, and General Amber Force Field was used for the force field parameters of adductions.[26] The force field parameters and partial charges for adenine, ribose and triphosphate in KTP and N^6-Me-ATP were taken from the corresponding parameters of ATP. A Mg^{2+} ion was placed between a ligand and protein to maintain the positions of ATP phosphates. Minimization of hydrogen atoms, the side chains and the entire system was executed for 500, 5000 and 3000 steps, respectively. The system was placed under implicit Hawkins, Cramer, Truhlar pairwise generalized Born (GB) model.[27, 28] The system was then slowly heated up to 50, 100, 150, 200, 250 K for 100 ps at each temperature. A 50 ns classical MD was subsequently performed with restrained ligand position at 298 K, which only allowed protein to move. Langevin Thermostat was used to maintain a temperature of 298K. Frames were saved in every 10 ps with a time step of 2 fs. Finally, SHAKE algorithm was applied to constrain the covalent bonds involving hydrogen atoms.[29]

4.3 Results and Discussion

4.3.1 A chemoproteomic approach for interrogating the N^6-Me-ATP- and KTP-binding capabilities of the human kinome

In this study, we aim to develop a targeted proteomic strategy to evaluate, at the proteome-wide scale, the binding capabilities of kinases toward two N^6-modified ATP derivatives, i.e., N^6-Me-ATP and KTP, relative to unmodified ATP. These two ATP analogs were chosen because the former represents the N^6-alkylated ATP derivative with the smallest alkyl group and KTP was previously found to be recognized by an endogenous kinase, i.e., PINK1.[11]

Our method is based on the use of the previously synthesized stable isotope-labeled desthiobiotin-ATP acyl phosphate probes (Figure 4.1a, Figure 4.2)[15] and a recently reported scheduled MRM method for high-throughput proteome-wide assessment of ATP-binding affinities of human kinases.[14] In particular, the kinome MRM library includes the normalized retention times (iRTs)[30] and at least three MRM transitions for each of the 818 non-redundant desthiobiotin-labeled light and heavy peptides derived from 474 kinases. The library encompasses 409 protein kinases and covers approximately 80% of the human kinome.

We first performed a competition experiment, where equal amounts of protein lysates of HEK293T cells were treated separately with 100 μM light and heavy desthiobiotin-C3-ATP probes in the absence or presence of the same concentration of N^6-Me-ATP or KTP. To achieve robust quantification, we conducted both forward and reverse chemical labeling experiments, where the lysates incubated with heavy and light probes were pre-incubated

with the ATP analogs in the forward and reverse labeling experiments, respectively (Figure 4.1b).

We manually checked all the quantified peptides to ensure that the signal intensity distributions of selected transitions match with those in the Skyline spectral library, with dotp values being larger than 0.7 (Figure 4.3a). On the basis of full-scan MS/MS acquired from shotgun proteomic experiments, transitions for the formation of the three most abundant y-ions were selected for the scheduled MRM analysis of the corresponding light- and heavy-labeled kinase peptides (e.g. the y_7, y_8, and y_9 ions in Figure 4.3b). The distribution patterns of the three y-ions derived from the light- and heavy-labeled peptides in both forward and reverse replicates were very similar to that in the full-scan MS/MS in the library (Figure 4.3c). In addition, the observed retention times and calculated iRTs for the 263 quantified desthiobiotin-conjugated peptides exhibited excellent linear correlations (Figure 4.3d). Moreover, the peptide intensity ratios obtained from forward and reverse labeling experiments are consistent for the quantified peptides (Figure 4.3e). These results together support the reliable identification and quantification of the kinase peptides.

When the ATP-binding pocket of putative ATP analog-targeted kinases is occupied by N^6-modified ATP derivatives, it will prevent the kinase from binding with the ATP acyl phosphate probe; hence, kinases exhibiting low labeling ratios of analog-treated/untreated samples are considered the candidate target kinases for the ATP analogs (Figure 4.3e). The target kinases of the tested ATP analogs are shown in the bottom quadrant of the scatter plots displaying the logarithmic ($\log_2$) ratios of the identified kinases in forward and reverse labeling experiments.

The LC-MRM results enabled the quantification of 300 unique peptides derived from 215 kinases from the N^6-Me-ATP competition experiments, and 306 distinct peptides representing 218 kinases from the KTP competition experiments. As expected, most kinases exhibit labeling ratios that are close to unity, revealing their preferential binding to ATP than its N^6-modified counterparts. We attempted to include all the putative kinases that can accept the two ATP analogs; hence, we employed a lenient criterion with the analog-treated/untreated ratio being lower than 0.75. With this criterion, our competition experiments led to the identifications of 27 and 18 kinases with capabilities in binding toward KTP (Table 4.1) and N^6-Me-ATP (Table 4.2), respectively. Among them, 9 kinases were targeted by both ATP analogs, including two isoforms of glycogen synthase kinase-3 (GSK3α and GSK3β), which display ratios of ~ 0.50 and ~0.65 for N^6-Me-ATP and KTP competition experiments, respectively (Figures 4.4 and 4.5). GSK3α and GSK3β are key kinases contributing to aberrant phosphorylation in Alzheimer's disease.[31, 32] In addition, these kinases can be inhibited by lithium[33, 34] and lithium is a mood stabilizer used in the treatment of bipolar disorder.[35]

Another targeted kinase, eukaryotic elongation factor 2 kinase (eEF2K), exhibits ratios of 0.60 and 0.69 in KTP and N^6-Me-ATP competition experiments, respectively. eEF2K is an atypical protein kinase that phosphorylates and inactivates eEF2, thereby negatively regulating translation elongation.[36] In addition, a tryptic peptide derived from CSKN2A2 (α2 subunit of casein kinase 2), i.e., VLGTEELYGYLKK, exhibit a ratio of 0.70 in the KTP competition experiment. This result is in keeping with a previous observation that

KTP can support the casein kinase 2-mediated phosphorylation of hungtingtin N17 domain.[37]

To further validate that the diminution in ATP probe labeling efficiency is dependent on the interaction between the ATP analogs and kinase, we next asked if the labeling efficiencies for the kinases decreases with the concentrations of the ATP analogs. To this end, we pre-incubated the same amount of protein lysate of HEK293T cells individually with 10, 100, or 200 μM ATP analog and subsequently labeled the lysate with 100 μM heavy ATP acyl phosphate probe, whereas the same amount of protein lysate without preincubation with the ATP analog was incubated with 100 μM light ATP probe. Our results showed that the ATP probe labeling efficiency of GSK3α gradually decreases with increasing concentrations of KTP or N^6-Me-ATP (Figure 4.6). For instance, the ATP probe labeling efficiencies for GSK3α decreased by 15%, 44% and 64% when the lysate was preincubated with 10, 100 and 200 μM N^6-Me-ATP, respectively (Figure 4.6 a-c). This result indicates that GSK3α exhibits similar binding affinities toward N^6-Me-ATP and ATP. The corresponding experiments with KTP revealed drops in labeling efficiencies by 14%, 35%, and 41%, respectively, suggesting that KTP can also bind to GSK3α, albeit at a weaker affinity than ATP (Figure 4.6 d-f). Similar results were obtained for GSK3β, though the extents of decreases differ somewhat from what we found for GSK3α (Figure 4.7). These results suggest that the alterations in ATP probe labeling efficiency can allow for determination of relative binding affinities of ATP analogs toward kinases. The quantitative results for other kinases are listed in Tables S1 and S2.

4.3.2 N^6-Me-ATP, but not KTP, could support the GSK3β-mediated phosphorylation reaction *in vitro*

Kinases catalyze the transfer of the γ phosphate in ATP to their substrate proteins or small molecules. The aforementioned chemoproteomic results revealed that the two ATP analogs exhibit appreciable binding affinities toward GSK3α and GSK3β. Hence, we next conducted *in-vitro* kinase assay to examine whether the two ATP analogs can serve as phosphate group donors for GSK3β.

While GSK3 can phosphorylate directly a few non-primed substrates with the target serine preceding a proline residue, most common targets for GSK3-mediated phosphorylation reside in a sequence of S/T-X-X-X-S/T(P), where the kinase phosphorylates a serine or threonine four residues N-terminal to a pre-phosphorylated serine/threonine.[38] Hence, we performed *in-vitro* phosphorylation reaction by incubating recombinant GST-GSK3β (Figure 4.8, Figure 4.9 see Experimental Section) with a primed substrate peptide, YRRAAVPPSPSLSRHSSPHQ(pS)EDEEE, which was derived from glycogen synthase 1, together with the same concentrations of ATP, N^6-Me-ATP or KTP. We subsequently monitored the reaction mixtures using LC-MS/MS (Figure 4.10 a-c), and compared the relative levels of phosphorylation based on the relative abundances of the $[M+3H]^{3+}$ ions of the mono-, di- , tri-phosphorylated forms of the GSK3β substrate peptide. Our results showed that KTP is not a desirable phosphate group donor, with only 6.6% substrate peptide being phosphorylated. Under the same experimental conditions, N^6-Me-ATP, however, enabled the phosphorylation of 67.8% of the substrate peptide, which is

very similar as ATP in enabling the GST-GSK3β-mediated phosphorylation of its substrate peptide (Figure 4.10 d). Controls were also been monitored. (Figure 4.11)

4.3.3 Molecular dynamics simulation for understanding the differences of N^6-Me-ATP and KTP in mediating the GSK3β-mediated phosphorylation

Our above proteomic data revealed that both N^6-Me-ATP and KTP are capable of competing with ATP in binding toward GSK3β; only the former ATP analog, however, could support the GSK3β-mediated *in-vitro* kinase reaction. To understand the molecular origins of these differences, we therefore conducted MD simulations to examine the activation loop dynamics and interactions between the kinase and the three ATP derivatives. In this vein, it is worth noting that, although molecular docking is a powerful method for examining protein-ligand interactions, protein dynamics provides molecular insights into kinase function.

We first examined if the three ligands can stably occupy the ATP-binding pocket of the kinase. Generally, proteins with an unfit ligand inside the binding pocket rearrange much more extensively than that bound tightly with a ligand. To quantify protein conformational flexibility, we calculated protein backbone root-mean-square-fluctuation (RMSF) with restrained ligand position during MD simulations. In this vein, increased protein motion leads to large RMSF values. As illustrated in Figure 4.12, GSK3β formed a stable complex with ATP and N^6-Me-ATP; upon binding with the two ligands, the kinase exhibited very similar fluctuation. GSK3β, however, fluctuates more substantially upon binding with KTP,

especially in the loop regions near amino acid residues Ile240 and Ala300. Because these two loop regions do not form interactions that are essential for stable ligand binding at the active site, the loop motion unlikely affects the binding of KTP with GSK3β.[39] Thus, the MD simulation results suggest that ATP, KTP and N^6-Me-ATP can all fit nicely into the ATP-binding pocket of GSK3β.

We next investigated the activation loop position to explore whether GSK3β can adopt an active conformation upon binding with the three nucleotides. When a kinase is catalytically competent, its activation loop needs to be in an active conformation for phosphate transfer. The active conformation is similar among different kinases; therefore, we used an active loop conformation of CDK2 (PDB ID: 1QMZ) as a reference structure for catalytically active form of kinases. We subsequently compared the loop of GSK3β complexes with the reference structure to examine the enzymatic activity (Figure 4.13a). During the entire course of MD run, the loop of the N^6-Me-ATP-GSK3β complex remained in an active position, similar as that for CDK2 (Figure 4.13b), suggesting that GSK3β can efficiently employ N^6-Me-ATP to phosphorylate its substrate.

Unlike the N^6-Me-ATP-GSK3β complex, active loop conformations of ATP-GSK3β and KTP-GSK3β existed, but the loop did not stay in the active configuration all the time. For example, in the first 32 ns of the MD run, the loop of the ATP-GSK3β complex was in an active configuration. Tyr216 of the loop subsequently formed hydrogen bond (H-bond) with ATP to hinder substrate binding for catalysis (Figure 4.13c). In this vein, it is worth noting that phosphorylated Tyr216, which increases the catalytic activity of GSK3β by approximately 200 fold,[40] is unlikely capable of forming H-bond with ATP. Moreover,

the H-bond between Tyr216 and ATP is exposed to solvent and can be disrupted by water molecules, which allows the loop to restore to an active position. Therefore, the ATP-GSK3β complex is catalytically active. The loop of KTP-GSK3β was also in the active form during the initial period of the MD run. After a 15 ns MD simulation, Cys218 of the loop started to form H-bond with KTP, which prohibits the loop from being accessed by a natural substrate (Figure 4.13d). The H-bond between Cys218 and KTP is more stable than that formed between Tyr216 and ATP. As a result, the dynamics of the activation loop suggests that, although the GSK3β-KTP complex can still be catalytically competent, its catalytic activity should be substantially reduced, which is in line with our results obtained from *in-vitro* biochemical assay. Close-up view of interactions between ATP, KTP and N^6-Me-ATP with GSK3β at different time steps (0 ns, 15 ns, 32 ns and 50 ns) can be found in Figures 4.14-4.16

4.4 Conclusions

We employed a chemoproteomic approach, based on the use of stable isotope-labeled ATP-affinity probes and scheduled MRM, to identify, at the proteome-wide level, kinase proteins that can bind to two N^6-modified ATP derivatives, i.e. N^6-Me-ATP and KTP. To the best of our knowledge, this is the first kinome-wide assessment about the interactomes of N^6-Me-ATP and KTP.

Kinases have long been shown to serve as central regulators of many crucial cellular processes including metabolism, membrane transport, gene expression, DNA repair, and cellular differentiation.[38] KTP was also identified as a neo-substrate for increasing the

activity of PINK1, which may constitute a novel approach for modulating kinase activity and drug discovery.[5] Here, our chemoproteomic approach led to the identifications of 27 and 18 putative kinases that can be targeted by KTP and N^6-Me-ATP, respectively. Many of these kinases are disease-related, e.g., GSK3α, GSK3β and eEF2K, as noted above. Therefore, our results revealed that the two ATP analogs can be recognized by a good number of endogenous kinases. This observation, hence, also calls for comprehensive characterizations of the interactions between the entire kinome and ATP/purine analogs when these analogs are employed in combination with engineered kinases for identifying direct substrates of kinases.

We also compared the relative binding affinities of kinases toward the ATP analogs and ATP. By employing different N^6-Me-ATP or KTP over ATP probe ratios of 1:10, 1:1 and 2:1, we observed progressive diminutions in ATP probe labeling efficiencies for many identified kinases. In this vein, the extent of labeling for GSK3α indicates that it has a very similar binding affinity to N^6-Me-ATP and ATP. With the *in-vitro* kinase assay, we further validated the ability of N^6-Me-ATP to activate endogenous GSK3β and catalyze the phosphorylation of its substrate.

By employing MD simulations, we found that N^6-Me-ATP-GSK3β, ATP-GSK3β and KTP-GSK3β complexes might all exist, but the activation loop position of KTP-GSK3β suggests the KTP-occupied GSK3β does not always assume an active conformation to support its enzymatic activity.

Together, our strategy allows for identification of the specific endogenous kinases that could accept unnatural N^6-modified ATP analogs, which provides important knowledge

for developing small-molecule inhibitors or activators of kinases, and for manipulating kinase-dependent bioorthogonal systems via specific ATP analog (s). It can be envisaged that the method can be adapted for uncovering kinases that can be targeted by other purine analogs and their nucleotide metabolites, e.g., 2-aminopurine and plant cytokinins.[41, 42] We expect that exploring further the biological effects modulated by ATP analogs may afford new venues for treating diseases resulting from dysregulated kinases.

Table 4. 1 Candidate kinase proteins with capabilities in binding toward KTP.

Protein Accession	Protein Gene	Mean Ratio	S.D.
Q16513	PKN2	0.538	0.221
O00418	EEF2K	0.599	0.19
P30876	POLR2B	0.605	0.298
P49840	GSK3A	0.607	0.157
Q9NWZ3	IRAK4	0.649	0.16
Q9Y2K2	SIK3	0.658	0.152
P48729	CSNK1A1	0.663	0.03
P41743	PRKCI	0.681	0.146
Q9H8X2	IPPK	0.686	0.05
Q9UJ70	NAGK	0.688	0.236
P78527	PRKDC	0.716	0.372
P19784	CSNK2A2	0.703	0.314
Q04759	PRKCQ	0.714	0.184
P49841	GSK3B	0.715	0.191
P54619	PRKAG1	0.717	0.224
P48426	PIP4K2A	0.719	0.344
Q96GX5	MASTL	0.719	0.512
Q8IWY7	TTBK2	0.726	0.016
Q5VST9	OBSCN	0.732	0.264
Q13263	TRIM28	0.732	0.31
P46734	MAP2K3	0.735	0.232
Q9P0L2	MARK1	0.736	0.196
Q9BYT3	STK33	0.738	0.008
P00558	PGK1	0.738	0.777
Q14680	MELK	0.74	0.149
P50750	CDK9	0.743	0.007
Q99538	LGMN	0.748	0.298

Table 4. 2 Candidate kinase proteins with capabilities in binding toward N^6-Me-ATP.

Protein Accession	Protein Gene	Mean Ratio	S.D.
Q9BRS2	RIOK1	0.501	0.057
P78527	PRKDC	0.513	0.238
P49841	GSK3B	0.54	0.168
P49840	GSK3A	0.549	0.164
P50613	CDK7	0.59	0.13
Q16513	PKN2	0.604	0.263
Q9P0L2	MARK1	0.625	0.029
Q14680	MELK	0.65	0.17
Q9H477	RBKS	0.68	0.078
O00418	EEF2K	0.694	0.188
Q96S44	TP53RK	0.702	0.128
P48729	CSNK1A1	0.704	0.116
Q9Y2K2	SIK3	0.711	0.091
Q15349	RPS6KA2	0.719	0.22
P53350	PLK1	0.722	0.475
O75676	RPS6KA4	0.726	0.5
P43405	SYK	0.736	0.194
Q9BZL6	PRKD2	0.736	0.115

Figure 4. 1 A chemoproteomic strategy for proteome-wide assessment of the binding capabilities of kinases toward ATP analogs. (a) The chemical structures of isotope-labeled ATP affinity probes; (b) A competition strategy for quantitative discovery of ATP analog-targeted kinases using stable isotope-coded ATP probes.

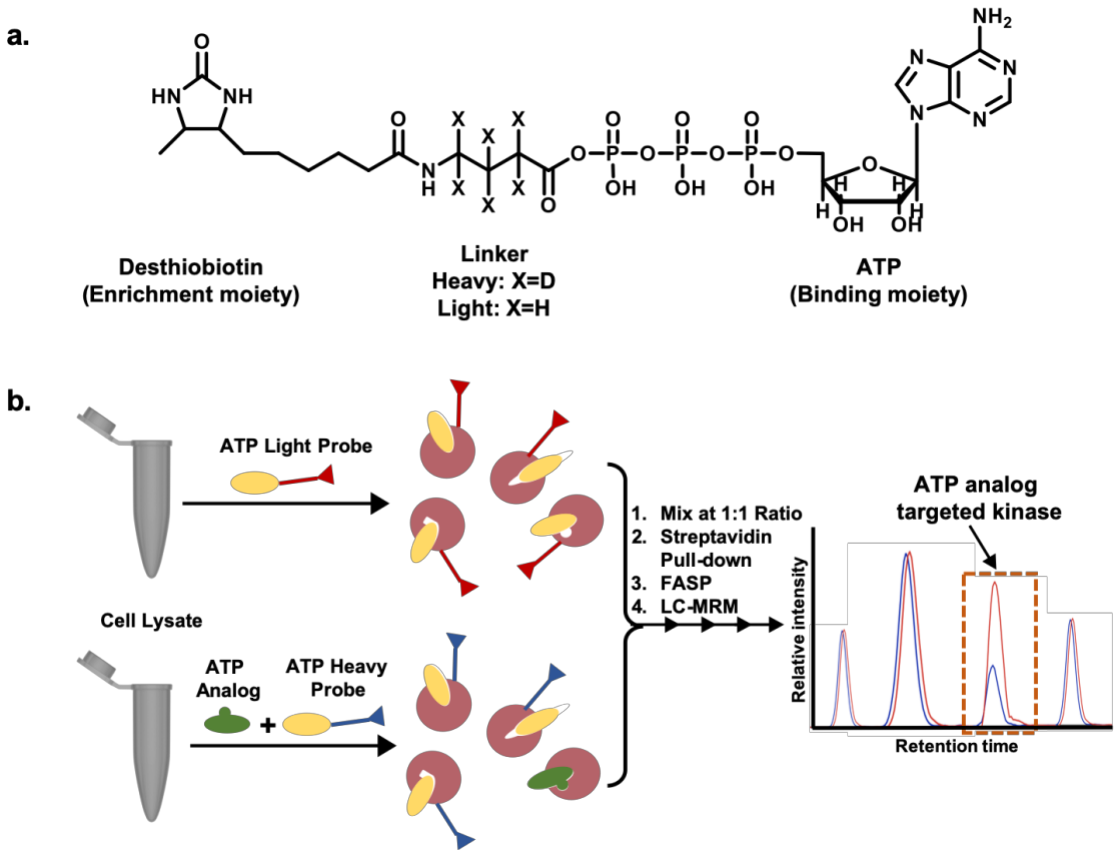


Figure 4. 2 Representative of ESI-MS of (a) the light desthiobiotin-C3 linker; (b) the heavy desthiobiotin-C3 linker; (c) the light desthiobiotin-C3-ATP probe; (d) the heavy desthiobiotin-C3-ATP probe.

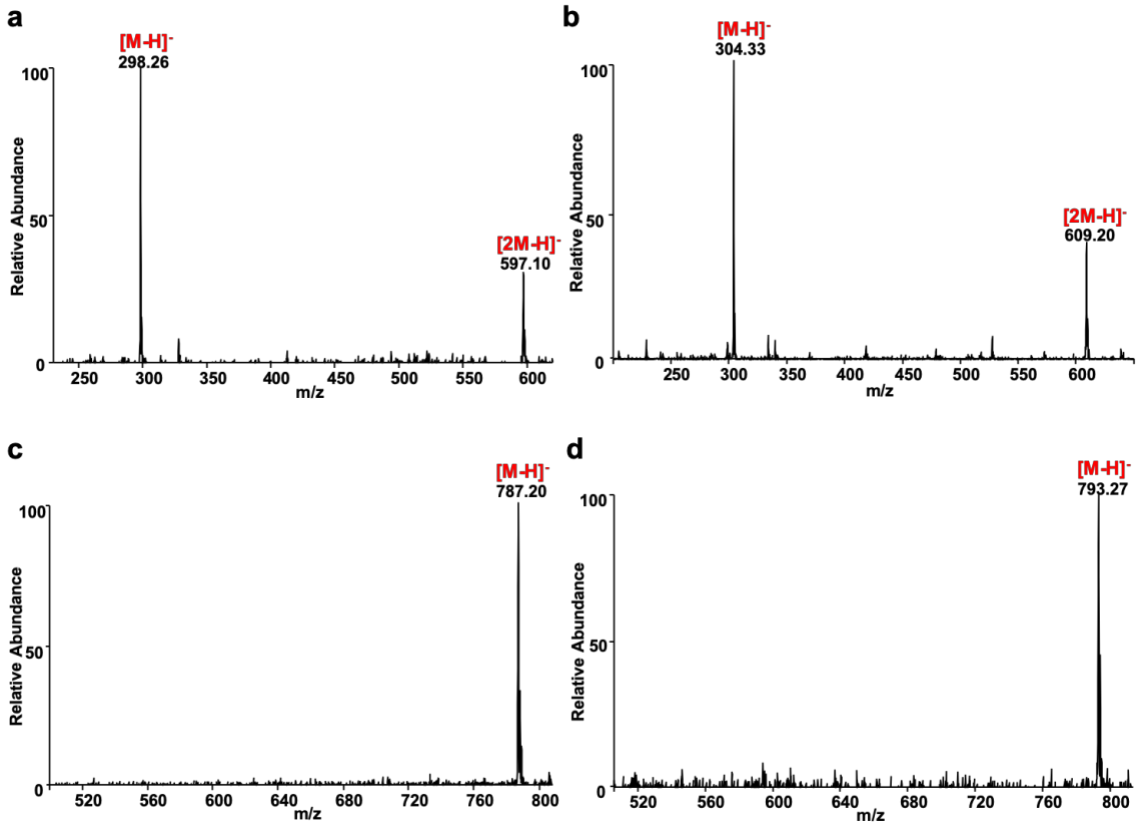


Figure 4. 3 Performance of the LC-MRM-based method for quantitative profiling of the binding capabilities of kinases toward N^6-Me-ATP and KTP. (a) Representative MS/MS of a probe-labeled tryptic peptide from GSK3β, y ions labeled in red represent the three transitions used for MRM analysis; (b) MS/MS of light and heavy precursors with the selected three transitions acquired from LC-MRM experiment; (c) A scatter plot showing the correlation between the observed retention times for desthiobiotin-labeled kinase peptides and their normalized retention times (iRTs) in the kinome MRM library; (d) The relative abundances of three fragment ions observed in DDA and MRM analyses(from forward and reverse probe labeling experiments); (e) Comparison of quantification results obtained from forward and reverse N^6-Me-ATP competition labeling experiments.

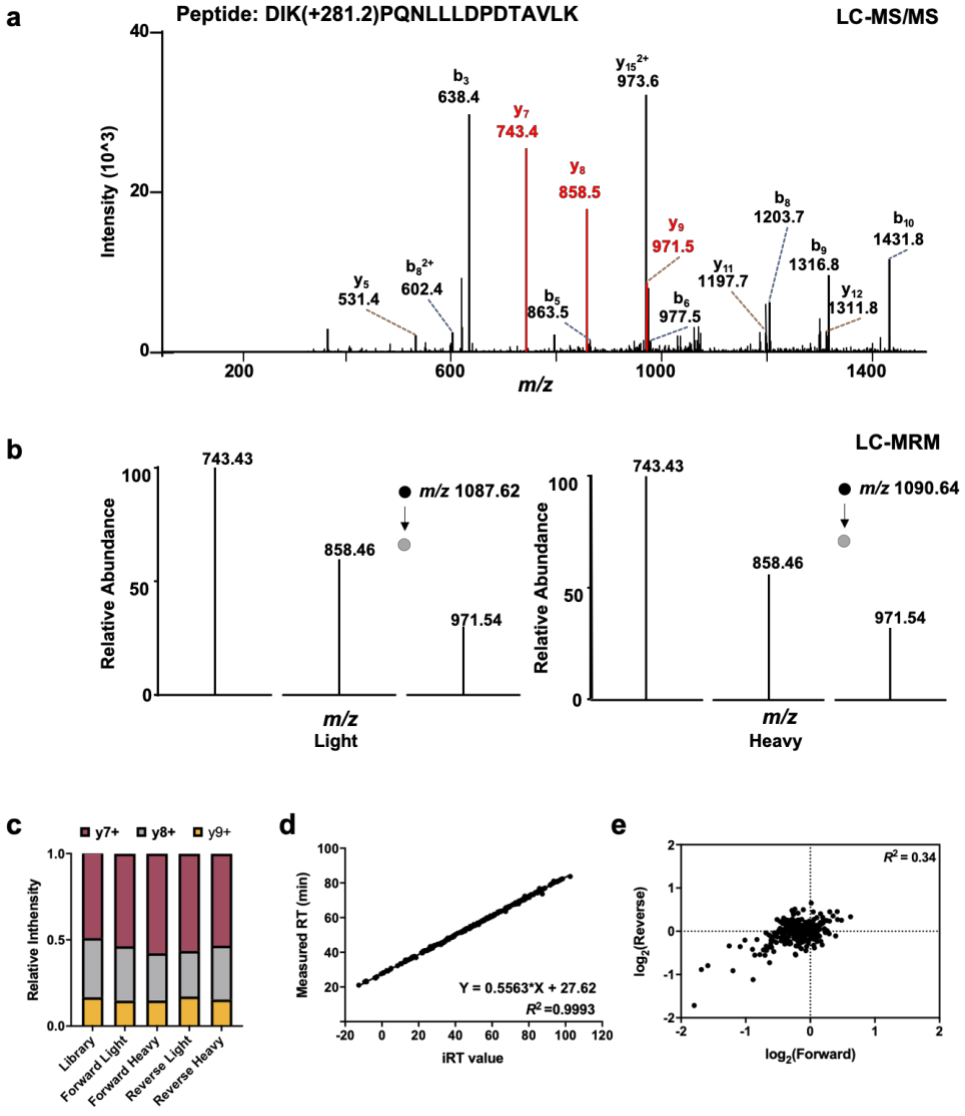


Figure 4. 4 MRM traces for the light and heavy forms of the probe-labeled peptide from GSK3α obtained from forward and reverse probe labeling experiments. (a-d) Selected-ion chromatograms for a tryptic peptide of GSK3α obtained from N^6-Me-ATP (a-b) and KTP (c-d) competition experiment. In forward and reverse labeling experiments, protein lysates pre-incubated with 100 μM N^6-Me-ATP or KTP were labeled with the heavy and light desthiobiotin-ATP acyl-phosphate probes, respectively, whereas the protein lysates without preincubation with the ATP analogs were labeled with the light and heavy desthiobiotin-ATP acyl-phosphate probe, respectively. The quantification results, as reflected by labeling ratios, obtained from LC-MRM analyses are listed in each pane

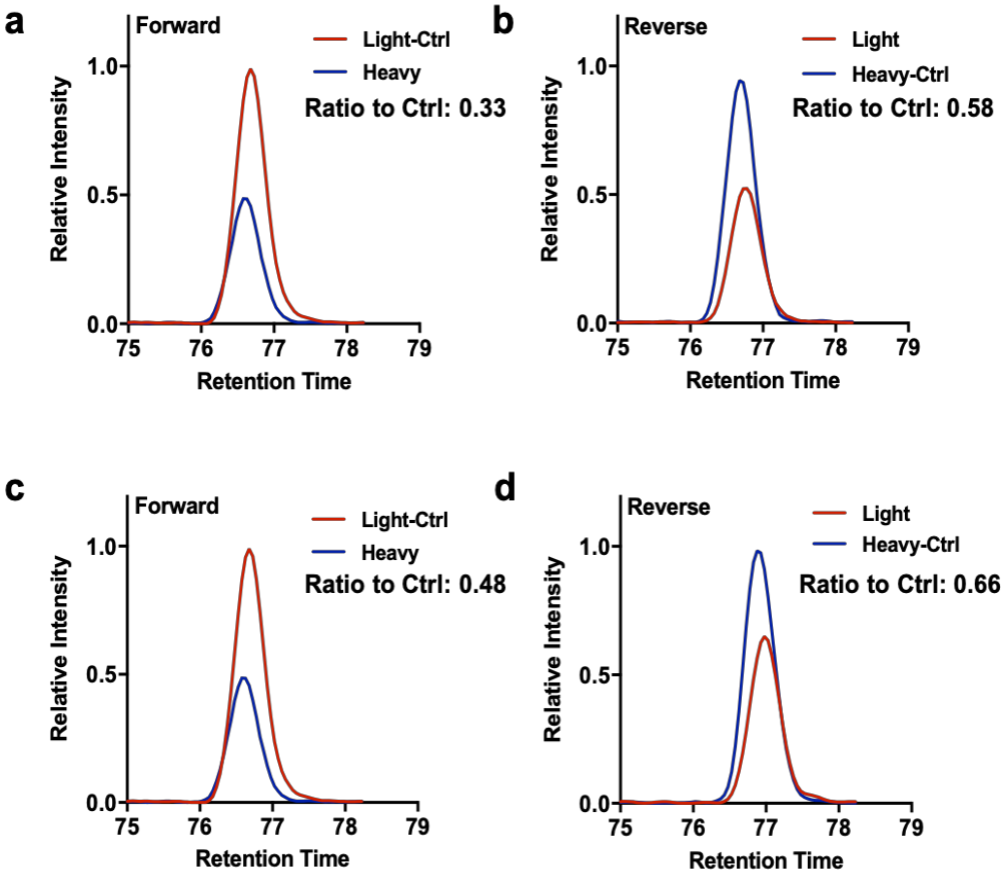


Figure 4. 5 MRM traces for the light and heavy forms of the probe-labeled peptide from GSK3β obtained from forward and reverse affinity probe labeling experiments. (a-d) Selected-ion chromatograms for a tryptic peptide of GSK3β obtained from N^6-Me-ATP (a-b) and KTP (c-d) competition experiments. In forward and reverse labeling experiments, protein lysates pre-incubated with 100 μM N^6-Me-ATP or KTP were labeled with the heavy and light desthiobiotin-ATP acyl-phosphate probes, respectively, whereas the protein lysates without preincubation with the ATP analogs were labeled with the light and heavy desthiobiotin-ATP acyl-phosphate probes, respectively. The quantification results, as reflected by labeling ratios, obtained from LC-MRM analyses are listed in each panel.

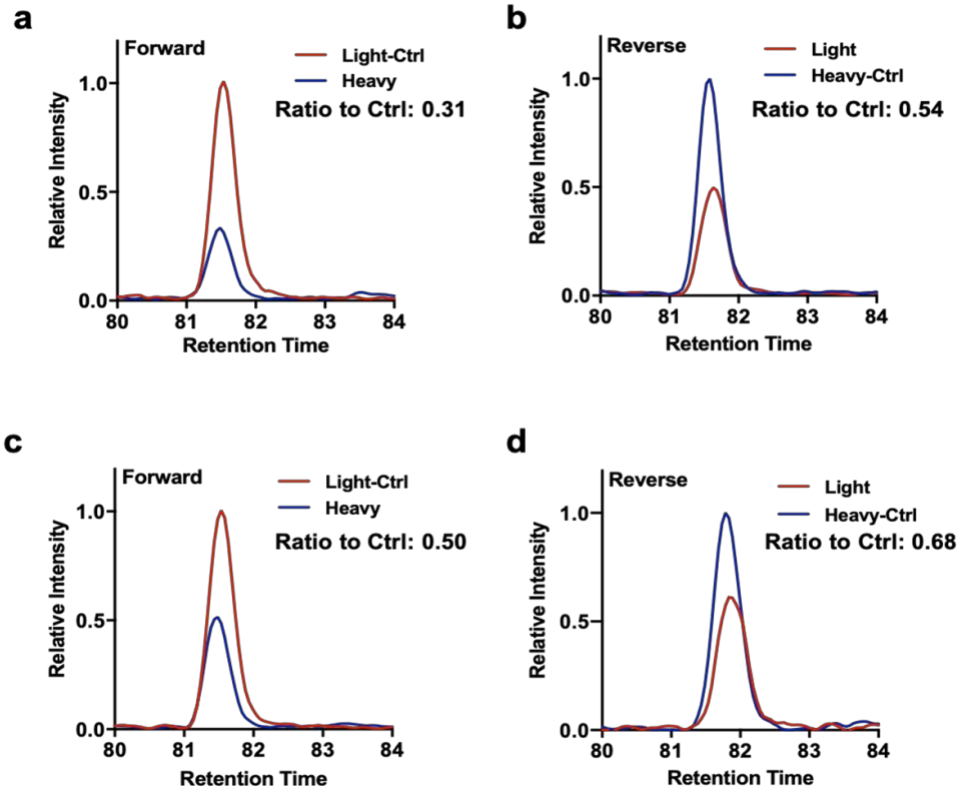


Figure 4. 6 MRM traces for probe-labeled peptide from GSK3α obtained from N^6-modified ATP analog concentration-dependent competition experiments. (a-f) Selected-ion chromatograms obtained from the competition experiments where the heavy lysate was pre-incubated with 10, 100, 200 μM N^6-Me-ATP (a-c) or KTP (d-f) prior to labeling with the heavy ATP acyl-phosphate probe, whereas the light lysate without the ATP analog was labeled with the light ATP acyl-phosphate probe. The quantification results, as reflected by labeling ratios, obtained from LC-MRM analyses are listed in each panel.

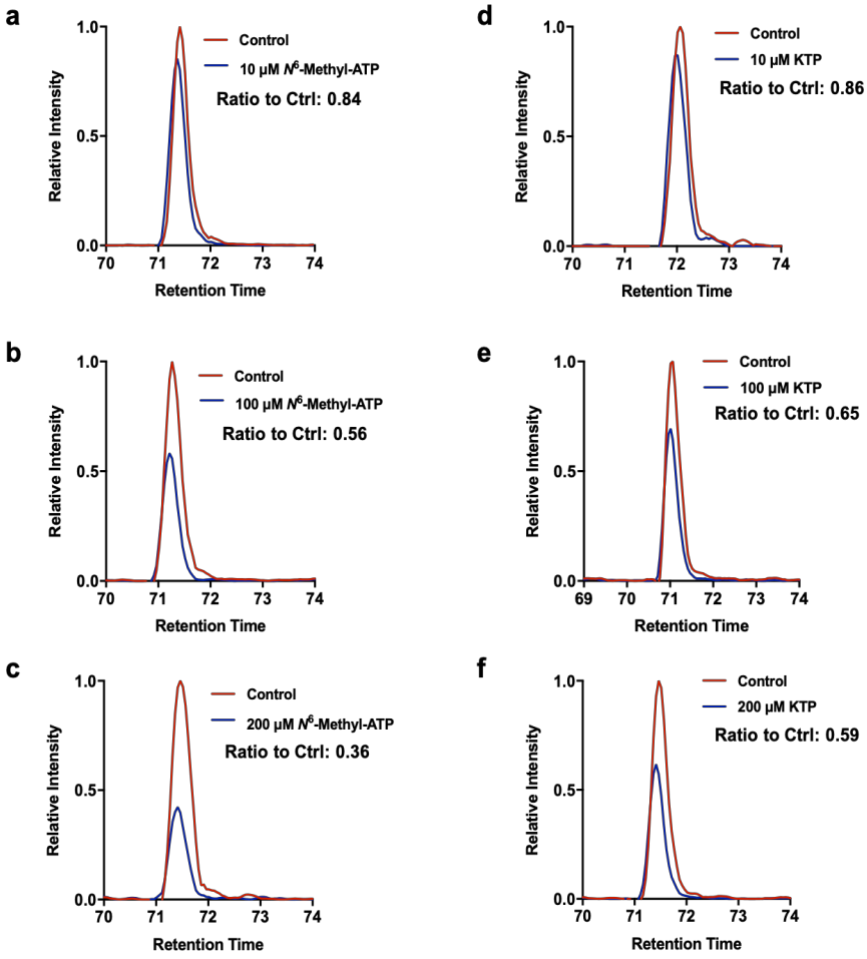


Figure 4. 7 MRM traces for probe-labeled peptide from GSK3β obtained from N^6-modified ATP analog concentration-dependent competition experiments. (a-f) Selected-ion chromatograms obtained from the competition experiments where the heavy lysate was pre-incubated with 10, 100, 200 μM N^6-Me-ATP (a-c) or KTP (d-f) prior to labeling with the heavy ATP acyl-phosphate probe, whereas the light lysate without the ATP analog was labeled with the light ATP acyl-phosphate probes. The quantification results, as reflected by labeling ratios, obtained from LC-MRM analyses are listed in each panel.

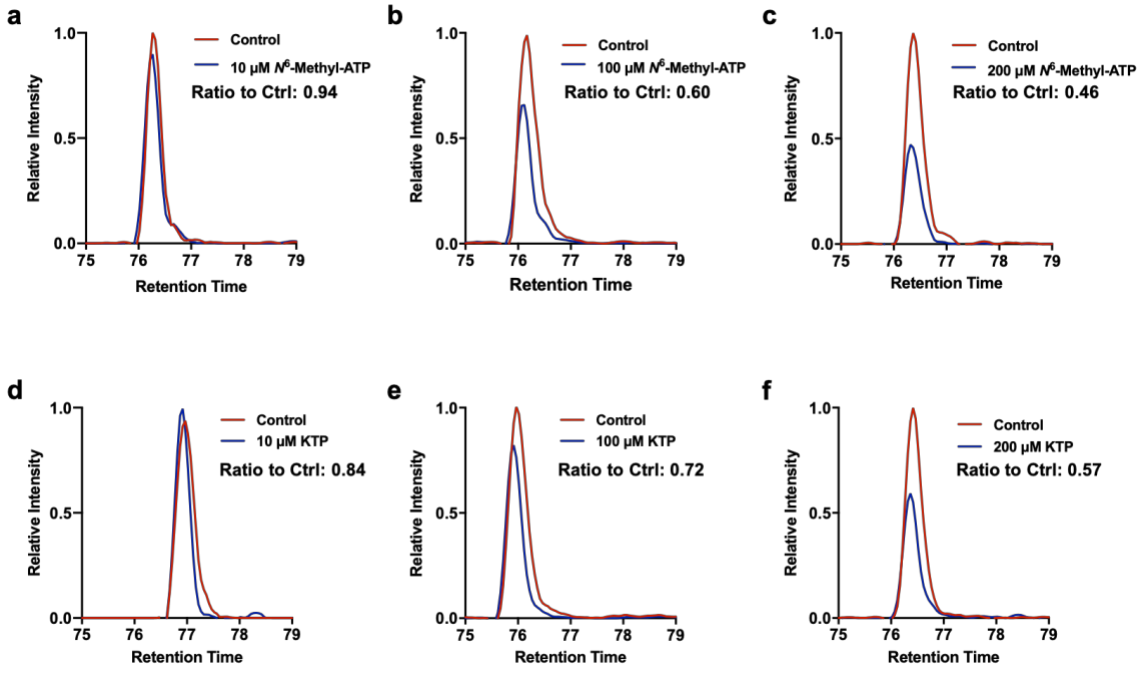


Figure 4. 8 Purification of recombinant GST-tagged GSK3β. The proteins were expressed in *E. coli* strain BL21. A lysate (L, lane 2) was made in lysis buffer and separated by centrifugation into pellet and supernatant fractions. The supernatant fraction was incubated with glutathione agarose beads and the flow through (F, lane 3), wash (lanes 4-6), and elution (lanes 7-10) fractions were collected. Samples of the fractions and markers (lane 1) were separated by SDS-PAGE and visualized by Coomassie staining.

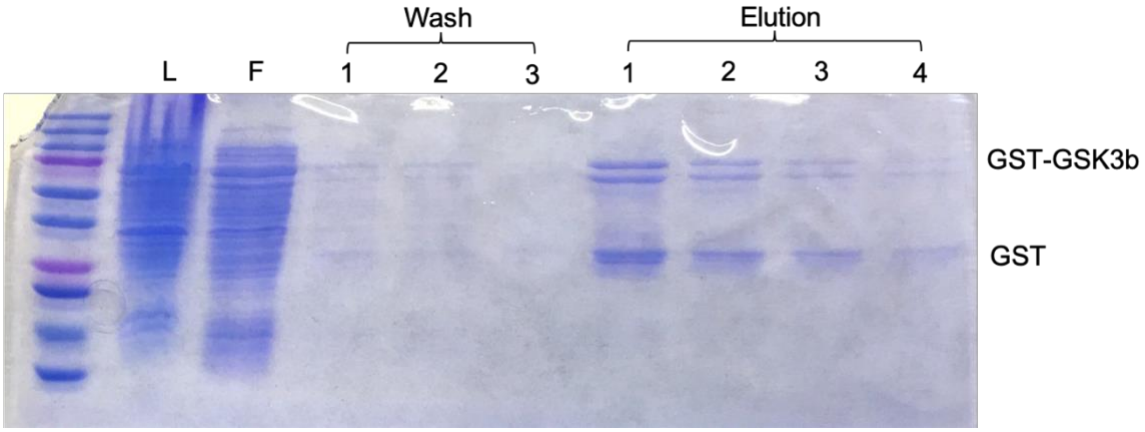


Figure 4. 9 FPLC purification of GST-GSK3β. Glutathione agarose beads purified GST-GSK3β was further purified through FPLC System, fragments have UV absorbance under 280nm at different time points were collected and labeled from F1-F4 (a) and separated by SDS-PAGE and visualized by Coomassie staining (b). Fragment F1 was used for kinase assay.

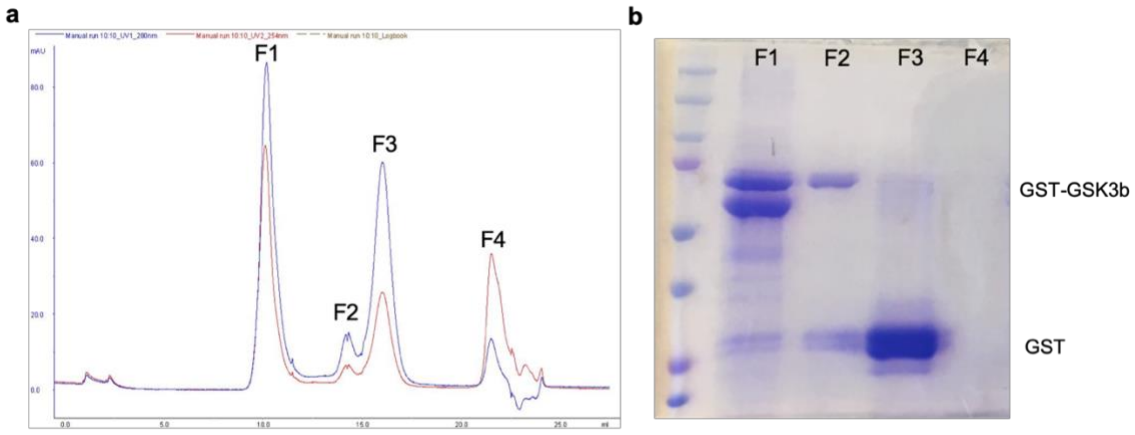


Figure 4. 10 *In vitro* kinase assay of GSK3β with a phosphopeptide substrate of the kinase. GST-GSK3β (1.0 μM) was incubated with 10 μM of its peptide substrate derived from glycogen synthase, along with 250 μM ATP, N^6-Me-ATP or KTP at 37°C for 2 h. GSK3β enzyme activity was assayed using the conversion ratio of the relative abundances of the mono-, di-, and tri-phosphorylated forms of the glycogen synthase peptide. (a-c) 'Ultra-Zoom' scan ESI-MS for the $[M+3H]^{3+}$ ions of the different phosphorylated forms of the glycogen synthase peptide from the *in vitro* kinase reactions with the use ATP (a), N^6-Me-ATP (b) or KTP (c) as phosphate group donors. (d) A comparison of relative GSK3β activities with the use of different ATP derivatives, as represented by different phosphorylated forms of its substrate peptide.

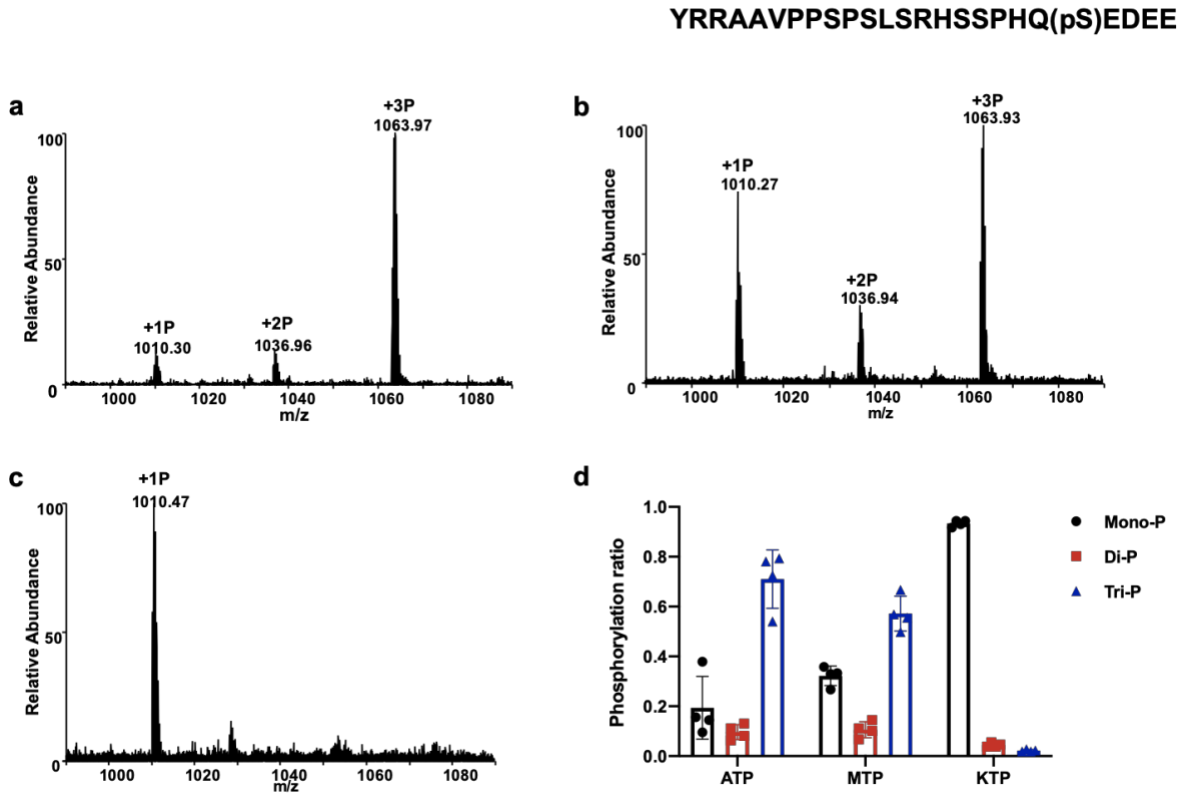


Figure 4. 11 Controls of in vitro kinase assay of GSK3β with phosphopeptide substrate. Heat-inactivated (boiled at 95°C for 5 min) GST-GSK3β (1.0 μM) was incubated with 10 μM of its peptide substrate, along with 250 μM ATP, N^6-Me-ATP or KTP at 37°C for 2 h. (a-c) 'Ultra-Zoom' scan ESI-MS for the $[M+3H]^{3+}$ ions of the different phosphorylated forms of the glycogen synthase peptide from the in vitro kinase reactions with the use ATP (a), N^6-Me-ATP (b) or KTP (c) as phosphate group donors.

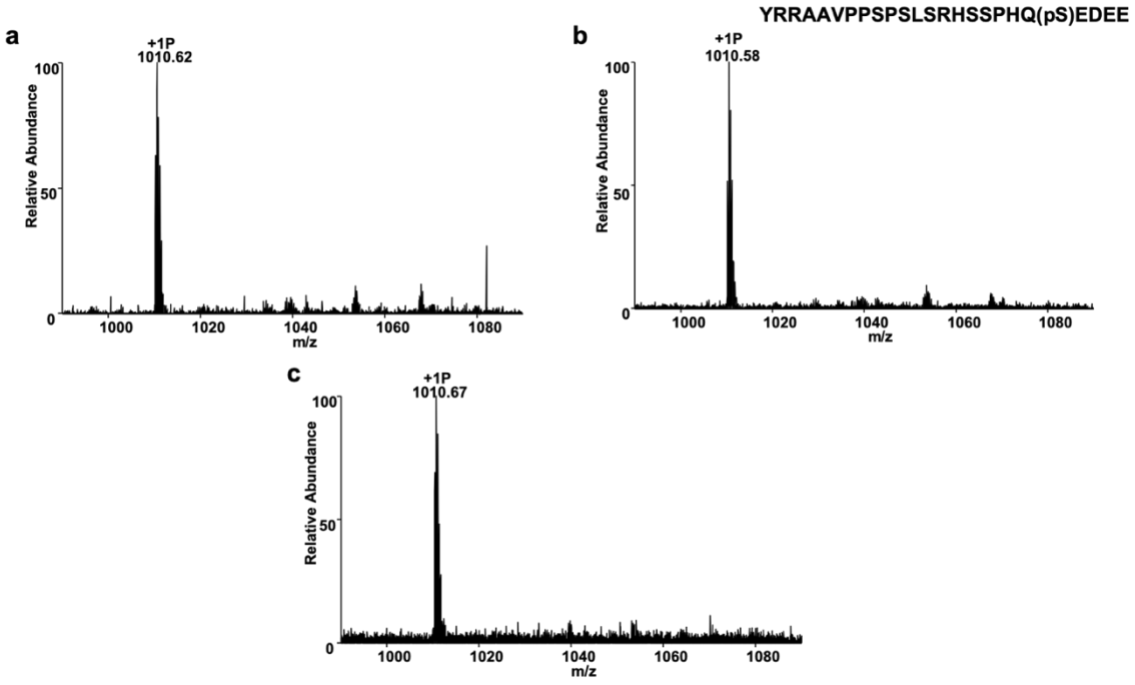


Figure 4. 12 Quantification results for the overall dynamics and conformational flexibility of GSK3β-ligand complexes. RMSF (Å) of C_α atoms over the 50 ns MD trajectory of ATP-GSK3β (red), KTP-GSK3β (black) and N^6-Me-ATP-GSK3β RMSF (blue).

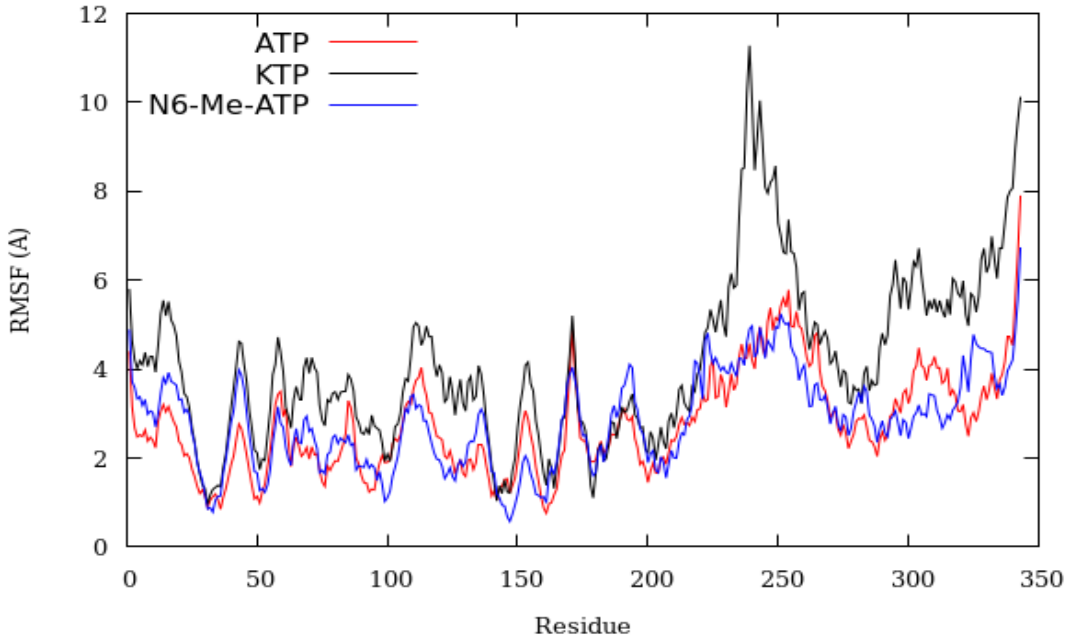


Figure 4. 13 Conformations of GSK3β-ligand complexes. (a) Superimposed GSK3β (purple) with CDK2 (grey). ATP (red) is in van der Waals representation. Two residues in GSK3β activation loop, Tyr216 (ice blue) and Cys218 (light green), are shown in Licorice representation. Activation loop in CDK2 is in cyan. (b) GSK3β (purple) with N^6-Me-ATP (blue), Tyr216 (ice blue) and Cys218 (light green). (c) GSK3β with ATP (red), Tyr216 and Cys218. (d) GSK3β with KTP (silver), Tyr216 and Cys218.

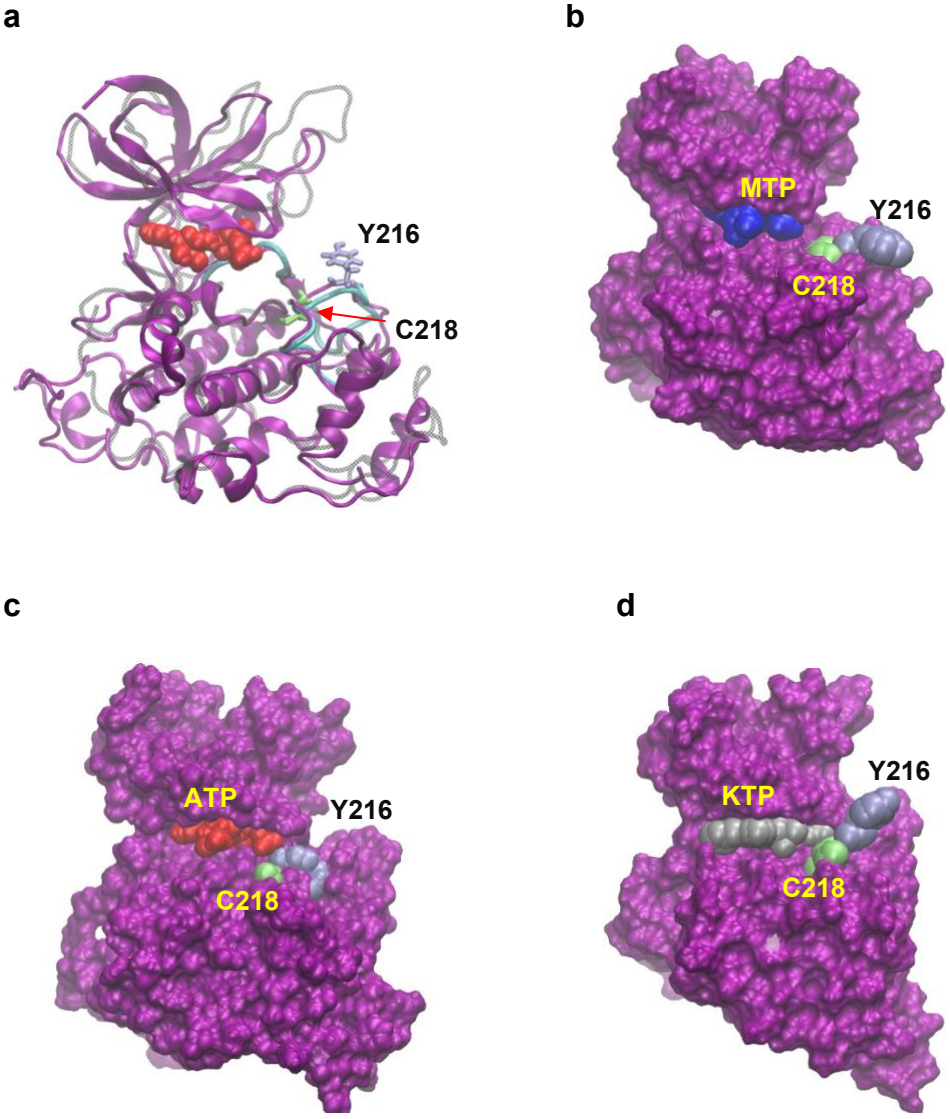


Figure 4. 14 Close-up view of interaction between ATP-GSK3β at different time steps: (a) 0 ns, (b) 15 ns, (c) 32 ns and (d) 50 ns. ATP is shown in licorice representation. Activation loop in CDK2 is colored in cyan with ribbon representation while activation loop in GSK3β is marked blue. The rest of protein is in black. K183, Y216 and C218 are from GSK3β and shown in blue line. Hydrogen bond between ATP and Y216 is shown in blue dash line and circled for better visualization

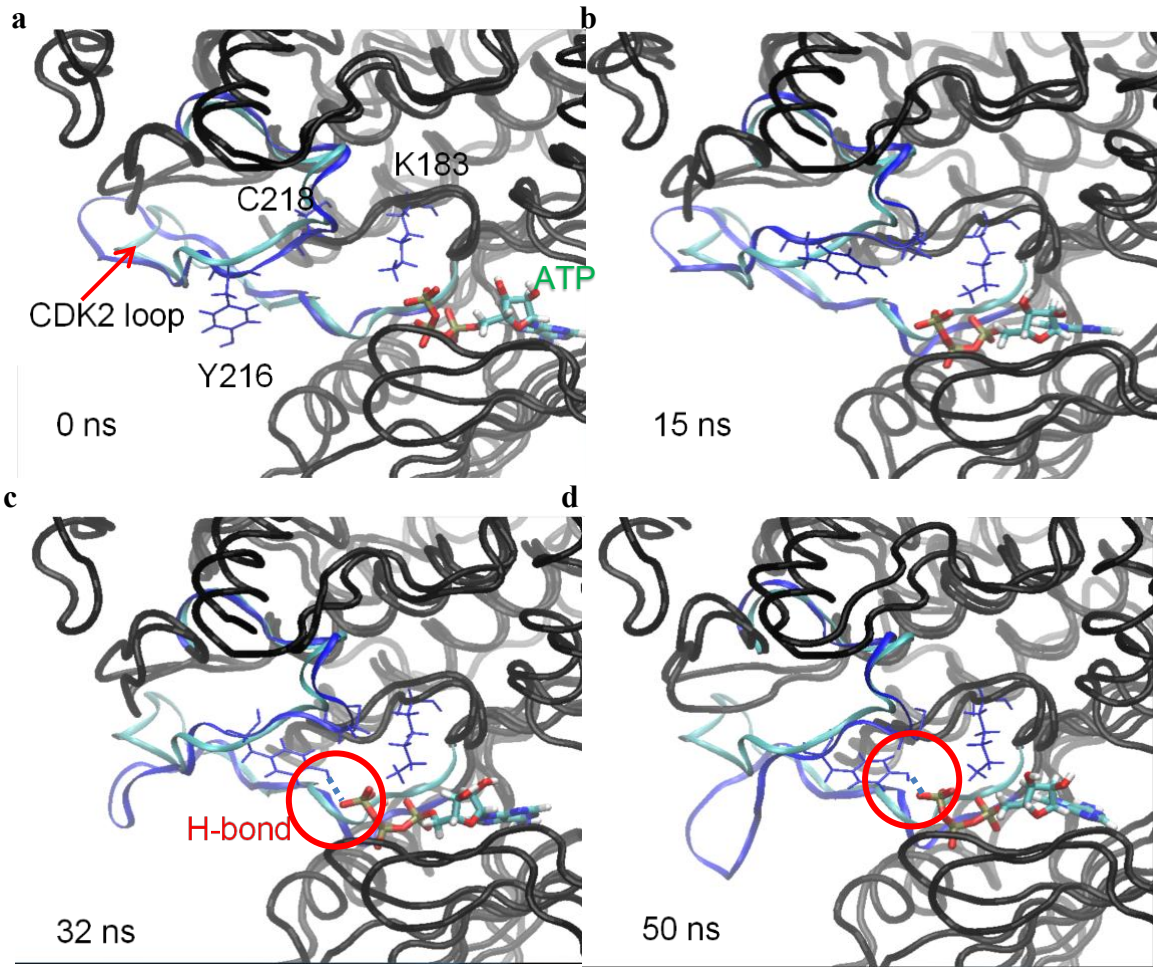


Figure 4. 15 Close-up view of interaction between KTP-GSK3β at different time steps: (a) 0 ns, (b) 15 ns, (c) 32 ns and (d) 50 ns. KTP is shown in licorice representation. Activation loop in CDK2 is colored in cyan with ribbon representation while activation loop in GSK3β is marked grey. The rest of protein is in black. K183, Y216 and C218 are from GSK3β and shown in grey line. Hydrogen bond between KTP and C218 is shown in blue dash line and circled for better visualization.

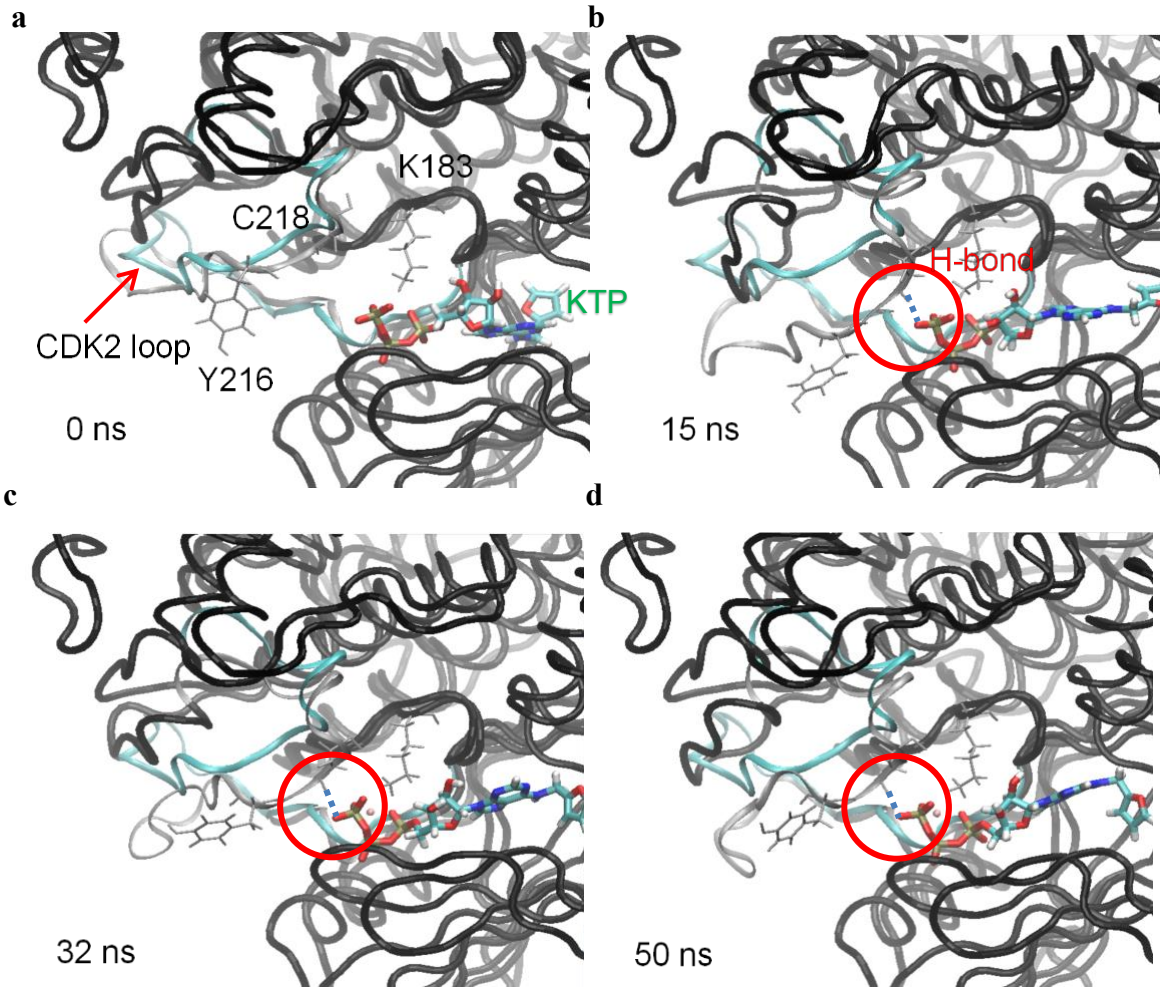
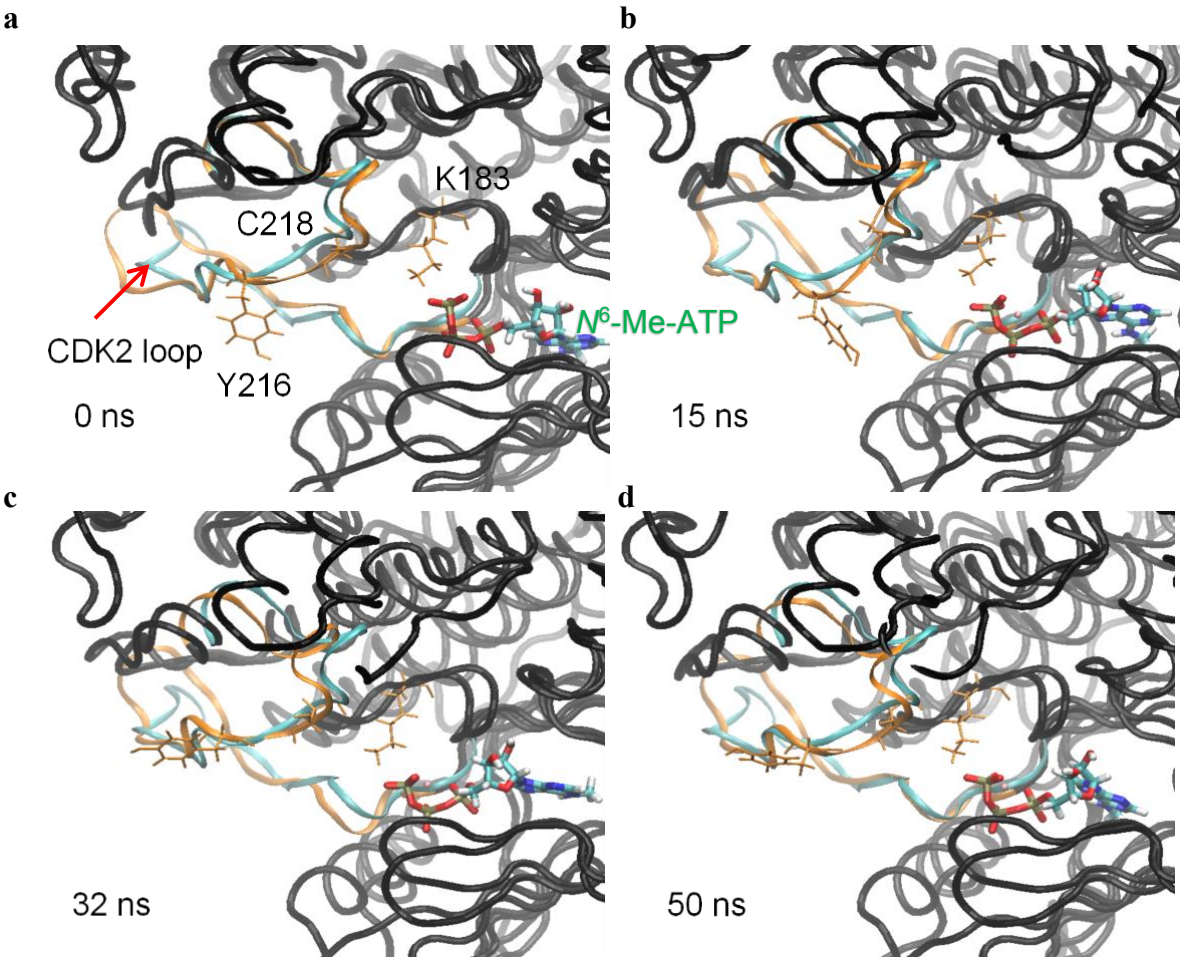


Figure 4. 16 Close-up view of interaction between N^6-Me-ATP-GSK3β at different timestep: (a) 0 ns, (b) 15 ns, (c) 32 ns and (d) 50 ns. N^6-Me-ATP is shown in licorice representation. Activation loop in CDK2 is colored in cyan with ribbon representation while activation loop in GSK3β is marked orange. The rest of protein is in black. K183, Y216 and C218 are from GSK3β and shown in orange line. No hydrogen bond observed between N^6-Me-ATP and Y216/C218.



References

1. Rauch, J.; Volinsky, N.; Romano, D.; Kolch, W., The secret life of kinases: Functions beyond catalysis. In *Cell Communication and Signaling*, BioMed Central: 2011; Vol. 9, pp 1-28.

2. Roskoski, R., A historical overview of protein kinases and their targeted small molecule inhibitors. In *Pharmacol. Res.*, Academic Press: 2015; Vol. 100, pp 1-23.

3. Wilson, L. J.; Linley, A.; Hammond, D. E.; Hood, F. E.; Coulson, J. M.; MacEwan, D. J.; Ross, S. J.; Slupsky, J. R.; Smith, P. D.; Eyers, P. A.; Prior, I. A., New Perspectives, opportunities, and challenges in exploring the human protein kinome. In *Cancer Res.*, American Association for Cancer Research Inc.: 2018; Vol. 78, pp 15-29.

4. Blume-Jensen, P.; Hunter, T., Oncogenic kinase signalling. In *Nature*, Nature Publishing Group: 2001; Vol. 411, pp 355-365.

5. Lemmon, M. A.; Schlessinger, J., Cell signaling by receptor tyrosine kinases. In *Cell*, Elsevier B.V.: 2010; Vol. 141, pp 1117-1134.

6. Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Muzio, L. L., The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review). In *Int. J. Mol. Med.*, Spandidos Publications: 2017; Vol. 40, pp 271-280.

7. Newman, R. H.; Hu, J.; Rho, H.-S.; Xie, Z.; Woodard, C.; Neiswinger, J.; Cooper, C.; Shirley, M.; Clark, H. M.; Hu, S.; Hwang, W.; Jeong, J. S.; Wu, G.; Lin, J.; Gao, X.; Ni, Q.; Goel, R.; Xia, S.; Ji, H.; Dalby, K. N.; Birnbaum, M. J.; Cole, P. A.; Knapp, S.; Ryazanov, A. G.; Zack, D. J.; Desiderio, S.; Pandey, A.; Turk, B. E.; Zhang, J.; Zhu, H.; Qian, J.; Zhang, J.; Zhu, H.; Qian, J., Construction of human activity-based phosphorylation networks. *Mol. Syst. Biol.* **2013,** *5*, 655.

8. Bodenmiller, B.; Aebersold, R., Phosphoproteome resource for systems biology research. *Methods in molecular biology (Clifton, N.J.)* **2011,** *694*, 307-322.

9. Prakash Damle, N.; Mohanty, D., Systems biology Deciphering kinase-substrate relationships by analysis of domain-specific phosphorylation network. **2014,** *30*, 1730-1738.

10. Allen, J. J.; Li, M.; Brinkworth, C. S.; Paulson, J. L.; Wang, D.; Hübner, A.; Chou, W.-H.; Davis, R. J.; Burlingame, A. L.; Messing, R. O.; Katayama, C. D.; Hedrick, S. M.; Shokat, K. M., A semisynthetic epitope for kinase substrates. *Nat. Methods* **2007,** *4*, 511-6.

11. Hertz, N. T.; Berthet, A.; Sos, M. L.; Thorn, K. S.; Burlingame, A. L.; Nakamura, K.; Shokat, K. M., A neo-substrate that amplifies catalytic activity of parkinson's-disease-related kinase PINK1. *Cell* **2013,** *154*, 737-747.

12. Picotti, P.; Aebersold, R., Selected reaction monitoring-based proteomics: Workflows, potential, pitfalls and future directions. In *Nat. Methods*, Nature Publishing Group: 2012; Vol. 9, pp 555-566.

13. Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J., Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* **2012,** *11*, 1475-1488.

14. Miao, W.; Xiao, Y.; Guo, L.; Jiang, X.; Huang, M.; Wang, Y., A High-Throughput Targeted Proteomic Approach for Comprehensive Profiling of Methylglyoxal-Induced Perturbations of the Human Kinome. *Anal. Chem.* **2016,** *88*, 9773-9779.

15. Xiao, Y.; Guo, L.; Wang, Y., Isotope-coded ATP probe for quantitative affinity profiling of ATP-binding proteins. *Anal. Chem.* **2013,** *85*, 7478-7486.

16. Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat. Methods* **2009,** *6* (5), 359-362.

17. Maclean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; Maccoss, M. J., Gene expression Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *BIOINFORMATICS APPLICATIONS NOTE* **2010,** *26*, 966-968.

18. Sherwood, C. A.; Eastham, A.; Lee, L. W.; Risler, J.; Vitek, O.; Martin, D. B., Correlation between y-type ions observed in ion trap and triple quadrupole mass spectrometers. *J. Proteome Res.* **2009,** *8*, 4243-4251.

19. Harper, S.; Speicher, D. W., Purification of proteins fused to glutathione S-transferase. *Methods in molecular biology (Clifton, N.J.)* **2011,** *681*, 259-280.

20. Tapia, J. A.; García-Marin, L. J.; Jensen, R. T., Cholecystokinin-stimulated protein kinase C-δ kinase activation, tyrosine phosphorylation, and translocation are mediated by Src tyrosine kinases in pancreatic acinar cells. *J. Biol. Chem.* **2003,** *278*, 35220-35230.

21. Tullai, J. W.; Chen, J.; Schaffer, M. E.; Kamenetsky, E.; Kasif, S.; Cooper, G. M., Glycogen synthase kinase-3 represses cyclic AMP Response element-binding protein (CREB)-targeted Immediate early genes in quiescent cells. *J. Biol. Chem.* **2007,** *282*, 9482-9491.

22. Bertrand, J. A.; Thieffine, S.; Vulpetti, A.; Cristiani, C.; Valsasina, B.; Knapp, S.; Kalisz, H. M.; Flocco, M., Structural characterization of the GSK-3β active site using selective and non-selective ATP-mimetic inhibitors. *J. Mol. Biol.* **2003,** *333*, 393-407.

23. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. In *J. Comput. Chem.*, J Comput Chem: 2005; Vol. 26, pp 1668-1688.

24. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.

25. Li, P.; Song, L. F.; Merz, K. M., Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water. *J. Phys. Chem. B* **2015,** *119*, 883-895.

26. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002,** *23*, 1623-1641.

27. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996,** *100*, 19824-19839.

28. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995,** *246*, 122-129.

29. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *JCoPh* **1977,** *23*, 327-341.

30. Escher, C.; Reiter, L.; MacLean, B.; Ossola, R.; Herzog, F.; Chilton, J.; MacCoss, M. J.; Rinner, O., Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **2012,** *12*, 1111-1121.

31. Hanger, D. P.; Hughes, K.; Woodgett, J. R.; Brion, J. P.; Anderton, B. H., Glycogen synthase kinase-3 induces Alzheimer's disease-like phosphorylation of tau: Generation of paired helical filament epitopes and neuronal localisation of the kinase. *Neurosci. Lett.* **1992,** *147*, 58-62.

32. Mandelkow, E.-M.; Drew&, G.; Biernat, J.; Custke, N.; Lintb, J. V.; Vandenheede, J. R.; Mandelkow, E., Glycogen synthase kinase-3 and the Alzheimer-like state of microtubule-associated protein tau. *European Biochemical Societies* **1992,** *314*, 315-321.

33. Stambolic, V.; Ruel, L.; Woodgett, J. R., Lithium inhibits glycogen synthase kinase-3 activity and mimics wingless signalling in intact cells. *Curr. Biol.* **1996,** *6*, 1664-1669.

34. Klein, P. S.; Melton, D. A., A molecular mechanism for the effect of lithium on development. *Proceedings of the National Academy of Sciences* **1996,** *93* (16), 8455-8459.

35. Machado-Vieira, R.; Manji, H. K.; Zarate, C. A., The role of lithium in the treatment of bipolar disorder: Convergent evidence for neurotrophic effects as a unifying hypothesis. In *Bipolar Disorders*, NIH Public Access: 2009; Vol. 11, pp 92-109.

36. Wang, X.; Xie, J.; Proud, C. G., Eukaryotic elongation factor 2 kinase (eEF2K) in cancer. In *Cancers (Basel)*, MDPI AG: 2017; Vol. 9, p 162.

37. Bowie, L. E.; Maiuri, T.; Alpaugh, M.; Gabriel, M.; Arbez, N.; Galleguillos, D.; Hung, C. L. K.; Patel, S.; Xia, J.; Hertz, N. T.; Ross, C. A.; Litchfield, D. W.; Sipione, S.; Truant, R., N6-Furfuryladenine is protective in Huntington's disease models by signaling huntingtin phosphorylation. *Proceedings of the National Academy of Sciences* **2018,** *115* (30), E7081-E7090.

38. Beurel, E.; Grieco, S. F.; Jope, R. S., Glycogen synthase kinase-3 (GSK3): Regulation, actions, and diseases. *Pharmacol. Ther.* **2015,** *148*, 114-131.

39. Brown, N. R.; Noble, M. E. M.; Endicott, J. A.; Johnson, L. N., The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nat. Cell Biol.* **1999,** *1*, 438-443.

40. Hughes, K.; Nikolakaki, E.; Plyte, S. E.; Totty, N. F.; Woodgett, J. R., Modulation of the glycogen synthase kinase-3 family by tyrosine phosphorylation. *EMBO J.* **1993,** *12*, 803-808.

41. HU;, Y.; CONWAY, T. W., 2-Aminopurine Inhibits the Double-Stranded RNA-Dependent Protein Kinase Both In Vitro and In Vivo. *J. Interferon Res.* **1993,** *13* (5), 323-328.

42. Mok;, D. W.; Mok, M. C., CYTOKININ METABOLISM AND ACTION. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **2001,** *52*, 89-118.

Chapter 5 Concluding Remarks

In this dissertation, we developed and utilized mass spectrometry-based chemical proteomic approaches for identification and quantitation of biological important small molecule binding proteins in HEK293T cells. By employing data dependent acquiring(DDA), multiple-reaction monitoring (MRM) and two different labeling strategies, i.e. stable isotope labeling by amino acids in cell culture (SILAC) and isotope-coded affinity-tag-based protein profiling(ICAT), we have successfully designed relative chemical probes and achieved high-through put profiling of novel specific ligand-binding proteins by assessing the altered level of pulled-down proteins in different biological systems or upon multiple treatments.

The first small molecule we studied is lysophosphatidic acid (LPA), it is a ubiquitous phospholipid that plays critical biological roles in various cellular processes. The emphasis of our study in Chapter 2 was placed on the development of a chemical proteomics approach for the proteome-wide identification of putative LPA-binding proteins using affinity-based proteome profiling. We designed and synthesized a series of desthiobiotin-tagged affinity-based LPA probes, and performed large-scale proteomic experiments. We proposed that the probe could provide potential LPA-binding proteins a desthiobiotin tag after the probes reacted with the lysine residue close to the LPA binding site, allows for subsequent affinity purification and MS identification. The probes are effective, in combination with LC-MS/MS analysis, for the proteome-wide identification of LPA-binding proteins. Over 1000 proteins with desthiobiotin modification were identified in

different cell lines through our experiments, and several proteins were previously reported to be LPA-interacting proteins. We further adapted the workflow to combined with SILAC by allowing a low and high concentration of the LPA probe comparation labeling experiment, to derive the peak intensity ratios of light and heavy desthiobiotin-labeled peptides which reflect the relative binding affinities of LPA toward specific lysine residues in individual proteins. Several putative LPA-binding proteins, such as PGK1 and ANXA5 were also identified with the validation of isothermal titration calorimetry(ITC). Additionally, based on the MS/MS, we were also able to predict the LPA binding sites in these proteins which formed the basis for the future characterizations of these proteins in LPA signaling.

In Chapter 3, we focused on arsenic as it is a wide spread environmental contaminant and long-term exposure to arsenic is known to be associated with the development of many human diseases. Previous studies suggest that trivalent arsenic, As(III), interacts preferentially with proteins harboring solvent-accessible and closely placed cysteine residues. Thus, we employed LC-MS/MS-based chemoproteomic method combined with SILAC to identify quantitatively As(III)-binding proteins. By quantifying peak intensity ratios of light- and heavy-labeled tryptic peptides, the selectivity of arsenic-binding proteins was systematically investigated, around 50 proteins exhibited preference in binding to As(III). Moreover, the validation by employing Western blot analysis revealed HSPA4, a member of heat shock protein 70 (HSP70) family, has the ability in binding to As(III) *in vitro*. Structural analysis further indicated that As(III) may target the cysteine residues located near the nucleotide-binding pocket of HSPA4. This method can provides

mechanistic insights into mechanistic studies of mechanisms underlying the toxic effects of arsenic species.

ATP and kinases are crucial elements of multiple cell signaling pathways. A chemical genetic approach, relying on synthetic ATP analogs and genetic manipulations of kinases that can accommodate these analogs, has been employed to identify direct substrates of protein kinases. It, however, has not been systematically analyzed whether these analog substrates can also be recognized by endogenous kinases. Curious about the specificity of the ATP analog binding protein and aiming to find more possible target kinases for better understanding the complicated signaling network which regulates kinase activity. In Chapter 4, We explored the analytical performance of scheduled LC-MRM analysis in conjunction with a chemical labeling strategy, which relies on the use of isotope-labeled desthiobiotin-ATP affinity probes. By calculation. Based on our lab previously generated kinome MRM library which include over 80% of human kinases, we were able to predict the RT of targe peptides and built relative MRM method for proteome-wide identification of endogenous kinases that can bind to two unnatural N^6-modified ATP derivatives, N^6-methyl-ATP (N^6-Me-ATP) and N^6-furfuryl-ATP (KTP). By integrating the peak area ratios of light- and heavy-ATP probe tagged target peptides in w/wo the ATP analogs competition conditions, we were able to assess quantitatively ~300 kinases, among which 28 and 18 are candidate kinases that can bind to KTP and N^6-Me-ATP, respectively. *In vitro* kinase activity assay further validated the top target we identified by LC-MRM, GSK3β, can accept the N^6-Me-ATP and remain its enzymatic activity to phosphorylate relative substrate.

Moreover, molecular modeling studies of these kinases provided insights about how selected kinases recognize these ATP analogs.

Together, by taking advantage of the MS bottom-up proteomic analysis (DDA and MRM) and different isotope labeling approaches (metabolic labeling and chemical labeling) , high-throughput discovery and targeted quantitative profiling of small molecule binding protein is enabled. The work in this dissertation presented novel and systematic chemoproteomic approaches to investigate the previously unrecognized roles of LPA, As(III) and ATP analogs in a wide array of biological events through the discovery of their novel binding proteins, which may shed the light on the mechanism study for understanding cell signaling pathways and cellular response upon exposure to arsenic. The methods described in this dissertation can be easily extended to the study the binding protein of other small molecules. MS-based proteomics strategies have the advantage of high through put and relatively high sensitivity, reproducibility and specificity; however, proteomic technique, which is highly dependent on sample preparation and instrument optimization and more susceptible to background interference resulting from sample matrix, at the same time can include a numerous false-positive results. Thus, the incorporation with precise validation methods is highly encouraged.